Evolution of non-recombining region in mating-type chromosome from the fungal genus Microbotryum

Fantin Carpentier

▶ To cite this version:

HAL Id: tel-02406590
https://theses.hal.science/tel-02406590v2

Submitted on 18 Dec 2019

Thèse de doctorat

Evolution des régions non-recombinantes sur les chromosomes de types sexuels chez les champignons du genre *Microbotryum*

Thèse de doctorat de l'Université Paris-Saclay
préparée à l'Université Paris-Sud

Thèse présentée et soutenue
à Orsay, le 19 Novembre 2019, par

Fantin Carpentier

Composition du Jury :

Mme Cécile Fairhead
Professeure Université Paris-Sud, laboratoire GQE   Présidente du Jury

Mme Susana Coelho
DR CNRS, Station Biologique de Roscoff   Rapportrice

M. Alex Widmer
Professeur, Institute of Integrative Biology (ETH Zürich)   Rapporteur (absent)

M. Bart Nieuwenhuis
Associate Professor, Faculty of Biology II (LMU Münich)   Examinateur

Mme Tatiana Giraud
DR CNRS, laboratoire ESE   Directrice de thèse

Acknowledgments

I want first to thank the members of my jury, Susana Coelho and Alex Widmer to have accepted to be referees, Cécile Fairhead and Bart Niewenhuis to have accepted to be examiners, and Tatiana Giraud my supervisor.

Many thanks to all the people with whom I share authorship for the studies presented in this thesis: Sara Branco, Ricardo Rodríguez de la Vega, Hélène Badouin, Jérôme Gouzy, Gabriela Aguileta, Sophie Siguenza, Jean-Tristan Brandenburg, Marco A. Coelho, Michael E. Hood, Alodie Snirc, Stéphanie Le Prieur, Damien M. de Vienne, Fanny E. Hartmann, Dominik Begerow, Anna Bazzicalupo and Sarah P. Otto. It has been a great pleasure to work with them to achieve such nice and thorough studies. I specially thank Michael Hood for having welcome me so warmly in Amherst, and Alodie Snirc for all the support for the lab work.

I kindly thank the members of my PhD committee, Aurélie Hua-Van, Xavier Vekemans, Nicolas Perrin and Gabriel Marais to have brought valuable insights to my work and for their attentive listening.

I warmly thank Tatiana Giraud. More than having supervised my work and taught me a great deal about writing, thinking, working and science (that is, A LOT), Tatiana has been so impressive, present, attentive and comforting to me. Thousand words could not tell how grateful I am.

Many thanks to Sara Branco, who took me under her wing the first day I arrived here and warned me with such kindliness about everything I was about to experience. Every day of this thesis, I remembered advices she gave me, and truly understood them little by little until now.

Many thanks to Ricardo Rodríguez de la Vega for the great and numerous scientific discussions we have, for had been such a work inspiration at so many levels, and for his everlasting smile, joy and growing thinking.

v

Résumé de la thèse

La reproduction sexuée est très répandue chez les eucaryotes, mais son succès a été et reste un mystère évolutif. À long terme, la recombinaison est plus avantageuse que la clonalité, car elle peut réunir des combinaisons d'allèles bénéfiques et permettre une sélection efficace qui empêche des mutations délétères de s'accumuler par interférences Hill-Robertson et par cliquet de Muller. Cependant, à court terme, la recombinaison peut être désavantageuse car elle brise les combinaisons bénéfiques d'allèles. La suppression de la recombinaison peut donc être sélectionnée pour préserver les combinaisons alléliques bénéfiques, qui sont ensuite transmises comme un seul locus et constituent un supergène. Les chromosomes sexuels sont considérés comme portant des supergènes, avec une suppression de recombinaison reliant les gènes déterminant le sexe. Cependant, dans de nombreux cas, la suppression de la recombinaison n'est pas limitée au locus déterminant le sexe et peut s'étendre sur presque toute la longueur des chromosomes sexuels en raison d'une extension progressive de la suppression de la recombinaison. Une théorie largement acceptée postule que la sélection a favorisé le lien entre des gènes sexuellement antagonistes et les gènes déterminant le sexe, générant des strates évolutives. Cependant, cette théorie manque d'un solide soutien expérimental. D'autre part, les grandes régions non recombinantes peuvent subir une dégénérescence génomique, avec l'accumulation de mutations délétères dues à une sélection moins efficace.

Chez les champignons basidiomycètes, certaines espèces portent des chromosomes de types sexuels présentant des caractéristiques similaires à celles des chromosomes sexuels, avec de grandes régions non-recombinantes qui subissent une dégénérescence génomique. La suppression de recombinaison sur ces chromosomes lie les locus de types sexuels et a évolué depuis un système dans lequel les locus de types sexuels sont non-liés et situés sur deux chromosomes distincts et recombinants. Avec un taux d'autofécondation élevé, la liaison des locus de types sexuels est avantageuse car elle augmente les chances de compatibilité des gamètes : lorsque les locus de types sexuels sont liés, deux types sexuels sont produits à l'issu d'un évènement de méiose contre quatre types sexuels lorsque les locus sont non-liés. La forte sélection en faveur de la liaison des locus de type sexuels aurait pu conduire l'évolution indépendante d'une telle liaison. L'étude de l'évolution du système de compatibilité des gamètes pourrait donc révéler des cas intéressants de convergence évolutive. Chez les champignons, la reproduction se produit entre gamètes portant des types sexuels différents, qui ne sont pas

associés à des fonctions mâles ou femelles. Peu de différences phénotypiques sont associées aux types sexuels, ce qui rend peu probable l'existence d'une sélection antagoniste entre les types sexuel (c'est-à-dire avec des gènes ayant des allèles bénéfiques à un type sexuel mais délétères à l'autre). Par conséquent, l'existence de strates évolutives sur les chromosomes de types sexuels impliquerait que des mécanismes évolutifs alternatifs à la sélection antagoniste doivent être considérés, et pourraient aussi être à l'œuvre sur les chromosomes sexuels. Cependant, il reste à déterminer si de fines différences peuvent être associées aux types sexuels, telles que des différences dans le niveau d'expression des gènes, ou un enrichissement de gènes différentiellement exprimés entre les types sexuels, ce qui indiquerait l'existence d'une sélection antagoniste entre types sexuels entraînant une suppression de recombinaison. Cependant, l'expression différentielle peut aussi être due à la dégénérescence dans les régions génomiques avec une ancienne suppression de recombinaison. Un moyen de vérifier si la suppression de recombinaison peut être induite par une sélection antagoniste entre types sexuels pourrait donc être d'évaluer si les gènes ayant une expression différentielle entre les types sexuels sont enrichis dans de jeunes strates évolutives.

Les chromosomes de types sexuels du champignon basidiomycète *Microbotryum lychnidis-dioicae* sont dimorphes, avec une grande région non-recombinante couvrant 90 % de la longueur des chromosomes. De plus, il a été démontré que les chromosomes de types sexuels de plusieurs espèces de *Microbotryum* ont subi une dégénérescence génomique sous la forme d'accumulation de substitutions non-synonymes ou d'accumulation de séquences répétées, comme dans les chromosomes sexuels. Enfin, un modèle théorique avait suggéré que le système de reproduction particulier des champignons *Microbotryum* pourrait promouvoir une extension progressive de la suppression de la recombinaison qui maintiendrait les allèles délétères récessifs à l'état hétérozygote. Le complexe d'espèces *Microbotryum* semblait donc être un bon modèle pour tester des hypothèses sur la formation des strates évolutives ainsi que les prédictions issues de la théorie de l'évolution des chromosomes sexuels.

Les travaux présentés dans cette thèse avaient pour but de répondre aux questions suivantes :

I.   Les strates évolutives peuvent-elles se former dans des chromosomes de type sexuel dans des organismes sans fonction mâle/femelle ?

II.   La sélection antagoniste entre types sexuels a-t-elle contribué à la formation de ces strates évolutives ?

III.   La liaison des locus de type sexuel a-t-elle évolué plusieurs fois indépendamment depuis un système sans liaison chez les espèces à taux d'autofécondation élevés ?

IV.   L'expression différentielle des gènes dans les régions non-recombinantes résulte-t-elle d'une dégénérescence ou d'une sélection antagoniste entre types sexuels ?

En plus d'aborder les quatre questions présentées ci-dessus sous la forme de cinq articles (dont quatre publiés), cette thèse présente en annexe deux documents : le premier est une étude de génomique des populations menée par Fanny Hartmann (post-doctorante dans l'équipe GEE) ; le second est une revue dirigée par Fanny Hartmann et Tatiana Giraud concernant les approches et les résultats obtenus sur l'adaptation, la coévolution et l'évolution du système de reproduction des champignons *Microbotryum* en utilisant des données génomiques.

Résumés des cinq articles correspondant aux quatre questions traitées dans le corps de la thèse

I.   Les strates évolutives peuvent-elles se former dans des chromosomes de type sexuel dans des organismes sans fonction mâle/femelle ?

Branco, S., Badouin, H., de la Vega, R. C. R., Gouzy, J., **Carpentier, F.**, Aguileta, G., Siguenza, S., Brandenburg, J. T., Coelho, M. A., Hood, M. E. & Giraud, T. (2017). Evolutionary strata on young mating-type chromosomes despite the lack of sexual antagonism. Proceedings of the National Academy of Sciences, 114(27), 7067-7072.

Les chromosomes sexuels peuvent présenter des étapes successives de suppression de recombinaison connues sous le nom de "strates évolutives", qui sont censées résulter de la liaison successive de gènes sexuellement antagonistes aux gènes déterminant le sexe. Toutefois, il y a peu de preuves à l'appui de cette théorie. Nous étudions ici si les strates évolutives peuvent évoluer sans antagonisme sexuel en utilisant des champignons qui présentent une suppression de recombinaison s'étendant au-delà des locus déterminant la compatibilité sexuelle malgré l'absence de rôles mâle/femelle associés aux types sexuels. En comparant les assemblages de chromosomes entiers de cinq champignons, causant la maladie du charbon des anthères, avec ou sans suppression de recombinaison entre leurs chromosomes de types sexuels, nous avons déduit l'ordre ancestral des gènes et les réarrangements chromosomiques dérivés dans ce groupe. Cette approche a mis en lumière la fusion chromosomique qui a permis la liaison des locus de types sexuels et a fourni des preuves de la formation de plusieurs strates évolutives à des temps différents (0,9-2,1 million d'années) dans les chromosomes de types sexuels. Plusieurs strates évolutives n'impliquaient pas les gènes déterminant la compatibilité sexuelle. L'existence de strates dépourvues de gènes de types sexuels, malgré l'absence d'antagonisme sexuel, appelle à une théorie unifiée de l'évolution des chromosomes liés au sexe, qui intégrerait, par exemple, l'influence de mutations délétères partiellement liées à la région sans recombinaison ou le maintien d'un polymorphisme de réarrangements neutres dû à une sélection équilibrante des sexes et des types sexuels.

II. La sélection antagoniste entre types sexuels a-t-elle contribué à la formation de ces strates évolutives ?

Bazzicalupo, A. L., **Carpentier, F.**, Otto, S. P., & Giraud, T. (2019). Little evidence of antagonistic selection in the evolutionary strata of fungal mating-type chromosomes (*Microbotryum lychnidis-dioicae*). G3: Genes, Genomes, Genetics, 9(6), 1987-1998.

La suppression de la recombinaison sur les chromosomes sexuels s'étend souvent de manière progressive, générant des strates évolutives de différenciation entre les chromosomes sexuels. L'antagonisme sexuel est une explication largement acceptée des strates évolutives, postulant que plusieurs gènes avec des allèles bénéfiques à un seul sexe sont successivement liés au locus déterminant le sexe. Le champignon causant la maladie du charbon des anthères, *Microbotryum lychnidis-dioicae*, possède des chromosomes de types sexuels avec des strates évolutives, dont certaines seulement impliquent les locus de types sexuels. Les rôles masculins et féminins sont inexistants dans ce champignon, mais une sélection antagoniste entre types sexuels pourrait aussi générer des strates évolutives, bien que le cycle biologique du champignon suggère qu'il devrait être limité à quelques caractères phénotypiques. Dans cette étude, nous avons testé l'hypothèse selon laquelle l'antagonisme entre types sexuels aurait pu induire une suppression de la recombinaison au-delà des gènes de types sexuels chez *M. lychnidis-dioicae* en recherchant les empreintes de la sélection antagoniste dans les strates évolutives qui n'impliquent pas les locus de types sexuels. Nous avons constaté que ces strates évolutives (i) n'étaient pas enrichies en gènes surexprimés en phase haploïde (en comparaison avec la phase dicaryotique) où les cellules sont de types sexuels différents, (ii) ne portaient aucun gène différentiellement exprimé entre les types sexuels, et (iii) ne portaient aucun gène présentant des empreintes de spécialisation en termes de séquences de protéines (dN/dS) entre types sexuels, après avoir filtré les données. Sans filtrage, onze gènes ont montré des signes de sélection positive dans les strates ne liant pas les gènes de types sexuels, ce qui constituait un enrichissement par rapport aux autosomes, mais leurs fonctions n'étaient vraisemblablement pas impliquées dans la sélection antagoniste. Ainsi, nous n'avons trouvé aucune preuve solide que la sélection antagoniste ait contribué à étendre la suppression de la recombinaison au-delà des gènes de types sexuels. D'autres hypothèses devraient donc être explorées pour améliorer notre compréhension de l'évolution des chromosomes liés au sexe.

III. La liaison des locus de type sexuel a-t-elle évolué plusieurs fois indépendamment depuis un système sans liaison chez les espèces à taux d'autofécondation élevés ?

Branco, S., **Carpentier, F.**, de la Vega, R. C. R., Badouin, H., Snirc, A., Le Prieur, S., Coelho, M. A., de Vienne, D. M., Hartmann, F. E., Begerow, D., Hood, M. E. & Giraud, T. (2018). Multiple convergent supergene evolution events in mating-type chromosomes. Nature communications, 9(1), 2000.

La convergence adaptative fournit des indications uniques sur la prévisibilité de l'évolution et sur les processus de diversification biologique. Les supergènes (liaison de gènes avec des combinaisons d'allèles bénéfiques) sont des exemples frappants d'adaptation, bien que leur prévalence ou leur évolution soient peu connues. Une étude récente sur les champignons causant la maladie du charbon des anthères a documenté la formation de supergènes par réarrangements chromosomiques ayant permis la liaison des locus de types sexuels *via* une suppression de recombinaison ; les locus de types sexuels contrôlent la compatibilité avant et après la reproduction. Dans cette étude, de nouveaux assemblages génomiques de haute qualité révèlent quatre autres cas indépendants de réarrangements chromosomiques menant à des liaisons des locus de types sexuels chez des espèces étroitement apparentées. De telles transitions convergentes dans l'architecture génomique de la détermination du type sexuel indiquent une forte sélection favorisant la liaison des locus de types sexuels, qui co-ségrègent alors et constituent des supergènes. Nous avons aussi mis en évidence des strates évolutives indépendantes (extensions progressives de la suppression de recombinaison) chez plusieurs espèces, avec de nombreux réarrangements, des pertes de gènes et des accumulations d'éléments transposables. Nous révélons donc ici une remarquable convergence dans l'évolution des chromosomes de type sexuels : la formation récurrente de supergènes et l'évolution répétée de phénotypes similaires par des changements génomiques différents.

Carpentier, F., de la Vega, R. C. R., Branco, S., Snirc, A., Coelho, M. A., Hood, M. E., & Giraud, T. (2019). Convergent recombination cessation between mating-type genes and centromeres in selfing anther-smut fungi. Genome research, 29(6), 944-953.

Le degré d'autofécondation a un impact majeur sur l'adaptabilité et est souvent contrôlé par des mécanismes moléculaires déterminant la compatibilité lors de la reproduction. Les changements dans les systèmes de compatibilité sont donc des événements évolutifs importants, mais leurs mécanismes génomiques sous-jacents sont souvent mal compris. Les champignons présentent des changements fréquents dans les systèmes de compatibilité et leurs génomes de petite taille facilitent l'étude des mécanismes impliqués. En particulier, la liaison entre les locus déterminant la compatibilité avant et après la reproduction a évolué à de nombreuses reprises, ce qui augmente les chances de compatibilité des gamètes sous autofécondation. Ici, nous avons étudié les chromosomes de types sexuels de deux champignons causant la malade du charbon des anthères avec des locus de types sexuels non liés malgré un fort taux d'autofécondation. Des analyses de ségrégation et des comparaisons d'assemblages génomiques de haute qualité ont révélé que ces deux espèces présentaient des liaisons entre les locus de types sexuels et leurs centromères respectifs. Cette configuration génomique confère les mêmes probabilités de compatibilité des gamètes que la liaison entre les locus de types sexuels, considérant le système de reproduction particulier de ces champignons (reproduction entre les gamètes issues d'un même évènement de méiose, appelée reproduction intra-tétrade). La suppression de recombinaison est associée à une grande inversion dans un seul des quatre événements de liaison. L'absence de polymorphisme trans-spécifique des gènes situés dans les régions non-recombinantes et les estimations de l'âge des suppressions de recombinaison indiquent que les événements de suppression de recombinaison se sont produits indépendamment chez les deux espèces sœurs. Notre étude montre que la sélection naturelle peut conduire à plusieurs reprises à des arrangements génomiques similaires et à des phénotypes similaires, et que différentes voies d'évolution peuvent conduire à des réponses à la sélection distinctes mais aussi bénéfiques. Notre étude souligne en outre que la reproduction intra-tétrade et la liaison de gènes avec les centromères ont d'importantes conséquences génétiques et évolutives, ce qui est encore peu reconnu, malgré leur présence dans de nombreux taxons.

IV.   L'expression différentielle des gènes dans les régions non-recombinantes résulte-t-elle d'une dégénérescence ou d'une sélection antagoniste entre types sexuels ?

La dégénérescence dans les régions non-recombinantes, telles que les supergènes ou les chromosomes sexuels, peut provoquer une expression génique sous-optimale, entraînant une expression différentielle entre les allèles si les mutations surviennent aléatoirement dans l'un ou l'autre allèle. Une diminution de l'expression allélique due à la dégénérescence a en effet été suggérée dans divers chromosomes sexuels. Cependant, l'existence d'une association entre les caractéristiques spécifiques de la dégénérescence et la réduction de l'expression allélique n'a pas été étudiée en détail et l'antagonisme sexuel peut également expliquer l'expression allélique différentielle sur les chromosomes sexuels. Le champignon causant la maladie du charbon des anthères, *Microbotryum lychnidis-dioicae*, est idéal pour tester les associations entre les signatures de dégénérescence spécifiques et l'expression différentielle car : 1) des cultures distinctes de cellules haploïdes de types sexuels opposés aident à identifier l'expression différentielle, 2) il y a de multiples strates évolutives sur ses chromosomes de types sexuels, reflétant des événements successifs de suppression de recombinaison, 3) il n'y a aucun antagonisme sexuel qui induirait une sélection pour une expression différentielle. Nous avons constaté que les gènes différentiellement exprimés étaient enrichis dans les strates évolutives les plus anciennes et que plusieurs signatures de dégénérescence étaient plus importantes dans les gènes différentiellement exprimés que dans ceux avec des niveaux d'expression similaires. En particulier, deux signatures de dégénérescence ont été trouvés associées de façon significative aux allèles les moins exprimés : l'insertion d'éléments transposable en amont des gènes et des indels et/ou des codons stop dans les séquences géniques. D'autres signatures de dégénérescence associées à l'expression différentielle incluaient des substitutions non-synonymes, ainsi qu'une altération du contenu en intron et en taux de GC. L'association entre l'expression différentielle des gènes et la dégénérescence des allèles est pertinente pour un large éventail de taxons où la compatibilité entre types sexuels ou le sexe est déterminée par des gènes situés dans de grandes régions sans recombinaison.

Table of content

1 General introduction

1.1 Evolution of sexual reproduction and the evolution of recombination suppression

1.1.1 Sexual reproduction versus asexual reproduction

Sexual reproduction can be defined as the process resulting in the production of offspring associated with an alternance between diploid and haploid stages and involving genetic exchange. From a diploid cell, meiosis results in gametes with haploid genomes resulting from the recombination and the segregation of homologous chromosomal pairs. Meiotic recombination, called hereafter recombination, corresponds to the genetic reshuffling of the genetic information carried by homologous pairs of chromosomes that occurs during meiosis through crossing-overs. In contrast, asexual reproduction refers to the production of offspring without changes in ploidy level and without genetic exchange. Offspring are thus genetically identical to their parents, except for mutations. Most eukaryotes reproduce strictly sexually (Bell 1982) while others alternate between sexual and asexual reproduction, as do many fungi (*e.g.* yeast; Nieuwenhuis and James 2016) and some animals (*e.g.* aphids; Moran 1992). While ubiquitous, the frequency of strict asexual lineages is low, being estimated to represent from 0% to 1% of all living organisms (Kearney 2003; Schaefer et al. 2006; De Meeûs et al. 2007), 1% in angiosperms (Whitton et al. 2008) and 0.1% of animals (White 1978; Vrijenhoek 1998). The high prevalence of sexual reproduction represents a puzzle in evolutionary biology, because sexual lineages bear high costs compared to clonal (i.e. asexual) lineages (Maynard Smith 1978; Hartfield and Keightley 2012).

Sexual reproduction implies a twofold cost relative to asexual reproduction in anisogamous organisms without paternal investment in offspring (Maynard Smith 1978). Considering a female who reproduces sexually and another who reproduce clonally, and assuming that each produces two offspring per generation on average, the number of descendants from the clonal female at the nth generation will be 2^n times that of the sexual female. This is because sexual females produce only 50% of females among their offspring and only females produce offspring. Another way of illustrating this twofold cost of sex is that a clonal female transmits twice as many of her genetic material than a female who reproduces sexually, whose offspring have half of his genetic material coming from a male, who does not

invest energy in their offspring in most species. The twofold cost of sex is thus also referred to as the cost of males. It raises the question of how sexual reproduction can be maintained over long evolutionary times, because a mutation causing clonality arising in a sexual population should quickly invade. Furthermore, sex has additional costs compared to asexuality as it requires time to find mates, and it may increase the risk of predation, result in disease transmission, and often requires a tremendous energy to produce sexual structures. One of the biggest challenges of evolutionary biology has therefore been to explain how sexual reproduction can be so widespread despite all these costs, by investigating what advantages compared to asexuality could compensate its costs, and in particular the twofold cost of males.

1.1.2 Evolutionary advantages of sexual reproduction

Many of the studied advantages of sexual reproduction result from recombination. During meiosis, homologous chromosomes pair and recombination repairs double-stranded DNA damage using an undamaged DNA copy as a template (Bernstein et al. 1981). This template may be located on a homologous chromosome and the repairing may result in a crossing-over *i.e.* reciprocal exchange of DNA segment between two homologous chromosomes (Otto and Lenormand 2002; Bernstein et al. 1981). Across long time scales, many crossovers occur between homologous chromosomes so that chromosomal segments are transmitted independently one from each other.

Recombination thus allows selection to act on each locus independently from the selection on nearby loci, which is more efficient in terms of adaptation. Otherwise, selection acting at one locus interferes with selection acting at a second linked locus. This phenomenon is known as Hill-Robertson interference, and several forms can be distinguished depending on the type of selection considered.

1. Genetic hitchhiking (Maynard Smith and Haigh 1974) occurs when an allele at a locus is selected for and drags linked alleles with it, resulting in the increase in frequency in a population for both the selected allele and the associated allele at another locus.
2. Background selection (Charlesworth et al. 1993; Charlesworth 1995; Hudson and Kaplan 1995) occurs when a newly arisen deleterious allele is selected against, decreasing in frequency together with alleles at nearby loci.

Recombination is also advantageous because it allows to generate chromosomes with fewer deleterious mutations than either parental ones, thus avoiding the Muller's ratchet. The Muller's ratchet (Muller 1932; Fisher 1930; Felsenstein 1974) is a metaphor illustrating that, in an asexual population, the chromosomes that contain the fewest deleterious mutations are regularly lost from any finite population through stochastic processes, resulting in a fixation of increasing numbers of deleterious mutations in the population; indeed, there is no way without recombination to re-generate a chromosome with fewer deleterious mutations from two homologous chromosomes carrying different deleterious mutations.

At the population level, recombination is advantageous because it generates genetic variability which increases adaptation rate. A clonal progeny bears the exact same genome as its mother, except for mutations, which is the only source of genetic novelty in asexual lineages. In contrast, in a progeny resulting from sexual reproduction, the genotypes result from a mixture between those of the parents. It has therefore been argued that sex could be maintained in environments fluctuating rapidly over time or in heterogeneous environments (Maynard Smith 1971; Charlesworth 1976).

However, environmental conditions may not change fast enough to prevent the invasion of clonal females in populations. The advantages of recombination in terms of DNA repair and faster adaptation rates occur over long evolutionary times while the costs of sexual reproduction are high on the short them, so that it has been argued that such advantages cannot explain why sexual populations are not invaded by asexual individuals on the short term. Yet, the biotic environment may impose heavy and short-term changing constraints, as proposed by the red queen hypothesis (Jaenike 1978; Hamilton 1980). Organisms are engaged in an arms race with their enemies, with a selection on parasites to better infect their hosts and a selection on hosts to better resist parasites. In this context, rapid adaptation as allowed by recombination can be advantageous on the short term, as the environment changes rapidly, from one generation to the other, with enemies having adapted to the most frequent genotypes of the previous generation.

Overall, recombination results in an increase of the adaptation rate which can be advantageous on both the short and long terms and increases selection efficiency on the long term. However, when selective pressures are stable, recombination may also be disadvantageous on the short term because it may break apart beneficial allelic combinations. Local recombination suppression can therefore be selected for avoiding breaking apart

beneficial allelic combinations, which leads to the genetic structure known as "supergene", where multiple genes are linked together and are transmitted as a single locus.

1.1.3 Evolutionary advantages of non-recombining regions in sexual lineages: the case of supergenes

A supergene is a system of linked genes transmitted as a single locus, allowing maintaining beneficial allelic combinations. The wing pattern in the *Heliconius* butterflies and some floral dimorphism and self-incompatibility systems are textbook examples of phenotypes controlled by supergenes and maintained polymorphic.

In butterflies from the *Heliconius* genus, several wing-pattern morphs are maintained within populations (Joron et al. 1999). Wing-pattern morphs are mimics of other locally abundant distantly related species (Saenko et al. 2019). The wing-pattern mimetism allows both mimetic species to decrease their risk of predation, as both mimetic species are unpalatable and their shared predators associate wing patterns to unpalatability faster when the morph is more abundant. In *Heliconus* species, the wing-pattern morphs differ notably in colour, size and shape of pattern elements (Nadeau 2016; Saenko et al. 2019). These phenotypes are regulated by genes located on the same chromosome and segregating as a single locus, which therefore constitute a supergene (Joron et al. 2006).

In the plant species *Primula vulgaris*, self-incompatibility is prevented at the molecular level by multiple genes linked together, forming a supergene. This supergene also encompasses several genes involved in floral dimorphism, controlling the length of the style and the anther location (Mather 1955; Ganders 1979). The supergene results in two self-incompatible morphs, one with a long style and short anthers ('pin' morph) and one with a short style and long anthers ('thrum' morph). Beyond the *Primula* genus, such floral dimorphism is often associated with a molecular self-incompatibility system (Kappel et al. 2017). A general model for explaining such association (Charlesworth and Charlesworth 1979) proposed that self-incompatibility evolved first to avoid high levels of homozygosity and the resulting inbreeding depression, *i.e.* reduced fitness due to the expression of deleterious recessive alleles at the homozygous state (Charlesworth and Charlesworth 1987). Following the evolution of such an self-incompatibility system, floral dimorphism evolved to avoid pollen wastage by its transfer to incompatible morphs (Charlesworth and Charlesworth 1979) as a pollinator leaving a pin morph flower

would carry pollen on its abdomen which will not reach the stigma of the thrum morph flower, and *vice versa*.

The establishment of a supergene in a population may be facilitated in regions with low rates of recombination. For instance, the loci controlling the self-incompatibility system in *Arabidopsis lyrata* are thought to have been physically close one to each other before the evolution of self-incompatibility (Goubet et al. 2012; Chantha et al. 2013; Charlesworth 2016), which would have decreased the probability they recombine. Supergenes have been frequently found in regions around centromeres, that are known to have low recombination rates (Lyon 2003; Larracuente and Presgraves 2012; Rick 1971; Van Der Gaag et al. 2000; Schwander et al. 2014). Structural variations can also affect recombination rate and be associated with supergene evolution. Inversions have been repeatedly found associated with supergenes, notably in the fire ant *Solenopsis invicta* (Wang et al. 2013). In this species, a supergene controls a suite of morphological and life-history traits (*e.g.* queen fecundity, odour of mature queens, worker size) associated to the presence of a single or several fertile queens per nest (Ross and Keller 1995; Wang et al. 2013). Recombination suppression has been proposed to result from multiple inversions in these ants (Wang et al. 2013; Huang et al. 2018), as well as in the *Heliconius* supergenes. However, the direction of the causal link between structural variations and recombination suppression is often difficult to infer because rearrangements are frequent after recombination suppression (Sun et al. 2017; Bergero et al. 2008; Badouin et al. 2015).

Sex chromosomes represent interesting and widespread cases of supergenes, having evolved independently multiple times, with some being maintained for millions of years, such as the mammalian sex chromosomes (Veyrunes et al. 2008). Moreover, non-recombining regions on sex chromosomes can span whole chromosomes (Borodin et al. 2012) or nearly whole chromosomes, in which case only short recombining regions remain at the edges of the chromosomes (Bachtrog and Charlesworth 2001; Balounova et al. 2019), *i.e.* pseudo-autosomal regions (PARs). Sex chromosomes are therefore ideal case studies to investigate the evolution of recombination suppression in sexual species.

1.1.4 Sex chromosomes and evolutionary strata

Sex chromosomes were discovered by Nettie M. Stevens in 1905 (Stevens 1905; Brush 1978; Abbott et al. 2017). She observed in male mealworms sperm cells with one chromosome shorter than others (later called the Y chromosome), which always resulted in the production of male offspring after fertilization with an egg, while the others always resulted in the production of female offspring (Stevens 1905). The XY sex-determining system she discovered was later found to be similar in mammals, with XY individuals being males and XX individuals being females. Other sex-determining systems have been discovered since then, such as the ZW sex-determining system in birds, with ZZ individuals being males and ZW being females (Morgan 1909), or the UV sex-determining system in bryophytes (Allen 1945), determining sex at the haploid stage.

The theory explaining sex chromosomes evolution has been formalized within the scope of the evolution of separate sexes (Charlesworth and Charlesworth 1978; Westergaard 1958). Sex chromosomes evolved from a pair of homologous chromosomes that acquired genes determining sex (Ohno 1967). Assuming that individuals bear both male and female functions in the ancestral state, one can imagine that two mutations occurred at distinct loci on homologous chromosomes, one resulting in male sterility and the other one in female sterility. The lack of recombination between the two loci should then be selected for: individuals are fertile only when heterozygote at both loci, and recombination in the germline of such individuals could result in neuter sterile individuals carrying both male-sterility and female-sterility loci. According to this model, the sex-determining locus is a supergene.

The sex-determining locus can be in a small non-recombining region, such as on the XY sex chromosomes of the plant *Carica papaya* (Yu et al. 2008) or the moss *Ceratodon purpureus* (Mcdaniel et al. 2007). Some other sex chromosomes do not recombine over most of their length (Bergero and Charlesworth 2009). In a XY sex-determining system such as in mammals, the XY chromosomes do not recombine in males while the XX chromosomes recombine in females; only the Y chromosome therefore bears a non-recombining region. Such recombination suppression results in a decreased selection efficiency due to Hill-Robertson interferences. Less efficient selection results in the accumulation of deleterious mutations on the Y chromosomes, such as non-synonymous substitutions, gene disruption, gene losses, and substitutions leading to suboptimal gene expression level. Genomic rearrangements, epigenetic

modifications and transposable elements may not always be deleterious *per se* but can sometimes disrupt gene function or expression and repeats that have accumulated in great numbers can impose strong energetic costs.

Once recombination is suppressed, the X- and Y-chromosome accumulate genetic changes independently. Synonymous substitutions are considered neutral because they do not impact the amino acid sequence in proteins. Assuming the effective size constant, the accumulation of synonymous substitutions thus only depends on the rate of mutation, which allows to estimate the age of recombination suppression based on the synonymous divergence between alleles from the Y- and the X-chromosome. As the X-chromosome recombines in females, it mostly retains the ancestral gene order (from before recombination suppression), while the Y-chromosome undergo genomic rearrangements. By plotting the X-Y allelic synonymous divergence along the X-chromosome gene order using human sex chromosomes, a strong correlation was found between the d_S values and the location of their respective X-allele. The age of recombination suppression decreased in a stepwise fashion along the X-chromosome gene order, a pattern called evolutionary strata of X-Y allelic divergence, indicating that recombination suppression expanded progressively along the sex chromosomes. Such evolutionary strata have been found in mammals (Lahn and Page 1999), birds (Handley et al. 2004; Nam and Ellegren 2008) and plants (Bergero et al. 2007). Such a stepwise expansion of recombination suppression has been suggested to be triggered by mutations with sexually antagonistic effects, *i.e.* being beneficial in one sex but deleterious in the other (Charlesworth and Charlesworth 1978; Nei 1969). Such sexually antagonistic genes should be selected for being linked to the sex-determining genes, so that only females carry the female beneficial alleles, and conversely for males, which would extend the regions without recombination in a stepwise manner.

The model based on sexual antagonism has been widely thought to explain the formation of evolutionary strata over the past decade. However, it lacks experimental support despite extensive research. Some sex chromosomes even seem to completely lack sexually antagonistic genes (Reichwald et al. 2015; Tennessen et al. 2016; Ponnikas et al. 2018). However, sexually antagonistic phenotypic traits are not easy to identify. Even when sex chromosomes are found enriched in sexually antagonistic factors, the sexually antagonistic mutations may have accumulated after recombination suppression (Rice 1984; Charlesworth et al. 2005; Ironside 2010; Ponnikas et al. 2018).

In a textbook model for the study of sex chromosomes, the fish *Poecilia reticulata,* colour is sexually antagonistic because males with bright colours attract more females but also more predators while females with bright colours gain no advantages while still suffering from higher risk of predation (Wright *et al.*, 2017; Lindholm & Breden, 2002). *Poecilia reticulata* has a XY sex-determining system, and ca. 80% of the male coloration traits were shown to be determined by genetic factors partially or fully linked to the male-determining loci on the Y chromosome (Lindholm and Breden 2002). The sex chromosomes of *P. reticulata* have therefore been promoted as an excellent model to show the role of sexually antagonistic genes in triggering the formation of evolutionary strata. However, the enrichment of male coloration traits near the male-determining gene does not allow any inference about the causative role of sexual antagonism in the formation of evolutionary strata as the enrichment can have followed recombination suppression rather than driving it (Bergero et al. 2019). Indeed, other mechanisms than sexual antagonism can trigger recombination suppression in sex chromosomes (Ironside 2010; Charlesworth 2018; Ponnikas et al. 2018) which may subsequently facilitate establishment of sexually antagonistic genes in linkage to the male-determining loci (Bergero et al. 2019). Furthermore, demonstrating the existence of evolutionary strata requires plotting the synonymous divergence between alleles from the Y and X chromosomes along the X chromosome gene order, which has not been done in *P. reticulata* (Charlesworth 2018). Recent studies actually suggested the absence of evolutionary strata in *P. reticulata,* although again without plotting synonymous divergence between alleles from the Y and X chromosomes (Bergero et al. 2019; Bergero and Charlesworth 2019). A genome-wide sexually dimorphic crossover pattern has been reported, with males only recombining at the tip of all chromosomes and females fully recombining. This creates *de facto* a large region which is in nearly complete linkage with the male-determining gene on the Y chromosome. Such a linkage would facilitate subsequent establishment of sexually antagonistic genes because male-beneficial alleles could be maintained in association with the male-determining loci. However, one cannot exclude that the restriction of crossing-overs in males at the tip of chromosomes has been selected for linking sexually antagonistic genes to the sex-determining genes. Despite the lack of clear and definitive evidence for evolutionary strata and for the role of sexual antagonism in triggering recombination suppression, the sex chromosomes of *P. reticulata* has widely been cited as a case supporting the theory based on sexual antagonism to explain evolutionary strata.

Other evolutionary mechanisms have been proposed to explain the extension of recombination suppression, relying for instance on heterozygote advantage or neutral rearrangements (Ironside 2010; Charlesworth 2018; Ponnikas et al. 2018), but they have been poorly considered. The hypothesis based on the heterozygote advantage propose that recessive deleterious mutations may accumulate at the margin of the non-recombining region and would favour an extension of the recombination suppression to prevent their expression at the homozygous state (Charlesworth and Wall 1999). Evolutionary strata could otherwise result from the fixation in population of neutral rearrangements by genetic drift in one of the sex chromosomes at the margin of the established non-recombining region, which would instantly suppress recombination.

Interestingly, it has been shown that large non-recombining regions evolved in several fungi near loci controlling mating types, while sexual reproduction occurs between isogamous individuals lacking male / female functions. If the recombination suppression extended stepwise in such organisms without any sexual antagonism, it means that other forces are able to generate evolutionary strata, and they may also apply to sex chromosomes. These organisms may thus constitute good models to study alternative mechanisms to sexual antagonism for understanding evolutionary strata.

1.2 Mating types in fungi

In animals and plants, sexual reproduction occurs between male and female individuals that are phenotypically differentiated. Sex is determined at the diploid stage by various mechanisms, such as temperature-dependent sex determination or genetic sex determination (Bachtrog et al. 2014). In fungi, mating compatibility is determined at the haploid stage by specialized loci called mating-type loci. Fungi do not have separated sexes in the sense of the existence of diploid individuals producing only one type of gamete as in animals and plants. Most ascomycete fungi can be seen as hermaphrodites because each haploid individual can produce both large and small gametes which correspond to female and male roles, respectively (Debuchy et al. 2010; Billiard et al. 2011). In basidiomycete fungi, there is no production of small and large differentiated gametes; however, in some species such as *Schizophyllum commune*, male-like and female-like roles can still be distinguished during mating: the female role is played by a nucleus-acceptor haploid mycelium and the male role by a nucleus-donor spore falling on the mycelium (Nieuwenhuis et al. 2011; Nieuwenhuis and Aanen 2018). In such cases, male and female-like roles are not genetically determined either (Casselton 2002; Billiard et al. 2011).

For most fungi, gametes can successfully fuse only when they carry functionally distinct mating-type alleles, which is called heterothallism, in opposition to homothallism, in which case no mating-type locus prevents the fusion between haploid clonemates (Billiard et al. 2012, 2011; Lin and Heitman 2007). Homothallism thus allows any mating system, including same-clone mating, as seen in the fungal genus *Cochliobolus* (Yun et al. 1999). It is worth noting that homothallism may not have evolved to promote same-clone mating (Billiard et al. 2012, 2011), in contrast to the prevailing opinion in the fungal literature. Indeed, same-clone mating has no or little advantage over asexuality as it leads to instant genome-wide homozygosity, which renders recombination ineffective to generate new allelic combinations. Homothallism may have rather evolved because it ensures compatibility with all other gametes in the population. Heterothallism prevents mating between haploid gametes with the same mating-type allele, but does not prevent selfing, i.e., mating between gametes resulting from the same meiosis event (intra-tetrad mating) or from distinct meiosis events (inter-tetrad mating) of the same diploid individual. Indeed, mating type is determined at the haploid stage, so that each diploid individual is heterozygous for mating type in heterothallic fungi. The union of two gametes resulting from meiosis events of distinct diploid individuals is defined as outcrossing, as in

other organisms. In heterothallic fungi, mating type can be determined by a single locus or by two unlinked loci. These two gamete compatibility systems are called bipolarity and tetrapolarity, respectively, because there are two mating-type alleles in tetrads produced by bipolar individuals, and four mating-type alleles in tetrads produced by tetrapolar individuals. In ascomycetes, only bipolar gamete compatibility systems occur. In most ascomycete species, bipolarity is regulated by a MAT locus whose alleles, MAT1-1 and MAT1-2, are so dissimilar that they are called idiomorphs rather than alleles. MAT1-1 encodes a high-mobility group protein and MAT1-2 a transcriptional activator protein with an alpha box domain (Coppin et al. 1997), and both functions are required for a successful mating in these species. A few ascomycete species have uncanonical gamete compatibility systems with additional proteins linked to the MAT locus (Coppin et al. 1997) or without MAT1-1 (Bennett and Johnson 2005).
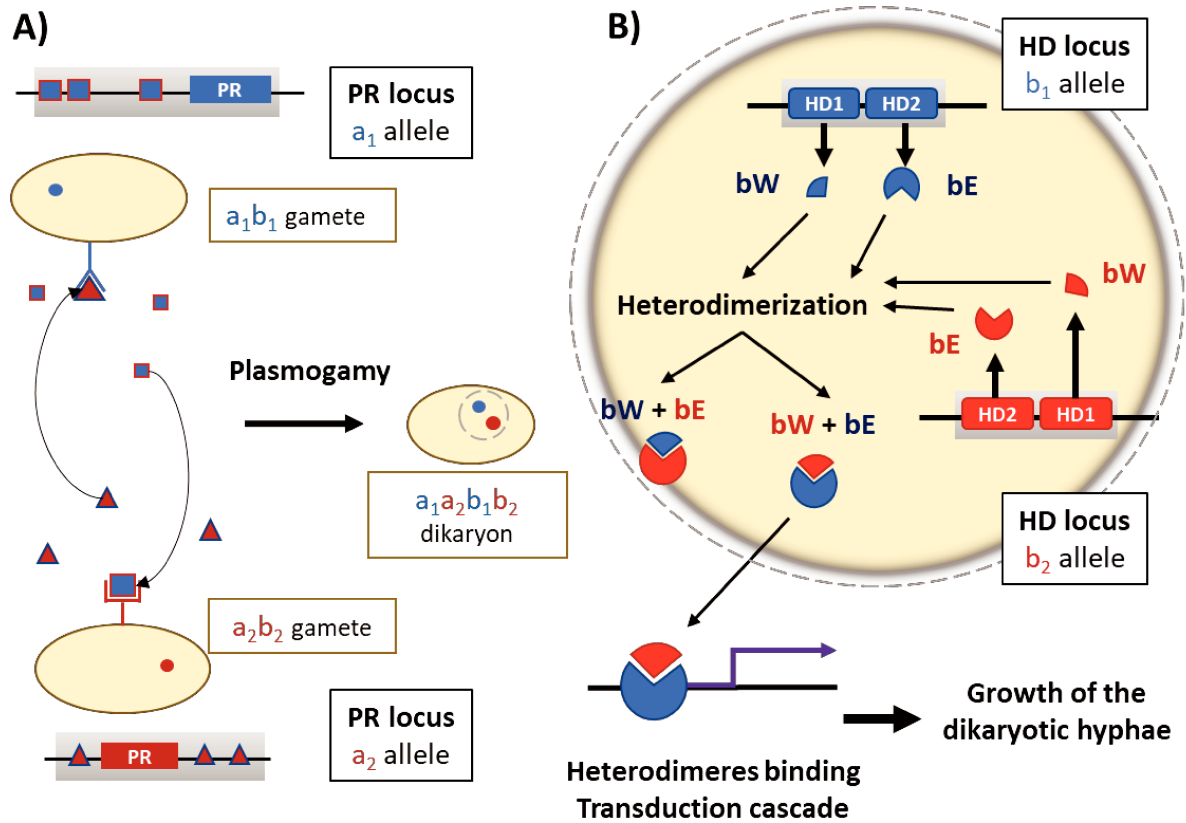
Figure 1: Mating-type determination in basidiomycetes. Successful mating is only possible between haploid cells carrying different alleles at each of the PR and HD locus. **A)** The PR locus – generically called a locus in basidiomycetes – is involved in gamete recognition and plasmogamy. In each mating-type, the pheromone receptor allele is linked to its incompatible pheromone alleles. **B)** The HD locus – generically called b locus in basidiomycetes – is involved in the maintenance and growth of the dikaryotic hyphae. This function is achieved by a transcriptional regulation resulting from the dimerization two homeodomain proteins produced by alternative alleles from distinct HD genes.

In basidiomycetes, most species are tetrapolar, with two mating-type loci generically called *a* and *b* loci. Successful mating is only possible between haploid cells carrying different alleles at both loci (Figure 1). The *a* locus is involved in gamete recognition and their fusion, only involving cytoplasm at this step, *i.e.* plasmogamy, and thus resulting in a cell with two nuclei, qualified as dikaryotic. The *a* locus is composed of a pheromone receptor gene and one or multiple pheromone genes. Gametes fuse when the pheromones produced by one gamete are recognized by the pheromone receptor of the other gamete (Figure 1A). In each mating-type, one pheromone receptor allele is linked to its incompatible pheromone alleles (Figure 1A). The *b* locus is involved in the maintenance and growth of the dikaryotic hyphae, in which two haploid nuclei remain separated in the hyphal cells after the plasmogamy of two compatible gametes (Figure 1B). This function is achieved by a transcriptional regulation resulting from the dimerization of two homeodomain proteins (bW and bE), encoded by two divergently transcribed genes of the *b* locus (Figure 1B). Recombination is suppressed at each of the *a* and *b* loci in tetrapolar species. Linking a pheromone receptor allele to its incompatible pheromone alleles and linking two homeodomain-coding alleles that cannot undergo dimerization is essential for ensuring the mating-type function.

The *a* and *b* loci are unlinked in tetrapolar species and can bear two alleles each or be highly polymorphic, yielding thousands of possible mating types, such as in *Coprinus cinerea* or *Schizophyllum commune*. In some basidiomycetes, the pheromone receptor gene lost its function in mating-type specificity, which resulted in transitions from tetrapolarity to bipolarity, as in *Coprinellus disseminatus* (James et al. 2006) and *Pholiota nameko* (Yi et al. 2009). Bipolarity in basidiomycetes can also result from the linkage of the *a* and *b* locus, with only a single pair of alternate alleles at each, and this linkage is achieved by a recombination suppression. Linking mating-type loci is beneficial under selfing-based mating systems because a diploid individual then produces only two mating types which increases the odds of gamete compatibility (50% compatibility) compared to four different mating types (25% compatibility; Nieuwenhuis et al. 2013). The mating-type loci linkage thus forms a supergene. Interestingly, recombination has been reported to be suppressed over large regions comprising the mating-type loci and many other genes in several fungi, such as in *Ustilago hordei* (Bakkeren et al. 2006), *Cryptococcus neoformans* (Fraser and Heitman 2004), *Malassezia globosa* (Xu *et al*., 2007), *Neurospora tetrasperma* (Gallegos *et al*., 2000) and *Microbotryum lychnidis-dioicae* (Hood 2002), which suggested there may be evolutionary strata on fungal mating-type chromosomes. Some studies claimed the existence of evolutionary strata on mating-type

chromosome of the fungal species *N. tetrasperma* (Menkis et al. 2008), *Cryptococcus neoformans* (Fraser et al. 2004) and *M. lychnidis-dioicae* (Votintseva and Filatov 2009), based on evidence for heterogeneous synonymous divergence values between alleles associated with the alternative mating types at genes located in non-recombining regions. However, heterogeneous synonymous divergence values are expected in non-recombining regions even in the absence of progressive extension of recombination suppression, due to gene conversion or stochasticity in coalescence, and no pattern of discrete evolutionary strata along the mating-type chromosomes had been reported. Definitive evidence for the existence of evolutionary strata on fungal mating-type chromosomes were thus lacking.

The mating-type chromosomes of the anther-smut fungus *Microbotryum lychnidis-dioicae* are dimorphic (Hood 2002), with a large region without recombination, spanning 90 % of the chromosome length (Hood et al. 2013). Moreover, mating-type chromosomes in multiple *Microbotryum* species were shown to have undergone genomic degeneration in the form of non-synonymous substitution accumulation or accumulated repetitive sequences (Fontanillas et al. 2015) as in sex chromosomes. Furthermore, a theoretical model had suggested that the particular automictic mating system of *Microbotryum* fungi could promote a stepwise extension of recombination suppression to shelter deleterious alleles (Antonovics and Abrams 2004) The *Microbotryum* species complex therefore seemed to be a good model to test hypotheses about the formation of evolutionary strata as well as expectations drawn from the sex chromosome evolution theory.
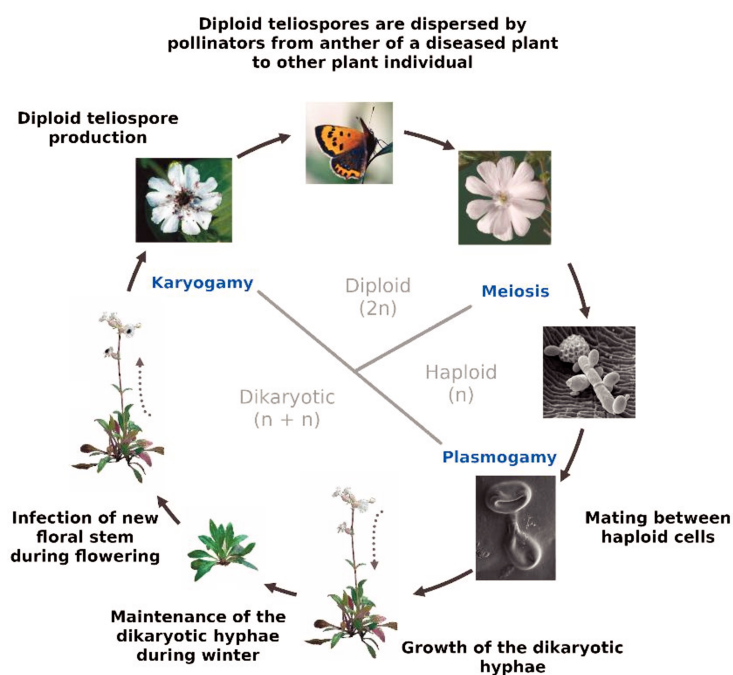
Figure 2: Life cycle of *Microbotryum violaceum*. Adapted from Lòpez-Villavicencio et al. (2007).

1.3 The *Microbotryum* species complex

The *Microbotryum* genus forms a complex of basidiomycete species that cause anther-smut disease on over hundreds of plants, from the *Caryophyllaceae* family (Thrall et al. 1993; Hood et al. 2010) and other plant families (*e.g.* Dipsacaceae, Lamiaceae; Kemler et al. 2006). Anther-smut fungi can be easily observed in nature under the form of purplish powder in the anthers of flowers, which gave the generic name of the *Microbotryum* species complex – *Microbotryum violaceum* (Pers.), originally called *Ustilago violacea* (Pers.) Roussel (Deml and Oberwinkler. 1982). Many *Microbotryum* species are thought to be have specialized on one host species each, and a few others parasitize a few closely related hosts (Refrégier et al. 2008). For this reason, I will name *Microbotryum* species by referring to the name of the host species which they were sampled, *e.g. Microbotryum violaceum* parasitizing *Silene paradoxa* abbreviated as *M. v. paradoxa*.

1.3.1 Life cycle

Microbotryum fungi can only sustain on perennial hosts given its life cycle (Thrall et al. 1993; Hood et al. 2010). During the host plant flowering, diploid teliospores are dispersed from the anther of a diseased plant individual to other plant individuals (Figure 2). Dispersal is pollinator-mediated for most species, though it can also be dispersed through wind or water (Shykoff and Kaltz 1997). Teliospores dispersed on leaf surfaces of the host plant undergo meiosis and produce linear and ordered tetrads. Compatible gametes undergo plasmogamy, which results in infectious dikaryotic hyphae that penetrate the host plant cell wall through the intercellular compartment in which the dikaryon spends the rest of the year. In the next infected flowering stems of the plant (which is often one year after the initial infection but can be in the later shoots the same year), dikaryotic cells undergo karyogamy, resulting in diploid teliospores in the anther of the host plant.
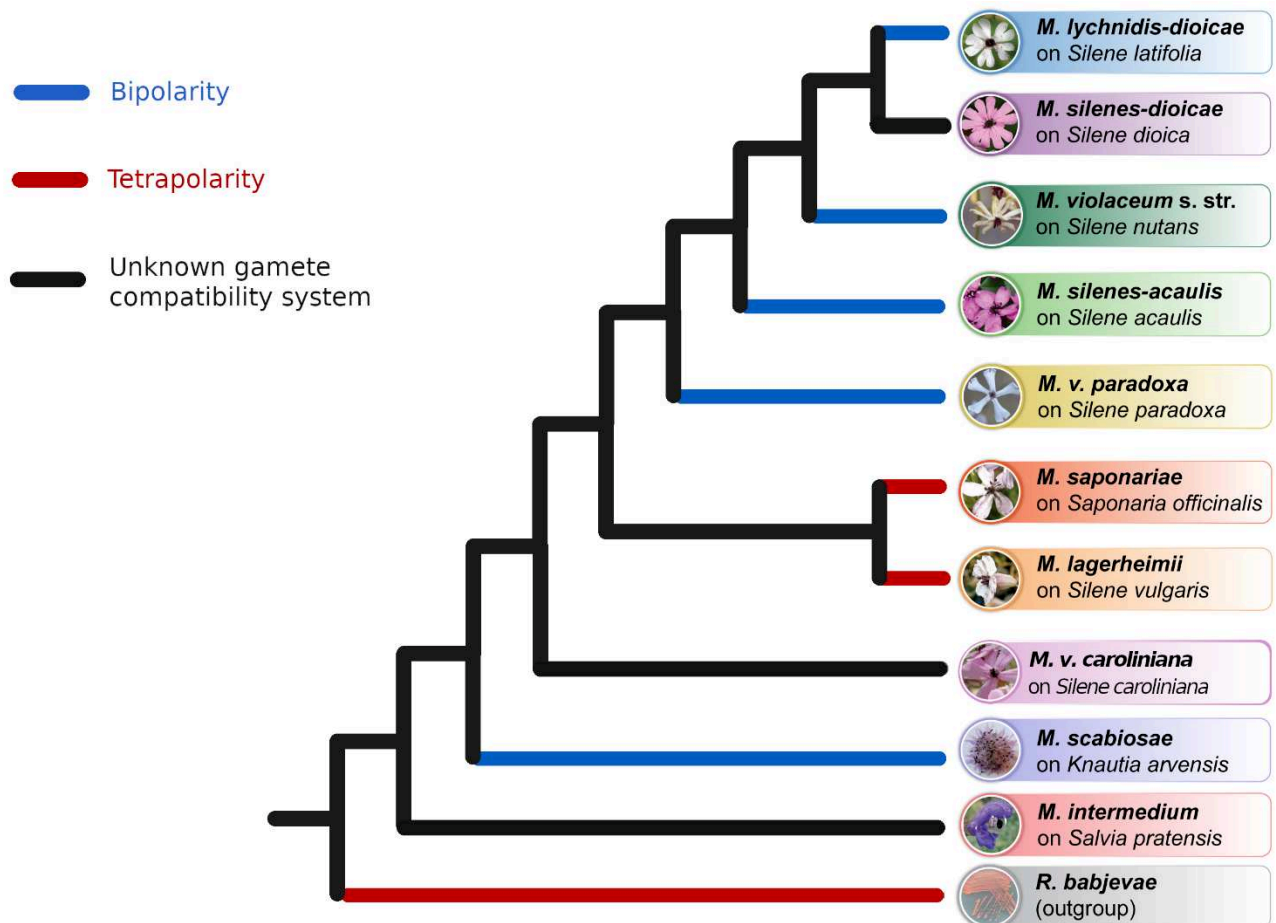
Figure 3: Cladogram of the *Microbotryum* genus (including here 10 *Microbotryum* species). *Rhodotorula babjevae* is used as an outgroup. Diseased host plants of *Microbotryum* fungi are shown. Colours on branches show the information we had on the gamete compatibility systems before the start of the PhD (Hood et al. 2015).

1.3.2 Mating system and mating types

Microbotryum fungi are heterothallic basidiomycetes and mainly reproduce through intra-tetrad mating *i.e.* between gametes produced through the same meiosis event and carrying compatibles alleles at both mating-type loci. The two mating-type loci are called PR (for pheromone receptor) and HD (for homeodomain) and are equivalent to the mating-type loci *a* and *b*, which are the generic terms used in basidiomycete (Figure 1). The PR mating-type locus has two alleles, called a_1 and a_2 (Day and Garber 1988; Garber and Day 1985). It has been showed that the a_1 and a_2 alleles of the PR locus are shared by multiple basidiomycetes species and started diverging ca. 370 million years ago (Devier et al. 2009). The diversity at the HD mating-type locus across the *Microbotryum* genus is less well known because gene genealogy trees are less clear (Petit et al. 2012) . The alternative alleles studied here at the HD mating-type locus will be called b_1 and b_2, always isolated from a single tetrad. The PR and HD mating-type loci were known to be linked by a recombination suppression in the bipolar species *M. lychnidis-dioicae*, which have thus only two mating types (Day and Garber 1988; Garber and Day 1985). By convenience, these two mating types are named in accordance to their PR allele - a_1 and a_2. A non-exhaustive historical review of discoveries and hypotheses related to mating system, mating types and recombination suppression in *Microbotryum* fungi is presented in Box 1.

In this PhD thesis, we studied the recombination suppression on mating-type chromosomes from the species *Microbotryum lychnidis-dioicae* and nine closely related *Microbotryum* species (Figure 3). All species studied in the thesis were sampled on *Silene* plant species, except *M. intermedium,* which was sampled on *Salvia pratensis*, *M. scabiosae* on *Knautia arvensis* and *M. saponariae* on *Saponariae officinalis*. The gamete compatibility system was known before the start of the PhD thesis for multiple *Microbotryum* species we studied (Figure 3; Box 1).

For the species studied in the thesis, one haploid genome of each of two compatible gametes from the same tetrad were sequenced (but *M. intermedium* for which only one genome could be sequenced) using the single molecule real-time sequencing technology proposed by the Pacific Bioscience company. The sampling, sequencing, assembly and gene prediction had been made before the start of my PhD by members of the GEE team and Michael Hood (Amherst College, USA).

Box 1: Historical review of discoveries and hypotheses related to mating system, mating types and recombination suppression in *Microbotryum*

Microbotryum violaceum (*sensu lato*) causes anther-smut disease in plants of the Caryophyllaceae family. These fungi are important research models in many fields of biology, including genomics, host-pathogen interactions and evolutionary ecology. In the early work on sexual compatibility, *Microbotryum violaceum*, called *Ustilago violacea* before 1982 (Deml and Oberwinkler. 1982), was the first fungi in which bipolar heterothallism was demonstrated (Kniep 1919).

More recently, the mating-type chromosomes of *M. lychnidis-dioicae* were the first to be found dimorphic, as initially revealed by electrophoretic karyotypes, and with extensive non-recombining regions, as supported by AFLP markers, being thus suggested to share genomic features with sex chromosomes (Hood 2002; Hood et al. 2013, 2004). A later study suggested that the non-recombining region on the mating-type chromosomes of *M. lychnidis-dioicae* was small but the size estimate was based on a few markers (Votintseva and Filatov 2009). This study also claimed the existence of evolutionary strata because they found heterogeneous synonymous divergence values in the non-recombining region, but no physical order of the loci had been provided and the variation in the degree of the allelic divergence did not look like discrete strata. A subsequent study reported an absence of correlation between the d_S level from (Votintseva and Filatov 2009) and an independent estimate of the age of the linkage of each marker to the mating-type based on gene genealogies (Petit et al. 2012).

The sequencing of cDNA libraries allowed the identification of the pheromone receptor gene (Yockteng et al. 2007) and later of the HD genes (Petit et al. 2012). STE3-like pheromone receptors have been characterized among sequences expressed during *Microbotryum lychnidis-dioicae* mating (Yockteng et al. 2007), and later amplified by PCR and Sanger-sequenced in multiple individuals with distinct mating-types in several *Microbotryum* species. Gene genealogies of the alternate alleles of the pheromone receptor gene from multiple basidiomycete species, including several *Microbotryum* species revealed in *Microbotryum* the deepest degree of trans-specific polymorphism ever reported. The *Microbotryum* a_1 and a_2 alleles had been estimated to diverge for 370 million years ago (Devier et al. 2009). The HD genes in *Microbotryum* were identified in the ESTs (Petit et al. 2012) by sequence similarity of previously characterized homeodomain in *Sporobolomyces* sp. (Coelho et al. 2010) which was the closest relative of *Microbotryum* with available relevant information. Incomplete trans-specific polymorphism signal has been found in HD gene trees, reconstructed using Sanger sequences from several *Microbotryum* species. Indeed, a_1 and a_2 alleles clustered by species in some groups and clustered by mating type with certain level of trans-specific polymorphism in other groups. While it is still not confirmed, the authors proposed that the recombination suppression at the HD genes was ancestral in the *Microbotryum* clade and that some rare crossing overs or gene conversion events could have reset the allelic divergence in some groups (Petit et al. 2012).

Box 1 (continued)

Pheromone genes have been identified in *Microbotryum* by sequence similarity with pheromone genes described in the Microbotryomycete *Rhodosporidium toruloides* fungus and their mating function has been experimentally validated (Xu et al. 2016). Studying pheromone genes and pheromone receptor gene sequences from a_1 and a_2 mating types of multiple *Microbotryum* species suggested the existence of distinct coevolutionary patterns between the two pairs of pheromones – pheromone receptors. The a_1 pheromone allele was present in many copies within species and showed little variation across *Microbotryum* species, as did the a2 pheromone receptor gene. In contrast, the a_2 pheromone allele was present in few copies within a genome and was found more diverse in sequence than the a_1 pheromone receptor a_1. These differences in coding sequence variation across *Microbotryum* species were shown to be associated with differences in the strength of purifying selection, which was found stronger on a_1 pheromone allele. These results are consistent with earlier observations on the mating behaviour in *Microbotryum* (Day 1976): a_1 cells initiate mating through a greater diffusion of a conserved a_1 signal while the a_2 cells play a responsive role through a_2 pheromones which do not need to be as conserved as the a_1 signal. The conjugation tube following the pheromone recognition indeed develops more rapidly and to a greater extent from a_2 than from a_1 cells, which supports the existence of differences in mating behaviour between mating types.

Further studies investigated degeneration features in the non-recombining regions of multiple *Microbotryum* species, sequenced using short-read Illumina sequences (Fontanillas et al. 2015). These studies reported an enrichment in degeneration features in the putative non-recombining regions, compared to pseudo-autosomal regions and autosomes, in the form of dN/dS and transposable element accumulation (Fontanillas et al. 2015). However, this study was based on a non-optimal assembly of the *M. lychnidis-dioicae* genome and assumed homology of all non-recombining regions across the *Microbotryum* species. Finally, long-read sequencing and high-quality assembly confirmed the existence of a large region of recombination suppression in *M. lychnidis-dioicae*, spanning around 90% of the mating-type chromosome length, with chaos of rearrangements (Badouin et al. 2015) and elevated synonymous divergence values between a_1- and a_2-associated alleles. Despite the heterogenous and elevated synonymous divergence values reported, no pattern of progressive extension of the non-recombining region had then been found. Still, uncertainty about evolutionary strata remained due to difficulties in inferring ancestral gene order. Indeed, ancestral gene order is essential when investigating evolutionary strata while, in opposition to the diploid XY sex determination system in which the X chromosome recombines in female, both a_1 and a_2 mating-type chromosomes lack recombination and therefore undergo rearrangements.

The distribution of the gamete compatibility system across the *Microbotryum* phylogeny led a previous study to infer that bipolarity evolved once at the basis of the *Microbotryum* clade and that there had been a reversal to tetrapolarity before the divergence of *M. saponariae* and *M. lagerheimii* (Hood et al. 2015). This hypothesis was the most parsimonious hypothesis to explain the evolution of bipolarity in terms of number of evolutionary transitions. However, it did not formally exclude possible independent acquisitions of bipolarity, which could be triggered by a selection acting on the odds of gamete compatibility. Bipolarity indeed confers higher odds of gamete compatibility than tetrapolarity, assuming high selfing rate. Formal tests using genomic data for investigating the evolution of the mating-type loci linkage remained to be performed.

1.4 Research questions

Sexual reproduction is widespread in eukaryotes, but its success has been and remains an evolutionary puzzle. On the long term, recombination is advantageous over asexuality as it may bring together beneficial allelic combination and allow an efficient selection to prevent the accumulation of deleterious mutation, which is due to Hill-Robertson interferences and Muller's ratchet. On the short term, however, recombination may be disadvantageous by breaking apart beneficial allelic combinations. Recombination suppression can therefore be selected for preserving beneficial allelic combinations, which are then transmitted as a single locus and constitute a supergene. Sex chromosomes are considered as supergenes, with a recombination suppression linking sex-determining genes. In many cases, however, recombination suppression is not restricted to the sex-determining locus and may span almost the whole sex chromosome length as a result of a stepwise extension of recombination suppression. Such large non-recombining regions may undergo genomic degeneration, with the accumulation of deleterious mutation due to less efficient selection. A largely accepted theory postulates that selection have favoured the linkage of sexually antagonistic genes to the sex-determining genes, generating evolutionary strata. However, this theory lacks strong experimental support.

In basidiomycete fungi, some bipolar species exhibit similar features to sex chromosomes on mating-type chromosomes such as large non-recombining regions which undergoes genomic degeneration. The recombination suppression on mating-type chromosomes links mating-type loci and evolved from tetrapolar species with unlinked mating-type loci. Under high selfing rate, bipolarity is advantaged over tetrapolarity as it increases the odds of gamete compatibility: only two mating types are produced in a progeny instead of four. The strong selection favouring bipolarity could have led to independent transitions from tetrapolarity to bipolarity. Studying the evolution of gamete compatibility system could thus reveal interesting cases of convergent evolution. Fungal mating occurs between gametes with different mating types which are not associated to male or female functions and little phenotypic differences are associated to mating-types, rending unlikely any antagonistic selection between mating types (*i.e.* with genes having alleles beneficial to one mating type but detrimental to the other; "mating-type antagonism"). Therefore, the existence of evolutionary strata on mating-type chromosomes would imply that alternative evolutionary mechanisms to antagonistic selection must be considered to explain the extension of recombination suppression. These mechanisms may then

play a role in sex chromosomes evolution as well. It remains to be investigated, however, whether fine differences may be associated to mating types, such as differences in gene expression level, whether genes with differential expression between mating types are enriched in regions linked to mating-type loci, which could indicate the existence of mating-type antagonistic selection driving recombination suppression. However, differential expression can also be due to degeneration in genomic regions with ancient recombination suppression. A means to test whether recombination suppression can be driven by mating-type antagonistic selection could thus be to assess whether genes with differential expression between mating types are enriched in young evolutionary strata.

In this thesis, I therefore contributed to address the following questions:

I.   Can evolutionary strata form in mating-type chromosomes in organisms with no male/female functions?
II. Did mating-type antagonistic selection contribute to the formation of such evolutionary strata?
III. Did bipolarity evolve multiple times independently from tetrapolarity in species with high selfing rates?
IV. Does differential expression of genes in non-recombining regions result from degeneration or from mating-type antagonistic selection?

Besides addressing these questions, I have been involved in two papers that can be found in Annexes. In a population genomic study conducted by Fanny Hartmann (post-doctoral researcher in the GEE team), I performed whole genome synteny analysis among several *Microbotryum* species. In a review conducted by Fanny Hartmann and Tatiana Giraud that reviewed the approaches and results obtained on the adaptation, coevolution and mating system evolution of *Microbotryum* fungi using genomic data, I wrote a section about the non-recombining regions on mating-type chromosomes. I am also collaborating with Marine Duhamel (PhD student), who has created a pipeline to detect de novo transposable elements.

2 Evolutionary strata on young mating-type chromosomes

Published in *PNAS* (July 3, 2017)

Sex chromosomes can display a large non-recombining region that can span almost the entire chromosome length (Bachtrog and Charlesworth 2001; Balounova et al. 2019). In some sex chromosomes, recombination has been suppressed in a stepwise manner, away from the supergene controlling the sex determination. The regions with different ages of recombination suppression are called evolutionary strata (Lahn and Page 1999). The most accepted theory so far postulates that evolutionary strata result from a selection favouring the linkage between the sex-determining genes and sexually antagonistic genes, i.e., with alleles beneficial in one sex that are detrimental in the other (Charlesworth and Charlesworth 1978; Nei 1969). However, the role of sexual antagonism in triggering the formation of evolutionary strata has received little support from empirical data, which greatly weakened the theory (Ironside 2010). Large non-recombining regions have been reported in several mating-type chromosomes of fungi. In basidiomycete fungi, mating-type chromosomes carry genes regulating mating compatibility, the PR and the HD genes. In species that mainly reproduce through intra-tetrad mating, the two mating-type genes can be linked by a recombination suppression as it increases the odds of compatibility between gametes. The existence of evolutionary strata had been suggested in some of these bipolar species without clear support (Menkis et al. 2008; Fraser et al. 2004; Votintseva and Filatov 2009). If evolutionary strata are observed on fungal mating-type chromosomes while there is no association between mating types and male/female functions, other mechanisms than sexual antagonism can explain why non-recombining regions extend stepwise (Ironside, 2010).

We investigated the existence of evolutionary strata on mating-type chromosomes with a large non-recombining region in the three closely related bipolar species *Microbotryum lychnidis-dioicae*, *M. silenes-dioicae* and *M. violaceum sensus stricto*. In sex chromosome evolution studies, evolutionary strata are identified by plotting the synonymous divergence calculated between the X-linked and Y-linked alleles along the X-chromosome gene order, as the X-chromosome retained the ancestral gene order due to recombination events in XX diploid individuals. In fungi, mating type is determined at the haploid stage so that both mating-type

chromosomes are always heterozygous and never recombine, which results in rearrangements in both mating-type chromosomes. To identify the ancestral gene order, we tried to find tetrapolar outgroups with syntenic mating-type chromosomes. We found that the two tetrapolar species *M. intermedium* and *M. lagerheimii* had collinear mating-type chromosomes despite their substantial phylogenetic distance; we therefore used as proxy of the ancestral state the gene order on the mating-type chromosomes from *M. lagerheimii*. We calculated the allelic synonymous divergence between a_1- and a_2-linked alleles from the three bipolar species and plotted them along the gene order from *M. lagerheimii*.

We identified multiple evolutionary strata in *M. lychnidis-dioicae* mating-type chromosomes. Three evolutionary strata involved mating-type genes, while five others did not. These findings show that evolutionary strata can form in the absence of sexual antagonism. Alternative hypotheses should therefore be considered to explain the formation of evolutionary strata in mating-type chromosomes, and they could act in sex chromosomes as well.

My contribution in this study has been to perform (i) the synonymous divergence analysis and (ii) the orthoMCL analysis to check mating-type chromosome synteny between mating types and between species. I also developed mating-type specific presence/absence markers to identify by PCR the allele at the PR mating-type locus - either a_1 or a_2 – on sporidia for identifying alternate mating types before genome sequencing.

Evolutionary strata on young mating-type chromosomes despite the lack of sexual antagonism

Sara Branco[a,1], Hélène Badouin[a,b,1], Ricardo C. Rodríguez de la Vega[a,1], Jérôme Gouzy[b], Fantin Carpentier[a], Gabriela Aguileta[a], Sophie Siguenza[b], Jean-Tristan Brandenburg[a], Marco A. Coelho[c], Michael E. Hood[d], and Tatiana Giraud[a,2]

[a]Ecologie Systématique Evolution, Univ. Paris Sud, AgroParisTech, CNRS, Université Paris-Saclay, 91400 Orsay, France; [b]Laboratoire des Interactions Plantes-Microorganismes, Université de Toulouse, Institut National de la Recherche Agronomique, CNRS, 31326 Castanet-Tolosan, France; [c]UCIBIO-REQUIMTE, Departamento de Ciências da Vida, Faculdade de Ciências e Tecnologia, Universidade NOVA de Lisboa, 2829-516 Caparica, Portugal; and [d]Department of Biology, Amherst College, Amherst, MA 01002-5000

Sex chromosomes can display successive steps of recombination suppression known as "evolutionary strata," which are thought to result from the successive linkage of sexually antagonistic genes to sex-determining genes. However, there is little evidence to support this explanation. Here we investigate whether evolutionary strata can evolve without sexual antagonism using fungi that display suppressed recombination extending beyond loci determining mating compatibility despite lack of male/female roles associated with their mating types. By comparing full-length chromosome assemblies from five anther-smut fungi with or without recombination suppression in their mating-type chromosomes, we inferred the ancestral gene order and derived chromosomal arrangements in this group. This approach shed light on the chromosomal fusion underlying the linkage of mating-type loci in fungi and provided evidence for multiple clearly resolved evolutionary strata over a range of ages (0.9–2.1 million years) in mating-type chromosomes. Several evolutionary strata did not include genes involved in mating-type determination. The existence of strata devoid of mating-type genes, despite the lack of sexual antagonism, calls for a unified theory of sex-related chromosome evolution, incorporating, for example, the influence of partially linked deleterious mutations and the maintenance of neutral rearrangement polymorphism due to balancing selection on sexes and mating types.

evolutionary strata | chromosomal rearrangements | fungi | genomic degeneration | mating-type chromosomes

Chromosomes carrying genes controlling mating compatibility may display suppressed recombination, as seen in sex chromosomes in some animals and plants (1, 2). When recombination is suppressed and one of the sex chromosomes is always in a heterozygous state (e.g., the Y chromosome in humans), extensive rearrangements of that chromosome typically occur. Our understanding of how sex chromosomes evolve remains incomplete, with many unanswered questions (3), including the reasons why these chromosomes display such large regions of suppressed recombination, extending well beyond the sex-determining genes. Sex chromosomes may bear discrete "evolutionary strata" of differentiation between alleles, with divergence decreasing with distance from sex-determining genes in the ancestral gene order, as inferred from the nonrearranged sex chromosome (1). It is generally accepted that such strata result from successive rounds of selection favoring the linkage of sexually antagonistic genes (with alleles beneficial in one sex but detrimental in the other) and sex-determining loci (1, 4). However, there is no definitive evidence to support a prominent role for sexually antagonistic genes in the formation of evolutionary strata in sex chromosomes (3). Other hypotheses have been put forward, including the permanent sheltering of deleterious alleles, which may accumulate in the margins of nonrecombining regions, provided that recombination rates are low in these regions. Another hypothesis is the fixation of neutral rearrangements

by drift in one of the two sex chromosomes with automatic recombination arrest and the maintenance of a polymorphic state due to balancing selection on sexes (2, 5–8) (*SI Appendix, SI Text and Fig. S1*).

Recombination suppression has been documented in fungal mating-type chromosomes (9, 10), which carry key loci determining mating compatibility. In basidiomycetes, mating type is typically controlled by two loci: (*i*) the pheromone receptor (PR) locus, which carries one gene encoding a mating-type–specific PR and one to several genes encoding pheromones, and (*ii*) the homeodomain gene (HD) locus, encoding two homeodomain-type transcription factors controlling postmating growth. Recombination suppression ensures full linkage of the mating-type genes within each of these two loci, which is required for correct mating-type determination (11), and typically does not extend beyond the mating-type genes (12, 13). Recombination suppression

Significance

Sex chromosomes can display divergent evolution, as seen in humans, in which the Y chromosome underlying maleness is smaller and contains much less information than the X chromosome. The differentiation between sex chromosomes can occur stepwise along their length, which is thought to result from the successive beneficial linkage of genes with different phenotype optima in the two sexes to sex-determining genes. However, there is little evidence to support this hypothesis. Here, we recovered ancestral chromosome structures and gathered evidence for stepwise differentiation between fungal mating-type chromosomes despite the absence of male/female roles. Our results suggest that the analogous features of sex chromosomes may not be due to differences in selection between males and females.

EVOLUTION

in fungal mating-type chromosomes can further link the two mating-type loci, HD and PR, to each other and/or the centromere (9, 14–18). Such linkage is beneficial in species with selfing-based mating systems, as it increases the odds of compatibility between gametes from a diploid parent (16, 19, 20) (*SI Appendix, SI Text and Fig. S2*). It has been repeatedly suggested that some fungal mating-type chromosomes have additional evolutionary strata extending the suppression of recombination beyond mating-type genes (9, 18, 21, 22) (*SI Appendix, SI Text*). However, no evidence has been provided of multiple discrete regions in which divergence between fungal mating-type–associated alleles decreases with increasing distance from sexual compatibility loci. Uncertainty about fungal evolutionary strata persists due to difficulties in inferring the ancestral gene order, which is essential for reconstructions of the evolution of mating-type chromosomes. If fungi have evolutionary strata not involving the linkage of mating-type genes, their presence cannot be explained by sexually antagonistic selection, as fungal mating types are not associated with male/female functions. Moreover, there are virtually no differences in ecological or life-history traits between fungal mating types (23), making it unlikely that there is any analogous mating-type–antagonistic selection improving the function of one mating type while having deleterious effects on the function of the other.

We investigated the existence of evolutionary strata in *Microbotryum lychnidis-dioicae*, a castrating anther-smut fungus. This species is particularly suited for determining whether the stepwise evolution of suppressed recombination can occur without sexual antagonism. It has two alternative mating-type chromosomes with a large region of suppressed recombination (10, 17, 24, 25), complete absence of male/female function, and virtually no opportunity for mating-type–antagonistic selection. Indeed, almost all matings occur within the tetrad (26–28) with the isogamous gametes fusing rapidly after meiosis without mitotic proliferation in a free haploid stage (Fig. 1*A*). There has been much debate about the existence of evolutionary strata in the *M. lychnidis-dioicae* mating-type chromosomes (17, 21, 22, 25) (*SI Appendix, SI Text*). Genes with different levels of synonymous divergence between alleles linked to the two mating types (called a$_1$ and a$_2$) are widely dispersed over 90% of the chromosome length, suggesting that, if there were any evolutionary strata, they have been erased by massive rearrangements (17). As in all fungi with distinct mating types, and in contrast to plant and animal sex chromosomes, *M. lychnidis-dioicae* mating-type chromosomes are always heterozygous in the diploid stage, so neither conserves the ancestral gene order (17). This precludes the typical use of the nonrearranged chromosome for the detection of evolutionary strata (1). The recombining regions in *M. lychnidis-dioicae*, such as the autosomes and the pseudoautosomal regions (PARs), are, however, highly collinear (17) (*SI Appendix,* Fig. S3), suggesting that the ancestral gene order should be retained in mating-type chromosomes undergoing recombination, even in selfing species. We use the term "autosome" as a contrast to "dimorphic" chromosomes, as originally intended (29, 30), rather than in opposition to sex chromosomes per se. Similarly, PAR is used to name the genomic regions at the edges of the mating-type chromosomes that are collinear and recombine like autosomes.

We compiled chromosome-length genome assemblies for five *Microbotryum* species and adopted an approach in which species carrying mostly recombining mating-type chromosomes were identified and used to recover the ancestral chromosomal state and gene order. Using this method, we were able to reconstruct the rearrangements leading to the current nonrecombining mating-type chromosomes in the selfing *M. lychnidis-dioicae*. One important step involved chromosomal rearrangements and fusion of the two separate ancestral mating-type chromosomes, linking the two mating-type loci in the same chromosome, which is beneficial under selfing (although not required) (*SI Appendix,*
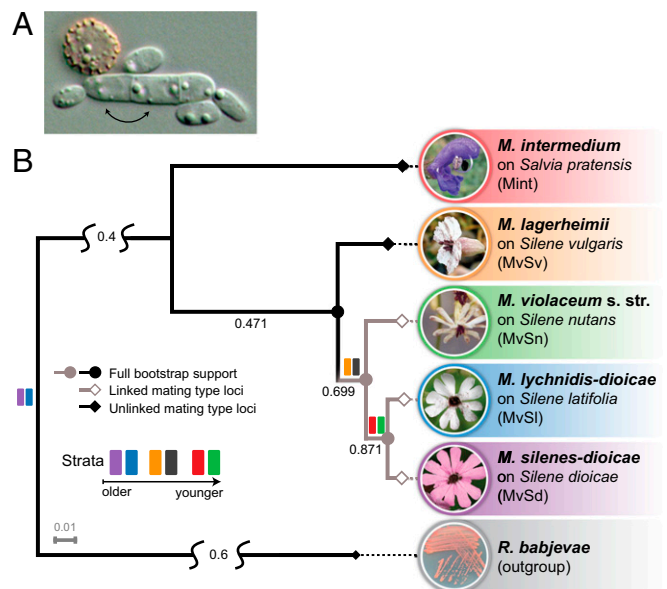

Fig. 1. Anther-smut fungi. (*A*) *M. lychnidis-dioicae* tetrad with a cell in the spore and three cells in a horizontal promycelium; mating occurs rapidly after meiosis within the tetrad (black arrows). Some haploid cells are budding, but almost 100% of matings occur within the promycelium. (*B*) Microbotryomycete phylogenetic tree based on 4,000 orthologous genes, including the studied *Microbotryum* species (shown in the anthers of their host plants) and the outgroup *Rhodotorula babjevae*. Branch color and symbol indicate linked (gray branches and open diamonds) or unlinked (black branches and closed diamonds) mating-type loci. The closed circles indicate full bootstrap support. Internode certainty with no conflict bipartitions is given below the branches (i.e., the normalized frequency of the most frequent bipartition across gene genealogies relative to the summed frequencies of the two most frequent bipartitions), indicating good support for the nodes. Relative tree certainty of the tree was 0.68. Colored rectangles along the branches indicate the evolution of the various evolutionary strata (see Fig. 3).

Fig. S2) (16). This finding sheds light on the genomic evolutionary steps leading to changes in breeding systems in fungi. We also provide compelling evidence for multiple, clearly differentiated and ancestrally adjacent evolutionary strata in fungal mating-type chromosomes with several strata devoid of genes controlling mating types. The existence of successive extensions of the region of recombination suppression beyond the linkage of mating-type genes in fungi without male/female functions indicates that evolutionary strata can occur without sexual antagonism. This finding has profound implications for our understanding of the evolutionary genomics of sexual eukaryotes across three kingdoms, highlighting the need for more serious consideration of alternative theoretical models of sex-related chromosome evolution beyond sexually antagonistic processes.

Results and Discussion

Ancestral Gene Order in *Microbotryum* Mating-Type Chromosomes.
We obtained high-quality assemblies of haploid genomes for two *Microbotryum* species expected to have retained the ancestral gene order: *Microbotryum intermedium*, a distant relative of *M. lychnidis-dioicae* (Fig. 1*B*) (31, 32), and *Microbotryum lagerheimii*, a species known to have unlinked HD and PR mating-type loci (32). As hypothesized, *M. intermedium* was found to have the PR and HD loci in two different chromosomes that were highly collinear with the mating-type chromosomes of *M. lagerheimii* (Fig. 2*A*). The collinearity between these distantly related species indicated a high degree of gene order conservation across multiple speciation events (31), therefore closely reflecting the
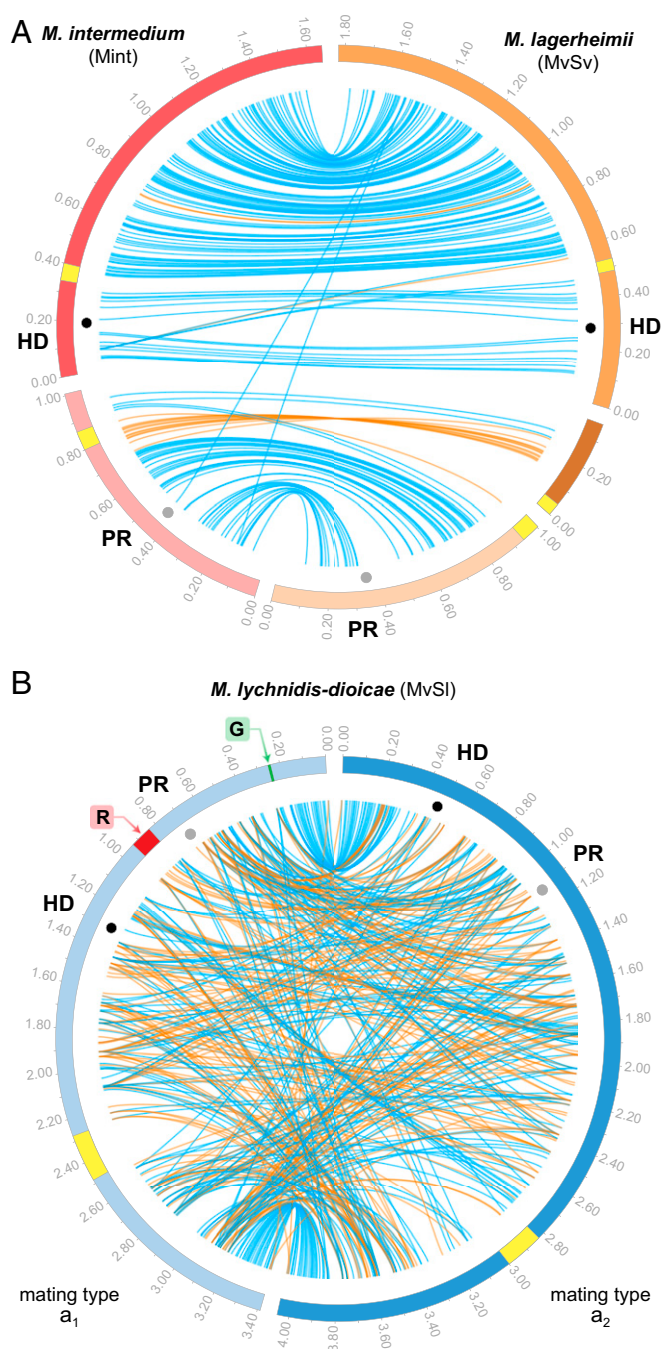
Fig. 2. Comparison of gene order across *Microbotryum* mating-type chromosomes. The mating-type HD and PR genes are indicated by small black and gray circles, respectively. Blue and orange lines link regions with collinearity extending over 2 kb, the latter corresponding to inversions. Areas without links correspond to highly rearranged regions. Yellow regions on the outer track indicate centromere-specific repeats (17). (*A*) Comparison of gene order between the mating-type chromosomes of *M. intermedium* (*Left*) and *M. lagerheimii* (*Right*). The darkest contig on the *Right* corresponds to MC16, as discussed in *SI Appendix, SI Text*. (*B*) Comparison of a$_1$ (light contig) and a$_2$ (dark contig) *M. lychnidis-doicae* mating-type chromosomes. The red and green genes in Fig. 4*B* are indicated in the a$_1$ genome.

ancestral state in this clade. The ancestral gene order has probably been maintained by regular recombination, as also indicated by the collinearity between the a$_1$ and a$_2$ mating-type chromosomes in *M. lagerheimii* (*SI Appendix*, Fig. S4), typical of recombining genomic regions (*SI Appendix*, Fig. S3).

Ancient Linkage of Mating-Type Genes at Each of the HD and PR Loci. We found footprints of ancient recombination suppression linking the essential pairs of interacting mating-type genes at each of the HD and PR loci, as is typical and ancestral in basidiomycete fungi (13). The nucleotide sequences of the a$_1$ and a$_2$ alleles of the PR gene were too differentiated to be aligned (Figs. 3*A* and 4) and a previous study inferred that these alleles had been diverging for about 370 million years (33). Recombination suppression linking the two HD mating-type genes was also very old (Fig. 3*B*), as shown by the second-highest level of synonymous divergence (d$_S$) between the a$_1$ and a$_2$ mating types at these genes in both *M. lagerheimii* and *M. lychnidis-dioicae* (Fig. 4) and by gene genealogies (34).

Chromosomal Fusion Led to the Large Nonrecombining Region Linking HD and PR Mating-Type Genes in *M. lychnidis-dioicae*. A more recent recombination suppression event involved the linkage of the HD and PR mating-type loci to each other (Fig. 3), which is favorable under selfing (15, 16) (*SI Appendix*, Fig. S2), the predominant mating system in *Microbotryum* (26, 28). Genome comparisons indicated that the *M. lychnidis-dioicae* mating-type chromosome was
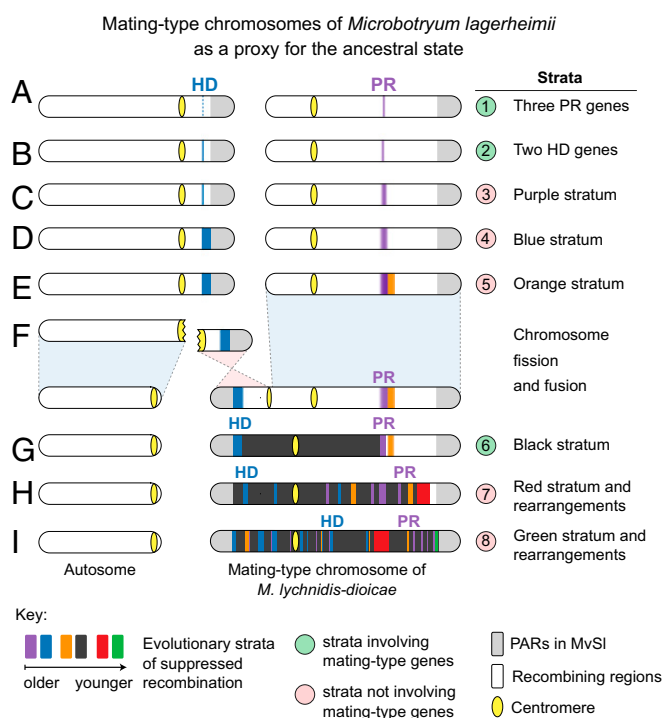
Fig. 3. Scenario of mating-type chromosome evolution in the genus *Microbotryum*. Strata involving the linkage of mating-type genes are numbered in green circles, whereas strata devoid of mating-type genes are numbered in red circles. (*A*) Ancestral state with separate HD and PR chromosomes and suppressed recombination restricted to the PR mating-type genes. (*B*) Suppressed recombination at all mating-type genes (as typically seen in basidiomycete fungi, with linked PR and pheromone genes at one locus and the two linked HD genes linked at a second locus). (*C*) Extension of suppressed recombination to purple nonmating-type genes. (*D*) Extension of suppressed recombination to blue nonmating-type genes. (*E*) Extension of suppressed recombination distal to the purple stratum, forming the orange stratum, devoid of mating-type genes. (*F*) Fission of the HD chromosome and fusion of one arm with the PR chromosome. (*G*) Complete linkage between the HD and PR mating-type loci, forming the black stratum. (*H*) Extension of suppressed recombination to nonmating-type genes forming the red stratum and rearrangements shuffling older evolutionary strata. (*I*) Further extension of suppressed recombination forming the green stratum and further rearrangements, leading to the current *M. lychnidis-dioicae* (MvSl) mating-type chromosome with the red stratum located between the PR and HD genes.

Fig. 4. Synonymous divergence (d_S) between alleles associated with the a_1 and a_2 mating types along mating-type chromosomes, as found in the PacBio-sequenced diploid individual in each *Microbotryum* species. Per-gene mean d_S values and SE values. The locations of the mating-type genes (PR and HD) are indicated. The divergence between the a_1 and a_2 PR alleles was too extensive (33) and could not be computed. It is plotted as 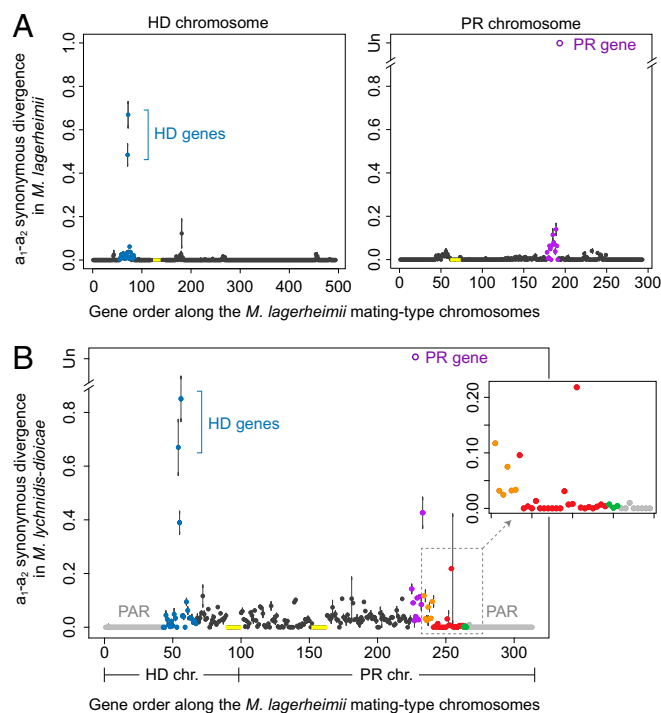an "unalignable" (Un) open purple circle. The *M. lagerheimii* centromeres are shown in yellow. (*A*) Synonymous divergence for all single-copy genes common to the *M. lagerheimii* a_1 and a_2 mating-type chromosomes plotted according to the genomic coordinates of its a_1 mating-type chromosomes. The HD chromosome is displayed on the left and the PR chromosome on the right. Genes with within-individual a_1–a_2 ds > 0 are shown in blue around HD and purple around PR. The mating-type loci are partially linked to the centromere in *M. lagerheimii* (32). (*B*) Synonymous divergence for all single-copy genes common to the *M. lychnidis-dioicae* mating-type chromosomes and *M. lagerheimii*, plotted in the ancestral gene order, as inferred from the genomic coordinates of the *M. lagerheimii* a_1 mating-type chromosomes. The limits of the *M. lagerheimii* PR and HD mating-type chromosomes are indicated, together with the positions of their centromeres. Blue and purple genes correspond to genes orthologous to the blue and purple regions in *A*. Other evolutionary strata are indicated in orange, black, red, and green. (*Inset*) Magnification of the youngest strata. The gray gene with nonzero d_S value at the right of the green stratum was located within the collinear PAR in Fig. 2*B* and was separated from the green genes by a small noncollinear region. The PARs in *M. lychnidis-dioicae* are shown in gray.

formed by the fusion of the complete ancestral PR chromosome with one arm of the ancestral chromosome carrying the HD locus (Fig. 3 *D–F* and *SI Appendix*, Fig. S5A). The remaining ancestral HD chromosome arm became an autosomal component in *M. lychnidis-dioicae* (Fig. 3 *D–F* and *SI Appendix*, Fig. S5A). Based on the inferred ancestral state, the two mating-type loci were initially some 800 kb apart following the fusion event and the fused chromosome was about 2 Mb long. The subsequent cessation of recombination between the PR and HD mating-type loci led to extensive rearrangements (*SI Appendix*, Fig. S3) and divergence between alleles associated with the a_1 and a_2 mating types in this region (in black in Fig. 4*B* and Fig. 3 *G–I*). Conversely, the chromosome regions ancestrally distal to the PR and HD loci largely constituted PARs that remained collinear (*SI Appendix*, Fig. S1), with a within-individual d_S of 0 (gray in Fig. 4*B*), as expected for recombining regions in organisms with high selfing rates.

For confirmation of the complete and ancient nature of recombination suppression in the region ancestrally between the HD and PR loci (black in Fig. 4*B*), we obtained and analyzed high-quality genome assemblies of *M. silenes-dioicae* and *M. violaceum s. str.*, two species closely related to *M. lychnidis-dioicae* (Fig. 1*B*). All three species had homologous mating-type chromosomes (*SI Appendix*, Fig. S5 *B* and *C*), each including the large black nonrecombining region ancestrally located between PR and HD, with a high degree of rearrangement between mating types and between species (*SI Appendix*, Fig. S5). Most of the genes (66%) in the black region displayed *trans*-specific polymorphism across *M. lychnidis-dioicae* and *M. silenes-dioicae*, with alleles associated with the a_1 mating type of both species branching together rather than clustering with the allele associated with the a_2 mating type from the same species. *Trans*-specific polymorphism also extended to the divergence from *M. violaceum s. str.* in 20% of the genes (*SI Appendix*, Figs. S6 and S7), which were dispersed along the ancestral gene order in the black stratum (corresponding to the black points with the highest d_S values in Fig. 4*B*). Such old *trans*-specific polymorphism associated with mating type could not have persisted without an ancient and complete cessation of recombination preceding the divergence of these three species, as even very low levels of recombination would have had time to break down the association between alleles at genes in the black stratum and mating-type genes. The less marked or absent *trans*-specific polymorphism for some of the genes within the black stratum probably results from occasional localized gene conversion events, as demonstrated for mating-type chromosomes in *Microbotryum* (35) and other fungi (36, 37). *Trans*-specific polymorphism in the genealogies of genes in the black stratum never extended to *M. lagerheimii*, consistent with the lack of linkage between the PR and HD mating-type loci in this species. When calibrated by the previous estimate for the date of the speciation event between *M. lychnidis-dioicae* and *M. silenes-dioicae* to 420,000 years ago (38), the divergence between a_1- and a_2-associated alleles in the black region led to an estimate for the date of linkage between the HD and PR loci of about 1.3 million years ago (95% confidence interval between 1.1 and 1.6 million years) (*SI Appendix, SI Text* and Table S1).

Evolutionary Strata Extending Suppressed Recombination Around Each of the PR and HD Mating-Type Loci. We detected further independent evolutionary strata, extending recombination suppression to nonmating-type genes in several successive steps (Fig. 3 *C* and *D* and *SI Appendix*, Table S1). Three strata (blue, purple, and orange strata, Figs. 1, 3, and 4), devoid of mating-type genes, evolved independently after the recombination suppression events at each of the PR and HD mating-type loci but before the HD and PR linkage via chromosome fusion (black stratum, Figs. 1, 3, and 4). Indeed, *M. lagerheimii*, which has unlinked HD and PR mating-type loci and therefore no black stratum, showed high divergence levels between alleles associated with the a_1 mating type and those associated with the a_2 mating type for 14 HD-proximal (blue) genes and 31 PR-proximal (purple) genes, none of which are involved in mating-type determination (Fig. 4*A* and *SI Appendix*, Tables S1, S2, and S3). These findings are in stark contrast with the near-zero within-individual d_S values along the rest of the mating-type chromosomes in this species. These same genes ancestrally located near the PR and HD loci also had the next-highest level of divergence after the mating-type genes themselves in *M. lychnidis-dioicae*, *M. silenes-dioicae*, and *M. violaceum s. str.* (purple and blue in Fig. 4*B*; *SI Appendix*, Tables S1 and S3 and Fig. S8). Most of these genes also displayed high levels of synonymous divergence between mating types in three other Microbotryomycetes (*SI Appendix*, Table S2 and figure 2*B* in ref. 39) and were also localized within noncollinear regions in these distant outgroups (figure 1*A* in ref. 39). These results

suggest that recombination suppression linking nonmating-type genes to key mating-type determinants is very old and occurred before these Microbotryomycetes species diverged (Fig. 1B).

Further supporting this inference, all genes in the purple stratum displayed *trans*-specific polymorphism with 50% displaying *trans*-specific polymorphism extending up to the divergence of *M. lychnidis-dioicae* and *M. lagerheimii* and 12.5% back to the root of the Microbotryomycetes (*SI Appendix*, Figs. S6 and S7). This confirms that a set of genes has been linked to the mating-type genes since long before the black stratum evolved, at least before the split between the common ancestor of *Microbotryum* and these outgroups. Indeed, such *trans*-specific polymorphisms separating alleles associated with the a_1 and a_2 mating types across species cannot be maintained without a full recombination suppression preceding all those speciation events. Additional strong evidence for an ancient and complete recombination cessation in the purple stratum is provided by the rearrangements between mating types present in *M. lagerheimii*, with the fully linked PR and pheromone genes of the a_1 genome located at opposite edges of the purple region (*SI Appendix*, Fig. S4). Recombination was suppressed in the purple stratum about 2.1 million years ago (95% confidence interval between 1.6 and 2.7 million years), as estimated based on the divergence between alleles associated with the alternative mating types (*SI Appendix*, SI Text and Table S1). This date confirms that the suppression of recombination predates the black stratum. We also found *trans*-specific polymorphism in the blue stratum in 75% of the genes (*SI Appendix*, Fig. S7). Recombination suppression in the purple and blue nonmating-type genes was more recent than that between the genes at each of the HD and PR mating-type loci, as shown by the lower d_S than for the mating-type genes and by the ancestral situation in basidiomycetes, with only mating-type genes linked at each of the PR and HD loci. Furthermore, the estimated dates of recombination suppression and *trans*-specific polymorphism in the purple stratum were much more recent than those for the PR gene (*SI Appendix*, Table S1 and Fig. S7) (33). The blue stratum may be more recent than the purple stratum, as suggested by its lower d_S values and less deep *trans*-specific polymorphism. However, even if equally ancient, the blue and purple strata constitute independent events of recombination suppression, given that they evolved in separate chromosomes, as seen in *M. lagerheimii* (Figs. 3 and 4).

We found another independent stratum ancestrally located distal to the purple stratum, toward the PARs (orange in Fig. 4B), with high d_S values in *M. lychnidis-dioicae*, *M. silenes-dioicae*, and *M. violaceum s. str.* (Fig. 4 and *SI Appendix*, Fig. S8). Genes in the orange stratum were not involved in mating-type determination and also displayed deep *trans*-specific polymorphism, indicating an ancient complete suppression of recombination (*SI Appendix*, Figs. S6 and S7). The orange stratum most likely evolved before the black stratum linking the HD and PR mating-type genes (Fig. 3E), as it showed a much higher mean d_S value (Figs. 4B and *SI Appendix*, Table S1). The orange stratum genes had zero d_S in *M. lagerheimii* (Fig. 4 and *SI Appendix*, Fig. S8), indicating that recombination suppression in this region evolved after the divergence with this species and therefore after the purple stratum (Fig. 3). Consistent with this inference, the genes of the orange stratum had lower d_S values than those of the purple stratum.

Additional, Younger Evolutionary Strata Evolved After the Linkage Between HD and PR Mating-Type Loci in *M. lychnidis-dioicae*. We found evidence for two additional and younger evolutionary strata in *M. lychnidis-dioicae*. These strata did not involve mating-type genes either and arose much later than the linkage between HD and PR mating-type genes (red and green in Fig. 4B and *SI Appendix*, Tables S1 and S3). The red stratum contains genes that were ancestrally located distal to the orange stratum

in the PAR and displayed intermediate d_S values in *M. lychnidis-dioicae* (*SI Appendix*, Table S1 and Fig. 4B, *Inset*). These findings indicate a further distal expansion of recombination suppression, after the evolution of the black stratum, in the lineage leading to *M. lychnidis-dioicae* (Fig. 3 H and I). The red stratum is currently located between the HD and PR loci in *M. lychnidis-dioicae* (Fig. 2B), where recombination has been shown to be completely suppressed using segregation analyses (10, 24). This additional recombination suppression step occurred before the divergence of *M. lychnidis-dioicae* and *M. silenes-dioicae*, as shown by the nonzero within-individual d_S values found in *M. silenes-dioicae* (*SI Appendix*, Fig. S8A) and the *trans*-specific polymorphism observed in almost 30% of the genes across these species (*SI Appendix*, Fig. S7). In *M. silenes-dioicae*, the red stratum genes have remained in close proximity but are separated from the PAR by a small noncollinear region in *M. silenes-dioicae* (*SI Appendix*, Fig. S5E). In contrast, the *trans*-specific polymorphism in the red stratum genes never extended to *M. violaceum s. str.* (*SI Appendix*, Fig. S7). In this later species, the red stratum genes were located in the PAR and had zero d_S values within the sequenced diploid strain (*SI Appendix*, Figs. S5 B and D and S7), demonstrating the regular occurrence of recombination. This indicates that the red stratum, which contains no mating-type genes, was formed more recently than the black stratum, in the lineage leading to *M. lychnidis-dioicae* and *M. silenes-dioicae*, after its divergence with *M. violaceum s. str.* (Fig. 1). We estimated the date of recombination suppression in the red stratum at about 0.9 millions years ago (95% confidence interval between 0.7 and 1.1 million years), which is much more recent than the black stratum linking the HD and PR loci (*SI Appendix*, SI Text and Table S1).

We found yet another small, very recent putative evolutionary stratum, also common only to *M. lychnidis-dioicae* and *M. silenes-dioicae* (green in Figs. 2B and 4B). In both species, this green stratum was located close to the PAR, but was separated from it by a region with an inversion between a_1 and a_2 (Fig. 2B and *SI Appendix*, Fig. S5E), and the genes had low but nonzero d_S values. These genes probably correspond to a very recent expansion of the nonrecombining region after the formation of the red stratum (Fig. 3I), as suggested by their lower d_S values (Fig. 4B and *SI Appendix*, Table S1), distinctive current location in the *M. lychnidis-dioicae* genome, and lack of *trans*-specific polymorphism (*SI Appendix*, Fig. S7). Alternatively, the red and green strata may have been generated in a single event, followed by physical separation due to rearrangements.

We further confirmed recombination suppression in the red and green strata using multiple available resequenced genomes for *M. lychnidis-dioicae* and *M. silenes-dioicae* (22, 40). Gene genealogies revealed that the alleles associated with the a_1 and a_2 mating types formed two different clades for at least one species in all of the gene genealogies, and *trans*-specific polymorphism was observed across all individuals in 23% of cases (*SI Appendix*, Fig. S9). These patterns can be explained only by full linkage to mating type before the two species split. Furthermore, the mean levels of polymorphism per mating type and per species were significantly lower in all of the evolutionary strata than in the PARs (Student's *t* tests, $P < 0.05$) (*SI Appendix*, Fig. S10), as expected in regions without recombination due to the lower effective population size. These findings indicated that the nonzero d_S in the evolutionary strata within the sequenced individuals was due to recombination suppression rather than elevated substitution rates. Maximum-likelihood Hudson–Kreitman–Agade tests further supported that the higher d_S levels in the purple, blue, orange, black, and red strata compared with the PARs in *M. lychnidis-dioicae* and *M. silenes-dioicae* were due to balancing selection (because of linkage to mating type) rather than to elevated mutation rates (*SI Appendix*, SI Text).

Conclusion. We found multiple evolutionary strata in *Microbotryum* mating-type chromosomes, several of which included no genes related to mating-type determination (*SI Appendix*, Tables S2 and S3). Only the two oldest strata (linking pairs of mating-type genes within each of the separate PR and HD loci, respectively) and the black stratum (linking the HD and PR loci together) involved mating-type genes. The linkage between HD and PR genes probably evolved because it increased the likelihood of compatibility between gametes in a selfing system (16) (*SI Appendix, SI Text* and Fig. S2). Mating-type loci linkage has been reported in several fungi (9, 14, 41, 42), and our study provides clues to the chromosomal rearrangements underlying this phenomenon. The five other *M. lychnidis-dioicae* evolutionary strata (purple, blue, orange, red, and green strata) did not involve mating-type genes. As this species has no male/female traits, our results indicate that sexual antagonism is insufficient to account for the occurrence of these evolutionary strata. Furthermore, given the *M. lychnidis-dioicae* life cycle, the existence of ecological differences between mating types that would enhance alternative mating-type functions is highly unlikely. Alternative hypotheses should therefore be considered, including the capture and shelter of deleterious alleles in a permanently heterozygous state, the fixation of neutral rearrangements by drift in one gametolog (see *SI Appendix, SI Text* and Fig. S1 for more details concerning these hypotheses). These general mechanisms apply to both sex and mating-type chromosomes across the plant, animal, and fungal kingdoms. Our results thus call for an expansion of theories concerning the evolution of eukaryotic sex-related chromosomes to forces beyond sexual antagonism.

Materials and Methods

Haploid genomes were sequenced with P6/C4 Pacific Biosciences SMRT technology (University of California, San Diego IGM Genomics Facility) (*SI Appendix*, Table S4). Orthologs were identified by orthomcl (43). Trees were inferred with RAxML (44). Strata divergence times were inferred with BEAST2 (45). More details about the materials and methods used are provided in *SI Appendix, SI Text*.

1. Bergero R, Charlesworth D (2009) The evolution of restricted recombination in sex chromosomes. *Trends Ecol Evol* 24:94–102.
2. Charlesworth D (2015) The status of supergenes in the 21st century: Recombination suppression in Batesian mimicry and sex chromosomes and other complex adaptations. *Evol Appl* 9:74–90.
3. Wright AE, Dean R, Zimmer F, Mank JE (2016) How to make a sex chromosome. *Nat Commun* 7:12087.
4. Charlesworth B (1991) The evolution of sex chromosomes. *Science* 251:1030–1033.
5. Ironside JE (2010) No amicable divorce? Challenging the notion that sexual antagonism drives sex chromosome evolution. *BioEssays* 32:718–726.
6. Charlesworth B, Wall JD (1999) Inbreeding, heterozygote advantage and the evolution of neo-X and neo-Y sex chromosomes. *Proc Biol Sci* 266:51–56.
7. Antonovics J, Abrams JY (2004) Intratetrad mating and the evolution of linkage relationships. *Evolution* 58:702–709.
8. Johnson LJ, Antonovics J, Hood ME (2005) The evolution of intratetrad mating rates. *Evolution* 59:2525–2532.
9. Menkis A, Jacobson DJ, Gustafsson T, Johannesson H (2008) The mating-type chromosome in the filamentous ascomycete *Neurospora tetrasperma* represents a model for early evolution of sex chromosomes. *PLoS Genet* 4:e1000030.
10. Hood ME (2002) Dimorphic mating-type chromosomes in the fungus *Microbotryum violaceum*. *Genetics* 160:457–461.
11. Kües U (2000) Life history and developmental processes in the basidiomycete *Coprinus cinereus*. *Microbiol Mol Biol Rev* 64:316–353.
12. Coelho MA, Bakkeren G, Sun S, Hood ME, Giraud T (2017) Fungal sex: The basidiomycota. *Microbiol Spectr*, 10.1128/microbiolspec.FUNK-0046-2016.
13. Kües U (2015) From two to many: Multiple mating types in Basidiomycetes. *Fungal Biol Rev* 29:126–166.
14. Bakkeren G, Kronstad JW (1994) Linkage of mating-type loci distinguishes bipolar from tetrapolar mating in basidiomycetous smut fungi. *Proc Natl Acad Sci USA* 91:7085–7089.
15. Idnurm A, Hood ME, Johannesson H, Giraud T (2015) Contrasted patterns in mating-type chromosomes in fungi: Hotspots versus coldspots of recombination. *Fungal Biol Rev* 29:220–229.
16. Nieuwenhuis BPS, et al. (2013) Evolution of uni- and bifactorial sexual compatibility systems in fungi. *Heredity (Edinb)* 111:445–455.
17. Badouin H, et al. (2015) Chaos of rearrangements in the mating-type chromosomes of the anther-smut fungus *Microbotryum lychnidis-dioicae*. *Genetics* 200:1275–1284.
18. Fraser JA, et al. (2004) Convergent evolution of chromosomal sex-determining regions in the animal and fungal kingdoms. *PLoS Biol* 2:e384.
19. James TY (2015) Why mushrooms have evolved to be so promiscuous: Insights from evolutionary and ecological patterns. *Fungal Biol Rev* 29:167–178.
20. Giraud T, Yockteng R, López-Villavicencio M, Refrégier G, Hood ME (2008) Mating system of the anther smut fungus Microbotryum violaceum: Selfing under heterothallism. *Eukaryot Cell* 7:765–775.
21. Votintseva AA, Filatov DA (2009) Evolutionary strata in a small mating-type-specific region of the smut fungus *Microbotryum violaceum*. *Genetics* 182:1391–1396.
22. Whittle CA, Votintseva A, Ridout K, Filatov DA (2015) Recent and massive expansion of the mating-type-specific region in the smut fungus *Microbotryum*. *Genetics* 199:809–816.
23. Billiard S, et al. (2011) Having sex, yes, but with whom? Inferences from fungi on the evolution of anisogamy and mating types. *Biol Rev Camb Philos Soc* 86:421–442.
24. Hood ME, Antonovics J, Koskella B (2004) Shared forces of sex chromosome evolution in haploid-mating and diploid-mating organisms: *Microbotryum violaceum* and other model organisms. *Genetics* 168:141–146.
25. Hood ME, Petit E, Giraud T (2013) Extensive divergence between mating-type chromosomes of the anther-smut fungus. *Genetics* 193:309–315.
26. Hood ME, Antonovics J (2000) Intratetrad mating, heterozygosity, and the maintenance of deleterious alleles in *Microbotryum violaceum* (=*Ustilago violacea*). *Heredity (Edinb)* 85:231–241.
27. Hood ME, Antonovics J (2004) Mating within the meiotic tetrad and the maintenance of genomic heterozygosity. *Genetics* 166:1751–1759.
28. Giraud T, Jonot O, Shykoff JA (2005) Selfing propensity under choice conditions in a parasitic fungus, *Microbotryum violaceum*, and parameters influencing infection success in artificial inoculations. *Int J Plant Sci* 166:649–657.
29. Montgomery T (1911) Are particular chromosomes sex determinants? *Biol Bull* 19:1–17.
30. Montgomery TH, Jr (1906) The terminology of aberrant chromosomes and their behavior in certain Hemiptera. *Science* 23:36–38.
31. Kemler M, Göker M, Oberwinkler F, Begerow D (2006) Implications of molecular characters for the phylogeny of the Microbotryaceae (Basidiomycota: Urediniomycetes). *BMC Evol Biol* 6:35.
32. Hood ME, Scott M, Hwang M (2015) Breaking linkage between mating compatibility factors: Tetrapolarity in *Microbotryum*. *Evolution* 69:2561–2572.
33. Devier B, Aguileta G, Hood ME, Giraud T (2009) Ancient trans-specific polymorphism at pheromone receptor genes in basidiomycetes. *Genetics* 181:209–223.
34. Petit E, et al. (2012) Linkage to the mating-type locus across the genus *Microbotryum*: Insights into nonrecombining chromosomes. *Evolution* 66:3519–3533.
35. Fontanillas E, et al. (2015) Degeneration of the nonrecombining regions in the mating-type chromosomes of the anther-smut fungi. *Mol Biol Evol* 32:928–943.
36. Menkis A, Whittle CA, Johannesson H (2010) Gene genealogies indicates abundant gene conversions and independent evolutionary histories of the mating-type chromosomes in the evolutionary history of *Neurospora tetrasperma*. *BMC Evol Biol* 10:234.
37. Sun S, Hsueh Y-P, Heitman J (2012) Gene conversion occurs within the mating-type locus of *Cryptococcus neoformans* during sexual reproduction. *PLoS Genet* 8:e1002810.
38. Gladieux P, et al. (2011) Maintenance of fungal pathogen species that are specialized to different hosts: Allopatric divergence and introgression through secondary contact. *Mol Biol Evol* 28:459–471.
39. Maia TM, et al. (2015) Evolution of mating systems in Basidiomycetes and the genetic architecture underlying mating-type determination in the yeast *Leucosporidium scottii*. *Genetics* 201:75–89.
40. Badouin H, et al. (2017) Widespread selective sweeps throughout the genome of model plant pathogenic fungi and identification of effector candidates. *Mol Ecol* 26:2041–2062.
41. Fraser JA, Heitman J (2005) Chromosomal sex-determining regions in animals, plants and fungi. *Curr Opin Genet Dev* 15:645–651.
42. Grognet P, et al. (2014) Maintaining two mating types: Structure of the mating type locus and its role in heterokaryosis in *Podospora anserina*. *Genetics* 197:421–432.
43. Li L, Stoeckert CJ, Jr, Roos DS (2003) OrthoMCL: Identification of ortholog groups for eukaryotic genomes. *Genome Res* 13:2178–2189.
44. Stamatakis A (2006) RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22:2688–2690.
45. Bouckaert R, et al. (2014) BEAST 2: A software platform for Bayesian evolutionary analysis. *PLOS Comput Biol* 10:e1003537.

3 Absence of antagonistic selection in *Microbotryum lychnidis-dioicae*

The previous chapter provides evidence for the formation of multiple evolutionary strata in the *Microbotryum* genus. There is no room for sexual antagonism to occur because *Microbotryum* fungi do not display male/female functions; the theory based on sexual antagonism therefore cannot explain the formation of evolutionary strata on their mating-type chromosomes. However, one may argue that antagonistic selection between mating types ("mating-type antagonism" *sensu* Abbate et al. 2010) could occur, even if not obvious based on the life cycle, and could explain the formation of evolutionary strata: selection would favour the linkage between mating-type determining genes and alleles beneficial to one mating-type but detrimental to the other. Given the few phenotypic or ecological traits associated to mating types in *Microbotryum* species, the mating-type antagonism hypothesis is unlikely to explain the formation of evolutionary strata on mating-type chromosomes. However, a previous study found differential gene expression between mating types in *Microbotryum lychnidis-dioicae* (Fontanillas et al. 2015); although this could be due solely to genomic degeneration, there remained uncertainties about the existence of mating-type antagonism and its possible role in driving evolutionary strata.

The present study therefore aimed at investigating whether mating-type antagonism could have triggered the emergence of evolutionary strata. If they exist, mating-type antagonistic genes were expected to (i) be down-regulated in dikaryotic cells relatively to haploid cells expressing a single mating type, (ii) be differentially expressed between mating types and/or (iii) show differentiation at the protein sequence level, and even possibly positive selection between mating types. Therefore, we used *Microbotryum lychnidis-dioicae*, for which expression data were available, to test whether its evolutionary strata were enriched in such differentially expressed genes compared to recombining regions (*i.e.* PARs and autosomes).

My contribution to this study was to (i) perform the expression level estimation and the differential expression analysis, (ii) perform the d_N/d_S calculation, (iii) participate in manuscript

writing. This study results from a collaboration with Anna Bazzicalupo (Montana State University, USA) and Sarah Perin Otto (University of British Columbia, Canada).

Little Evidence of Antagonistic Selection in the Evolutionary Strata of Fungal Mating-Type Chromosomes (*Microbotryum lychnidis-dioicae*)

Anna Liza Bazzicalupo,*[,1,2] Fantin Carpentier,[†,2] Sarah Perin Otto,[‡] and Tatiana Giraud[†,3]

*Department of Botany, 3200-6270 University Blvd., University of British Columbia, Vancouver, BC V6T 1Z4, Canada, [†]Ecologie Systématique Evolution, Univ. Paris-Sud, CNRS, AgroParisTech, Université Paris-Saclay, 91400 Orsay, France, and [‡]Department of Zoology & Biodiversity Research Centre, 6270 University Blvd., University of British Columbia, Vancouver, BC V6T 1Z4, Canada

ORCID IDs: 0000-0001-5845-9517 (A.L.B.); 0000-0003-2985-159 (F.C.); 0000-0003-3042-0818 (S.P.O.); 0000-0002-2685-6478 (T.G.)

ABSTRACT Recombination suppression on sex chromosomes often extends in a stepwise manner, generating evolutionary strata of differentiation between sex chromosomes. Sexual antagonism is a widely accepted explanation for evolutionary strata, postulating that sets of genes beneficial in only one sex are successively linked to the sex-determining locus. The anther-smut fungus *Microbotryum lychnidis-dioicae* has mating-type chromosomes with evolutionary strata, only some of which link mating-type genes. Male and female roles are non-existent in this fungus, but mating-type antagonistic selection can also generate evolutionary strata, although the life cycle of the fungus suggests it should be restricted to few traits. Here, we tested the hypothesis that mating-type antagonism may have triggered recombination suppression beyond mating-type genes in *M. lychnidis-dioicae* by searching for footprints of antagonistic selection in evolutionary strata not linking mating-type loci. We found that these evolutionary strata (i) were not enriched in genes upregulated in the haploid phase, where cells are of alternative mating types, (ii) carried no gene differentially expressed between mating types, and (iii) carried no genes displaying footprints of specialization in terms of protein sequences (d_N/d_S) between mating types after recommended filtering. Without filtering, eleven genes showed signs of positive selection in the strata not linking mating-type genes, which constituted an enrichment compared to autosomes, but their functions were not obviously involved in antagonistic selection. Thus, we found no strong evidence that antagonistic selection has contributed to extending recombination suppression beyond mating-type genes. Alternative hypotheses should therefore be explored to improve our understanding of the sex-related chromosome evolution.

Recombination between different genotypes can generate beneficial allelic combinations and purge deleterious mutations (Otto 2009).

Paradoxically, the genomic regions involved in promoting such genetic exchanges are often substantially excluded from these benefits, having evolved recombination suppression (Idnurm *et al.* 2015). Examples of such genomic regions include the sex chromosomes determining male and female phenotypes in many plants and animals, self-incompatibility loci in plants, and mating-type loci in fungi (Uyenoyama 2005; Beukeboom and Perrin 2014; Idnurm *et al.* 2015; Charlesworth 2016). The suppression of recombination in these genomic regions maintains the allelic combinations required for correct sex or mating-type determinism by linking, for example, pheromone and pheromone receptor alleles. This represents a case of beneficial allelic associations of multiple genes through linkage, more generally called "supergenes" (Charlesworth 2016; Branco *et al.* 2018). On sex-related chromosomes (considered generically to include chromosomes bearing mating-type genes, as in the

Microbotryum fungus studied here), recombination suppression often extends well beyond the genes involved in sex determination, which is puzzling, given the reduced efficiency of selection in the absence of recombination (Marais *et al.* 2008; Otto 2009). The lack of recombination is consequently thought to lead to the accumulation of deleterious alleles through Muller's ratchet (Charlesworth and Charlesworth 1997; Immler and Otto 2015) and contribute to the degeneration of the non-recombining chromosome in many species (*e.g.*, Marais *et al.* 2008; Otto 2009). Extensive differentiation between the non-recombining sex chromosomes (*e.g.*, between the X and Y) often arises through a series of recombination suppression steps. With each round of recombination suppression, an "evolutionary stratum" is generated, with the amount of genetic differentiation between the sex chromosomes within the region recording the time since the cessation of genetic exchange.

Understanding the nature of selection underlying the stepwise differentiation of sex chromosomes is a field of active research in evolutionary biology (Beukeboom and Perrin 2014; Wright *et al.* 2016). Sexually antagonistic selection, in which selection favors different alleles in the two sexes, is thought to be a particularly important driver of sex chromosome differentiation and recombination suppression (Bergero and Charlesworth 2009; Charlesworth 2016; Wright *et al.* 2016). However, empirical support for the hypothesis that sexually antagonistic genes accumulate near sex-determining regions, causing the successive steps of recombination suppression for their linkage to sex-determining loci, is slim, despite numerous studies in diverse plant and animal systems (Beukeboom and Perrin 2014; Wright *et al.* 2016). This lack of support may reflect the challenge of identifying sexually antagonistic loci (*e.g.*, Kasimatis *et al.* 2017; Mank 2017) and/or the relative lack of studies focusing on young sex chromosomes, where the signals of selection are not yet affected by degeneration.

Alternative hypotheses have also been put forward to explain the progressive expansion of recombination suppression between the sex chromosomes (Ironside 2010; Ponnikas *et al.* 2018). This may involve other forms of selection besides sexually antagonistic selection, such as conflicting selection pressures between haploid and diploid phases (Scott and Otto 2017) and heterozygote advantage (Antonovics and Abrams 2004; Otto 2014; Immler and Otto 2015). In a recent study on *Rumex*, for example, sex-linked genes were found to be more highly expressed in the haploid (pollen) phase, suggesting that conflicts between haploid and diploid phases may drive the evolution of sex chromosomes (Sandler *et al.* 2018). In inbreeding organisms, like the *Microbotryum* fungus studied here, linkage to the sex- or mating-type-determining region is a potent mechanism to preserve heterozygosity, which can favor recombination suppression (Charlesworth and Wall 1999). On the other hand, such inbreeding restricts the conditions under which a polymorphism can be maintained (*e.g.*, Charlesworth and Wall 1999), suggesting that selectively-driven differentiation of sex-related chromosomes may be rare in such inbred organisms. As an alternative to selective explanations, neutral inversions may accumulate near sex-determining regions (Ironside 2010; Ponnikas *et al.* 2018), or reduced recombination may occur as a side-consequence of silencing transposable elements that spread in and near sex-determining regions (Kent *et al.* 2017).

Most theories about the evolution of sex chromosomes are based on studies of organisms in which sex is determined at the diploid stage (*e.g.*, animals) (Bergero and Charlesworth 2009; Charlesworth 2016; Wright *et al.* 2016). However, the suppression of recombination in sex-determining regions can also occur in organisms in which sex is determined at the haploid stage, such as algae or bryophytes (Coelho *et al.* 2018). A progressive spread of recombination

suppression on sex chromosomes can occur when there is antagonistic selection between the two sexes (Immler and Otto 2015). In principle, this could also apply to fungi, in which mating type is controlled at the haploid stage, provided that there are traits for which optima differ between mating types. However, there may be few adaptive differences between haploid cells of different mating types other than mating-type determination itself and sometimes mitochondrial inheritance (Xu 2005; Billiard *et al.* 2011). Differences in gene expression levels have been found between haploid cells of different mating types in some fungi (Samils *et al.* 2013; Grognet *et al.* 2014; Fontanillas *et al.* 2015). However, such differences in expression levels between fungal mating types may be due to degeneration following recombination suppression, rather than adaptive differences between the mating types (Fontanillas *et al.* 2015).

The anther-smut fungus *Microbotryum lychnidis-dioicae* has mating-type chromosomes with large regions devoid of recombination. The cessation of recombination occurred in several successive steps, with six different evolutionary strata dating from 0.9 to 2.1 million years ago. One of the evolutionary strata, known as the black stratum (Branco *et al.* 2017), evolved to link the two mating-type loci controlling pre- and post-mating compatibility, respectively (Figure 1). Such linkage is beneficial when organisms undergo selfing as their main mode of sexual reproduction, as it maximizes the chances of compatibility among the haploid products of meiosis from a single diploid genotype: only two mating types are produced in a progeny with linked mating-type loci against four mating types with unlinked mating types (Nieuwenhuis *et al.* 2013; Branco *et al.* 2017). The other evolutionary strata, all given names based on colors (Branco *et al.* 2017), occurred in successive steps and link genes not involved in mating-type determination to the mating-type genes (Figure 1). The purple, blue and orange strata predate the black stratum, and occurred while the mating-type loci were still located on different chromosomes, at the basis of the *Microbotryum* clade, while the red and green strata are younger than the event linking the two mating type loci through the black stratum (Branco *et al.* 2017). A similar stepwise progression of recombination suppression along mating-type chromosomes also occurred independently in other *Microbotryum* fungi, trapping different gene sets (Branco *et al.* 2018). Only small recombining regions remained at both ends of mating-type chromosomes, called pseudo-autosomal regions (PARs).

These fungi have no 'female' or 'male' functions (displaying isogamous gametes), and sexual antagonism should not, therefore, have driven the development of these evolutionary strata. In addition, the selfing mating system in these fungi may render challenging to maintain sexually antagonistic variation, if any (Gregorius 1982; Charlesworth and Wall 1999; Jordan and Connallon 2014). Furthermore, the brevity of the haploid stage suggests that there would be little opportunity for ploidally antagonistic selection (Immler and Otto 2015).

Antagonistic selection could instead act between haploid mating types, but there is currently little evidence for such antagonism in *Microbotryum* fungi. Our understanding of the life cycle (Figure 2) suggests that the haploid phase is very brief which limits the possibility of mating-type specific antagonistic selection or ploidally antagonistic selection (*i.e.*, differential selection between haploid and dikaryotic phases). Most mating events occur rapidly after meiosis, between products of the same meiotic tetrad, causing high levels of inbreeding (Hood and Antonovics 1998; Hood and Antonovics 2000; Hood and Antonovics 2004; Schäfer *et al.* 2010). Mating generates infectious dikaryotic (n+n) hyphae that penetrate the plant. While haploid sporidia can multiply *in vitro* on media with a high sugar content and in flower
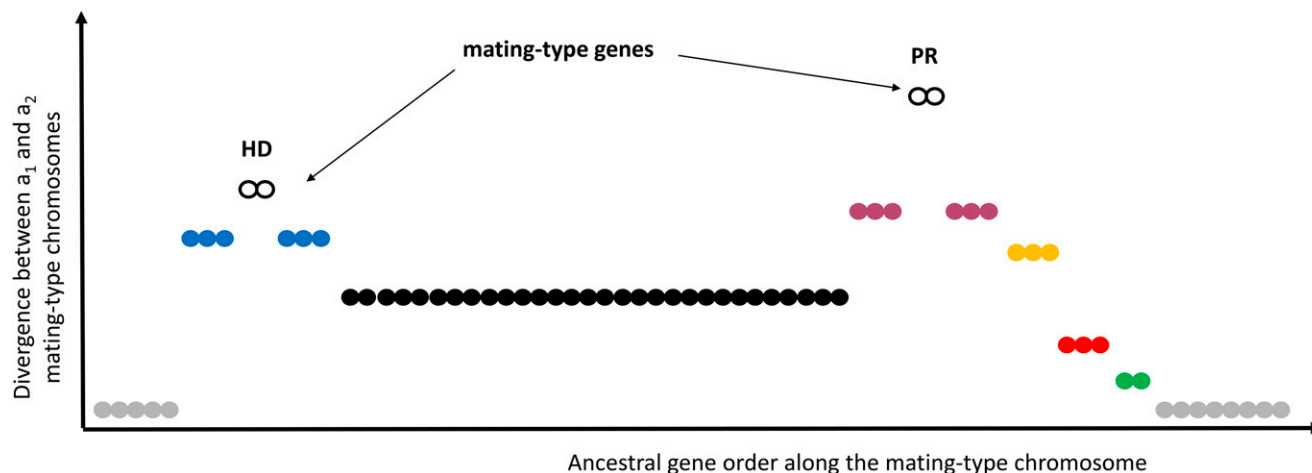
Figure 1 Schematic representation of the evolutionary strata on the mating-type chromosome of *Microbotryum lychnidis-dioicae*. The per-gene synonymous divergence between alleles (y-axis) represents relative timing of the suppression of recombination steps plotted along the ancestral gene order (x-axis). The PR and HD gene clusters (open black circles) show the most ancient divergences. They control pre- and post-mating compatibility, respectively, and encompass several ancestrally linked mating-type genes. The sequence of suppression of recombination begins around each of the mating-type loci, generating the blue and purple evolutionary strata. Recombination suppression then spread distally to the PR locus, creating the orange stratum. The event that linked the two mating-type loci and their surrounding strata generated the black stratum. The suppression of recombination then spread further outwards distal to the PR locus, creating the red and then the green strata. The pseudo-autosomal regions, which are still recombining, are shown in gray. Only the evolutionary strata shown in black (open or closed black circles) involve linking mating-type genes.

nectar (Schäfer *et al.* 2010; Golonka and Vilgalys 2013), it remains unclear whether this stage is of any biological relevance (Hood and Antonovics 2000; Hood and Antonovics 2004; Granberg *et al.* 2008). Dikaryotic hyphae are not thought to invade the plant via flowers but rather by penetration at the junction of the anticlinal epidermis on vegetative tissues with low sugar content (Schäfer *et al.* 2010). It is often the case that spore-bearing anthers are present early in the season in the first flower (Alexander and Maltby 1990; Alexander and Antonovics 1995), implying that infection occurred at the vegetative stage prior to the development of the first flower. Furthermore, male plants develop disease at least as frequently as female plants (Zillig 1921; Alexander 1989; Alexander 1990; Thrall and Jarosz 1994; Alexander and Antonovics 1995; Biere and Antonovics 1996; Biere and Honders

1998), despite the fact that male flowers fall rapidly after pollinator visits (Kaltz and Shykoff 2001), consistent with the view that infection does not occur through floral tissues where haploid cells might grow.

Most tellingly, alleles that are lethal in the haploid phase and linked to mating-type loci have been found in up to 50% of *M. lychnidis-dioicae* strains in natural populations (Kaltz and Shykoff 1997; Oudemans *et al.* 1998; Thomas *et al.* 2003). The maintenance of such alleles can only be explained by a lack of the free-living haploid phase in nature and mating within the tetrad. Consistent with this inference, estimates of selfing rates are extremely high in natural populations, at about 95% (Giraud *et al.* 2005; Vercken *et al.* 2010), as shown by the almost complete homozygosity of the autosomes (Vercken *et al.* 2010; Branco *et al.* 2018). The limited nature of a free-living haploid phase
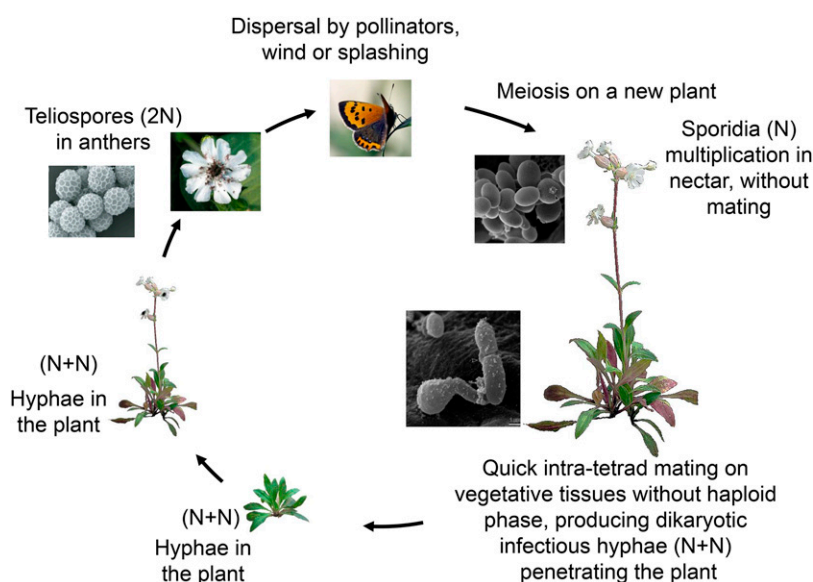
Figure 2 Life cycle of the anther-smut fungus *Microbotryum lychnidis-dioicae*. Diploid (2N) teliospores are produced in anthers of diseased plants, being heterozygous at all mating-type genes. The teliospores are dispersed to healthy plants by pollinators, wind or splashing. Once on a new plant, teliospores undergo meiosis. In the nectar of flowers, haploid sporidia (N) multiply clonally without mating until the flower wilts, after which flowers fall in male plants. On vegetative tissues where teliospores fall from flowers or by splashing, quick intra-tetrad mating occurs, preventing any haploid phase, producing dikaryotic infectious hyphae (N+N) penetrating the plant. In dikaryotic hyphae, there is exactly one nucleus of each mating type in each cell, preventing competition between mating types for replication and transmission. The flowers produced from infected meristems will produced diseased flowers. Pictures from López-Villavicencio *et al.* (2007) and Schäfer *et al.* (2010) © Canadian Science Publishing or its licensors.

in *Microbotryum* fungi raises a challenge to any hypothesis for the successive spread of suppressed recombination on the sex chromosomes that relies on antagonistic selection (given that the mating types are only separate in the haploid phase) or ploidally antagonistic selection.

Even if there is a free-living haploid phase, few traits other than those determined by mating-type genes may make a "better a_1" but a "worse a_2", or vice versa, which would be a prerequisite for recombination suppression by antagonistic selection. Similarly, we hypothesized that there would be little evidence for alleles favored in one of the mating types but disfavored in the dikaryotic phase, a prerequisite for recombination suppression by ploidally antagonistic selection (Immler and Otto 2015). Nevertheless, even transient selection acting on a few loci subject to antagonistic selection involving the haploid phase could drive the spread of suppressed recombination on the mating-type chromosomes. For example, inheritance of mitochondria may be asymmetric between mating types in *M. lychnidis-dioicae*, although not completely uniparental (Wilch *et al.* 1992) where the fungus was still named *Ustilago violacea*), potentially introducing selection in the haploid phase. We thus examined footprints of selection in *M. lychnidis-dioicae* for signs of antagonistic selection between the mating types or between the haploid and the dikaryotic phase, focusing on genes belonging to the evolutionary strata that are not involved in mating-type locus linkage, known as the "color strata" (Branco *et al.* 2017) (Figure 1). For the "black" stratum that evolved for linking mating-type loci (Figure 1), no further evolutionary explanations are required. Genes under antagonistic selection between mating types would, by definition, have important and different roles in the haploid a_1 or a_2 mating-type cells. Such genes could therefore be expressed at higher levels in the haploid stage (in at least one mating type) than in the dikaryotic stage (heterozygous for mating type). It is indeed common that genes important in a particular life stage are upregulated in that stage, especially in fungi, in which, for example, genes involved in pathogenicity are often upregulated in the host plant (Seitner *et al.* 2018; Zhang *et al.* 2018). If genes undergoing antagonistic selection drive stratum evolution beyond mating-type genes, these genes should be present in the "color strata". We thus sought to determine whether genes are more often haploid-upregulated in color strata compared to autosomes (contrasting expression in haploid a_1 and/or a_2 cells *vs.* in dikaryons), although we recognize that a single gene with antagonistic selection may be sufficient to drive recombination cessation in each stratum, which would be difficult to identify. This rationale is similar to using an enrichment in sex-biased genes in young evolutionary strata as evidence for sexually antagonistic selection driving sex chromosome evolution (Charlesworth 2017). Antagonistic selection between mating types may also drive differences in gene expression patterns between the two mating types, allowing the opposing selection pressures to be partially or fully resolved. We thus asked whether genes upregulated in the haploid phase also displayed significant differential expression between mating types more often in color strata than in autosomes, indicating that such strata have witnessed more antagonistic selection in the past.

Genes under antagonistic selection may alternatively show signs of divergent selection between mating types in terms of protein sequence. We tested whether the color evolutionary strata were enriched in genes with signs of divergent selection by looking at the ratio of nonsynonymous *vs.* synonymous substitutions. Two main issues in such enrichment tests have to be kept in mind, however: (i) finding no enrichment in genes under differential expression or divergent selection does not constitute definitive evidence for the lack of antagonistic selection; (ii) finding enrichment in genes under differential

expression or divergent selection does not constitute definitive evidence for antagonistic selection; (iii) divergent selection in expression or sequence may occur after recombination suppression. More generally, identifying loci under antagonistic selection is notoriously difficult even under ideal scenarios. Nevertheless, such tests contribute to our global understanding of selective pressures occurring in evolutionary strata.

We used published expression data to address these questions (Fontanillas *et al.* 2015; Perlin *et al.* 2015), together with stratum delimitation based on high-quality genome data (Branco *et al.* 2017). Gene expression in various stages in *M. lychnidis-dioicae* (Perlin *et al.* 2015) and in haploid cells of different mating types (Fontanillas *et al.* 2015) were produced in previous studies. The life stages investigated were: (i) haploid yeast forms of separate mating types grown on water agar, under conditions inducing mating *in vitro* when mating types are mixed (Hood and Antonovics 2004) and, thus, simulating the stage at which separate mating types are present on the meristem before mating and plant infection; (ii) haploid yeast forms of separate mating types grown in medium with a high glucose content where they multiply by mitosis; these conditions simulate the stage at which haploid sporidia undergo mitosis in flowers; (iii) the dikaryotic parasitic stage, called "n+n" for dikaryotic stage with two unfused nuclei per cell, before karyogamy, studied in infected plants (Perlin *et al.* 2015). We used these gene expression data, as well as sequence data, to test whether antagonistic selection between mating types played a role in the spread of recombination suppression beyond coupling the mating-type genes in *M. lychnidis-dioicae*.

MATERIALS AND METHODS

Datasets
We used published gene expression data to test for antagonistic selection in *M. lychnidis-dioicae* (Fontanillas *et al.* 2015; Perlin *et al.* 2015): Suppl. Table S1 https://trace.ncbi.nlm.nih.gov/Traces/study/?acc=+PRJNA246470&go=go).

We also used published gene predictions, assignations to genomic compartments and orthologous group reconstruction (Branco *et al.* 2017; Branco *et al.* 2018).

Differential expression analyses
We performed differential expression analyses to identify genes evolving under antagonistic selection between mating types, as we expected such genes to be (i) upregulated in the haploid phase compared to the dikaryotic phase and/or (ii) upregulated in one mating-type compared to the opposite mating-type. We performed a pseudo-alignment of each read set from the published RNAseq experiments (Suppl. Table S1) against each of the predicted set of coding-sequences from the a_1 and a_2 haploid genomes of the Lamole *M. lychnidis-dioicae* strain (Badouin *et al.* 2015) (Suppl. Table S2), using algorithms implemented in Kallisto v. 0.45.0 (Bray *et al.* 2016). Kallisto performs an RNAseq read count quantification through a k-mer and De Bruijn graph approach, allowing ultra-fast quantification and calculation of standard errors using bootstraps. We ran Kallisto with 100 bootstraps samples and a sequence-based bias correction.

We then used the pseudo-alignment outputs to perform differential expression analyses using the Sleuth R package (Pimentel *et al.* 2017). The statistical methods implemented in Sleuth allow accurate estimation of differential expression levels and of their significance using the entire quantification variance from the bootstraps. To estimate the significance of the differential expression levels, we used the likelihood-ratio test (LRT) implemented in Sleuth to compare linear models assuming different parameters to explain the variance from the

abundance estimates for each sample. Specifically, we compared a model assuming the variance to be explained only by biological replicates to models that take into account the replicates and either the life stage (*i.e.*, haploid or dikaryotic) or the mating-type (either a_1 or a_2). After the LRT, the Sleuth package returns a *q-value* per coding-sequence, *i.e.*, the corrected *p-value* following the Benjamini-Hochberg correction for reducing the false discovery rate (FDR) due to multiple testing; we chose a threshold of 0.001 for the *q-value* (corresponding to a FDR of 0.001), below which we considered the differential expression to be significant (analyses with higher thresholds did not change the global patterns and appeared not stringent enough, with most genes differentially expressed, Figure S1). We performed analyses of upregulation in the haploid stage using the set of coding-sequences from either the a_1 or a_2 haploid genome of the Lamole *M. lychnidis-dioicae* strain as reference. We only present the analyses based on the a_1 reference as results based on the a_2 reference yielded the same patterns and conclusions. The Suppl. Tables S1 and S2 present the accession numbers of the data used as well as pseudo-alignment statistics.

We investigated whether the non-recombining regions, and, more specifically, the color strata not involving mating-type genes, contained a higher proportion of genes upregulated in the haploid phase than autosomes, by performing chi-squared tests. We thus tested whether the genes more strongly expressed in at least one haploid condition/mating type than at the dikaryotic stage were more frequent on the mating-type chromosome than on autosomes. We first performed a global chi-squared test comparing the various genomic compartments (autosomes, PARs, black and color strata). We then compared the proportion of genes upregulated in the haploid phase in the color strata and the autosomes. As the question of the evolutionary origin of the strata not involving mating-type genes was the same for all the color strata and there were few in the color strata (Table 1), we pooled the genes from all color strata for the various tests to improve power.

For the genes upregulated in the haploid compared to the dikaryotic phase, we then investigated whether the differential expression levels were greater in color strata than on the autosomes. We hypothesized that, if the genes on the color strata had particularly important roles in the haploid phase, the haploid upregulation relative to the dikaryon might be stronger in the color strata than in autosomes. We performed pairwise Wilcoxon signed-rank tests to compare the distribution of the differential expression level of genes upregulated in the haploid phase between autosomes, the black stratum and the pooled color strata, for each mating type and each medium. The Sleuth package returns the differential expression level as a "beta" value. In order to get a

differential expression level index similar to the classically used log2 fold-change, we used the option "transform_fun_counts = function(x) log2(x + 0.5)" in the data preparation steps, as advocated in the Kallisto and Sleuth methods (sleuth_prep() R function; https://github.com/pachterlab/sleuth/issues/59).

We also investigated whether the genes upregulated in the haploid phase displayed differential expression between the a_1 and a_2 mating types, and whether the difference level, if any, was greater than for other genes. We only compared genes from the same genomic compartment (black stratum or pooled color strata) in order to compare genes with similar levels of degeneration due to recombination suppression, as degeneration can lead to differential expression between mating types without selection for it (Fontanillas *et al.* 2015). We used the differential expression levels as given by the beta values between the a_1 and a_2 mating types for genes upregulated in the haploid phase. We compared the distributions of differential expression level of genes upregulated during the haploid phase with those of all other pooled differentially expressed genes in the same genomic compartment, in a Wilcoxon signed-rank test in R. We also identified the genes significantly upregulated in one mating type compared to the other, either in water or rich medium. We also investigated genes differentially expressed between mating types in haploid phases at the 0.001 threshold.

Positive selection tests

If antagonistic selection drove the spread of recombination suppression beyond mating-type genes, generating the color evolutionary strata, then we could alternatively expect the genes involved in functions specific to alternative mating types to show specialization in the a_1 or a_2 mating type, with alleles encoding proteins with different sequences. We therefore tested whether the genes present in color strata displayed footprints of divergent selection between the alleles associated with the alternative mating types, in the form of significantly higher non-synonymous divergence (dN) than would be expected from the synonymous (neutral) substitution rate (dS). Ratios of dN/dS, known as ω, and the significance of positive selection assessed by comparing the likelihood of different sequence evolution models, were inferred with CODEML in the PAML package (Yang 2007). For all genes for which both a_1 and a_2 alleles were present, in the mating-type chromosome and in autosomes as control, we performed branch-site tests of positive selection ("Test 2" in the PAML manual, 2017 version). For each gene, we tested whether a model of evolution allowing sites to evolve with a different ω, and possibly >1, in the branch of the alleles associated with the a_1 or a_2 mating type in *M. lychnidis-dioicae* was more likely than a

■ **Table 1 Counts of genes in *Microbotryum lychnidis-dioicae* with expression data that could be assigned to the various chromosomal locations and those with differential expression (threshold 0.001) between life stages, respectively for the autosomes, pseudoautosomal regions (PARs) of the mating-type chromosome, and non-recombining region (NRR) of the mating-type chromosome, separated into the different evolutionary strata (blue, purple, black, orange, red and green)**

	Number of assigned genes	Number of genes upregulated in at least one haploid stage/mating type	Percentage of genes upregulated in at least one haploid stage/mating type
Total	9983	1534	15.37%
Autosomes	9083	1454	16.01%
PAR	137	13	9.49%
NRR black stratum	698	52	7.45%
NRR color strata (sum)	65	15	23.08%
Blue stratum	21	4	19.05%
Green stratum	3	1	33.33%
Orange stratum	7	0	0.00%
Purple stratum	9	4	44.44%
Red stratum	25	6	24.00%

model with $\omega \leq 1$ in all branches and sites and with similar values in all branches. For modeling sequence evolution and the direction of nucleotide changes, we used, as background branches, the sequences of genes in the a_1 and a_2 mating types in *M. lagerheimii* and *M. saponariae* (with recombining mating-type chromosomes), and, as the focal (foreground) branch, either the a_1 or a_2 sequence of *M. lychnidis-dioicae*. The input tree was the focal gene tree, as recombination suppression and gene conversion lead to gene-specific genealogies with more or less *trans*-specific polymorphism for a_1 and a_2 alleles (Branco *et al.* 2017; Branco *et al.* 2018). To build the gene trees, we used RAxML (Stamatakis 2006) under a GTRGAMMA model, with the orthologous sequences from both mating types in each of the three species *M. lychnidis-dioicae*, *M. lagerheimii* and *M. saponariae*, aligned using the codon-based approach implemented in translatorX (Abascal *et al.* 2010). In the first model (model A, specified in the PAML control file as: model = 2, NSsites = 2, omega = 0, fix_omega = 0.2), different classes of site were allowed: class 0 with $0 < \omega < 1$ in the background and foreground branches, class 2a with $0 < \omega < 1$ in the background branch and $\omega > 1$ in the foreground branch, class 2b with $\omega = 1$ in the background branch and $\omega > 1$ in the foreground branch. In the null model without positive selection (null model A, specified in the PAML control file as: model = 2, NSsites = 2, omega = 1, fix_omega = 1), three classes of sites were allowed, with either $\omega < 1$ or $\omega = 1$ in both the background and foreground branches, or $\omega < 1$ in the background branches and $\omega = 1$ in the foreground branch. Likelihood ratio tests (LRTs) were performed to compare the two models, with one degree of freedom, as suggested in the PAML documentation, and ω values were inferred for the different classes of sites. Because low levels of synonymous divergence can artificially inflate ω estimates, which would then be unreliable, we discarded the likelihood values resulting from the A model when the ω estimated of the foreground branch was higher than five in either the 2a or 2b site class, as typically recommended (Pond and Muse 2005; Chamary *et al.* 2006; Stoletzki and Eyre-Walker 2011). Such biases can be particularly problematic in non-recombining regions where non-synonymous substitutions may accumulate due to relaxed selection. We nevertheless also present results without filtering.

In addition, we plotted the per-gene d_N/d_S between the a_1 and a_2-associated alleles along the *M. lagerheimii* ancestral-like gene order of the mating-type chromosome, using the d_N and d_S values computed in the yn00 program (Yang and Nielsen 2000; Yang 2007).

Data availability

RESULTS

Upregulation in the haploid stage and/or in one mating type

Among the 12,254 predicted genes in *M. lychnidis-dioicae*, we were able to assign 9,983 genes studied for expression to chromosomal locations, *i.e.*, on autosomes, pseudo-autosomal regions (PARs) of the mating-type chromosome, the 'black' stratum of the non-recombining region (NRR) linking the two mating-type loci, and the various evolutionary "color strata" of the non-recombining regions not involving mating-type genes (the 'blue', 'green', 'orange', 'red' and 'purple' strata, Figure 1). Genes differentially expressed (q-value ≤ 0.001) between at least two life stages represented 15% of the assigned genes and were distributed among the various genome compartments (Figure 3 and Table 1). Most (95%) of the genes upregulated in the haploid phase nevertheless resided on autosomes (Table 1).

Significant differences were detected in the proportion of genes upregulated in a haploid phase compared to the dikaryotic phase among the various genomic compartments, *i.e.*, autosomes, PARs, black and color strata (Figure 3; chi-squared = 43.122, df = 3, p-value = $< 2.319e-09$). The significance was however mainly driven by the black stratum and the PARs being depleted in genes upregulated in the haploid phase (Figure 3 and Table 1). The color strata displayed no significant enrichment relative to autosomes in genes upregulated in the haploid phase (chi-squared = 2.3925, df = 1, p-value = 0.1219). Only 15 genes were found to be upregulated in the haploid phase and residing in the color strata (Table 1), and they all were upregulated in the rich medium, none in the water medium (Figure 4). Using higher thresholds for significant differences in expression levels did not change the patterns notably regarding the relative proportions of
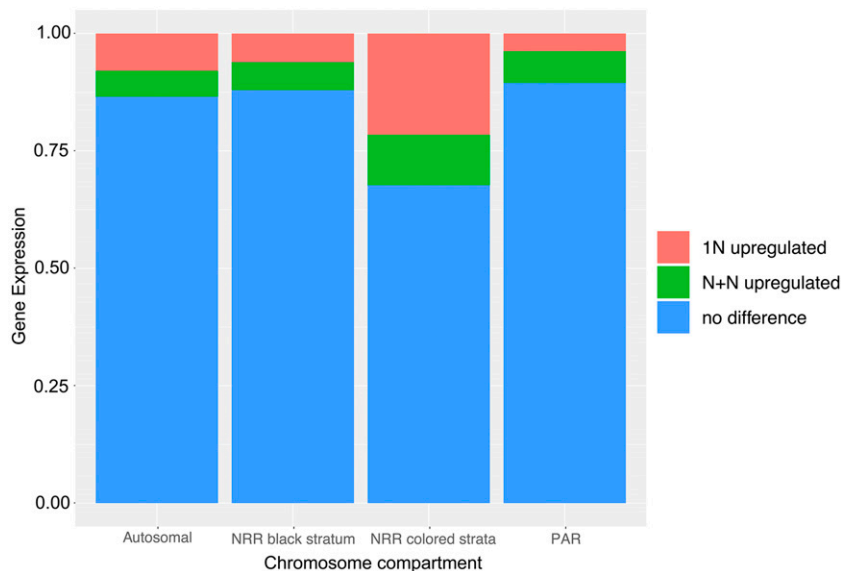
Figure 3 Differential expression in *Microbotryum lychnidis-dioicae*. Proportions of genes upregulated in at least one haploid stage (in red, 1N upregulated), upregulated at the dikaryotic stage (in green, N+N upregulated) or showing no differential expression (in blue). Expression level was considered significantly different at the 0.001 threshold. Different thresholds for significance did not change the patterns notably (Suppl. Figure S1). Genes are separated according to their genomic compartment: autosomes, pseudoautosomal regions (PARs) of the mating-type chromosome, non-recombining region (NRR) of the mating-type chromosome, and into the black *vs.* color evolutionary strata.

genes upregulated in a haploid stage among genomic compartments and the high proportions of genes differentially expressed between life history stages suggested that these thresholds were not stringent enough (Suppl. Figure S1). The putative functions of the genes upregulated in at least one haploid stage/mating type in the black or color strata did not correspond to functions reasonably expected to be advantageous in one mating type but not the other, as could be functions related to mitochondria inheritance (Suppl. Table S3). These findings altogether provide little support that antagonistic selection was a major driver of the spread of recombination suppression beyond mating-type genes.

We then tested whether the genes identified above as upregulated in the haploid cells had particularly large differences in expression between the dikaryotic phase and the haploid phases, comparing haploid expression, in either rich or water media, to dikaryotic expression *in planta*. The differential expression level for haploid-upregulated genes was not stronger for genes in the color strata than for genes in autosomes (Suppl. Table S4; Figure 4). Genes in the color strata even showed less variation and fewer extreme values in differential expression (Figure 4). Haploid upregulation was stronger in the black stratum

than in autosomes in the water medium, although the test was not significant anymore when applying a Bonferroni correction for multiple testing (Figure 4B; Suppl. Table S4).

We found that the genes in the color or black strata that were upregulated in the haploid compared to the dikaryotic phase displayed no greater difference either in expression level between the a_1 and a_2 mating types than other genes (Figure 5; Suppl. Table S5), which is again not consistent with the antagonistic selection hypothesis. The color strata outlier (MvSl-1064-A1-R4_A1g01162) that showed higher haploid differential expression between a_1 and a_2 in both water agar and rich medium belonged to the blue stratum, one of the oldest strata. This gene had no putative function, and a BLASTp search did not yield any further insight into its function.

There were only eight genes in the genome with significant differential expression between mating types, all in the water medium, none residing in color strata. The genes with significant differential expression between mating types included the pheromone receptor gene itself, two genes in autosomes and five in the black stratum, with no obvious function that can be related to antagonistic selection (Suppl. Table S6; the enrichment in the black stratum relative to
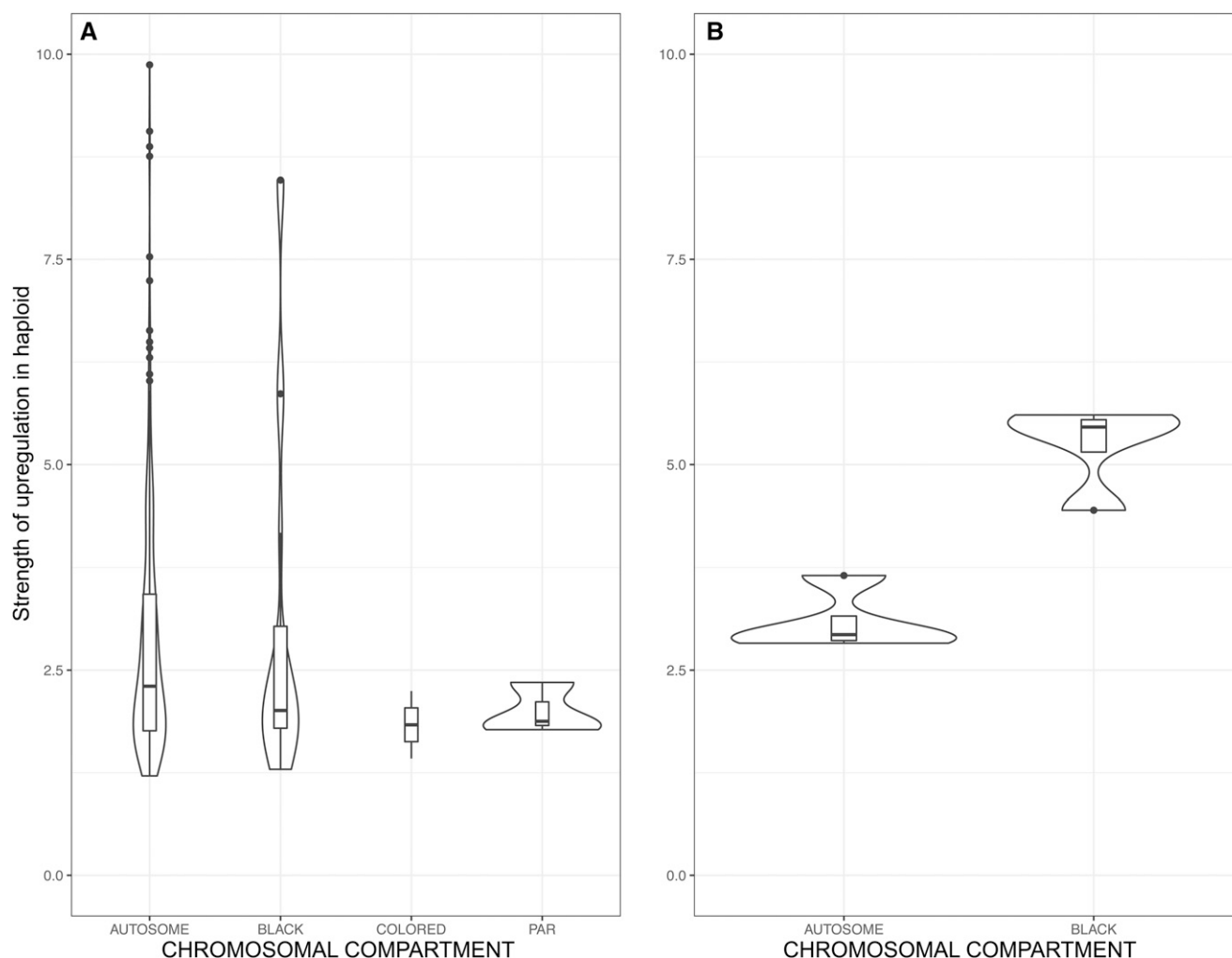
Figure 4 Strength of haploid upregulation in the different genomic compartments in *Microbotryum lychnidis-dioicae*. Violin plots of the differential expression levels (beta values, calculated in a similar way as the usual log$_2$ fold-change) of the haploid compared to the dikaryotic phase for the various genomic compartments (autosomes, black stratum, color strata and PARs), with haploids grown in: A 'rich' or B 'water' media (dikaryotic expression was measured *in planta*). In panel B, no genes were found upregulated in water in color strata or PARs.
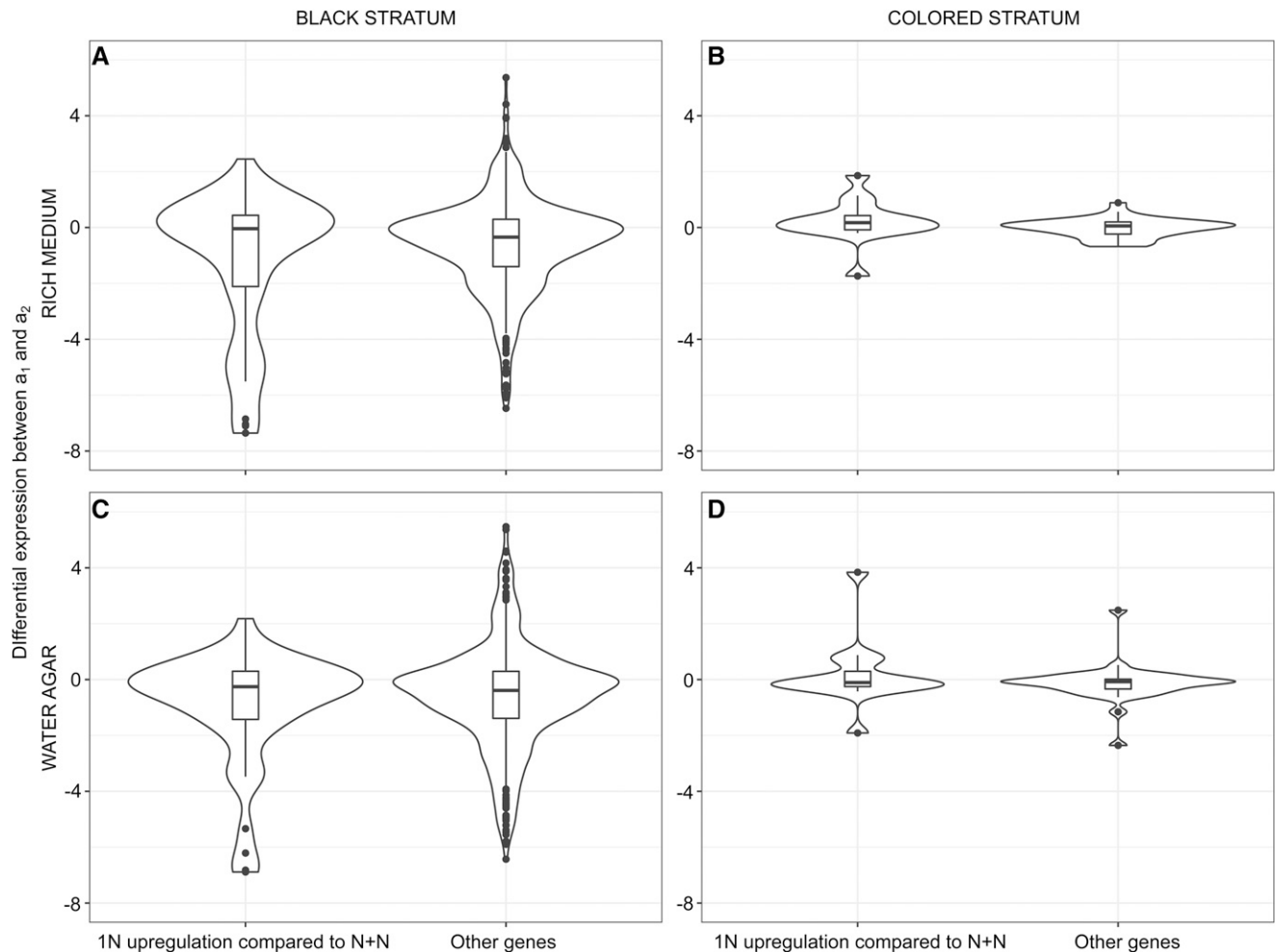
Figure 5 Strength of differential expression between mating-types in *Microbotryum lychnidis-dioicae*. Violin plot of differential expression (beta values, calculated in a similar way as the usual $\log_2$ fold-change) between the a_1 and a_2 mating types at haploid stages in *Microbotryum lychnidis-dioicae*, for genes found upregulated in at least one haploid stage compared to the dikaryotic stage and for the other genes (either upregulated in the dikaryon or without differential expression), in the black stratum (A and C) or the color strata (B and D), on rich medium (A and B) or water (C and D).

autosomes was not significant; Fisher exact test, $P = 0.08$) The finding of a lack of genes with differential expression between mating types in color strata again does not support the hypothesis that mating-type antagonistic selection would drive evolutionary strata of recombination suppression.

Divergent selection between mating types

We then investigated whether the color strata were enriched in genes with signs of divergent selection between mating types, with amino-acid substitutions more frequent between the a_1 and a_2 mating types than would be expected on the basis of neutral (synonymous) substitution rates (*i.e.*, $\omega = dN/dS > 1$). For each gene, we considered, as background branches, the sequences of genes in the a_1 and a_2 mating types of the outgroup *M. lagerheimii* and *M. saponariae* (species with recombining mating-type chromosomes), and as the focal (foreground) branch, either the a_1 or a_2 sequence of *M. lychnidis-dioicae* (Suppl. Figure S2). We determined whether models allowing sites with $\omega > 1$ in the foreground branches were more likely (Suppl. Figure S2). For comparison, we also ran the analyses on autosomal genes, PAR genes, and genes in the black stratum.

Likelihood ratio tests, performed after filtering to remove low d_S values that may generate unreliable d_N/d_S estimates (Pond and Muse 2005; Chamary *et al.* 2006; Stoletzki and Eyre-Walker 2011), indicated that the model with divergent selection between mating types was significantly more likely than the alternative in very few genes in non-recombining regions (Table 2): only two genes in the black stratum were found to evolve under positive selection, one in each of the a_1 and a_2 mating-type chromosome, and none in the color strata. The color strata were thus not enriched in genes under positive selection compared to the autosomes, they even seemed depleted in genes under diversifying selection between mating types. The putative functions of the genes under positive selection did not appear likely involved in antagonistic selection between mating types (Suppl. Table S3).

Without filtering for low d_S values, there were 11 genes with signs of positive selection in one of the mating types in color strata (four in the blue stratum, three in the red stratum, two in the orange stratum and two in the purple stratum; Suppl. Table S7). Only one of these 11 genes, located in the purple stratum, was significantly upregulated in the haploid phase (Suppl. Table S7). There were 47 genes with signs of positive selection in the black stratum, three being haploid upregulated,

a_1 mating-type	Null model	Model with selection	Model with selection among upregulated
Autosomal genes (MC02)	6286	28	3
PARs	89	0	0
Black stratum	118	1	0
Purple stratum	8	0	0
Blue stratum	18	0	0
Orange stratum	7	0	0
Red stratum	23	0	0
Green stratum	3	0	0
Color strata (pooled)	59	0	0
a_2 mating-type			
Autosomal genes (MC02)	6262	41	7
PARs	89	0	0
Black stratum	105	1	0
Purple stratum	8	0	0
Blue stratum	17	0	0
Orange stratum	5	0	0
Red stratum	24	0	0
Green stratum	3	0	0
Color strata (pooled)	57	0	0

and 140 in autosomes, 17 being haploid upregulated. There was a significant enrichment in the both the black and color strata compared to autosomes for genes with signs of positive selection (Fisher tests, $P = 2.2 \times 10^{-16}$ for the black stratum, $P = 1.36 \times 10^{-8}$ for the color strata). The enrichment in genes with both signs of positive selection and haploid upregulation was significant only for the black stratum and not for the color strata (Fisher tests, d.f.=1, $P = 0.0025$ for the black stratum, $P = 0.1218$ for the color strata). The putative functions of the genes with signs of positive selection (Suppl. Table S7) did not suggest any role for antagonistic selection, except perhaps a function linked to mitochondria stability in the black stratum (MvSl-1064-A1-R4_A1g00541, see discussion).

The d_N/d_S values between alleles associated with a_1 *vs.* a_2 mating types were not higher in the color evolutionary strata (Figure 6). Altogether these findings provide little support for the notion that the specialization of genes with important haploid roles to a_1 or a_2 mating types is the predominant force driving recombination suppression in the various color strata.

DISCUSSION

Our findings that the color strata of the non-recombining mating-type chromosomes were not enriched in genes upregulated in the haploid phase and carried no gene differentially expressed between mating types or under divergent selection after filtering provide little support for the hypothesis that the spread of recombination suppression beyond mating-type genes in *M. lychnidis-dioicae* was due to antagonistic selection between mating types.

Relaxing the filtering of d_N/d_S for high values, we found 11 genes with significant signs of positive selection in color strata. However, the finding that a single one was upregulated in the haploid phase (where mating types are expressed) and their putative functions provided little evidence for antagonistic selection. The single haploid upregulated gene with significant positive selection without

filtering among color strata was in the purple stratum and appeared involved in histone deposition, which could be related to recombination suppression (MvSl-1064-A1-R4_A1g00230); this is relevant for the evolution of mating-type chromosomes but unlikely to involve antagonistic selection. Note that a very high d_N/d_S ratio most often represents a biased estimate (when d_S is very low, d_N/d_S estimates are not reliable) and filtering them out is recommended and typically done (Pond and Muse 2005; Chamary *et al.* 2006; Stoletzki and Eyre-Walker 2011; Villanueva-Cañas *et al.* 2013). This issue is likely particularly problematic in the highly selfing *M. lychnidis-dioicae*, in which differentiation is very low between a_1 and a_2 genomes in a given diploid individual in young evolutionary strata, and in which non-synonymous substitutions accumulate in non-recombining regions due to relaxed selection (Fontanillas *et al.* 2015). In addition, signs of positive selection in evolutionary strata may be due to dominant beneficial mutations that have appeared after recombination suppression and cannot spread to the alternative allele due to lack of recombination and may thus not correspond to divergent selection between mating types. Most of the putative functions of genes with significant signs of positive selection without filtering in fact did not correspond to roles that can be imagined to be under antagonistic selection between mating types. The only function that could be related to antagonistic selection in non-recombining regions was involved in mitochondria stability (MvSl-1064-A1-R4_A1g00541). Such a function may be involved in mitochondria inheritance, which is asymmetric between mating types in *M. lychnidis-dioicae*. Mating type a_2 progeny inherit ca 90% of a_2 parental mitochondria, while equal proportions of parental mitochondria are inherited by a_1 progeny (Wilch *et al.* 1992). This gene was however located in the black stratum and therefore cannot explain the evolution of recombination suppression in the color strata flanking the mating-type loci. Another interesting function among genes with significant signs of positive selection without filtering in the black stratum appeared related to the negative regulation of mitosis (MvSl-1064-A1-R4_A1g00541), which may be important in the mating stage, but for both mating types. In the future, when more high-quality genome assemblies will be available across the *Microbotryum* genus, it may be worth testing for positive selection including more species to estimate dN/dS along shorter branches just before the evolution of the various color strata.

Genes upregulated in the haploid phase (likely to have important roles in this stage where cells are of distinct mating types) similarly appeared unlikely to have evolved under antagonistic selection as they seem to have similar roles in the a_1 and a_2 mating types, particularly for the ones located in the color strata. Genes upregulated in the haploid phase did not differ markedly between the a_1 and a_2 mating types, in terms of protein sequence or gene expression level. These findings, together with the location of most (95%) of the genes upregulated in the haploid phase on autosomes, support the view that the functions important in the haploid phase are similar between a_1 and a_2 cells, which does not meet the expectations of antagonistic selection acting between mating types. The genes of importance at the haploid stage for roles other than mating-type determinism, such as genes involved in haploid mitotic division or the functional process of mating, are likely to perform the same function in both mating types and would therefore not be expected to be selected for linkage to mating-type genes. The putative functions assigned to the genes upregulated in the haploid phase were in fact all general in nature, with no obvious reason for selection for different aspects in the two alternative mating types.

While antagonistic selection is an attractive and theoretically plausible hypothesis for explaining the spread of recombination suppression
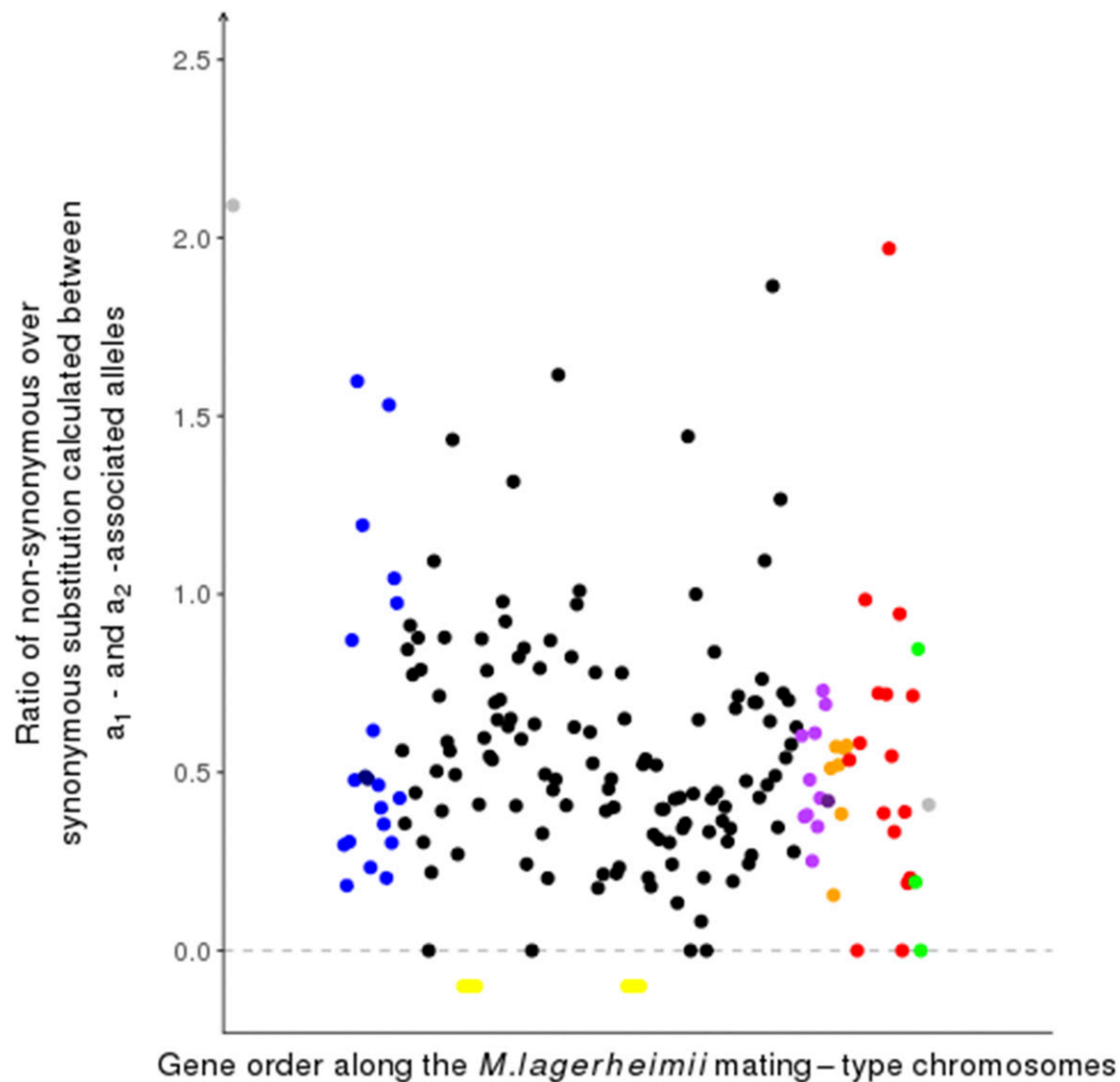
Figure 6 Per-gene non-synonymous over synonymous (d_N/d_S) differences between a_1-a_2 associated mating types along the mating-type chromosomes in *Microbotryum lychnidis-dioicae*. Genes are located according to the ancestral-like gene order (*i.e.*, gene order from *M. lagerheimii*) and evolutionary strata are indicated by their colors. Ancestral location of centromeres (before chromosomal fusion) are indicated in yellow; d_N/d_S values could not be computed for most of the genes in pseudo-autosomal regions (in gray), as most had null d_S values.

on sex chromosomes, decades of research have uncovered limited evidence in support of this theory (Beukeboom and Perrin 2014; Wright *et al.* 2016). Alternative hypotheses have been put forward (Ironside 2010; Ponnikas *et al.* 2018), including the successive linkage of genes accumulating deleterious recessive mutations in the margins of the non-recombining region, due to linkage disequilibrium with sex-determining loci. Complete linkage fixes heterozygosity, thereby permanently sheltering heterozygous deleterious recessive mutations. This process has been hypothesized to play a role in *Microbotryum* fungi (Hood and Antonovics 2000; Antonovics and Abrams 2004; Hood and Antonovics 2004; Johnson *et al.* 2005), as the haploid phase is of limited relevance under natural conditions. Alternatively, chromosomal inversions may arise and fix by drift on one sex chromosome but fail to spread to the other sex chromosome if the inversion completely suppresses recombination with the sex determining locus (Ironside 2010). It has also been suggested that transposable element (TE) accumulation in or near the non-recombining portion of sex

chromosomes can suppress recombination further, through genomic silencing of the TEs by DNA methylation and/or chromatin modifications (Kent *et al.* 2017). These hypotheses have been very little studied to date, despite their potentially important roles in the spread of recombination suppression on sex and mating-type chromosomes. The existence of evolutionary strata in a fungus without male and female roles, and the lack of evidence for antagonistic selection between mating types, highlights the need to investigate these alternative hypotheses.

Of course, lack of evidence for widespread antagonistic selection is not evidence that antagonistic selection had no role in stratum evolution, and we had little power to detect antagonistic selection acting on only one or a few sites. In particular, the lack of enrichment in genes upregulated in the haploid phase in the color strata still allows the possibility that a single gene under antagonistic selection, in each stratum, drove recombination suppression. Conversely, finding genes under divergent selection in evolutionary strata does

not provide strong evidence that antagonistic selection caused recombination cessation, given that alternative hypotheses for stepwise recombination suppression also predict such footprints of divergent selection. For example, heterozygote advantage in the diploid or dikaryotic phase (potentially applying also to associative overdominance, with different deleterious mutations associated with the two sex chromosomes) can promote recombination cessation and lead to a pattern of divergent selection between alleles associated with alternative mating types (Otto 2014 ; Immler and Otto 2015). Furthermore deleterious allele accumulation after recombination cessation is expected to generate patterns of differential expression between mating types (Fontanillas et al. 2015). These issues in testing the role of antagonistic selection in stratum evolution are also reasons why alternative hypotheses to antagonistic selection are worth exploring and disentangling. Investigating genomic patterns of transposable element accumulation and their silencing, as well as deleterious mutation accumulation, at the margin of regions lacking recombination, could allow testing if these mechanisms have promoted the spread of recombination cessation. Silencing-based mechanisms would generate new strata without any discernible chromosomal rearrangement and may lead to a more gradual divergence and loss of recombination than an inversion-based hypothesis, for example. The spread of recombination suppression appears more continuous than discrete in some cases, including in some *Microbotryum* species (Branco et al. 2018) and other organisms (Bergero and Charlesworth 2009). In *Microbotryum lychnidis-dioicae*, however, discrete strata have been inferred (Figure 1). The evolutionary forces driving these strata remain to be determined. The work presented here finds little evidence that either mating-type antagonistic selection and/or ploidally antagonistic selection are responsible.

ACKNOWLEDGMENTS

LITERATURE CITED

Abascal, F., R. Zardoya, and M. Telford, 2010   TranslatorX: multiple alignment of nucleotide sequences guided by amino acid translations. Nucleic Acids Res. 38: W7–W13. https://doi.org/10.1093/nar/gkq291

Alexander, H. M., 1989   An experimental field-study of anther-smut disease of *Silene alba* caused by *Ustilago violacea*- Genotypic variation and disease incidence. Evolution 43: 835–847.

Alexander, H. M., 1990   Epidemiology of anther-smut infection of *Silene alba* caused by *Ustilabo violacea* - patterns of spore deposition and disease incidence. J. Ecol. 78: 166–179. https://doi.org/10.2307/2261043

Alexander, H. M., and J. Antonovics, 1995   Spread of anther-smut disease (*Ustilago violacea*) and character correlations in a genetically variable population of *Silene alba*. J. Ecol. 83: 783–794. https://doi.org/10.2307/2261415

Alexander, H. M., and A. Maltby, 1990   Anther-smut infection of *Silene alba* caused by *Ustilago violacea* - Factors determining fungal reproduction. Oecologia 84: 249–253. https://doi.org/10.1007/BF00318280

Antonovics, J., and J. Y. Abrams, 2004   Intratetrad mating and the evolution of linkage relationships. Evolution 58: 702–709. https://doi.org/10.1111/j.0014-3820.2004.tb00403.x

Badouin, H., M. E. Hood, J. Gouzy, G. Aguileta, S. Siguenza et al., 2015   Chaos of rearrangements in the mating-type chromosomes of the anther-smut fungus *Microbotryum lychnidis-dioicae*. Genetics 200: 1275–1284. https://doi.org/10.1534/genetics.115.177709

Bergero, R., and D. Charlesworth, 2009   The evolution of restricted recombination in sex chromosomes. Trends Ecol. Evol. 24: 94–102. https://doi.org/10.1016/j.tree.2008.09.010

Beukeboom, L., and N. Perrin, 2014, pp. 92–93 in *The evolution of sex determination*, Oxford University Press, Oxford. https://doi.org/10.1093/acprof:oso/9780199657148.001.0001

Biere, A., and J. Antonovics, 1996   Sex-specific costs of resistance to the fungal pathogen *Ustilago violacea (Microbotryum violaceum)* in *Silene alba*. Evolution 50: 1098–1110.

Biere, A., and S. C. Honders, 1998   Anther smut transmission in *Silene latifolia* and *Silene dioica*: Impact of host traits, disease frequency, and host density. Int. J. Plant Sci. 159: 228–235. https://doi.org/10.1086/297543

Billiard, S., M. Lopez-Villavicencio, B. Devier, M. Hood, C. Fairhead et al., 2011   Having sex, yes, but with whom? Inferences from fungi on the evolution of anisogamy and mating types. Biol. Rev. Camb. Philos. Soc. 86: 421–442. https://doi.org/10.1111/j.1469-185X.2010.00153.x

Branco, S., H. Badouin, R. Rodríguez de la Vega, J. Gouzy, F. Carpentier et al., 2017   Evolutionary strata on young mating-type chromosomes despite lack of sexual antagonism. Proc. Natl. Acad. Sci. USA 114: 7067–7072. https://doi.org/10.1073/pnas.1701658114

Branco, S., F. Carpentier, R. Rodríguez de la Vega, H. Badouin, A. Snirc et al., 2018   Multiple convergent supergene evolution events in mating-type chromosomes. Nat. Commun. 9: 2000. https://doi.org/10.1038/s41467-018-04380-9

Bray, N. L., H. Pimentel, P. Melsted, and L. Pachter, 2016   Near-optimal probabilistic RNA-seq quantification. Nat. Biotechnol. 34: 525–527. Erratum: 34: 888. https://doi.org/10.1038/nbt.3519

Chamary, J. V., J. L. Parmley, and L. D. Hurst, 2006   Hearing silence: non-neutral evolution at synonymous sites in mammals. Nat. Rev. Genet. 7: 98–108. https://doi.org/10.1038/nrg1770

Charlesworth, B., and D. Charlesworth, 1997   Rapid fixation of deleterious alleles can be caused by Muller's ratchet. Genet. Res. 70: 63–73. https://doi.org/10.1017/S0016672397002899

Charlesworth, B., and J. D. Wall, 1999   Inbreeding, heterozygote advantage and the evolution of neo-X and neo-Y sex chromosomes. Proc. Biol. Sci. 266: 51–56. https://doi.org/10.1098/rspb.1999.0603

Charlesworth, D., 2016   The status of supergenes in the 21st century: recombination suppression in Batesian mimicry and sex chromosomes and other complex adaptations. Evol. Appl. 9: 74–90. https://doi.org/10.1111/eva.12291

Charlesworth, D., 2017   Evolution of recombination rates between sex chromosomes. Philos. Trans. R. Soc. Lond. B Biol. Sci. 372: 20160456. https://doi.org/10.1098/rstb.2016.0456

Coelho, S. M., J. Gueno, A. P. Lipinska, J. M. Cock, and J. G. Umen, 2018   UV Chromosomes and Haploid Sexual Systems. Trends Plant Sci. 23: 794–807. https://doi.org/10.1016/j.tplants.2018.06.005

Fontanillas, E., M. Hood, H. Badouin, E. Petit, V. Barbe et al., 2015   Degeneration of the non-recombining regions in the mating type chromosomes of the anther smut fungi. Mol. Biol. Evol. 32: 928–943. https://doi.org/10.1093/molbev/msu396

Giraud, T., O. Jonot, and J. A. Shykoff, 2005   Selfing propensity under choice conditions in a parasitic fungus, *Microbotryum violaceum*, and parameters influencing infection success in artificial inoculations. Int. J. Plant Sci. 166: 649–657. https://doi.org/10.1086/430098

Golonka, A. M., and R. Vilgalys, 2013   Nectar inhabiting yeasts in Virginian populations of *Silene latifolia* (Caryophyllaceae) and coflowering species. Am. Midl. Nat. 169: 235–258. https://doi.org/10.1674/0003-0031-169.2.235

Granberg, A., U. Carlsson-Graner, P. Arnqvist, and B. E. Giles, 2008   Variation in breeding system traits within and among populations of *Microbotryum violaceum* on *Silene dioica*. Int. J. Plant Sci. 169: 293–303. https://doi.org/10.1086/523964

Gregorius, H. R., 1982   Selection in plant populations of effectively infinite size. 2. Protectedness of a biallelic polymorphism. J. Theor. Biol. 96: 689–705. https://doi.org/10.1016/0022-5193(82)90237-5

Grognet, P., F. Bidard, C. Kuchl, L. Chan Ho Tong, E. Coppin et al., 2014   Maintaining two mating types: structure of the mating type locus and its role in heterokaryosis in *Podospora anserina*. Genetics 197: 421–432. https://doi.org/10.1534/genetics.113.159988

Hood, M. E., and J. Antonovics, 1998   Two-celled promycelia and mating-type segregation in *Ustilago violacea (Microbotryum violaceum)*. Int. J. Plant Sci. 159: 199–205. https://doi.org/10.1086/297539

Hood, M. E., and J. Antonovics, 2000   Intratetrad mating, heterozygosity, and the maintenance of deleterious alleles in *Microbotryum violaceum (=Ustilago violacea)*. Heredity 85: 231–241. https://doi.org/10.1046/j.1365-2540.2000.00748.x

Hood, M. E., and J. A. Antonovics, 2004   Mating within the meiotic tetrad and the maintenance of genomic heterozygosity. Genetics 166: 1751–1759. https://doi.org/10.1534/genetics.166.4.1751

Idnurm, A., M. E. Hood, H. Johannesson, and T. Giraud, 2015   Contrasted patterns in mating-type chromosomes in fungi: Hotspots *vs.* coldspots of recombination. Fungal Biol. Rev. 29: 220–229. https://doi.org/10.1016/j.fbr.2015.06.001

Immler, S., and S. P. Otto, 2015   The evolution of sex chromosomes in organisms with separate haploid sexes. Evolution 69: 694–708. https://doi.org/10.1111/evo.12602

Ironside, J. E., 2010   No amicable divorce? Challenging the notion that sexual antagonism drives sex chromosome evolution. BioEssays 32: 718–726. https://doi.org/10.1002/bies.200900124

Johnson, L. J., J. Antonovics, and M. E. Hood, 2005   The evolution of intratetrad mating rates. Evolution 59: 2525–2532. https://doi.org/10.1111/j.0014-3820.2005.tb00966.x

Jordan, C. Y., and T. Connallon, 2014   Sexually antagonistic polymorphism in simultaneous hermaphrodites. Evolution 68: 3555–3569. https://doi.org/10.1111/evo.12536

Kaltz, O., and J. A. Shykoff, 1997   Sporidial mating-type ratios of teliospores from natural populations of the anther smut fungus *Microbotryum* (equals *Ustilago*) *violaceum*. Int. J. Plant Sci. 158: 575–584. https://doi.org/10.1086/297470

Kaltz, O., and J. A. Shykoff, 2001   Male and female *Silene latifolia* plants differ in per-contact risk of infection by a sexually transmitted disease. J. Ecol. 89: 99–109. https://doi.org/10.1046/j.1365-2745.2001.00527.x

Kasimatis, K. R., T. C. Nelson, and P. C. Phillips, 2017   Genomic signatures of sexual conflict. J. Hered. 108: 780–790. https://doi.org/10.1093/jhered/esx080

Kent, T. V., J. Uzunović, and S. I. Wright, 2017   Coevolution between transposable elements and recombination. Philos. Trans. R. Soc. Lond. B Biol. Sci. 372. pii: 20160458

López-Villavicencio, M., O. Jonot, A. Coantic, M. E. Hood, J. Enjalbert *et al.*, 2007   Multiple infections by the anther Smut pathogen are frequent and involve related strains. PLoS Pathog. 3: e176. https://doi.org/10.1371/journal.ppat.0030176

Mank, J. E., 2017   Population genetics of sexual conflict in the genomic era. Nat. Rev. Genet. 18: 721–730. https://doi.org/10.1038/nrg.2017.83

Marais, G. A. B., M. Nicolas, R. Bergero, P. Chambrier, E. Kejnovsky *et al.*, 2008   Evidence for degeneration of the Y chromosome in the dioecious plant *Silene latifolia*. Curr. Biol. 18: 545–549. https://doi.org/10.1016/j.cub.2008.03.023

Nieuwenhuis, B. P. S., S. Billiard, S. Vuilleumier, E. Petit, M. E. Hood *et al.*, 2013   Evolution of uni- and bifactorial sexual compatibility systems in fungi. Heredity 111: 445–455. https://doi.org/10.1038/hdy.2013.67

Otto, S. P., 2009   The evolutionary enigma of sex. Am. Nat. 174: S1–S14. https://doi.org/10.1086/599084

Otto, S. P., 2014   Selective maintenance of recombination between the sex chromosomes. 27: 1431–1442.

Oudemans, P. V., H. M. Alexander, J. Antonovics, S. Altizer, P. H. Thrall *et al.*, 1998   The distribution of mating-type bias in natural populations of the anther-smut *Ustilago violacea* on *Silene alba* in Virginia. Mycologia 90: 372–381. https://doi.org/10.1080/00275514.1998.12026921

Perlin, M., J. Amselem, E. Fontanillas, S. Toh, Z. Chen *et al.*, 2015   Sex and parasites: Genomic and transcriptomic analysis of *Microbotryum lychnidis-dioicae*, the biotrophic and plant-castrating anther smut fungus. BMC Genomics 16: 461. https://doi.org/10.1186/s12864-015-1660-8

Pimentel, H., N. L. Bray, S. Puente, P. Melsted, and L. Pachter, 2017   Differential analysis of RNA-seq incorporating quantification uncertainty. Nat. Methods 14: 687–690. https://doi.org/10.1038/nmeth.4324

Pond, S. K., and S. V. Muse, 2005   Site-to-site variation of synonymous substitution rates. Mol. Biol. Evol. 22: 2375–2385. https://doi.org/10.1093/molbev/msi232

Ponnikas, S., H. Sigeman, J. K. Abbott, and B. Hansson, 2018   Why Do Sex Chromosomes Stop Recombining? Trends Genet. 34: 492–503. https://doi.org/10.1016/j.tig.2018.04.001

Samils, N., A. Gioti, M. Karlsson, Y. Sun, T. Kasuga *et al.*, 2013   Sex-linked transcriptional divergence in the hermaphrodite fungus *Neurospora tetrasperma*. Proceedings of the Royal Society B-Biological Sciences 280. https://doi.org/10.1098/rspb.2013.0862

Sandler, G., F. E. G. Beaudry, S. C. H. Barrett, and S. I. Wright, 2018   The effects of haploid selection on Y chromosome evolution in two closely related dioecious plants. Evolution Letters 197: 865.

Schäfer, A. M., M. Kemler, R. Bauer, and D. Begerow, 2010   The illustrated life cycle of *Microbotryum* on the host plant *Silene latifolia*. Botany-Botanique 88: 875–885. https://doi.org/10.1139/B10-061

Scott, M. F., and S. P. Otto, 2017   Haploid selection favors suppressed recombination between sex chromosomes despite causing biased sex ratios. Genetics 207: 1631–1649.

Seitner, D., S. Uhse, M. Gallei, and A. Djamei, 2018   The core effector Cce1 is required for early infection of maize by Ustilago maydis. Mol. Plant Pathol. 19: 2277–2287. https://doi.org/10.1111/mpp.12698

Stamatakis, A., 2006   RAxML-VI-HPC: Maximum Likelihood-based Phylogenetic Analyses with Thousands of Taxa and Mixed Models. Bioinformatics 22: 2688–2690. https://doi.org/10.1093/bioinformatics/btl446

Stoletzki, N., and A. Eyre-Walker, 2011   The positive correlation between dN/dS and dS in mammals is due to runs of adjacent substitutions. Mol. Biol. Evol. 28: 1371–1380. https://doi.org/10.1093/molbev/msq320

Thomas, A., J. Shykoff, O. Jonot, and T. Giraud, 2003   Sex-ratio bias in populations of the phytopathogenic fungus *Microbotryum violaceum* from several host species. Int. J. Plant Sci. 164: 641–647. https://doi.org/10.1086/375374

Thrall, P. H., and A. M. Jarosz, 1994   Host-pathogen dynamics in experimental populations of *Silene alba* and *Ustilago violacea*. 1. experimental tests of theoretical models. J. Ecol. 82: 561–570. https://doi.org/10.2307/2261264

Uyenoyama, M. K., 2005   Evolution under tight linkage to mating type. New Phytol. 165: 63–70. https://doi.org/10.1111/j.1469-8137.2004.01246.x

Vercken, E., M. Fontaine, P. Gladieux, M. Hood, O. Jonot *et al.*, 2010   Glacial refugia in pathogens: European genetic structure of anther smut pathogens on *Silene latifolia* and *S. dioica*. PLoS Pathog. 6: e1001229. https://doi.org/10.1371/journal.ppat.1001229

Villanueva-Cañas, J., S. Laurie, and M. Albà, 2013   Improving genome-wide scans of positive selection by using protein isoforms of similar length. Genome Biol. Evol. 5: 457–467. https://doi.org/10.1093/gbe/evt017

Wilch, G., S. Ward, and A. Castle, 1992   Transmission of mitochondrial DNA in *Ustilago violacea*. Curr. Genet. 22: 135–140. https://doi.org/10.1007/BF00351473

Wright, A., R. Dean, F. Zimmer, and J. Mank, 2016   How to make a sex chromosome. Nat. Commun. 7: 12087. https://doi.org/10.1038/ncomms12087

Xu, J. P., 2005   The inheritance of organelle genes and genomes: patterns and mechanisms. Genome Biol. Evol. 48: 951–958.

Yang, Z., 2007   PAML 4: phylogenetic analysis by maximum likelihood. Mol. Biol. Evol. 24: 1586–1591. https://doi.org/10.1093/molbev/msm088

Yang, Z. H., and R. Nielsen, 2000   Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. Mol. Biol. Evol. 17: 32–43. https://doi.org/10.1093/oxfordjournals.molbev.a026236

Zhang, M., H. Feng, Y. H. Zhao, L. L. Song, C. Gao *et al.*, 2018   *Valsa mali* pathogenic effector VmPxE1 contributes to full virulence and interacts with the host peroxidase MdAPX1 as a potential target. Front. Microbiol. 9: 11.

Zillig, H., 1921   Über spezialisierte Formen beim Antherenbrand, *Ustilago violacea* (Pers.) Fuck. Zentralbl. Bakteriol. 53: 33–74.

Communicating editor: E. Betran

4 Convergent evolution of recombination suppression

The linkage of mating-type loci in *Microbotryum* fungi is likely driven by a selection for optimizing gamete compatibility under selfing. The *Microbotryum* genus encompasses both tetrapolar species, with unlinked PR and HD mating-type genes located on two distinct mating-type chromosomes and bipolar species, with PR and HD genes linked by a recombination suppression onto the same chromosome. A previous study analysed the transitions between tetrapolarity and bipolarity in a phylogenetic context (Hood et al. 2015). The outgroup to our studied species in the *Microbotryum* genus was *M. intermedium*, a tetrapolar fungus; *M. lagerheimii* was also tetrapolar, while internal in the phylogeny. All the other species considered in the phylogeny were bipolar. The most parsimonious hypothesis was therefore that bipolarity evolved once after the split of *M. intermedium* and has been lost in *M. lagerheimii*. Another hypothesis would be that the ancestral tetrapolarity state has been retained in *M. lagerheimii* from the common ancestor, which would imply that bipolarity evolved multiple times independently. While the convergence hypothesis seems less parsimonious a priori, the convergent evolution of bipolarity could have occurred if there was a strong selection to link the PR and the HD genes. As a matter of fact, HD and PR linkage greatly increases the odds of gamete compatibility compared with unlinked PR and HD genes under intra-tetrad mating which is known to occur at high frequency in *Microbotryum* fungi.

In the present study, we tested the hypothesis of convergent evolution of mating-type loci linkage by analyzing the mating-type chromosomes in the bipolar fungi *M. silenes-acaulis*, *M. v. paradoxa*, *M. v. caroliniana* and *M. scabiosae*. We checked the synteny among their mating-type chromosomes and those from the previously studied species *M. lychnidis-dioicae*, *M. silenes-dioicae*, *M. lagerheimii* and *M. intermedium*. We plotted the synonymous divergence between the a_1 and a_2 mating-type chromosomes for each species to identify non-recombining regions and possible evolutionary strata. In addition, we tested whether the recombination suppression event linking the PR and HD genes arose in a common ancestor of all bipolar species by estimating the level of trans-specific polymorphism in gene genealogies, *i.e.* alleles segregating by mating type rather than by species. Indeed, if a gene became linked to the

mating-type genes in a common ancestor of two species, the a_1- and a_2-linked alleles would diverge before the speciation event. In order to get further insights into the question of whether the HD-PR linkage was ancestral to all bipolar species studied, we compared the gene content of the mating-type chromosomes among bipolar species and to the inferred ancestral gene content on separate HD and PR mating-type chromosomes.

We found that *M. lagerheimii* retained the same gene order as *M. intermedium* which strongly suggested that tetrapolarity has been retained in *M. lagerheimii*, and therefore that there had been multiple independent HD-PR linkage events. Moreover, from the ancestral tetrapolar state with unlinked PR and HD genes on two distinct mating-type chromosomes, we found that bipolarity evolved through distinct fission/fusion events of the PR and HD mating-type chromosomes at least five times. Among the studied species, only *M. lychnidis-dioicae, M. violaceum sensu stricto* and *M. silenes-dioicae* shared the same ancestral linkage event. In agreement with the genomic comparison, we only found trans-specific polymorphism for the genes between the PR and HD genes for these three species. We also reported two young evolutionary strata in *M. v. paradoxa* and one in *M. v. caroliniana*.

The multiple convergent supergene evolution reported in this study challenges evolution as being "utterly unpredictable and quite unrepeatable" (Gould 1990). Instead, we showed that a common and strong selection can lead to similar phenotypes which can result from different genomic changes. To some extent, this shows that evolution can be repeatable.

My contribution to this study was to perform (i) the synteny analysis and (ii) the d_S calculation with Sara Branco and Ricardo Rodriguez C. de la Vega to reconstruct evolutionary scenarios of mating-type chromosome evolution for the four newly sequenced bipolar *Microbotryum* species. We were asked by a referee to add population data to further support the inference of recombination suppression in the young evolutionary strata. To do so, we investigated whether, for genes inferred to be in regions linked to mating type loci, alleles associated to a given mating-type were more similar with each other than with alleles associated with the alternative mating type. I first designed primers to sequence some targeted genes, that I amplified by PCR after sporidia isolation on petri dishes and DNA extraction. The resulting sequences poorly aligned onto the targeted genes or elsewhere on the genome while the DNA samples were of good quality and no contamination was suspected. I therefore re-extracted DNA from *Microbotryum* strains previously sampled to send whole genomes to be sequenced using Illumina technology. I analysed the resulting reads by cleaning, trimming and mapping. I

performed gene genealogies using these population data and worked with Damien de Vienne (Université de Lyon 1) to think and code a method to quantify in an automatic way the degree of clustering of a_1 and a_2-associated alleles in gene genealogies.

Multiple convergent supergene evolution events in mating-type chromosomes

Sara Branco[1,2], Fantin Carpentier[1], Ricardo C. Rodríguez de la Vega [1], Hélène Badouin[1,3], Alodie Snirc[1], Stéphanie Le Prieur[1], Marco A. Coelho[4], Damien M. de Vienne[3], Fanny E. Hartmann[1], Dominik Begerow[5], Michael E. Hood[6] & Tatiana Giraud [1]

Convergent adaptation provides unique insights into the predictability of evolution and ultimately into processes of biological diversification. Supergenes (beneficial gene linkage) are striking examples of adaptation, but little is known about their prevalence or evolution. A recent study on anther-smut fungi documented supergene formation by rearrangements linking two key mating-type loci, controlling pre- and post-mating compatibility. Here further high-quality genome assemblies reveal four additional independent cases of chromosomal rearrangements leading to regions of suppressed recombination linking these mating-type loci in closely related species. Such convergent transitions in genomic architecture of mating-type determination indicate strong selection favoring linkage of mating-type loci into cosegregating supergenes. We find independent evolutionary strata (stepwise recombination suppression) in several species, with extensive rearrangements, gene losses, and transposable element accumulation. We thus show remarkable convergence in mating-type chromosome evolution, recurrent supergene formation, and repeated evolution of similar phenotypes through different genomic changes.

[1] Ecologie Systématique Evolution, Bâtiment 360, Univ. Paris-Sud, AgroParisTech, CNRS, Université Paris-Saclay, 91400 Orsay, France. [2] Department of Microbiology and Immunology, Montana State University, Bozeman, MT 59717, USA. [3] Univ Lyon, Université Lyon 1, CNRS, Laboratoire de Biométrie et Biologie Evolutive UMR5558, F-69622 Villeurbanne, France. [4] UCIBIO-REQUIMTE, Departamento de Ciências da Vida, Faculdade de Ciências e Tecnologia, Universidade NOVA de Lisboa, 2829-516 Caparica, Portugal. [5] Ruhr-Universitat Bochum, AG Geobotanik Gebaude ND 03/174 Universitatsstraße, 15044780 Bochum, Germany. [6] Department of Biology, University of Virginia, Gilmer 051, Charlottesville, VA 22903, USA. These authors contributed equally: Sara Branco, Fantin Carpentier, Ricardo C. Rodríguez de la Vega. These authors jointly supervised this work: Michael E. Hood, Tatiana Giraud. Correspondence and requests for materials should be addressed to T.G. (email: tatiana.giraud@u-psud.fr)

Gould's view that evolution is "utterly unpredictable and quite unrepeatable"[1] has long prevailed. It is difficult to test the repeatability of evolution, but such tests are essential for understanding biological diversification in response to selection[2,3]. Cases of convergent evolution following similar selective pressures provide ideal opportunities for assessing the repeatability of evolutionary processes and unraveling the proximal and ultimate mechanisms generating diversity[2,4]. Examples of convergent evolution include ecological morphs in Nicaraguan crater lake cichlid fishes[5], cave morphs in Mexican cavefishes[6], resistance to toxic compounds in animals[7], and lactase persistence in humans[8]. However, few examples have been studied in detail and many unresolved questions remain, including the frequency of convergent evolution, the genetic mechanisms underlying convergent trait evolution, whether convergence is widespread in organisms other than plants and animals, and the phylogenetic scales at which it occurs.

Supergenes (the beneficial linkage of genes controlling different ecological traits by recombination suppression) are striking cases of adaptation, arising by conspicuous changes in genomic architecture. As such, supergenes can be good models for assessing the predictability and proximate/ultimate causes of convergent evolution. Although we still know little about supergene prevalence and evolutionary importance[9–11], interesting cases have been reported, including the non-recombining genomic regions controlling multiple wing color patterns in butterflies[12] and polymorphic social behavior in ants[10,11]. Chromosomes involved in sexual compatibility often also have large non-recombining regions, and the early stages in development of these regions can be considered to constitute supergenes[13]. Recombination suppression linking different traits involved in sexual compatibility has been documented in the sex chromosomes of animals and plants[13,14], the mating-type chromosomes of algae[15] and fungi[16–21], and self-incompatibility loci in plants[22]. Recombination cessation not only maintains beneficial allelic

combinations but also reduces selection efficacy, leading to genomic decay and the accumulation of transposable elements (TEs)[23]. The frequency and proximal mechanisms of recombination suppression and the tempo of genomic degeneration remain unclear[13,14,24]. Chromosomes with recent recombination suppression events are ideal for investigating the initial steps of supergene formation and degeneration[25–27].

Fungi provide excellent systems for studying the causes, consequences, and frequency of recombination cessation, as they have diverse mating-type-determining systems involving multiple genes[28] and mating-type chromosomes with recent recombination suppression events[16–21]. Most basidiomycetes (mushrooms, rusts, and smut fungi) have two independently segregating loci controlling mating type at the haploid stage: (1) the PR locus, containing a pheromone receptor gene and one to several mating pheromone genes involved in pre-fertilization compatibility, and (2) the HD locus, encoding two homeodomain transcription factors responsible for post-fertilization compatibility[28]. Linkage between PR and HD loci underlies a major transition determining reproductive compatibility in these fungi[17,20,29]. Such linkage was long thought to be rare but is beneficial in selfing mating systems[30,31] (Supplementary Fig. 1). Linkage between PR and HD loci results in large regions controlling multiple mating-type functions (pre- and post-fertilization compatibility), which can be described as supergenes, as they represent a beneficial allelic combination where linkage increases fitness.

Here we describe a remarkable case of multiple convergent events of beneficial linkage between the PR and HD mating-type loci, corresponding to the repeated formation of supergenes in multiple young mating-type chromosomes across closely related fungi. We studied species of anther-smut fungi (*Microbotryum violaceum* complex; Fig. 1), a group of selfing pathogens. The mating-type chromosomes in this group were first described in *Microbotryum lychnidis-dioicae*[32], in which the mating-type loci are linked by a large region without recombination resulting from
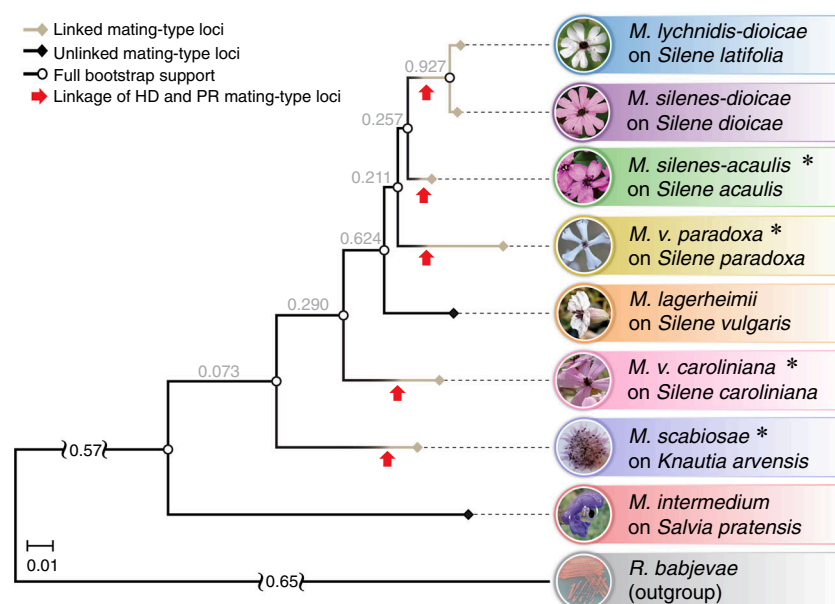
Fig. 1 Phylogenies of anther-smut fungi and their breeding systems. Phylogenetic tree of the studied *Microbotryum* species (shown in the anthers of their host plants) and the outgroup *Rhodotorula babjevae*, based on 4229 orthologous genes. Species whose genomes were obtained in the present study are indicated by asterisks. Branch color and symbol indicate linked (gray branches and diamonds) or unlinked (black branches and diamonds) mating-type loci. The white circles indicate full bootstrap support. Red arrows indicate independent mating-type locus linkage events. Tree internode certainty with no conflict bipartitions (the normalized frequency of the most frequent bipartition across gene genealogies relative to the summed frequencies of the two most frequent bipartitions) is provided below the branches, indicating good support for the nodes. Relative certainty for this tree is 0.397

the fusion of the entire ancestral *PR* chromosome and one arm of the ancestral *HD* chromosome (Fig. 2)[16]. The two mating-type loci were precisely the ancestral limits of the initial recombination suppression event, with their linkage resulting in a supergene. Initial mating-type loci linkage was followed by further stepwise

expansions of suppressed recombination beyond mating-type genes[16]. These successive steps expanded the linked region and led to the formation of evolutionary strata with decreasing allelic divergence between mating-type chromosomes at increasing distance from the mating-type-determining genes[16], as observed
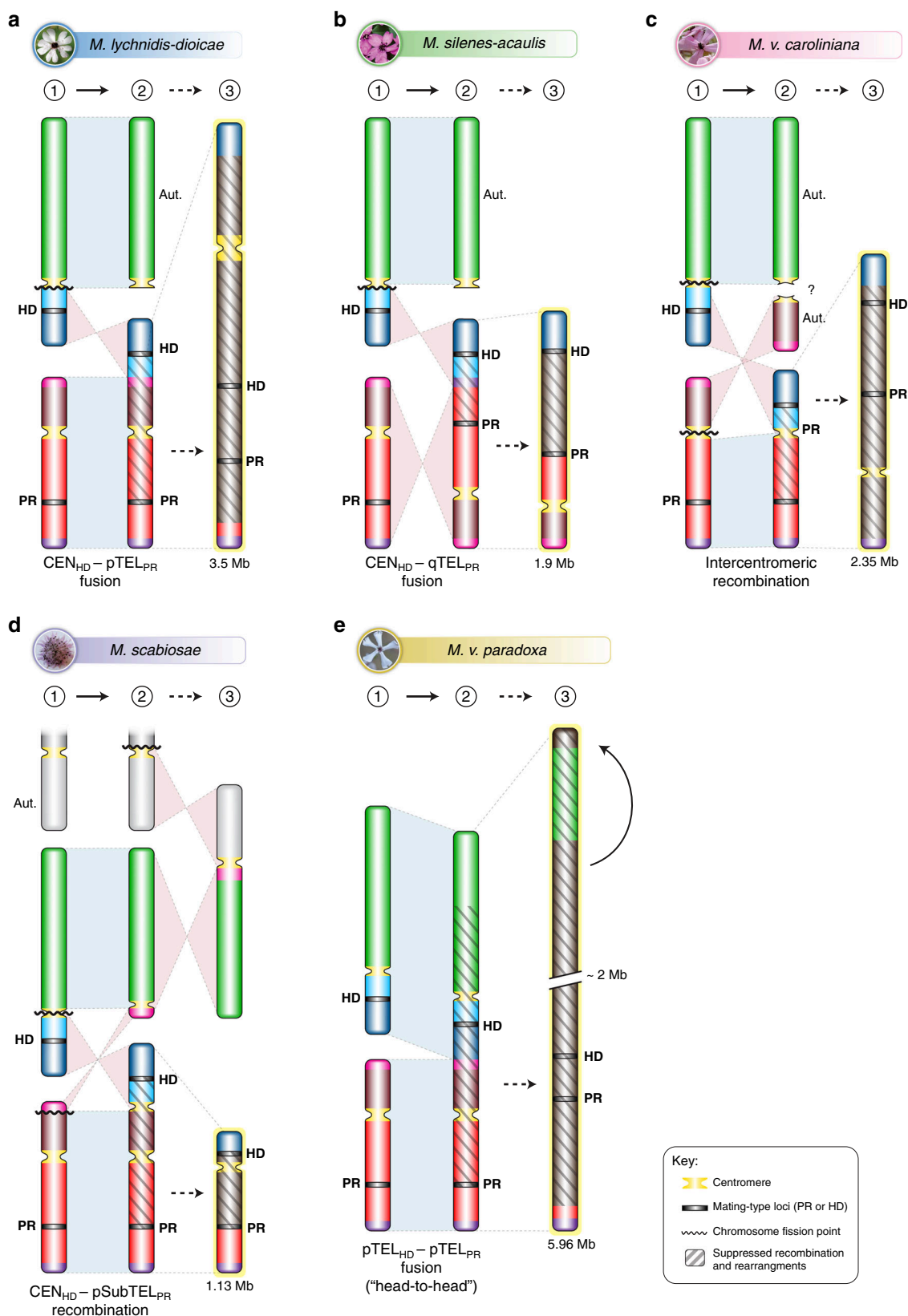
in animal and plant sex chromosomes[33] and likely other fungal mating-type chromosomes[17,18]. In contrast to the initial mating-type loci linkage event, these additional evolutionary strata did not control any traits for which linkage to mating-type would be beneficial[16]. Instead, the recombination suppression events occurring after mating-type locus linkage probably evolved to shelter deleterious alleles or through neutral rearrangements[16,24].

The majority of *Microbotryum* anther-smut fungi species display linked mating-type loci, but *M. lagerheimii* and *M. intermedium* have mating-type loci located on separate chromosomes[16,34] (Fig. 1). Given the phylogeny of anther-smut fungi (Fig. 1), a previous study based on parsimony inferred ancestral linkage between the *PR* and *HD* mating-type loci in the *Microbotryum* clade, with a reversal to unlinked mating-type loci in *M. lagerheimii*[34]. However, the distantly related species *M. lagerheimii* and *M. intermedium* have highly collinear mating-type chromosomes[16], whereas gene order is rearranged in non-recombining regions across species with linked *PR* and *HD* mating-type loci[16,19]. The collinearity between *M. lagerheimii* and *M. intermedium* mating-type chromosomes suggests that the ancestral state and gene order have been retained in these species. This raises the alternative hypothesis of a remarkable number of independent transitions linking the mating-type loci across anther-smut fungi (Fig. 1).

Using high-quality assemblies of closely related species and the ancestral gene order retained in *M. lagerheimii*[16], we uncovered four independent events of mating-type locus linkage in addition to the previously identified supergene[16]. The various *Microbotryum* species achieved mating-type locus linkage through different chromosomal rearrangements and have non-recombining regions of different sizes, ages, and gene contents. Our results show that supergenes can evolve frequently and that natural selection can repeatedly lead to similar phenotypes through multiple evolutionary trajectories and different genomic changes, consistent with repeatable evolution. We also document repeated and independent formation of evolutionary strata, with stepwise expansions of non-recombining regions beyond mating-type genes and provide evidence for increasing genomic decay in regions with a longer history of recombination suppression.

Results

Five independent routes for linking mating-type loci. We inferred the evolutionary histories of mating-type chromosomes in multiple anther-smut fungi by comparing high-quality genome assemblies of eight *Microbotryum* species (Fig. 1). We obtained haploid genome assemblies for both mating types (a_1 and a_2) of four *Microbotryum* species with full linkage of *PR* and *HD* mating-type loci, as previously shown by progeny segregation[34]. We also studied available haploid genome sequences of four additional *Microbotryum* species[16] (Fig. 1; Supplementary Table 1). We used the *M. lagerheimii* genome as a proxy for ancestral gene order[16] due to its unlinked *PR* and *HD* loci and very few rearrangements relative to the distantly related *M. intermedium* species[16]. Whole-genome BLAST comparisons

revealed five different chromosomal rearrangements and fusions underlying the linkage between the *HD* and *PR* loci, one in each of the four newly assembled genomes and the fifth at the base of the previously analyzed clade containing *M. lychnidis-dioicae* and *M. silenes-dioicae*[16] (Figs. 1 and 2).

The different rearrangements led to variation across species in mating-type chromosome size and composition, as well as in non-recombining region length and captured gene content (Supplementary Table 1). Recombination ceased at different times, as shown by the different levels of synonymous divergence (d_S) between alleles associated with the a_1 and a_2 mating types for genes ancestrally located between the *HD* and *PR* loci (Figs. 3 and 4). For genes linked to mating-type loci, alleles associated with alternative mating types accumulate differences over time since the linkage event, whereas genes unlinked to mating-type loci are highly homozygous in these selfing fungi (with virtually no divergence between alleles present in the a_1 or a_2 haploid genomes within diploid individuals; Supplementary Figs. 2a–c). The absence of trans-specific polymorphism in genes ancestrally located between the *PR* and *HD* loci following chromosome fusion provided further evidence for the existence of five independent fusion events. Specifically, alleles of genes between mating-type loci clustered by species and not by mating type, demonstrating that their linkage to mating-type loci occurred after speciation events (Fig. 4; Supplementary Fig. 3). Only *M. lychnidis-dioicae* and *M. silenes-dioicae* displayed trans-specific polymorphism in the genomic regions ancestrally located between the *PR* and *HD* loci. Alleles associated with the a_1 mating type of both species consistently clustered together, as did alleles associated with the a_2 mating type, indicating that *PR–HD* linkage predated the speciation event in this clade (Fig. 4; Supplementary Fig. 3). Unlike mating-type chromosomes, autosomes were highly collinear between mating types and with no evidence of widespread interchromosomal rearrangements across species (Supplementary Fig. 4) or trans-specific polymorphism (only 1 of the 4229 single-copy shared autosomal genes displayed trans-specific polymorphism, and even then, only between *M. lychnidis-dioicae* and *M. silenes-dioicae*). The interchromosomal rearrangements and recombination suppression were thus restricted to the mating-type chromosomes and repeatedly led to regions of suppressed recombination bordered by the *HD* and *PR* loci. This indicates that the mating-type chromosome rearrangements linking mating-type genes were selected for, forming adaptive supergenes.

In *M. lychnidis-dioicae*, mating-type locus linkage was achieved by the fusion of the putative centromere end of the *HD* chromosome short arm to the distal end of the *PR* chromosome short arm (Fig. 2a). This evolutionary transition occurred before the divergence of *M. lychnidis-dioicae* and *M. silenes-dioicae*, as shown by their similar chromosome structures and trans-specific polymorphism, as previously reported[16]. Genealogies of genes ancestrally located between the *HD* and *PR* loci and calibrated using the date of speciation between *M. lychnidis-dioicae* and *M. silenes-dioicae*[35] indicated that the mating-type loci in this clade became linked 1.2 million years (MY) ago (Fig. 4). The timing of

Fig. 2 Routes of mating-type chromosome evolution in *Microbotryum*. Model for mating-type chromosomal rearrangement events, as inferred from comparisons with the two mating-type chromosomes of *M. lagerheimii* (used as a proxy for the ancestral mating-type chromosomes in the genus[16]). Mating-type chromosome content across *Microbotryum* species is illustrated by colors referring to different parts of the two *M. lagerheimii* mating-type chromosomes (Supplementary Figs. 5–8). The inferred ancestral locations of putative centromeres and mating-type loci are indicated in yellow and black, respectively, and the regions of suppressed recombination are dashed. Chromosome sizes are indicated by their relative scales; the last stage in the evolution of recombination suppression often involves increases in chromosome size due to the accumulation of repetitive elements. Mating-type chromosome evolution in **a** *M. lychnidis-dioicae*, **b** *M. silenes-acaulis*, **c** *M. violaceum caroliniana*, **d** *M. scabiosae*, and **e** *M. v. paradoxa*, in which the top edge of the mating-type chromosome corresponds to a rearrangement from the middle of the chromosome, supporting complete recombination suppression up to the end of the chromosome
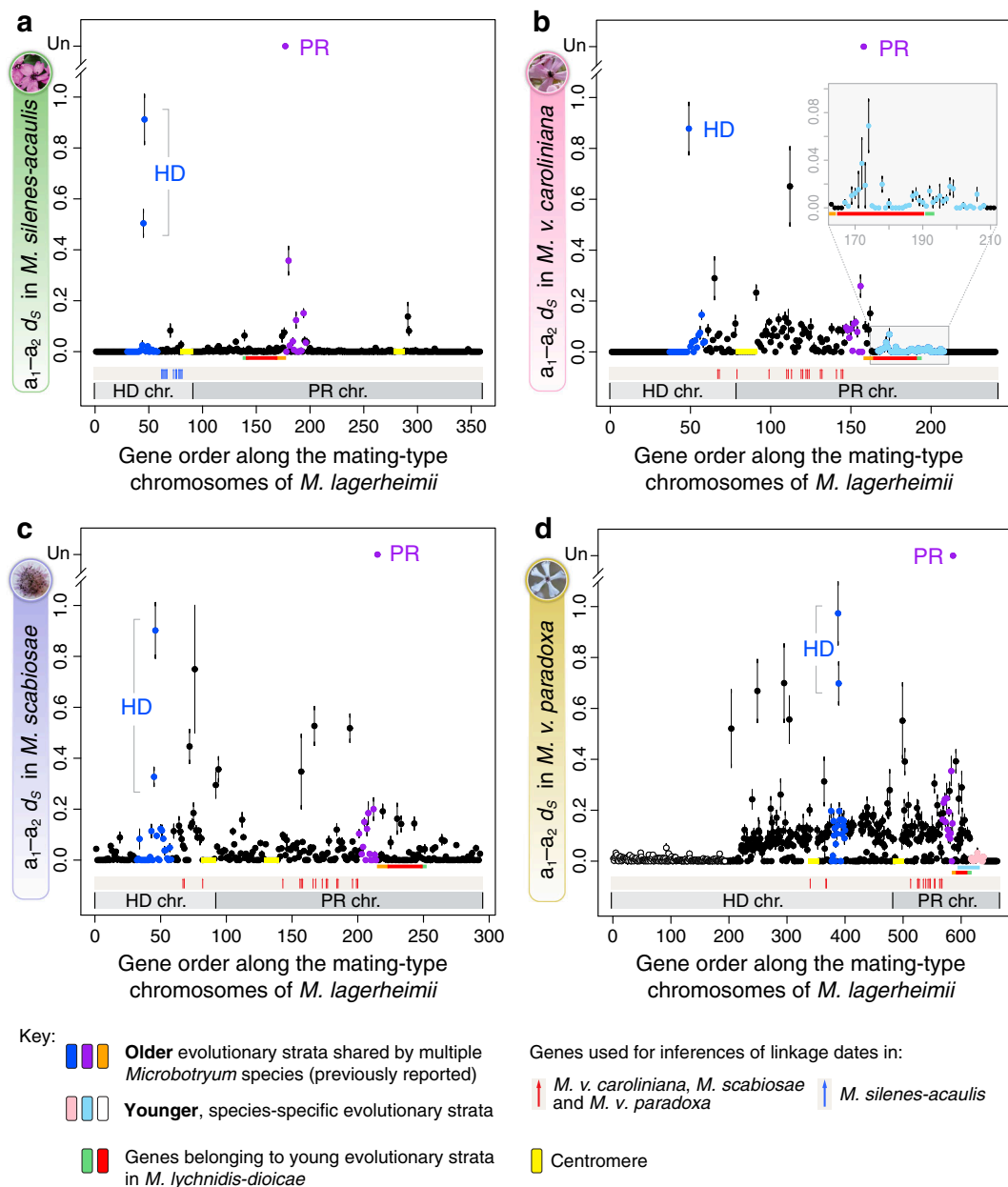
Fig. 3 Divergence between a_1- and a_2-associated alleles. Per-gene synonymous divergence and standard error ($d_S \pm SE$) between alleles associated with the a_1 and a_2 mating types within *Microbotryum* diploid individuals, following the ancestral gene order for the mating-type chromosome. Synonymous divergence is plotted against the genomic coordinates of the a_1 mating-type chromosomes of *M. lagerheimii* for all single-copy genes common to both mating-type chromosomes. The limits of the PR and HD *M. lagerheimii* mating-type chromosomes are indicated and oriented according to the fusion in each species (i.e., not in the same orientation in all species). Divergence between the a_1 and a_2 pheromone receptor (PR) genes was too extensive and d_S could not be calculated (depicted as "Un" for unalignable). The yellow boxes indicate the positions of *M. lagerheimii* putative centromeres. The red vertical arrows at the bottom indicate the 17 genes used for inferring HD–PR linkage dates in all species except for *M. silenes-acaulis*, for which we used a restricted set of 13 genes ancestrally located between the HD locus and the putative centromere (blue vertical arrows). Ancient evolutionary strata that evolved at the base of the *Microbotryum* clade are indicated in purple (around PR) and blue (around HD), as in the previous study in which they were discovered[16]. The genes involved in the more recent evolutionary strata previously identified in *M. lychnidis-dioicae*[16] are indicated with red and green bars at the bottom. **a** *M. silenes-acaulis*; **b** *M. v. caroliniana*, with a recent stratum indicated in light blue and enlarged in an inset; the current location of these genes is indicated in Supplementary Fig. 6a; **c** *M. scabiosae*; **d** *M. v. paradoxa*, with recent strata depicted in pink and white (the current location of these genes is indicated in Supplementary Fig. 8a). The light blue bar at the bottom indicates the genes involved in the young evolutionary stratum of *M. v. caroliniana*

PR–HD linkage was inferred from the divergence between alleles associated with the a_1 and a_2 mating types at 17 genes (red vertical arrows in Fig. 3). In these regions linked to mating-type loci, the alleles remained associated with the a_1 or a_2 mating type and diverged progressively with time since recombination suppression (Fig. 3). In *M. lychnidis-dioicae*, autosomes were highly syntenic and without differentiation between alleles in the sequenced a_1 and a_2 genomes derived from a single diploid individual[19].

In *M. silenes-acaulis*, PR–HD linkage resulted from a different and more recent chromosomal rearrangement. As in *M. lychnidis-dioicae*, the short arm of the ancestral *HD* chromosome

5
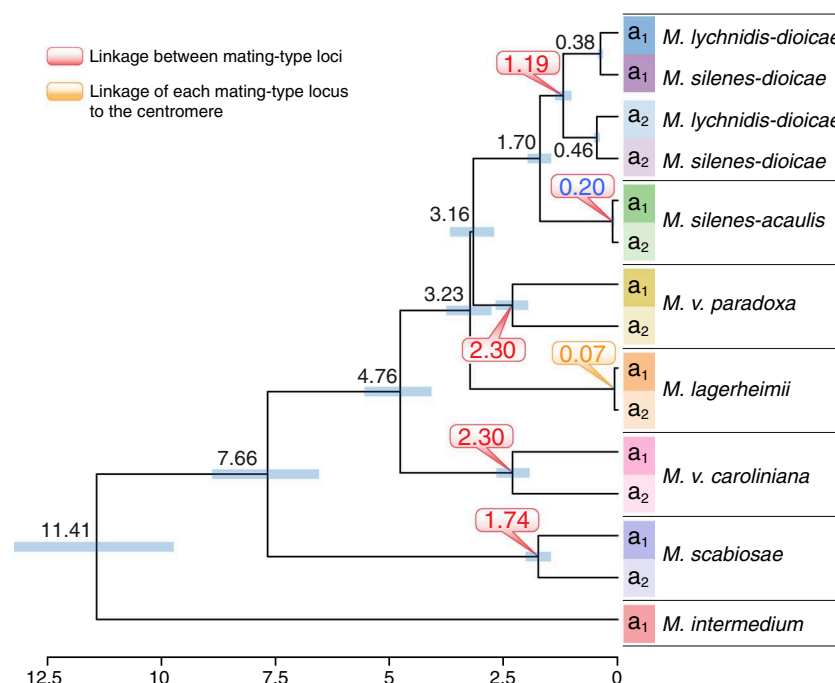
Fig. 4 Dates of mating-type loci linkage. Reconstructed phylogenetic tree based on 17 concatenated genes ancestrally located between the mating-type loci after chromosomal fusion in all the studied species with linked mating-type loci but *M. silenes-acaulis* and including alleles from both a_1 and a_2 genomes. Numbers on tree nodes indicate the inferred dates of speciation (in black) and the events of mating-type loci linkage, either one to each other or to their respective putative centromeres (in red and orange, respectively). The blue bars correspond to 95% confidence intervals. The scale at the bottom indicates the time before present (million years). None of the individual genes showed trans-specific polymorphism, except between the sister species *M. lychnidis-dioicae* and *M. silenes-dioicae* (Supplementary Fig. 3). We used a restricted set of 13 genes (Fig. 3) for estimating the *M. silene-acaulis* mating-type loci linkage because not all 17 genes were located in its non-recombining region

fused with the entire ancestral *PR* chromosome, but the putative centromere end of the *HD* chromosome arm fused to the opposite end of the *PR* chromosome (Fig. 2b; Supplementary Fig. 5b, c). The much lower levels of synonymous divergence (d_S) between the alleles of genes positioned between the *PR* and *HD* mating-type loci (Fig. 3a; Supplementary Fig. 5a and Supplementary Table 2)[16] indicated a more recent fusion event. Because not all the 17 genes used for dating mating-type loci linkage in the other species were ancestrally located between the mating-type loci in *M. silenes-acaulis* (due to the PR fusion in the opposite direction in this species, Fig. 2), we used a restricted set of 13 genes ancestrally located between the *HD* locus and the putative centromere (blue vertical arrows in Fig. 3) for dating *HD–PR* linkage in this species. We estimated that mating-type loci linkage occurred ca. 0.2 MYs ago in *M. silenes-acaulis* (Figs. 3 and 4). Further evidence for a more recent chromosome fusion was also provided by the small number of rearrangements between the a_1 and a_2 mating-type chromosomes in *M. silenes-acaulis* (Supplementary Fig. 5a, b), contrasting with the extensive rearrangements observed in the *M. lychnidis-dioicae* non-recombining region[19]. In *M. silenes-acaulis*, a large inversion encompassing the region between the *PR* and *HD* loci (Supplementary Fig. 5a) may have contributed directly to the recombination suppression linking the two mating-type loci. All *M. silenes-acaulis* autosomes displayed high levels of synteny and d_S values of zero between the sequenced a_1 and a_2 genomes originating from a single diploid individual (Supplementary Figs. 2a and 4a).

Other specific rearrangements led to mating-type locus linkage in the remaining species. All these rearrangements appeared older than those occurring in the ancestor of *M. lychnidis-dioicae* and *M. silenes-dioicae*, as shown by the higher d_S levels (Fig. 3;

Supplementary Table 2) and the inferred earlier occurrence of recombination suppression based on the 17 gene set (Fig. 4). One of the oldest events was estimated to have occurred in *M. v. caroliniana*, about 2.3 MYs ago (Fig. 4). Unlike those of *M. lychnidis-dioicae* and *M. silenes-acaulis*, the *M. v. caroliniana* mating-type chromosome contained a single ancestral *PR* chromosome arm (Fig. 2c; Supplementary Fig. 6b)[16]. Higher d_S values between the alleles of genes ancestrally positioned between the *PR* and *HD* loci (Fig. 3b, Supplementary Table 2) and massive rearrangements in the non-recombining region (Supplementary Fig. 6a) provided further evidence for an earlier onset of recombination cessation in *M. v. caroliniana* than in *M. silenes-acaulis*. All *M. v. caroliniana* autosomes were syntenic and with d_S values of zero between the sequenced a_1 and a_2 genomes isolated from a single diploid individual (Supplementary Figs. 2b and 4b).

In *M. scabiosae*, a more recent event (1.7 MY old; Fig. 4) linked the mating-type loci following a chromosomal rearrangement similar to that in *M. lychnidis-dioicae* and *M. silenes-dioicae* (Fig. 2d) but with one extremity of the ancestral PR chromosome becoming incorporated into an autosome (black fragment in the outer track in Supplementary Fig. 7b; Fig. 2d). This particular configuration suggests ectopic recombination within a chromosome arm rather than rearrangement at putative centromeres as described above (Fig. 2). Consistent with the more recent recombination suppression, *M. scabiosae* displayed less extensive rearrangements between the a_1 and a_2 mating-type chromosomes than *M. v. caroliniana* (Supplementary Fig. 7a). Several large inversions nevertheless occurred between mating-type chromosomes. We were unable to sequence two meiotic products of a single diploid individual for *M. scabiosae*, which probably

explains the allelic variation observed between the sequenced a_1 and a_2 genomes even for pseudo-autosomal regions (PARs) and autosomes (Fig. 3c; Supplementary Figs. 2d and 7a). Nevertheless, the *M. scabiosae* mating-type chromosomes still appear to be exceptional in terms of rearrangements and divergence between the a_1 and a_2 genomes compared to autosomes (Fig. 3c; Supplementary Figs. 2d, 4c, and 7a).

Unlike those of all other species considered, *M. v. paradoxa* mating-type chromosomes resulted from the fusion of the entire ancestral *PR* and *HD* chromosomes (Fig. 2e; Supplementary Fig. 5k). This species experienced one of the earliest mating-type locus linkage events, with recombination suppression occurring about 2.3 MY ago (Fig. 4). This estimated age of recombination suppression is consistent with the high levels of rearrangements (Supplementary Fig. 8b) and high d_S values (Fig. 3d) observed for *M. v. paradoxa* mating-type chromosomes. We obtained non-zero d_S values across one side in most *M. v. paradoxa* autosomes, with zero d_S values along the remaining length, as expected after an outcrossing event followed by a selfing event (Supplementary Fig. 2e). The d_S values on autosomes remained much lower than those between the *HD* and *PR* loci on mating-type chromosomes (Fig. 3d; Supplementary Fig. 2e). There was also a very high degree of autosome synteny between the a_1 and a_2 genomes (Supplementary Fig. 4d), suggestive of ongoing recombination, as well as inter-species synteny, contrasting with the interchromosomal and intrachromosomal rearrangements observed for mating-type chromosomes (Supplementary Fig. 8a).

Gene genealogies provided further support for the existence of five independent mating-type locus linkage events. No trans-specific polymorphism was found for any gene in the genomic regions ancestrally located between the *PR* and *HD* mating-type loci, other than in the sister species *M. lychnidis-dioicae* and *M. silenes-dioicae* (Fig. 4). Furthermore, with the exception of these two species, the inferred divergence date of alleles associated with the a_1 and a_2 mating types at genes ancestrally located between the mating-type loci following chromosome fusion was younger than at speciation events (Fig. 4).

Unlike all species described above, *M. lagerheimii* has unlinked mating-type loci (Fig. 1) despite also having a selfing mating system, as shown by the values of zero for d_S obtained for all autosomes (Supplementary Fig. 2e). In this species, each mating-type locus is instead linked to the putative centromere of its chromosome[34], yielding the same odds of gamete compatibility as mating-type locus linkage under intra-tetrad selfing (Supplementary Fig. 1b, c). The linkage between the mating-type loci and putative centromeres was inferred to be very recent, occurring only ca. 0.07 MY ago (Fig. 4).

Independent evolution of evolutionary strata. Along with repeated and independent evolution of mating-type loci linkage by distinct genome rearrangements, we also observed the convergent evolution of subsequent expansion of the non-recombining regions forming evolutionary strata beyond the mating-type genes across multiple species. Such young evolutionary strata were defined as genomic regions with non-zero divergence between the alleles found in a_1 and a_2 genomes but with lower levels of differentiation than for the genomic region ancestrally located between the PR and HD loci. We identified these young evolutionary strata by plotting d_S levels between the alternate alleles along the inferred ancestral mating-type chromosome gene order[16]. In organisms with high levels of selfing, such as *Microbotryum* fungi, d_S is zero or very low in most diploid individuals (reflecting very high homozygosity levels, Supplementary Fig. 2). Non-recombining regions are a notable exception, where the degree of differentiation between alleles associated with the a_1 and a_2 mating types constitutes a proxy for time since linkage to

mating-type loci. Using this approach, we detected evolutionary strata extending recombination suppression beyond mating-type genes, including the two known ancient strata around the *HD* and *PR* loci common to all *Microbotryum* species (blue and purple strata[16], Fig. 3), as well as younger clade-specific strata. Some of the genes in the ancient (blue and purple) strata had low d_S levels in some species, probably due to occasional gene-conversion events that reset the signal of divergence, as known to occur in fungal mating-type chromosomes[36,37].

We identified a young evolutionary stratum in *M. v. caroliniana* (light blue in Fig. 3b). This genomic region was located distally to the *PR* mating-type locus and had non-zero d_S values significantly lower than the mean d_S for genes ancestrally located between the *PR* and *HD* loci (Fig. 3d, Supplementary Table 2). The limit of the light-blue stratum was set at the most distal gene with a non-zero d_S value, as all autosomes had zero d_S values in the sequenced *M. v. caroliniana* diploid individual (Supplementary Fig. 2a). The light-blue stratum extended farther into the PAR than the most recent evolutionary strata in *M. lychnidis-dioicae* (red and green bars in Fig. 3b). The mean d_S value in this evolutionary stratum was not significantly different from that in the PARs (Supplementary Table 2), indicating that mating-type locus linkage was recent. Such stretches of within-individual non-zero d_S genes were restricted to non-recombining regions in the *M. v. caroliniana* diploid individual sequenced (Supplementary Fig. 2b), providing strong evidence for recombination suppression in the light-blue region. Gene order in this region was largely conserved between the a_1 and a_2 mating-type chromosomes (Supplementary Fig. 5b), demonstrating that recombination can be halted in the absence of inversions. A small localized inversion within this stratum (orange links in Supplementary Fig. 5d) provided further evidence for the lack of recombination in this region. The autosomes were completely collinear between the two haploid genomes in the sequenced diploid individual (Supplementary Fig. 4b).

Evolutionary strata extending beyond the genes involved in mating-type determination were also detected in *M. v. paradoxa*. In this species, d_S values were also highest in the non-recombining region ancestrally located between the *PR* and *HD* loci (Fig. 3d; Supplementary Table 2) and were lower, but non-zero, in the two distal regions. These regions thus likely constitute two additional young evolutionary strata (white and pink, Fig. 3d). The region of recombination suppression extended farther into the PARs than in any of the other studied species, to the extent that only a single, very small PAR was retained (Fig. 3d and Supplementary Fig. 8a). We confirmed the complete suppression of recombination suppression in the white stratum by identifying a small region at the extremity of the *M. v. paradoxa* mating-type chromosome corresponding to rearranged genes ancestrally located in the center of the chromosome (Fig. 2e, and shown in gray in the outer track in Supplementary Fig. 8b). The pink region on the other side of the mating-type chromosome, distal to the *PR* locus (Fig. 3d), had high d_S values but not higher than those of autosomes in the sequenced *M. v. paradoxa* individual and without inversions or rearrangements.

We confirmed recombination suppression in the pink, white, and light blue regions by sequencing multiple genomes for *M. v. caroliniana* and *M. v. paradoxa* (Supplementary Table 3). For genes linked to the mating-type loci, a_1- and a_2-associated alleles will differentiate to the point of forming two distinct clades in gene genealogies within species. Gene genealogies revealed such pattern for genes in the pink, white, and light blue regions, where the alleles associated with a given mating type were significantly more clustered than for genes in the PARs or in autosomes (Supplementary Fig. 9, Supplementary Table 4). In all three regions, all a_1 alleles branched in one clade and all a_2 alleles in

another clade in multiple gene genealogies (Supplementary Fig. 9), strongly supporting full linkage to mating type. Furthermore, the mean levels of polymorphism per mating type and per species were significantly lower in all the evolutionary strata than in the PARs (Supplementary Fig. 10, Supplementary Table 5), as expected in regions without recombination due to the lower effective population size. These findings indicated that the non-zero d_S values in the evolutionary strata within the sequenced individuals were due to recombination suppression rather than higher polymorphism levels.
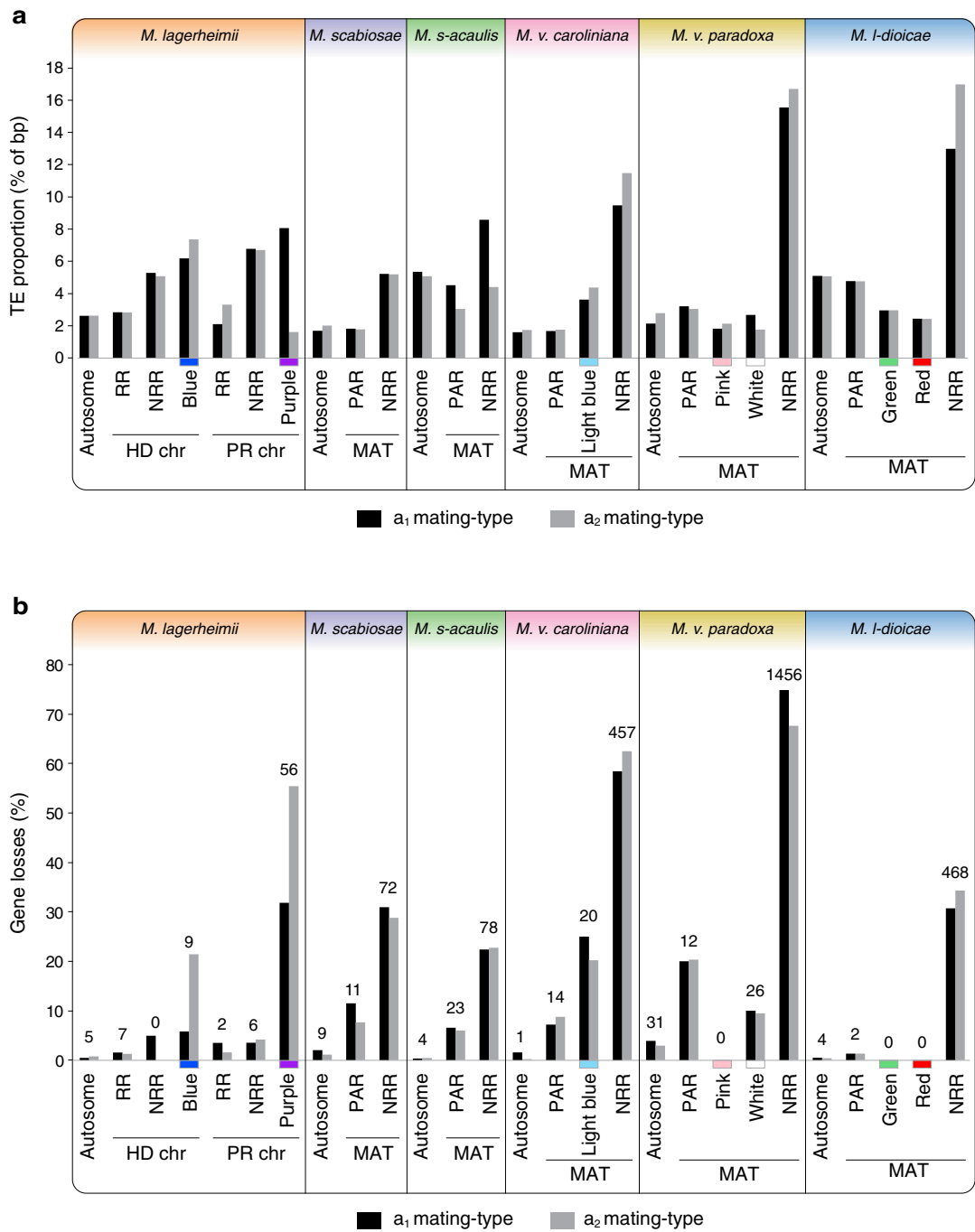
Fig. 5 Differential degeneration across strata and species. We quantified the TE content and gene loss in genomes of both mating types (a_1 and a_2) of all species under study. For each species, we measured the TE accumulation separately for one fully assembled autosome (as a control), recombining regions (RR), and non-recombining regions (NRR) on mating-type chromosomes (MAT), separating the youngest evolutionary strata (light blue, red, green pink, and white strata) from the remaining of the NRR where applicable. Strata were ordered from the youngest to the oldest per species. In *M. lagerheimii*, the NRRs correspond to the regions between the mating-type loci and the putative centromeres, while in the other species they mostly correspond to the regions ancestrally between the HD and PR loci. The purple and blue strata were too rearranged within the large non-recombining region to quantify their specific gene loss or TE content except in *M. lagerheimii*. **a** Transposable element (TE) content (percent of base pairs); **b** Gene loss (genes with an allele present in the genome of one mating type but absent from the genome of the opposite mating type). Numbers at the top of the bars indicate the numbers of genes missing in the a_2 mating-type chromosome, present only in the a_1 mating-type chromosome

Gene loss and TE accumulation. We found evidence of differential gene loss and TE accumulation across species and evolutionary strata. Levels of gene loss and TE content increased with the age of recombination suppression, both within and across species (Fig. 5). Even the youngest evolutionary strata showed footprints of genomic decay relative to recombining regions (Fig. 5). The regions ancestrally located between the PR and HD loci displayed higher levels of degeneration in species with older recombination suppression events than in species with more recent mating-type locus linkage. Higher TE loads resulted in chromosomes larger than those in the ancestral state and contributed to differences in size between mating-type chromosomes (Figs. 2 and 5). The PARs and youngest strata displayed little evidence of TE accumulation compared to autosomes and showed low but non-negligible levels of gene loss (Fig. 5). Mating type (a₁ versus a₂) had no significant effect on gene loss or TE content, while differences between species were significant (Supplementary Table 6). The onset of genomic degeneration is thus rapid, with further gradual accumulation of TEs. Gene loss was extensive on both mating-type chromosomes: the two species with the most ancient recombination suppression between the mating-type loci lost between 60 and 70% of genes in this region within 2.3 MY, and M. silenes-acaulis has already lost >20% of genes within 0.20 MY.

Discussion

We report an unprecedented case of convergent evolution, with five parallel recombination suppression events independently linking PR and HD mating-type loci in anther-smut fungi and generating megabase-long supergenes beneficial under selfing mating systems. We also reveal the convergent evolution of young evolutionary strata in multiple closely related species. Furthermore, our unique dataset suggests a progression of genomic decay in non-recombining mating-type chromosomes, with older regions of suppressed recombination displaying higher levels of gene loss and TE accumulation.

The convergent evolution of megabase-long supergenes in Microbotryum mating-type chromosomes, through distinct genomic rearrangements, has repeatedly led to the beneficial co-segregation of different mating-type functions (pheromones and pheromone receptors controlling pre-mating compatibility encoded by the PR locus, and the homeodomain proteins controlling post-mating compatibility encoded by the HD locus). A previous parsimony analysis indicated a very low probability of independent mating-type loci linkage events in anther-smut fungi and instead inferred reversal to the ancestral state of unlinked PR and HD loci in M. lagerheimii[34]. However, our analyses based on well-assembled chromosomes contradict these earlier inferences and reveal the striking occurrence of repeated convergent events linking mating-type loci through distinct genomic rearrangements. Further evidence for convergent linkage events is provided by differences in synonymous divergence between alleles associated with alternative mating types, the absence of trans-specific polymorphism in genes ancestrally located between the PR and HD loci, and differences in inferred linkage dates, with linkage occurring after speciation in at least five lineages. The existence of numerous other species with linked PR and HD mating-type loci across the Microbotryum genus[34,38] suggests the existence of many more independent events of mating-type locus linkage and convergent supergene formation in this genus. The occurrence of recombination suppression linking mating-type loci to each other or to centromere has been documented in a few interspersed lineages across fungi[17,18,21,29,31], but we provide here the first report of repeated convergent mating-type locus linkage events in multiple closely related species.

Repeated evolution of mating-type locus linkage in multiple closely related species implies very strong selection, which is supported by the lack of similar large-scale rearrangements or recombination suppression in autosomes. In selfing-based mating systems, as observed in all species of Microbotryum studied to date[39,40], HD–PR linkage increases the odds of compatibility between the gametes of a given diploid individual (Supplementary Figs. 1a, c) and is expected to be favored by selection. Interestingly, although the selfing species M. lagerheimii has unlinked mating-type loci[34], the HD and PR loci are linked to the putative centromeres of their corresponding chromosomes[34], which also increases gamete compatibility under selfing via intra-tetrad mating (automixis) (Supplementary Fig. 1b).

Our study provides invaluable insight into the frequency and importance of supergene evolution, a topic currently under explored[9]. Other examples of supergenes include the linkage of genes involved in mating types in plants and algae, mimicry wing patterns in butterflies, and complex social behavior in ants[10,11,13,15,22]. The repeated convergent evolution of non-recombining regions in anther-smut fungi provides strong support for the view that chromosomal rearrangements and the formation of supergenes are frequent and play an important role in evolution and adaptation[10,13,41]. The existence of repeated transitions in the genomic architecture of mating-type determination, following different chromosomal rearrangements, illustrates the power of natural selection and high genomic fluidity in shaping adaptation. Our findings suggest that natural selection drives evolution along trajectories leading to similar phenotypic outcomes through different genomic changes[2,4]. Very few instances of convergent evolution have been documented beyond the textbook examples of Nicaraguan crater lake cichlid fishes[5], cave morphs in Mexican cavefishes[6], and non-recombining chromosomes controlling polymorphic social behavior in ants[10,11].

We also found striking convergence in the stepwise extension of recombination suppression beyond mating-type genes, generating independent evolutionary strata in several Microbotryum species. The evolutionary causes leading to such strata devoid of mating-type genes are more likely to be the sheltering of deleterious alleles or neutral rearrangements than beneficial gene linkage[16,24,42]. Finding independent evolutionary strata in multiple closely related species provides further support for the occurrence of repeated evolution towards similar chromosomal states.

Our study provides important clues to the proximal mechanisms underlying the evolution of recombination suppression. Contrary to the common view that chromosomal inversions play a major role in preventing recombination[14], we found that recombination cessation can occur with the conservation of collinearity, as previously documented in fungi[16,20,21]. The occurrence of mating-type locus linkage via different routes reveals a high degree of genomic fluidity. Chromosomal rearrangements occurred frequently at putative centromeres, as in a recently reported case in the human fungal pathogen Cryptococcus neoformans[20]. These repeat-rich regions are highly labile and some of the inferred fusions in anther-smut fungi encompassed two ancestral putative centromeres in the same chromosome, as in M. v. paradoxa.

By the examination of repeated supergene formation events of contrasting ages, our results add further insights on the tempo of genomic decay and TE accumulation after recombination suppression. We found that gene loss occurred more rapidly than rearrangements or repeat accumulation. TE accumulation rates differed between species, probably because the Microbotryum species-specific TE loads[43] affect transposition rates. The high levels of degeneration observed in the two mating types probably

resulted from less efficient selection, due to lack of recombination and the sheltering of deleterious mutations in a permanently heterozygous state, with only very brief periods of haploid selection restricted to the meiotic tetrad stage[39]. Gene loss in these fungal mating-type chromosomes was more rapid than in the Y chromosome of the plant *Silene latifolia*[44], likely because plant sex chromosome degeneration is delayed by haploid purifying selection, unlike in animals or *Microbotryum* fungi[23]. Contrasting with sex chromosomes, genomic degeneration in *Microbotryum* mating-type chromosomes was not asymmetric (with no significant effect of mating-type on TE content or gene loss), as expected for organisms with an obligate heterozygous mating-type or sex chromosomes[45,46]. Lethal alleles linked to mating type were found at relatively high frequencies in natural populations of several *Microbotryum* species[47], preventing haploid growth in vitro but maintained through high levels of intra-tetrad mating. We also detected non-negligible levels of gene loss in the PARs. This may reflect the existence of very recent, undetected evolutionary strata or lower rates of recombination in the PARs compared to fully recombining autosomes[48]. This second hypothesis is consistent with the suggestion that partial deleterious allele sheltering in the PARs may account for evolutionary strata[16,24,42]: low recombination rates in the PARs would allow the accumulation of deleterious alleles, leading to selection for further recombination suppression and the permanent sheltering of these deleterious alleles.

In conclusion, our findings reveal remarkable repeated convergence in young mating-type chromosomes in closely related species, with supergenes evolving rapidly and frequently. Furthermore, our study shows that natural selection can repeatedly lead to similar phenotypes through multiple different evolutionary trajectories and genomic changes, rendering evolution predictable. The very recent advances in sequencing technologies yielding high-quality genome assemblies are allowing in-depth studies of chromosomal architecture and documenting the importance and prevalence of supergenes[9], as well as the high degree of genomic fluidity and convergence[4]. Future studies will certainly lead to the identification of many more cases of beneficial gene linkage, as predicted from evolutionary theory[41].

Methods

Strains, DNA extraction, and sequencing. *Microbotryum violaceum* is a plant pathogen species complex that includes recently recognized cryptic and host specialized species, which have not all been formally named yet. For species without Latin names, we used *M. violaceum* (the terminology used to denote the whole species complex) followed by the name of the host plant, as is typically done in phytopathology for host races or *formae speciales*. We isolated a₁ and a₂ haploid cells from the following species: *M. violaceum caroliniana* parasitizing *Silene caroliniana* (strain 1250, Virginia Beach, USA, GPS Coord.: 36°54′36.0″N 76°02′24.0″W), *M. violaceum paradoxa* parasitizing *Silene paradoxa* (strain 1252, near Florence, Italy, GPS Coord.: 43°32′35.7″N 11°21′35.1″E), *M. silenes-acaulis* parasitizing *S. acaulis* (strain 1248, La Grave, France, GPS Coord.: 45°01′32.9″N 6°16′22.9″), and *M. scabiosae* parasitizing *Knautia arvensis* (strain 1118, Vosges, Retournemer lake, near Colmar, France, GPS Coord.: 48°03′00.0″N 6°59′00.0″E). The a₁ and a₂ haploid cells were isolated from single tetrads using micro-manipulation for all strains except *M. scabiosae*, in which they were isolated from different teliospores.

DNA was extracted using a Carver hydraulic press (reference 3968, Wabash, IN, USA) for breaking cell walls and the Qiagen Anion-exchange columns Ref 10243 together with the buffers Ref 19060 (Courtaboeuf, France) for purifying DNA while avoiding fragmenting DNA. Haploid genomes were sequenced using the P6/C4 Pacific Biosciences SMRT technology (UCSD IGM Genomics Facility La Jolla, CA, USA).

Assembly and annotation. Genome assemblies were generated with the wgs-8.2 version of the PBcR assembler[49] with the following parameters: genomeSize = 30000000, assembleCoverage = 50. Assemblies were polished with quiver algorithm of smrtanalysis suite 2.3.0 (https://github.com/PacificBiosciences/

GenomicConsensus). A summary of raw data and assembly statistics for mating-type chromosomes is reported in Supplementary Table 1.

The protein-coding gene models were predicted with EuGene[50], trained for *Microbotryum*. Similarities to the fungal subset of the uniprot database plus the *M. lychnidis-dioicae* Lamole proteome[19] were integrated into EuGene for the prediction of gene models.

Mating-type chromosomes were identified by: (1) identifying the contigs carrying the PR and HD mating-type genes, (2) blasting the a₁ against the a₂ haploid genomes and visualizing the output using Circos[51] to identify contigs lacking collinearity, (3) blasting the haploid genomes against the completely assembled mating-type chromosomes of *M. lychnidis-dioicae*[19] and *M. lagerheimii*[16], (4) blasting the identified a₁ contigs to the whole a₂ haploid genome, and vice-versa, to detect which additional alternative mating-type contigs were linked to the previously identified mating-type contigs, and (5) re-doing steps (3) and (4) until no additional contig was identified. These contigs were then orientated in comparison to each other by: (1) using the putative centromere-specific repeats, as initial assemblies often yielded chromosome arms broken at the putative centromeres with identifiable putative centromere-specific repeats on each separated contig (e.g., Supplementary Figs. 5–8), and (2) blasting the a₁ and a₂ mating-type contigs against each other for identifying the PAR as the collinear regions that were then assigned to the edges of the chromosomes. The center contigs without centromeric repeats at any of their edges could not be oriented and were plotted in an arbitrary orientation.

Orthologous groups, species tree, and d_S plots. To study the evolution of suppressed recombination in a phylogenetic context, we reconstructed the relationships between the nine *Microbotryum* species and a closest outgroup (*Rhodotorula babjevae*) for which genomes were available. The genomes of these species were either sequenced for this study or obtained from previous studies[16,19] (Fig. 1). A previously published genome of *M. intermedium* was used from a strain collected on the plant *Salviae pratensis*, while its usual hosts belong to Dipsacaceae; species identity has, however, been double-checked using ITS sequences. We compared the translated gene models of the *Microbotryum* species and the outgroup with blastp 2.2.30+. The output was used to obtain orthologous groups by Markov clustering as implemented in orthAgogue[52]. We aligned the protein sequences of 4229 fully conserved single-copy genes with muscle v3.8.31[53] and obtained the codon-based CDS alignments with TranslatorX[54]. We used RAxML 8.2.7[55] to obtain maximum likelihood gene trees for all fully conserved single-copy genes and a species tree with the concatenated alignment of 2,172,278 codons with no gaps (trimal -nogaps option) under the GTRGAMMA substitution model. We estimated the branch support values by bootstrapping the species tree based on the concatenated alignment and by estimating the relative internode and tree certainty scores based on the frequency of conflicting bipartitions for each branch in the species tree among the fully conserved single-copy genes[56].

For d_S plots, we identified alleles using orthologous groups with a single sequence in each haploid genome for a given species. We used MUSCLE[53] embedded in TranslatorX[54] to align the two alleles per gene per species. Synonymous divergence and its standard error were estimated with the yn00 program of the PAML package[57].

Figures and statistical tests. Supplementary Figs. 4–8 were prepared using Circos[51]. We analyzed gene order after removing TEs to identify larger blocks of synteny. We identified syntenic blocks by searching all one-to-one gene correspondences between pairs of haploid genomes based on the orthologous groups reconstruction (see above). Statistical tests (Student's *t*-test, analysis of variance, and Wilcoxon rank tests) were performed using JMP v7 (SAS Institute).

Date estimates for recombination suppression and genealogies. For dating HD/PR linkage, we used alignments including a₁- and a₂-associated alleles at 17 single-copy orthologous groups that were located between the PR and HD loci following chromosomal fusion in all species but *M. silenes-acaulis* and that had both alleles retained (red vertical arrows in Fig. 3). The divergence between alleles associated with the a₁ versus a₂ mating types in these genes corresponds to the date of their linkage to mating-type loci. Indeed, genes linked to mating-type loci maintain one allele associated with a₁ and another allele associated with a₂ over time and the differentiation between these two alleles increases with the time since linkage to mating type. We only used the genes that displayed trans-specific polymorphism between *M. lychnidis-dioicae*, *M. silenes-dioicae* and *M. violaceum sensu stricto*[16] to avoid biasing estimates to younger dates because of gene conversion. For *M. silenes-acaulis*, we used a restricted set of 13 genes ancestrally located between the *HD* locus and the putative centromere (blue vertical arrows in Fig. 3a) because most genes located in its non-recombining region were in recombining regions in other species (Fig. 2). Divergence times were estimated using BEAST v2.4.0[58], with XML inputs generated using BEAUTi, and the default parameters except for unlinked substitution (HKY+G with empirical frequencies for each codon position) and clock models, Yule process to model speciation, and 10,000,000 mcmc generations sampled every 1000. We used a single calibration prior at 0.42 MY for all runs, corresponding to the divergence between

M. lychnidis-dioicae and *M. silenes-dioicae*[35], with a normal distribution and a sigma of 0.05. In some of the 17 individual gene trees ancestrally located between the HD and PR loci, some basal nodes were different from those in the species trees (Supplementary Fig. 3). However, the incongruent nodes were weakly supported so that gene genealogies were actually not significantly different from the species tree ($P > 0.35$, AU test[59]). We therefore forced the tree resulting from the concatenation of these 17 genes to the species tree topology for the date estimate analysis in BEAST. Genealogies of these 17 genes were inferred for codon-based alignments of genes in the different strata using RAxML[55] version 8.2.7, assuming the GTRGAMMA model and rapid bootstrap (options: -f a and -# 100).

Identification of TEs. Repetitive DNA content was analyzed with RepeatMasker[60], using REPBASE v19.11[61]. We used the RepeatMasker output to compute the percentage of base pairs occupied by TEs across the different evolutionary strata and PARs. For these counts, putative centromeres were filtered out. For plotting d_S along chromosomes, repeats were removed. Further filtering of repeats was performed by blasting (tBLASTx), with removal of repeats matching to more than five locations in the genome.

Detection of centromeric repeats. We identified centromeric-specific repeats using a method specifically designed for this purpose[62], based on the observation that in most species studied to date putative centromeres contain the most abundant tandem repeats, are gene poor, and repeat rich. For identifying centromeric repeats, we used Tandem-Repeat Finder (TRF v. 4.07b[63]) on assembled Illumina reads of the *M. lagerheimii* strain as the one sequenced using the Pacific Bioscience technology. We performed the assemblies as follows: we randomly chose 500,000 Illumina reads that we assembled with PRICE v1.2[64] using a random set of 1,000,000 reads as seed file and using the following command line arguments: -fpp (or -mpp when using mate-pair reads) inputFile_R1 inputFile_R2 650 90 -picf 20000 seedFile 500 2 25 -nc 10 -mpi 85 -MPI 95 – tpi 85 -TPI 95 -logf logfile -o outputFile. PRICE works by rounds of assembly: in the first round, it maps randomly picked reads onto contigs (provided by the "seedFile), assembles the reads that did not mapped, and then extends the contig with the unmapped assembled sequences. For the second and following rounds, PRICE considers the extended contigs as the reference to restart the process of picking, mapping reads, assembling the unmapped reads, and extending the reference contigs. We analyzed the presence of tandem repeats in each of the 10 assembly cycle outputs using the following parameters in a TRF wrapper perl script[62]: match = 1, mismatch = 1, indel = 2, probability of match = 80, probability of indel = 5, min score = 200, max period = 2000. We performed these steps 15 times, picking randomly 500,000 input reads and 1,000,000 reads for the seed file. The repeats detected in the Illumina genomes were blasted against the corresponding high-quality genomes. We identified the putative centromeres in *M. lagerheimii* as the most gene-poor and repeat-rich regions and with the most abundant tandem repeats. The delimitations of the centromeric regions using this method yielded a single region per contig and were congruent with those using BLAST of the centromeric repeats identified previously[19] in *M. lychnidis-dioicae*. Putative centromeres were identified in the other species by blasting the identified centromeric-specific repeats (the tandem repeats identified here and previously[19] gave congruent results, see Supplementary Figure S4b).

Gene loss. Alleles were identified by applying orthomcl[65] to the protein data sets for unique a_1 and a_2 orthologs, discarding orthologous groups containing more than one protein-coding gene per mating type. The loss of a gene was inferred when a protein-coding gene in one mating type did not have any match in the orthomcl output in the opposite mating type within a diploid genome. We computed the number of gene losses across the PARs and the evolutionary strata that were not too rearranged to be delimited.

Polymorphism data and analyses. To rule out high levels of polymorphism as the cause for the observed high d_S values in the youngest strata of the *Microbotryum* mating-type chromosomes, we assessed the level of polymorphism and a_1–a_2 allelic segregation in gene genealogies. When genes are linked to mating-type loci, alleles associated with the a_1 versus a_2 alleles accumulate differences until completely segregating according to mating-type allele rather than according to strain in gene genealogies. To test for this pattern, we re-sequenced multiple strains of *M. v. paradoxa* and of *M. v. caroliniana* (4 strains for *M. v. paradoxa* and 11 strains for *M. v. caroliniana*, Supplementary Table 3, strains collected before 2014 and thus not falling under the Nagoya protocol) from a_1 and a_2 haploid sporidia isolated from single tetrads using micromanipulation. For *M. v. caroliniana*, we also used spores collected from *S. virginica*, as this plant species is parasitized by the same anther-smut species as the one parasitizing *S. caroliniana*. Haploid sporidia were cultured on potato dextro agar and DNA was extracted using the Nucleospin Soil Kit (Macherey-Nagel, Germany). Haploid genomes of identified mating type were sequenced (Illumina paired-end sequencing with 46× mean coverage).

After trimming and filtering for quality (length >50; quality base >10) using cutadapt[66], reads were mapped against the high-quality PacBio reference genome of the same mating type and species. We used bowtie2[67] in the "very-sensitive-local" mode with the default parameters. Mapped reads were filtered for PCR

duplicates using picard-tools (http://broadinstitute.github.io/picard) and realigned on the PacBio reference genome using GATK[68]. Single-nucleotide polymorphisms (SNPs) were called with GATK HaplotypeCaller, which provide a gVCF per strain. For each strain, we filtered on a quality above 100 and other parameters (QD, FS, MQ, MQRankSum and ReadPosRankSum) for which the threshold was the fifth percentile (95th for FS parameter). The subsequent SNPs having >90% of missing data among strains were excluded from the dataset.

We generated pseudo-sequences per species and mating type by substituting reference nucleotides by their variants in the reference sequence, using the predicted CDS of the PacBio reference genome and the VCF file produced by GATK GenotypeVCF that combine gVCF into one file. We then computed the θπ statistic of diversity with EggLib version 2[69] for each species and mating type.

We generated alignments of a_1 and a_2 alleles per species and per predicted coding sequence and with a_1 and a_2 alleles from the predicted orthologous genes of the *M. lagerheimii* reference genome. Codon-based alignments were performed for each single-copy gene in the mating-type chromosome, for the various evolutionary strata, and the pseudo-autosomal regions, as well as for a well-assembled autosome, using TranslatorX[54]. Trees were computed with the RAxML rapid bootstrap mode (-f a -m GTRGAMMA -# 100) and plotted and rooted on the *M. lagerheimii* a_1 strain using the R ape package[70].

Clustering of mating-type-associated alleles in genealogies. For genes linked to the mating type loci, a_1- and a_2-associated alleles will differentiate to the point of forming two distinct clades in gene genealogies. To assess the level of clustering of alleles retrieved from a_1 and a_2 genomes in the gene genealogies, we designed and computed the following index: for each tree, we successively sampled the closest pairs of a_1- and a_2-associated alleles—in terms of nodal distance—until no pair was left (leaving out singletons in cases where the numbers of a_1 and a_2 alleles were not identical). This was performed 10 times in order to remove a possible effect of the order in which the pairs were sampled. We then computed the mean of these minimum values and compared it to a null distribution of the same index obtained by randomly permuting 1000 times the leaves of the input tree. The index was defined as the proportion of the random permutations giving a mean minimum value smaller or equal to the observed one. Index values close to 1 mean a_1 and a_2 alleles were completely separated in the tree (i.e., permutations always brought a_1- and a_2-associated alleles closer than they actually were). Conversely, an index value close to 0 meant that the a_1- and a_2-associated alleles were all forming pairs ("cherries") in the tree (i.e., random permutations always increased their distances). Nodes with low bootstrap support were collapsed prior to the analysis. A custom script for the computation of this index was written in R using the ape package for tree manipulations[70] and is available in Supplementary Note 1.

References

1. Gould, S. *Wonderful Life - The Burgess Shale and the Nature of History* (Norton, New York, 1989).
2. Wake, D. B., Wake, M. H. & Specht, C. D. Homoplasy: from detecting pattern to determining process and mechanism of evolution. *Science* **331**, 1032–1035 (2011).
3. Stern, D. L. & Orgogozo, V. Is genetic evolution predictable? *Science* **323**, 746–751 (2009).
4. Elmer, K. R. & Meyer, A. Adaptation in the age of ecological genomics: insights from parallelism and convergence. *Trends Ecol. Evol.* **26**, 298–306 (2011).

5. Elmer, K. R. et al. Parallel evolution of Nicaraguan crater lake cichlid fishes via non-parallel routes. *Nat. Commun.* **5**, 5168 (2014).

6. Duboue, E. R., Keene, A. C. & Borowsky, R. L. Evolutionary convergence on sleep loss in cavefish populations. *Curr. Biol.* **21**, 671–676 (2011).

7. Dobler, S., Dalla, S., Wagschal, V. & Agrawal, A. A. Community-wide convergent evolution in insect adaptation to toxic cardenolides by substitutions in the Na,K-ATPase. *Proc. Natl. Acad. Sci. USA* **109**, 13040–13045 (2012).

8. Tishkoff, S. A. et al. Convergent adaptation of human lactase persistence in Africa and Europe. *Nat. Genet.* **39**, 31–40 (2007).

9. Pennisi, E. 'Supergenes' drive evolution. *Science* **357**, 1083 (2017).

10. Purcell, J., Brelsford, A., Wurm, Y., Perrin, N. & Chapuisat, M. Convergent genetic architecture underlies social organization in ants. *Curr. Biol.* **24**, 2728–2732 (2014).

11. Wang, J. et al. A Y-like social chromosome causes alternative colony organization in fire ants. *Nature* **493**, 664–668 (2013).

12. Joron, M. et al. Chromosomal rearrangements maintain a polymorphic supergene controlling butterfly mimicry. *Nature* **477**, 203–206 (2011).

13. Charlesworth, D. The status of supergenes in the 21st century: recombination suppression in Batesian mimicry and sex chromosomes and other complex adaptations. *Evol. Appl.* **9**, 74–90 (2016).

14. Wright, A., Dean, R., Zimmer, F. & Mank, J. How to make a sex chromosome. *Nat. Commun.* **7**, 12087 (2016).

15. Ahmed, S. et al. A haploid system of sex determination in the brown alga *Ectocarpus* sp. *Curr. Biol.* **24**, 1945–1957 (2014).

16. Branco, S. et al. Evolutionary strata on young mating-type chromosomes despite lack of sexual antagonism. *Proc. Natl. Acad. Sci. USA* **114**, 7067–7072 (2017).

17. Fraser, J. A. et al. Convergent evolution of chromosomal sex-determining regions in the animal and fungal kingdoms. *PLoS Biol.* **2**, 2243–2255 (2004).

18. Menkis, A., Jacobson, D. J., Gustafsson, T. & Johannesson, H. The mating-type chromosome in the filamentous ascomycete *Neurospora tetrasperma* represents a model for early evolution of sex chromosomes. *PLoS Genet.* **4**, e1000030 (2008).

19. Badouin, H. et al. Chaos of rearrangements in the mating-type chromosomes of the anther-smut fungus *Microbotryum lychnidis-dioicae*. *Genetics* **200**, 1275–1284 (2015).

20. Sun, S. et al. Fungal genome and mating system transitions facilitated by chromosomal translocations involving intercentromeric recombination. *PLoS Biol.* **15**, e2002527 (2017).

21. Grognet, P. et al. Maintaining two mating types: structure of the mating type locus and its role in heterokaryosis in *Podospora anserina*. *Genetics* **197**, 421–432 (2014).

22. Goubet, P. M. et al. Contrasted patterns of molecular evolution in dominant and recessive self-incompatibility haplotypes in. *Arab. PLoS Genet.* **8**, e1002495 (2012).

23. Bachtrog, D. Y-chromosome evolution: emerging insights into processes of Y-chromosome degeneration. *Nat. Rev. Genet.* **14**, 113–124 (2013).

24. Ironside, J. E. No amicable divorce? Challenging the notion that sexual antagonism drives sex chromosome evolution. *Bioessays* **32**, 718–726 (2010).

25. Charlesworth, D. Plant contributions to our understanding of sex chromosome evolution. *New Phytol.* **208**, 52–65 (2015).

26. Muyle, A. et al. Rapid *de novo* evolution of X chromosome dosage compensation in *Silene latifolia*, a plant with young sex chromosomes. *PLoS Biol.* **10**, e1001308 (2012).

27. Bachtrog, D. A dynamic view of sex chromosome evolution. *Curr. Opin. Genet. Dev.* **16**, 578–585 (2006).

28. Coelho, M. A., Bakkeren, G., Sun, S., Hood, M. E. & Giraud, T. in *The Fungal Kingdom, Section 2: Life of Fungi* (eds Gow, N. & Heitman, J.) (ASM Press, Washington, 2017).

29. Bakkeren, G. & Kronstad, J. W. Linkage of mating type loci distinguishes bipolar from tetrapolar mating in basidiomycetous smut fungi. *Proc. Natl. Acad. Sci. USA* **91**, 7085–7089 (1994).

30. James, T. Y. Why mushrooms have evolved to be so promiscuous: Insights from evolutionary and ecological patterns. *Fung. Biol. Rev.* **29**, 167–178 (2015).

31. Nieuwenhuis, B. P. S. et al. Evolution of uni- and bifactorial sexual compatibility systems in fungi. *Heredity* **111**, 445–455 (2013).

32. Hood, M. E. Dimorphic mating-type chromosomes in the fungus *Microbotryum violaceum*. *Genetics* **160**, 457–461 (2002).

33. Bergero, R., Forrest, A., Kamau, E. & Charlesworth, D. Evolutionary strata on the X chromosomes of the dioecious plant *Silene latifolia*: evidence from new sex-linked genes. *Genetics* **175**, 1945–1954 (2007).

34. Hood, M. E., Scott, M. & Hwang, M. Breaking linkage between mating compatibility factors: tetrapolarity in *Microbotryum*. *Evolution* **69**, 2561–2572 (2015).

35. Gladieux, P. et al. Maintenance of fungal pathogen species that are specialized to different hosts: allopatric divergence and introgression through secondary contact. *Mol. Biol. Evol.* **28**, 459–471 (2011).

36. Fontanillas, E. et al. Degeneration of the non-recombining regions in the mating type chromosomes of the anther smut fungi. *Mol. Biol. Evol.* **32**, 928–943 (2015).

37. Sun, S., Hsueh, Y.-P. & Heitman, J. Gene conversion occurs within the mating-type locus of *Cryptococcus neoformans* during sexual reproduction. *PLoS Genet.* **8**, e1002810 (2012).

38. Kemler, M., Goker, M., Oberwinkler, F. & Begerow, D. Implications of molecular characters for the phylogeny of the Microbotryaceae (Basidiomycota: Urediniomycetes). *BMC Evol. Biol.* **6**, 35 (2006).

39. Hood, M. E. & Antonovics, J. Intratetrad mating, heterozygosity, and the maintenance of deleterious alleles in *Microbotryum violaceum* (=*Ustilago violacea*). *Heredity* **85**, 231–241 (2000).

40. Bueker, B. et al. Distribution and population structure of the anther smut *Microbotryum silenes-acaulis* parasitizing an arctic-alpine plant. *Mol. Ecol.* **25**, 811–824 (2016).

41. Kirkpatrick, M. & Barton, N. Chromosome inversions, local adaptation and speciation. *Genetics* **173**, 419–434 (2006).

42. Antonovics, J. & Abrams, J. Y. Intratetrad mating and the evolution of linkage relationships. *Evolution* **58**, 702–709 (2004).

43. Horns, F., Petit, E. & Hood, M. E. Massive expansion of *Gypsy*-like retrotransposons in *Microbotryum* fungi. *Genome Biol. Evol.* **9**, 363–371 (2017).

44. Bergero, R. & Charlesworth, D. Preservation of the Y transcriptome in a 10-million-year-old plant sex chromosome system. *Curr. Biol.* **21**, 1470–1474 (2011).

45. Bull, J. J. Sex chromosomes in haploid dioecy - unique contrast to Mullers theory for diploid dioecy. *Am. Nat.* **112**, 245–250 (1978).

46. Hood, M. E., Antonovics, J. & Koskella, B. Shared forces of sex chromosome evolution in haploids and diploids. *Genetics* **168**, 141–146 (2004).

47. Thomas, A., Shykoff, J., Jonot, O. & Giraud, T. Sex-ratio bias in populations of the phytopathogenic fungus *Microbotryum violaceum* from several host species. *Int. J. Plant. Sci.* **164**, 641–647 (2003).

48. Otto, S. P. et al. About PAR: The distinct evolutionary dynamics of the pseudoautosomal region. *Trends Genet.* **27**, 358–367 (2011).

49. Koren, S. et al. Hybrid error correction and de novo assembly of single-molecule sequencing reads. *Nat. Biotechnol.* **30**, 693–700 (2012).

50. Foissac, S. et al. Genome annotation in plants and fungi: EuGène as a model platform. *Curr. Bioinform.* **3**, 87–97 (2008).

51. Krzywinski, M. I. et al. Circos: an information aesthetic for comparative genomics. *Genome Res.* **19**, 1639–1645 (2009).

52. Ekseth, O. K., Kuiper, M. & Mironov, V. orthAgogue: an agile tool for the rapid prediction of orthology relations. *Bioinformatics* **30**, 734–736 (2014).

53. Edgar, R. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).

54. Abascal, F., Zardoya, R. & Telford, M. TranslatorX: multiple alignment of nucleotide sequences guided by amino acid translations. *Nucleic Acids Res.* **38**, W7–W13 (2010).

55. Stamatakis, A. RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22**, 2688–2690 (2006).

56. Salichos, L., Stamatakis, A. & Rokas, A. Novel information theory-based measures for quantifying incongruence among phylogenetic trees. *Mol. Biol. Evol.* **31**, 1261–1271 (2014).

57. Yang, Z. PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24**, 1586–1591 (2007).

58. Bouckaert, R. et al. Beast 2: a software platform for Bayesian evolutionary analysis. *PLoS Comput. Biol.* **10**, e1003537 (2014).

59. Shimodaira, H. An approximately unbiased test of phylogenetic tree selection. *Syst. Biol.* **51**, 492–508 (2002).

60. Smit, A. & Green, P. RepeatMasker at http://www.repeatmasker.org/.

61. Jurka, J. Repeats in genomic DNA: mining and meaning. *Curr. Opin. Struct. Biol.* **8**, 333–337 (1998).

62. Melters, D. P. et al. Comparative analysis of tandem repeats from hundreds of species reveals unique insights into centromere evolution. *Genome Biol.* **14**, R10 (2013).

63. Benson, G. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res.* **27**, 573–580 (1999).

64. Ruby, J. G., Bellare, P. & DeRisi, J. L. PRICE: software for the targeted assembly of components of (Meta) genomic sequence data. *G3 (Bethesda)* **3**, 865–880 (2013).

65. Li, L., Stoeckert, C. J. & Roos, D. OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res.* **13**, 2178–2189 (2003).

66. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J.* **17**, 10–12 (2011).

67. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).

68. McKenna, A. et al. The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).

69. De Mita, S. & Siol, M. EggLib: processing, analysis and simulation tools for population genetics and genomics. *BMC Genet.* **13**, 27 (2012).

70. Paradis, E., Claude, J. & Strimmer, K. APE: analyses of phylogenetics and evolution in R language. *Bioinformatics* **20**, 289–290 (2004).

Acknowledgements

Author contributions

T.G. and M.E.H. designed and supervised the study. T.G., M.E.H., D.B., and S.B. contributed to obtain funding. D.B. provided strains. H.B., M.E.H., S.L., A.S., and T.G. obtained the genomes. S.B., F.C., R.C.R.d.l.V., H.B., F.H., and M.A.C. performed the genomic analyses. D.M.d.V. performed analyses on gene genealogy allele clustering. T.G., S.B., and M.E.H. wrote the manuscript with contributions from all other authors.

Additional information

Supplementary Information accompanies this paper at https://doi.org/10.1038/s41467-018-04380-9.

Competing interests: The authors declare no competing interests.

Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

5 Convergent recombination cessation between mating-type genes and centromeres

Published in *Genome Research* (May 1, 2019)

In the previous chapter, we demonstrated that the strong selection to increase the odd of gamete compatibility under intra-tetrad mating repeatedly selected for bipolarity. However, previous studies estimated a very high level of selfing also in the tetrapolar sister-species *M. lagerheimii* and *M. saponariae*. Why these two species retained tetrapolarity was therefore puzzling, and it weakened our conclusions on the predictability of evolution. In *M. saponariae*, it had been showed that each of the PR and HD loci segregated with the centromere of their respective chromosome. Under intra-tetrad mating, this genomic conformation confers the same odds of gamete compatibility as bipolarity.

In the present study, we therefore investigated whether the PR and HD genes were also linked to their respective centromere in *M. lagerheimii* through segregation analysis. We also looked for additional evidence for recombination suppression between the mating-type genes and their centromere in both *M. lagerheimii* and *M. saponariae*, in the form of elevated synonymous divergence and genomic rearrangements between compatible gametes. We also investigated whether the recombination suppression event linking each mating-type gene to its respective centromere arose in the common ancestor of these two sister species, considering the previously reported remarkable frequency of convergence in terms of recombination suppression events.

I conducted this study as the main author. I performed the (i) synteny analysis and (ii) the d_S calculation, (iii) generated raw plots that were then adorned by Marco Coehlo, (iv) redefined centromeric repeats in *M. lagerheimii* and *M. saponariae* using a method based on Illumina reads which identifies high-copy tandem repeats (Melters et al. 2013), (v) wrote the manuscript with the help of Tatiana Giraud and Michael E. Hood, and (vi) was the corresponding author for the journal.

Research

Convergent recombination cessation between mating-type genes and centromeres in selfing anther-smut fungi

Fantin Carpentier,[1,4] Ricardo C. Rodríguez de la Vega,[1,4] Sara Branco,[1,6] Alodie Snirc,[1] Marco A. Coelho,[2,7] Michael E. Hood,[3,5] and Tatiana Giraud[1,5]

[1]Ecologie Systématique Evolution, Bâtiment 360, Univ. Paris-Sud, AgroParisTech, CNRS, Université Paris-Saclay, 91400 Orsay, France; [2]UCIBIO-REQUIMTE, Departamento de Ciências da Vida, Faculdade de Ciências e Tecnologia, Universidade NOVA de Lisboa, 2829-516 Caparica, Portugal; [3]Department of Biology, Amherst College, Amherst, Massachusetts 01002, USA

The degree of selfing has major impacts on adaptability and is often controlled by molecular mechanisms determining mating compatibility. Changes in compatibility systems are therefore important evolutionary events, but their underlying genomic mechanisms are often poorly understood. Fungi display frequent shifts in compatibility systems, and their small genomes facilitate elucidation of the mechanisms involved. In particular, linkage between the pre- and postmating compatibility loci has evolved repeatedly, increasing the odds of gamete compatibility under selfing. Here, we studied the mating-type chromosomes of two anther-smut fungi with unlinked mating-type loci despite a self-fertilization mating system. Segregation analyses and comparisons of high-quality genome assemblies revealed that these two species displayed linkage between mating-type loci and their respective centromeres. This arrangement renders the same improved odds of gamete compatibility as direct linkage of the two mating-type loci under the automictic mating (intratetrad selfing) of anther-smut fungi. Recombination cessation was found associated with a large inversion in only one of the four linkage events. The lack of *trans*-specific polymorphism at genes located in nonrecombining regions and linkage date estimates indicated that the events of recombination cessation occurred independently in the two sister species. Our study shows that natural selection can repeatedly lead to similar genomic patterns and phenotypes, and that different evolutionary paths can lead to distinct yet equally beneficial responses to selection. Our study further highlights that automixis and gene linkage to centromeres have important genetic and evolutionary consequences, while being poorly recognized despite being present in a broad range of taxa.

[Supplemental material is available for this article.]

Mating systems reflect the degree of selfing/outcrossing in natural populations and impact gene flow, the accumulation of deleterious alleles, and adaptability (Lande and Schemske 1985; Charlesworth and Charlesworth 1987; Charlesworth et al. 1990; Charlesworth 2002; Igic et al. 2008; Hereford 2010; Lande 2015). Outcrossing can promote gene flow and therefore the rapid spread of beneficial alleles as well as the purge of deleterious alleles, whereas selfing is often associated with reproductive assurance and can help maintain favorable combinations of alleles at different loci. There is a wide diversity of mating systems in nature that strongly impacts the evolution of organisms. Automixis with mating among products of a given meiosis that separated in the first meiosis division ("fusion of nonsister second division products") (Lewis and John 1963), for example, is a little-known form of self-fertilization (Mogie 1986); such automixis is often called central fusion in animals, the term "fusion" referring to the union of gametes and the term "central" referring to the placement of the fusing gametes in an ordered tetrad (Suomalainen 1950; Goudie

and Oldroyd 2014). This kind of automixis maintains heterozygosity at all loci for which there has been no recombination with the centromere (Hood and Antonovics 2000, 2004; Zakharov 2005; Lenormand et al. 2016; Engelstädter 2017). This effect can extend over large portions of the genome when there are low levels of crossing-over (Hood and Antonovics 2000, 2004). Automixis with central fusion can thus maintain long-term heterozygosity, which can lead to sheltering deleterious alleles or may be beneficial in cases of advantageous overdominance (i.e., heterozygote advantage) (Engelstädter 2017). Automixis and its genetic and evolutionary consequences are poorly studied despite being relatively frequent (Mogie 1986) across a variety of taxa such as in fungi (Hood and Antonovics 2000; Zakharov 2005; Menkis et al. 2008; Grognet et al. 2014), plants (Asker 1980; Walker 1985; Antonius and Nybom 1995; Cruden and Lloyd 1995; Schön et al. 2009), reptiles (Watts et al. 2006; Booth et al. 2011; Booth and Schuett 2015), fishes (Chapman et al. 2007; Dudgeon et al. 2017; Feldheim et al. 2017), birds (Schut et al. 2008), crustaceans (Nougué et al. 2015), nematodes (Van der Beek et al. 1998), and insects (Suomalainen et al. 1976; Normark 2003; Oldroyd et al. 2008).

[4]These authors contributed equally to this work.
[5]These authors supervised the study.
Present addresses: [6]Department of Microbiology and Immunology, Montana State University, Bozeman, MT 59717, USA; [7]Department of Molecular Genetics and Microbiology, Duke University Medical Center, Durham, NC 27710, USA
Corresponding author: fantin.carpentier@gmail.com

Evolutionary transitions between mating systems are known to be relatively frequent (Goldberg et al. 2010; Goldberg and Igić 2012; Chantha et al. 2013; Nieuwenhuis et al. 2013; Hanschen et al. 2018). Changes in the genetic determination of gamete production or compatibility often underlie transitions in mating systems, such as the evolution of a self-incompatibility system. For example, in many species mating can only occur between males and females, which enforces outcrossing, and sexes are often determined by sex chromosomes (Beukeboom and Perrin 2014). In angiosperms, mating can also be restricted by a self-incompatibility locus, which promotes outcrossing in hermaphroditic species by preventing mating between genotypes carrying identical alleles (Vekemans et al. 2014). In most fungi, gamete compatibility is controlled at the haploid stage, and only cells carrying different alleles at the mating-type loci can successfully mate (Billiard et al. 2011, 2012).

Fungi provide excellent eukaryotic models for studying the genomic changes involved in gamete compatibility transition, because they display highly diverse and labile mate-recognition systems (Billiard et al. 2011, 2012; Nieuwenhuis et al. 2013) as well as relatively small and compact genomes that allow for high-quality genome assemblies (Gladieux et al. 2014; Badouin et al. 2015; Faino et al. 2015; Sonnenberg et al. 2016; Branco et al. 2017, 2018; Sun et al. 2017a,b). In basidiomycete fungi (e.g., rusts, smuts, and mushrooms), mating type is most often controlled by two loci: (1) the PR locus, determining gamete fusion compatibility with a pheromone receptor and neighboring pheromone genes; and (2) the HD locus, determining compatibility for postmating development with two homeodomain genes (Coelho et al. 2017). To successfully mate and produce offspring, two gametes must carry different alleles at both loci. In most basidiomycetes, the PR and HD loci segregate independently (Raper 1966; Nieuwenhuis et al. 2013). Multiple independent events of linkage of the two mating-type loci have been documented in several fungal species (Bakkeren and Kronstad 1994; Nieuwenhuis et al. 2013; Branco et al. 2017; Sun et al. 2017b). Such control of gamete compatibility inherited as a single locus is advantageous under selfing because it increases the odds of gamete compatibility among the gametes of a given diploid individual (Fig. 1; Coelho et al. 2017).

The plant-castrating anther-smut fungi belonging to the highly selfing basidiomycete genus *Microbotryum* are particularly good systems for studying the genomic changes underlying shifts in gamete compatibility systems. Before the radiation of this genus, recombination suppression extended around each of the PR and HD loci (Branco et al. 2017). Several *Microbotryum* species underwent independent transitions to complete linkage between the mating-type loci through various chromosomal rearrangements that brought the HD and PR loci onto the same chromosome (Branco et al. 2017, 2018). In some of these species, the cessation of recombination subsequently expanded far beyond the mating-type loci in several successive steps to include the majority of the mating-type chromosomes (Branco et al. 2017, 2018). Recombining pseudoautosomal regions (PARs) remained at both edges of the mating-type chromosomes in many lineages (Branco et al. 2018).

The majority of studied anther-smut fungi undergo selfing by automixis (Hood and Antonovics 2000; Giraud et al. 2008; Vercken et al. 2010; Gladieux et al. 2011; Bueker et al. 2016) and have linked mating-type loci (Branco et al. 2018). Here, we studied two closely related species, *Microbotryum lagerheimii* and *Microbotryum saponariae,* that have retained unlinked PR and HD mating-type loci located on different chromosomes (Fig. 1; Hood et al. 2015), despite

selfing mating systems (Fortuna et al. 2016, 2018; Abbate et al. 2018). However, *M. saponariae* displays the PR and HD mating-type loci completely linked to the centromere of their respective chromosomes, which induces central fusion automixis and ensures the same odds of compatibility under selfing by automixis as would linkage between the mating-type loci (Fig. 1; Hood et al. 2015). Although the mating-type loci are also known to be located on different chromosomes in *M. lagerheimii* (Branco et al. 2017), it is unclear whether the HD and PR loci are linked to the centromeres. In case they are linked to centromeres and given that *M. lagerheimii* and *M. saponariae* are sister species in available phylogenies (Fig. 2), recombination cessation with the centromeres could potentially predate their speciation event. An alternative hypothesis would be independent linkage events, with convergence for complete centromere linkage occurring in the two species after their divergence. In this study, we used segregation analyses and high-quality genome assemblies to investigate (1) whether HD and PR loci are linked to the centromeres in *M. lagerheimii*, (2) whether linkage predates speciation between *M. lagerheimii* and *M. saponariae* or constitutes independent events, and (3) whether the PR and HD loci became linked to centromeres at similar dates in each species.

Results

Linkage of mating-type loci to centromeres in *M. saponariae* and *M. lagerheimii*

To test whether recombination was suppressed between the mating-type loci and their respective centromere in *M. lagerheimii*, we analyzed PR and HD mating-type loci segregation within ordered linear tetrads using allele-specific PCR markers for each mating-type locus. When there is complete centromere linkage, alleles at both mating types always segregate at the first meiotic division, leading to the ordered linear *Microbotryum* tetrad with cells derived from opposite poles of meiosis I carrying alternate alleles at both the PR and HD loci (Fig. 1; Hood and Antonovics 2000; Hood et al. 2015). Conversely, when there is no centromere linkage of the mating-type loci, mating-type alleles segregating at the second meiotic division results in only half the tetrads carrying alternate alleles at both loci in the opposite cells of the ordered linear tetrad (Fig. 1). We found evidence supporting centromere linkage of mating-type loci in *M. lagerheimii*, with isolated meiotic products derived from opposite poles of meiosis I showing alternate alleles at both the PR and HD loci in all of the 78 meioses analyzed. Given the number of tetrads analyzed, the 95% confidence interval for the occurrence of recombination between at least one mating-type locus and its centromere was 0%–5%. This indicates that the *M. lagerheimii* PR and HD loci are completely or nearly completely linked to their respective centromere, as in its sister species *M. saponariae* (Hood et al. 2015).

We compared the sequences of the mating-type chromosomes to investigate whether inversions could have contributed to linkage between mating-type loci and their centromeres. While the alternate HD mating-type chromosomes were collinear within both *M. saponariae* and *M. lagerheimii* (Fig. 3A,C), we observed a 51.4 kbp inversion between HD and the centromere in the *M. saponariae* lineage compared to the ancestral state, shared by *Microbotryum intermedium* and *M. lagerheimii* (Supplemental Figs. S2A, S3A). In both *M. saponariae* and *M. lagerheimii*, the HD locus was located close to the centromere (distant by 138 kbp in *M. saponariae* and by 162 kbp in *M. lagerheimii*) (Fig. 3A,C). In *M. lagerheimii* the alternate PR chromosomes (a_1 and a_2) also showed
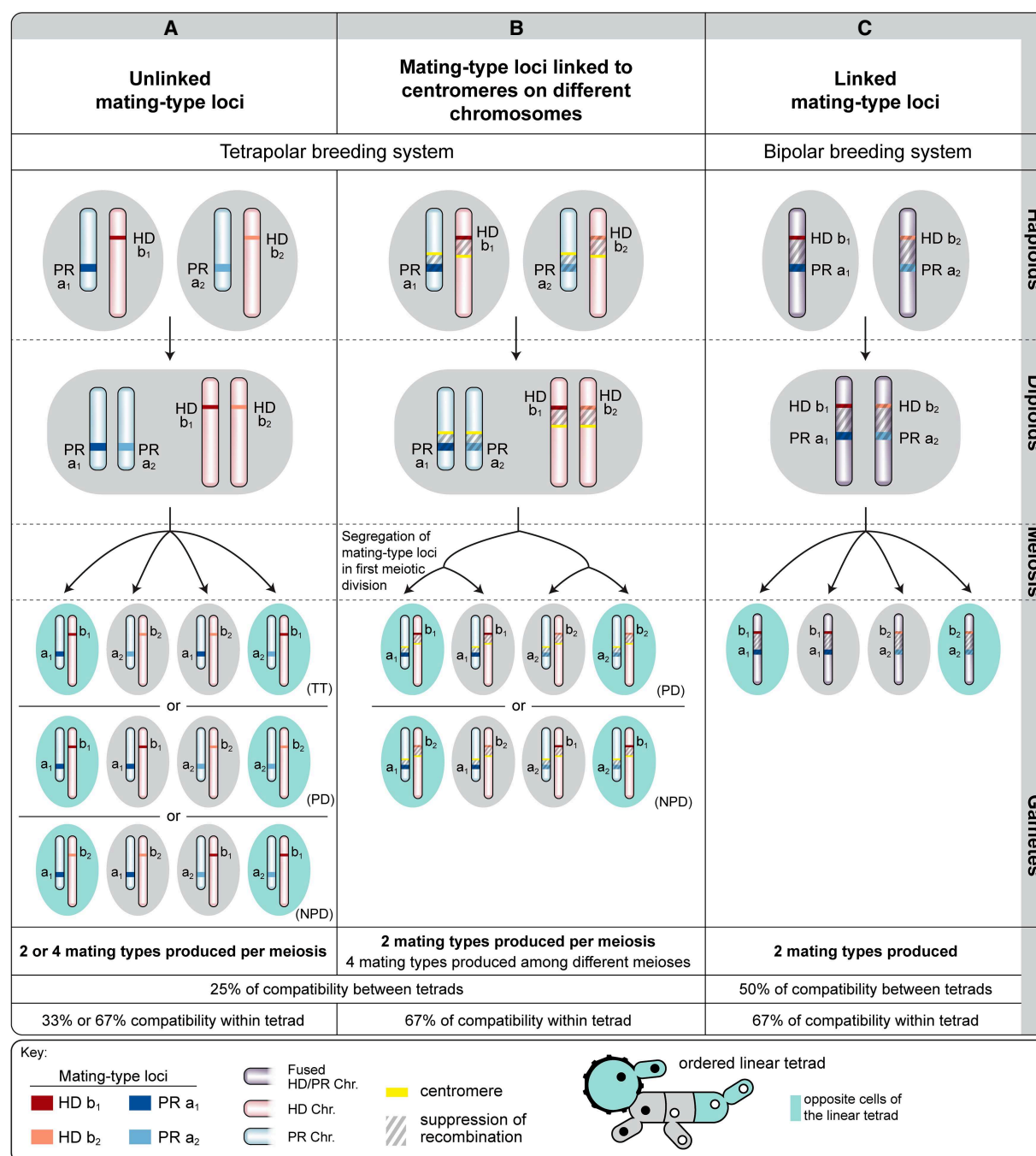
Figure 1. Odds of compatibility among gametes of a diploid individual in basidiomycete fungi. Gametes are fully compatible only if they carry different alleles at both mating-type loci, the PR (including pheromone receptor and pheromone genes, with a_1 and a_2 alleles) and HD (including homeodomain genes, with b_1 and b_2 alleles) loci. (A) With PR and HD mating-type loci unlinked from each other and from the centromeres (shown here located in different chromosomes in blue and red), the percentage of compatibility of a given gamete among the other gametes produced by the same diploid individual is 25% across multiple meioses (a given gamete is compatible with one of every four gametes), and the percentage is 33% within tetrad (a given gamete is compatible with one of the other three gametes in the tetrad) or 67% (a given gamete is compatible with two of the three remaining gametes in the tetrad), depending on segregation of the mating-type alleles. The different types of gametes produced are tetratypes (TT), parental ditypes (PD), or nonparental ditypes (NPD), which depend on allele segregation and on whether a crossing-over occurred between one of the two loci and the centromere. (B) With PR and HD mating-type genes linked to the centromeres of different chromosomes (blue and red), the percentage of compatibility of a given gamete among the other gametes produced by the same diploid individual is 25% across multiple meioses but 67% within a tetrad (a given gamete is compatible with two of the three other gametes in the tetrad) due to the segregation of the variation occurring only at meiosis I for both mating-type loci. The different types of gametes produced are parental ditypes (PD) or nonparental ditypes (NPD), which depend on segregation. (C) With HD and PR loci fully linked to each other on the same chromosome, the percentage of compatibility of a given gamete among the other gametes produced by the same diploid individual is 50% across multiple meioses (a given gamete is compatible with one of every two gametes), and 67% within a single meiotic tetrad (a given gamete is compatible with two of the three other gametes in the tetrad). The light blue background shows the opposite cells of ordered tetrads, both in a *Microbotryum* linear tetrad representation (*bottom*) and in the different types of possible tetrads depicted depending on mating-type locus linkage.
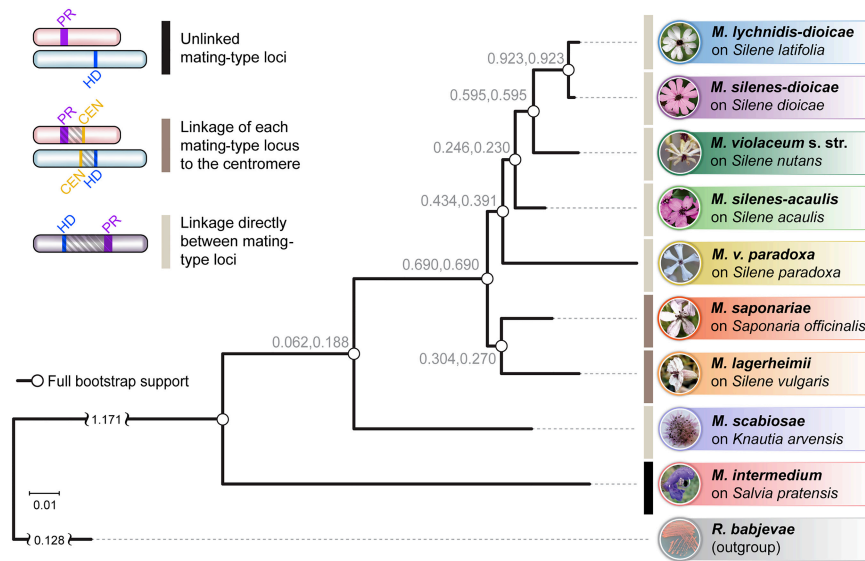
Figure 2. Phylogenies of anther-smut fungi and their mating-type loci linkage. *Microbotryomycete* phylogenetic tree based on 780 orthologous genes, including the studied *Microbotryum* species (shown in the anthers of their host plants) and the outgroup *Rhodotorula babjevae*. The empty circles indicate full bootstrap support. Tree internode certainty with no conflict bipartitions (the normalized frequency of the most frequent bipartition across gene genealogies relative to the summed frequencies of the two most frequent bipartitions) is given *above* the branches, indicating good support for the bipartitions. Black bars at *right* indicate unlinked mating-type loci, dark gray linkage of mating-type loci to centromeres, and light gray mating-type loci linkage.

may blur historical steps (Branco et al. 2017). We used the mating-type chromosomes of an outgroup species with independently segregating mating types (*M. intermedium*) as a proxy for the ancestral gene order; however, using the *M. lagerheimii* gene order gave similar conclusions given the few rearrangements observed (Supplemental Figs. S2, S3; Branco et al. 2017). We observed very high levels of synonymous divergence around the PR and HD mating-type loci both in *M. lagerheimii* and *M. saponariae* (purple and blue genomic regions, respectively) (Fig. 4). This result was expected given ancient recombination cessation proximal to each of the PR and HD mating-type loci (Branco et al. 2017). The nonzero d_S values between the PR-proximal region and the centromere in *M. saponariae* and *M. lagerheimii* supported complete linkage to the centromere. In highly selfing organisms such as anther-smut fungi, homozygosity is high at almost all genes (i.e., $d_S = 0$ between alleles on autosomes in a diploid individual) (Supplemental Fig. S4) except in regions linked to mating types (Branco et al. 2018). The lower d_S values

nearly complete collinearity (Fig. 3B), with only rearrangements around the PR locus, as typical in *Microbotryum* due to very old recombination suppression in this region (Branco et al. 2017). In contrast, *M. saponariae* displayed a large pericentric inversion distinguishing the alternate PR chromosomes. This inversion involved 593 kbp in the a_1 and 701 kbp in the a_2 PR mating-type chromosomes, representing 50% and 52% of a_1 and a_2 PR chromosome lengths (Fig. 3D). The two edges of the inversion were very close to the centromere and the PR locus (inversion boundaries were distant by 35 kbp from centromeres and by 0 kbp from the edge of the PR-proximal region with ancient recombination suppression) (Fig. 3D). The inversion appeared derived in the a_2 *M. saponariae* PR chromosome, as the *M. saponariae* a_1 PR chromosome was highly collinear to the *M. lagerheimii* a_1 and a_2 PR chromosomes (Supplemental Fig. S1). No further rearrangements were present within the large inversion beyond those located in the PR-proximal region (Fig. 3D). A small additional inversion was observed toward the PAR in the short arm of the *M. saponariae* a_2 PR mating-type chromosome (green region in Fig. 3D, involving 11 kbp and 20 kbp on the a_1 and a_2 PR mating-type chromosomes). This inversion may correspond to an additional step extending further recombination cessation toward the PAR.

Synonymous divergence (d_S) between alleles associated to the alternative mating types at the genes located between the PR-proximal region and the centromere in *M. lagerheimii* and *M. saponariae* provided further evidence for complete centromere linkage of the PR mating-type loci. For all genes linked to a mating-type locus, the same allele remains associated with the same mating-type, so that alleles associated to the alternate mating types accumulate independent mutations, showing increasing divergence (d_S) with time since the complete recombination cessation. To recover the history of recombination cessation, we plotted allelic divergence (d_S) along the ancestral gene order, as subsequent rearrangements

closer to the centromere rather than closer to the purple region, in both *M. saponariae* and *M. lagerheimii* (Fig. 4A,C), suggest stepwise extension of recombination cessation farther from the PR locus and to eventually reach the centromere. The synonymous divergence between the HD-proximal blue region and its centromere in both species was almost zero, although some genes exhibited nonzero d_S value (Fig. 4B,D).

We found increased transposable element content in the HD and PR chromosomes in *M. saponariae* and *M. lagerheimii* compared to their autosomes. There were differences within each species between a_1 and a_2 mating-type chromosomes (Supplemental Fig. S5). Such transposable element accumulation and differences between homologous chromosomes further supported complete recombination cessation.

Absence of *trans*-specific polymorphism between centromeres and the HD- and PR-proximal regions in *M. saponariae* and *M. lagerheimii*

We used genealogies of genes located between centromeres and the HD- and PR-proximal blue and purple regions, respectively, to assess whether linkage of mating-type loci to centromeres in *M. saponariae* and *M. lagerheimii* derived from a single event predating their speciation or from independent events in each lineage. If recombination cessation predates speciation, the alleles associated to the alternative mating types will cluster by mating type rather than by species (which is called *trans*-specific polymorphism), because the alleles will have been linked to mating types since before the speciation. In contrast, if linkage is more recent than speciation, alleles will cluster by species because recombination will have broken any allelic association with mating-type within species after speciation. None of the orthologous groups corresponding to genes located between the PR- or the HD-proximal purple
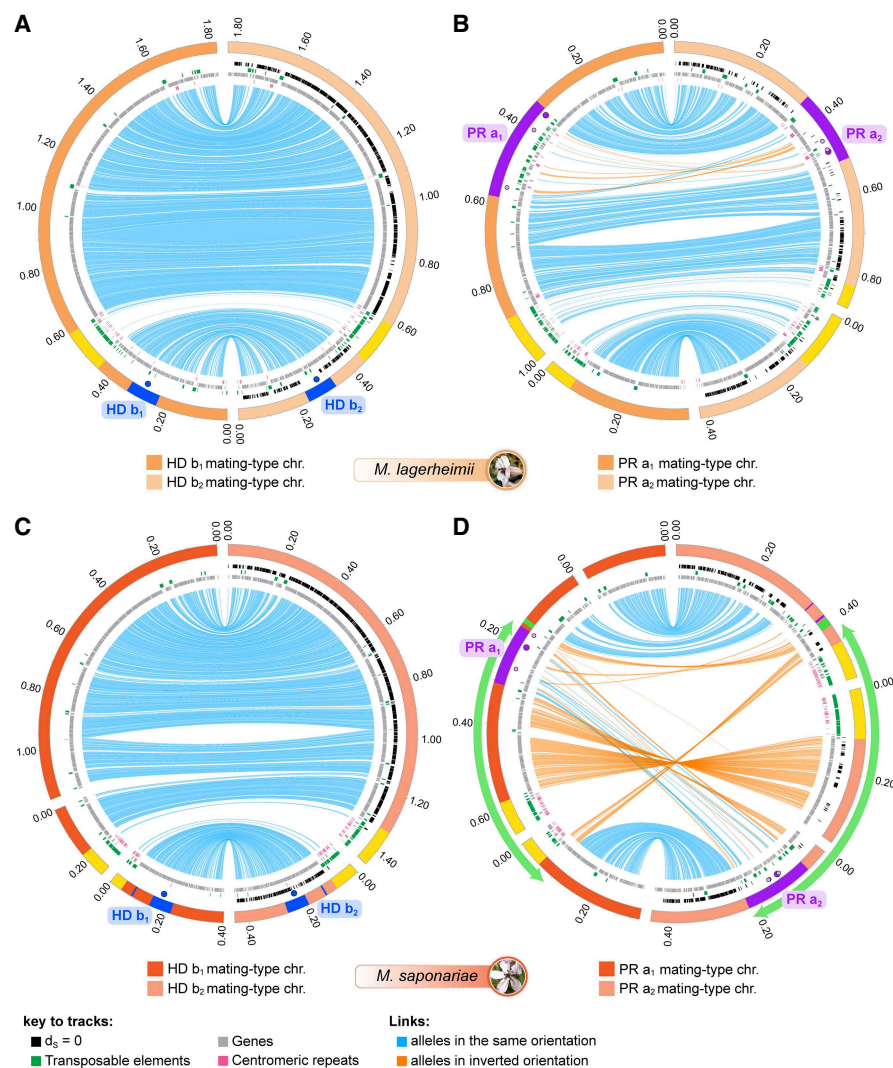
Figure 3. Intraspecific comparison of gene order between mating-type chromosomes. Comparison of gene order between HD and PR chromosome pairs in *M. lagerheimii* (*A,B*) and *M. saponariae* (*C,D*). The *outer* tracks represent contigs, staggered every 200 kb. The HD, PR, and pheromone genes are indicated by blue, dark purple, and small light purple circles, respectively. Blue and orange lines link alleles, the latter corresponding to inversions. The link width is proportional to the corresponding gene length. Yellow regions on the contig track indicate the centromeres (regions with low gene density, high TE density, and enriched in tandem-repeats marked in pink). The black marks along the *right* contigs track indicate genes that have no synonymous substitutions between a_1 and a_2 alleles within species ($d_S = 0$). Green marks indicate transposable elements (TEs), and gray marks non-TE genes. The ancient regions of recombination suppression are indicated on the *outer* track in blue for the HD locus and in purple for the PR locus. (*A*) Comparison of the b_1 (*left*, orange) and b_2 (*right*, light orange) HD *M. lagerheimii* mating-type chromosomes. (*B*) Comparison of the a_1 (*left*, orange) and a_2 (*right*, light orange) PR *M. lagerheimii* mating-type chromosomes. (*C*) Comparison of the b_1 (*left*, red) and b_2 (*right*, light red) HD *M. saponariae* mating-type chromosomes. (*D*) Comparison of the a_1 (*left*, red) and a_2 (*right*, light red) PR *M. saponariae* mating-type chromosomes. The large green arrow indicates the large inversion between the two mating-type chromosomes encompassing the mating-type locus and the centromere. The green regions on the contig track of each mating-type chromosome indicate the small inversion that likely occurred after the large inversion linking the PR locus to the centromere, extending the region of suppressed recombination.

and blue regions and their centromeres displayed *trans*-specific polymorphism shared by *M. saponariae* and *M. lagerheimii*. We could use nine genes in the HD mating-type chromosome (Supplemental Fig. S6A) and 10 genes in the PR mating-type chromosomes (Supplemental Fig. S6B), for which both alleles were available in all species with available genomes (the genes are indicated by red arrows on Fig. 4). These findings indicate independent

events of complete mating-type-loci-centromere linkage in *M. saponariae* and *M. lagerheimii*.

We found further support for recombination suppression occurring after *M. saponariae* and *M. lagerheimii* speciation by dating the differentiation between alleles associated with a_1 versus a_2 mating types in gene genealogies. We computed a phylogenetic tree using the concatenated alignments of the nine and 10 genes with both alleles available in all genomes and located between the HD-proximal region and the centromere, or the PR-proximal region and the centromere, respectively. We calibrated the tree nodes using the speciation date between *Microbotryum lychnidis-dioicae* and *Microbotryum silenes-dioicae*, previously estimated at 420 ky (Gladieux et al. 2011). Although these estimates are not robust absolute dates, they are useful to obtain relative dates of speciation and chromosome evolution events. Recombination cessation between the PR-proximal purple region and the centromere was younger in *M. saponariae* and *M. lagerheimii* (95% confidence interval 171–302 and 80–158 ky, respectively) (Fig. 5A) than their speciation event (95% confidence interval 2997–4386 ky) (Fig. 5B). The date of recombination cessation between the HD-proximal blue region and the centromere was even younger in both species (95% confidence interval 0.1–18 and 3–26 ky, respectively) (Fig. 5A).

Discussion

Here, we document convergent evolution of increased odds of gamete compatibility under automixis by independent linkage events of mating-type loci and centromeres in two closely related fungal species. Such linkage represents further convergence in gamete compatibility patterns with other congeneric lineages, which were previously shown to have achieved similar gamete compatibility odds through multiple independent direct linkage events between PR and HD mating-type loci (Branco et al. 2018). Linkage of the two mating-type loci, one to each other or to their centromere, are equally beneficial under automixis in terms of gamete compatibility odds (Fig. 1; Hood et al. 2015). We found here that the two mating-type loci in *M. lagerheimii*, although located on separate chromosomes, are completely linked to their respective centromeres, as in *M. saponariae*. Furthermore, we showed that in these sister species of anther-smut fungi such linkage occurred through independent recombination cessation events. Convergence of mating-type
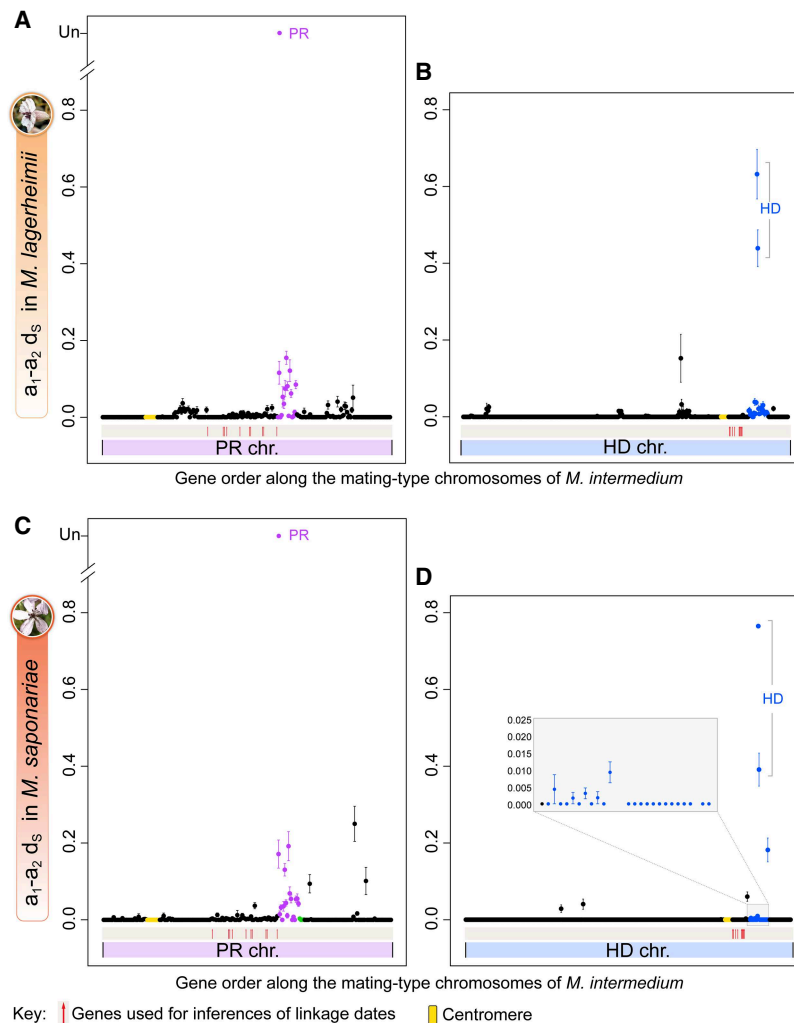
Figure 4. Per-gene synonymous divergence and respective standard error ($d_S \pm$ SE) between alleles associated to the a_1b_1-a_2b_2 mating types along the mating-type chromosomes within diploid *M. lagerheimii* and *M. saponariae* individuals. Synonymous divergence is plotted against the genomic coordinates of the $a_1 b_1$ mating-type chromosomes of *M. intermedium* for all single-copy genes shared by the mating-type chromosomes, as a proxy for ancestral gene order. Divergence between the a_1 and a_2 pheromone receptor (PR) was too extensive (Devier et al. 2009) and could not be computed (noted as unalignable [UN]). The positions of the centromeres are indicated by yellow dots. Genes with $d_S > 0$ between mating types around the PR and HD mating-type loci in *M. lagerheimii* are in purple and blue, respectively. The purple and blue regions correspond to the older PR and HD regions of suppressed recombination that evolved before and at the base of the radiation of the clade, respectively (Branco et al. 2017, 2018). Red arrows indicate the genes used for dating recombination cessation events. (*A*) Per-gene synonymous divergence between mating types in *M. lagerheimii* along the gene order of the a_2 PR *M. intermedium* mating-type chromosome. (*B*) Per-gene synonymous divergence between mating types in *M. lagerheimii* along the gene order of the b_2 HD *M. intermedium* mating-type chromosome. (*C*) Per-gene synonymous divergence between mating types in *M. saponariae* along the gene order of the a_2 PR *M. intermedium* mating-type chromosome. (*D*) Per-gene synonymous divergence between mating types in *M. saponariae* along the gene order of the b_2 HD *M. intermedium* mating-type chromosome.

loci and centromere linkage seems to have also occurred at much larger phylogenetic scale within fungi. *Cryptococcus amylolentus*, a distant *Microbotryum* fungal relative, also displays both HD and PR genes linked to different centromeres (Sun et al. 2017a). Our study thus shows that natural selection can lead repeatedly to similar genomic changes but also to distinct and equally beneficial solutions under a shared evolutionary pressure. These findings contribute to our understanding of evolution and the degree to which it is repeatable.

Segregation analyses showed that the HD and PR loci are linked to their respective centromere in *M. lagerheimii*, as previously shown in its sister species *M. saponariae* (Hood et al. 2015). Although the inclusion of a finite number of analyzed meiotic tetrads leaves the possibility that linkage to centromeres is only nearly complete, our genomic results support complete linkage in the PR chromosome. We found substantial differentiation between alleles associated to alternative mating types at genes between the PR-proximal region and its centromere in both species, as well as a large inversion in *M. saponariae*. Additional evidence of complete recombination cessation is that the HD chromosome in *M. saponariae* and the PR chromosomes in both species are size dimorphic, with size cosegregating with mating-type alleles (Hood et al. 2015). Chromosome size dimorphism is likely due to the differential transposable element amounts we found between alternate HD and PR chromosomes in each species.

The absence of *trans*-specific polymorphism and the more recent inferred linkage dates compared to the speciation event between *M. saponariae* and *M. lagerheimii* strongly support that recombination cessation was independent in the two species. The recent origins of complete linkage between the two mating-type loci and their centromeres in both species are corroborated by the low synonymous divergence values between a_1- and a_2-associated alleles and the lack of extensive rearrangements in the regions without recombination between the mating-type locus proximal regions and the centromeres.

For linkage of mating-type loci to centromeres to be beneficial under selfing by automixis, both HD and PR mating-type genes need to be linked to their respective centromere (Zakharov 1986, 2005). However, for both *M. lagerheimii* and *M. saponariae*, the PR linkage to the centromere evolved long before the HD-centromere linkage. The PR locus-centromere linkage alone provides no advantage concerning gamete compatibility odds when mating occurs within a tetrad; however, the PR-centromere linkage may have been generated in several steps extending the recombination cessation region beyond mating-type genes, as previously described (Branco et al. 2017). This would be consistent with the apparent heterogeneity in the d_S values in the genes between the PR-proximal region and the centromere, with highest d_S values for genes closer to PR than to the centromere (Fig. 4). Under this hypothesis, expansion of the regions of suppressed recombination would have
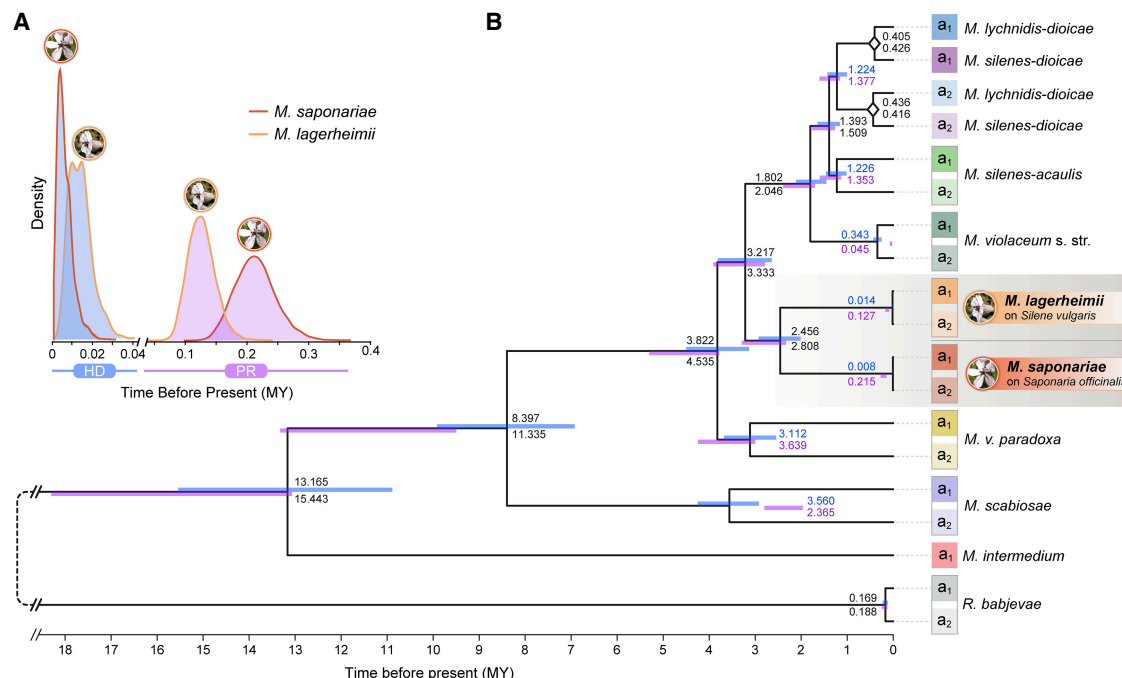
Figure 5. Linkage date estimates across *Microbotryum*. Linkage between mating-type loci (PR and HD) and centromeres, or between PR and HD loci, was inferred from dates of divergence between alleles associated to the a_1 and a_2 mating types at genes linked to the mating-type loci. In gene genealogies, nodes separating alleles associated to the a_1 and a_2 mating types within species correspond to the date of linkage to the mating-type loci (to the HD locus in blue and to the PR locus in purple). Genealogies of alleles associated to the alternative mating types (a_1 and a_2) were reconstructed based on a concatenated alignment of both alleles at nine genes ancestrally located between the centromere and the HD-proximal region (8525 aligned codons) and of 10 genes ancestrally located between the centromere and the PR-proximal region (10,200 aligned codons). The genes used for this analysis, with both alleles in all species, are indicated by red arrows in Figure 4. (*A*) Marginal posterior densities for the most recent common ancestor date (MRCA time) estimated with BEAST v2.4.0 based on multiple sequence alignments of genes located between the HD-proximal (blue) or the PR-proximal (light purple) region and the corresponding centromeres in the ancestral gene order. (*B*) Time-calibrated tree of a_1 and a_2 alleles with nodes drawn at the mean date (in million years [MY]) for the genes on the HD chromosome. Inferred divergence dates for the genes located between the HD-proximal or the PR-proximal regions and the corresponding centromeres are shown in blue or purple fonts, respectively, to the *right* of the nodes. Light blue and light purple bars correspond to 95% confidence intervals. Speciation dates as inferred from each data set are shown in black font on the *right* side of the nodes (*bottom*, PR set; *top*, HD set).

occurred through processes unrelated to the mating system, such as the accumulation and methylation of transposable elements in nonrecombining regions and in their margins (Ponnikas et al. 2018). Once the PR-centromere linkage was achieved, selection for HD-centromere linkage may have occurred and been selected for increasing odds of compatibility under automixis.

Alternatively, the close physical proximity of the HD locus to the centromere may be sufficient to render recombination events infrequent enough that recombination cessation between the PR locus and its centromere would be immediately beneficial for increasing the odds of gamete compatibility under automixis. This hypothesis of rare recombination between the centromere and the HD would explain the very low d_S values in this region and the collinearity between b_1 and b_2 HD mating-type chromosomes in *M. saponariae* despite the inversion that occurred between HD and the centromere since its speciation from *M. lagerheimii*. This hypothesis is not incompatible with the PR-centromere linkage having evolved by successive evolutionary steps. Low recombination rates have been invoked in sex chromosomes to explain low differentiation between alleles on X and Y Chromosomes in some animals (Stöck et al. 2013).

Although inversions are often thought to play a major role in suppressing recombination (Lemaitre et al. 2009; Wang et al. 2012; Wright et al. 2016), nonrecombining regions with conserved collinearity have been reported in several fungi (Jacobson 2005; Grognet et al. 2014; Branco et al. 2017, 2018; Sun et al. 2017b).

In this study, we only found inversions in the region with recent recombination cessation between the *M. saponariae* PR chromosomes. Finding that the limits of this inversion are precisely the centromere and the PR-proximal region is consistent with a role of inversions in recombination suppression, although we cannot exclude that the inversion occurred as a subsequent rearrangement after recombination cessation. The remaining mating-type loci and centromere linkage events occurred via different proximal mechanisms not involving rearrangements. Elucidating the proximal mechanisms suppressing recombination by exploring, for example, changes in DNA methylation and heterochromatin marks, as well as Spo11-dependent formation of double-strand breaks (Keeney 2008; Termolino et al. 2016), will be interesting in future studies.

Our results have general implications beyond mating systems in fungi and evolutionary convergence, providing an excellent illustration of the benefits of mating via central fusion automixis. Broader implication of central fusion automixis has been rarely considered, despite occurring in a wide range of taxa such as in fungi (Hood and Antonovics 2000; Zakharov 2005; Menkis et al. 2008; Grognet et al. 2014), plants (Asker 1980; Walker 1985; Antonius and Nybom 1995; Cruden and Lloyd 1995; Schön et al. 2009), and insects (Suomalainen et al. 1976; Normark 2003; Oldroyd et al. 2008). Theoretical models and reviews have highlighted the evolutionary and genetic consequences of central fusion automixis in maintaining heterozygosity (Zakharov 1986,

2005; Antonovics and Abrams 2004; Hood et al. 2005; Engelstädter 2017), but few cases have been experimentally studied so far. Our findings illustrate how linkage to centromere under central fusion automixis can generate a sort of pseudolinkage among genes on different chromosomes and preserve heterozygosity. Such maintenance of heterozygosity can be beneficial in a variety of cases (Ferreira and Amos 2006), such as the sheltering of deleterious alleles (Hood and Antonovics 2000) or overdominance in immune systems (Hraber et al. 2007) and other functions, as suggested at several loci in the case of the central fusion automictic Cape honeybees (Goudie et al. 2014).

Methods

To conduct segregation analyses, we isolated *M. lagerheimii* haploid cells from opposite poles of meiosis I across replicate meioses from the same diploid parent using micromanipulation (Hood et al. 2015). We investigated mating-type segregation by PCR amplification of allele-specific markers. The *M. lagerheimii* strain used for segregation analyses was collected on *Lychnis flos-jovis* in Valle Pesio, Italy (GPS 44.188400, 7.670650).

Genome analyses were conducted in the *M. saponariae* and *M. lagerheimii* assemblies. We used alternative mating types isolated from a single diploid spore of *M. saponariae* parasitizing *Saponariae officinalis* (cell 1268, PRAT 47, a_1 b_1, and cell 1269, PRAT 48, a_2 b_2) collected near Chiusa di Pesio, Italy (GPS coordinates 44.31713297, 7.622967437 on July 8, 2012). We used the *M. lagerheimii* genome previously published (a_1 b_1 and a_2 b_2 assemblies GCA_ 900015505.1 and GCA_900013405.1, respectively) (Branco et al. 2017). DNA was extracted with the QIAGEN Genomic-tip 100/G (catalog number 10243) and Genomic DNA Buffer Set (catalog number 19060) following the manufacturer's instructions and using a Carver hydraulic press (catalog number 3968). Haploid genomes were sequenced using the P6/C4 Pacific Biosciences SMRT technology (UCSD IGM Genomics Facility). Assemblies of the genomes were generated with the wgs-8.2 version of the PBcR assembler (Koren et al. 2012). Contigs were aligned with optical maps of the two mating-type chromosomes obtained previously (Hood et al. 2015), with MapSolver software (OpGen); see statistics on assemblies in Supplemental Tables S1 and S2. We obtained orthologous groups with orthAgogue (Ekseth et al. 2014) based on BLASTP+2.2.30 followed by Markov clustering (Van Dongen 2000). We aligned the protein sequences of 780 fully conserved single-copy genes with MAFFT v7.388 (Katoh and Standley 2013) and obtained the codon-based CDS alignments with TranslatorX (Abascal et al. 2010). We used RAxML 8.2.7 (Stamatakis 2006) to obtain maximum likelihood gene trees for all 780 fully conserved single-copy genes and a species tree with the concatenated alignment. We estimated synonymous divergence (d_S) and its standard error with the yn00 program of the PAML package (Yang 2007).

We used nine orthologous groups (8525 aligned codons) for dating the recombination cessation between the HD-proximal region and the centromere and 10 orthologous (10,200 aligned codons) groups for dating recombination cessation between the PR-proximal region and the centromere. Divergence times were estimated using BEAST v2.4.0 (Drummond and Rambaut 2007), with the XLM inputs being generated using BEAUTi (Drummond et al. 2012).

Transposable elements were identified and annotated de novo in the high-quality genome assemblies, using both LTR-harvest (defaults parameters) (Ellinghaus et al. 2008); and RepeatModeler (defaults parameters) (Smit and Hubley 2015). We identified de novo centromeric-specific repeats (Melters et al.

2013) using Tandem-Repeat Finder (TRF v. 4.07b) (Benson 1999); see Supplemental File 1.

Data access

The Pacific Biosciences (PacBio) genome assemblies from this study have been submitted to the European Nucleotide Archive (ENA; https://www.ebi.ac.uk/ena) under accession number GCA_900015975 for the a_1 genome and GCA_900015475 for the a_2 genome of *Microbotryum saponariae* from *Saponaria officinalis*.

References

Abascal F, Zardoya R, Telford MJ. 2010. TranslatorX: multiple alignment of nucleotide sequences guided by amino acid translations. *Nucleic Acids Res* **38**(Web Server issue)**:** W7–W13. doi:10.1093/nar/gkq291

Abbate J, Gladieux P, Hood ME, De Vienne DM, Antonovics J, Snirc A, Giraud T. 2018. Co-occurrence among three divergent plant-castrating fungi in the same *Silene* host species. *Mol Ecol* **27:** 3357–3370. doi:10.1111/mec.14805

Antonius K, Nybom H. 1995. Discrimination between sexual recombination and apomixes/automixis in a *Rubus plant* breeding programme. *Hereditas* **123:** 205–213. doi:10.1111/j.1601-5223.1995.00205.x

Antonovics J, Abrams JY. 2004. Intratetrad mating and the evolution of linkage relationships. *Evolution* **58:** 702–709. doi:10.1111/j.0014-3820.2004.tb00403.x

Asker S. 1980. Gametophytic apomixis: elements and genetic regulation. *Hereditas* **93:** 277–293. doi:10.1111/j.1601-5223.1980.tb01367.x

Badouin H, Hood ME, Gouzy J, Aguileta G, Siguenza S, Perlin MH, Cuomo CA, Fairhead C, Branca A, Giraud T. 2015. Chaos of rearrangements in the mating-type chromosomes of the anther-smut fungus *Microbotryum lychnidis-dioicae*. *Genetics* **200:** 1275–1284. doi:10.1534/genetics.115.177709

Bakkeren G, Kronstad JW. 1994. Linkage of mating-type loci distinguishes bipolar from tetrapolar mating in basidiomycetous smut fungi. *Proc Natl Acad Sci* **91:** 7085–7089. doi:10.1073/pnas.91.15.7085

Benson G. 1999. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res* **27:** 573–580. doi:10.1093/nar/27.2.573

Beukeboom LW, Perrin N. 2014. *The evolution of sex determination*. Oxford University Press, New York.

Billiard S, López-Villavicencio M, Devier B, Hood ME, Fairhead C, Giraud T. 2011. Having sex, yes, but with whom? Inferences from fungi on the evolution of anisogamy and mating types. *Biol Rev* **86:** 421–442. doi:10.1111/j.1469-185X.2010.00153.x

Billiard S, López-Villavicencio M, Hood ME, Giraud T. 2012. Sex, outcrossing and mating types: unsolved questions in fungi and beyond. *J Evol Biol* **25:** 1020–1038. doi:10.1111/j.1420-9101.2012.02495.x

Booth W, Schuett GW. 2015. The emerging phylogenetic pattern of parthenogenesis in snakes. *Biol J Linn Soc* **118:** 176–186. doi:10.1111/bij.12744

Booth W, Johnson DH, Moore S, Schal C, Vargo EL. 2011. Evidence for viable, non-clonal but fatherless Boa constrictors. *Biol Lett* **7:** 253–256. doi:10.1098/rsbl.2010.0793

Branco S, Badouin H, Rodríguez de la Vega RC, Gouzy J, Carpentier F, Aguileta G, Siguenza S, Brandenburg JT, Coelho MA, Hood ME, et al. 2017. Evolutionary strata on young mating-type chromosomes despite the lack of sexual antagonism. *Proc Natl Acad Sci* **114:** 7067–7072. doi:10.1073/pnas.1701658114

Branco S, Carpentier F, Rodríguez de la Vega RC, Badouin H, Snirc A, Le Prieur S, Coelho MA, de Vienne DM, Hartmann FE, Begerow D, et al. 2018. Multiple convergent supergene evolution events in mating-type chromosomes. *Nat Commun* **9:** 2000. doi:10.1038/s41467-018-04380-9

Bueker B, Eberlein C, Gladieux P, Schaefer A, Snirc A, Bennett DJ, Begerow D, Hood ME, Giraud T. 2016. Distribution and population structure of the anther smut *Microbotryum silenes-acaulis* parasitizing an arctic-alpine plant. *Mol Ecol* **25:** 811–824. doi:10.1111/mec.13512

Chantha SC, Herman AC, Platts AE, Vekemans X, Schoen DJ. 2013. Secondary evolution of a self-incompatibility locus in the *Brassicaceae* genus *Leavenworthia*. *PLoS Biol* **11:** e1001560. doi:10.1371/journal.pbio.1001560

Chapman DD, Shivji MS, Louis E, Sommer J, Fletcher H, Prodöhl PA. 2007. Virgin birth in a hammerhead shark. *Biol Lett* **3:** 425–427. doi:10.1098/rsbl.2007.0189

Charlesworth D. 2002. Plant sex determination and sex chromosomes. *Heredity (Edinb)* **88:** 94–101. doi:10.1038/sj.hdy.6800016

Charlesworth D, Charlesworth B. 1987. Inbreeding depression and its evolutionary consequences. *Annu Rev Ecol Evol Syst* **18:** 237–268. doi:10.1146/annurev.es.18.110187.001321

Charlesworth D, Morgan MT, Charlesworth B. 1990. Inbreeding depression, genetic load, and the evolution of outcrossing rates in a multilocus system with no linkage. *Evolution* **44:** 1469–1489. doi:10.1111/j.1558-5646.1990.tb03839.x

Coelho MA, Bakkeren G, Sun S, Hood ME, Giraud T. 2017. Fungal sex: the Basidiomycota. *Microbiol Spectr* **5:** FUNK-0046-2016. doi:10.1128/microbiolspec.FUNK-0046-2016

Cruden RW, Lloyd RM. 1995. Embryophytes have equivalent sexual phenotypes and breeding systems: why not a common terminology to describe them? *Am J Bot* **82:** 816–825. doi:10.1002/j.1537-2197.1995.tb15694.x

Devier B, Aguileta G, Hood ME, Giraud T. 2009. Ancient *trans*-specific polymorphism at pheromone receptor genes in basidiomycetes. *Genetics* **181:** 209–223. doi:10.1534/genetics.108.093708

Drummond AJ, Rambaut A. 2007. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol* **7:** 214. doi:10.1186/1471-2148-7-214

Drummond AJ, Suchard MA, Xie D, Rambaut A. 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol Biol Evol* **29:** 1969–1973. doi:10.1093/molbev/mss075

Dudgeon CL, Coulton L, Bone R, Ovenden JR, Thomas S. 2017. Switch from sexual to parthenogenetic reproduction in a zebra shark. *Sci Rep* **7:** 40537. doi:10.1038/srep40537

Ekseth OK, Kuiper M, Mironov V. 2014. OrthAgogue: an agile tool for the rapid prediction of orthology relations. *Bioinformatics* **30:** 734–736. doi:10.1093/bioinformatics/btt582

Ellinghaus D, Kurtz S, Willhoeft U. 2008. LTRharvest, an efficient and flexible software for de novo detection of LTR retrotransposons. *BMC Bioinformatics* **9:** 18. doi:10.1186/1471-2105-9-18

Engelstädter J. 2017. Asexual but not clonal: evolutionary processes in automictic populations. *Genetics* **206:** 993–1009. doi:10.1534/genetics.116.196873

Faino L, Seidl MF, Datema E, van den Berg GC, Janssen A, Wittenberg AH, Thomma BP. 2015. Single-molecule real-time sequencing combined with optical mapping yields completely finished fungal genome. *MBio* **6:** e00936-15. doi:10.1128/mBio.00936-15

Feldheim KA, Clews A, Henningsen A, Todorov L, McDermott C, Meyers M, Bradley J, Pulver A, Anderson E, Marshall A. 2017. Multiple births by a captive swellshark *Cephaloscyllium ventriosum* via facultative parthenogenesis. *J Fish Biol* **90:** 1047–1053. doi:10.1111/jfb.13202

Ferreira ÁG, Amos W. 2006. Inbreeding depression and multiple regions showing heterozygote advantage in *Drosophila melanogaster* exposed to stress. *Mol Ecol* **15:** 3885–3893. doi:10.1111/j.1365-294X.2006.03093.x

Fortuna TM, Snirc A, Badouin H, Gouzy J, Siguenza S, Esquerre D, Le Prieur S, Shykoff JA, Giraud T. 2016. Polymorphic microsatellite markers for the tetrapolar anther-smut fungus *Microbotryum saponariae* based on genome sequencing. *PLoS One* **11:** e0165656. doi:10.1371/journal.pone.0165656

Fortuna TM, Namias A, Snirc A, Branca A, Hood ME, Raquin C, Shykoff JA, Giraud T. 2018. Multiple infections, relatedness and virulence in the anther-smut fungus castrating *Saponaria* plants. *Mol Ecol* **27:** 4947–4959. doi:10.1111/mec.14911

Giraud T, Yockteng R, Lo M. 2008. Mating system of the anther smut fungus *Microbotryum violaceum*: selfing under heterothallism. *Eukaryot Cell* **7:** 765–775. doi:10.1128/EC.00440-07

Gladieux P, Vercken E, Fontaine MC, Hood ME, Jonot O, Couloux A, Giraud T. 2011. Maintenance of fungal pathogen species that are specialized to different hosts: allopatric divergence and introgression through secondary contact. *Mol Biol Evol* **28:** 459–471. doi:10.1093/molbev/msq235

Gladieux P, Ropars J, Badouin H, Branca A, Aguileta G, de Vienne DM, Rodríguez de la Vega RC, Branco S, Giraud T. 2014. Fungal evolutionary genomics provides insight into the mechanisms of adaptive divergence in eukaryotes. *Mol Ecol* **23:** 753–773. doi:10.1111/mec.12631

Goldberg EE, Igić B. 2012. Tempo and mode in plant breeding system evolution. *Evolution* **66:** 3701–3709. doi:10.1111/j.1558-5646.2012.01730.x

Goldberg EE, Kohn JR, Lande R, Robertson KA, Smith SA, Igic B. 2010. Species selection maintains self-incompatibility. *Science* **330:** 493–495. doi:10.1126/science.1194513

Goudie F, Oldroyd B. 2014. Thelytoky in the honey bee. *Apidologie* **45:** 306–326. doi:10.1007/s13592-013-0261-2

Goudie F, Allsopp MH, Oldroyd BP. 2014. Selection on overdominant genes maintains heterozygosity along multiple chromosomes in a clonal lineage of honey bee. *Evolution* **68:** 125–136. doi:10.1111/evo.12231

Grognet P, Bidard F, Kuchly C, Ho Tong LC, Coppin E, Benkhali JA, Couloux A, Wincker P, Debuchy R, Silar P. 2014. Maintaining two mating types: structure of the mating type locus and its role in heterokaryosis in *Podospora anserina*. *Genetics* **197:** 421–432. doi:10.1534/genetics.113.159988

Hanschen ER, Herron MD, Wiens JJ, Nozaki H, Michod RE. 2018. Repeated evolution and reversibility of self-fertilization in the volvocine green algae. *Evolution* **72:** 386–398. doi:10.1111/evo.13394

Hereford J. 2010. Does selfing or outcrossing promote local adaptation? *Am J Bot* **97:** 298–302. doi:10.3732/ajb.0900224

Hood ME, Antonovics J. 2000. Intratetrad mating, heterozygosity, and the maintenance of deleterious alleles in *Microbotryum violaceum* (= *Ustilago violacea*). *Heredity (Edinb)* **85:** 231–241. doi:10.1046/j.1365-2540.2000.00748.x

Hood ME, Antonovics J. 2004. Mating within the meiotic tetrad and the maintenance of genomic heterozygosity. *Genetics* **166:** 1751–1759. doi:10.1534/genetics.166.4.1751

Hood ME, Katawczik M, Giraud T. 2005. Repeat-induced point mutation and the population structure of transposable elements in *Microbotryum violaceum*. *Genetics* **170:** 1081–1089. doi:10.1534/genetics.105.042564

Hood ME, Scott M, Hwang M. 2015. Breaking linkage between mating compatibility factors: tetrapolarity in *Microbotryum*. *Evolution* **69:** 2561–2572. doi:10.1111/evo.12765

Hraber P, Kuiken C, Yusim K. 2007. Evidence for human leukocyte antigen heterozygote advantage against hepatitis C virus infection. *Hepatology* **46:** 1713–1721. doi:10.1002/hep.21889

Igic B, Lande R, Kohn JR. 2008. Loss of self-incompatibility and its evolutionary consequences. *Int J Plant Sci* **169:** 93–104. doi:10.1086/523362

Jacobson DJ. 2005. Blocked recombination along the mating-type chromosomes of *Neurospora tetrasperma* involves both structural heterozygosity and autosomal genes. *Genetics* **171:** 839–843. doi:10.1534/genetics.105.044040

Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* **30:** 772–780. doi:10.1093/molbev/mst010

Keeney S. 2008. Spo11 and the formation of DNA double-strand breaks in meiosis. *Genome Dyn Stab* **2:** 81–123. doi:10.1007/7050_2007_026

Koren S, Schatz MC, Walenz BP, Martin J, Howard JT, Ganapathy G, Wang Z, Rasko DA, McCombie RW, Jarvis ED, et al. 2012. Hybrid error correction and de novo assembly of single-molecule sequencing reads. *Nat Biotechnol* **30:** 693–700. doi:10.1038/nbt.2280

Lande R. 2015. Evolution of phenotypic plasticity in colonizing species. *Mol Ecol* **24:** 2038–2045. doi:10.1111/mec.13037

Lande R, Schemske DW. 1985. The evolution of self-fertilization and inbreeding depression in plants. I. Genetic models. *Evolution* **39:** 24–40. doi:10.1111/j.1558-5646.1985.tb04077.x

Lemaitre C, Braga MD, Gautier C, Sagot MF, Tannier E, Marais GA. 2009. Footprints of inversions at present and past pseudoautosomal boundaries in human sex chromosomes. *Genome Biol Evol* **1:** 56–66. doi:10.1093/gbe/evp006

Lenormand T, Engelstädter J, Johnston SE, Wijnker E, Haag CR. 2016. Evolutionary mysteries in meiosis. *Philos Trans R Soc B* **371:** 20160001. doi:10.1098/rstb.2016.0001

Lewis KR, John B. 1963. *Chromosome marker*. J & A Churchill Ltd., London.

Melters DP, Bradnam KR, Young HA, Telis N, May MR, Ruby JG, Sebra R, Peluso P, Eid J, Rank D, et al. 2013. Comparative analysis of tandem repeats from hundreds of species reveals unique insights into centromere evolution. *Genome Biol* **14:** R10. doi:10.1186/gb-2013-14-1-r10

Menkis A, Jacobson DJ, Gustafsson T, Johannesson H. 2008. The mating-type chromosome in the filamentous ascomycete *Neurospora tetrasperma* represents a model for early evolution of sex chromosomes. *PLoS Genet* **4:** e1000030. doi:10.1371/journal.pgen.1000030

Mogie M. 1986. Automixis: its distribution and status. *Biol J Linn Soc* **28:** 321–329. doi:10.1111/j.1095-8312.1986.tb01761.x

Nieuwenhuis BP, Billiard S, Vuilleumier S, Petit E, Hood ME, Giraud T. 2013. Evolution of uni- and bifactorial sexual compatibility systems in fungi. *Heredity (Edinb)* **111:** 445–455. doi:10.1038/hdy.2013.67

Normark BB. 2003. The evolution of alternative genetic systems in insects. *Annu Rev Entomol* **48:** 397–423. doi:10.1146/annurev.ento.48.091801.112703

Nougué O, Rode NO, Jabbour-Zahab R, Ségard A, Chevin LM, Haag CR, Lenormand T. 2015. Automixis in Artemia: solving a century-old controversy. *J Evol Biol* **28:** 2337–2348. doi:10.1111/jeb.12757

Oldroyd BP, Allsopp MH, Gloag RS, Lim J, Jordan LA, Beekman M. 2008. Thelytokous parthenogenesis in unmated queen honeybees (*Apis mellifera capensis*): central fusion and high recombination rates. *Genetics* **180:** 359–366. doi:10.1534/genetics.108.090415

Ponnikas S, Sigeman H, Abbott JK, Hansson B. 2018. Why do sex chromosomes stop recombining? *Trends Genet* **34:** 492–503. doi:10.1016/j.tig.2018.04.001

Raper JR. 1966. *Genetics of sexuality in higher fungi.* Ronald Press, New York.

Schön I, Martens K, Van Dijk P. 2009. Apomixis: basics for non-botanists. In *Lost sex: the evolutionary biology of parthenogenesis*, pp. 47–62. Springer, Dordrecht, Netherlands.

Schut E, Hemmings N, Birkhead TR. 2008. Parthenogenesis in a passerine bird, the zebra finch *Taeniopygia guttata*. *Ibis (Lond 1859)* **150:** 197–199. doi:10.1111/j.1474-919X.2007.00755.x

Smit AF, Hubley RR. 2015. RepeatModeler Open-1.0. http://www.repeatmasker.org/.

Sonnenberg AS, Gao W, Lavrijssen B, Hendrickx P, Sedaghat-Tellgerd N, Foulongne-Oriol M, Kong WS, Schijlen EG, Baars JJ, Visser RG. 2016. A detailed analysis of the recombination landscape of the button mushroom *Agaricus bisporus* var. *bisporus*. *Fungal Genet Biol* **93:** 35–45. doi:10.1016/j.fgb.2016.06.001

Stamatakis A. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22:** 2688–2690. doi:10.1093/bioinformatics/btl446

Stöck M, Savary R, Betto-Colliard C, Biollay S, Jourdan-Pineau H, Perrin N. 2013. Low rates of X-Y recombination, not turnovers, account for homomorphic sex chromosomes in several diploid species of palearctic green toads (*Bufo viridis* subgroup). *J Evol Biol* **26:** 674–682. doi:10.1111/jeb.12086

Sun S, Yadav V, Billmyre RB, Cuomo CA, Nowrousian M, Wang L, Souciet JL, Boekhout T, Porcel B, Wincker P, et al. 2017a. Fungal genome and mating system transitions facilitated by chromosomal translocations involving intercentromeric recombination. *PLoS Biol* **15:** e2002527. doi:10.1371/journal.pbio.2002527

Sun Y, Svedberg J, Hiltunen M, Corcoran P, Johannesson H. 2017b. Large-scale suppression of recombination predates genomic rearrangements in *Neurospora tetrasperma*. *Nat Commun* **8:** 1140. doi:10.1038/s41467-017-01317-6

Suomalainen E. 1950. Parthenogenesis in animals. *Adv Genet* **3:** 193–253. doi:10.1016/S0065-2660(08)60086-3

Suomalainen E, Saura A, Lokki J. 1976. Evolution of parthenogenetic insects. In *Evolutionary biology*, pp. 209–257. Springer, Boston.

Termolino P, Cremona G, Consiglio MF, Conicella C. 2016. Insights into epigenetic landscape of recombination-free regions. *Chromosoma* **125:** 301–308. doi:10.1007/s00412-016-0574-9

Van der Beek JG, Los JA, Pijnacker LP. 1998. Cytology of parthenogenesis of five *Meloidogyne* species. *Fundam Appl Nematol* **21:** 393–399.

Van Dongen SM. 2000. "Graph clustering by flow simulation." PhD thesis, Utrecht University, Utrecht.

Vekemans X, Poux C, Goubet PM, Castric V. 2014. The evolution of selfing from outcrossing ancestors in Brassicaceae: What have we learned from variation at the *S*-locus? *J Evol Biol* **27:** 1372–1385. doi:10.1111/jeb.12372

Vercken E, Fontaine MC, Gladieux P, Hood ME, Jonot O, Giraud T. 2010. Glacial refugia in pathogens: European genetic structure of anther smut pathogens on *Silene latifolia* and *Silene dioica*. *PLoS Pathog* **6:** e1001229. doi:10.1371/journal.ppat.1001229

Walker T. 1985. Some aspects of agamospory in ferns—the Braithwaite system. *Proc R Soc Edinb Sect B Biol Sci* **86:** 59–66. doi:10.1017/S026972700000796X

Wang J, Na J, Yu Q, Gschwend AR, Han J, Zeng F, Aryal R, VanBuren R, Murray JE, Zhang W, et al. 2012. Sequencing papaya X and Y^h chromosomes reveals molecular basis of incipient sex chromosome evolution. *Proc Natl Acad Sci* **109:** 13710–13715. doi:10.1073/pnas.1207833109

Watts PC, Buley KR, Sanderson S, Boardman W, Ciofi C, Gibson R. 2006. Parthenogenesis in Komodo dragons. *Nature* **444:** 1021–1022. doi:10.1038/4441021a

Wright AE, Dean R, Zimmer F, Mank JE. 2016. How to make a sex chromosome. *Nat Commun* **7:** 12087. doi:10.1038/ncomms12087

Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol* **24:** 1586–1591. doi:10.1093/molbev/msm088

Zakharov IA. 1986. Some principles of the gene localization in eukaryotic chromosomes. Formulation of the problem and analysis of non-random localization of the mating-type loci in some fungi. *Sov Genet* **22:** 1415–1419.

Zakharov IA. 2005. Intratetrad mating and its genetic and evolutionary consequences. *Russ J Genet* **41:** 402–411. doi:10.1007/s11177-005-0103-z

6 Differential gene expression is associated with degeneration and not sexual antagonism in mating-type chromosomes of anther-smut fungi

Under revision in *Molecular Biology and Evolution* after a first round of review

In organisms with separate and differentiated sexes, genes differentially expressed between sexes (*i.e.* sex-biased genes) can be interpreted as resolving sexual antagonistic conflicts and they have been suggested to drive the evolution of evolutionary strata. Indeed, one type of evidence that it is looked for support of the sexual antagonism theory is to check whether evolutionary strata are enriched in sex-biased genes (Ellegren and Parsch 2007; Parsch and Ellegren 2013; Cox and Calsbeek 2009). Sex-biased genes located in non-recombining region could however alternatively result from degeneration following recombination suppression rather than from sexual antagonistic selection. Indeed, the absence of recombination results in genomic degeneration because deleterious mutations accumulate due to the Hill-Robertson effects. Gene expression may be altered by mutations in sequences involved in the regulation of gene expression or by non-synonymous mutations as changes in amino acid may result in a modulation of the mRNA translation. In *Microbotryum lychnidis-dioicae*, we showed that mating-type antagonism is very unlikely. Any gene differentially expressed found in *M. lychnidis-dioicae* is unlikely to result from mating-type antagonism but rather from degenerative mutations.

In the present study, we therefore investigated whether genes differentially expressed between mating-types (*i.e.* mating-type biased genes) were associated to signatures of degeneration, and if the regions that spent more times without recombination contained more mating-type biased genes. We found an enrichment of mating-type biased genes on mating-type chromosomes compared to autosomes and differential expression was significantly associated with putatively deleterious mutations. Although the causal relationship between degenerative mutations and differential gene expression cannot be demonstrated with the study design, we assessed a directional relationship between the differential expression and putatively deleterious mutations and found that the least expressed alleles was significantly more associated with some types of

degenerative mutations. These findings call for considering a possible important role of degeneration in driving the differential expression of genes located in non-recombining regions of sex-related chromosomes.

My contribution to this study was to (i) add an additional genome to the pre-existing orthologous reconstruction (the missing alternative mating-type from *M. intermedium*, the species most external to the phylogeny and tetrapolar), (ii) perform GC calculation among the predicted coding sequences in a homemade awk script, (iii) and participate in the manuscript writing, discussion on statistics and revision of the manuscript.

Differential gene expression between mating types is associated with sequence degeneration in the absence of sexual antagonism

Wen-Juan Ma[1]*, Fantin Carpentier[2], Tatiana Giraud[2], Michael Hood[1]*

1. Department of Biology, Amherst College, Amherst, 01002 Massachusetts, USA

2. Ecologie Systématique Evolution, Bâtiment 360, Univ. Paris-Sud, AgroParisTech, CNRS, Université Paris-Saclay, 91400, Orsay, France

* Corresponding authors: wenjuanma84@gmail.com, mhood@amherst.edu

Running title: Differential gene expression correlates with sequence degeneration

Abstract (249 words)

In animals and plants, differential expression of genes on differentiated sex chromosomes is widespread and is considered to arise in the context of sexually antagonistic selection. However, there is potential for differential expression in non-recombining regions to be associated with degenerative mutations that occur stochastically in one or the other allele, which has been little studied. The anther-smut fungus *Microbotryum lychnidis-dioicae* is ideal for testing the hypothesis that differential expression is associated with degeneration because: 1) separate haploid cultures of opposite mating types help identify differential expression, 2) there are multiple evolutionary strata on its mating-type chromosomes, reflecting successive events of gene linkage to mating-type loci, and 3) antagonistic selection is unlikely between the isogamous haploid mating types. We found that genes showing differential expression between haploid mating types were enriched on the oldest evolutionary strata, and that several signatures of sequence degeneration were greater in differentially expressed than non-differentially expressed genes within genomic compartments. Two degenerative signatures were significantly predictive of lower expression levels between allele pairs: upstream insertion of transposable elements and acquisition of indels and/or early stop codons. Other associated degenerative mutations included non-synonymous substitutions, altered intron and GC content. Differential gene expression in the absence of sexual antagonism may result from less effective selection in non-recombining regions that contributes to mutation accumulation. The association between differential gene expression and allele degeneration is relevant for a broad range of taxa where mating compatibility or sex is determined by genes located in large regions of recombination suppression.

Introduction

Sexual antagonism occurs when trait values that increase gene transmission through the male function decrease gene transmission through the female function, or conversely (Lande 1980; Rice 1987; Charlesworth et al. 2014; Dean and Mank 2014). Such antagonistic conflict can be resolved by the differential gene expression between sexes, resulting in "sex-biased genes" (Ellegren and Parsch 2007; Parsch and Ellegren 2013). Sex-biased genes can be autosomal and hormonally regulated, or they can be linked to the sex-determining genes on sex chromosomes (Yang et al., 2006; Parsch & Ellegren 2013; Pointer et al., 2013; Shen et al, 2017; Catalán et al., 2018). The adaptive hypothesis of sexual antagonism explaining the presence of sex-biased genes on sex chromosomes is commonly found in the literature (Van Doorn and Kirkpatrick 2007; Ellegren and Parsch 2007; Kitano and Peichel 2012; Parsch and Ellegren 2013; Perry et al. 2014; Darolti et al. 2018).

However, non-recombining regions such as those found on sex chromosomes often undergo degenerative changes that may also contribute to differential gene expression. Recombination suppression renders selection less effective due to reduced effective population size (Charlesworth 2012), genetic hitchhiking of deleterious mutations with beneficial ones (Rice 1987), and deleterious mutation sheltering (Rice 1996; Bachtrog 2005; Wright et al. 2016). Although receiving little consideration to date, sex-biased expression could thus result from mutation accumulation in non-recombining regions independent of sexually antagonistic selection, with one of the alleles being more degenerated by chance (but see Lindholm and Breden 2002; Wright et al. 2017). For example, genes with differential expression on neo-sex chromosomes of the passerine bird *Sylvia communis* (Sigeman et al. 2018), and in the analogous mating-type chromosomes in the hermaphroditic fungus *Neurospora tetrasperma* (Samils et al. 2013) were found to have greater sequence divergence between alleles than non-differentially expressed genes. Those studies pointed to sexual antagonism as the likely cause of differential expression between sexes and the relationship to sequence divergence between alleles. However,

100   accumulated mutations with degenerative effects remain an alternative possibility for explaining

101   differential expression.

102   Several different types of mutations can cause sequence degeneration, and some have the

103   potential to alter gene expression. Base pair substitutions and indels (insertion or deletion

104   mutations) can change amino acid sequences which can affect gene expression through

105   modulation of the mRNA translation (Kimball and Jefferson 2004), or disrupt promoter regions

106   that impact transcriptional regulation (Wray et al. 2003). Induction of early stop codons that

107   truncate protein length can lead to post-transcriptional regulatory negative feedbacks upon

108   expression (e.g. nonsense mediated decay; Montgomery et al. 2013). Transposable element

109   insertion in upstream promoter regions, or internal to genes, has long been recognized for effects

110   on expression (Mcclintock 1942; Feschotte 2008; Cordaux and Batzer 2009; Tirosh et al. 2009;

111   Lee and Young 2013). Epigenetic modifications, particularly cytosine methylation, contribute

112   both to heterochromatin formation and elevating mutation rates that reduce GC content (Bird

113   1980; Grummt and Pikaard 2003); thus reduced GC content could represent a signature of

114   methylation-induced gene silencing. Shorter introns are more efficient for correcting transcription

115   (Marais et al. 2005), such that changes in introns can influence transcription rates, nuclear export,

116   and transcript stability (Heyn et al. 2015). These forms of degenerative changes are expected to

117   accumulate under the reduced selection efficacy in non-recombining regions. Yet, very few

118   studies have addressed the association between sequence degeneration and differential gene

119   expression on sex chromosomes (Ellegren and Parsch 2007; Graves 2010; Grath and Parsch

120   2016).

121   Fungi can provide valuable insights into the relationship between sequence degeneration and

122   differential gene expression in mating-type chromosomes that share many features with sex

123   chromosomes (Fraser and Heitman 2004; Hood et al. 2004). The benefits of fungi relative to

124   other types of organisms for such studies include easy access to the haploid phase where alternate

125   mating types are expressed, the existence of young events of recombination suppression in

126 successive evolutionary strata, and the low potential for sexually antagonistic traits in many

127 species (Giraud et al. 2008; Fontanillas et al. 2015; Branco et al. 2017; Branco et al. 2018). The

128 anther-smut fungi, in the genus *Microbotryum*, undergo mating in the haploid phase via

129 isogamous yeast-like cells of opposite mating types (a_1 and a_2), which can be cultured separately

130 to analyze expression levels of alleles (Perlin et al. 2015). The species *Microbotryum lychnidis-*

131 *dioicae,* causing anther-smut disease on the plant *Silene latifolia*, carry dimorphic mating-type

132 chromosomes that have been assembled at the chromosome-level scale (Hood 2002; Hood et al.

133 2013; Fontanillas et al. 2015; Branco et al. 2017). These mating-type chromosomes (a_1, ~3.3Mb,

134 and a_2, ~4.0Mb, respectively) lack recombination across 90% of their length (Hood 2002; Hood

135 et al. 2013). Importantly, evolutionary strata of different ages have been identified, i.e., regions

136 with different levels of differentiation between mating types as a result of an expanding process

137 of recombination suppression over the past ca. 1.5 million years (Branco et al. 2017; Branco et al.

138 2018). The non-recombining regions of the mating-type chromosomes in *M. lychnidis-dioicae* are

139 flanked by small recombining pseudo-autosomal regions (PARs).

140   In *M. lychnidis-dioicae*, any differential gene expression between the alternative haploid

141 mating types would not likely be due to 'mating-type antagonistic selection' (*sensu* Abbate and

142 Hood 2010). The existence of genes under mating-type antagonistic selection in fungi would

143 indeed require fitness differences associated with mating-type dimorphic traits, while previous

144 studies on *M. lychnidis-dioicae* have shown that differences between the mating types, either

145 developmental or ecological, are lacking outside of the immediate process of gamete fusion (Day

146 1979; Garber and Day 1985; Hood and Antonovics 2000; Hood and Antonovics 2004; Giraud et

147 al. 2008). Moreover, a recent study on gene expression and positive selection detected no

148 evidence for mating-type antagonistic selection (Bazzicalupo et al. 2019). This model system is

149 therefore ideal to investigate the association of degenerative mutations with differential gene

150 expression between chromosomes determining reproductive compatibility without the

151 confounding effect of sexual antagonism.

152   In this study, we therefore investigated whether mating-type specific differential gene

153   expression was related to differences between alleles for various signatures of degeneration in the

154   genome of *M. lychnidis-dioicae*. As prior work indicated that non-recombining regions of the

155   mating-type chromosomes were enriched for signatures of sequence degeneration compared to

156   autosomes (Fontanillas et al. 2015), we determined whether differential gene expression varied

157   among genomic compartments defined as autosomes, pseudo-autosomal regions (PARs),

158   youngest evolutionary strata on non-recombining regions of the mating-type chromosomes

159   (including previously identified red and green evolutionary strata, Branco et al. 2017), and oldest

160   evolutionary strata on non-recombining regions (blue, purple, orange and black evolutionary

161   strata, Branco et al. 2017). We studied differential gene expression between mating types only in

162   the haploid stage because the a_1 and a_2 mating types are determined at the haploid stage: mating

163   can only occur between haploid sporidia of opposite mating types. In addition, almost all genes

164   on autosomes and PARs are homozygous due to a high selfing rate (Badouin et al. 2017; Branco

165   et al. 2017), and expression levels only for highly differentiated alleles on oldest evolutionary

166   strata could be assigned to a_1 or a_2 mating types in the diploid or dikaryotic stages, which would

167   profoundly limit and likely bias the analyses by taking into account only degenerated genes. We

168   determined whether differential expression was associated with differences in degenerative

169   mutations between alleles, including comparisons for each degenerative trait for differentially

170   versus non-differentially expressed genes within each genomic compartment. We thus assessed

171   the possibility that the allele showing lower expression levels would have higher levels of

172   degeneration footprints. The investigated degeneration signatures included differences between

173   alleles in the levels of non-synonymous sequence divergence, transposable element (TE)

174   insertions, alteration of predicted protein length (via mechanisms including acquisition of indels

175   and/or early stop codons), intron content, and GC content (as a predicted consequence of

176   epigenetic gene silencing) (Hoof and Green 1996; Marais et al. 2005; Feschotte 2008; Bedford

177   and Hartl 2009; Cordaux and Batzer 2009). Associations between mating-type specific

178   differential expression and signatures of degeneration, while requiring further study to establish

179   the nature of causality, would reflect an important component of gene evolution aside from

180   antagonistic selection.

181

182   **Results**

183   ***Allele identification and differential gene expression between a_1 and a_2 haploid genomes***

184   Alleles of single-copy genes in *M. lychnidis-dioicae* were identified using the criterion of 1:1

185   reciprocal best BLASTp as well as OrthoMCL, giving nearly (99.95%) identical results, between

186   a_1 and a_2 haploid genomes, based on the previously published genome assembly and gene

187   annotation (Branco et al. 2017; Branco et al. 2018). After filtering out TE-related gene sequences,

188   we identified 371 single-copy allelic pairs in mating-type chromosomes and 9,025 in autosomes

189   (Table S1).

190   We used whole-genome RNA-seq data from each of two replicate cultures for separate a_1

191   and a_2 haploid mating-type sporidia of *M. lychnidis-dioicae,* under low nutrient conditions, that

192   resemble the natural haploid growth environment (Schäfer et al. 2010; Perlin et al. 2015). We

193   retained 8,549 single-copy genes with significantly detectable expression (see filtering criteria

194   and details in Materials and Methods section) for further analysis (342 on mating-type

195   chromosomes and 8,207 on autosomes). The threshold of differential gene expression profile (i.e.

196   ($|Log2(a_1/a_2)|$) significantly greater than zero with false discovery rate (FDR) < 0.050, Fig. 1;

197   Table S2) revealed 392 genes (4.59% out of the 8,549 genes analyzed for expression) that were

198   significantly more highly expressed in the a_1 haploid culture, and 203 (2.37%) that were

199   significantly more highly expressed in a_2 haploid culture (Fig. S1).

200

201   ***Differential gene expression and multiple signatures of sequence degeneration***

202   Regression analysis (generalized linear model, GLM) revealed that the degree of differential

203   expression (DE) between allele pairs of the two haploid mating types significantly increased with

7

204   increasing differences between alleles (using absolute values) in the various degeneration traits

205   examined (Table 1). The significant main-effect predictors of differential expression included

206   genomic compartment and differences between alleles in non-synonymous divergence (dN),

207   transposable element (TE) insertion number within 20kb (up and downstream), intron content

208   (proportional to coding sequence length), and overall GC content (GC0). Differences between

209   alleles in predicted protein length was not a significant main-effect predictor but was strongly

210   significant as an interaction term with genomic compartment and all other traits except intron

211   content (Table 1; Fig. 2). Differential expression indeed increased with differences between

212   alleles in predicted protein length, but only in oldest evolutionary strata and when associated with

213   higher differences between alleles in dN, TE content, and GC0 (Table 1; Fig. 2). Genes with

214   differential expression between mating types showed significant enrichment in the oldest

215   evolutionary strata compared to autosomes, but there was no enrichment in the youngest

216   evolutionary strata or the PARs (Table 2). Similar patterns were observed for the comparisons in

217   each of the a_1 or a_2 haploid genomes separately (Table S3). Further *post hoc* assessments of

218   degenerative traits are presented in the following sections, including whether difference between

219   alleles is oriented such that the more affected allele is less expressed.

220

221   *Relationship between differential expression and elevated substitution rates*

222   Differentially expressed genes had greater sequence divergence between alleles than non-

223   differentially expressed (non-DE) genes within genomic compartments, specifically within the

224   oldest evolutionary strata of the mating-type chromosomes. DE genes had significantly higher

225   non-synonymous mutation rate (dN) and synonymous mutation rate (dS) between alleles than

226   non-DE genes within oldest evolutionary strata (Wilcoxon rank sum test for independent

227   samples, dN: W = 1433, $P < 0.001$, dS: W = 1422, $P < 0.001$) (Fig. 3A, Fig. S2, Table S4). There

228   was almost no sequence divergence (dN or dS) between alleles on either autosomes or PARs for

229   DE or non-DE genes. The youngest evolutionary strata had only one DE gene, precluding

230 comparison to non-DE genes within this compartment. The oldest strata pattern held for genes in

231 a_1 or a_2 cells considered separately (Fig. S3A and S3B). A tendency of higher dN/dS for DE

232 genes compared to non-DE genes was not significant within oldest strata (W = 1946, P = 0.611)

233 (Fig. S4).

234   Although DE genes were associated with higher dN than non-DE genes, this trait was not

235 directionally related to the difference in expression levels between allele pairs. The allele having

236 lower expression levels in *M. lychnidis-dioicae* did not have significantly greater accumulation of

237 non-synonymous changes than the allele with higher expression. To test this possibility, each

238 allele in *M. lychnidis-dioicae* was compared for sequence divergence with their ortholog in *M.*

239 *lagerheimii*, which has retained largely collinear and homozygous mating-type chromosomes, as

240 the inferred ancestral state in the *Microbotryum* genus (Branco et al. 2017; Branco et al. 2018).

241 Alleles in a_1 haploid genome or in a_2 haploid genome of *M. lychnidis-dioicae* were compared for

242 dN divergence accordingly with alleles from the *M. lagerheimii* genome of the same mating type;

243 dN divergence from their ortholog in *M. lagerheimii* was not greater for alleles in *M. lychnidis-*

244 *dioicae* having lower expression levels than alleles with higher expression levels (W = 1,267,

245 smallest P = 0.909) (Fig. S5, Table S5).

246

247 ***Relationship between differential expression and TE insertions***

248   Differentially expressed genes were associated with greater differences between alleles for

249 TE insertions (within 20kb up and downstream) than alleles of non-DE genes across genomic

250 compartments. However, the difference was significant only in the autosomes (W = 313879, P <

251 0.001, Fig. 3B), not in the PARs (W = 546, P < 0.192) or the oldest evolutionary strata (W =

252 4062, P = 0.173); the comparison was not possible in youngest evolutionary strata (as noted

253 above).

254   The alleles with lower expression did show a pattern of more TE insertions than the alleles

255 with higher expression. To assess this pattern, differences in TE insertion numbers between

256   alleles were calculated as the TE number for the allele with lower expression minus the TE

257   number for the allele with higher expression; a positive value thus represented an excess of TEs

258   in the less expressed allele. This oriented TE number difference between alleles was tested as a

259   predictor of the expression ratio $|Log2(a_1/a_2)|$ using a sliding window approach with a 15kb

260   window size overlapping by 5kb. Among DE genes, oriented TE insertion difference was a

261   significant predictor of the express ratio only in the window covering from 10kb upstream to the

262   gene (Fig. 4A, Fig. S6); alleles with more TE insertions having reduced expression (for this

263   window, Wald $X^2 = 6.674$, $P = 0.010$, statistics of remaining windows in Table S6). Among non-

264   DE genes, none of the windows was a significant predictor of variation in the expression ratio

265   (Table S6).

266

267   ***Relationship between differential expression and differences in predicted protein length***

268   Differential gene expression was associated with the mutational changes that affect the

269   predicted protein length, including altered stop codon positions, indels, and indels causing

270   frameshifts. Within genomic compartment, alternate alleles of DE genes were significantly more

271   likely to produce proteins of different lengths than alleles of non-DE genes, particularly within

272   the oldest evolutionary strata (two-proportion Z test, z = 2.186, $P = 0.029$) and autosomes (z =

273   4.64, $P = 8.78e-06$, Fig. 3C, Table S7); there were too few DE genes on PARs and youngest

274   evolutionary strata for statistical comparisons.

275   The various types of mutational changes that caused protein length variation between alleles

276   differed between DE and non-DE genes, as well as among genomic compartments. Among the

277   258 genes with different protein sequence lengths between alleles, all had indels. However, DE

278   genes in the oldest evolutionary strata and autosomes had significantly more indels than non-DE

279   genes; oldest strata mean indel number differed between alleles by 2.64 for DE genes and by 1.85

280   for non-DE genes (W = 2453.5, $P = 0.013$), and in autosomes alleles differed by a mean of 1.19

281   indels for DE genes and 1.03 for non-DE genes (W = 490.5, P = 0.025, Fig. 5A); PARs and

282   youngest evolutionary strata could not be analyzed.

283   Similarly, differences in the positions of stop codons contributed to protein length variation

284   more for DE genes than non-DE genes. Among genes with different protein lengths between

285   alleles in the oldest evolutionary strata, 44.6% (N = 56) of DE genes had different stop codon

286   positions between alleles, which was significantly higher than the 24.0% (N = 75) of non-DE

287   genes (two-proportion z-test, P = 0.018, Fig. 5B). Similarly, DE genes in the autosomes were

288   marginally significantly more likely to have different stop codon positions between alleles than

289   non-DE genes, with 33.3% (N = 21) vs 10% (N = 40), respectively (P = 0.057, Fig. 5B). Only

290   three frameshift mutations were observed among the 258 of genes examined with different

291   protein/coding sequence lengths, and thus frameshifts were not distinguishing features of DE

292   versus non-DE genes.

293   The alleles with lower expression did show a pattern of truncation of protein length

294   compared to the alleles with higher expression (i.e. by early stop codons or deletions). To assess

295   this pattern, differences in protein length between alleles were calculated as the ratio for the allele

296   with higher expression divided by the allele with lower expression; a larger ratio thus represented

297   a shorter length for the allele with lower expression. Among DE genes, this oriented metric of

298   protein length differences was a significant predictor of the differential expression degree as the

299   ratio $|Log2(a_1/a_2)|$, with alleles producing shorter proteins being less expressed (Wald X^2 =

300   19.326, two-tailed P < 0.001, Fig. 4B). No significant relationship to expression level ratio was

301   found for the length ratios of non-DE genes (Wald X^2 = 0.222, P = 0.638, Fig. 4B).

302

303   ***Relationship between differential expression and intron content***

304   Differential gene expression was associated with differences between alleles in intron

305   content, considering lower intron content to be favored by selection (Marais et al. 2005). There

306   were significantly greater intron content differences between alleles for DE genes than for non-

307   DE genes; considering the ratio of intron to coding sequence lengths, alleles of DE gene overall

308   differed on average by 0.008 and alleles of non-DE genes differed by 0.002 (W = 2102758, $P <$

309   0.001). Alleles differed in intron content more for DE than non-DE genes within the autosomes

310   (W = 1920124, $P = 0.033$) and oldest evolutionary strata (W = 3205, $P = 0.001$) (Fig. 3D, Table

311   S8), but not within the PARs (W = 605, $P = 0.888$); the comparison in youngest evolutionary

312   strata was not possible.

313   The alleles with lower expression did not show a greater intron content (i.e. as a signature of

314   degeneration) than the alleles with higher expression. To assess this possibility, differences

315   between alleles were calculated as the value for the less expressed allele minus the value for the

316   more expressed allele; a positive value thus represented greater intron content for the less

317   expressed allele. This oriented metric of intron content differences between alleles was not a

318   significant predictor of differential expression level among DE genes (Wald $X^2 = 0.350$, $P =$

319   0.554), or among non-DE genes (Wald $X^2 = 0.216$, $P = 0.642$).

320

321   ***Relationship between differential expression and GC content***

322   Consistent with gene silencing by cytosine methylation possibly contributing to decreased

323   GC content (Bird 1980; Grummt and Pikaard 2003; Mugal et al. 2015), DE genes had

324   significantly greater overall GC0 differences between their alleles than non-DE genes within the

325   autosomes (W = 1907831, $P < 0.001$) and oldest evolutionary strata (W = 3010, $P < 0.001$) (Fig.

326   3E). The comparison within the PARs was not significant (W = 578, $P = 0.318$); the comparison

327   for youngest evolutionary strata was not possible. Analysis of third codon position GC3 provided

328   similar patterns and levels of significance (Fig. S7, Table S9).

329   The alleles with lower expression did not show lower GC content than the alleles with higher

330   expression. To assess this possibility, GC0 or GC3 differences between alleles were calculated as

331   the value for allele with higher expression minus the value for allele with lower expression; a

332   positive value thus represented reduced GC content for the allele with lower expression. Among

333 DE genes, neither the oriented GC0 or GC3 differences between alleles were significant

334 predictors of the level of differential expression (GC0: Wald $X^2 = 1.039$, $P = 0.308$, and GC3:

335 Wald $X^2 = 2.226$, $P = 0.136$).

336

Discussion

338   Genomic regions controlling mating compatibility, whether non-recombining sex or mating-

339 type chromosomes, have been subject of intense research. This is because these regions

340 determine traits essential for fitness and because of their rapid evolutionary dynamics. Sequence

341 degeneration of non-recombining regions has been linked to major genomic features such as

342 chromosomal heteromorphism, dosage compensation or gene trafficking between non-

343 recombining regions and autosomes (Mank 2013; Wright et al. 2016). To our knowledge, our

344 study is the first to reveal an association between differential gene expression and a variety of

345 degenerative mutations, doing so in the anther-smut fungus *M. lychnidis-dioicae* where

346 antagonistic selection is unlikely (Hood and Antonovics 2004; Giraud et al. 2008; Branco et al.

347 2017; Branco et al. 2018; Bazzicalupo et al. 2019), and thus does not constitute a confounding

348 factor. Genes differentially expressed between the haploid mating types were enriched only on

349 the oldest strata of the mating-type chromosomes, where they displayed various forms of

350 sequence (*dN*, *dS*, or GC content) or structural (TE insertions, introns content, or protein length)

351 heterozygosity at levels higher than non-differentially expressed genes within the genomic

352 compartments. These results show that differential gene expression is strongly associated with

353 sequence degeneration, which can result either from a direct effect of the studied degenerative

354 mutations or from overall higher rates of mutation accumulation in these regions. Our results

355 suggest that differential expression should therefore be interpreted in a context that includes

356 degenerative mutations in addition to antagonistic selection when studying sex-related non-

357 recombining chromosomes.

358

13

Differential gene expression between haploid mating types

359
360 The proportion of genes with differential expression between haploid mating types of *M.*

361 *lychnidis-dioicae* was low but slightly higher than in a previous study based on the same dataset

362 (Fontanillas et al. 2015), likely due to an improved genome assembly and non-recombining

363 region identification. The 2.4~4.6% of genes with differential expression between mating types

364 of *M. lychnidis-dioicae* was similar to plant and animal non-reproductive tissues, e.g. liver,

365 spleen, leaves, roots (Yang et al. 2006; Ayroles et al. 2009; Perry et al. 2014; Haselman et al.

366 2015; Meisel and Ph 2017; Ma, Veltsos, Sermier, et al. 2018; Ma, Veltsos, Toups, et al. 2018).

367 However, this is much lower than in reproductive tissues (e.g. ovaries or testes) of most animals

368 and plants (Ellegren and Parsch 2007; Parsch and Ellegren 2013). There is an overall positive

369 relationship between proportions of genes with sex-biased expression and levels of sexual

370 dimorphism across taxa (Mank 2017), suggesting a correlation between phenotypic difference

371 between sexes and underlying transcriptional architecture.

372 Nevertheless, differentially expressed (DE) genes were enriched in the mating-type

373 chromosomes of the isogamous *M. lychnidis-dioicae*, which is also consistent with studies in

374 anisogamous animals and plants having differentiated sex chromosomes (reviewed by Ellegren

375 and Parsch 2007). Similarly, in the anisogamous fungus *N. tetrasperma*, DE genes were more

376 frequently detected on mating-type chromosomes (Samils et al. 2013). In animals and plants,

377 linkage of sexually-antagonistic genes to the sex-determining genes in non-recombining regions

378 is considered fundamental to formation of evolutionary strata and the resolution of sexual

379 conflict, for example by allowing for sex-specific or sex-biased gene expression (Rice 1987;

380 Charlesworth et al. 2005; Otto et al. 2011; Lipinska et al. 2017). However, decades of research

381 have uncovered little genetic evidence directly supporting sexually antagonistic selection as the

382 driving force for evolutionary strata (Van Doorn and Kirkpatrick 2007; Innocenti and Morrow

383 2010; Wright et al. 2017; Ma, Veltsos, Sermier, et al. 2018). It is important to note that the

384 potential for degenerative mutations to contribute to differential expression in non-recombining

14

385   regions of sex chromosomes is not exclusive of, and should be considered in addition to, the

386   hypothesis invoking sexual antagonism.

387   Our results encourage a broader view of evolutionary forces, aside from sexual antagonism,

388   that may be related to the occurrence of DE genes and their enrichment on chromosomes

389   determining reproductive compatibility. The relationship of sequence degeneration with

390   differential expression has so far been largely understudied but is likely the primary association

391   in *M. lychnidis-dioicae*. The lack of female and male functions in *Microbotryum* likely explains

392   the absence of genes experiencing mating-type antagonistic selection (Bazzicalupo et al. 2019).

393   Yet, gene degeneration on non-recombining sex or mating-type chromosomes is expected to

394   occurs commonly due to reduced efficacy of selection (Rice 1987; Charlesworth et al. 2005;

395   Ellegren and Parsch 2007; Otto et al. 2011; Fontanillas et al. 2015; Lipinska et al. 2017). The

396   resulting mutation accumulation may generate contrasting expression levels between differently

397   affected alleles or may accumulate where differential expression decreases purifying selection on

398   the less expressed allele. Again, sexually antagonistic selection and degeneration are not mutually

399   exclusives processes, especially for the species having separate sexes where the general

400   relationship between amounts of sexual dimorphism and DE genes is observed (Ellegren and

401   Parsch 2007; Parsch and Ellegren 2013). Still, the potential role of degenerative mutations

402   remains worth considering in all systems. The association found between multiple types of

403   degeneration footprints and differential expression in *M. lychnidis-dioicae*, in the absence of

404   sexual antagonism, suggest the possibility of similar associations across diverse types of

405   organisms. Finally, previous studies show there was overall very low genetic variation within *M.*

406   *lychnidis-dioicae*, both on autosomes and within each of a_1 and a_2 mating-type chromosomes, due

407   to high selfing rate and small effective population sizes (Badouin et al. 2017; Branco et al. 2017).

408   Therefore, we expect most a_1 mating types to be similar in expression and most a_2 mating types

409   likewise.

410

Various forms of degeneration

411

412   The properties of non-recombining regions that reduce the efficiency of selection (reduced

413 N_e, hitchhiking and sheltering) can lead to the fixation of various mutations having degenerative

414 effects, several of which were significant predictors in the overall regression model of differential

415 expression between mating types of *M. lychnidis-dioicae*. Even within genomic compartments,

416 DE genes had significantly higher levels of degenerative mutations distinguishing the alleles than

417 non-DE genes. Some signatures of degeneration, such as TE insertions, indels and/or premature

418 stop codons, may be most plausibly conceived as mechanisms that reduce transcription levels. In

419 particular, transposable element (TE) insertion into genes or upstream have long been recognized

420 to alter gene expression (Mcclintock 1942; Britten and Davidson 1971). TEs can disrupt

421 promoter regions or other regulatory sequences internal to genes (Feschotte 2008; Cordaux and

422 Batzer 2009). In addition, epigenetic silencing, as a defense against TE proliferation, can tighten

423 local chromatin structure and inhibit access of transcriptional machinery (Eichten et al. 2014;

424 King 2015). Consistent with a direct effect upon differential expression, the relative excess of TE

425 insertions between alleles, specifically upstream of genes, was associated with a lower expression

426 level between alleles of DE genes. Similarly, the introduction of early stop or non-sense codons is

427 expected to reduce expression. Transcripts from alleles with premature stop codons are affected

428 by nonsense mediated decay, involving degradation of mRNA and further components of the

429 RNAi pathway than down-regulate expression (Hoof and Green 1996). Importantly, the

430 differential expression between alleles was explained by the shorter allele having lower

431 expression. While TEs, indels and/or premature stop codons are potentially important mutations

432 affecting differential expression between alleles, further studies are needed to directly test the

433 nature of causality between differential expression and specific degenerative mutations. Reduced

434 purifying selection on the less expressed allele in non-recombining regions may have also

435 allowed greater accumulation of the degenerative mutations studied here.

436   Other signature of degeneration in *M. lychnidis-dioicae* were not directionally predictive of

437   lower allele expression but were nevertheless more associated with DE than non-DE genes. Most

438   substantial among these characteristics was the degree of sequence divergence between alleles.

439   Alleles of DE genes were distinguished by markedly more non-synonymous and synonymous

440   base pair differences than alleles of non-DE genes. Similar results were demonstrated in the

441   anisogamous, hermaphroditic ascomycete *N. tetrasperma*, showing that differential gene

442   expression was positively correlated with sequence divergence between alleles of genes on

443   mating-type chromosomes (Samils et al. 2013).

444   The remaining signatures of degeneration may be most informative of how genes might

445   evolve as a consequence of differential expression. In general, gene expression levels are

446   expected to positively correlate with the strength of selection (Bedford and Hartl 2009); stronger

447   selection and higher expression levels have been correlated with reduced intron content (Marais

448   et al. 2005). Our data show that differences between alleles in intron content is positively

449   associated with differential expression level, but the directional hypothesis testing for a possible

450   role in reducing expression levels was not significant. Additionally, our data showed changes in

451   GC content that are consistent with the consequences of suppression of gene expression,

452   specifically as predicted to result from epigenetic silencing through cytosine methylation. Two

453   important drivers of GC content changes are biased gene conversion and methylcytosine-driven

454   C-to-T mutation rates (Grummt and Pikaard 2003; Mugal et al. 2015); however, gene conversion

455   that relies on meiotic pairing is unlikely to cause the different mutation rates between alleles of

456   single copy genes that were studied here (Chen et al. 2007).

457

458   ***Degeneration across genomic compartments***

459   The different forms of genetic degeneration in *M. lychnidis-dioicae* were not equally

460   represented across genomic compartments, perhaps reflecting the history of recombination

461   suppression. In this system, enrichment of DE genes on the mating-type chromosomes is unlikely

462 to be due to antagonistic selection. As a matter of fact, enrichment of DE genes was significant

463 only in the oldest evolutionary strata and not in the younger strata, indicating it is a consequence

464 and not a driver of recombination suppression.

465   Mating between different haploid sexes or mating types ensures that all diploids are

466 heterogametic (Bull 1978), and it has long been recognized that regions linked to mating type can

467 preserve heterozygosity (Mather 1942). In *M. lychnidis-dioicae*, the large non-recombining

468 regions are in fact highly heterozygous (Branco et al. 2017). In contrast, the autosomes and PARs

469 are largely homozygous, due to the selfing mating system of *M. lychnidis-dioicae* (Giraud et al.

470 2008; Hood et al. 2013). Consistent with mating-type linkage preserving heterozygosity, nearly

471 the full range of mutational changes or footprints of degeneration showed lowest levels in the

472 autosomes and PARs and increasing through the youngest evolutionary strata to highest levels in

473 the oldest evolutionary strata. Importantly, however, comparisons within genomic compartments

474 repeatedly showed that allele-distinguishing mutations occurred more in association with DE

475 genes than non-DE genes or in the manner positively associated with levels of differential

476 expression. Therefore, strong evidence is shown for these degenerative changes being associated

477 with changes in expression levels between alleles. Finally, The recent discovery of multiple

478 independent mating-type linkage events across the *Microbotryum* genus (Branco et al. 2018)

479 should allow further assessment of mutation accumulation and its consequences for gene

480 functions.

481

482 ***Conclusions***

483   Our findings on differential gene expression being more frequent in oldest evolutionary strata

484 and being associated with various types of sequence degeneration, shed new lights on how

485 differentially expressed genes might evolve. In animals and plants, it is widely accepted that

486 differential gene expression on sex chromosomes is driven by sexually antagonistic selection

487 (Cox and Calsbeek 2009; Mank 2013). Our study shows that the accumulation of degenerative

488   mutations between alleles is significantly associated with the degree of differential gene

489   expression, in a system where sexual antagonistic selection is unlikely to occur as a confounding

490   factor. Furthermore, the genes with differential expression were highly enriched on mating-type

491   chromosomes, similar to diverse organisms where the separate sex functions have been cited as

492   the primary cause. We further found evidence of a directional relationship between differential

493   gene expression and some types of mutational changes, in particular TE insertions and premature

494   stop codons, being greater in the alleles with lower expression levels, although a causal

495   relationship remains to be demonstrated. Our results suggest an important relationship between

496   mutation accumulation and differential expression between alleles, which is relevant to a broad

497   range of taxa where reproductive compatibility is determined in extensive regions of

498   recombination suppression.

499

Materials and Methods

501   *Allele identification between a_1 and a_2 haploid genomes*

502   In order to quantify differentially expressed genes between the two haploid genomes, the

503   alleles between a_1 and a_2 haploid genomes need to be identified for those genes. The genome

504   assembly and annotation of the same strain of *M. lychnidis-dioicae* have been published (Branco

505   et al. 2017). To identify 1:1 single copy homologs in each haploid genome, the Reciprocal Best

506   BLAST(p) Hits (RBBH) python script

507   (github.com/peterjc/galaxy_blast/tree/master/tools/blast_rbh) was applied (Camacho et al. 2009),

508   with 50 percentage of length coverage. RBBH scripts also identified paralogs within each haploid

509   genome. A number of protein sequence alignment identity thresholds were tested, in order to

510   identify for the best strategy of maximizing the number of allele pair identification on the non-

511   recombining regions and while avoiding spurious BLAST results with low identity percent.

512   Increasing the percent of protein sequence identity threshold from >70% to >85% resulted in a

513   decrease from 12.2% to 9.9% of single-copy genes on the mating-type chromosomes being

514   identified as differentially expressed genes (detailed below), while decreasing the threshold from

515   >70% to >30% resulted in only a marginal increase from 12.2% to 12.7%. The change in the

516   percentages of identified alleles that were differentially expressed on autosomes was negligible,

517   being 1.0%, 1.1% and 1.1% respectively for 80%, 70% and 30% thresholds (Fig. S8). Therefore,

518   the threshold of >70% protein sequence identity was used. Additionally, we also have detected

519   orthologs between a_1 and a_2 genomes using OrthoMCL (Li et al. 2003), and the set of identified

520   single-copy orthologous genes and their alleles was almost identical, differing in only five out of

521   9396 genes compared to reciprocal best BLASTp analysis. To avoid potential bias due to

522   paralogs for identifying differential gene expression and other downstream analysis, genes with

523   paralogs within each haploid genome were filtered out and only single-copy allele pairs were

524   retained for downstream analysis. Genes were located to genomic compartments, including

525   autosomes, pseudo-autosomal regions (PARs), youngest evolutionary strata of the mating type

526   chromosomes (including previously identified red and green strata; Branco et al. 2017) and oldest

527   evolutionary strata (blue, purple, orange and black strata; Branco et al. 2017).

528

529   ***Transposable element filtering***

530   Transposable element (TE) annotation of both haploid genomes of *M. lychnidis-dioicae* was

531   published previously (Hartmann et al. 2018), and was used for analysis in this study. The coding

532   sequence of each gene from both a_1 and a_2 haploid genomes was search by BLAST(n) against the

533   published annotated TE consensus sequences of the same species, and alignment >80 percent of

534   query coverage (coding sequences) was used for identifications of TEs. The BLASTn output was

535   parsed using BASH scripts, and the coding sequences identified as TEs were removed from the

536   gene list for all further downstream analysis.

537

538   ***Identification of differentially expressed genes***

539   RNAseq data and summary statistics of the datasets were described previously (Perlin et al.

540   2015), and the raw data of haploid culture growing separately in water agar conditions were

541   downloaded from the deposited NCBI database

542   (https://trace.ncbi.nlm.nih.gov/Traces/study/?acc=+PRJNA246470&go=go). Briefly, haploid

543   sporidial strains of the original isolate (same as the reference genome "Lamole strain") were

544   generated from the meiotic products of a single tetrad. Then haploid fungal cells of either haploid

545   a_1 or a_2 strain grew separately on 2% water agar, each strain grew in two replicates, with nutrient

546   free environment without the mating partner for two days, which essentially mimicked the natural

547   conditions on the plant before mating and infection (Perlin et al. 2015). First, the RNAseq raw

548   reads were quality assessed using FastQC v0.11.2

549   (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/), and quality trimmed using

550   Trimmomatic v0.33 with default parameters for paired-end reads (Bolger et al. 2014). We filtered

551   reads containing adaptor sequences and trimmed reads if the sliding window average Phred score

552   over four bases was < 15 or if the leading/trailing bases had a Phred score < 3. Reads were then

553   removed post filtering if either read pair was < 36 bases.

554   To avoid possible bias for calling differential gene expression due to differences in homolog

555   length between a_1 and a_2, gaps differing between alleles by greater than 3bp were trimmed to

556   keep the same length, using published custom Python script (Parker 2016). This trimming

557   includes the gaps from the ends of the alignment and inside the alignment, with inside gaps

558   starting with the closest to the end of the alignment (greater than the minimum gap size) until

559   there are no gaps larger than minimum gap size (Parker 2016). The trimmed allele pairs with

560   equal length were used for read mapping and calling differential gene expression.

561   To quantify gene expression, we mapped the trimmed reads of haploid samples to the

562   trimmed homolog sequences of each haploid genome respectively with Kallisto v.0.43.0 (Bray et

563   al. 2016). Read counts of the output from Kallisto mapping (e.g. using pseudo-alignment) were

564   imported for gene expression analysis in EdgeR v3.4 (Robinson et al. 2010; McCarthy et al.

565   2012). We filtered low counts and kept genes with average Log(CPM) > 0 per sample, and CPM

566   (count per million) > 1 in half of the total samples per haploid culture. We then normalized the

567   expression using the weighted trimmed mean of M-values (TMM) implemented in EdgeR, which

568   is a scaling factor for library sizes that minimizes the log-fold change between samples. We

569   explored the libraries of both haploid cultures in pairwise correlation of raw counts between

570   replicates (Fig. S9), and two dimensions using multi-dimensional scaling (MDS) plots (Fig. S10).

571   Normalized expression counts for each sample were used to calculate differential expression

572   between mating types using standard measures. We first identified genes with differential

573   expression between mating types based on overall expression of the comparison group, and using

574   Benjamini-Hochberg correction for multiple-testing with false discovery rate (FDR) of 5%.

575   Differential expression between mating types was classified into four categories of fold changes,

576   namely 2 (low), 2-4 (mild), 4-8 (high), and > 8 (very high), and expressed as $\log_2$ ratio of a_1-to-a_2

577   expression (which has negative values for genes with higher a_2 expression and positive values for

578   higher a_1 expression). As suggested by (Montgomery and Mank 2016), fold changes > 0 will be

579   interpreted throughout, because we are working on haploid cell cultures and there are no possible

580   scaling nor allometry issues due to whole-body sampling. Thus, unless stated otherwise, both

581   conditions FDR < 0.05 and $|\log_2 FC| > 0$ will be met when calling mating-type bias. Finally, to

582   investigate the expression level of differentially (with a_1 or a_2 mating-type bias) and non-

583   differentially expressed genes, we compared normalized read counts (Transcripts Per Million,

584   Log2TPM, obtained from EdgeR v3.4) of significantly expressed genes at autosomal and mating-

585   type chromosomes (filtering criteria is the same as described above) from a_1 and a_2 samples (Fig.

586   S1).

587   The classification of genes as having differential expression between mating types or the

588   absolute values of gene expression ratio $|\text{Log2}(a_1/a_2)|$ was used to assess relationships to various

589   forms of mutational changes. Generalized linear model (GLM) analysis was used to assess the

590   predictors of absolute values of expression ratio $|\text{Log2}(a_1/a_2)|$, with main effect variables and all

591   two-way interactions terms for genomic compartments and the absolute value of differences

592   between alleles for sequence divergence (dN), transposable element insertions number within

593   20kb (up and downstream), predicted protein length, intron content and GC content. The absolute

594   value of the differences between alleles was calculated for each trait as detailed below. Model

595   family comparison was based upon minimizing Akaike's Information Criterion and over/under-

596   dispersion using ratio of deviance/df; Tweedie, power 1.7 (approaching gamma distribution)

597   provided the best available fit for the expression ratio response variable. A best fit model was

598   selected using stepwise model selection, following removal of non-significant interaction terms.

599   Other *post hoc* tests evaluating individual degeneration trait are described below. All statistical

600   analyses were conducted in SPSS v23 (IBM Corp 2015) and R v3.4.3 (R Core Team 2017).

601

602   ***Relationship between differential expression and elevated substitution rates***

603   Pairs of alleles between a_1 and a_2 mating types were aligned with PRANK (v170427) using

604   the codon model (Löytynoja and Goldman 2010). Each pair of allele alignment was then

605   analyzed with codeml or yn00 in PAML (Yang 2007) (runmode -2) to calculate the number of

606   nonsynonymous substitutions per nonsynonymous site (dN), the number of synonymous

607   substitutions per synonymous site (dS), and the ratio of the two (dN/dS), the latter excluding

608   genes with dS value of zero. We then compared sequence divergence between alleles using non-

609   parametric Wilcoxon rank sum tests for DE versus non-DE genes within genomic compartments.

610   Also, the allele sequences were compared between *M. lychnidis-dioicae* and their orthologs in

611   *M. lagerheimii*, which has retained largely collinear and recombining mating-type chromosomes,

612   as the inferred ancestral state in the *Microbotryum* genus (Branco et al. 2017; Branco et al. 2018).

613   The single-copy orthologs for a_1 or a_2 genomes between *M. lychnidis-dioicea* and *M. lagerheimii*

614   were identified using RBBH with 70 percent protein sequence coverage identity

615   (github.com/peterjc/galaxy_blast/tree/master/tools/blast_rbh, Camacho et al. 2009). Wilcoxon

616   rank sum test was used to assess dN between orthlogs in *M. lychnidis-dioicea* and *M. lagerheimii*

617 to evaluate the hypothesis that the alleles with lower expression levels would have greater

618 sequence divergence.

619

620 *Relationship between differential expression and TE insertions*

621   The TE annotation of the *M. lychnidis-dioicae* genome published previously (Hartmann et al.

622 2018) was used for the analysis in this study. First, the TE insertion sites were assessed for each

623 given focal gene, upstream 0-5k, 5-10kb, 10-15kb, 15-20kb distance intervals, and downstream

624 0-5kb, 5-10kb, 10-15kb and 15-20kb distance intervals using Bedtools window function for each

625 indicated distance window (https://bedtools.readthedocs.io/en/latest/content/tools/window.html).

626 Both annotation GFF3 files of gene models and TE annotations of *M. lychnidis-dioicae* were

627 provided as input files. The output files were parsed using Bash scripts. Wilcoxon rank sum tests

628 were used to compare TE insertions for DE and non-DE genes within genomic compartments.

629   Also, a limited GLM model was used to assess the hypothesized directional association of TE

630 insertions and reducing allele expression ($|Log2(a_1/a_2)|$); this model contained genomic

631 compartment and oriented TE differences between alleles as main effects and their interaction

632 term. Oriented TE differences between alleles were calculated as the TE number for the allele

633 with lower expression minus the TE number for the higher expressed allele; a positive value thus

634 represented an excess of TEs in the lower expressed allele. A sliding window approach was used

635 with a window size of three adjacent intervals, progressing from upstream to downstream of the

636 genes.

637

638 *Relationship between differential expression and differences in predicted protein length*

639   We first verified whether there was bias in the gene prediction model across genomic

640 compartments, using the ratio of predicted coding sequence length divided by three times protein

641 sequence length, assessed using linear regression model. Coding sequencing length divided by

642 the length of predicted protein multiplied by 3 was consistently close to 1 and did not differ

643　among genomic compartments (autosome, PAR, youngest strata and oldest evolutionary strata;

644　Linear model, R^2 = -5.50e-05, F-statistic = 0.869, P= 0.530, Fig. S11). We therefore calculated

645　the ratio of predicted protein length between allele pairs, and compared the proportions of genes

646　in DE and non-DE categories that had unequal lengths using two-proportion Z test for genes

647　within genomic compartments. The mutational causes of unequal protein lengths was assessed by

648　manually quantifying premature stop codons or indels using Geneious v8.1.7 (Kearse et al. 2012).

649　A limited GLM model was used to assess the hypothesized directional association of protein

650　truncation and reducing allele expression ($|Log2(a_1/a_2)|$); this model contained genomic

651　compartment and oriented predicted protein length differences between alleles as main effects

652　and their interaction term. Oriented predicted protein length differences between alleles were

653　calculated as the ratio for the allele with higher expression divided by the allele with lower

654　expression; a larger ratio thus represented a shorter length for the allele with lower expression.

655

656　*Relationship between differential expression and intron content*

657　　Using the published annotation gene models and coding sequences, we extracted the intron

658　number and mean intron length information from the annotation gff3 file, using Perl script

659　(https://bioops.info/2012/11/intron-size-gff3-perl/). We investigated the proportional differences

660　of the intron content for both DE and non-DE genes within genomic compartments using

661　Wilcoxon rank sum test. We also used a limited GLM model to test the hypothesized directional

662　association of greater intron content and reducing allele expression ($|Log2(a_1/a_2)|$); this model

663　contained genomic compartments and oriented intron content differences between alleles as main

664　effects and their interaction term. Oriented intron content differences between alleles were

665　calculated as the value for the lower expressed allele minus the value for the higher expressed

666　allele; a positive value thus represented greater intron content for the lower expressed allele.

667

668　*Relationship between differential expression and GC content*

669   We calculated the total GC percentage (GC0) and the GC percentage at the third position of

670   aminol acid (GC3) for alleles of each gene coding sequence using homemade awk scripts. We

671   investigated the differences of GC0 and GC3 for both DE and non-DE genes within genomic

672   compartments using Wilcoxon rank sum test. We also used a limited GLM model to test the

673   hypothesized directional association of reduced GC content and reducing allele expression

674   ($|Log2(a_1/a_2)|$); this model contained genomic compartments and oriented GC content differences

675   between alleles as main effects and their interaction term. Oriented GC content differences

676   between alleles were calculated as the value for the allele with higher expression minus the value

677   for the allele with lower expression; a positive value thus representing reduced GC content for the

678   allele with lower expression.

679

680   **Ethics Statement**

681

682   N/A

Data availability

We used published gene expression data to investigate the association of sequence degeneration and differential gene expression in *Microbotryum lychnidis-dioicae* (Fontanillas et al. 2015; Perlin et al. 2015, https://trace.ncbi.nlm.nih.gov/Traces/study/?acc=+PRJNA246470&go=go). We used published genome assembly, gene predictions and assignations to genomic compartments (Branco et al. 2017; Branco et al. 2018). We also used published transposable elements identification in *M. lychnidis-dioicae* (Hartmann et al. 2018). All relevant scripts and data files to perform these analyses are deposited in Zenodo and Github (https://github.com/Wen-Juan/Differential_expression_associateswith_degeneration_Microbotryum_fungus), which will be released immediately upon manuscript acceptance.

References

Abbate JL, Hood ME. 2010. Dynamic linkage relationships to the mating-type locus in automictic fungi of the genus *Microbotryum*. J. Evol. Biol. 23:1800–1805.

Ayroles JF, Carbone MA, Stone EA, Jordan KW, Lyman RF, Magwire MM, Rollmann SM, Duncan LH, Lawrence F, Anholt RRH, et al. 2009. Systems genetics of complex traits in *Drosophila melanogaster*. Nat. Genet. 41:299–307.

Bachtrog D. 2005. Sex chromosome evolution: molecular aspects of Y-chromosome degeneration in *Drosophila*. Genome Res. 15:1393–1401.

714 Badouin H, Gladieux P, Gouzy J, Siguenza S, Aguileta G, Snirc A, Le Prieur S, Jeziorski C,
715   Branca A, Giraud T. 2017. Widespread selective sweeps throughout the genome of model
716   plant pathogenic fungi and identification of effector candidates. Mol. Ecol. 26:2041–2062.

717 Bazzicalupo AL, Carpentier F, Otto SP, Giraud T. 2019. Little evidence of antagonistic selection
718   in the evolutionary strata of fungal mating-type chromosomes (*Microbotryum lychnidis-*
719   *dioicae*). Genes|Genomes|Genetics 9:1987-1998.

720 Bedford T, Hartl DL. 2009. Optimization of gene expression by natural selection. Proc. Natl.
721   Acad. Sci. U. S. A. 106:1133–1138.

722 Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: A flexible trimmer for Illumina sequence
723   data. Bioinformatics 30:2114–2120.

724 Branco S, Badouin H, Rodríguez RC, Vega D, Gouzy J, Carpentier F. 2017. Evolutionary strata
725   on young mating-type chromosomes despite the lack of sexual antagonism. Proc. Natl.
726   Acad. Sci. U. S. A. 114:7367–7072.

727 Branco S, Carpentier F, Rodríguez RC, Vega D, Badouin H, Snirc A, Prieur S Le, Coelho MA,
728   Vienne DM De, Hartmann FE, et al. 2018. Multiple convergent supergene evolution events
729   in mating-type chromosomes. Nat. Commun. 9:2000.

730 Bray N, Pimentel H, Melsted P, Pachter L. 2016. Near-optimal RNA-Seq quantification. Nat.
731   Biotechnol. 34:525–528.

732 Britten RJ, Davidson EH. 1971. Repetitive and non-repetitive DNA sequences and a speculation
733   on the origins of evolutionary novelty. Q. Rev. Biol. 46:111–138.

734 Bull JJ. 1978. Sex chromosomes in haploid dioecy: a unique contrast to Muller's theory for
735   dipoid dioecy. Am. Nat. 112:245–250.

736 Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009.
737   BLAST+: Architecture and applications. BMC Bioinformatics 10:1–9.

738 Charlesworth B. 2012. The effects of deleterious mutations on evolution at linked sites. Genetics
739   190:5–22.

740 Charlesworth B, Jordan CY, Charlesworth D. 2014. The evolutionary dynamics of sexually
741   antagonistic mutations in pseudoautosomal regions of sex chromosomes. Evolution
742   68:1339–1350.

743 Charlesworth D, Charlesworth B, Marais G. 2005. Steps in the evolution of heteromorphic sex
744   chromosomes. Heredity 95:118–128.

745 Chen JM, Cooper DN, Chuzhanova N, Férec C, Patrinos GP. 2007. Gene conversion:
746   mechanisms, evolution and human disease. Nat. Rev. Genet. 8:762–775.

747 Cordaux R, Batzer MA. 2009. The impact of retrotransposons on human genome evolution. Nat.

748   Rev. Genet. 10:691–703.

749   Cox RM, Calsbeek R. 2009. Sexually antagonistic selection, sexual dimorphism, and the

750   resolution of intralocus sexual conflict. Am. Nat. 173:176–187.

751   Darolti I, Wright AE, Pucholt P, Berlin S, Mank JE. 2018. Slow evolution of sex-biased genes in

752   the reproductive tissue of the dioecious plant *Salix viminalis*. Mol. Ecol. 27:694–708.

753   Day AW. 1979. Mating type and morphogenesis in *Ustilago violacea*. Bot. Gaz. 140:94–101.

754   Dean R, Mank JE. 2014. The role of sex chromosomes in sexual dimorphism: discordance

755   between molecular and phenotypic data. J. Evol. Biol. 27:1443–1453.

756   Van Doorn GS, Kirkpatrick M. 2007. Turnover of sex chromosomes induced by sexual conflict.

757   Nature 449:909–912.

758   Eichten SR, Schmitz RJ, Springer NM. 2014. Epigenetics: beyond chromatin modifications and

759   complex genetic regulation. Plant Physiol. 165:933–947.

760   Ellegren H, Parsch J. 2007. The evolution of sex-biased genes and sex-biased gene expression.

761   Nat. Rev. Genet. 8:689.

762   Feschotte C. 2008. The contribution of transposable elements ot the evolution of regulatory

763   networks. Nat. Rev. Genet. 9:397–405.

764   Fontanillas E, Hood ME, Badouin H, Petit E, Barbe V, Gouzy J, De Vienne DM, Aguileta G,

765   Poulain J, Wincker P, et al. 2015. Degeneration of the nonrecombining regions in the

766   mating-type chromosomes of the anther-smut fungi. Mol. Biol. Evol. 32:928–943.

767   Fraser JA, Heitman J. 2004. MicroReview Evolution of fungal sex chromosomes. Mol. Microbiol.
768   51:299–306.
769   Garber ED, Day AW. 1985. Genetic mapping of a phytopathogenic Basidiomycete, *Ustilago*

770   *violacea*. Bot. Gaz. 146:449–459.

771   Giraud T, Yockteng R, Lo M. 2008. Mating system of the anther smut fungus *Microbotryum*

772   *violaceum*: selfing under heterothallism. Eukaryot. Cell 7:765–775.

773   Grath S, Parsch J. 2016. Sex-biased gene expression. Annu. Rev. Genet. 50:29–44.

774   Graves JAM. 2010. Review: sex chromosome evolution and the expression of sex-specific genes

775   in the placenta. Placenta 31:S27–S32.

776   Grummt I, Pikaard CS. 2003. Epigenetic silencing of RNA polymerase I transcription. Nat. Rev.

777   Mol. Cell Biol. 4:641–649.

778   Hartmann FE, De La Vega RCR, Brandenburg JT, Carpentier F, Giraud T. 2018. Gene presence-

779   absence polymorphism in castrating anther-smut fungi: recent gene gains and

780   phylogeographic structure. Genome Biol. Evol. 10:1298–1314.

781   Haselman JT, Olmstead AW, Degitz SJ. 2015. General and comparative endocrinology global

782   gene expression during early differentiation of *Xenopus (Silurana) tropicalis* gonad tissues.

783   Gen. Comp. Endocrinol. 214:103–113.

784 Heyn P, Kalinka AT, Tomancak P, Neugebauer KM. 2015. Introns and gene expression: cellular

785   constraints, transcriptional regulation, and evolutionary consequences. BioEssays 37:148–

786   154.

787 Hood M, Antonovics J. 2000. Intratetrad mating, heterozygosity, and the maintenance of

788   deleterious alleles. Heredity 85:231–241.

789 Hood ME. 2002. Dimorphic mating-type chromosomes in the fungus *Microbotryum violaceum*.

790   Genetics 160:457–461.

791 Hood ME, Antonovics J. 2004. Mating within the meiotic tetrad and the maintenance of genomic.

792   Genetics 166:1751–1759.

793 Hood ME, Antonovics J, Koskella B. 2004. Shared forces of sex chromosome evolution in

794   haploid-mating and diploid-mating organisms: *Microbotryum violaceum* and other model

795   organisms. Genetics 168:141–146.

796 Hood ME, Petit E, Giraud T. 2013. Extensive divergence between mating-type chromosomes of

797   the anther-smut fungus. Genetics 193:309–315.

798 Hoof A van, Green PJ. 1996. Premature nonsense codons decrease the stability of

799   phytohemagglutinin mRNA in a position-dependent manner. Plant J. 10:415–424.

800 IBM Corp. 2015. IBM SPSS Statistics for Macintosh, Version 23.0. Version 23:Armonk, NY:

801   IBM Corp.

802 Innocenti P, Morrow EH. 2010. The sexually antagonistic genes of *Drosophila melanogaster*.

803   PLoS Biol. 8:e1000335.

804 Kearse M, Moir R, Wilson A, Stones-havas S, Sturrock S, Buxton S, Cooper A, Markowitz S,

805   Duran C, Thierer T, et al. 2012. Geneious basic: an integrated and extendable desktop

806   software platform for the organization and analysis of sequence data. Bioinforma. Appl.

807   Note 28:1647–1649.

808 Kimball SR, Jefferson LS. 2004. Amino acids as regulators of gene expression. Nutr. Metab. 1:3.

809 King GJ. 2015. Crop epigenetics and the molecular hardware of genotype × environment

810   interactions. Front. Plant Sci. 6:1–19.

811 Kitano J, Peichel CL. 2012. Turnover of sex chromosomes and speciation in fishes. Environ.

812   Biol. Fishes 94:549–558.

813 Lande R. 1980. Sexual dimorphism, sexual selection, and adaptation in polygenic characters.

814   Evolution 34:292–305.

815 Lee TI, Young RA. 2013. Transcriptional regulation and its misregulation in disease. Cell

816   152:1237–1251.

817　Li L, Stoeckert Jr CJ, Roos DS. 2003. OrthoMCL : Identification of Ortholog Groups for

818　　　　Eukaryotic Genomes. Genome Res. 13:2178–2189.

819　Lindholm A, Breden F. 2002. Sex chromosomes and sexual selection in poeciliid fishes. Am.

820　　　　Nat. 160:S214–S224.

821　Lipinska AP, Toda NRT, Heesch S, Peters AF, Cock JM, Coelho SM. 2017. Multiple gene

822　　　　movements into and out of haploid sex chromosomes. Genome Biol. 18:104.

823　Löytynoja A, Goldman N. 2010. webPRANK: a phylogeny-aware multiple sequence aligner with

824　　　　interactive alignment browser. BMC Bioinformatics 11:579.

825　Ma W-J, Veltsos P, Sermier R, Parker DJ, Perrin N. 2018. Evolutionary and developmental

826　　　　dynamics of sex-biased gene expression in common frogs with proto-Y chromosomes.

827　　　　Genome Biol. 19:156.

828　Ma W-J, Veltsos P, Toups MA, Rodrigues N, Sermier R, Jeffries DL, Perrin N. 2018. Tissue

829　　　　specificity and dynamics of sex-biased gene expression in a common frog population with

830　　　　differentiated, yet homomorphic, sex chromosomes. Genes 9:294.

831　Mank JE. 2013. Sex chromosome dosage compensation: Definitely not for everyone. Trends

832　　　　Genet. 29:677–683.

833　Mank JE. 2017. Population genetics of sexual conflict in the genomic era. Nat. Rev. Genet.

834　　　　18:721-730.

835　Marais G, Nouvellet P, Keightley PD, Charlesworth B. 2005. Intron size and exon evolution in

836　　　　*Drosophila*. Genetics 170:481–485.

837　Mather K. 1942. Heterothally as an outbreeding mechanism in fungi. Nature 149:54–56.

838　McCarthy DJ, Chen Y, Smyth GK. 2012. Differential expression analysis of multifactor RNA-

839　　　　Seq experiments with respect to biological variation. Nucleic Acids Res. 40:4288–4297.

840　Mcclintock B. 1942. Maize genetics. In: Carnegie institution of Washington year book. Vol. 41.

841　　　　p. 181–186.

842　Meisel RP, Ph D. 2017. The house fly Y chromosome is young and undifferentiated from its

843　　　　ancient X chromosome partner. Genome Res. 27:1–10.

844　Montgomery SB, Goode DL, Kvikstad E, Albers CA, Zhang ZD, Mu XJ, Ananda G, Howie B,

845　　　　Karczewski KJ, Smith KS, et al. 2013. The origin, evolution, and functional impact of short

846　　　　insertion-deletion variants identified in 179 human genomes. Genome Res. 23:749–761.

847　Montgomery SH, Mank JE. 2016. Inferring regulatory change from gene expression: the

848　　　　confounding effects of tissue scaling. Mol. Ecol. 25:5114–5128.

849　Mugal CF, Arndt PF, Holm L, Ellegren H. 2015. Evolutionary Consequences of DNA

850　　　　Methylation on the GC Content in Vertebrate Genomes. Genes|Genomes|Genetics 5:441–

851   447.

852   Otto SP, Pannell JR, Peichel CL, Ashman TL, Charlesworth D, Chippindale AK, Delph LF,

853   Guerrero RF, Scarpino S V., McAllister BF. 2011. About PAR: The distinct evolutionary

854   dynamics of the pseudoautosomal region. Trends Genet. 27:358–367.

855   P.Bird A. 1980. DNA methylation and the frequency of CpG in animal DNA. Nucleic Acids Res.

856   8:1499–1504.

857   Parker DJ. 2016. Fastatools.pdf. Zenodo:https://zenodo.org/record/59775.

858   Parsch J, Ellegren H. 2013. The evolutionary causes and consequences of sex-biased gene

859   expression. Nat. Rev. Genet.14:83–87.

860   Perlin MH, Amselem J, Fontanillas E, Toh SS, Chen Z, Goldberg J, Duplessis S, Henrissat B,

861   Young S, Zeng Q, et al. 2015. Sex and parasites: genomic and transcriptomic analysis of

862   Microbotryum lychnidis-dioicae, the biotrophic and plant-castrating anther smut fungus.

863   BMC Genomics 16:461.

864   Perry JC, Harrison PW, Mank JE. 2014. The ontogeny and evolution of sex-biased gene

865   expression in *Drosophila melanogaster*. Mol. Biol. Evol. 31:1206–1219.

866   R Core Team. 2017. R: a language and environment for statistical computing. Vienna, Austria: R

867   Foundation for Statistical Computing

868   Rice WR. 1987. The accumulation of sexually antagonistic genes as a selective agent promoting

869   the evolution of reduced recombination between primitive sex chromosomes. Evolution

870   41:911–914.

871   Rice WR. 1996. Evolution of the Y Sex chromosome in animals. Bioscience 46:331–343.

872   Robinson MD, Mccarthy DJ, Smyth GK. 2010. edgeR: a Bioconductor package for

873   differential expression analysis of digital gene expression data. Appl. Note 26:139–140.

874   Samils N, Gioti A, Karlsson M, Sun Y, Kasuga T, Bastiaans E, Wang Z, Li N, Townsend JP,

875   Johannesson H. 2013. Sex-linked transcriptional divergence in the hermaphrodite fungus

876   *Neurospora tetrasperma*. Proc. R. Soc. B Biol. Sci. 280:20130862.

877   Schäfer AM, Kemler M, Bauer R, Begerow D. 2010. The illustrated life cycle of *Microbotryum*

878   on the host plant Silene latifolia. Botany 88:875–885.

879   Sigeman H, Ponnikas S, Videvall E, Zhang H, Chauhan P, Naurin S, Hansson B. 2018. Insights

880   into avian incomplete dosage compensation: Sex-biased gene expression coevolves with sex

881   chromosome degeneration in the common whitethroat. Genes 9:373.

882   Tirosh I, Barkai N, Verstrepen KJ. 2009. Promoter architecture and the evolvability of gene

883   expression. J. Biol. 8:1–6.

884   Wray GA, Hahn MW, Abouheif E, Balhoff JP, Pizer M, Rockman M V., Romano LA. 2003. The

885   evolution of transcriptional regulation in eukaryotes. Mol. Biol. Evol. 20:1377–1419.

886   Wright AE, Darolti I, Bloch NI, Oostra V, Sandkam B, Buechel SD, Kolm N, Breden F, Vicoso

887   B, Mank JE. 2017. Convergent recombination suppression suggests role of sexual selection

888   in guppy sex chromosome formation. Nat. Commun. 8:14251.

889   Wright AE, Dean R, Zimmer F, Mank JE. 2016. How to make a sex chromosome. Nat. Commun.

890   7:12087.

891   Yang X, Schadt EE, Wang S, Wang H, Arnold AP, Ingram-drake L, Drake TA, Lusis AJ. 2006.

892   Tissue-specific expression and regulation of sexually dimorphic genes in mice. Genome

893   Res. 16:995–1004.

894   Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. Mol. Biol. Evol.

895   24:1586–1591.

896

897

898

899

900

901

902

903

904

905

906

907

908

909

910

911

912

913

914

915

916

917

918

919 **Table 1.** Output of a reduced best-fit generalized linear model (GLM) with differential gene

920 expression ($|Log2(a_1/a_2)|$) as the response variable and as predictable variables genomic

921 compartment and various degeneration traits, i.e., non-synonymous mutation rate (dN),

922 transposable element (TE) insertions, protein length, intron content and GC content. P values

923 <0.05 are in bold. NA: not applicable.

924

Predictable variables and interaction terms	GLM model output parameter			
	Wald Chi-Square	**Degree of freedom (df)**	**P value**	**Regression coefficient**
(Intercept)	496.78	1	**<0.001**	NA
Compartment	20.151	3	**<0.001**	NA
dN	13.21	1	**<0.001**	5.081
TE insertions	8.405	1	**0.004**	0.044
Protein length	0.41	1	0.522	10.612
Intron content	10.209	1	**0.001**	0.768
GC content	4.233	1	**0.040**	0.499
Compartment * Protein length	24.662	3	**<0.001**	NA
dN * Protein length	13.36	1	**<0.001**	-50.726
TE insertions * Protein Length	8.398	1	**0.004**	-0.37
GC content * Protein length	10.801	1	**0.001**	-3.962

925

926

927

928

929

930

931

932

933

934

935

936

Table 2. Numbers and percentages of genes with differential expression (DE) on the different genomic compartments on mating-type chromosomes and autosomes, and Fisher's exact test for even distribution between DE genes on autosomes versus other genomic compartments, including pseudo-autosomal regions (PARs), youngest evolutionary strata (previously identified red and green strata; Branco et al. 2017) and oldest evolutionary strata (blue, purple, orange and black strata; Branco et al. 2017). *P* values <0.05 are in bold. NA: not applicable.

	Autosomes	**PAR**	**Youngest strata**	**Oldest Strata**
DE gene number	507	12	1	74
Total number	8207	114	29	198
Percentage	6.18%	10.53%	3.45%	37.37%
Fisher's exact test (*P* value)	NA	0.085	1	**2.20E-16**

Figure Legends

968 **Fig. 1. Heatmap and hierarchical clustering of differentially expressed genes (false**

969 **discovery rate, FDR < 0.05) between haploid a_1 and a_2 cultures of *Microbotryum lychnidis-***

970 ***dioicae* under low nutrient condition.**

971 Each column shows a replicated sample for each haploid cell culture. Z-score denotes the relative

972 gene expression level, i.e. blue and red represent high and low expression, respectively. On each

973 node of the clustering tree, bootstrap support values are shown based on 10,000 replicates.

974

975 **Fig. 2. Interaction plots for pairs of predictor variables in overall GLM of differential gene**

976 **expression between mating types of *Microbotryum lychnidis-dioicae*.**

977 Y-axes are GLM-predicted response values of differential expression ratio between allele pairs in

978 a_1 and a_2 haploid genomes, and x-axes are allele differences between allele pairs in a_1 and a_2

979 haploid genomes in predicted protein length as the predictor variable, then binned into levels of

980 interacting categorical predictor variable (i.e. panel **A**, genomic compartment) or other interacting

981 continuous predictor variables (i.e. panels **B-D**; the lowest bin being no differences between

982 alleles, and low and high bins being split at the median value among genes with non-zero

983 differences between alleles, respectively). (**A**) Interaction plot between protein length differences

984 and genomic compartment. Genomic compartments include autosomes, pseudo-autosomal

985 regions (PARs), youngest and oldest evolutionary strata. (**B**) Interaction plots between protein

986 length differences and differences in transposable elements (TEs) insertions. (**C**) Interaction plots

987 between protein length differences and non-synonymous mutation (*dN*) rate differences. (**D**)

988 Interaction plots between protein length differences and GC content differences.

989

990 **Fig. 3. Comparisons of differentially expressed (DE) versus non-differentially expressed**

991 **(non-DE) genes between mating types of *Microbotryum lychnidis-dioicae* for various**

992 **degeneration-associated traits within genomic compartments.**

993 (**A**) Non-synonymous sequence divergence, *dN*, between alleles of DE and non-DE genes. (**B**)

994 Transposable element (TE) insertion number differences between alleles within 20kb (up and

995 downstream) of DE and non-DE genes. (**C**) Proportions of differentially expressed (DE) and non-

996 differentially expressed (non-DE) genes with different protein lengths between alleles. (**D**) Intron

997 content proportional differences between alleles of DE and non-DE genes. (**E**) Total GC content

998 (GC0) proportional differences between alleles of DE and non-DE genes. Analyzed allele

999 differences represent absolute values comparisons (i.e. unoriented with regard to allele expression

1000 levels). Comparisons in panels A, C-E reflect Wilcoxon rank sum tests; panel B reflects a two-

proportion z-test. Significance levels shown as, ***: $P < 0.001$, *: $P < 0.05$; non-significant test results shown in Supplementary Tables S4, S6-S9. Genomic compartments include autosomes, pseudo-autosomal regions (PARs), youngest and oldest evolutionary strata. The notation "*a*" indicates that the youngest evolutionary strata contained only one DE gene, precluding comparisons to non-DE genes within this compartment.

Fig. 4. Significant predictors of the degree of differential expression between mating types of *Microbotryum lychnidis-dioicae* testing directional effects of degeneration-associated traits.

(**A**) Relationship of expression ratio to oriented TE insertion differences in the region from 10kb upstream to the gene, where the trait was calculated as TE number for the allele with lower expression minus the TE number for the higher expressed allele; a positive value thus represented an excess of TEs in the lower expressed allele. (**B**) Relationship of expression ratio to oriented predicted protein length differences, where the trait was calculated as the ratio for the allele with higher expression divided by the allele with lower expression; a larger ratio thus represented a shorter length for the allele with lower expression.

Fig. 5. Average indel numbers and proportions of genes with different stop codon positions between alleles of differentially expressed genes of *Microbotryum lychnidis-dioicae*.

Among genes having alleles with different predicted protein lengths, boxplot of average indel numbers for both differentially expressed (DE in black) and non-DE genes (in grey) across various genomic compartments (**A**), and barplot for proportions of genes with different stop codon positions for both DE and non-DE genes across genomic compartments (**B**). **: P < 0.01, *: P < 0.05, '.': P < 0.1, NS: not significant. Genomic compartments include autosomes, pseudo-autosomal regions (PARs), youngest and oldest evolutionary strata.

Fig 1

Row Z−Score

Density

clusters
2
1

100 100   100 100

a₂ Sample1   a₂ Sample2   a₁ Sample1   a₁ Sample2

Fig 2

Fig 3

Fig 4

A

B
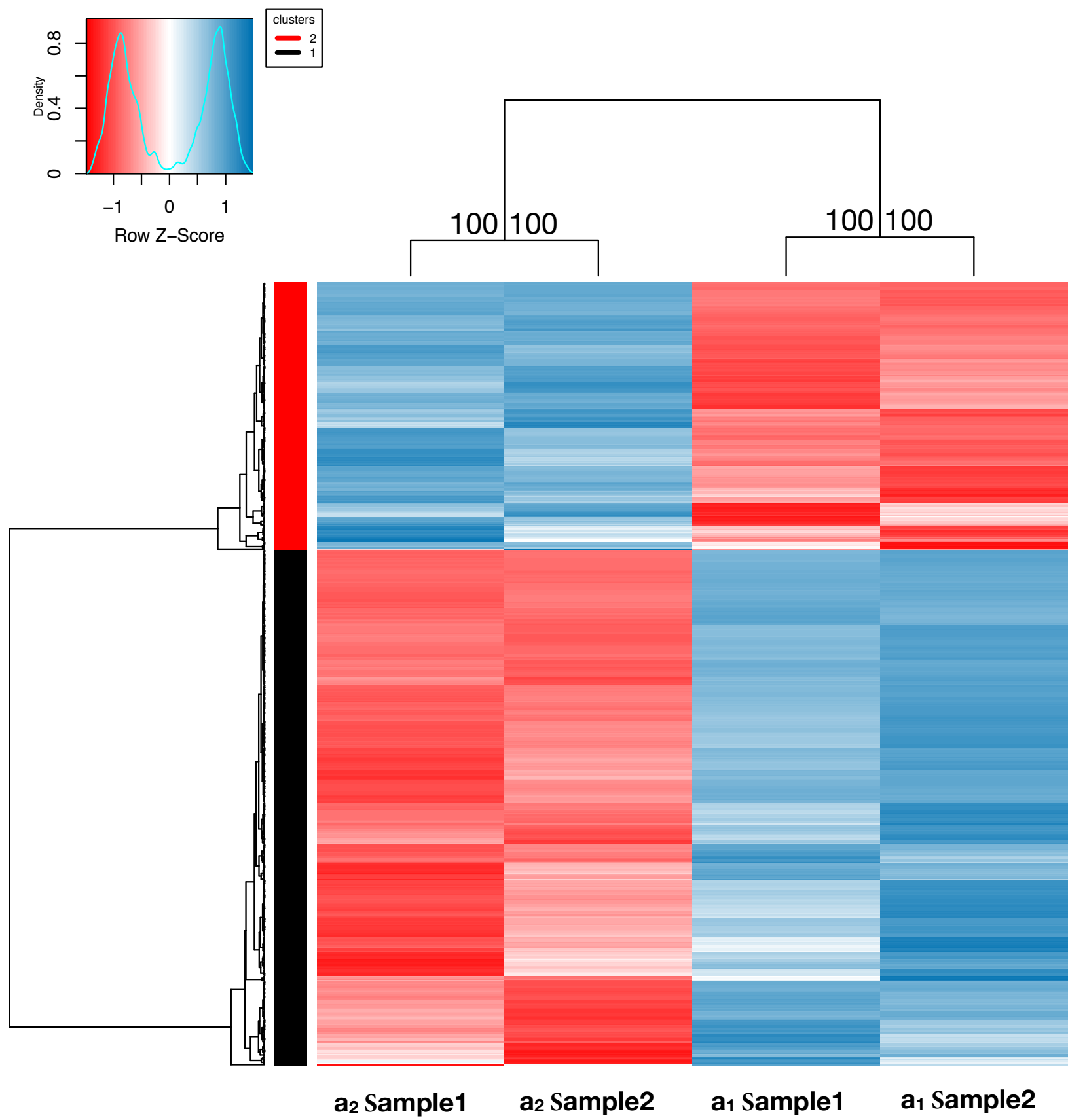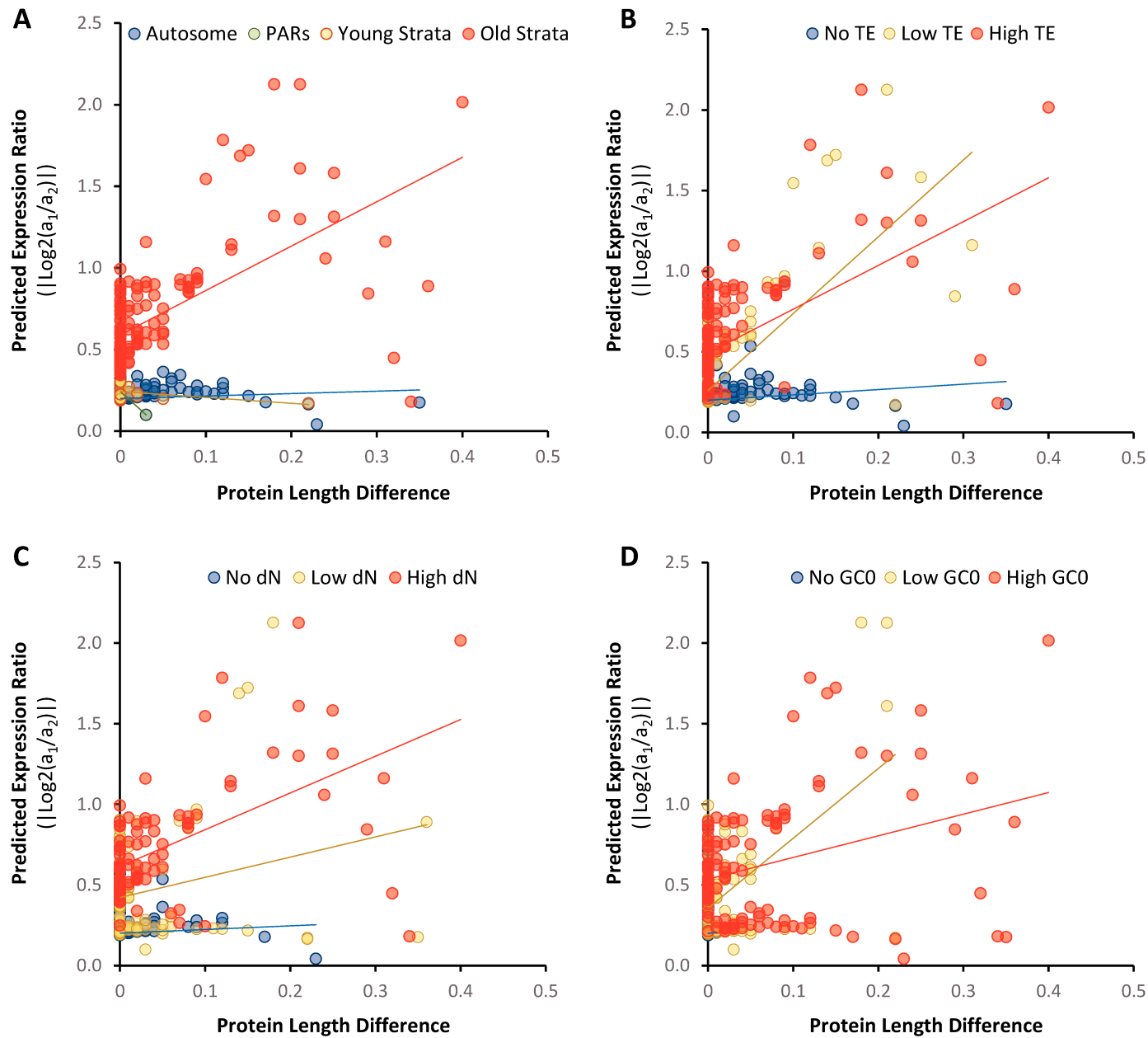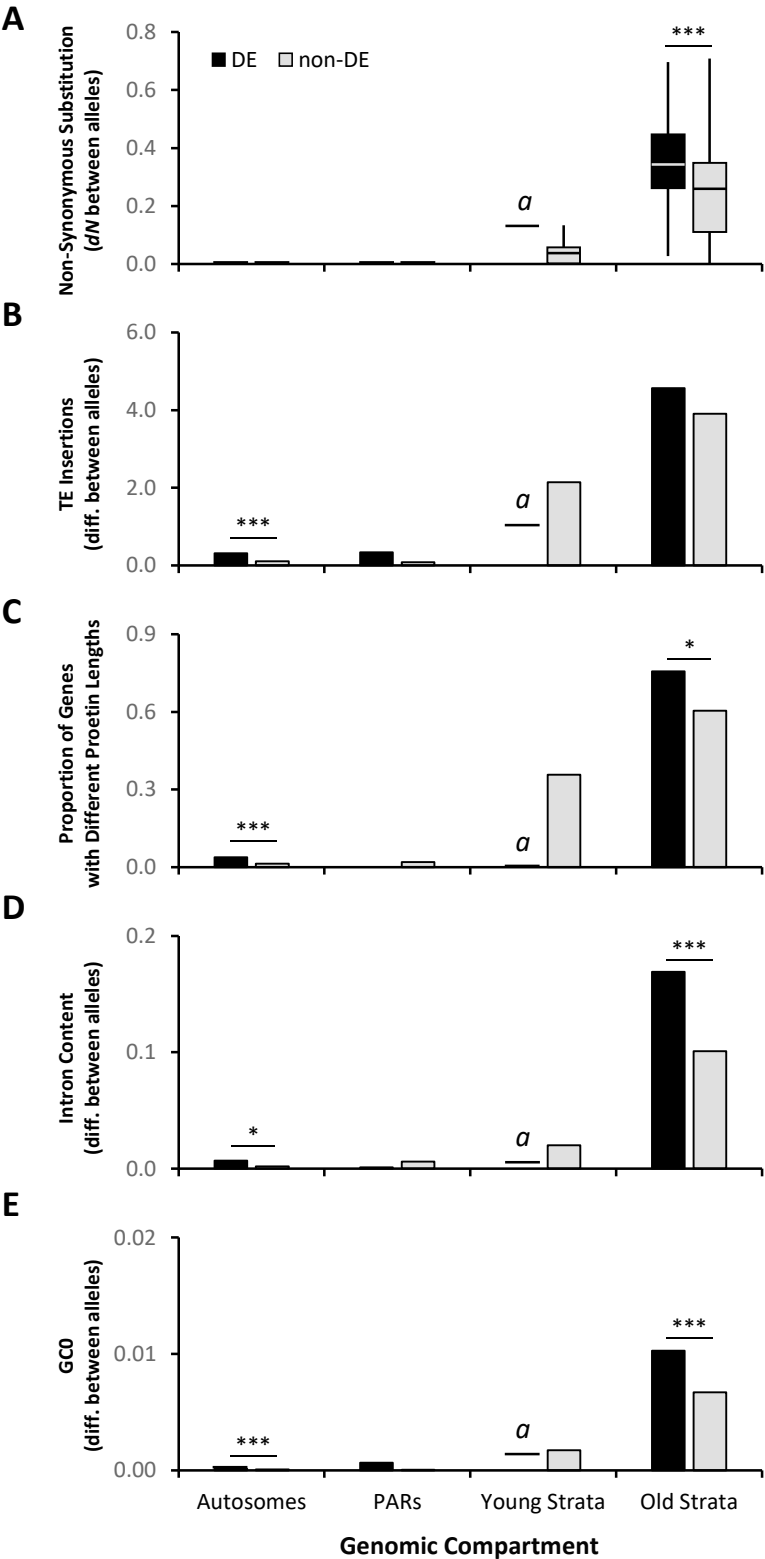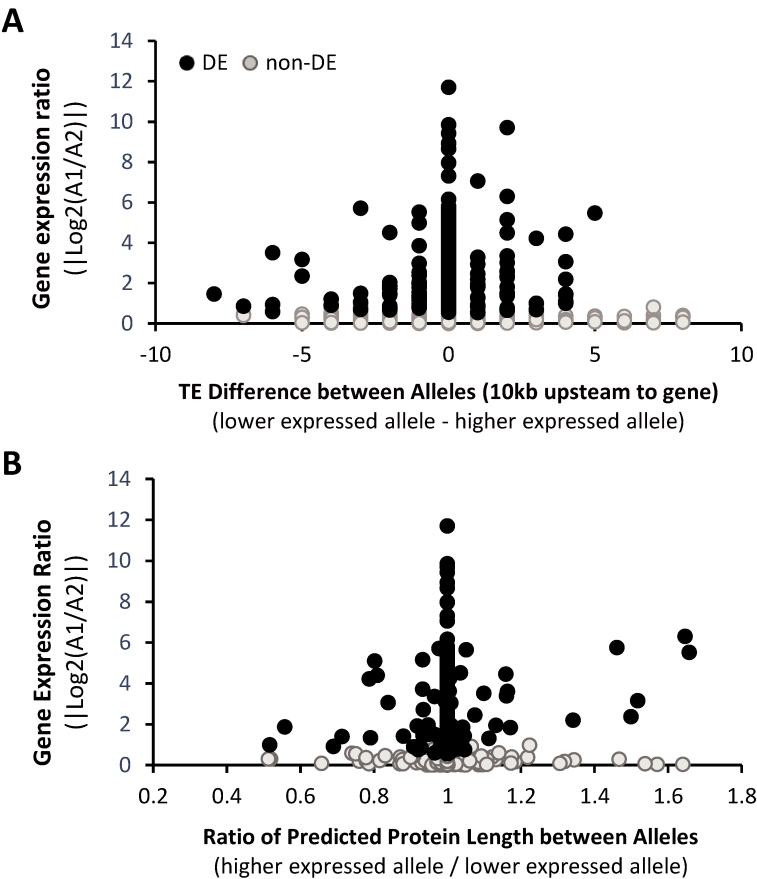
Fig 5
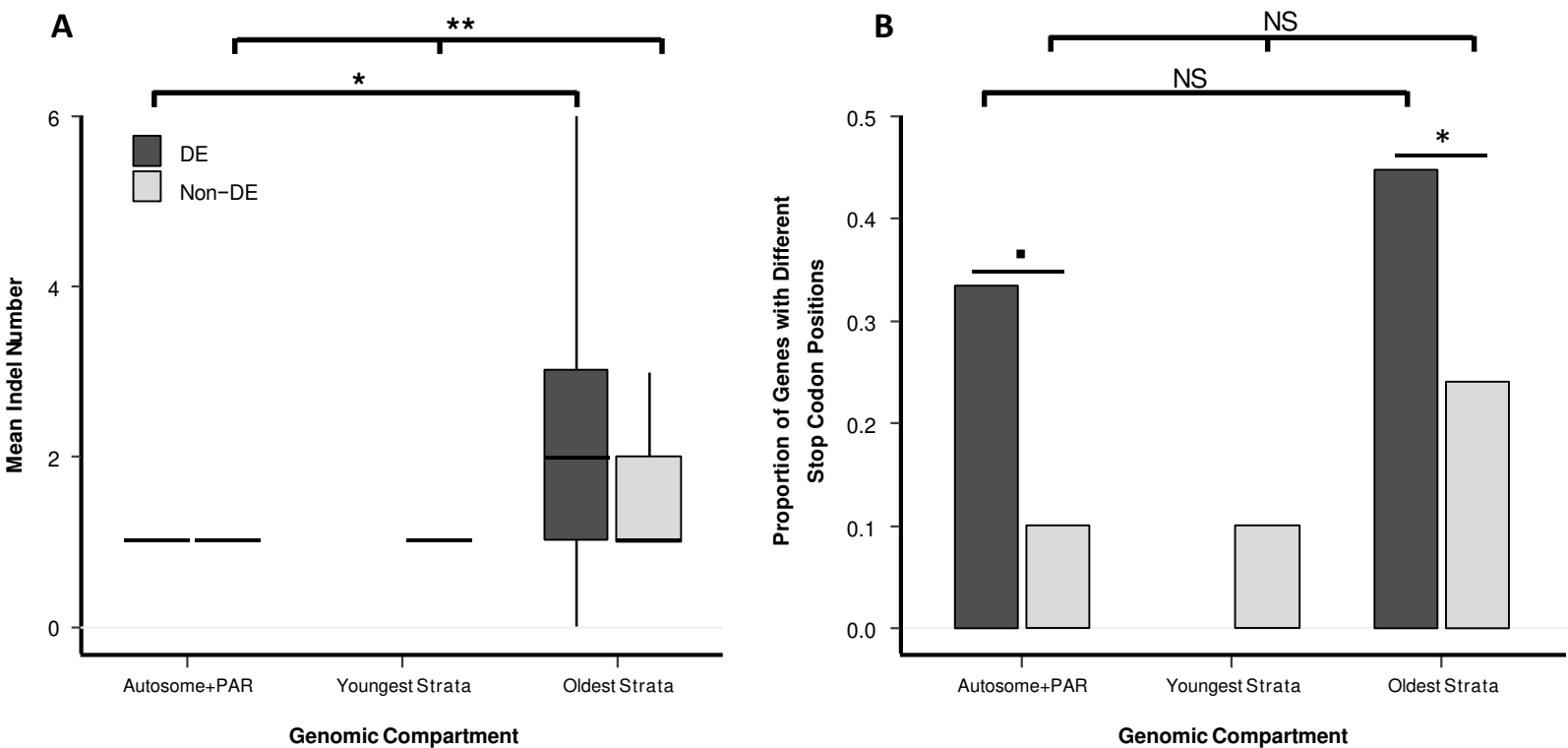
7 General discussion

In this thesis, I studied the evolution of recombination suppression on mating-type chromosomes in multiple fungal species of the *Microbotryum* genus. We reported the existence of evolutionary strata on fungal mating-type chromosomes and provided evidence that they formed without involving antagonistic selection between compatible sexes or mating-types. These findings challenge the long-standing view that evolutionary strata formation on sex chromosome is triggered by sexual antagonism and call for the consideration of other mechanisms. Recent studies on the fish *Poecilia reticulata*, which was yet *a priori* an excellent model for supporting the role of sexual antagonism in sex chromosome evolution, could not actually disentangle causes from consequences (Bergero et al. 2019). Several alternatives to the sexual antagonism hypothesis have been proposed, involving neutral rearrangements, heterozygote advantage or spread of transposable elements and their silencing marks (Ironside 2010; Ponnikas et al. 2018; Kent et al. 2017). Moreover, if sexual antagonism is not required for the formation of evolutionary strata, non-recombining regions may extend progressively also in autosomes. For instance, three polymorphic inversions have been reported at the supergene involved in the mimicry in *Heliconius* butterflies and evolved successively, being thus likely evolutionary strata, although the authors never use this term (Jay et al. 2019, 2018; Joron et al. 2011). A recent study on the fire ant social chromosome supergene did not detect evolutionary strata (Pracana et al. 2017); nevertheless, three different inversions have been detected in this supergene, although sequentially close in time (Wang et al. 2013).

We reported multiple convergent recombination suppression events across the *Microbotryum* genus. In two species with unlinked mating-type chromosome, recombination suppression linked each mating-type locus to their respective centromere. Across the other studied species, we published that recombination suppression evolved independently at least five times to link mating-type loci after they were brought through distinct genomic rearrangements onto the same chromosome, which constitutes a striking instance of evolutionary convergence. Gould viewed evolution as "unrepeatable" and "unpredictable", and argued that, owing to stochastic processes, the same outcome "would probably never rise again even if life's tape could be replayed a thousand time" (Gould 1990). The evolutionary convergence we reported in *Microbotryum* beautifully contradicts this view of evolution, because the transition from tetrapolarity to bipolarity occurred repeatedly through distinct

rearrangements. It is likely that chromosomal fusions regularly arise stochastically in *Microbotryum* without being selected for, and that those bringing the mating-type loci onto the same chromosome have been selected for.

Recently, I characterized the mating-type chromosomes of three additional bipolar species, *M. v. tatarinowii*, *M. scopulorum* and *M. coronariae* (Figure 4). I found that the rearrangement bringing the mating-type loci onto the same chromosome in these three species was similar as the one in *M. lychnidis-dioicae*. Given the placement of *M. v. tatarinowii* in the species tree (Figure 4), the recombination suppression event linking its mating-type loci should be independent from the other recombination suppression events. Given the placement of *M. scopulorum* and *M. coronariae* in the species tree (Figure 4), the recombination suppression event linking their mating-type loci could have occurred in the common ancestor of *M. scopulorum*, *M. coronariae*, *M. violaceum s.s.*, *M. silenes-dioicae* and *M. lychnidis-dioicae*. A more comprehensive analysis of the trans-specific polymorphism in gene genealogies encompassing these five species is needed to elucidate whether the corresponding recombination suppression event arose only once in the common ancestor of these four species.

Natural selection equally favours bipolarity and tetrapolarity with each mating-type locus linked to the centromere of distinct chromosomes under high intra-tetrad selfing rates as both yields the same odds of gamete compatibility under intra-tetrad mating. However, *Microbotryum* fungi can also undergo selfing through inter-tetrad mating, under which bipolarity gives a higher rate of gamete compatibility than linkage to centromeres. Indeed, bipolarity yields only two mating types within a progeny while linkage of mating-type loci to centromeres yields four mating types across different meiosis of a given diploid individual. The rate of intra-tetrad mating has been estimated at about 57% and the rate of self inter-tetrad mating at 21.5% (Giraud et al. 2005). Given such occurrence of both inter- and intra-tetrad mating, it is possible that tetrapolarity with mating-type loci linked to centromere has been an intermediate step in the evolution of bipolarity from tetrapolarity. Evidence for a centromere linkage of each mating-type locus predating their fusion on a single mating-type chromosome could be found in genomic data by studying further bipolar *Microbotryum* species, e.g., if the genes ancestrally located between each mating-type locus and their respective centromere were found in a species to show an older recombination suppression than the genes between ancient centromeres.
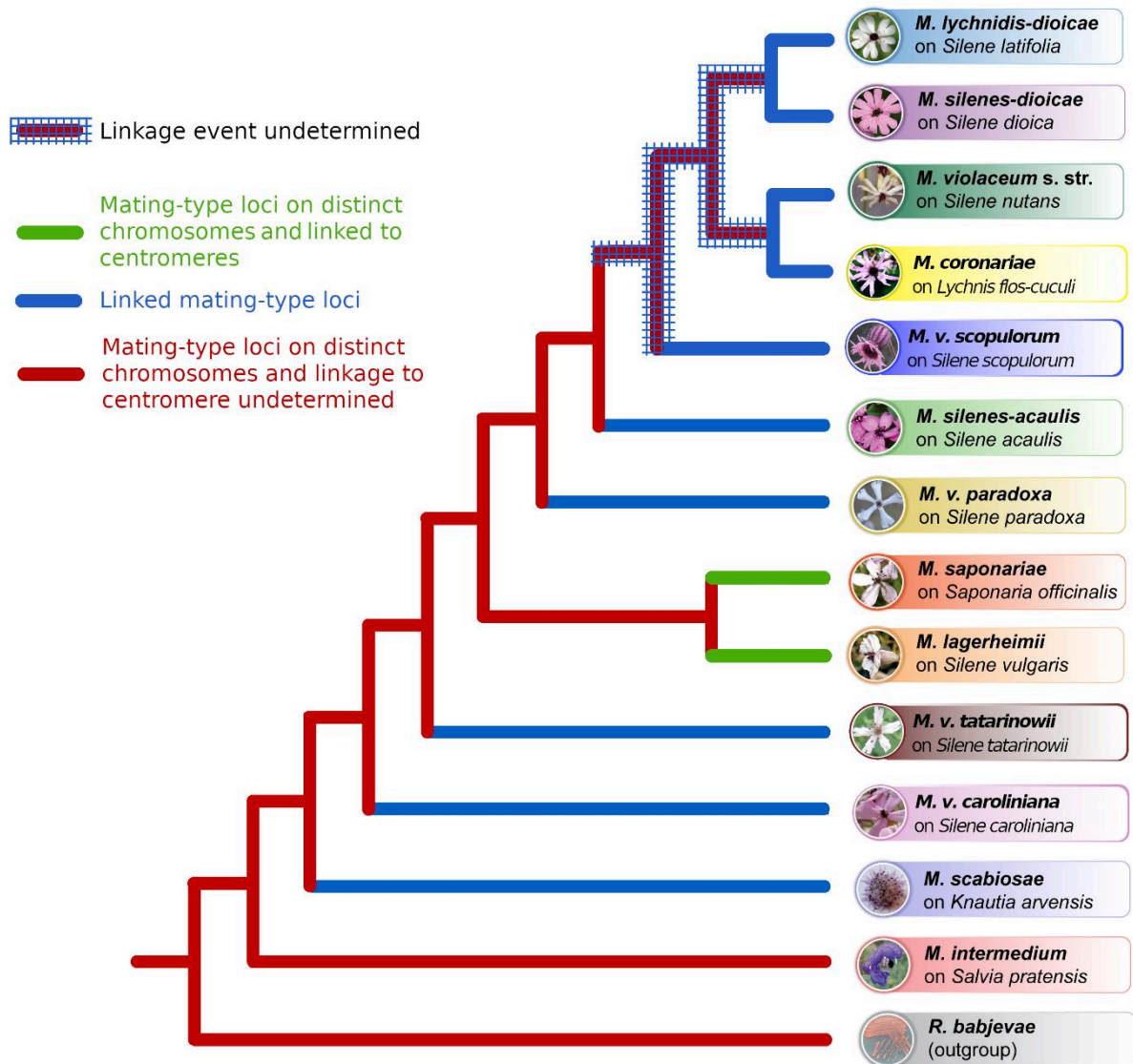
Figure 4: Cladogram of the *Microbotryum* genus (included here 13 *Microbotryum* species). *Rhodotorula babjevae* is used as an outgroup. Diseased host plants of *Microbotryum* fungi are shown. Colours on branches show the information we have on the gamete compatibility systems at the end of the PhD.

The evolutionary convergence we reported in *Microbotryum* strongly suggests that a transition from low to high selfing rate triggered transitions from tetrapolarity to bipolarity. A transition toward high selfing rates could have been very ancient and before the radiation of the *Microbotryum* clade. Such ancient transition toward selfing would imply a certain delay for the transition from tetrapolarity to bipolarity. This delay could result from a low probability of ectopic recombination or a low probability for the fixation in a population of the rearrangement bringing the mating-type loci onto the same chromosome. Indeed, it seems unlikely the same rearrangement occurs simultaneously in the two alternative mating-types and an intermediate state with linkage in a single mating-type chromosome may be deleterious. A model for the evolution of tetrapolarity toward bipolarity has been proposed in *Cryptococcus neoformans* (Fraser et al. 2004). From an ancestral tetrapolar population, the mating-type chromosomes from one mating-type may fuse, resulting in the coexistence in the population of diploid individuals with a tetrapolar gamete compatibility system and diploid individuals with a "tripolar" gamete compatibility system *i.e.* with a fused mating-type chromosome for one mating type allele and two unfused mating-type chromosomes for the other allele. In such a diploid tripolar individual, a double crossing-over located between the mating-type loci may occur between the fused and unfused mating-type chromosomes, resulting in a diploid "tripolar" individual with the fused mating-type chromosome carrying alleles that were unlinked before the double crossing-over event. Although this scenario seems mechanistically plausible, fused and unfused mating-type chromosome may result in abnormal meiosis and impose deleterious effect until bipolarity evolves.

We reported evidence suggesting that differential expression for genes in the non-recombining region on mating-type chromosomes have been triggered by genomic degeneration. While the differential expression of genes in non-recombining regions of sex chromosomes are usually interpreted as resulting from sexual antagonism, they might instead result from degeneration. Genomic degeneration has been little studied, and reports are classically limited to the quantification of its features. Genomic degeneration has tremendous impacts on non-recombining regions. Moreover, degeneration can trigger further evolutionary consequences with whole genome impacts, such as dosage compensation and gene movements out of the non-recombining region. It may also generate a "reservoir" of transposable elements that subsequently disperse in the genome. It is therefore primordial to consider the effect of degeneration when studying evolution in organisms with ancient recombination suppression. Little is known about the tempo of degeneration, *i.e.* the pace of each deleterious mutations

accumulation once recombination is suppressed. I am currently studying the tempo of degeneration. To do so, I am computing degeneration index and assessing their values across the various evolutionary strata and *Microbotryum* species, and I will plot them against the time since recombination cessation, either by using synonymous divergence calculated between a_1 and a_2 associated alleles or by estimating absolute ages using a calibration point. I will then try to fit a statistical model explaining the relationships between the degeneration features and the time since recombination cessation. This will allow studying the tempo of degeneration, i.e., at what speed deleterious mutations accumulate and whether some types of deleterious mutations accumulate faster than others. The degeneration features I am analysing are the number of transposable element copies, the number non-synonymous substitutions calculated between alleles associated to the a_1 and a_2 mating types, and the frequency of optimal codons.

Synonymous mutations are often assumed to be neutral because they do not change amino acids in proteins. However, evidence from multiple organisms suggest that some synonymous codons are selected over others because they increase the efficiency of translation, which should be the more crucial as the gene is more expressed. In this context, optimal codons are defined as codons used preferentially in highly expressed genes, leading to a codon bias. Because selection is less efficient in non-recombining regions, a codon bias resulting from a selection for increased translation efficiency is expected to be less strong in non-recombining regions. The decrease in the frequency of optimal codons in a non-recombining region can thus be considered as a degeneration feature. Compared to the deleterious effects resulting from the accumulation of transposable elements or non-synonymous substitutions, the deleterious effects of non-optimal codon usage may be less severe. Degeneration in codon usage has been little documented and investigating differences in the accumulation dynamic of various mutations that are differentially impacted by selection will bring valuable insights to understand genomic degeneration.

Overall, the results and conclusions drawn in this thesis show that non-recombining regions share common features in sex chromosomes and mating-type chromosomes, suggesting that the role of sexual antagonism may not be as predominant as commonly thought. This suggests that other evolutionary mechanisms and features associated with recombination suppression may be ubiquitous among organisms. it will be interesting in future studies to investigate whether evolutionary strata can be observed in other systems without sexual antagonisms, such as around plant self-incompatibility systems or other supergenes. It will also be worthwhile testing the alternative hypotheses to sexual antagonism, e.g., the heterozygote

advantage, transposable elements or neutral rearrangements. This could be done by investigating at the population level patterns of accumulation of transposable elements, neutral rearrangements and deleterious mutations.

Cited literature

Abbott JK, Nordén AK, Hansson B. 2017. Sex chromosome evolution: historical insights and future perspectives. *Proc R Soc B* **284**: 20162806.

Allen CE. 1945. The genetics of bryophytes II. *Bot Rev* **11**: 260–287.

Antonovics J, Abrams JY. 2004. Intratetrad mating and the evolution of linkage relationships. *Evolution (N Y)* **58**(4): 702–709.

Bachtrog D, Charlesworth B. 2001. Towards a complete sequence of the human Y chromosome. *Genome Biol* **2**(5): reviews1061.1-1016.5.

Bachtrog D, Mank JE, Peichel CL, Kirkpatrick M, Otto SP, Ashman T-L, Hahn MW, Kitano J, Mayrose I, Ming R, et al. 2014. Sex determination: why so many ways of doing it? *PLoS Biol* **12**(7): e1001899.

Badouin H, Hood ME, Gouzy J, Aguileta G, Siguenza S, Perlin MH, Cuomo CA, Fairhead C, Branca A, Giraud T. 2015. Chaos of rearrangements in the mating-type chromosomes of the anther-smut fungus *Microbotryum lychnidis-dioicae*. *Genetics* **200**(4): 1275–1284.

Bakkeren G, Jiang RL, Warren Y, Butterfield H, Shin R, Chiu R, Linning J, Schein N, Lee GH. 2006. Mating factor linkage and genome evolution in basidiomycetous pathogens of cereals. *Fungal Genet Biol* **43**(9): 655–666.

Balounova V, Gogela R, Cegan R, Cangren P, Zluvova J, Safar J, Kovacova V, Bergero R, Hobza R, Vyskot B, et al. 2019. Evolution of sex determination and heterogamety changes in section *Otites* of the genus *Silene*. *Sci Rep* **9**: 1045.

Bell G. 1982. *The masterpiece of nature: the evolution and genetics of sexuality*. University of Califorlia Press, Berkeley.

Bennett RJ, Johnson AD. 2005. Mating in *Candida albicans* and the search for a sexual cycle. *Annu Rev Microbiol* **59**: 233–255.

Bergero R, Charlesworth D. 2019. Reply to Writght et al.: How to explain the absence of extensive Y-specific regions in the guppy sex chromosomes. *PNAS* **116**(26): 12609–12610.

Bergero R, Charlesworth D. 2009. The evolution of restricted recombination in sex chromosomes. *Trends Ecol Evol* **24**(2): 94–102.

Bergero R, Charlesworth D, Filatov DA, Moore RC. 2008. Defining regions and rearrangements of the *Silene latifolia* Y chromosome. *Genetics* **178**: 2045–2053.

Bergero R, Forrest A, Kamau E, Charlesworth D. 2007. Evolutionary strata on the X chromosomes of the dioecious plant *Silene latifolia*: evidence from new sex-linked genes. *Genetics* **175**(4): 1945–1954.

Bergero R, Gardner J, Bader B, Yong L, Charlesworth D. 2019. Exaggerated heterochiasmy in a fish with sex-linked male coloration polymorphisms. *PNAS* **116**(14): 6924–6931.

Bernstein H, Byers GS, Michod RE. 1981. Evolution of sexual reproduction: importance of DNA repair, complementation, and variation. *Am Nat* **117**(4): 537–549.

Billiard S, López-Villavicencio M, Devier B, Hood ME, Fairhead C, Giraud T. 2011. Having sex, yes, but with whom? Inferences from fungi on the evolution of anisogamy and mating types. *Biol Rev* **86**(2): 421–442.

Billiard S, López-Villavicencio M, Hood ME, Giraud T. 2012. Sex, outcrossing and mating types: unsolved questions in fungi and beyond. *J Evol Biol* **25**: 1020–1038.

Borodin PM, Basheva EA, Torgasheva AA, Dashkevich OA, Golenishchev FN, Kartavtseva I V., Mekada K, Dumont BL. 2012. Multiple independent evolutionary losses of XY pairing at meiosis in the grey voles. *Chromosom Res* **20**: 259–268.

Brush SG. 1978. Nettie M. Stevens and the discovery of sex determination by chromosomes. *Isis* **69**(2): 162–172.

Casselton LA. 2002. Mate recognition in fungi. *Heredity (Edinb)* **88**: 142–147.

Chantha SC, Herman AC, Platts AE, Vekemans X, Schoen DJ. 2013. Secondary evolution of a self-incompatibility locus in the *Brassicaceae* genus *Leavenworthia*. *PLoS Biol* **11**(5): e1001560.

Charlesworth B. 1976. Recombination modification in a fluctuating environment. *Genetics* (3): 181–195.

Charlesworth B, Charlesworth D. 1978. A model for the evolution of dioecy and gynodioecy. *Am Nat* **112**(988): 975–997.

Charlesworth B, Morgan MT, Charlesworth D. 1993. The effect of deleterious mutations on neutral molecular variation. *Genetics* **134**: 1289–1303.

Charlesworth B, Wall JD. 1999. Inbreeding, heterozygote advantage and the evolution of neo-X and neo-Y sex chromosomes. *Proc R Soc B Biol Sci* **266**: 51–56.

Charlesworth D. 2018. The Guppy Sex Chromosome System and the for Y Chromosome Recombination Suppression. *Genes (Basel)* **9**: 264–282.

Charlesworth D. 1995. The Pattern of Neutral Molecular Variation Under the Background Selection Model.

Charlesworth D. 2016. The status of supergenes in the 21st century: Recombination suppression in Batesian mimicry and sex chromosomes and other complex adaptations. *Evol Appl* **9**: 74–90.

Charlesworth D, Charlesworth B. 1987. Inbreeding depression and its evolutionary consequences. *Annu Rev Ecol Evol Syst* **18**: 237–268.

Charlesworth D, Charlesworth B. 1979. The evolutionary genetics of sexual systems in flowering plants. *Proc R Soc B* **205**(1161): 513–530.

Charlesworth D, Charlesworth B, Marais G. 2005. Steps in the evolution of heteromorphic sex chromosomes. *Heredity (Edinb)* **95**: 118–128.

Coelho MA, Sampaio JP, Gonc P. 2010. A deviation from the bipolar-tetrapolar mating paradigm in an early diverged Basidiomycete. *PLoS Genet* **6**(8): e1001052.

Coppin E, Debuchy R, Arnaise S, Picard M. 1997. Mating types and sexual development in filamentous Ascomycetes. *Microbiol Mol Biol Rev* **61**(4): 411–428.

Cox RM, Calsbeek R. 2009. Sexually antagonistic selection, sexual dimorphism, and the resolution of intralocus sexual conflict. *Am Nat* **173**(2): 176–187.

Day AW. 1976. Communication through fimbriae during conjugation in a fungus. *Nature* **262**: 583–584.

Day AW, Garber ED. 1988. *Ustilago violacea*, anther smut of the *Caryophyllaceae*. In *Genetics of plant pathogenic fungi*, pp. 457–482.

De Meeûs T, Prugnolle F, Agnew P. 2007. Asexual reproduction: genetics and evolutionary aspects. *Cell Mol Life Sci* **64**(11): 1355–1372.

Debuchy R, Berteaux-Lecellier V, Silar P. 2010. Mating systems and sexual morphogenesis in Ascomycetes. In *Cellular and Molecular Biology of Filamentous Fungi* (eds. K.A. Borkovich, D.J. Ebbole, and M. Momany), pp. 501–535, Washington DC: ASM.

Deml G, Oberwinkler. F. 1982. Studies in Heterobasidiomycetes. On *Ustilago violacea* (Pers.) Rouss. from *Saponaria officinalis*. *L Phytopathol Z* **104**: 345–356.

Devier B, Aguileta G, Hood ME, Giraud T. 2009. Ancient trans-specific polymorphism at pheromone receptor genes in basidiomycetes. *Genetics* **181**(1): 209–223.

Ellegren H, Parsch J. 2007. The evolution of sex-biased genes and sex-biased gene expression. *Nat Rev Genet* **8**(9): 689–98.

Felsenstein J. 1974. The evolutionary advantage of recombination. *Genetics* **78**: 737–756.

Fisher RA. 1930. *The genetical theory of natural selection*. The Clarendon Press, Oxford.

Fontanillas E, Hood ME, Badouin H, Petit E, Barbe V, Gouzy J, De Vienne DM, Aguileta G, Poulain J, Wincker P, et al. 2015. Degeneration of the nonrecombining regions in the mating-type chromosomes of the anther-smut fungi. *Mol Biol Evol* **32**(4): 928–943.

Fraser JA, Diezmann S, Subaran RL, Allen A, Lengeler KB, Dietrich FS, Heitman J. 2004. Convergent evolution of chromosomal sex-determining regions in the animal and fungal kingdoms. *PLoS Biol* **2**(12): e384.

Fraser JA, Heitman J. 2004. Evolution of fungal sex chromosomes. *Mol Microbiol* **51**: 299–306.

Ganders FR. 1979. The biology of heterostyly. *New Zeal J Bot* **17**: 607–635.

Garber ED, Day AW. 1985. Genetic mapping of aphytopathogenic basidiomycete, *Ustilago violacea*. *Bot Gaz* **146**: 449–459.

Giraud T, Jonot O, Shykoff JA. 2005. Selfing propensity under choice conditions in a parasitic fungus, Microbotryum violaceum, and parameters influencing infection success in artificial inoculations. *Int J Plant Sci* **166**(4): 649–657.

Goubet PM, Bergès H, Bellec A, Prat E, Helmstetter N, Mangenot S, Gallina S, Holl A-C, Fobis-Loisy I, Vekemans X, et al. 2012. Contrasted patterns of molecular evolution in dominant and recessive self-incompatibility haplotypes in *Arabidopsis*. *PLoS Genet* **8**(3): e1002495.

Gould SJ. 1990. *Wonderful life: the Burgess Shale and the nature of history*. WW Norton & Company.

Hamilton WD. 1980. Sex versus non-sex versus parasite. *Oikos* **35**: 282–290.

Handley LL, Ceplitis H, Ellegren H. 2004. Evolutionary strata on the chicken Z chromosome: implications for sex chromosome evolution. *Genetics* **167**: 367–376.

Hartfield M, Keightley PD. 2012. Current hypotheses for the evolution of sex and recombination. *Integr Zool* **7**: 192–209.

Hood ME. 2002. Dimorphic mating-type chromosomes in the fungus *Microbotryum violaceum*. *Genetics* **160**(2): 457–461.

Hood ME, Antonovics J, Koskella B. 2004. Shared forces of sex chromosome evolution in haploid-mating and diploid-mating organisms: *Microbotryum violaceum* and other model organisms. *Genetics* **168**(1): 141–146.

Hood ME, Mena-alı JI, Gibson AK, Oxelman B, Giraud T, Yockteng R, Arroyo MTK, Conti F, Pedersen AB, Gladieux P, et al. 2010. Distribution of the anther-smut pathogen *Microbotryum* on species of the *Caryophyllaceae*. *New Phytol* **187**(1): 217–229.

Hood ME, Petit E, Giraud T. 2013. Extensive divergence between mating-type chromosomes of the anther-smut fungus. *Genetics* **193**(1): 309–315.

Hood ME, Scott M, Hwang M. 2015. Breaking linkage between mating compatibility factors: tetrapolarity in *Microbotryum*. : 2561–2572.

Huang Y, Dang VD, Chang N, Wang J. 2018. Multiple large inversions and breakpoint rewiring of gene expression in the evolution of the fire ant social supergene. *Proc R Soc B* **285**: 20180221.

Hudson RR, Kaplan NL. 1995. Deleterious background selection with recombination. *Genetics* (141): 1605–1617.

Ironside JE. 2010. No amicable divorce? Challenging the notion that sexual antagonism drives sex chromosome evolution. *BioEssays* **32**(8): 718–726.

Jaenike J. 1978. An hypothesis to account for the maintenance of sex within populations. *Evol theory* **3**: 191–194.

James TY, Srivilai P, Kües U, Vilgalys R. 2006. Evolution of the bipolar mating system of the mushroom *Coprinellus disseminatus* from its tetrapolar ancestors involves loss of mating-type-specific pheromone receptor function. *Genetics* **172**: 1877–1891.

Jay P, Chouteau M, Whibley A, Bastide H, Llaurens V, Parrinello H. 2019. Mutation accumulation in chromosomal inversions maintains wing pattern polymorphism in a butterfly. *bioRxiv* : doi: http://dx.doi.org/10.1101/736504.

Jay P, Whibley A, Frézal L, Rodríguez de Cara MÁ, Nowell RW, Mallet J, Dasmahapatra KK, Joron M. 2018. Supergene evolution triggered by the introgression of a chromosomal inversion. *Curr Biol* **28**: 1–7.

Joron M, Frezal L, Jones RT, Chamberlain NL, Lee SF, Haag CR, Whibley A, Becuwe M, Baxter SW, Ferguson L, et al. 2011. Chromosomal rearrangements maintain a polymorphic supergene controlling butterfly mimicry. *Nature* **477**(7363): 203–206.

Joron M, Papa R, Beltra M, Chamberlain N, Mavarez J, Baxter S, Abanto M, Bermingham E, Humphray SJ, Rogers J, et al. 2006. A conserved supergene locus controls colour pattern diversity in *Heliconius* butterflies. *PLoS Biol* **4**(10): e303.

Joron M, Wynne I. R, Lamas G, Mallet J. 1999. Variable selection and the coexistence of multiple mimetic forms of the butterfly *Heliconius numata*. *Evol Ecol* **13**: 721–754.

Kappel C, Huu CN, Lenhard M. 2017. A short story gets longer: recent insights into the molecular basis of heterostyly. *J Exp Bot* **68**(21–22): 5719–5730.

Kearney MR. 2003. Why is sex so unpopular in the Australian desert? *Trends Ecol Evol* **18**(12): 605–607.

Kemler M, Göker M, Oberwinkler F, Begerow D. 2006. Implications of molecular characters for the phylogeny of the Microbotryaceae (Basidiomycota: Urediniomycetes). **6**: 35.

Kent T V., Uzunović J, Wright SI. 2017. Coevolution between transposable elements and recombination. *Philos Trans R Soc B Biol Sci* **372**(1736): 20160458.

Kniep H. 1919. Untersuchungen uber den Antherenbrand (Ustilago violacea Pers.). Ein Beitrag zum Sexualitatsproblem. *Z Bot* **11**: 257–284.

Lahn BT, Page DC. 1999. Four evolutionary strata on the human X chromosome. *Science* **286**: 964–967.

Larracuente AM, Presgraves DC. 2012. The selfish segregation distorter gene complex of *Drosophila melanogaster*. *Genetics* **192**: 33–53.

Lin X, Heitman J. 2007. Mechanisms of homothallism in fungi and transitions between heterothallism and homothallism. In *Sex in fungi: molecular determination and evolutionary implications* (eds. J. Heitman, J.W. Kronstad, J.W. Taylor, and L.A. Casselton), Washington, DC: ASM Press.

Lindholm A, Breden F. 2002. Sex Chromosomes and Sexual Selection in Poeciliid Fishes. *Am Nat* **160**: S214–S224.

Lopez-Villavicencio M, Jonot O, Coantic A, Hood ME, Enjalbert J, Giraud T. 2007. Multiple infections by the anther smut pathogen are frequent and involve related strains. *PLoS Pathog* **3**(11): e176.

Lyon MF. 2003. Transmission ratio distortion in mice. *Annu Rev Genet* **37**: 393–408.

Mather K. 1955. Polymorphism as an outcome of disruptive selection. *Evolution (N Y)* **9**: 52–61.

Maynard Smith J. 1978. *The evolution of sex*. Cambrdige University Press, Cambridge (UK).

Maynard Smith J. 1971. What use is sex? *J theor Biol* **30**: 319–335.

Maynard Smith J, Haigh J. 1974. The hitch-hiking effect of a favourable gene. *Genet Res* **23**: 22–35.

Mcdaniel SF, Willis JH, Shaw AJ. 2007. A linkage map reveals a complex basis for segregation distortion in an interpopulation cross in the moss *Ceratodon purpureus*. *Genetics* **176**: 2489–2500.

Melters DP, Bradnam KR, Young HA, Telis N, May MR, Ruby JG, Sebra R, Peluso P, Eid J, Rank D, et al. 2013. Comparative analysis of tandem repeats from hundreds of species reveals unique insights into centromere evolution. *Genome Biol* **14**(1): R10.

Menkis A, Jacobson DJ, Gustafsson T, Johannesson H. 2008. The mating-type chromosome in the filamentous ascomycete *Neurospora tetrasperma* represents a model for early evolution of sex chromosomes. *PLoS Genet* **4**(3): e1000030.

Moran N. 1992. The evolution of Aphid life cycles. *Annu Rev Entomol* **37**(1): 321–348.

Morgan TH. 1909. A biological and cytological study of sex determination in phylloxerans and aphids. *J Exp Zool* **7**: 239–351.

Muller HJ. 1932. Some genetic aspects of sex. *Am Nat* **66**(703): 118–138.

Nadeau NJ. 2016. Genes controlling mimetic colour pattern variation in butterflies. *Curr Opin Insect Sci* **17**: 24–31.

Nam K, Ellegren H. 2008. The chicken (*Gallus gallus*) Z chromosome contains at least three nonlinear evolutionary strata. *Genetics* **180**: 1131–1136.

Nei M. 1969. Linkage modification and sex difference in recombination. *Genetics* **63**: 681–699.

Nieuwenhuis BPS, Aanen DK. 2018. Nuclear arms races: Experimental evolution for mating success in the mushroom-forming fungus *Schizophyllum commune*. *PLoS One* **13**(12): e0209671.

Nieuwenhuis BPS, Billiard S, Vuilleumier S, Petit E, Hood ME, Giraud T. 2013. Evolution of uni- and bifactorial sexual compatibility systems in fungi. *Heredity (Edinb)* **111**: 445–455.

Nieuwenhuis BPS, James TY. 2016. The frequency of sex in fungi. *Phil Trans R Soc B* **371**(1706): 20150540.

Nieuwenhuis S, Debets AJM, Aanen DK. 2011. Sexual selection in mushroom-forming basidiomycetes. *Proc R Soc B* **278**: 152–157.

Ohno S. 1967. *Sex chromosomes and sex-linked genes*. Springer-V.

Otto SP, Lenormand T. 2002. Resolving the paradox of sex and recombination. *Nat Rev Genet* **3**: 252–261.

Parsch J, Ellegren H. 2013. The evolutionary causes and consequences of sex-biased gene expression. *Nat Rev Genet* **14**(2): 83–87.

Petit E, Giraud T, de Vienne DM, Coelho MA, Aguileta G, Amselem J, Kreplak J, Poulain J, Gavory F, Wincker P, et al. 2012. Linkage to the mating-type locus across the genus *Microbotryum*: insights into nonrecombining chromosomes. *Evolution (N Y)* **66**(11): 3519–3533.

Ponnikas S, Sigeman H, Abbott JK, Hansson B. 2018. Why do sex chromosomes stop recombining ? *Trends Genet* **34**(7): 492–503.

Pracana R, Priyam A, Levantis I, Nichols RA. 2017. The fire ant social chromosome supergene variant Sb shows low diversity but high divergence from SB. *Mol Ecol* **26**(11): 2864–2879.

Refrégier G, Gac M Le, Jabbour F, Widmer A, Shykoff JA, Yockteng R, Hood ME, Giraud T. 2008. Cophylogeny of the anther smut fungi and their caryophyllaceous hosts: prevalence of host shifts and importance of delimiting parasite species for inferring cospeciation. *BMC Evol Biol* **8**: 100.

Reichwald K, Petzold A, Koch P, Cellerino A, Englert C, Platzer M. 2015. Insights into sex chromosome evolution and aging from the genome of a short-lived fish resource. *Cell* **163**: 1527–1538.

Rice WR. 1984. Sex chromosomes and the evolution of sexual dimorphism. *Evolution (N Y)* **38**(4): 735–742.

Rick CM. 1971. The tomato *Ge* locus: linkage relations and geographic distribution of alleles. *Genetics* **67**: 75–85.

Ross G, Keller L. 1995. Ecology and evolution of social organization: insights from fire ants and other highly eusocial insects. *Annu Rev Ecol Syst* **26**: 631–656.

Saenko S V, Chouteau M, Prunier FP, Blugeon C, Joron M. 2019. Unravelling the genes forming the wing pattern supergene in the polymorphic butterfly *Heliconius numata*. *Evodevo* **10**: 16.

Schaefer I, Domes K, Heethoff M, Schneider K, Schön I, Norton RA, Scheu S, Maraun M. 2006. No evidence for the "Meselson effect" in parthenogenetic oribatid mites (Oribatida, Acari). *J Evol Biol* **19**: 184–193.

Schwander T, Libbrecht R, Keller L. 2014. Supergenes and complex phenotypes. *Curr Biol* **24**(7): R288–R294.

Shykoff JA, Kaltz O. 1997. Effects of the anther smut fungus *Microbotryum violaceum* on host life-history patterns in *Silene latifolia* (Caryophyllaceae). *Int J Plant Sci* **158**(2): 164–171.

Stevens NM. 1905. A study of the germ cells of *Aphis rosae* and *Aphis oenotherae*. *J Exp Zool* **2**: 313–333.

Sun Y, Svedberg J, Hiltunen M, Corcoran P, Johannesson H. 2017. Large-scale suppression of recombination predates genomic rearrangements in *Neurospora tetrasperma*. *Nat Commun* **8**(1): 1140.

Tennessen JA, Govindarajulu R, Liston A, Ashman T. 2016. Homomorphic ZW chromosomes in a wild strawberry show distinctive recombination heterogeneity but a small sex-determining region. *New Phytol* **211**: 1412–1423.

Thrall PH, Biere A, Antonovics J. 1993. Plant life-history and disease susceptibility - the occurrence of *Ustilago violacea* on different species within the Caryophyllaceae. *J Ecol* **81**(3): 489–498.

Van Der Gaag M, Debets AJM, Oosterhof J, Slakhorst M, Thijssen JAGM, Hoekstra RF. 2000. Spore-killing meiotic drive factors in a natural population of the fungus *Podospora anserina*. *Genetics* **156**: 593–605.

Veyrunes F, Waters PD, Miethke P, Rens W, Mcmillan D, Alsop AE, Grützner F, Deakin JE, Whittington CM, Schatzkamer K, et al. 2008. Bird-like sex chromosomes of platypus imply recent origin of mammal sex chromosomes. *Genome Res* **18**: 965–973.

Votintseva AA, Filatov DA. 2009. Evolutionary strata in a small mating-type-specific region of the smut fungus *Microbotryum violaceum*. *Genetics* **182**(4): 1391–1396.

Vrijenhoek RC. 1998. Animal clones and diversity. *Bioscience* **48**(8): 617–628.

Wang J, Wurm Y, Nipitwattanaphon M, Riba-grognuz O, Huang Y, Shoemaker D, Keller L. 2013. A Y-like social chromosome causes alternative colony organization in fire ants. *Nature* **493**(7434): 664.

Westergaard M. 1958. The mechanism of sex determination in dioecious flowering plants. *Adv Genet* **9**: 217–281.

White M. 1978. *Modes of speciation*. W.H. Freeman, San Francisco.

Whitton J, Sears CJ, Baack EJ, Otto SP. 2008. The dynamic nature of apomixis in the angiosperms. *Int J Plant Sci* **169**(1): 169–182.

Xu L, Petit E, Hood ME. 2016. Variation in mate-recognition pheromones of the fungal genus *Microbotryum. Heredity (Edinb)* **116**: 44–51.

Yi R, Tachikawa T, Ishikawa M, Mukaiyama H, Bao D, Aimi T. 2009. Genomic structure of the A mating-type locus in a bipolar basidiomycete, *Pholiota nameko. Mycol Res* **113**(2): 240–248.

Yockteng R, Marthey S, Chiapello H, Gendrault A, Hood ME, Rodolphe F, Devier B, Wincker P, Dossat C, Giraud T. 2007. Expressed sequences tags of the anther smut fungus, *Microbotryum violaceum*, identify mating and pathogenicity genes. *BMC Genomics* **8**: 272.

Yu Q, Hou S, Feltus FA, Jones MR, Murray JE, Veatch O, Lemke C, Saw JH, Moore RC, Thimmapuram J, et al. 2008. Low X/Y divergence in four pairs of papaya sex-linked genes. *Plant J* **53**: 124–132.

Yun S-H, Berbee ML, Yoder OC, Turgeon BG. 1999. Evolution of the fungal self-fertile reproductive life style from self-sterile ancestors. *Proc Natl Acad Sci USA* **96**: 5592–5597.

List of figures

Photo credits to Michael E. Hood, and electronic microscopy photo credits to Dominik Bergerow.

Cladograms realized using iToL with a newick format provided by Ricardo Rodriguez de la Vega. Photo credit to Michael E. Hood. Species box models inspired by figures from Marco Coelho.

Supplemental information

S.1   Evolutionary strata on young mating-type chromosomes

SI Appendix

Evolutionary strata on young mating-type chromosomes despite the lack of sexual antagonism

Sara Branco[1,a], Hélène Badouin[1,a,b], Ricardo C. Rodríguez de la Vega[1a], Jérôme Gouzy[b], Fantin Carpentier[a], Gabriela Aguileta[a], Sophie Siguenza[b], Jean-Tristan Brandenburg[a], Marco A. Coelho[c], Michael E. Hood[2,d], Tatiana Giraud[2,a,*]

[1] These authors contributed equally to the study

[2] These authors jointly supervised the study

[a] Ecologie Systématique Evolution, Bâtiment 360, Univ. Paris-Sud, AgroParisTech, CNRS, Université Paris-Saclay, 91400 Orsay, France.
[b] LIPM, Université de Toulouse, INRA, CNRS, 31326 Castanet-Tolosan, France.
[c] UCIBIO-REQUIMTE, Departamento de Ciências da Vida, Faculdade de Ciências e Tecnologia, Universidade NOVA de Lisboa, 2829-516 Caparica, Portugal.
[d] Department of Biology, Amherst College, 01002-5000 Amherst, Massachusetts, United State of America.

* Corresponding author: Tatiana Giraud (tatiana.giraud@u-psud.fr)

SI Appendix Includes:

SI Text
SI Materials and Methods
SI References
SI Figures S1 – S10
SI Tables S1 – S7

SI Text

Mating-type chromosomes in fungi

There are several hypotheses for the evolution of recombination suppression in fungal mating-type chromosomes, some also holding for animal, plant and algal sex chromosomes. These include:

1) Linkage of two mating-type loci: Most basidiomycete fungi (i.e., mushrooms, rusts and smut fungi) contain two loci determining pre- and post-mating compatibility, respectively. The PR locus controls gamete fusion, with pheromone receptor and pheromone precursor genes (the latter sometimes being present in multiple copies), and the HD locus controls compatibility during growth of after fusion of haploids cells, with two adjacent homeodomain genes (1, 2 , 3). Both PR and HD mating-type loci in basidiomycetes encompass two to several genes, and recombination is suppressed within each of these two loci, in all species, ensuring proper mating-type determination (3, 4). Linkage of these two mating-type loci is beneficial (but not required) under selfing mating systems in basidiomycete fungi as it increases the probability of compatibility among gametes from the same diploid parent (5-7) (Fig. S2). In contrast, the segregation of the two mating-type loci in association with multi-allelism increases discrimination against selfing and mating among closely related individuals, which promotes outcrossing (5-9). The PR and HD loci segregate independently in most basidiomycetes, but these loci are linked in some species (6, 10-13). Independent meiotic segregation of the two mating-type loci in basidiomycetes is called tetrapolarity (because meiotic segregation can result in four possible mating types among gametes of a diploid parent, Fig. S2A), while linkage of the mating-type loci in basidiomycetes leads to bipolarity (segregation of only two mating types among gametes of a diploid parent, Fig. S2C).

2) A second explanation for suppressed recombination in fungal mating-type chromosomes is the linkage of mating-type genes to the centromere. In ascomycete and basidiomycete automictic species (*i.e.,* undergoing selfing among products of the same meiotic tetrad (11, 14, 15)), such linkage is particularly common and thought to be favorable (16, 17), resulting in mating-type segregation and reformation of the diploid between meiotic products separated at the first meiotic division. In basidiomycetes, linkage of both HD and PR mating-type genes to their respective centromeres increases the odds of compatibility within tetrads (Fig. S2B). Note that in basidiomycete fungi with linked HD and PR genes, linkage to the centromere is not beneficial with regard to the odds of compatibility between gametes, as it will not further increase mating compatibility among gametes in a diploid individual (Fig. S2C).

3) In analogy to the sexually antagonistic selection hypothesis, which is the main theory for explaining evolutionary strata in sex chromosome evolution, there could be "mating-type antagonistic selection" (18) if some phenotypes were beneficial for improving the a_1 function while being detrimental

to the a_2 function, or vice versa. However, there are not many traits that could have different optima in the two mating types.

4) Epistatic interactions can also favor linkage for protecting beneficial combinations of alleles, unrelated to mating type or sex functions. Unlike hypothesis 3 above that also deals with epistasis (19), this fourth hypothesis postulates that beneficial allelic combinations do not improve the a_1 and a_2 mating type functions in themselves, and therefore does not correspond to "mating-type antagonistic selection". Epistatic interactions can also occur on autosomes but the maintenance of alternative alleles should be facilitated when linked to the mating-type or sex-determining genes, which are under strong balancing selection.

5) Another evolutionary explanation for suppressed recombination on fungal mating-type chromosomes, that could also apply to sex chromosomes, involves linkage of deleterious alleles to the mating-type loci, favoring permanent sheltering in an heterozygous state, as has been theoretically modeled (20, 21). This may arise specifically in sex and mating-type chromosomes if recombination frequency gradually decreases from the non-recombining into the PAR, so that partial linkage (linkage disequilibrium) to mating-type or sex-determining genes makes selection against recessive deleterious alleles less efficient. This would allow deleterious alleles to increase in frequency in these PAR edges at the margin of the non-combining region. Rare recombination events would then generate individuals homozygous for deleterious alleles and could therefore be selected against. Complete linkage of these PAR margins in disequilibrium with mating type may thus be favorable and selected for (Fig. S1A). Permanent sheltering would thus be more easily achieved than purging if recombination is rare at the PAR margins. Further theoretical models are needed to explore the general conditions under which such a mechanism can generate evolutionary strata.

6) A last mechanism explaining suppressed recombination involves neutral inversions or rearrangements extending the region of suppressed recombination (22); these could be fixed by drift in one of the sex or mating-type chromosomes and arrest recombination between sex or mating-type chromosomes. For example, an inversion trapping a part of the PAR at the margin of the non-recombining region could be neutral in the short term, and could be fixed by drift in one of the sex or mating-type chromosomes, and not the other, because of the recombination suppression and because of mating occurring only between mating types (Fig. S1B and C). Such phenomena can occur on any chromosome, but linkage in mating-type and sex chromosomes could be maintained longer in a polymorphic state (and therefore be more detectable), due to balancing selection retaining alternate alleles determining sexes and mating types (22). As for the hypothesis above, theoretical models could be developed for validating whether such a mechanism can generate evolutionary strata.

The *Microbotryum lychnidis-dioicae* system

Contributing to understanding of these various evolutionary models, the anther-smut fungus *Microbotryum lychnidis-dioicae* has been important for the study of fungal mating-type chromosomes. It was central to the first studies describing bipolar breeding systems in fungi (23) and the first case of size-dimorphic mating-type chromosomes in fungi (24). Suppressed recombination encompasses ca. 90% of the mating-type chromosome lengths (sized 3.5 Mb for a_1 and 4.0 Mb for a_2) with 614 and 683 predicted non-transposable element genes, respectively (12, 24-27). The non-recombining region shows chaos of rearrangements and gene losses, transposable element accumulation, non-synonymous substitutions and impaired gene expression (12, 27). The non-recombining region is flanked at both sides by two small recombining and collinear regions, therefore called pseudo-autosomal regions (PARs).

The existence of evolutionary strata has been debated in *M. lychnidis-dioicae*. Varying levels of synonymous divergence between a_1 and a_2 alleles in a handful of genes in the unassembled non-recombining region led to the claim that evolutionary strata existed (28). Based on highly fragmented assemblies, it was then speculated that an oldest stratum evolved for linking pheromone precursor and pheromone receptor genes, a second small (60 kb) old stratum for linking the HD and PR locus, and a third recent and large stratum for linking the MAT genes to the centromeres (29). However, the full assemblies of the mating-type chromosomes recently showed that the HD and PR locus are 600 kb apart, that no clear discrete strata patterns were discernable using the current gene orders, as genes with different a_1-a_2 synonymous divergence levels (a proxy for time since recombination suppression) were dispersed chaotically along the chromosomes (12). These observations could be explained by a lack of evolutionary strata, extensive rearrangements of original strata, and/or gene conversion, which is known to occur in fungal mating-type chromosomes (30, 31). A recent study suggested the existence of five evolutionary strata along the mating-type chromosomes in *M. lychnidis-dioicae* using a clustering method based on putative oligonucleotide compositional differences between strata (32); however, this method did not seem to yield reliable inference because it could not retrieve the pseudo-autosomal boundary in one mating-type chromosome. Furthermore, each inferred stratum included genes with highly heterogeneous levels of divergence (12). Our study shows that the recovered putative strata in this previous study (32) in fact did not correspond to genuine evolutionary strata.

Elucidating whether evolutionary strata exist and which genomic regions they include is important for understanding how and why suppressed recombination evolved. The different mechanisms for explaining suppression of recombination outlined above do not all predict the existence of evolutionary strata. For example, linking the two mating-type loci would not generate multiple strata, while the successive linkage of deleterious alleles to the mating-type locus would (20). In this regard, an interesting feature of *M. lychnidis-dioicae* is that frequent deleterious alleles linked to mating types have been reported, that prevent haploid growth (14, 33, 34). In fact, many genes have been lost from

one or the other mating type (12) and transposable elements and non-synonymous substitutions have accumulated in non-recombining regions (12, 27).

Genome assemblies and comparisons

We used Pacific Bioscience sequencing technologies to generate very high quality genome assemblies for the mating-type chromosomes of several *Microbotryum* species. The a_1 and a_2 haploid genomes of *M. lagerheimii* and one haploid genome of *M. intermedium* were well assembled, with 42 and 37 contigs for a_1 and a_2 *M. lagerheimii*, respectively, 24 contigs for *M. intermedium*, for an estimated number of chromosomes of 12 for haploid *Microbotryum* genomes (24). The mating-type chromosomes were well assembled; the HD chromosomes of *M. intermedium* and *M. lagerheimii* appeared assembled in a single contig each, and were homologous and syntenic between the species; the complete PR chromosome of *M. intermedium* matched two contigs in the a_1 *M. lagerheimii* assembly, with a small inversion internal to one contig (the MC16 contig, in dark purple in Figure 2A; the inversion is represented by the orange links).

In addition to collinearity with *M. intermedium*, collinearity between the a_1 and a_2 mating-type chromosomes within *M. lagerheimii* (Fig. S4) indicates that recombination is frequent enough to conserve gene order along most of their lengths. In *M. lychnidis-dioicae* the recombining pseudo-autosomal regions of the mating-type chromosomes and autosomes are also highly collinear between a_1 and a_2 genomes, while the non-recombining regions of the mating-type chromosomes are thoroughly rearranged (Fig. S3).

The two distinct contigs carrying the PR and HD genes in *M. lagerheimii* (Fig. S4) are consistent with the species being tetrapolar, as inferred earlier based on meiotic segregation analysis of recombination between these mating-type loci (35). The high degree of collinearity between the mating-type chromosomes of *M. lagerheimii* and *M. intermedium,* and the finding of PR and HD genes in different contigs, each harbouring centromere-specific repeats (12) (Fig. 2A), indicate that *M. intermedium* is also tetrapolar.

We included comparisons of the mating-type chromosomes of *M. lychnidis-dioicae* and *M. violaceum s. str.* (Fig. S5B), whose haploid genomes were also well assembled (Table S4). Three contigs in the a_1 *M. violaceum sens. str.* genome matched the a_1 mating-type chromosomes of *M. lychnidis-dioicae*. In addition, these three contigs had some fragments joined together in the assembly of the *M. violaceum sens. str.* a_2 genome (Fig. S5D), indicating that all three belonged to the mating-type chromosome. The non-recombining region was highly rearranged between *M. lychnidis-dioicae* and *M. violaceum s. str.* (Fig. S5B) in all four possible comparisons, as well as between a_1 and a_2 in both species (Fig. S5D).

We also compared the mating-type chromosomes of the sister species *M. lychnidis-dioicae* and *M. silenes-dioicae* (Fig. S5C), the latter being assembled in more contigs than other species (Table S4;

6 contigs for both a_1 and a_2 mating-type chromosomes). These different contigs had some fragments joined together in the assembly of the alternative mating type (Fig. S5E) and in *M. lychnidis-dioicae* (Fig. S5C), allowing assigning to mating-type chromosomes. The non-recombining region appeared homologous but already highly rearranged between *M. lychnidis-dioicae* and *M. silenes-dioicae* (Fig. S5C) in all four possible comparisons, as well as between a_1 and a_2 in *M. silenes-dioicae* (Fig. S5E).

SI Materials and Methods

DNA extraction and sequencing

DNA was extracted using the Qiagen Kit 10243 (Courtaboeuf, France) following manufacturer instructions and using a Carver hydraulic press (reference 3968, Wabash, IN, USA) for breaking cell walls. A haploid genome of *M. intermedium* (strain 1389-BM-12-12, collected on *Salvia pratensis*, Italy, Coord. GPS : 44.33353 & 7.13637)*,* as well as the genomes of a_1 and a_2 haploid cells of *M. lagerheimii* (strain 1253, collected on *S. vulgaris*, near Chambery, France, Coord. GPS : 45.4 & 6.11), *M. violaceum s. str.* (strain 1249, collected on *S. nutans*, Guarda, Switzerland, Coord. GPS : 46.777 & 10.16) and *M. silenes-dioicae* (strain 1303, collected on *S. dioica*, Bois Carre, Lac de Puy Vachier 45.026485 & 6.276612) were sequenced using the P6/C4 Pacific Biosciences SMRT technology (UCSD IGM Genomics Facility La Jolla, CA, USA) (Table S4).

Assembly and annotation

We assembled the haploid genomes of a_1 and a_2 *M. lagerheimii, M. violaceum s. str., M. silenes-dioicae,* and a_1 *M. intermedium,* and re-assembled the reference genome of *M. lychnidis-dioicae* Lamole (12) into separated a_1 and a_2 haploid genomes. Assemblies of the genomes were generated with the wgs-8.2 version of the PBcR assembler (36) with the following parameters: genomeSize=30000000, assembleCoverage=50. Assemblies were polished with quiver software (https://github.com/PacificBiosciences/GenomicConsensus). A summary of raw data and assembly statistics is reported in Table S4.

The protein-coding gene models were predicted with EuGene (37), trained for *Microbotryum*. Similarities to the fungi subset of the uniprot database (38) plus the *M. lychnidis-dioicae* Lamole proteome (12) were integrated into EuGene for the prediction of gene models.

Mating-type chromosomes were identified by: 1) finding the contigs carrying the PR or HD mating-type genes in both a_1 and a_2 haploid genomes, 2) using BLAST to search the a_1 against the a_2 haploid genomes and visualizing the output using Circos (39) for identifying contigs with lack of collinearity, 3) blasting the haploid genomes against the completely assembled mating-type chromosomes of *M. lychnidis-dioicae* (12), 4) blasting the a_1 contigs identified in the first steps to the whole a_2 haploid genome, and reciprocally, to identify additional mating-type contigs, 5) repeating this process from step 3 until no additional contig was identified. These contigs were then orientated by: 1) using the centromere-specific repeats (12), as initial assemblies often yielded chromosome arms broken at their centromeres, with identifiable centromere-specific repeats on each separated contig (e.g., Fig. 2A); and 2) blasting the a_1 and a_2 mating-type contigs against each other for identifying the PARs as the collinear regions, that were then assumed to be at the edges of the chromosomes. When the mating-type

chromosomes were fragmented and internal contigs did not include centromeric repeats at one of its edges, it was impossible to recover contig orientation (as indicated in the figure captions).

Figures and statistical tests

The Figures 2, S3, S4 and S5 were prepared using Circos (39). We analyzed gene order to identify larger blocks of synteny. Alleles were assigned between the two mating-type chromosomes and between species by applying orthomcl (40) to the protein data sets. The Student t test was performed using JMP v7 (SAS Institute).

Species tree

We compared the translated gene models of five *Microbotryum* species and the red yeast *Rhodotorula babjevae* with blastp+ 2.30, and the output was used to obtain orthologous groups by Markov clustering (41) as implemented in OrthoMCL v1.4 (40). We aligned the protein sequences of 4,000 single-copy autosomal genes with MUSCLE v3.8.31 (42), and obtained the codon-based CDS alignments with TranslatorX (43). We used RAxML 8.2.7 (44) to obtain maximum likelihood gene trees for all 4,000 fully conserved single-copy genes and a species tree with the concatenated alignment of 2,100,301 codons with no gaps (trimal -nogaps option (45)) under the GTRGAMMA substitution model. We estimated the branch support values by bootstrapping the species tree based on the concatenated alignment and by estimating the relative internode and tree certainty scores based on the frequency of conflicting bipartitions for each branch in the species tree among the fully conserved single-copy genes (46).

Gene genealogies

Gene genealogies were inferred for codon-based alignments of genes in the different strata using RAxML (44) version 8.2.7, assuming the GTRGAMMA model and rapid bootstrap (options: -f a and -# 100). We analyzed for trans-specific polymorphism levels all single-copy genes for which we had both alleles and all species.

Date estimates for recombination suppression

In order to estimate dates of recombination suppression, we leveraged the codon-based alignments of single-copy orthologous groups yielding gene trees displaying trans-specific polymorphism in each evolutionary stratum, as the divergence between alleles associated to the a_1 versus a_2 mating types then corresponds to the time since recombination suppression. We could use three gene families in the purple stratum (1,744 aligned codons), 22 gene families in the black stratum (17,776 aligned codons) and four

gene families in the red stratum (2,726 aligned codons), with alignments including the five *Microbotryum* species and the red yeast used as outgroup, and gene genealogies showing trans-specific polymorphism. There was not enough single-copy gene families showing trans-specific polymorphism for reliable date estimate in the other evolutionary strata. Divergence times were estimated using BEAST v2.4.0 (47), the xml inputs being generated using BEAUTi, and setting the following parameters (others left as default values): unlinked substitution (HKY+G with empirical frequencies for each codon position) and clock models, Yule process to model speciation, and 10,000,000 mcmc generations sampled every 1,000. For all runs we used a single calibration prior at 0.42 million years, corresponding to the divergence between *M. lychnidis-dioicae* and *M. silenes-dioicae* (48), with a normal distribution and a sigma of 0.05.

Transposable element identification

Repetitive DNA content was analyzed with RepeatMasker (49), using REPBASE v19.11 (50). For plotting d_S along chromosomes, these repeats were removed and further filtering of repeats was performed by blasting (tBLASTx) and removing those matching at more than five locations in the genome.

Polymorphism data and analyses

To rule out high levels of polymorphism as the cause for the observed high d_S values in the *Microbotryum* mating-type chromosomes, we assessed the level of polymorphism across *M. silenes-dioicae* and *M. lychnidis-dioicae* by using multiple available genomes (Illumina paired-end 60x-100x), from individuals collected across Europe (29, 51). We only used genomes including a single mating-type chromosome (a_1 or a_2) either because a haploid strain was sequenced or because a lethal allele was linked to one mating type in a diploid strain so that only haploid cells carrying the alternative mating-type chromosome could be cultivated *in vitro* and sequenced (34, 51, 52). We compiled sequences for a total of ten a_1 and fourteen a_2 mating-type chromosomes of *M. lychnidis-dioicae,* and six a_1 and eight a_2 mating-type chromosomes of *M. silenes-dioicae* (Table S5).

We used two approaches to determine which mating-type chromosome(s) were present in genomes sequences (a_1 or a_2), by checking the presence/absence of the a_1 and a_2 pheromone receptor alleles. First, each genome was assembled *de novo* using SOAPdenovo (53) with default parameters, and searched using BLAST for the sequences of the pheromone receptor alleles of *M. lychnidis-dioicae* and *M. silenes-dioicae*, which are highly differentiated between a_1 and a_2 (54). In a second validation step, we counted the reads mapped against the two alleles of the mating-type pheromone receptors. Strains that displayed reads mapping against both alleles were excluded from further analyses.

We performed SNP calling as described previously (51). Reads were mapped with GLINT (T. Faraut & E. Courcelle, available at http://lipm-bioinfo.toulouse.inra.fr/download/glint/) with parameters

set as follows: minimum length of the high-scoring pair hsp $\geq$ 90, with $\leq$ 5 mismatches, no gap allowed, only best-scoring hits taken into account. For each strain, we used as a reference for mapping the haploid PacBio genome assembly corresponding to its species and mating-type. Between 92% and 97% of reads were mapped as proper pairs, and the mapping coverage ranged from 50 to 171 (Table S5). SNPs were called with VarScan (55) (min-coverage 10, min-reads2 5, min-avg-qual 30, min-var-freq 0.3, p-value 0.01), and filtered to remove SNPs located near indels, strand-biased SNPs and heterozygote SNPs.

To compute statistics of diversity, we generated pseudo-alignments of CDS in FASTA format from the VCF file produced by VarScan as previously (51). Pseudo-alignments are obtained by substituting reference nucleotides by their variants in the reference sequence. We then computed the θ_π statistic of diversity with EggLib version 2 (56) for each species and mating-type.

We generated alignments combining the *M. lychnidis-dioicae* and *M. silenes-dioicae* sequences generated above, as well as sequences of the other species. Codon-based alignments of single-copy genes in mating-type evolutionary strata and autosomes were obtained with MUSCLE and translatorX (see above). We used a custom python script to incorporate the sequences of the multiple genomes of *M. lychnidis-dioicae* and *M. silenes-dioicae* while taking into account gaps in the translatorX alignments. Trees were computed with the RAxML rapid bootstrap mode (-f a -m GTRGAMMA -x -# 100). To help visualization, trees were rooted using *R. babjevae* and *M. intermedium*, and strains were colored by mating-type with the ETE3 python toolkit (57).

For further checking that the observed differences in diversity levels between the pseudo-autosomal regions and the different strata in *M. lychnidis-dioicae* and *M. silenes-dioicae* (SI Appendix, Table S6) were not due to differences in mutation rates, we performed maximum-likelihood Hudson-Kreitman-Agade (MLHKA) tests as implemented in the MLHKA software (58) using the multiple genomes described above. The MLHKA test is designed to detect signatures of departure from neutrality (either balancing or positive selection) while controlling for mutation rate by the use of divergence data, and it allows multilocus comparisons (58). MLHKA tests can thus be used to test whether differences in diversity levels between two sets of genes is most likely due to elevated mutation rates or to balancing selection (in our case due to linkage of genes in evolutionary strata to mating-type genes), by comparing a model where all loci in the dataset evolve neutrally to an alternative model where *N* loci depart from neutrality (59, 60). Because MLHKA tests require a set of control loci that evolve neutrally, we applied these tests to pairwise comparisons between each stratum and the pseudo-autosomal region. To perform the MLHKA tests, we considered the pooled CDS sequences associated with the a_1 and a_2 mating-types for each stratum in each of *M. lychnidis-dioicae* and *M. silenes-dioicae* and used the genome of *M. violaceum sens. str.* as an outgroup. In order to improve divergence estimates, alignments were re-aligned with MACSE (61). Only synonymous sites were taken into account, and singletons were excluded. Unless stated otherwise, tests were run with 1,000,000 iterations of the Markov chain, and repeated at least three times with different random seed numbers to ensure convergence. We arbitrarily chose a starting value of 10 for divergence time, as this value is of little importance when the number of

iterations is high enough for the Markov chains to converge (58). Synonymous nucleotidic diversity (θ_π) was computed with EggLib version 2 (56) and used as starting value for theta. As the test assumes independence between loci, we concatenated the alignments of the CDS for each evolutionary stratum. For the largest stratum (the black stratum), we randomly sampled and concatenated CDS for 20 genes, as the high number of genes in this stratum resulted in unreasonable running times. For the two most recent strata (red and green strata), we also ran the tests without concatenating the alignments, with 2,000,000 million of iterations of the Markov chains. This yielded similar results as concatenating the CDS. *Microbotryum lychnidis-dioicae* displays a strong population structure in Europe (51, 62), which can lead to over-estimate the diversity in neutral loci and result in false negative tests. In fact, visual inspection of the alignments showed that most polymorphic sites in the PARs in this species corresponded to fixed differences between genetic clusters. We therefore also ran tests separately on the two genetic clusters for which we had several strains of each mating-type, i.e., the Southern and North-Western clusters (51). We corrected p-values for multiple testing within each species with a Benjamini and Hochberg correction (R function p.adjust). In *M. silenes-dioicae*, tests for all strata were significant (adjusted p-values < 0.05, Table S7A), indicating that balancing selection was more likely than elevated mutation rates for explaining the higher diversity in strata than in the PARs. In *M. lychnidis-dioicae*, tests were also significant for the purple, blue, black and orange strata (adjusted p-values < 0.05, Table S7B-D). The difference in diversity between the red stratum and the PAR was marginally significant when considering the whole data set, and significant when the tests were run separately on two genetic clusters (Table S7B-D). There was no significant difference in diversity between the green stratum and the PAR in *M. lychnidis-dioicae*, suggesting that suppressed recombination is not complete there in this species, or that it is too recent to have allowed an accumulation of differences between mating types.

SI References

1. Feldbrugge M, Kamper J, Steinberg G, & Kahmann R (2004) Regulation of mating and pathogenic development in *Ustilago maydis*. *Curr Op Microbiol* 7:666-672.

2. Kuees U (2015) From two to many: Multiple mating types in Basidiomycetes. *Fung Biol Rev* 29(3-4):126-166.

3. Casselton LA & Kues U (2007) The origin of multiple mating types in the model mushrooms *Coprinopsis cinerea* and *Schizophyllum commune*. *Sex in Fungi*, eds Heitman J, Kronstad J, Taylor J, & Casselton L (ASM Press, Washington, DC), Vol **88**, pp 142 - 147.

4. Kues U (2000) Life history and developmental processes in the basidiomycete *Coprinus cinereus*. *Microbiol Mol Biol Rev* 64(2):316-453.

5. James TY (2015) Why mushrooms have evolved to be so promiscuous: Insights from evolutionary and ecological patterns. *Fung Biol Rev* 29(3-4):167-178.

6. Nieuwenhuis BPS*, et al.* (2013) Evolution of uni- and bifactorial sexual compatibility systems in fungi. *Heredity* 111:445-455.

7. Giraud T, Yockteng R, Lopez-Villavicencio M, Refregier G, & Hood ME (2008) The mating system of the anther smut fungus, *Microbotryum violaceum*: selfing under heterothallism. *Euk Cell* 7:765-775.

8. Heitman J, Sun S, & James TY (2013) Evolution of fungal sexual reproduction. *Mycologia* 105(1):1-27.

9. Kües U, James T, & Heitman J (2011) The mycota XIV : Mating type in basidiomycetes: Unipolar, bipolar, and tetrapolar paterns of sexuality. *Evolution of fungi and fungal-like organisms*, eds Pöggeler S & Wöstemeyer J (Springer-Verlag, Berlin, Heidelberg Germany).

10. Bakkeren G & Kronstad JW (1994) Linkage of mating type loci distinguishes bipolar from tetrapolar mating in basidiomycetous smut fungi. *Proc Natl Acad Sci USA* 91:7085-7089.

11. Idnurm A, Hood ME, Johannesson H, & Giraud T (2015) Contrasted patterns in mating-type chromosomes in fungi: Hotspots versus coldspots of recombination. *Fung Biol Rev* 29(3-4):220-229.

12. Badouin H*, et al.* (2015) Chaos of rearrangements in the mating-type chromosomes of the anther-smut fungus *Microbotryum lychnidis-dioicae*. *Genetics* 200(4):1275-1284.

13. Fraser JA*, et al.* (2004) Convergent evolution of chromosomal sex-determining regions in the animal and fungal kingdoms. *PLoS Biol* 2:2243-2255.

14. Hood ME & Antonovics J (2000) Intratetrad mating, heterozygosity, and the maintenance of deleterious alleles in *Microbotryum violaceum (=Ustilago violacea)*. *Heredity* 85:231-241.

15. Menkis A, Jacobson DJ, Gustafsson T, & Johannesson H (2008) The mating-type chromosome in the filamentous ascomycete *Neurospora tetrasperma* represents a model for early evolution of sex chromosomes. *PLoS Genet* 4(3):e1000030.

16. Zakharov IA (1986) Some principles of the gene localization in eukaryotic chromosomes. Formation of the problem and analysis of nonrandom localization of the mating-type loci in some fungi. *Genetika* 22:2620-2624.

17. Zakharov IA (2005) Intratetrad mating and its genetic and evolutionary consequences. *Russ J Genet* 41:508-519.

18. Abbate JL & Hood ME (2010) Dynamic linkage relationships to the mating-type locus in automictic fungi of the genus *Microbotryum*. *J. Evol. Biol* 23(8):1800-1805.

19. Charlesworth D (2016) The status of supergenes in the 21st century: recombination suppression in Batesian mimicry and sex chromosomes and other complex adaptations. *Evol App* 9(1):74-90.

20. Antonovics J & Abrams JY (2004) Intratetrad mating and the evolution of linkage relationships. *Evolution* 58(4):702-709.

21. Johnson LJ, Antonovics J, & Hood ME (2005) The evolution of intratetrad mating rates. *Evolution* 59(12):2525-2532.

22. Ironside JE (2010) No amicable divorce? Challenging the notion that sexual antagonism drives sex chromosome evolution. *Bioessays* 32(8):718-726.

23. Kniep H (1919) Untersuchungen uber den Antherenbrand (*Ustilago violacea* Pers.). Ein Beitrag zum Sexualitatsproblem. *Z Bot* 11:257-284.

24. Hood ME (2002) Dimorphic mating-type chromosomes in the fungus *Microbotryum violaceum*. *Genetics* 160:457-461.

25. Hood ME, Antonovics J, & Koskella B (2004) Shared forces of sex chromosome evolution in haploids and diploids. *Genetics* 168:141-146.

26. Hood ME, Petit E, & Giraud T (2013) Extensive divergence between mating-type chromosomes of the anther-smut fungus. *Genetics* 193(1):309-315.

27. Fontanillas E, *et al.* (2015) Degeneration of the non-recombining regions in the mating type chromosomes of the anther smut fungi. *Mol Biol Evol* 32:928-943.

28. Votintseva AA & Filatov DA (2009) Evolutionary strata in a small mating-type-specific region of the smut fungus *Microbotryum violaceum*. *Genetics* 182(4):1391-1396.

29. Whittle CA, Votintseva A, Ridout K, & Filatov DA (2015) Recent and massive expansion of the mating-type specific region in the smut fungus *Microbotryum*. *Genetics* 199:809-816.

30. Menkis A, Whittle C, & Johannesson H (2010) Gene genealogies indicates abundant gene conversions and independent evolutionary histories of the mating-type chromosomes in the evolutionary history of *Neurospora tetrasperma*. *BMC Evol Biol* 10(1):234.

31. Sun S, Hsueh Y-P, & Heitman J (2012) Gene conversion occurs within the mating-type locus of *Cryptococcus neoformans* during sexual reproduction. *PLoS Genet* 8(7):e1002810.

32. Pandey RS & Azad RK (2016) Deciphering evolutionary strata on plant sex chromosomes and fungal mating-type chromosomes through compositional segmentation. *Plant Mol Biol* 90(4-5):359-373.

33. Hood ME & Antonovics JA (2004) Mating within the meiotic tetrad and the maintenance of genomic heterozygosity. *Genetics* 166:1751-1759.

34. Thomas A, Shykoff J, Jonot O, & Giraud T (2003) Sex-ratio bias in populations of the phytopathogenic fungus *Microbotryum violaceum* from several host species. *Int J Plant Sci* 164(4):641-647.

35. Hood ME, Scott M, & Hwang M (2015) Breaking linkage between mating compatibility factors: Tetrapolarity in *Microbotryum*. *Evolution* 69(10):2561-2572.

36. Koren S, *et al.* (2012) Hybrid error correction and de novo assembly of single-molecule sequencing reads. *Nat Biotechnol* 30:693-700.

37. Foissac S, *et al.* (2008) Genome annotation in plants and fungi: EuGène as amodel platform. *Current Bioinformatics* 3:87-97.

38. Consortium TU (2011) Ongoing and future developments at the Universal Protein Resource. *Nucleic Acids Res* 39:D214-D219.

39. Krzywinski MI, *et al.* (2009) Circos: An information aesthetic for comparative genomics. *Genom Res* 19:1639-1645.

40. Li L, Stoeckert CJ, & Roos D (2003) OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Research* 13:2178-2189.

41. van Dongen S (2000) Graph clustering by flow simulation. PhD (University of Utrecht, The Netherlands).

42. Edgar R (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32 1792-1797.

43. Abascal F, Zardoya R, & Telford M (2010) TranslatorX: multiple alignment of nucleotide sequences guided by amino acid translations. *Nucleic Acids Res* 38:W7-13.

44. Stamatakis A (2006) RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22:2688-2690.

45. Capella-Gutierrez S, Silla-Martinez JM, & Gabaldon T (2009) trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25:1972-1973.

46. Salichos L, Stamatakis A, & Rokas A (2014) Novel Information Theory-Based Measures for Quantifying Incongruence among Phylogenetic Trees. *Molyang Biol Evol* 31(5):1261-1271.

47. Bouckaert R, *et al.* (2014) BEAST 2: a software platform for Bayesian evolutionary analysis. *PLoS Comput Biol* 10:e1003537.

48. Gladieux P, *et al.* (2011) Maintenance of fungal pathogen species that are specialized to different hosts: allopatric divergence and introgression through secondary contact. *Mol Biol Evol* 28:459-471.

49. Smit A & Green P (RepeatMasker at http://www.repeatmasker.org/.

50. Jurka J (1998) Repeats in genomic DNA: mining and meaning. *Curr Opin Struct Biol* 8:333-337.

51. Badouin H, *et al.* (2017) Widespread selective sweeps throughout the genome of model plant pathogenic fungi and identification of effector candidates. *Mol Ecol* 26:2041-2062.

52. Oudemans P, *et al.* (1998) The distribution of mating type bias in natural populations of the anther smut *Ustilago violacea (=Microbotryum violaceum)* on *Silene alba (= S. latifolia)* in Virginia. *Mycologia* 90:372-381.

53. Luo R, *et al.* (2012) SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *GigaScience* 1(1):18.

54. Devier B, Aguileta G, Hood M, & Giraud T (2009) Ancient trans-specific polymorphism at pheromone receptor genes in basidiomycetes. *Genetics* 181:209–223.

55. Koboldt D, *et al.* (2012) VarScan 2: Somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genom Res* 22:568-576.

56. De Mita S & Siol M (2012) EggLib: processing, analysis and simulation tools for population genetics and genomics. *BMC Genetics* 13(1):27.

57. Huerta-Cepas J, Serra F, & Bork P (2016) ETE 3: Reconstruction, Analysis, and Visualization of Phylogenomic Data. *Mol Biol Evol* 33(6):1635-1638.

58. Wright SI & Charlesworth B (2004) The HKA test revisited: A maximum-likelihood-ratio test of the standard neutral model. *Genetics* 168(2):1071-1076.

59. Qiu S, Bergero R, & Charlesworth D (2013) Testing for the Footprint of Sexually Antagonistic Polymorphisms in the Pseudoautosomal Region of a Plant Sex Chromosome Pair. *Genetics* 194(3):663-+.

60. Wright SI, *et al.* (2006) Testing for effects of recombination rate on nucleotide diversity in natural populations of Arabidopsis lyrata. *Genetics* 174(3):1421-1430.

61. Ranwez V, Harispe S, Delsuc F, & Douzery E (2011) MACSE: Multiple alignment of coding sequences accounting for frameshifts and stop codons. *PLoS ONE* 6:e22594.

62. Vercken E, *et al.* (2010) Glacial refugia in pathogens: European genetic structure of anther smut pathogens on *Silene latifolia* and *S. dioica. PloS Path* 6:e1001229.

List of SI Figures

Fig. S1. Hypotheses for the generation of evolutionary strata on sex or mating-type chromosomes.

Fig. S2. Odds of compatibility among gametes of a diploid individual in basidiomycete fungi.

Fig. S3. Comparison of gene order between *Microbotryum lychnidis-dioicae* chromosomes.

Fig. S4. Synteny between a_1 and a_2 mating-type chromosomes of *Microbotryum lagerheimii*.

Fig. S5. Comparison of gene order between mating-type chromosomes.

Fig. S6. Examples of gene genealogies at non-mating type genes showing different levels of trans-specific polymorphism, i.e., with more or less ancient linkage to mating type.

Fig. S7. Cumulative percentage of genes showing trans-specific polymorphism across strata and age of trans-specific polymorphism relative to speciation events.

Fig. S8. Synonymous divergence between alleles associated with a_1 and a_2 mating types in the sequenced diploid individual in each of *Microbotryum silenes-dioicae* and *M. violaceum s. str.*, plotted along the ancestral gene order.

Fig. S9. Patterns of segregation of alleles at non-mating-type genes, associated with the a_1 and a_2 mating types, respectively, in gene genealogies using multiple resequenced genomes of *M. silenes-dioicae* (MvSd) and *M. lychnidis-dioicae* (MvSl).

Fig. S10. Genetic diversity in *Microbotryum lychnidis-dioicae* and *M. silenes-dioicae* ($\theta\pi$) based on multiple genomes.
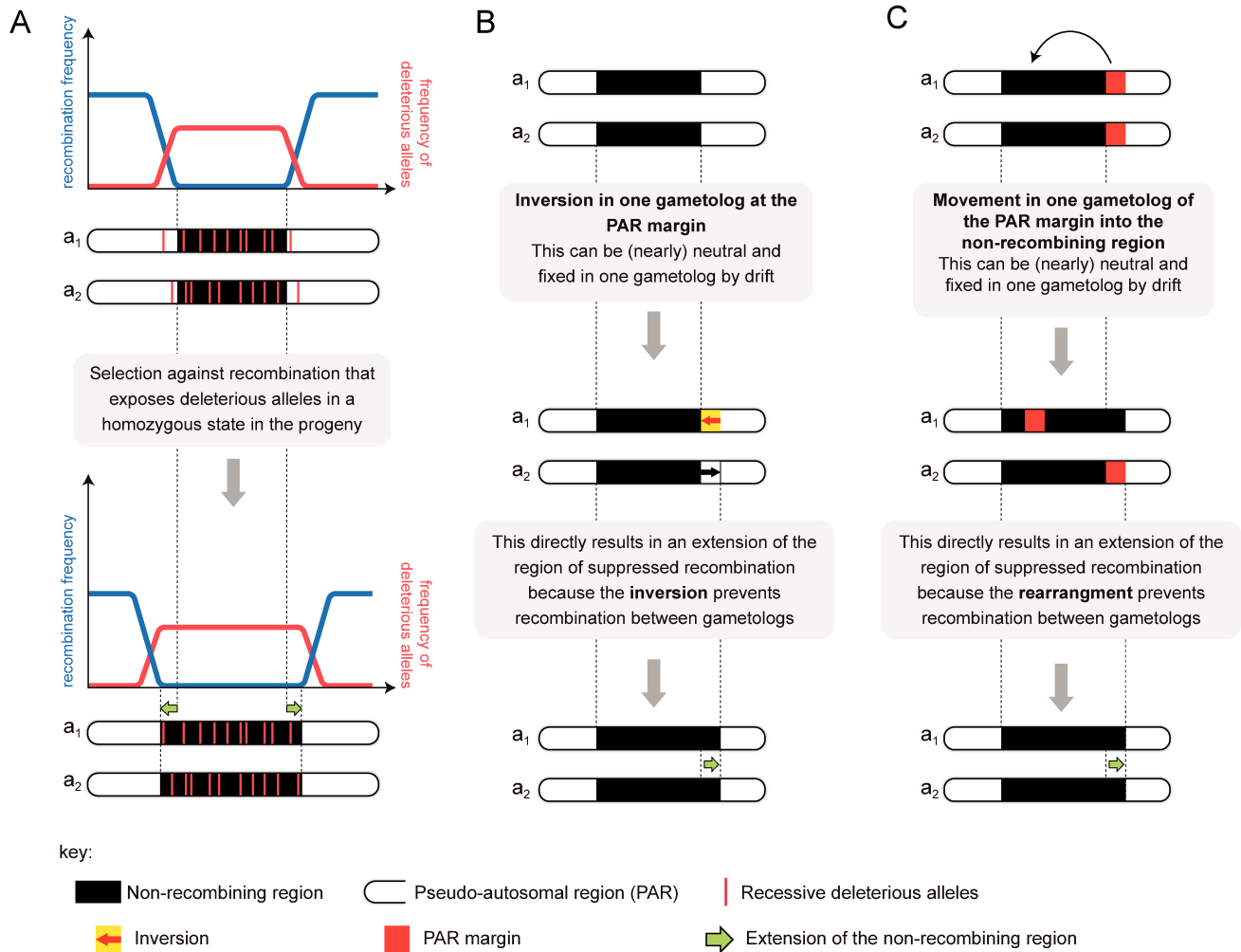
Fig. S1. **Hypotheses for the generation of evolutionary strata in sex or mating-type chromosomes. (A)** Expansion of linkage to the mating-type loci that encompasses partially linked genetic load loci, favoring permanent sheltering in a heterozygous state. This may occur if recombination frequency gradually decreases from the non-recombining region into the PAR, so that partial linkage (linkage disequilibrium) to mating-type or sex-determining genes makes selection against recessive deleterious alleles less efficient. This may allow deleterious alleles to increase in frequency in these PAR regions at the margin of the non-recombining region, further selecting for cessation of recombination. Rare recombination events would generate individuals homozygous for deleterious alleles and would therefore be selected against. Complete linkage of these PAR regions in linkage disequilibrium with mating type would be selected for. Alternative mechanisms for explaining suppressed recombination involve neutral inversions **(B)** or rearrangements **(C)** extending the region of suppressed recombination; these could be fixed by drift in one of the sex or mating-type chromosomes and arrest recombination between sex or mating-type chromosomes.

A	B	C
Unlinked mating-type loci	**Mating-type loci linked to centromeres on different chromosomes**	**Linked mating-type loci**
Tetrapolar breeding system		Bipolar breeding system

Haploids

Diploids

Meiosis

Segregation of mating-type loci in first meiotic division

Gametes

(TT)

a_1b_1 or a_2b_2 or a_1b_2 or a_2b_1

(PD)

a_1b_1 or a_1b_1 or a_2b_2 or a_2b_2

(NPD)

a_1b_2 or a_1b_2 or a_2b_1 or a_2b_1

(PD)
a_1b_1 a_1b_1 or a_2b_2 a_2b_2

(NPD)
a_1b_2 a_1b_2 a_2b_1 a_2b_1

b_1a_1 b_1a_1 b_2a_2 b_2a_2
a_1b_1 a_1b_1 a_2b_2 a_2b_2

2 or 4 mating types produced per meiosis	**2 mating types produced per meiosis** 4 mating types produced among different meioses	**2 mating types produced**
25% of compatibility between tetrads		50% of compatibility between tetrads
33% or 67% compatibility within tetrad	67% of compatibility within tetrad	67% of compatibility within tetrad

Key: Mating-type loci
- HD b_1
- HD b_2
- PR a_1
- PR a_2
- PR Chr.
- HD Chr.
- Fused HD/PR Chr.
- centromere
- suppression of recombination

Fig. S2. Odds of compatibility among gametes of a diploid individual in basidiomycete fungi. Gametes are fully compatible only if they carry different alleles at both mating-type loci, i.e., the PR (including pheromone receptor and pheromone genes, with a_1 and a_2 alleles) and HD (including homeodomain genes, with b_1 and b_2 alleles) loci. (A) Under tetrapolarity with PR and HD mating-type loci unlinked from each other and from the centromeres (shown here located in different chromosomes, blue and red, respectively), the percentage of compatibility of a given gamete among the other gametes produced by the same diploid individual is 25% across multiple meioses (a given gamete is compatible with one of every four gametes), and the percentage is 33% within tetrad (a given gamete is compatible with one of the other three gametes in the tetrad) or 67% (a given gamete is compatible with two of the three remaining gametes in the tetrad) depending on segregation of the mating type alleles. The different types of gametes produced are tetratypes (TT), parental ditypes (PD) or non-parental ditypes (NPD), which depends on allele segregation and on whether a crossing-over occurred between one of the two loci and the centromere. (B) Under tetrapolarity with PR and HD mating-type genes linked to the centromeres of different chromosomes (blue and red, respectively), the percentage of compatibility of a given gamete among the other gametes produced by the same diploid individual is 25% across multiple meioses but 67% within a tetrad (a given gamete is compatible with two of the three other gametes in the tetrad) due to the segregation of the variation occurring only at meiosis I for both mating type loci. The different types of gametes produced are parental ditypes (PD) or non-parental ditypes (NPD), which depends on segregation. (C) Under bipolarity, i.e., with HD and PR loci fully linked to each other on the same chromosome, the percentage of compatibility of a given gamete among the other gametes produced by the same diploid individual is 50% across multiple meioses (a given gamete is compatible with two of every four gametes), and 67% within a single meiotic tetrad (a given gamete is compatible with two of the three other gametes in the tetrad).
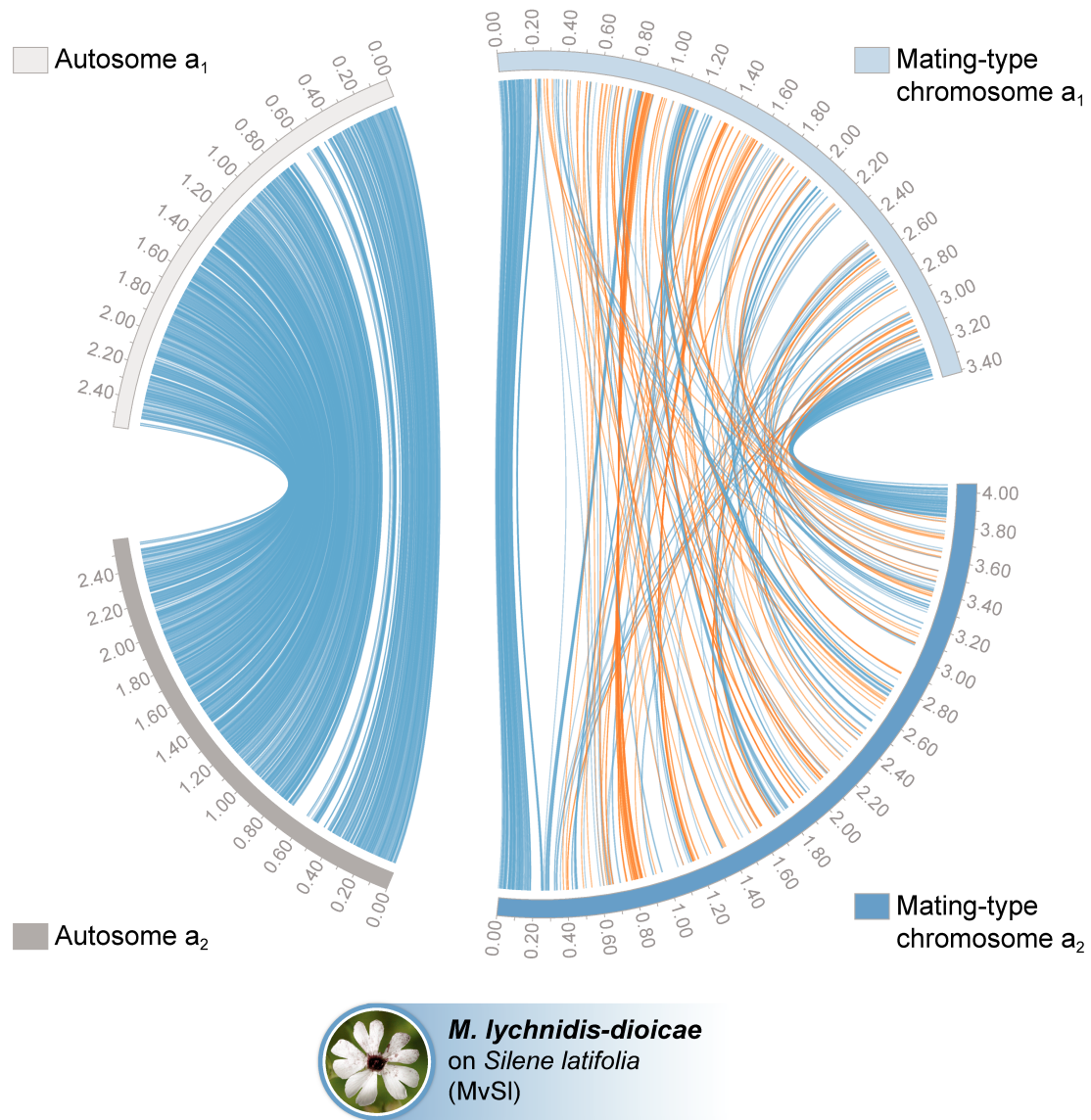
Fig. S3. **Comparison of gene order between *Microbotryum lychnidis-dioicae* chromosomes.** a_1 (top) and a_2 (bottom) mating chromosomes of *M. lychnidis-dioicae*; an autosome pair is shown at the left and the mating-type chromosomes at the right. Blue and orange lines link regions with collinearity across > 5 kb, with the orange links corresponding to inversions. The genomic regions without links correspond to highly rearranged regions. The recombining regions (the autosome and the pseudo-autosomal regions, PARs) are highly collinear between a_1 and a_2 chromosomes, despite a highly selfing mating system, while the non-recombining region on the mating-type chromosome is highly rearranged.
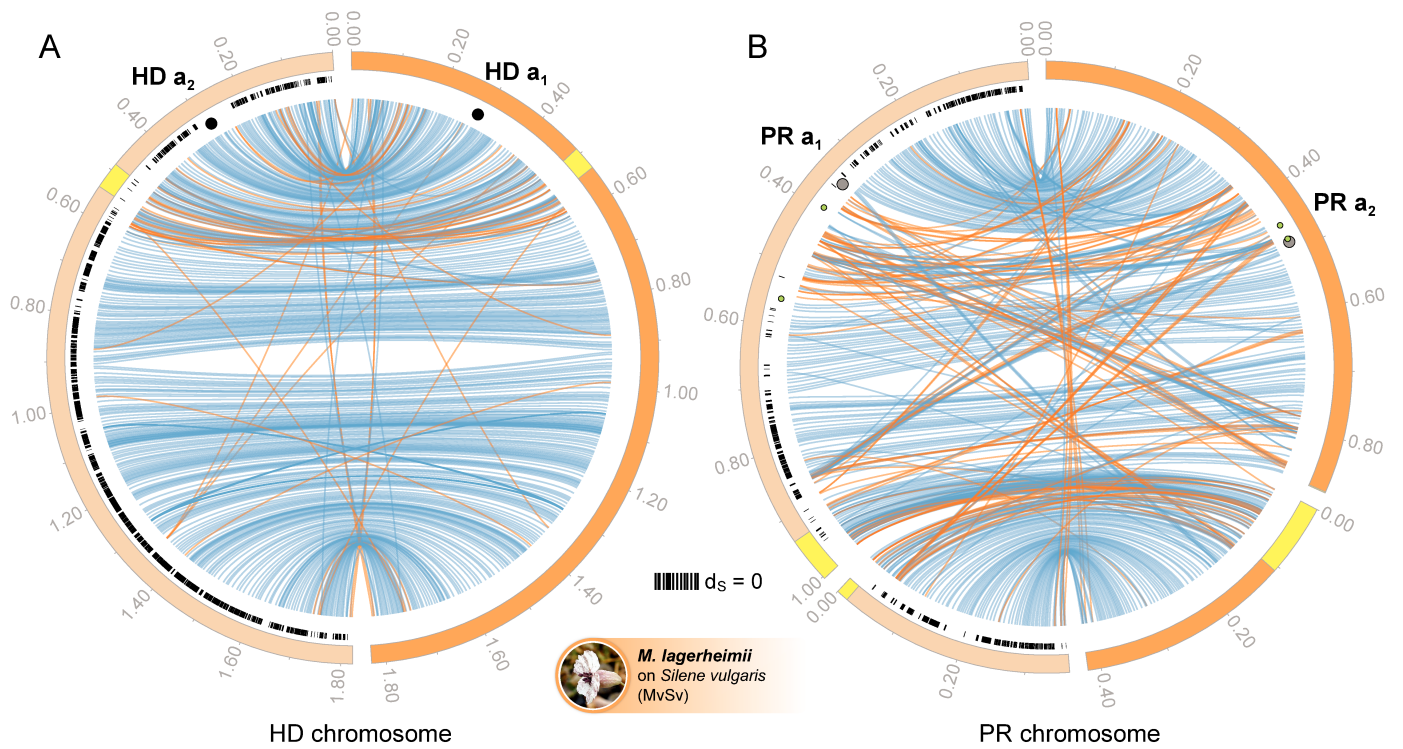
Fig. S4. **Synteny between a₁ and a₂ mating-type chromosomes of _Microbotryum lagerheimii_.** The homeodomain (HD) mating-type chromosomes are displayed in the left circle and the pheromone receptor (PR) mating-type chromosomes in the right circle. The HD genes are indicated by black circles, the pheromone receptor gene by grey circles and the two pheromone genes by small green circles. Blue and orange lines link regions with collinearity across > 1 kb, with the orange links corresponding to inversions. The genomic region between the pheromone receptor and pheromone genes in the a₁ mating-type chromosomes is rearranged compared to the a₂ mating-type chromosomes. Black marks along the contig circles indicate genes with no synonymous differentiation between a₁ and a₂ alleles within the sequenced individual (d$_S$ = 0). Yellow regions on the outer track indicate centromere-specific repeats (12).
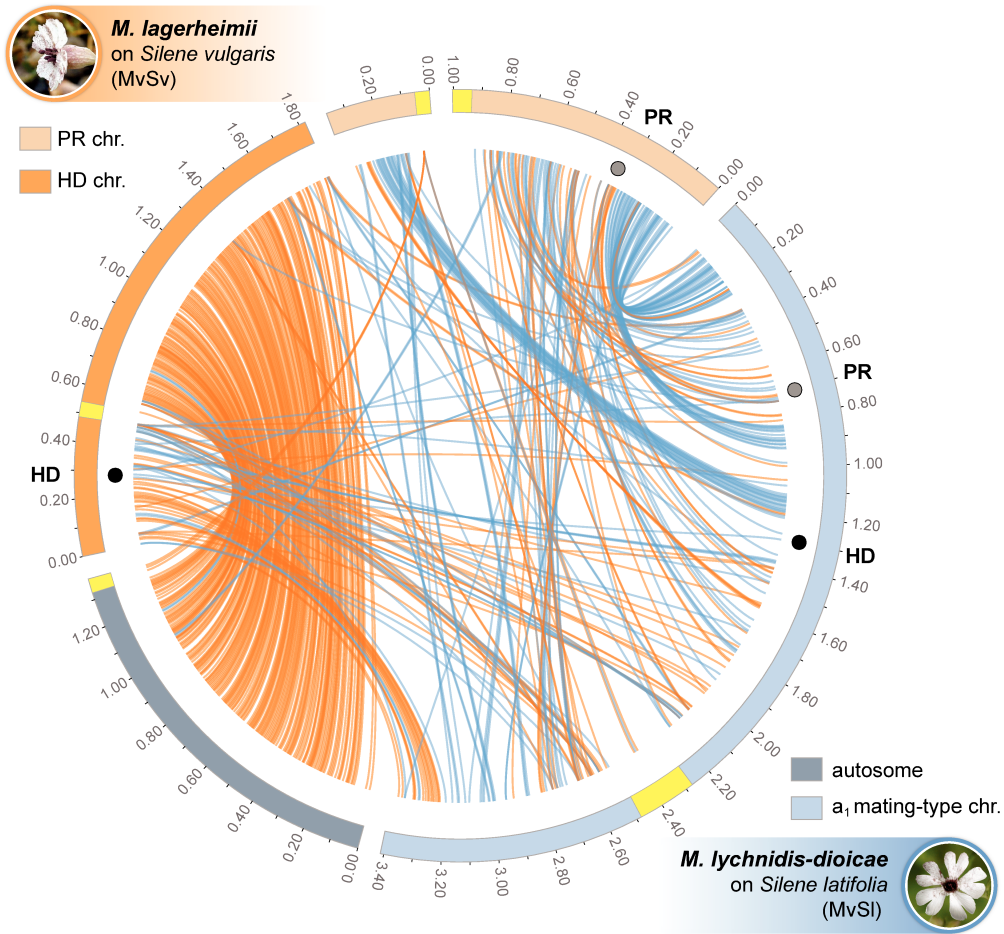
Fig. S5. Comparison of gene order between mating-type chromosomes. In all panels (A-E), the outer tracks illustrate assembled chromosomes/contigs and are color coded by species as in Fig. 1. The positions of HD and PR genes are indicated by the black and grey small circles, respectively. Yellow regions on the outer track indicate the centromere-specific repeats (12) and, when represented, red ("R") and green ("G") regions indicate the genomic location of the red and green genes as described in Fig. 4B. **(A)** Comparison between the a_1 mating-type chromosomes (chr.) of *Microbotryum lagerheimii* (orange contigs, with light orange for the PR mating-type chromosome and dark orange for the HD mating-type chromosome) and the mating-type chromosome of *M. lychnidis-dioicae* (light blue contig), as well as an autosomal contig of *M. lychnidis-dioicae* (dark blue contig). Blue and orange lines link regions with collinearity across > 2 kb, with orange corresponding to inversions. Genomic regions without links correspond to highly rearranged regions.
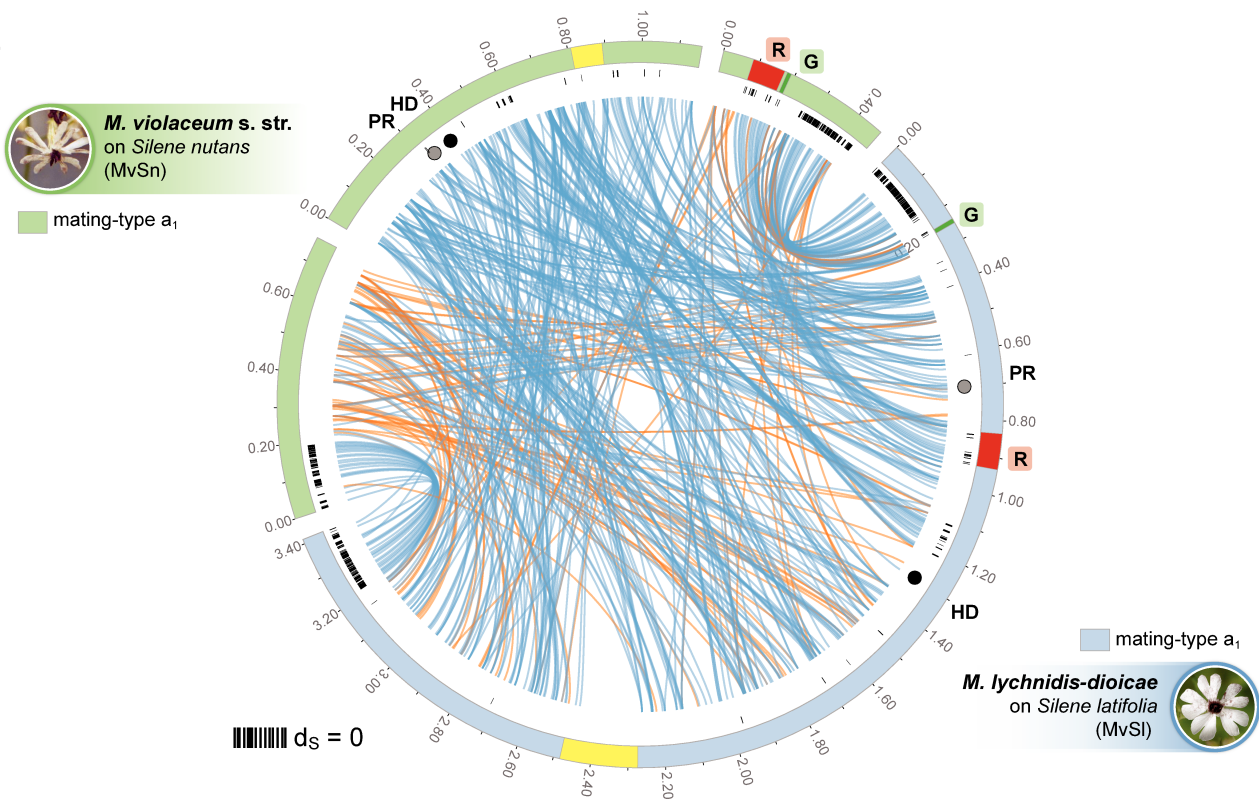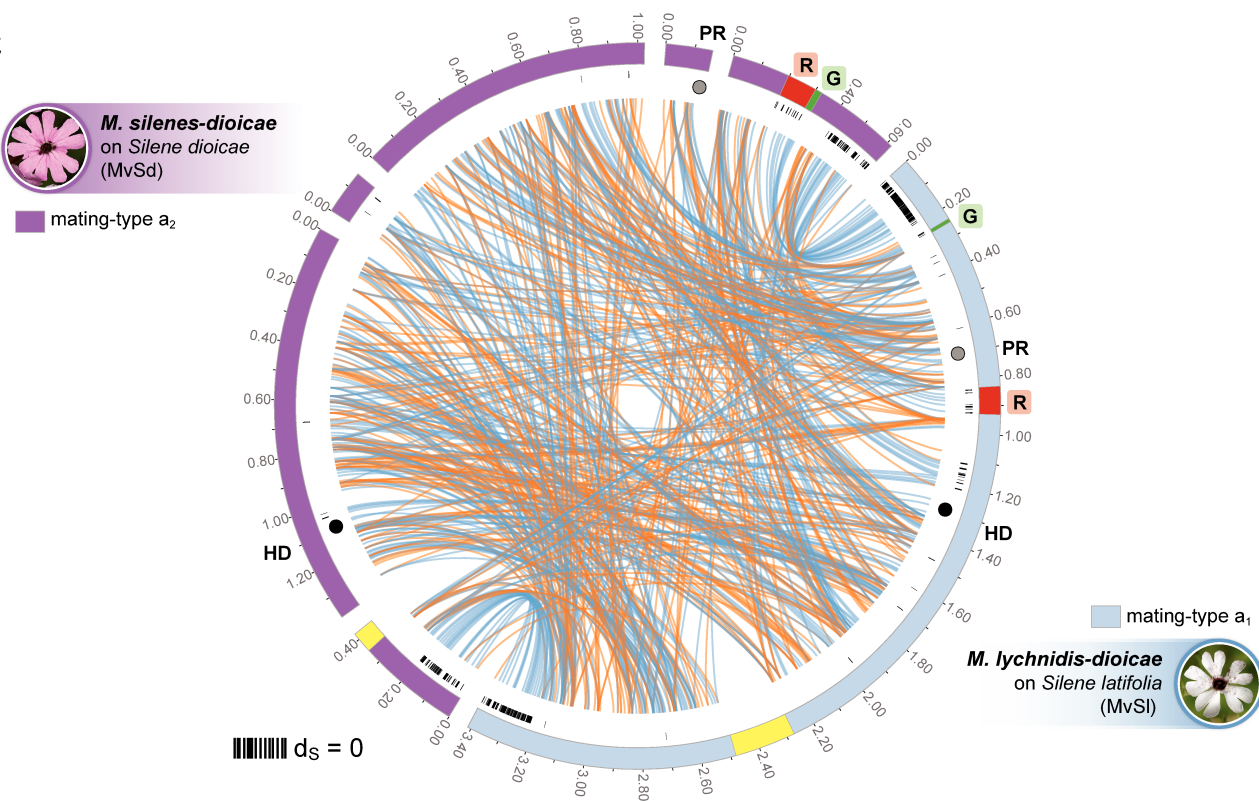
Fig. S5. Continued. (B) Comparison of a₁ mating-type chromosomes of *Microbotryum lychnidis-dioicae* (right, light blue contig) and *M. violaceum s. str.* (left, light green contigs). It should be noted that the orientation of the central contig in the a₁ *M. violaceum s. str.* mating-type chromosome could not be assessed, so all its links were drawn blue. The black marks along the contig circle indicate genes that have no synonymous substitution between a₁ and a₂ alleles (d$_S$ = 0) in *M. lychnidis-dioicae*. **(C)** Comparison between the a₁ mating-type chromosomes of *M. silenes-dioicae* (dark purple contigs on the left) and *M. lychnidis-dioicae* (light blue contig on the right). Blue and orange lines link regions with collinearity across > 3 kb, with orange corresponding to inversions (note that the *M. silenes-dioicae* central contigs could not be ordered or oriented). All the remaining features are as described in panel (A).
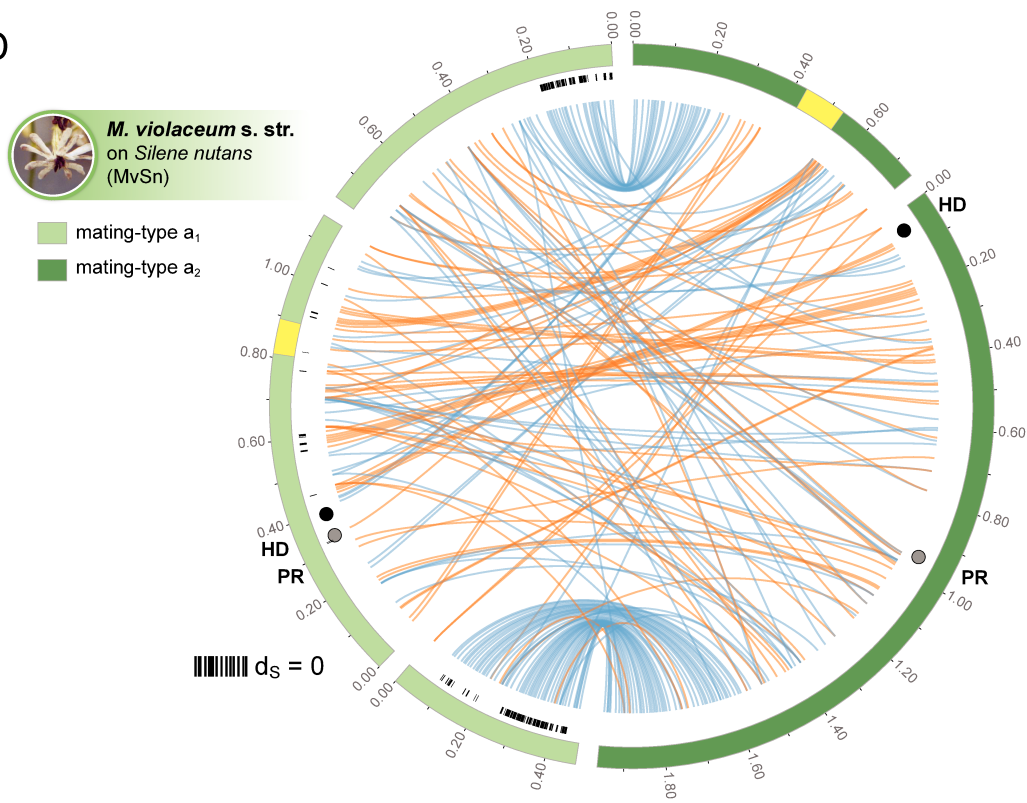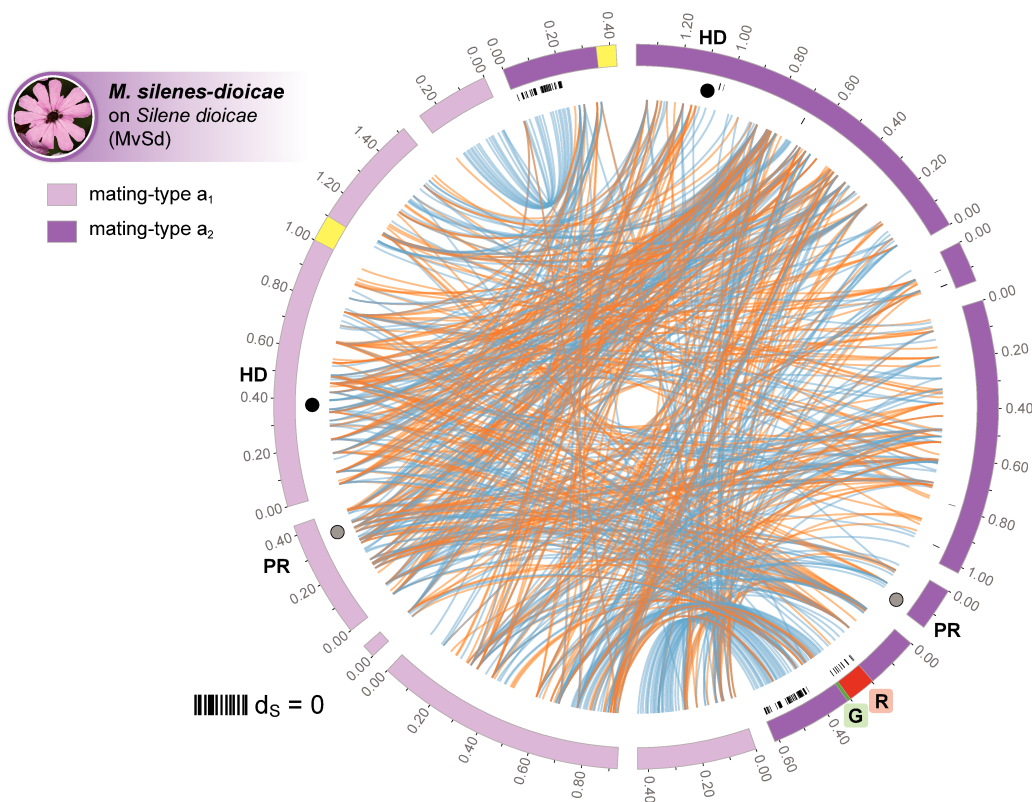
Fig. S5. Continued. **(D)** Comparison of a_1 (left, light green) and a_2 (right, dark green) mating-type chromosomes of *M. violaceum s. stricto*. Blue and orange lines link regions with collinearity across > 3 kb, with orange corresponding to inversions. It should be noted that the orientation of the central contig in the a_1 mating-type chromosome could not be assessed. The black marks along the contig circle indicate the genes with no synonymous substitutions between a_1 and a_2 alleles (d_S=0) in *M. violaceum s. stricto*. **(E)** Comparison of a_1 (left, light purple) and a_2 (right, dark purple) mating-type chromosomes of *M. silenes-dioicae*. Blue and orange lines link regions with collinearity across >3 kb, with orange corresponding to inversions. Note that the central contigs could not be ordered or oriented. The location of the red ("R") and green ("G") genes in Fig. 4B are indicated in the a_2 genome. The black marks along the contig circle indicate the genes with no synonymous substitutions between a_1 and a_2 alleles (d_S = 0). The remaining features are as described in panel (A).
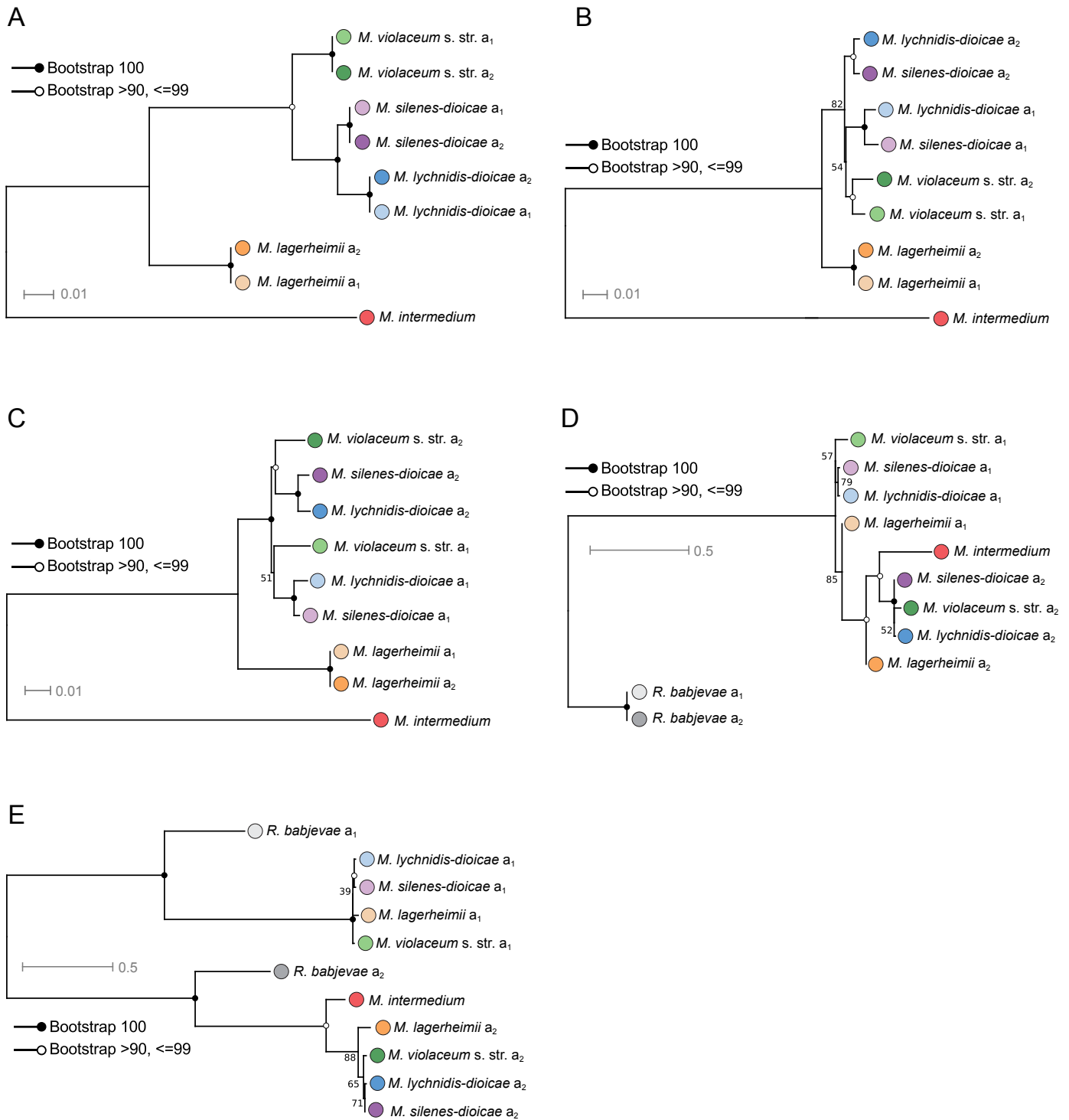
Fig. S6. Examples of gene genealogies at non-mating type genes showing different levels of trans-specific polymorphism, i.e., with more or less ancient linkage to mating type. Bootstraps are shown at nodes and the branch length scale is indicated. Strains of the same species are indicated by circles of the same color (a_1, lighter variant; a_2, darker variant) and according to Fig. 1. **(A)** No trans-specific polymorphism (orthogroup belonging to the *M. lychnidis-dioicae* PAR). **(B)** Trans-specific polymorphism across *M. lychnidis-dioicae* and *M. silenes-dioicae* (orthogroup belonging to the *M. lychnidis-dioicae* black stratum). **(C)** Trans-specific polymorphism across *M. lychnidis-dioicae, M. silenes-dioicae* and *M. violaceum* s. str. (orthogroup belonging to the *M. lychnidis-dioicae* black stratum). **(D)** Trans-specific polymorphism across *M. lychnidis-dioicae, M. silenes-dioicae, M. violaceum s. str.* and *M. lagerheimii* (orthogroup belonging to the *M. lychnidis-dioicae* PR stratum). **(E)** Trans-specific polymorphism across *M. lychnidis-dioicae, M. silenes-dioicae, M. violaceum s. str. M. lagerheimii* and *Rhodotorula babjevae* (orthogroup belonging to the *M. lychnidis-dioicae* orange stratum).
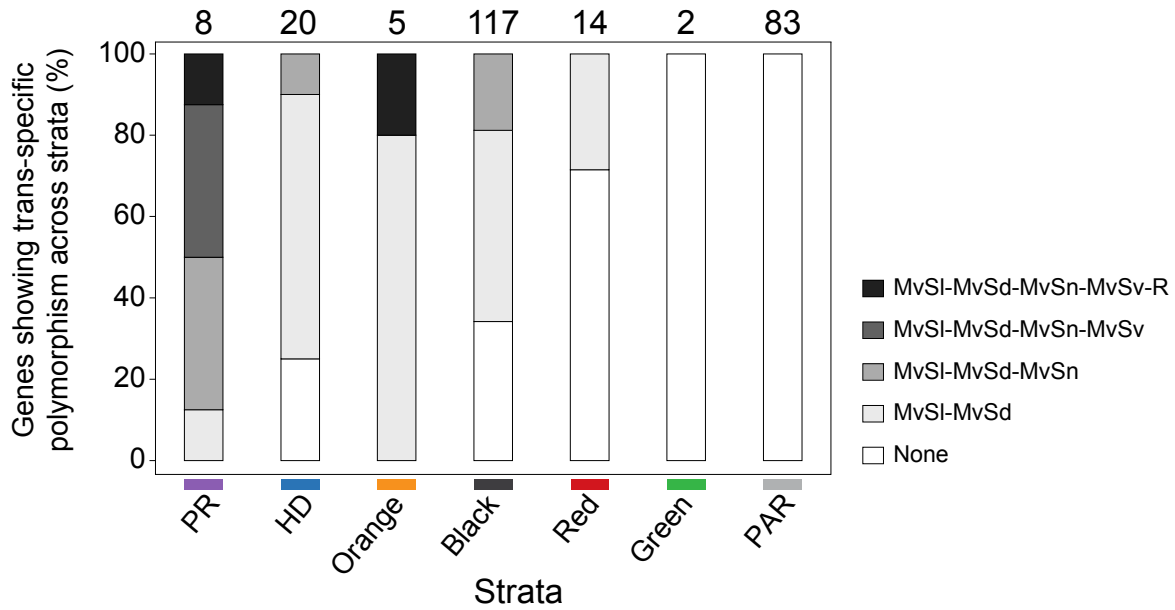
23

Fig. S7. **Cumulative percentage of genes showing trans-specific polymorphism across strata and age of trans-specific polymorphism relative to speciation events.** Black - single-copy genes with deep trans-specific polymorphism including all species, *M. lychnidis-dioicae* (MvSl), *M. silenes-dioicae* (MvSd), *M. violaceum* s. str. (MvSn), *M. lagerheimii* (MvSv), and even *Rhodotorula babjevae* (R); dark grey - genes showing trans-specific polymorphism including only *M. lychnidis-dioicae, M. silenes-dioicae, M. violaceum* s. str., and *M. lagerheimii*; grey - genes showing trans-specific polymorphism with only *M. lychnidis-dioicae, M. silenes-dioicae* and *M. violaceum s. str.;* light grey - genes with trans-specific polymorphism only between *M. lychnidis-dioicae* and *M. silenes-dioicae*; white - genes with no trans-specific polymorphism. Numbers on the top indicate genes for which gene genealogies with all species and alleles were computed. See Figure S6 for examples of gene genealogies with different levels of trans-specific polymorphism.
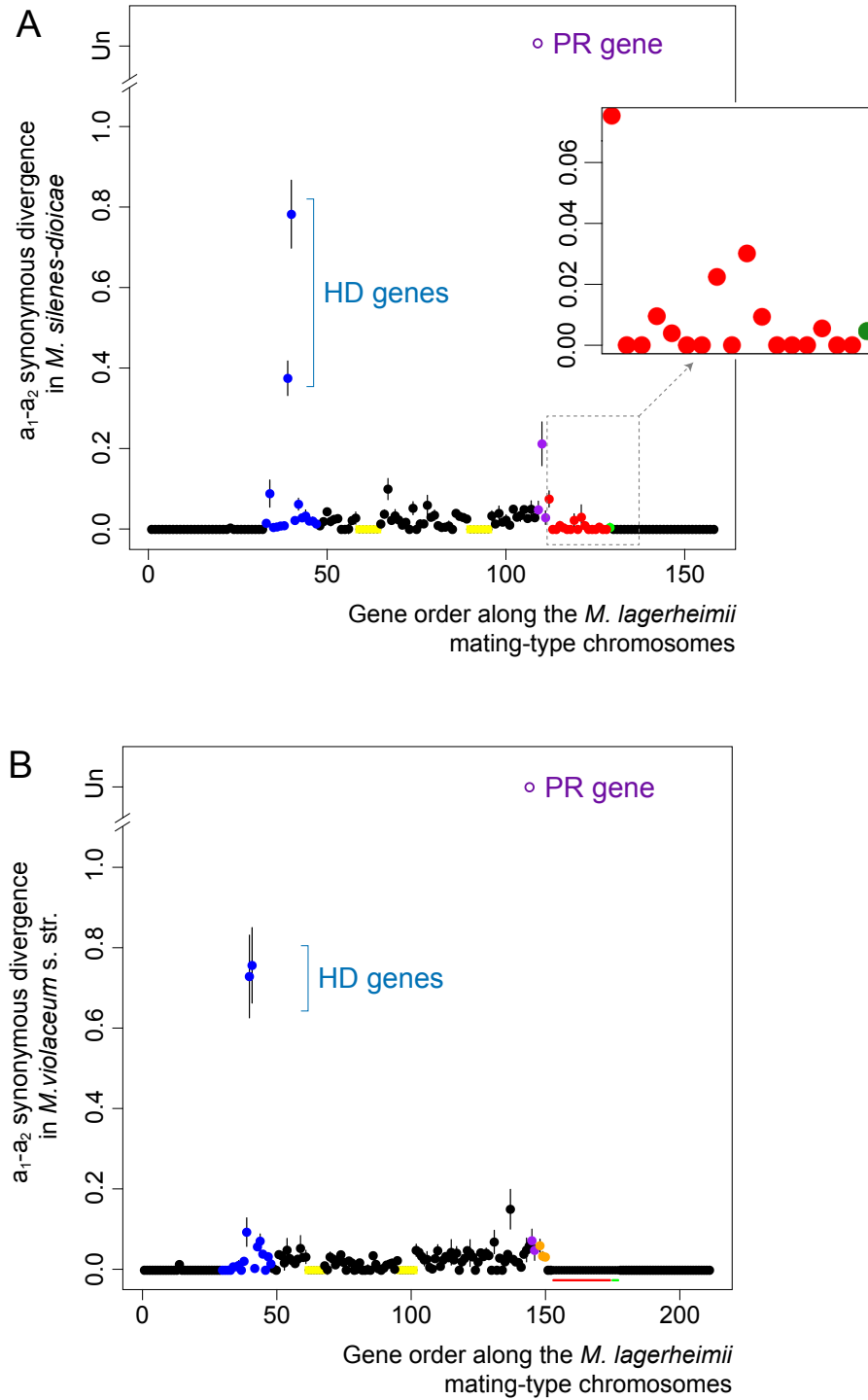
Fig. S8. Synonymous divergence between alleles associated with a₁ and a₂ mating types in the sequenced diploid individual in each of *Microbotryum silenes-dioicae* and *M. violaceum s. stricto,* plotted along the ancestral gene order. Synonymous divergence between a_1-a_2 in *M. silenes-dioicae* **(A)** and *M. violaceum s. str.* **(B)** plotted in the ancestral genomic coordinates, as inferred from the *M. lagerheimii* a_1 mating-type chromosome. In both panels, points are mean d_S values per gene and error bars the standard errors. All non-transposable element genes shared by the mating-type chromosomes are depicted. Divergence between the a_1 and a_2 pheromone receptor (PR) was too extensive (54) and could not be computed in both species (it is plotted as an "unalignable" open circle to ease visualization of the various strata). The positions of the inferred ancestral centromeres are indicated by yellow boxes. Genes with non-zero d_S around the PR and HD mating-type loci within the sequenced individual of *M. lagerheimii* are colored in purple and blue, respectively. More recent putative evolutionary strata in *M. silenes-dioicae* (panel A) are indicated in red and green, and zoomed in on the inset shown on the upper right. In panel (B), the genes belonging to more recent evolutionary strata in *M. silenes-dioicae* and *M. lychnidis-dioicae* are indicated by red and green bars.

Fig. S9. Patterns of segregation of alleles at non-mating-type genes, associated with the a_1 and a_2 mating types, respectively, in gene genealogies using multiple resequenced genomes of *M. silenes-dioicae* (MvSd) and *M. lychnidis-dioicae* (MvSl). (A) Percentage (%) of gene genealogies showing different levels of segregation of a_1 and a_2 alleles: white - gene genealogies with mixed a_1 and a_2 *M. silenes-dioicae* and *M. lychnidis-dioicae* alleles; light grey - gene genealogies with a_1 and a_2 alleles separated in *M. silenes-dioicae* or *M. lychnidis-dioicae*, but not in both species; dark grey - gene genealogies with a_1 and a_2 alleles separated in both *M. silenes-dioicae* or *M. lychnidis-dioicae,* but without trans-specific polymorphism across the two species; black - gene genealogies with full trans-specific polymorphism. The numbers at the top indicate the number of genes for which information was available. **(B-E)** Examples of gene genealogies showing different levels of segregation of alleles at non-mating-type genes, associated with the a_1 and a_2 mating types, respectively (shown in lighter and darker color variants, respectively) based on the analyses of multiple *M. silenes-dioicae* (MvSd, purple) and *M. lychnidis-dioicae* (MvSl, blue) genomes. Alleles of *M. lagerheimii* (MvSv, orange), *M. violaceum s. str.* (MvSn, green), *M. intermedium* (Minter, red) and *Rhodotorula babjevae* (sfc, gray) are also included; **(B)** gene genealogy with full trans-specific polymorphism (red statum); **(C)** alleles at non-mating-type genes, associated with the a_1 and a_2 mating types, respectively, separated in both *M. silenes-dioicae* and *M. lychnidis-dioicae,* but without trans-specific polymorphism across the two species (red stratum); **(D)** gene genealogy with alleles at non-mating-type genes, associated with the a_1 and a_2 mating types, respectively, separated in *M. lychnidis-dioicae* only (red stratum); **(E)** intermingled alleles at non-mating-type genes, associated with the a_1 and a_2 mating types, respectively, of both *M. silenes-dioicae* and *M. lychnidis-dioicae* (pseudo-autosomal region, PAR).

Fig. S10. *Microbotryum lychnidis-dioicae* **and** *M. silenes-dioicae* **genetic diversity (θπ) based on multiple genomes.** Diversity (±SE) was estimated in a_1 or a_2 mating type separately, in the pseudo-autosomal regions (PARs) and the different evolutionary strata (green, red, black, orange, HD-proximal and PR-proximal). *M. silenes-dioicae* is known to display lower genetic diversity than *M. lychnidis-dioicae* (48, 51, 62).

List of SI Tables

Table S1. Synonymous substitution (d_S) means within diploid individuals between alleles associated to the a_1 and a_2 mating types for genes in the different putative strata and the pseudo-autosomal regions (PARs), number of genes and estimated dates for recombination suppression.

Table S2. Synonymous substitution (d_S) between alleles at non-mating-type genes, associated with the a_1 and a_2 mating types, respectively, in three outgroup species for the genes belonging to the purple and blue strata in *Microbotryum lagerheimii*.

Table S3. Gene ID and putative function of the genes located in the different evolutionary strata in *Microbotryum lychnidis-dioicae* (MvSl) (shared with *M. lagerheimii* (MvSv) mating-type chromosomes).

Table S4. Raw data summary and assembly statistics for the genomes generated in this study.

Table S5. *Microbotryum lychnidis-dioicae* and *M. silenes-dioicae* strains used for performing polymorphism analyses.

Table S6. Nucleotide diversity (θ_π) and divergence (mean number of pairwise differences) estimates in pseudo-autosomal regions and evolutionary strata in *M. lychnidis-dioicae* and *M. silenes-dioicae*.

Table S7. Results of maximum-likelihood Hudson-Kreitman-Agade (MLHKA) tests for assessing whether differences in diversity between each evolutionary stratum and the pseudo-autosomal regions (PARs) is due to balancing selection or elevated mutation rates.

Table S1. Synonymous substitution (d_S) means within diploid individuals between alleles associated to the a1 and a2 mating types for genes in the different putative strata and the pseudo-autosomal regions (PARs), number of genes and estimated dates for recombination suppression. Mean d_S and standard error (SE) across genes between the large putative strata, numbered in their presumed order of origin of suppressed recombination, whether they involve mating-type genes, the *Microbotryum* species in which recombination is suppressed in these strata, the estimated date of recombination suppression in strata where inference was possible, and number of genes ancestrally in *M. lychnidis-dioicae* in the strata (many genes have been lost from one or the other mating types following recombination suppression).

Stratum order	Stratum	Mean d_S ± SE	Involving mating-type genes?	Species in which recombination is suppressed	Estimated date recombination suppression	Number of genes ancestrally in *M. lychnidis-dioicae*
1	PR mating-type gene	Not alignable	Yes	All species	370 MYA (54)	3
2	HD mating-type genes	0.62 ± 0.326	Yes	All species		2
3	PR-proximal genes (purple)	0.112 ± 0.040	No	All species	2.1 MYA	31
4	HD-proximal genes (blue)	0.103 ± 0.045	No	All species		14
5	Black (linking PR and HD mating-type loci)	0.032 ± 0.002	Yes	*M. lychnidis-dioicae, M. silenes-dioicae* and *M. violaceum s. str.*	1.3 MYA	85
6	Orange	0.052 ± 0.015	No	*M. lychnidis-dioicae, M. silenes-dioicae* and *M. violaceum s. str.*		*7*
7	Red	0.018 ± 0.049	No	*M. lychnidis-dioicae* and *M. silenes-dioicae*	0.9 MYA	22
8	Green	0.004 ± 0.003	No	*M. lychnidis-dioicae* and *M. silenes-dioicae*		3
	pPAR (from the ancestral HD chromosome)	<0.0001	No	None		73
	qPAR (from the ancestral PR chromosome)	<0.0001	No	None		80

Table S2. Synonymous substitution (d_S) between alleles associated with the a_1 and a_2 mating types in three individuals of outgroup species for the genes belonging to the *Microbotryum lagerheimii* purple and blue strata. (A) *Sporobolomyces salmonicolor* (Ss) strains CBS 6832 vs. ML 2241; (B) *Rhodotorula toruloides* (Rt) strains JCM 10020 (A1B1) vs. JCM 10021 (A2B11); (C) *Leucosporidium scottii* (Ls) strains CBS 5930 vs. CBS 5931. The ID of the query gene from the reference *M. lychnidis-dioicae* genome, the ID in the outgroup genome, the d_S value and the putative function inferred from similarity are given. The only functions associated with mating or mating-type functions are the mating-type genes (highlighted in bold). The pheromone receptor genes are not shown as they could not be aligned between mating types.

(A)

MvSI-Query ID	Paired ID in the outgroup	d_S	Functional annotation
MvSI-1064-A1-R4_A1g00025	Ss_CBS6832|g3490.t1;Ss_ML2241|g967.t1	0.0253	glycoside hydrolase family 28 protein
MvSI-1064-A1-R4_A1g00175	Ss_CBS6832|g4453.t1;Ss_ML2241|g4966.t1	0.0177	ABC1-A atypical protein kinase
MvSI-1064-A1-R4_A1g00230	Ss_CBS6832|g4436.t1;Ss_ML2241|g4982.t1	0.0045	ASF1-like histone chaperone
MvSI-1064-A1-R4_A1g00478	Ss_CBS6832|g2690.t1;Ss_ML2241|g5108.t1	0.0167	protein disulfide-isomerase
MvSI-1064-A1-R4_A1g00495	Ss_CBS6832|g6519.t1;Ss_ML2241|g6611.t1	0.0232	nucleotide-binding protein/RRM_SF
MvSI-1064-A1-R4_A1g00591	**Ss_CBS6832|g5519.t1;Ss_ML2241|g1781.t1**	**0.3037**	**homeodomain transcription factor HD1**
MvSI-1064-A1-R4_A1g00592	**Ss_CBS6832|g5518.t1;Ss_ML2241|g1780.t1**	**0.3041**	**homeodomain transcription factor HD2**
MvSI-1064-A1-R4_A1g00593	Ss_CBS6832|g2692.t1;Ss_ML2241|g5106.t1	0.0223	5'-3' exoribonuclease 2
MvSI-1064-A1-R4_A1g00659	Ss_CBS6832|g307.t1;Ss_ML2241|g1640.t1	0.0285	alpha-mannosyltransferase, glycosyltransferase family 71 protein
MvSI-1064-A1-R4_A1g00726	Ss_CBS6832|g4454.t1;Ss_ML2241|g4962.t1	0.0173	defective in Cullin neddylation protein 1/UBA_like_SF
MvSI-1064-A1-R4_A1g00772	Ss_CBS6832|g2693.t1;Ss_ML2241|g5105.t1	0.0260	U3 snoRNP-associated protein Esf2 / RRM_ABT1_like
MvSI-1064-A1-R4_A1g00800	Ss_CBS6832|g2586.t1|corr;Ss_ML2241|g6295.t1|corr	0.0368	S-(hydroxymethyl)glutathione dehydrogenase / alcohol dehydrogenase
MvSI-1064-A1-R4_A1g00801	Ss_CBS6832|g6218.t1;Ss_ML2241|g6054.t1	0.0308	exosome complex protein LRP1
MvSI-1064-A1-R4_A1g00844	Ss_CBS6832|g3596.t1;Ss_ML2241|g741.t1	0.0609	hypothetical protein
MvSI-1064-A1-R4_A1g00854	Ss_CBS6832|g4447.t1;Ss_ML2241|g4977.t1	0.0555	karyopherin kap95
MvSI-1064-A1-R4_A1g00970	Ss_CBS6832|g2670.t1;Ss_ML2241|g3808.t1	0.0287	T-complex protein 1 subunit beta
MvSI-1064-A1-R4_A1g00971	Ss_CBS6832|g2640.t1;Ss_ML2241|g6034.t1	0.0593	dolichyldiphosphatase / PAP2_like
MvSI-1064-A1-R4_A1g00972	Ss_CBS6832|g2641.t1;Ss_ML2241|g6035.t1	0.0449	hypothetical protein
MvSI-1064-A1-R4_A1g00997	Ss_CBS6832|g3124.t1;Ss_ML2241|g781.t1	0.0024	hypothetical protein
MvSI-1064-A1-R4_A1g00998	Ss_CBS6832|g2641.t1|corr;Ss_ML2241|g6036.t1|corr	0.0293	U3 small nucleolar ribonucleoprotein IMP3 / Ribosomal protein S4/S9
MvSI-1064-A1-R4_A1g01017	Ss_CBS6832|g4458.t1;Ss_ML2241|g4960.t1	0.0349	phosphatidylethanolamine N-methyltransferase (PEMT)
MvSI-1064-A1-R4_A1g01154	Ss_CBS6832|g2758.t1;Ss_ML2241|g2001.t1	0.0684	ribosomal protein L1
MvSI-1064-A1-R4_A1g01159	Ss_CBS6832|g5531.t1;Ss_ML2241|g1794.t1	0.0478	bZIP transcription factor (AP-1)
MvSI-1064-A1-R4_A1g01160	Ss_CBS6832|g2542.t1;Ss_ML2241|g1752.t1	0.0300	COP9 signalosome complex subunit 2
MvSI-1064-A1-R4_A1g01161	Ss_CBS6832|g2659.t1;Ss_ML2241|g3820.t1	0.0212	mitochondrial carrier protein
MvSI-1064-A1-R4_A1g01162	Ss_CBS6832|g5512.t1;Ss_ML2241|g1774.t1	0.0186	hypothetical protein
MvSI-1064-A1-R4_A1g01163	Ss_CBS6832|g2660.t1;Ss_ML2241|g3819.t1	0.0627	ADP-ribosylation factor-like protein 2
MvSI-1064-A1-R4_A1g01388	Ss_CBS6832|g1143.t1;Ss_ML2241|g6591.t1	0.0183	ribosome assembly protein Noc2
MvSI-1064-A1-R4_A1g01440	Ss_CBS6832|g2708.t1;Ss_ML2241|g5917.t1	0.0458	autophagy-related protein 7 / E1-like protein-activating enzyme Gsa7p/Apg7p
MvSI-1064-A1-R4_A1g01441	Ss_CBS6832|g2773.t1;Ss_ML2241|g1987.t1	0.0234	zinc finger, C3HC4 RING-type protein
MvSI-1064-A1-R4_A1g01457	Ss_CBS6832|g2652.t1;Ss_ML2241|g3827.t1	0.0269	unnamed protein product
MvSI-1064-A1-R4_A1g01458	Ss_CBS6832|g2652.t1;Ss_ML2241|g3827.t1	0.0269	DNA-directed RNA polymerase III subunit RPC1

Table S2. Continued.

(B)

MvSI-Query ID	Paired ID in the outgroup	dS	Functional annotation
MvSI-1064-A1-R4_A1g00175	Rt_JCM10020_06889-R1;Rt_JCM10021_05528-R1	0.338	ABC1-A atypical protein kinase
MvSI-1064-A1-R4_A1g00230	Rt_JCM10020_06874-R1;Rt_JCM10021_05544-R1	0.382	ASF1-like histone chaperone
MvSI-1064-A1-R4_A1g00478	Rt_JCM10020_01934-R1;Rt_JCM10021_07790-R1	0.618	protein disulfide-isomerase
MvSI-1064-A1-R4_A1g00495	Rt_JCM10020_01942-R1;Rt_JCM10021_02138-R1	0.363	nucleotide-binding protein/RRM_SF
MvSI-1064-A1-R4_A1g00591	**Rt_JCM10020_02036-R1;Rt_JCM10021_02036-R1**	**0.692**	**homeodomain transcription factor HD1**
MvSI-1064-A1-R4_A1g00592	**Rt_JCM10020_02035-R1;Rt_JCM10021_02037-R1**	**0.648**	**homeodomain transcription factor HD2**
MvSI-1064-A1-R4_A1g00593	Rt_JCM10020_01962-R1;Rt_JCM10021_02108-R1	0.423	5'-3' exoribonuclease 2
MvSI-1064-A1-R4_A1g00659	Rt_JCM10020_01652-R1;Rt_JCM10021_00547-R1	0.357	alpha-mannosyltransferase, glycosyltransferase family 71 protein
MvSI-1064-A1-R4_A1g00726	Rt_JCM10020_06891-R1;Rt_JCM10021_05538-R1	0.210	defective in Cullin neddylation protein 1/UBA_like_SF
MvSI-1064-A1-R4_A1g00772	Rt_JCM10020_01963-R1;Rt_JCM10021_02107-R1	0.554	U3 snoRNP-associated protein Esf2 / RRM_ABT1_like
MvSI-1064-A1-R4_A1g00800	Rt_JCM10020_02223-R1;Rt_JCM10021_01846-R1	0.361	S-(hydroxymethyl)glutathione dehydrogenase / alcohol dehydrogenase
MvSI-1064-A1-R4_A1g00801	Rt_JCM10020_01998-R1;Rt_JCM10021_02076-R1	0.334	exosome complex protein LRP1
MvSI-1064-A1-R4_A1g00854	Rt_JCM10020_06883-R1;Rt_JCM10021_05536-R1	0.195	karyopherin kap95
MvSI-1064-A1-R4_A1g00970	Rt_JCM10020_02405-R1;Rt_JCM10021_03563-R1	0.472	T-complex protein 1 subunit beta
MvSI-1064-A1-R4_A1g00971	Rt_JCM10020_02275-R1;Rt_JCM10021_01791-R1	0.442	dolichyldiphosphatase / PAP2_like
MvSI-1064-A1-R4_A1g00972	Rt_JCM10020_02276-R1;Rt_JCM10021_01790-R1	0.367	hypothetical protein
MvSI-1064-A1-R4_A1g00997	Rt_JCM10020_04030-R1;Rt_JCM10021_02518-R1	0.423	hypothetical protein
MvSI-1064-A1-R4_A1g00998	Rt_JCM10020_02282-R1;Rt_JCM10021_01784-R1	0.496	U3 small nucleolar ribonucleoprotein IMP3 / Ribosomal protein S4/S9
MvSI-1064-A1-R4_A1g01017	Rt_JCM10020_06899-R1;Rt_JCM10021_05518-R1	0.523	phosphatidylethanolamine N-methyltransferase (PEMT)
MvSI-1064-A1-R4_A1g01154	Rt_JCM10020_02129-R1;Rt_JCM10021_01938-R1	0.223	ribosomal protein L1
MvSI-1064-A1-R4_A1g01159	Rt_JCM10020_02186-R1;Rt_JCM10021_01883-R1	0.270	bZIP transcription factor (AP-1)
MvSI-1064-A1-R4_A1g01160	Rt_JCM10020_02367-R1;Rt_JCM10021_03528-R1	0.537	COP9 signalosome complex subunit 2
MvSI-1064-A1-R4_A1g01161	Rt_JCM10020_02382-R1;Rt_JCM10021_03542-R1	0.334	mitochondrial carrier protein
MvSI-1064-A1-R4_A1g01162	Rt_JCM10020_02317-R1;Rt_JCM10021_01748-R1	0.370	hypothetical protein
MvSI-1064-A1-R4_A1g01163	Rt_JCM10020_02358-R1;Rt_JCM10021_01709-R1	0.425	ADP-ribosylation factor-like protein 2
MvSI-1064-A1-R4_A1g01388	Rt_JCM10020_04617-R1;Rt_JCM10021_04683-R1	0.394	ribosome assembly protein Noc2
MvSI-1064-A1-R4_A1g01440	Rt_JCM10020_05792-R1;Rt_JCM10021_07432-R1	0.565	autophagy-related protein 7 / E1-like protein-activating enzyme Gsa7p/Apg7p
MvSI-1064-A1-R4_A1g01441	Rt_JCM10020_02100-R1;Rt_JCM10021_01969-R1	0.380	zinc finger, C3HC4 RING-type protein
MvSI-1064-A1-R4_A1g01458	Rt_JCM10020_02343-R1;Rt_JCM10021_01724-R1	0.559	DNA-directed RNA polymerase III subunit RPC1

Table S2. Continued.

(C)

MvSI-Query ID	Paired ID in the outgroup	dS	Functional annotation
MvSI-1064-A1-R4_A1g00025	Ls_CBS5930|g7897.t1;Ls_CBS5931|g4047.t1	0.0000	glycoside hydrolase family 28 protein
MvSI-1064-A1-R4_A1g00175	Ls_CBS5930|g8243.t1;Ls_CBS5931|g3668.t1	0.0205	ABC1-A atypical protein kinase
MvSI-1064-A1-R4_A1g00230	Ls_CBS5930|g8248.t1;Ls_CBS5931|g3672.t1	0.1701	ASF1-like histone chaperone
MvSI-1064-A1-R4_A1g00478	Ls_CBS5930|g3772.t1|corr;Ls_CBS5931|g4597.t1|corr	0.0000	protein disulfide-isomerase
MvSI-1064-A1-R4_A1g00495	Ls_CBS5930|g3767.t1;Ls_CBS5931|g4602.t1	0.0000	nucleotide-binding protein/RRM_SF
MvSI-1064-A1-R4_A1g00591	**Ls_CBS5930|g2009.t1;Ls_CBS5931|g4188.t1**	**0.5762**	**homeodomain transcription factor HD1**
MvSI-1064-A1-R4_A1g00592	**Ls_CBS5930|g2008.t1;Ls_CBS5931|g4189.t1**	**0.5376**	**homeodomain transcription factor HD2**
MvSI-1064-A1-R4_A1g00593	Ls_CBS5930|g3769.t1;Ls_CBS5931|g4600.t1	0.0000	5'-3' exoribonuclease 2
MvSI-1064-A1-R4_A1g00659	Ls_CBS5930|g5144.t1;Ls_CBS5931|g1185.t1	0.0117	alpha-mannosyltransferase, glycosyltransferase family 71 protein
MvSI-1064-A1-R4_A1g00726	Ls_CBS5930|g3411.t1;Ls_CBS5931|g3664.t1	0.0155	defective in Cullin neddylation protein 1/UBA_like_SF
MvSI-1064-A1-R4_A1g00772	Ls_CBS5930|g3768.t1;Ls_CBS5931|g4601.t1	0.0000	U3 snoRNP-associated protein Esf2 / RRM_ABT1_like
MvSI-1064-A1-R4_A1g00800	Ls_CBS5930|g1990.t1|corr;Ls_CBS5931|g4209.t1|corr	0.0000	S-(hydroxymethyl)glutathione dehydrogenase / alcohol dehydrogenase
MvSI-1064-A1-R4_A1g00844	Ls_CBS5930|g3965.t1;Ls_CBS5931|g6949.t1	0.0058	hypothetical protein
MvSI-1064-A1-R4_A1g00854	Ls_CBS5930|g8369.t1|corr;Ls_CBS5931|g3662.t1	0.0159	karyopherin kap95
MvSI-1064-A1-R4_A1g00970	Ls_CBS5930|g6624.t1|corr;Ls_CBS5931|g3773.t1|corr	0.0000	T-complex protein 1 subunit beta
MvSI-1064-A1-R4_A1g00971	Ls_CBS5930|g5099.t1;Ls_CBS5931|g239.t1	0.0488	dolichyldiphosphatase / PAP2_like
MvSI-1064-A1-R4_A1g00972	Ls_CBS5930|g5100.t1;Ls_CBS5931|g238.t1	0.0065	hypothetical protein
MvSI-1064-A1-R4_A1g00997	Ls_CBS5930|g1916.t1;Ls_CBS5931|g98.t1	0.0062	hypothetical protein
MvSI-1064-A1-R4_A1g00998	Ls_CBS5930|g3566.t1;Ls_CBS5931|g147.t1	0.1334	U3 small nucleolar ribonucleoprotein IMP3 / Ribosomal protein S4/S9
MvSI-1064-A1-R4_A1g01017	Ls_CBS5930|g3407.t1;Ls_CBS5931|g8158.t1	0.0516	phosphatidylethanolamine N-methyltransferase (PEMT)
MvSI-1064-A1-R4_A1g01154	Ls_CBS5930|g5411.t1|corr;Ls_CBS5931|g3145.t1	0.0786	ribosomal protein L1
MvSI-1064-A1-R4_A1g01159	Ls_CBS5930|g3515.t1;Ls_CBS5931|g197.t1	0.0000	bZIP transcription factor (AP-1)
MvSI-1064-A1-R4_A1g01160	Ls_CBS5930|g5395.t1;Ls_CBS5931|g3159.t1	0.0239	COP9 signalosome complex subunit 2
MvSI-1064-A1-R4_A1g01161	Ls_CBS5930|g6628.t1;Ls_CBS5931|g3769.t1	0.0000	mitochondrial carrier protein
MvSI-1064-A1-R4_A1g01162	Ls_CBS5930|g5385.t1;Ls_CBS5931|g3169.t1	0.0986	hypothetical protein
MvSI-1064-A1-R4_A1g01163	Ls_CBS5930|g6627.t1;Ls_CBS5931|g3770.t1	0.0000	ADP-ribosylation factor-like protein 2
MvSI-1064-A1-R4_A1g01388	Ls_CBS5930|g4892.t1|corr;Ls_CBS5931|g6975.t1|corr	0.0948	ribosome assembly protein Noc2
MvSI-1064-A1-R4_A1g01440	Ls_CBS5930|g3762.t1;Ls_CBS5931|g4607.t1	0.0000	autophagy-related protein 7 / E1-like protein-activating enzyme Gsa7p/Apg7p
MvSI-1064-A1-R4_A1g01441	Ls_CBS5930|g3507.t1;Ls_CBS5931|g205.t1	0.0000	zinc finger, C3HC4 RING-type protein
MvSI-1064-A1-R4_A1g01457	Ls_CBS5930|g6847.t1;Ls_CBS5931|g3790.t1	0.0022	unnamed protein product
MvSI-1064-A1-R4_A1g01458	Ls_CBS5930|g6847.t1;Ls_CBS5931|g3790.t1	0.0022	DNA-directed RNA polymerase III subunit RPC1

Table S3. Gene ID and putative function of the genes located in the different evolutionary strata in *Microbotryum lychnidis-dioicae* (MvSl) (shared with *M. lagerheimii* (MvSv) mating-type chromosomes). The two HD genes are highlighted in bold and grey. The d_S corresponds to synonymous divergence between alleles associated to the alternative mating types of the sequenced individual.

MvSl gene ID (a₁)	MvSv gene ID (a₁)	d_S in MvSl	d_S in MvSv	Stratum	Putative function
MvSl-1064-A1-R4_A1g01509	MvSv-1253-A1-R1_MC03g02122	0	0	PAR	"AutoIPR: IPR008926:Ribonucleotide reductase R1 subunit, N-terminal; IPR013346:Ribonucleotide reductase, class I , alpha subunit; IPR013350:Ribonucleotide reductase alpha chain;"
MvSl-1064-A1-R4_A1g01508	MvSv-1253-A1-R1_MC03g02123	0	0	PAR	AutoIPR: IPR010482:Peroxin/Dysferlin domain;
MvSl-1064-A1-R4_A1g01502	MvSv-1253-A1-R1_MC03g02124	0,002787	0	PAR	AutoIPR: IPR009053:Prefoldin;
MvSl-1064-A1-R4_A1g01501	MvSv-1253-A1-R1_MC03g02125	0	0	PAR	AutoIPR: IPR001805:Adenosine kinase;
MvSl-1064-A1-R4_A1g01500	MvSv-1253-A1-R1_MC03g02126	0	0	PAR	AutoIPR: IPR016024:Armadillo-type fold; IPR027159:Nuclear cap-binding protein subunit 1;
MvSl-1064-A1-R4_A1g01498	MvSv-1253-A1-R1_MC03g02130	0	0	PAR	AutoIPR: IPR005821:Ion transport domain;
MvSl-1064-A1-R4_A1g01497	MvSv-1253-A1-R1_MC03g02134	0	0	PAR	AutoIPR: IPR011992:EF-hand domain pair;
MvSl-1064-A1-R4_A1g01492	MvSv-1253-A1-R1_MC03g02142	0	0	PAR	
MvSl-1064-A1-R4_A1g01491	MvSv-1253-A1-R1_MC03g02144	0	0	PAR	
MvSl-1064-A1-R4_A1g01489	MvSv-1253-A1-R1_MC03g02148	0	0	PAR	"AutoIPR: IPR006110:RNA polymerase, subunit omega/K/RPABC2;"
MvSl-1064-A1-R4_A1g01488	MvSv-1253-A1-R1_MC03g02149	0	0	PAR	"AutoIPR: IPR022591:Transcription initiation factor TFIID subunit 1, domain of unknown function;"
MvSl-1064-A1-R4_A1g01487	MvSv-1253-A1-R1_MC03g02150	0	0	PAR	AutoIPR: IPR021709:Protein of unknown function DUF3292;
MvSl-1064-A1-R4_A1g01478	MvSv-1253-A1-R1_MC03g02153	0	0	PAR	
MvSl-1064-A1-R4_A1g01486	MvSv-1253-A1-R1_MC03g02154	0	0	PAR	AutoIPR: IPR022440:Conserved hypothetical protein CHP03788;
MvSl-1064-A1-R4_A1g01485	MvSv-1253-A1-R1_MC03g02155	0	0	PAR	AutoIPR: IPR001237:43kDa postsynaptic protein; IPR001623:DnaJ domain; IPR026901:DnaJ homologue subfamily C member 3;
MvSl-1064-A1-R4_A1g01474	MvSv-1253-A1-R1_MC03g02159	0	0	PAR	
MvSl-1064-A1-R4_A1g01476	MvSv-1253-A1-R1_MC03g02160	0	0	PAR	
MvSl-1064-A1-R4_A1g01477	MvSv-1253-A1-R1_MC03g02164	0	0	PAR	AutoIPR: IPR006994:Transcription factor 25;
MvSl-1064-A1-R4_A1g01484	MvSv-1253-A1-R1_MC03g02165	0	0	PAR	
MvSl-1064-A1-R4_A1g01473	MvSv-1253-A1-R1_MC03g02166	0	0	PAR	
MvSl-1064-A1-R4_A1g01472	MvSv-1253-A1-R1_MC03g02167	0	0	PAR	AutoIPR: IPR003162:Transcription initiation factor TAFII31; IPR009072:Histone-fold;
MvSl-1064-A1-R4_A1g01471	MvSv-1253-A1-R1_MC03g02168	0	0	PAR	AutoIPR: IPR007900:Transcription initiation factor TFIID component TAF4;
MvSl-1064-A1-R4_A1g01470	MvSv-1253-A1-R1_MC03g02170	0	0	PAR	
MvSl-1064-A1-R4_A1g01468	MvSv-1253-A1-R1_MC03g02172	0	0	PAR	
MvSl-1064-A1-R4_A1g01466	MvSv-1253-A1-R1_MC03g02174	0	0	PAR	
MvSl-1064-A1-R4_A1g01463	MvSv-1253-A1-R1_MC03g02180	0	0	PAR	
MvSl-1064-A1-R4_A1g01462	MvSv-1253-A1-R1_MC03g02181	0	0	PAR	"AutoIPR: IPR014624:Predicted 26S proteasome regulatory complex, non-ATPase subcomplex, subunit s5a, Plasmodium; IPR027040:Proteasome subunit Rpn10;"
MvSl-1064-A1-R4_A1g01461	MvSv-1253-A1-R1_MC03g02182	0	0	PAR	AutoIPR: IPR000456:Ribosomal protein L17;
MvSl-1064-A1-R4_A1g01460	MvSv-1253-A1-R1_MC03g02183	0	0	PAR	
MvSl-1064-A1-R4_A1g01458	MvSv-1253-A1-R1_MC03g02185	0	0,01838	PAR	"AutoIPR: IPR009010:Aspartate decarboxylase-like domain; IPR012754:DNA-directed RNA polymerase, subunit beta-prime; IPR015700:DNA-directed RNA polymerase III largest subunit;"
MvSl-1064-A1-R4_A1g01457	MvSv-1253-A1-R1_MC03g02186	0	0,02487	PAR	
MvSl-1064-A1-R4_A1g01456	MvSv-1253-A1-R1_MC03g02187	0	0	PAR	
MvSl-1064-A1-R4_A1g01455	MvSv-1253-A1-R1_MC03g02188	0	0	PAR	
MvSl-1064-A1-R4_A1g01454	MvSv-1253-A1-R1_MC03g02189	0	0	PAR	
MvSl-1064-A1-R4_A1g01453	MvSv-1253-A1-R1_MC03g02190	0	0	PAR	"AutoIPR: IPR005828:General substrate transporter; IPR016196:Major facilitator superfamily domain, general substrate transporter;"
MvSl-1064-A1-R4_A1g01451	MvSv-1253-A1-R1_MC03g02191	0	0	PAR	"AutoIPR: IPR011042:Six-bladed beta-propeller, TolB-like;"
MvSl-1064-A1-R4_A1g01450	MvSv-1253-A1-R1_MC03g02192	0	0	PAR	"AutoIPR: IPR009082:Signal transduction histidine kinase, homodimeric domain; IPR011006:CheY-like superfamily; IPR013655:PAS fold-3; IPR014285:Nitrogen fixation negative regulator NifL;"
MvSl-1064-A1-R4_A1g01447	MvSv-1253-A1-R1_MC03g02193	0	0	PAR	"AutoIPR: IPR017923:Transcription factor IIS, N-terminal;"
MvSl-1064-A1-R4_A1g01446	MvSv-1253-A1-R1_MC03g02194	0	0	PAR	
MvSl-1064-A1-R4_A1g01445	MvSv-1253-A1-R1_MC03g02195	0	0	PAR	
MvSl-1064-A1-R4_A1g01444	MvSv-1253-A1-R1_MC03g02196	0	0	PAR	"AutoIPR: IPR002226:Catalase haem-binding site; IPR010582:Catalase immune-responsive domain; IPR011614:Catalase core domain; IPR018028:Catalase, mono-functional, haem-containing; IPR020835:Catalase-like domain; IPR024708:Catalase active site;"

MvSl gene ID (a₁)	MvSv gene ID (a₁)	d_S in MvSl	d_S in MvSv	Stratum	Putative function
MvSl-1064-A1-R4_A1g01441	MvSv-1253-A1-R1_MC03g02203	0	0,00184	blue	"AutoIPR: IPR013083:Zinc finger, RING/FYVE/PHD-type;"
MvSl-1064-A1-R4_A1g01440	MvSv-1253-A1-R1_MC03g02204	0	0,008478	blue	"AutoIPR: IPR006285:Ubiquitin-like modifier-activating enzyme Atg7; IPR018075:Ubiquitin-activating enzyme, E1;"
MvSl-1064-A1-R4_A1g01163	MvSv-1253-A1-R1_MC03g02211	0,04476	0,009024	blue	"AutoIPR: IPR005225:Small GTP-binding protein domain; IPR006689:Small GTPase superfamily, ARF/SAR type; IPR027417:P-loop containing nucleoside triphosphate hydrolase;"
MvSl-1064-A1-R4_A1g01162	MvSv-1253-A1-R1_MC03g02212	0,02929	0,0101	blue	
MvSl-1064-A1-R4_A1g01161	MvSv-1253-A1-R1_MC03g02213	0,00892	0,01558	blue	AutoIPR: IPR023395:Mitochondrial carrier domain;
MvSl-1064-A1-R4_A1g01160	MvSv-1253-A1-R1_MC03g02214	0,002244	0,03034	blue	AutoIPR: IPR000717:Proteasome component (PCI) domain; IPR011990:Tetratricopeptide-like helical; IPR013143:PCI/PINT associated module;
MvSl-1064-A1-R4_A1g01159	MvSv-1253-A1-R1_MC03g02215	0,02168	0,0225	blue	AutoIPR: IPR004827:Basic-leucine zipper domain; IPR013910:Transcription factor PAP1; IPR023167:Yap1 redox domain;
MvSl-1064-A1-R4_A1g01154	MvSv-1253-A1-R1_MC03g02216	0,005776	0,0116	blue	
MvSl-1064-A1-R4_A1g00659	MvSv-1253-A1-R1_MC03g02219	0,04854	0,008824	blue	AutoIPR: IPR022751:Alpha-mannosyltransferase;
MvSl-1064-A1-R4_A1g00800	MvSv-1253-A1-R1_MC03g02223	0,03416	0,007407	blue	"AutoIPR: IPR002085:Alcohol dehydrogenase superfamily, zinc-type; IPR016040:NAD(P)-binding domain; IPR020843:Polyketide synthase, enoylreductase;"
MvSl-1064-A1-R4_A1g00801	MvSv-1253-A1-R1_MC03g02224	0	0,0222	blue	AutoIPR: IPR007146:Sas10/Utp3/C1D; IPR011082:Exosome-associated factor Rrp47/DNA strand repair C1D;
MvSl-1064-A1-R4_A1g00808	MvSv-1253-A1-R1_MC03g02226	0,6698	0,02125	blue	
MvSl-1064-A1-R4_A1g00591	**MvSv-1253-A1-R1_MC03g02227**	**0,3894**	**0,4844**	**blue**	**AutoIPR: IPR008422:Homeobox KN domain; IPR009057:Homeodomain-like; HD gene**
MvSl-1064-A1-R4_A1g00592	**MvSv-1253-A1-R1_MC03g02228**	**0,8506**	**0,669**	**blue**	**HD gene**
MvSl-1064-A1-R4_A1g00593	MvSv-1253-A1-R1_MC03g02229	0,04204	0,03775	blue	AutoIPR: IPR027073:5'-3' exoribonuclease;
MvSl-1064-A1-R4_A1g00772	MvSv-1253-A1-R1_MC03g02230	0,03975	0,005973	blue	"AutoIPR: IPR012677:Nucleotide-binding, alpha-beta plait;"
MvSl-1064-A1-R4_A1g00025	MvSv-1253-A1-R1_MC03g02235	0	0,01151	blue	"AutoIPR: IPR000743:Glycoside hydrolase, family 28; IPR011050:Pectin lyase fold/virulence factor;"
MvSl-1064-A1-R4_A1g00478	MvSv-1253-A1-R1_MC03g02236	0,09438	0,03602	blue	AutoIPR: IPR012336:Thioredoxin-like fold;
MvSl-1064-A1-R4_A1g00495	MvSv-1253-A1-R1_MC03g02237	0,06234	0,01041	blue	"AutoIPR: IPR012677:Nucleotide-binding, alpha-beta plait;"
MvSl-1064-A1-R4_A1g01431	MvSv-1253-A1-R1_MC03g02238	0,03381	0	blue	"AutoIPR: IPR001752:Kinesin, motor domain; IPR008984:SMAD/FHA domain; IPR011993:Pleckstrin homology-like domain; IPR022164:Kinesin-like; IPR027417:P-loop containing nucleoside triphosphate hydrolase; IPR027640:Kinesin-like protein;"
MvSl-1064-A1-R4_A1g00997	MvSv-1253-A1-R1_MC03g02240	0,03601	0,01086	blue	
MvSl-1064-A1-R4_A1g00998	MvSv-1253-A1-R1_MC03g02241	0,03152	0,004411	blue	AutoIPR: IPR022801:Ribosomal protein S4/S9;
MvSl-1064-A1-R4_A1g00972	MvSv-1253-A1-R1_MC03g02242	0,01747	0,0052	blue	
MvSl-1064-A1-R4_A1g00971	MvSv-1253-A1-R1_MC03g02243	0,02441	0,02234	blue	AutoIPR: IPR000326:Phosphatidic acid phosphatase type 2/haloperoxidase;
MvSl-1064-A1-R4_A1g00970	MvSv-1253-A1-R1_MC03g02244	0,01335	0,01239	blue	AutoIPR: IPR002423:Chaperonin Cpn60/TCP-1; IPR027409:GroEL-like apical domain; IPR027410:TCP-1-like chaperonin intermediate domain; IPR027413:GroEL-like equatorial domain;
MvSl-1064-A1-R4_A1g00960	MvSv-1253-A1-R1_MC03g02248	0,03457	0	black	AutoIPR: IPR005024:Snf7;
MvSl-1064-A1-R4_A1g00958	MvSv-1253-A1-R1_MC03g02249	0,0192	0	black	AutoIPR: IPR004000:Actin-related protein;
MvSl-1064-A1-R4_A1g00957	MvSv-1253-A1-R1_MC03g02250	0,0344	0	black	AutoIPR: IPR013926:CGI121/TPRKB;
MvSl-1064-A1-R4_A1g00956	MvSv-1253-A1-R1_MC03g02251	0,05355	0	black	
MvSl-1064-A1-R4_A1g00955	MvSv-1253-A1-R1_MC03g02252	0,1164	0	black	
MvSl-1064-A1-R4_A1g00922	MvSv-1253-A1-R1_MC03g02253	0,04901	0	black	
MvSl-1064-A1-R4_A1g00921	MvSv-1253-A1-R1_MC03g02254	0,02868	0,003722	black	AutoIPR: IPR008942:ENTH/VHS;
MvSl-1064-A1-R4_A1g00920	MvSv-1253-A1-R1_MC03g02255	0,03374	0	black	
MvSl-1064-A1-R4_A1g00913	MvSv-1253-A1-R1_MC03g02256	0,06733	0,000814	black	AutoIPR: IPR012943:Spindle associated; IPR024545:Mto2p-binding domain;
MvSl-1064-A1-R4_A1g00912	MvSv-1253-A1-R1_MC03g02260	0	0	black	AutoIPR: IPR003782:Copper chaperone SCO1/SenC; IPR012336:Thioredoxin-like fold;
MvSl-1064-A1-R4_A1g00911	MvSv-1253-A1-R1_MC03g02261	0,01205	0	black	"AutoIPR: IPR008698:NADH:ubiquinone oxidoreductase, B18 subunit;"
MvSl-1064-A1-R4_A1g00794	MvSv-1253-A1-R1_MC03g02262	0,02735	0	black	AutoIPR: IPR009846:Splicing factor 3B subunit 5/RDS3 complex subunit 10;
MvSl-1064-A1-R4_A1g00644	MvSv-1253-A1-R1_MC03g02267	0,03817	0	black	"AutoIPR: IPR000571:Zinc finger, CCCH-type;"
MvSl-1064-A1-R4_A1g00645	MvSv-1253-A1-R1_MC03g02268	0,0177	0	black	AutoIPR: IPR001138:Zn(2)-C6 fungal-type DNA-binding domain;
MvSl-1064-A1-R4_A1g00646	MvSv-1253-A1-R1_MC03g02269	0,01013	0	black	AutoIPR: IPR012908:GPI inositol-deacylase PGAP1-like;
MvSl-1064-A1-R4_A1g00647	MvSv-1253-A1-R1_MC03g02270	0,07508	0	black	"AutoIPR: IPR007541:Uncharacterised protein family, basic secretory protein;"
MvSl-1064-A1-R4_A1g00648	MvSv-1253-A1-R1_MC03g02271	0,02229	0	black	"AutoIPR: IPR019339:CBF1-interacting co-repressor CIR, N-terminal domain; IPR022209:Pre-mRNA splicing factor;"
MvSl-1064-A1-R4_A1g01129	MvSv-1253-A1-R1_MC03g02273	0,05174	0	black	AutoIPR: IPR008978:HSP20-like chaperone;
MvSl-1064-A1-R4_A1g00311	MvSv-1253-A1-R1_MC03g02276	0,02807	0	black	AutoIPR: IPR015915:Kelch-type beta propeller;
MvSl-1064-A1-R4_A1g00312	MvSv-1253-A1-R1_MC03g02277	0,02654	0	black	"AutoIPR: IPR019191:Essential protein Yae1, N-terminal;"
MvSl-1064-A1-R4_A1g00886	MvSv-1253-A1-R1_MC03g02280	0,03442	0	black	
MvSl-1064-A1-R4_A1g00869	MvSv-1253-A1-R1_MC03g02282	0,04332	0	black	"AutoIPR: IPR006003:Carbohydrate kinase, FGGY-related;"

MvSl gene ID (a_1)	MvSv gene ID (a_1)	d_S in MvSl	d_S in MvSv	Stratum	Putative function
MvSl-1064-A1-R4_A1g01135	MvSv-1253-A1-R1_MC16g09520	0,05381	0	black	"AutoIPR: IPR019400:Peptidase C65, otubain;"
MvSl-1064-A1-R4_A1g01136	MvSv-1253-A1-R1_MC16g09519	0,0594	0	black	
MvSl-1064-A1-R4_A1g01137	MvSv-1253-A1-R1_MC16g09518	0,02957	0	black	
MvSl-1064-A1-R4_A1g00732	MvSv-1253-A1-R1_MC16g09516	0,002396	0	black	
MvSl-1064-A1-R4_A1g01407	MvSv-1253-A1-R1_MC16g09514	0,02369	0	black	
MvSl-1064-A1-R4_A1g01406	MvSv-1253-A1-R1_MC16g09512	0,002257	0	black	AutoIPR: IPR007757:MT-A70-like;
MvSl-1064-A1-R4_A1g01405	MvSv-1253-A1-R1_MC16g09511	0,01406	0	black	AutoIPR: IPR027450:Alpha-ketoglutarate-dependent dioxygenase AlkB-like;
MvSl-1064-A1-R4_A1g01404	MvSv-1253-A1-R1_MC16g09510	0,05471	0	black	
MvSl-1064-A1-R4_A1g00326	MvSv-1253-A1-R1_MC16g09508	0,04389	0	black	
MvSl-1064-A1-R4_A1g00327	MvSv-1253-A1-R1_MC16g09507	0,02724	0	black	
MvSl-1064-A1-R4_A1g00448	MvSv-1253-A1-R1_MC16g09503	0,06255	0	black	
MvSl-1064-A1-R4_A1g00749	MvSv-1253-A1-R1_MC16g09501	0,02509	0	black	
MvSl-1064-A1-R4_A1g00750	MvSv-1253-A1-R1_MC16g09500	0,002806	0	black	
MvSl-1064-A1-R4_A1g00751	MvSv-1253-A1-R1_MC16g09499	0,01061	0	black	
MvSl-1064-A1-R4_A1g00752	MvSv-1253-A1-R1_MC16g09498	0,07687	0	black	
MvSl-1064-A1-R4_A1g00755	MvSv-1253-A1-R1_MC16g09497	0,07173	0	black	
MvSl-1064-A1-R4_A1g00757	MvSv-1253-A1-R1_MC16g09496	0,05103	0	black	AutoIPR: IPR001461:Peptidase A1; IPR021109:Aspartic peptidase;
MvSl-1064-A1-R4_A1g00544	MvSv-1253-A1-R1_MC16g09494	0,007251	0	black	"AutoIPR: IPR010032:FAD-linked oxidoreductase; IPR016166:FAD-binding, type 2; IPR023595:L-gulonolactone/D-arabinono-1,4-lactone oxidase;"
MvSl-1064-A1-R4_A1g00543	MvSv-1253-A1-R1_MC16g09493	0	0	black	AutoIPR: IPR016460:Coatomer beta subunit (COPB1); IPR026739:AP complex subunit beta;
MvSl-1064-A1-R4_A1g00542	MvSv-1253-A1-R1_MC16g09492	0,002593	0	black	AutoIPR: IPR001737:Ribosomal RNA adenine methylase transferase; IPR025814:18S rRNA dimethylase DIM1;
MvSl-1064-A1-R4_A1g00541	MvSv-1253-A1-R1_MC16g09488	0,01331	0	black	AutoIPR: IPR024990:Anaphase-promoting complex subunit 1;
MvSl-1064-A1-R4_A1g00534	MvSv-1253-A1-R1_MC16g09487	0	0	black	AutoIPR: IPR001557:L-lactate/malate dehydrogenase;
MvSl-1064-A1-R4_A1g00533	MvSv-1253-A1-R1_MC16g09486	0,07944	0	black	"AutoIPR: IPR011234:Fumarylacetoacetase, C-terminal-related;"
MvSl-1064-A1-R4_A1g00532	MvSv-1253-A1-R1_MC16g09485	0,0261	0	black	
MvSl-1064-A1-R4_A1g00528	MvSv-1253-A1-R1_MC16g09482	0,02956	0	black	
MvSl-1064-A1-R4_A1g00525	MvSv-1253-A1-R1_MC16g09479	0,02199	0,01648	black	
MvSl-1064-A1-R4_A1g00522	MvSv-1253-A1-R1_MC16g09478	0,04318	0,01311	black	"AutoIPR: IPR011701:Major facilitator superfamily; IPR016196:Major facilitator superfamily domain, general substrate transporter;"
MvSl-1064-A1-R4_A1g00521	MvSv-1253-A1-R1_MC16g09477	0,002238	0,002248	black	AutoIPR: IPR013126:Heat shock protein 70 family;
MvSl-1064-A1-R4_A1g00520	MvSv-1253-A1-R1_MC16g09476	0	0,02046	black	
MvSl-1064-A1-R4_A1g00519	MvSv-1253-A1-R1_MC16g09475	0	0,01423	black	"AutoIPR: IPR011016:Zinc finger, RING-CH-type; IPR013083:Zinc finger, RING/FYVE/PHD-type;"
MvSl-1064-A1-R4_A1g00518	MvSv-1253-A1-R1_MC16g09474	0,002431	0,01173	black	AutoIPR: IPR011047:Quinonprotein alcohol dehydrogenase-like superfamily; IPR015943:WD40/YVTN repeat-like-containing domain;
MvSl-1064-A1-R4_A1g00517	MvSv-1253-A1-R1_MC16g09473	0	0	black	
MvSl-1064-A1-R4_A1g00516	MvSv-1253-A1-R1_MC16g09472	0,02563	0,01878	black	"AutoIPR: IPR007087:Zinc finger, C2H2; IPR015940:Ubiquitin-associated/translation elongation factor EF1B, N-terminal, eukaryote; IPR017346:Uncharacterised conserved protein UCP037991, UAS/UBX;"
MvSl-1064-A1-R4_A1g00515	MvSv-1253-A1-R1_MC16g09471	0,01326	0,02447	black	"AutoIPR: IPR006293:DNA helicase, ATP-dependent, RecQ type, bacterial; IPR027417:P-loop containing nucleoside triphosphate hydrolase;"
MvSl-1064-A1-R4_A1g00514	MvSv-1253-A1-R1_MC16g09470	0,02296	0,02073	black	AutoIPR: IPR008936:Rho GTPase activation protein;
MvSl-1064-A1-R4_A1g00513	MvSv-1253-A1-R1_MC16g09469	0,0009531	0,01438	black	AutoIPR: IPR008521:Magnesium transporter NIPA;
MvSl-1064-A1-R4_A1g00512	MvSv-1253-A1-R1_MC16g09468	0	0,02597	black	"AutoIPR: IPR007305:Vesicle transport protein, Got1/SFT2-like;"
MvSl-1064-A1-R4_A1g00511	MvSv-1253-A1-R1_MC16g09467	0	0,03902	black	AutoIPR: IPR023395:Mitochondrial carrier domain;
MvSl-1064-A1-R4_A1g00510	MvSv-1253-A1-R1_MC16g09466	0	0,01317	black	"AutoIPR: IPR003754:Tetrapyrrole biosynthesis, uroporphyrinogen III synthase;"
MvSl-1064-A1-R4_A1g00509	MvSv-1253-A1-R1_MC16g09465	0,09326	0,02499	black	AutoIPR: IPR003817:Phosphatidylserine decarboxylase-related;
MvSl-1064-A1-R4_A1g00508	MvSv-1253-A1-R1_MC16g09464	0,1019	0,01383	black	"AutoIPR: IPR002938:Monooxygenase, FAD-binding; IPR020946:Flavin monooxygenase-like;"
MvSl-1064-A1-R4_A1g00506	MvSv-1253-A1-R1_MC16g09461	0,004505	0	black	AutoIPR: IPR002791:Domain of unknown function DUF89;
MvSl-1064-A1-R4_A1g00505	MvSv-1253-A1-R1_MC16g09460	0	0,01273	black	"AutoIPR: IPR005097:Saccharopine dehydrogenase / Homospermidine synthase; IPR007698:Alanine dehydrogenase/PNT, NAD(H)-binding domain; IPR007886:Alanine dehydrogenase/pyridine nucleotide transhydrogenase, N-terminal;"
MvSl-1064-A1-R4_A1g00503	MvSv-1253-A1-R1_MC16g09459	0,009071	0,01064	black	AutoIPR: IPR008942:ENTH/VHS; IPR018205:VHS subgroup;
MvSl-1064-A1-R4_A1g01351	MvSv-1253-A1-R1_MC16g09458	0,003649	0	black	AutoIPR: IPR001189:Manganese/iron superoxide dismutase;
MvSl-1064-A1-R4_A1g01350	MvSv-1253-A1-R1_MC16g09457	0,01323	0	black	AutoIPR: IPR024338:Stretch-activated cation channel Mid1;
MvSl-1064-A1-R4_A1g01348	MvSv-1253-A1-R1_MC16g09456	0,0105	0	black	AutoIPR: IPR000387:Protein-tyrosine/Dual specificity phosphatase; IPR008343:Mitogen-activated protein (MAP) kinase phosphatase; IPR024950:Dual specificity phosphatase;

MvSl gene ID (a_1)	MvSv gene ID (a_1)	d_S in MvSl	d_S in MvSv	Stratum	Putative function
MvSl-1064-A1-R4_A1g01347	MvSv-1253-A1-R1_MC16g09455	0,02952	0	black	"AutoIPR: IPR012954:BP28, C-terminal domain; IPR016024:Armadillo-type fold; IPR022125:U3 small nucleolar RNA-associated protein 10;"
MvSl-1064-A1-R4_A1g01415	MvSv-1253-A1-R1_MC16g09444	0,02235	0	black	"AutoIPR: IPR008352:Mitogen-activated protein (MAP) kinase, p38; IPR020777:Tyrosine-protein kinase, neurotrophic receptor;"
MvSl-1064-A1-R4_A1g01393	MvSv-1253-A1-R1_MC16g09443	0,03313	0	black	
MvSl-1064-A1-R4_A1g01392	MvSv-1253-A1-R1_MC16g09442	0,02447	0	black	
MvSl-1064-A1-R4_A1g01479	MvSv-1253-A1-R1_MC16g09439	0	0	black	
MvSl-1064-A1-R4_A1g00211	MvSv-1253-A1-R1_MC12g07980	0,01398	0	black	"AutoIPR: IPR008010:Membrane protein,Tapt1/CMV receptor;"
MvSl-1064-A1-R4_A1g00779	MvSv-1253-A1-R1_MC12g07965	0,04375	0	black	AutoIPR: IPR005378:Vacuolar protein sorting-associated protein 35;
MvSl-1064-A1-R4_A1g00778	MvSv-1253-A1-R1_MC12g07964	0,04144	0	black	"AutoIPR: IPR015408:Zinc finger, Mcm10/DnaG-type;"
MvSl-1064-A1-R4_A1g00241	MvSv-1253-A1-R1_MC12g07960	0,03193	0	black	
MvSl-1064-A1-R4_A1g00242	MvSv-1253-A1-R1_MC12g07959	0,033	0	black	AutoIPR: IPR007653:Signal peptidase 22kDa subunit;
MvSl-1064-A1-R4_A1g00245	MvSv-1253-A1-R1_MC12g07956	0,07408	0	black	
MvSl-1064-A1-R4_A1g00269	MvSv-1253-A1-R1_MC12g07952	0,04395	0	black	"AutoIPR: IPR003111:Peptidase S16, lon N-terminal; IPR014252:Sporulation protease LonC; IPR014721:Ribosomal protein S5 domain 2-type fold, subgroup; IPR015947:PUA-like domain; IPR027065:Lon protease; IPR027417:P-loop containing nucleoside triphosphate hydrolase;"
MvSl-1064-A1-R4_A1g00270	MvSv-1253-A1-R1_MC12g07950	0,04929	0	black	"AutoIPR: IPR006219:DHAP synthase, class 1; IPR013785:Aldolase-type TIM barrel;"
MvSl-1064-A1-R4_A1g00292	MvSv-1253-A1-R1_MC12g07947	0,02908	0	black	AutoIPR: IPR001394:Ubiquitin carboxyl-terminal hydrolases family 2; IPR019955:Ubiquitin supergroup;
MvSl-1064-A1-R4_A1g00290	MvSv-1253-A1-R1_MC12g07946	0,02661	0	black	AutoIPR: IPR027815:Domain of unknown function DUF4463;
MvSl-1064-A1-R4_A1g00693	MvSv-1253-A1-R1_MC12g07943	0	0	black	AutoIPR: IPR002042:Uricase;
MvSl-1064-A1-R4_A1g00694	MvSv-1253-A1-R1_MC12g07942	0,01544	0	black	"AutoIPR: IPR002100:Transcription factor, MADS-box;"
MvSl-1064-A1-R4_A1g00719	MvSv-1253-A1-R1_MC12g07935	0,1062	0	black	"AutoIPR: IPR004589:DNA helicase, ATP-dependent, RecQ type; IPR018973:DEAD/DEAH-box helicase, putative; IPR027417:P-loop containing nucleoside triphosphate hydrolase;"
MvSl-1064-A1-R4_A1g00718	MvSv-1253-A1-R1_MC12g07934	0,03731	0	black	"AutoIPR: IPR007577:Glycosyltransferase, DXD sugar-binding motif;"
MvSl-1064-A1-R4_A1g00254	MvSv-1253-A1-R1_MC12g07933	0,03514	0	black	AutoIPR: IPR003386:Lecithin:cholesterol/phospholipid:diacylglycerol acyltransferase;
MvSl-1064-A1-R4_A1g00359	MvSv-1253-A1-R1_MC12g07932	0,02437	0	black	AutoIPR: IPR020831:Glyceraldehyde/Erythrose phosphate dehydrogenase family;
MvSl-1064-A1-R4_A1g00281	MvSv-1253-A1-R1_MC12g07931	0,04624	0	black	"AutoIPR: IPR003959:ATPase, AAA-type, core; IPR027417:P-loop containing nucleoside triphosphate hydrolase;"
MvSl-1064-A1-R4_A1g00280	MvSv-1253-A1-R1_MC12g07930	0,02891	0	black	"AutoIPR: IPR008973:C2 calcium/lipid-binding domain, CaLB; IPR019558:Mammalian uncoordinated homology 13, subgroup, domain 2;"
MvSl-1064-A1-R4_A1g00433	MvSv-1253-A1-R1_MC12g07928	0,02267	0,001815	black	"AutoIPR: IPR022042:snRNA-activating protein complex, subunit 3;"
MvSl-1064-A1-R4_A1g00432	MvSv-1253-A1-R1_MC12g07927	0,1063	0	black	AutoIPR: IPR009071:High mobility group box domain;
MvSl-1064-A1-R4_A1g00349	MvSv-1253-A1-R1_MC12g07926	0,09006	0	black	AutoIPR: IPR000791:GPR1/FUN34/yaaH;
MvSl-1064-A1-R4_A1g00350	MvSv-1253-A1-R1_MC12g07925	0,02792	0	black	AutoIPR: IPR015915:Kelch-type beta propeller;
MvSl-1064-A1-R4_A1g00351	MvSv-1253-A1-R1_MC12g07924	0,03002	0,0007955	black	"AutoIPR: IPR007320:Programmed cell death protein 2, C-terminal; IPR015940:Ubiquitin-associated/translation elongation factor EF1B, N-terminal, eukaryote; IPR017346:Uncharacterised conserved protein UCP037991, UAS/UBX;"
MvSl-1064-A1-R4_A1g00352	MvSv-1253-A1-R1_MC12g07923	0,03106	0,006247	black	"AutoIPR: IPR016656:Transcription initiation factor TFIIE, beta subunit;"
MvSl-1064-A1-R4_A1g01272	MvSv-1253-A1-R1_MC12g07922	0,01626	0	black	AutoIPR: IPR008417:B-cell receptor-associated 31-like;
MvSl-1064-A1-R4_A1g01271	MvSv-1253-A1-R1_MC12g07921	0,01543	0	black	AutoIPR: IPR008011:Complex 1 LYR protein;
MvSl-1064-A1-R4_A1g01173	MvSv-1253-A1-R1_MC12g07920	0,03657	0,007907	black	"AutoIPR: IPR012677:Nucleotide-binding, alpha-beta plait;"
MvSl-1064-A1-R4_A1g01174	MvSv-1253-A1-R1_MC12g07919	0,02616	0	black	AutoIPR: IPR014729:Rossmann-like alpha/beta/alpha sandwich fold;
MvSl-1064-A1-R4_A1g01220	MvSv-1253-A1-R1_MC12g07918	0,09423	0	black	"AutoIPR: IPR001781:Zinc finger, LIM-type;"
MvSl-1064-A1-R4_A1g01219	MvSv-1253-A1-R1_MC12g07917	0,06024	0	black	AutoIPR: IPR018803:Stress-responsive protein Ish1;
MvSl-1064-A1-R4_A1g00165	MvSv-1253-A1-R1_MC12g07913	0,03698	0,00352	black	"AutoIPR: IPR006153:Cation/H+ exchanger; IPR018422:Cation/H+ exchanger, CPA1 family;"
MvSl-1064-A1-R4_A1g00157	MvSv-1253-A1-R1_MC12g07911	0,03635	0,01021	black	"AutoIPR: IPR016196:Major facilitator superfamily domain, general substrate transporter;"
MvSl-1064-A1-R4_A1g00155	MvSv-1253-A1-R1_MC12g07910	0,04423	0	black	AutoIPR: IPR001645:Folylpolyglutamate synthetase;
MvSl-1064-A1-R4_A1g00153	MvSv-1253-A1-R1_MC12g07908	0,03147	0,003478	black	"AutoIPR: IPR010222:RNA helicase, ATP-dependent DEAH box, HrpA-type; IPR022307:DEAD/DEAH-box helicase, putative, actinobacteria; IPR027417:P-loop containing nucleoside triphosphate hydrolase;"
MvSl-1064-A1-R4_A1g01212	MvSv-1253-A1-R1_MC12g07907	0,03602	0,002255	black	"AutoIPR: IPR001208:Mini-chromosome maintenance, DNA-dependent ATPase; IPR027417:P-loop containing nucleoside triphosphate hydrolase; IPR027925:MCM N-terminal domain;"
MvSl-1064-A1-R4_A1g01184	MvSv-1253-A1-R1_MC12g07904	0,03576	0,006311	black	AutoIPR: IPR001708:Membrane insertase OXA1/ALB3/YidC;
MvSl-1064-A1-R4_A1g01189	MvSv-1253-A1-R1_MC12g07903	0,02386	0	black	"AutoIPR: IPR011598:Myc-type, basic helix-loop-helix (bHLH) domain;"
MvSl-1064-A1-R4_A1g00016	MvSv-1253-A1-R1_MC12g07897	0	0	black	
MvSl-1064-A1-R4_A1g00017	MvSv-1253-A1-R1_MC12g07896	0	0,01191	black	"AutoIPR: IPR011701:Major facilitator superfamily; IPR016196:Major facilitator superfamily domain, general substrate transporter;"
MvSl-1064-A1-R4_A1g00144	MvSv-1253-A1-R1_MC12g07894	0,05262	0	black	"AutoIPR: IPR006789:ARP2/3 complex, 16kDa subunit (p16-Arc);"

MvSl gene ID (a₁)	MvSv gene ID (a₁)	d_S in MvSl	d_S in MvSv	Stratum	Putative function
MvSl-1064-A1-R4_A1g01389	MvSv-1253-A1-R1_MC12g07893	0,03083	0	black	
MvSl-1064-A1-R4_A1g01388	MvSv-1253-A1-R1_MC12g07892	0,04141	0,00527	black	AutoIPR: IPR005343:Nucleolar complex protein 2;
MvSl-1064-A1-R4_A1g01385	MvSv-1253-A1-R1_MC12g07891	0,02823	0,001958	black	"AutoIPR: IPR009668:RNA polymerase I associated factor, A49-like;"
MvSl-1064-A1-R4_A1g01384	MvSv-1253-A1-R1_MC12g07890	0,0293	0,00419	black	AutoIPR: IPR001060:FCH domain; IPR018808:Muniscin C-terminal mu homology domain;
MvSl-1064-A1-R4_A1g01369	MvSv-1253-A1-R1_MC12g07886	0,02333	0,001781	black	AutoIPR: IPR005792:Protein disulphide isomerase; IPR012336:Thioredoxin-like fold;
MvSl-1064-A1-R4_A1g01368	MvSv-1253-A1-R1_MC12g07885	0,02638	0,005801	black	AutoIPR: IPR019021:Methyl methanesulphonate-sensitivity protein 22;
MvSl-1064-A1-R4_A1g01367	MvSv-1253-A1-R1_MC12g07884	0,01551	0,00982	black	
MvSl-1064-A1-R4_A1g01366	MvSv-1253-A1-R1_MC12g07883	0,0362	0,00599	black	AutoIPR: IPR006565:Bromodomain transcription factor;
MvSl-1064-A1-R4_A1g00845	MvSv-1253-A1-R1_MC12g07881	0,03648	0,001712	black	"AutoIPR: IPR008937:Ras guanine nucleotide exchange factor; IPR023578:Ras guanine nucleotide exchange factor, domain;"
MvSl-1064-A1-R4_A1g00844	MvSv-1253-A1-R1_MC12g07880	0,008035	0	black	
MvSl-1064-A1-R4_A1g00843	MvSv-1253-A1-R1_MC12g07879	0,0116	0,002791	black	AutoIPR: IPR002060:Squalene/phytoene synthase;
MvSl-1064-A1-R4_A1g00842	MvSv-1253-A1-R1_MC12g07878	0,06322	0	black	"AutoIPR: IPR002018:Carboxylesterase, type B;"
MvSl-1064-A1-R4_A1g00840	MvSv-1253-A1-R1_MC12g07874	0,04014	0,004762	black	
MvSl-1064-A1-R4_A1g00832	MvSv-1253-A1-R1_MC12g07872	0,02415	0,001244	black	AutoIPR: IPR011009:Protein kinase-like domain;
MvSl-1064-A1-R4_A1g00575	MvSv-1253-A1-R1_MC12g07870	0,02905	0,003192	black	AutoIPR: IPR002058:PAP/25A-associated; IPR002934:Nucleotidyl transferase domain;
MvSl-1064-A1-R4_A1g00569	MvSv-1253-A1-R1_MC12g07869	0,1057	0,04197	black	
MvSl-1064-A1-R4_A1g00093	MvSv-1253-A1-R1_MC12g07864	0,04154	0,006273	black	
MvSl-1064-A1-R4_A1g00186	MvSv-1253-A1-R1_MC12g07863	0,03298	0,01127	black	AutoIPR: IPR005013:Dolichyl-diphosphooligosaccharide--protein glycosyltransferase subunit WBP1;
MvSl-1064-A1-R4_A1g00185	MvSv-1253-A1-R1_MC12g07862	0,02174	0,002111	black	AutoIPR: IPR011047:Quinonprotein alcohol dehydrogenase-like superfamily; IPR026895:ER membrane protein complex subunit 1;
MvSl-1064-A1-R4_A1g00184	MvSv-1253-A1-R1_MC12g07861	0	0	black	"AutoIPR: IPR002222:Ribosomal protein S19/S15; IPR023575:Ribosomal protein S19, superfamily;"
MvSl-1064-A1-R4_A1g01038	MvSv-1253-A1-R1_MC12g07859	0,01992	0	black	AutoIPR: IPR000836:Phosphoribosyltransferase domain; IPR023031:Orotate phosphoribosyltransferase;
MvSl-1064-A1-R4_A1g01021	MvSv-1253-A1-R1_MC12g07856	0,02939	NA	black	
MvSl-1064-A1-R4_A1g00175	MvSv-1253-A1-R1_MC12g07851	0,1428	0,06338	purple	AutoIPR: IPR004147:UbiB domain;
MvSl-1064-A1-R4_A1g00279	MvSv-1253-A1-R1_MC12g07848	0,09067	NA	purple	
MvSl-1064-A1-R4_A1g00854	MvSv-1253-A1-R1_MC12g07846	0,02438	0,04877	purple	AutoIPR: IPR016024:Armadillo-type fold; IPR027140:Importin subunit beta;
MvSl-1064-A1-R4_A1g01227	MvSv-1253-A1-R1_MC12g07836	0,03955	0	purple	AutoIPR: IPR021138:Ribosomal protein L18a; IPR023573:Ribosomal protein L18a/LX;
MvSl-1064-A1-R4_A1g00230	MvSv-1253-A1-R1_MC12g07828	0,109	0,1148	purple	"AutoIPR: IPR006818:Histone chaperone, ASF1-like;"
MvSl-1064-A1-R4_A1g01017	MvSv-1253-A1-R1_MC12g07823	0,02769	0,07915	purple	AutoIPR: IPR007318:Phospholipid methyltransferase;
MvSl-1064-A1-R4_A1g00726	MvSv-1253-A1-R1_MC12g07815	0,1145	0,07136	purple	AutoIPR: IPR009060:UBA-like; IPR014764:Defective-in-cullin neddylation protein;
MvSl-1064-A1-R4_A1g00624	MvSv-1253-A1-R1_MC12g07792	0,08473	0,03743	purple	"AutoIPR: IPR016071:Staphylococcal nuclease (SNase-like), OB-fold;"
MvSl-1064-A1-R4_A1g00126	MvSv-1253-A1-R1_MC12g07780	0,4265	0,1397	purple	AutoIPR: IPR019166:Apolipoprotein O;
MvSl-1064-A1-R4_A1g00306	MvSv-1253-A1-R1_MC12g07758	NA	NA	orange	
MvSl-1064-A1-R4_A1g00307	MvSv-1253-A1-R1_MC12g07757	0,1172	0,004983	orange	AutoIPR: IPR000702:Ribosomal protein L6;
MvSl-1064-A1-R4_A1g00441	MvSv-1253-A1-R1_MC12g07755	0,03188	0,004631	orange	"AutoIPR: IPR009025:DNA-directed RNA polymerase, RBP11-like dimerisation domain; IPR011262:DNA-directed RNA polymerase, insert domain;"
MvSl-1064-A1-R4_A1g01253	MvSv-1253-A1-R1_MC12g07753	0,02466	0	orange	"AutoIPR: IPR009053:Prefoldin; IPR016655:Prefoldin, subunit 3;"
MvSl-1064-A1-R4_A1g01252	MvSv-1253-A1-R1_MC12g07752	0,07508	0,006777	orange	
MvSl-1064-A1-R4_A1g00669	MvSv-1253-A1-R1_MC12g07750	0,03217	0,002827	orange	"AutoIPR: IPR001876:Zinc finger, RanBP2-type; IPR007286:EAP30; IPR011991:Winged helix-turn-helix DNA-binding domain; IPR021648:Vacuolar protein sorting protein 36, GLUE domain;"
MvSl-1064-A1-R4_A1g01261	MvSv-1253-A1-R1_MC12g07749	0,03357	0,0004392	orange	"AutoIPR: IPR007309:B-block binding subunit of TFIIIC; IPR017956:AT hook, DNA-binding motif;"
MvSl-1064-A1-R4_A1g00407	MvSv-1253-A1-R1_MC12g07746	0,09576	0	red	AutoIPR: IPR000999:Ribonuclease III domain;
MvSl-1064-A1-R4_A1g00406	MvSv-1253-A1-R1_MC12g07745	0	0,01146	red	AutoIPR: IPR000307:Ribosomal protein S16; IPR023803:Ribosomal protein S16 domain;
MvSl-1064-A1-R4_A1g00405	MvSv-1253-A1-R1_MC12g07744	0,003614	0	red	AutoIPR: IPR015943:WD40/YVTN repeat-like-containing domain;
MvSl-1064-A1-R4_A1g00404	MvSv-1253-A1-R1_MC12g07742	0	0	red	AutoIPR: IPR013216:Methyltransferase type 11;
MvSl-1064-A1-R4_A1g00402	MvSv-1253-A1-R1_MC12g07740	0,01347	0,0006517	red	AutoIPR: IPR025279:Stress response protein NST1;
MvSl-1064-A1-R4_A1g00401	MvSv-1253-A1-R1_MC12g07739	0	0	red	AutoIPR: IPR002942:RNA-binding S4 domain; IPR022801:Ribosomal protein S4/S9;
MvSl-1064-A1-R4_A1g00399	MvSv-1253-A1-R1_MC12g07737	0	0	red	AutoIPR: IPR001147:Ribosomal protein L21e;
MvSl-1064-A1-R4_A1g00398	MvSv-1253-A1-R1_MC12g07736	0	0	red	"AutoIPR: IPR019711:ATPase, F0 complex, subunit H;"
MvSl-1064-A1-R4_A1g00397	MvSv-1253-A1-R1_MC12g07735	0	0	red	"AutoIPR: IPR007648:ATPase inhibitor, IATP, mitochondria;"
MvSl-1064-A1-R4_A1g00396	MvSv-1253-A1-R1_MC12g07734	0	0	red	"AutoIPR: IPR007482:Protein-tyrosine phosphatase-like, PTPLA;"
MvSl-1064-A1-R4_A1g00394	MvSv-1253-A1-R1_MC12g07732	0	0		

MvSl gene ID (a_1)	MvSv gene ID (a_1)	d_S in MvSl	d_S in MvSv	Stratum	Putative function
MvSl-1064-A1-R4_A1g00392	MvSv-1253-A1-R1_MC12g07731	0,03094	0	red	
MvSl-1064-A1-R4_A1g00391	MvSv-1253-A1-R1_MC12g07730	0,006953	0	red	"AutoIPR: IPR001392:Clathrin adaptor, mu subunit; IPR027059:Coatomer delta subunit;"
MvSl-1064-A1-R4_A1g00390	MvSv-1253-A1-R1_MC12g07729	0,008075	0	red	
MvSl-1064-A1-R4_A1g00384	MvSv-1253-A1-R1_MC12g07725	0,2181	0		
MvSl-1064-A1-R4_A1g00374	MvSv-1253-A1-R1_MC12g07723	0,001469	0,008644	red	AutoIPR: IPR009075:Acyl-CoA dehydrogenase/oxidase C-terminal; IPR012258:Acyl-CoA oxidase;
MvSl-1064-A1-R4_A1g00373	MvSv-1253-A1-R1_MC12g07722	0	0,02706	red	"AutoIPR: IPR020869:Exosome complex exonuclease 2, probable; IPR027408:PNPase/RNase PH domain;"
MvSl-1064-A1-R4_A1g00372	MvSv-1253-A1-R1_MC12g07721	0,002337	0,002394	red	"AutoIPR: IPR000924:Glutamyl/glutaminyl-tRNA synthetase, class Ib; IPR010987:Glutathione S-transferase, C-terminal-like;"
MvSl-1064-A1-R4_A1g00370	MvSv-1253-A1-R1_MC12g07719	0	0,01147	red	"AutoIPR: IPR007246:Gaa1-like, GPI transamidase component;"
MvSl-1064-A1-R4_A1g00369	MvSv-1253-A1-R1_MC12g07718	0,003175	0,001599	red	AutoIPR: IPR016161:Aldehyde/histidinol dehydrogenase;
MvSl-1064-A1-R4_A1g00368	MvSv-1253-A1-R1_MC12g07717	0,007024	0,002525	red	AutoIPR: IPR003358:tRNA (guanine-N-7) methyltransferase;
MvSl-1064-A1-R4_A1g00367	MvSv-1253-A1-R1_MC12g07716	0,003783	0	red	"AutoIPR: IPR011701:Major facilitator superfamily; IPR016196:Major facilitator superfamily domain, general substrate transporter;"
MvSl-1064-A1-R4_A1g00113	MvSv-1253-A1-R1_MC12g07715	0,007056	0	green	"AutoIPR: IPR002290:Serine/threonine- / dual specificity protein kinase, catalytic domain; IPR020777:Tyrosine-protein kinase, neurotrophic receptor;"
MvSl-1064-A1-R4_A1g00110	MvSv-1253-A1-R1_MC12g07712	0,001184	0	green	
MvSl-1064-A1-R4_A1g00109	MvSv-1253-A1-R1_MC12g07711	0,004364	0,04003	green	"AutoIPR: IPR012677:Nucleotide-binding, alpha-beta plait;"
MvSl-1064-A1-R4_A1g00080	MvSv-1253-A1-R1_MC12g07706	0	0	PAR	
MvSl-1064-A1-R4_A1g00078	MvSv-1253-A1-R1_MC12g07704	0	0	PAR	AutoIPR: IPR002672:Ribosomal protein L28e;
MvSl-1064-A1-R4_A1g00077	MvSv-1253-A1-R1_MC12g07703	0,009973	0,03064	PAR	AutoIPR: IPR004087:K Homology domain;
MvSl-1064-A1-R4_A1g00076	MvSv-1253-A1-R1_MC12g07702	0	0,002377	PAR	
MvSl-1064-A1-R4_A1g00075	MvSv-1253-A1-R1_MC12g07701	0	0,004902	PAR	"AutoIPR: IPR008928:Six-hairpin glycosidase-like; IPR010905:Glycosyl hydrolase, family 88;"
MvSl-1064-A1-R4_A1g00074	MvSv-1253-A1-R1_MC12g07700	0	0,0205	PAR	AutoIPR: IPR002791:Domain of unknown function DUF89;
MvSl-1064-A1-R4_A1g00073	MvSv-1253-A1-R1_MC12g07699	0	0,01967	PAR	AutoIPR: IPR013763:Cyclin-like;
MvSl-1064-A1-R4_A1g00072	MvSv-1253-A1-R1_MC12g07698	0	0,01032	PAR	AutoIPR: IPR002123:Phospholipid/glycerol acyltransferase;
MvSl-1064-A1-R4_A1g00071	MvSv-1253-A1-R1_MC12g07697	0	0	PAR	"AutoIPR: IPR012951:Berberine/berberine-like; IPR016166:FAD-binding, type 2;"
MvSl-1064-A1-R4_A1g00070	MvSv-1253-A1-R1_MC12g07696	0	0	PAR	AutoIPR: IPR009449:GDPGTP exchange factor Sec2p;
MvSl-1064-A1-R4_A1g00069	MvSv-1253-A1-R1_MC12g07695	0	0,0039	PAR	"AutoIPR: IPR020461:Tyrosine-protein kinase, neurotrophic receptor, type 1;"
MvSl-1064-A1-R4_A1g00068	MvSv-1253-A1-R1_MC12g07694	0	0	PAR	
MvSl-1064-A1-R4_A1g00067	MvSv-1253-A1-R1_MC12g07693	0	0,0214	PAR	
MvSl-1064-A1-R4_A1g00065	MvSv-1253-A1-R1_MC12g07691	0	0	PAR	AutoIPR: IPR012696:Phosphonate metabolism PhnM;
MvSl-1064-A1-R4_A1g00064	MvSv-1253-A1-R1_MC12g07683	0	0	PAR	
MvSl-1064-A1-R4_A1g00063	MvSv-1253-A1-R1_MC12g07682	0	0	PAR	"AutoIPR: IPR009244:Mediator complex, subunit Med7;"
MvSl-1064-A1-R4_A1g00062	MvSv-1253-A1-R1_MC12g07681	0	0	PAR	"AutoIPR: IPR012677:Nucleotide-binding, alpha-beta plait;"
MvSl-1064-A1-R4_A1g00061	MvSv-1253-A1-R1_MC12g07680	0	0	PAR	AutoIPR: IPR001806:Small GTPase superfamily; IPR002041:Ran GTPase; IPR005225:Small GTP-binding protein domain; IPR027417:P-loop containing nucleoside triphosphate hydrolase;
MvSl-1064-A1-R4_A1g00057	MvSv-1253-A1-R1_MC12g07678	0	0	PAR	
MvSl-1064-A1-R4_A1g00056	MvSv-1253-A1-R1_MC12g07677	0	0	PAR	"AutoIPR: IPR005828:General substrate transporter; IPR016196:Major facilitator superfamily domain, general substrate transporter;"
MvSl-1064-A1-R4_A1g00055	MvSv-1253-A1-R1_MC12g07676	0	0	PAR	AutoIPR: IPR013960:DASH complex subunit Duo1;role in spindle attachment, chromosome segregation and spindle stability.
MvSl-1064-A1-R4_A1g00054	MvSv-1253-A1-R1_MC12g07675	0	0	PAR	AutoIPR: IPR001269:tRNA-dihydrouridine synthase; IPR013785:Aldolase-type TIM barrel;
MvSl-1064-A1-R4_A1g00053	MvSv-1253-A1-R1_MC12g07674	0	0	PAR	AutoIPR: IPR012336:Thioredoxin-like fold;
MvSl-1064-A1-R4_A1g00052	MvSv-1253-A1-R1_MC12g07673	0	0	PAR	
MvSl-1064-A1-R4_A1g00051	MvSv-1253-A1-R1_MC12g07672	0	0	PAR	AutoIPR: IPR000760:Inositol monophosphatase;
MvSl-1064-A1-R4_A1g00050	MvSv-1253-A1-R1_MC12g07671	0	0	PAR	"AutoIPR: IPR002092:DNA-directed RNA polymerase, phage-type; IPR024075:DNA-directed RNA polymerase, helix hairpin domain;"
MvSl-1064-A1-R4_A1g00049	MvSv-1253-A1-R1_MC12g07670	0	0	PAR	AutoIPR: IPR002761:DUF71 domain; IPR013813:Endoribonuclease L-PSP/chorismate mutase-like;
MvSl-1064-A1-R4_A1g00048	MvSv-1253-A1-R1_MC12g07669	0	0	PAR	
MvSl-1064-A1-R4_A1g00047	MvSv-1253-A1-R1_MC12g07668	0	0	PAR	
MvSl-1064-A1-R4_A1g00046	MvSv-1253-A1-R1_MC12g07667	0	0	PAR	"AutoIPR: IPR011330:Glycoside hydrolase/deacetylase, beta/alpha-barrel;"
MvSl-1064-A1-R4_A1g00044	MvSv-1253-A1-R1_MC12g07665	0	0	PAR	"AutoIPR: IPR011330:Glycoside hydrolase/deacetylase, beta/alpha-barrel;"
MvSl-1064-A1-R4_A1g00040	MvSv-1253-A1-R1_MC12g07664	0	0	PAR	"AutoIPR: IPR011330:Glycoside hydrolase/deacetylase, beta/alpha-barrel;"
MvSl-1064-A1-R4_A1g00039	MvSv-1253-A1-R1_MC12g07663	0	0	PAR	
MvSl-1064-A1-R4_A1g00038	MvSv-1253-A1-R1_MC12g07661	0	0	PAR	

MvSl gene ID (a_1)	MvSv gene ID (a_1)	d_S in MvSl	d_S in MvSv	Stratum	Putative function
MvSl-1064-A1-R4_A1g00037	MvSv-1253-A1-R1_MC12g07660	0	0	PAR	"AutoIPR: IPR001567:Peptidase M3A/M3B; IPR024077:Neurolysin/Thimet oligopeptidase, domain 2; IPR024079:Metallopeptidase, catalytic domain;"
MvSl-1064-A1-R4_A1g00036	MvSv-1253-A1-R1_MC12g07659	0	0	PAR	AutoIPR: IPR006565:Bromodomain transcription factor;This recognition is often a prerequisite for protein-histone association and chromatin remodeling.
MvSl-1064-A1-R4_A1g00035	MvSv-1253-A1-R1_MC12g07658	0	0	PAR	
MvSl-1064-A1-R4_A1g00033	MvSv-1253-A1-R1_MC12g07656	0	0	PAR	
MvSl-1064-A1-R4_A1g00032	MvSv-1253-A1-R1_MC12g07655	0	0	PAR	"AutoIPR: IPR002035:von Willebrand factor, type A;"
MvSl-1064-A1-R4_A1g00031	MvSv-1253-A1-R1_MC12g07654	0	0	PAR	"AutoIPR: IPR003083:S-crystallin; IPR017933:Glutathione S-transferase/chloride channel, C-terminal;"
MvSl-1064-A1-R4_A1g00030	MvSv-1253-A1-R1_MC12g07653	0	0	PAR	
MvSl-1064-A1-R4_A1g00029	MvSv-1253-A1-R1_MC12g07652	0	0	PAR	
MvSl-1064-A1-R4_A1g00009	MvSv-1253-A1-R1_MC12g07644	0	0	PAR	"AutoIPR: IPR006785:Peroxisome membrane anchor protein Pex14p, N-terminal; IPR025655:Peroxisomal membrane protein 14;"
MvSl-1064-A1-R4_A1g00008	MvSv-1253-A1-R1_MC12g07643	0	0	PAR	AutoIPR: IPR013087:Zinc finger C2H2-type/integrase DNA-binding domain;
MvSl-1064-A1-R4_A1g00007	MvSv-1253-A1-R1_MC12g07642	0	0	PAR	AutoIPR: IPR008984:SMAD/FHA domain;
MvSl-1064-A1-R4_A1g00006	MvSv-1253-A1-R1_MC12g07641	0	0	PAR	"AutoIPR: IPR003204:Cytochrome c oxidase, subunit Va/VI;"
MvSl-1064-A1-R4_A1g00005	MvSv-1253-A1-R1_MC12g07640	0	0	PAR	
MvSl-1064-A1-R4_A1g00004	MvSv-1253-A1-R1_MC12g07639	0	0	PAR	AutoIPR: IPR003121:SWIB/MDM2 domain;

Table S4. Raw data summary and assembly statistics for the genomes generated in this study.

sample	Raw data summary							Assembly statistics										
	# Smart cell	# Subreads	Max Subread Length (bp)	N50 (bp)	Mean (bp)	Median (bp)	BP	Accession numbers	# Contigs	Min length (bp)	Max length (bp)	N50 BP	L50 (#contigs)	N90 (bp)	L90 (#contigs)	mean length (bp)	Median Length (bp)	Assembly size (bp)
Microbotryum lagerheimii 1253 A1	4	589978	41394	9526	7290	7059	4 301 220 311	ERS1013677	42	6001	3235273	1584903	7	462124	17	614444	172652	25 806 628
Microbotryum lagerheimii 1253 A2	4	587933	43134	9660	7394	7202	4 347 011 348	ERS1013678	37	7917	3230111	1585144	7	534025	17	693725	474693	25 667 824
Microbotryum violaceum s. str. 1249 A1	4	548702	44168	11951	8759	8361	4 806 182 398	ERS1013671	166	5956	2602331	1183707	10	65951	47	191449	31698	31 780 530
Microbotryum violaceum s. str. 1249 A2	4	456117	46941	13250	9516	8962	4 340 608 458	ERS1013672	61	8455	2644084	1643691	7	606103	17	456993	30510	27 876 544
Microbotryum lychnidis-dioicae Lamole A1	Raw data from Badouin et al. 2015 (12)							ERS1013679	48	12190	3412169	1736850	6	648002	16	622941	44679	29 901 156
Microbotryum lychnidis-dioicae Lamole A2	Raw data from Badouin et al. 2015 (12)							ERS459551	37	15308	4051571	1730088	6	1040457	14	819414	312349	30 318 316
Microbotryum intermedium 1389	3	562278	62677	11040	8512	8220	4 786 123 623	ERS1324257	24	16897	2982376	1644950	6	1002499	13	977543	1002499	23 461 035
Microbotryum silenes-dioicae 1303_A1	7	938931	46274	10834	7946	7576	7 460 889 114	ERS1436592	144	6283	2425171	936137	12	153077	45	233285	32003	33 593 023
Microbotryum silenes-dioicae 1303_A2	6	490924	46480	12620	8406	7474	4 126 642 950	ERS1436593	128	8587	3251277	1321747	10	359665	30	265304	29362	33 958 966

Table S5. *Microbotryum lychnidis-dioicae* and *M. silenes-dioicae* strains used for performing polymorphism analyses. Reference to strain ID, species, mating-type, accession number, number and percentage of reads mapped as proper pairs and reference of the original genome publication, are given.

Strain ID	Species	Mating-type	Accession number	Number of reads mapped as proper pairs	% of reads mapped as proper pairs	Reference
MvSd-335-H3-A1	*M. silenes-dioicae*	a_1	SRS1072004	102696842	96.28%	(48)
MvSd-336-01-A1	*M. silenes-dioicae*	a_1	SRS1072005	127697676	96.85%	(48)
MvSd-578-2-A2	*M. silenes-dioicae*	a_2	SRS1072006|SRS1072007	78559190	96.54%	(48)
MvSd-69-05-A2	*M. silenes-dioicae*	a_2	SRS1072010	121301046	97.35%	(48)
MvSd-707-1-A2	*M. silenes-dioicae*	a_2	SRS1072011	138917854	96.93%	(48)
MvSd-72-A2	*M. silenes-dioicae*	a_2	SRS1072012|SRS1072013	181254142	97.40%	(48)
MvSd-851-6-A2	*M. silenes-dioicae*	a_2	SRS1072015	113257210	97.21%	(48)
MvSd-900-1-A1	*M. silenes-dioicae*	a_1	SRS1072016	143422672	97.33%	(48)
MvSd-932-2-A2	*M. silenes-dioicae*	a_2	SRS1072017	107796114	96.87%	(48)
MvSd-937-2-A2	*M. silenes-dioicae*	a_2	SRS1072018	134111998	97.24%	(48)
MvSd-949-2-A1	*M. silenes-dioicae*	a_1	SRS1072019	105576460	96.66%	(48)
MvSd-IT02-A2	*M. silenes-dioicae*	a_2	SRS1072020	126925450	96.74%	(48)
MvSd-sp002-A1	*M. silenes-dioicae*	a_1	SRS1072021	133835232	96.73%	(48)
MvSd-sp003-A1	*M. silenes-dioicae*	a_1	SRS1072023	110615508	97.32%	(48)
MvSl-00-10-A2	*M. lychnidis-dioicae*	a_2	SRS1072024	71471502	95.32%	(48)
MvSl-100-6-A1	*M. lychnidis-dioicae*	a_1	SRS1072025	70881632	95.69%	(48)
MvSl-140-01-A2	*M. lychnidis-dioicae*	a_2	SRS1072033	157160988	95.07%	(48)
MvSl-141-01-A2	*M. lychnidis-dioicae*	a_2	SRS1072034	66844444	94.89%	(48)
MvSl-40-01-A2	*M. lychnidis-dioicae*	a_2	SRS1072035	65532992	94.18%	(48)
MvSl-446-2-A2	*M. lychnidis-dioicae*	a_2	SRS1072038|SRS1072039	68358472	94.97%	(48)
MvSl-466-3-A1	*M. lychnidis-dioicae*	a_1	SRS1072042|SRS1072043	75508212	94.92%	(48)
MvSl-576-A2	*M. lychnidis-dioicae*	a_2	SRS1072044|SRS1072045	56128954	95.40%	(48)
MvSl-641-A1	*M. lychnidis-dioicae*	a_1	SRS1072046|SRS1072047	87723406	95.25%	(48)
MvSl-661-A2	*M. lychnidis-dioicae*	a_2	SRS1072050	84294192	94.85%	(48)
MvSl-665-A2	*M. lychnidis-dioicae*	a_2	SRS1072048|SRS1072049	37283112	94.87%	(48)
MvSl-687-5-A2	*M. lychnidis-dioicae*	a_2	SRS1072051	46854784	94.81%	(48)
MvSl-830-2-A1	*M. lychnidis-dioicae*	a_1	SRS1072056	81814412	95.81%	(48)
MvSl-856-2-A1	*M. lychnidis-dioicae*	a_1	SRS1072057	37685654	94.57%	(48)
MvSl-I00-3_A2	*M. lychnidis-dioicae*	a_2	SRS1072064	78297378	95.19%	(48)
MvSl-IOA-A1	*M. lychnidis-dioicae*	a_1	SRS1072065	61186726	95.57%	(48)
MvSl-W-1069-A1	*M. lychnidis-dioicae*	a_1	SRS781520	46860062	92.35%	(26)
MvSl-W-1088-A1	*M. lychnidis-dioicae*	a_1	SRS781522	38830336	92.42%	(26)
MvSl-W-1089-A2	*M. lychnidis-dioicae*	a_2	SRS781488	38918966	91.94%	(26)
MvSl-W-1090-A2	*M. lychnidis-dioicae*	a_2	SRS781491	31476556	93.72%	(26)
MvSl-W-1103-A2	*M. lychnidis-dioicae*	a_2	SRS781483	45555798	91.93%	(26)
MvSl-W-405-A1	*M. lychnidis-dioicae*	a_1	SRS780332	88330498	94.39%	(26)
MvSl-W-769-A1	*M. lychnidis-dioicae*	a_1	SRS781521	42520222	91.74%	(26)
MvSl-W-920-A2	*M. lychnidis-dioicae*	a_2	SRS781485	42908076	92.50%	(26)

Table S6. Nucleotide diversity (θ_π) and divergence (mean number of pairwise differences) estimates in pseudo-autosomal regions (PARs) and evolutionary strata in:

(A) *Microbotryum silenes-dioicae*

Gene set	Mean diversity ± SE		Divergence from *M. violaceum s. str.*	
	All sites	Silent sites	All sites	Silent sites
Purple stratum	3.24e-02 ± 1.53e-02	5.23e-02 ± 1.98e-02	4.71e-02 ± 2.04e-02	6.09e-02 ± 1.42e-02
Blue stratum	2.43e-02 ± 1.09e-02	3.72e-02 ± 1.72e-02	5.61e-02 ± 2.09e-02	6.69e-02 ± 1.45e-02
Orange stratum	6.25e-02 ± 4.75e-02	8.88e-02 ± 6.2e-02	7.87e-02 ± 5.16e-02	6.25e-02 ± 2.02e-02
Black stratum	9e-03 ± 8.79e-04	1.45e-02 ± 1.29e-03	2.33e-02 ± 1.13e-03	3.57e-02 ± 1.11e-03
Red stratum	2.47e-03 ± 1.56e-03	4.03e-03 ± 2.41e-03	2.26e-02 ± 3.12e-03	3.88e-02 ± 3.8e-03
Green stratum	2.22e-03 ± 3.6e-04	3.4e-03 ± 8.92e-04	2.39e-02 ± 6.44e-03	4.53e-02 ± 1.27e-02
PARs	2.46e-04 ± 3.37e-05	4.11e-04 ± 8.56e-05	4.1e-02 ± 2.11e-03	8.18e-02 ± 3.54e-03

(B) *Microbotryum lychnidis-dioicae* (whole data set)

Gene set	Mean diversity ± SE		Divergence from *M. violaceum s. str.*	
	All sites	Silent sites	All sites	Silent sites
Purple stratum	3.6e-02 ± 1.48e-02	5.53e-02 ± 2.02e-02	5.28e-02 ± 2.05e-02	6.46e-02 ± 1.47e-02
Blue stratum	2.44e-02 ± 9.91e-03	3.92e-02 ± 1.58e-02	5.66e-02 ± 1.98e-02	7.01e-02 ± 1.39e-02
Orange stratum	6.16e-02 ± 4.51e-02	8.94e-02 ± 6.01e-02	8.06e-02 ± 5.18e-02	6.5e-02 ± 2.04e-02
Black stratum	9.88e-03 ± 6.78e-04	1.61e-02 ± 1.08e-03	2.57e-02 ± 1.02e-03	3.88e-02 ± 1.13e-03
Red stratum	3.11e-03 ± 1.34e-03	3.99e-03 ± 2.36e-03	2.32e-02 ± 2.88e-03	4.03e-02 ± 4.13e-03
Green stratum	1.47e-03 ± 4.39e-04	2.49e-03 ± 1.2e-04	2.48e-02 ± 6.92e-03	4.75e-02 ± 1.12e-02
PARs	2.61e-03 ± 2.22e-04	3.16e-03 ± 2.64e-04	4.41e-02 ± 2.24e-03	8.78e-02 ± 3.79e-03

(C) the North-Western cluster of *Microbotryum lychnidis-dioicae*

Gene set	Mean diversity ± SE		Divergence from *M. violaceum s. str.*	
	All sites	Silent sites	All sites	Silent sites
Purple stratum	3.92e-02 ± 1.61e-02	5.77e-02 ± 2.03e-02	5.23e-02 ± 1.98e-02	6.34e-02 ± 1.36e-02
Blue stratum	2.68e-02 ± 1.09e-02	3.99e-02 ± 1.6e-02	5.69e-02 ± 1.98e-02	7.03e-02 ± 1.38e-02
Orange stratum	6.72e-02 ± 4.95e-02	9.03e-02 ± 6.11e-02	7.8e-02 ± 4.93e-02	6.41e-02 ± 1.93e-02
Black stratum	1.05e-02 ± 7.57e-04	1.63e-02 ± 1.12e-03	2.58e-02 ± 1.02e-03	3.89e-02 ± 1.15e-03
Red stratum	3.4e-03 ± 1.45e-03	4.85e-03 ± 2.41e-03	2.36e-02 ± 2.93e-03	4.07e-02 ± 4.18e-03
Green stratum	4.72e-04 ± 3.69e-04	9.37e-04 ± 9.37e-04	2.5e-02 ± 7.24e-03	4.75e-02 ± 1.22e-02
PARs	9.05e-04 ± 1.23e-04	1.49e-03 ± 2.19e-04	4.45e-02 ± 2.29e-03	8.76e-02 ± 3.79e-03

(D) the Southern cluster of *Microbotryum lychnidis-dioicae*

Gene set	Mean diversity ± SE		Divergence from *M. violaceum s. str.*	
	All sites	Silent sites	All sites	Silent sites
Purple stratum	4.23e-02 ± 1.76e-02	5.54e-02 ± 2e-02	4.73e-02 ± 1.63e-02	5.64e-02 ± 1.07e-02
Blue stratum	2.85e-02 ± 1.18e-02	3.82e-02 ± 1.57e-02	5.66e-02 ± 1.99e-02	6.93e-02 ± 1.36e-02
Orange stratum	7.29e-02 ± 5.34e-02	8.82e-02 ± 5.9e-02	6.26e-02 ± 3.53e-02	5.88e-02 ± 1.41e-02
Black stratum	1.1e-02 ± 8.12e-04	1.54e-02 ± 1.07e-03	2.53e-02 ± 9.53e-04	3.85e-02 ± 1.11e-03
Red stratum	2.8e-03 ± 1.64e-03	3.57e-03 ± 2.24e-03	2.35e-02 ± 3.16e-03	4.01e-02 ± 4.28e-03
Green stratum	8.3e-04 ± 5.2e-04	1.2e-03 ± 1.2e-03	2.5e-02 ± 6.54e-03	4.81e-02 ± 9.62e-03
PARs	1.07e-03 ± 1.29e-04	1.66e-03 ± 2.58e-04	4.39e-02 ± 2.19e-03	8.71e-02 ± 3.78e-03

Table S7. Results of maximum-likelihood Hudson-Kreitman-Agade (MLHKA) tests for assessing whether differences in diversity between each evolutionary stratum and the pseudo-autosomal regions (PARs) is due to balancing selection or elevated mutation rates. Genes were concatenated per stratum unless specified otherwise. Statistically significant results before and after correction for multiple testing at the significance level of 0.05 are indicated in bold; d.f.: degrees of freedom; Selection coefficient: maximum likelihood estimate of the selection coefficient.

(A) In *Microbotryum silenes-dioicae*

Test	d.f.	Selection coefficient	ln likelihood null model	ln likelihood alternative model	Chi2 statistics	p-values	Adjusted p-values
Purple stratum vs PAR	1	60.4123	-284.798	-370.048	170.5	**5.75e-39**	**4.03e-38**
Blue stratum vs PAR	1	44.6019	-286.815	-370.684	167.74	**2.31e-38**	**1.38e-37**
Orange stratum vs PAR	1	75.4768	-286.583	-391.715	210.26	**1.20e-47**	**9.62e-47**
Black stratum vs PAR	1	24.844	-287.805	-340.116	104.62	**1.48e-24**	**7.39e-24**
Red stratum vs PAR	1	7.6661	-288.224	-300.015	23.58	**1.20e-06**	**3.59e-06**
Green stratum vs PAR	1	4.74454	-285.191	-289.094	7.81	**5.21e-03**	**5.21e-03**
Red stratum vs PAR (no concatenation)	16	56.0665	-335.885	-380.386	89	**3.82e-12**	**1.53e-11**
Green stratum vs PAR (no concatenation)	3	10.3826	-289.177	-297.46	16.57	**8.68e-04**	**1.74e-03**

(B) In *Microbotryum lychnidis-dioicae* (whole data set)

Test	d.f.	Selection coefficient	ln likelihood null model	ln likelihood alternative model	Chi2 statistics	p-values	Adjusted p-values
Purple stratum vs PAR	1	12.1267	-383.916	-412.85	57.87	**2.80e-14**	**1.40e-13**
Blue stratum vs PAR	1	10.7939	-386.523	-412.316	51.59	**6.85e-13**	**2.74e-12**
Orange stratum vs PAR	1	18.0431	-384.52	-424.17	79.3	**5.34e-19**	**3.20e-18**
Black stratum vs PAR	1	6.39375	-385.591	-402.312	33.44	**7.34e-09**	**2.20e-08**
Red stratum vs PAR	1	1.94212	-383.672	-385.361	3.38	6.61e-02	1.32e-01
Green stratum vs PAR	1	0.795155	-384.517	-380.695	-7.64	1.00e+00	1.00e+00

(C) In the North-Western cluster of *Microbotryum lychnidis-dioicae*

Test	d.f.	Selection coefficient	ln likelihood null model	ln likelihood alternative model	Chi2 statistics	p-values	Adjusted p-values
Red stratum vs PAR	1	4.96955	-344.795	-336.841	15.91	**6.65e-05**	**1.99e-04**
Green stratum vs PAR	1	0.941793	-333.064	-336.68	-7.23	1.00e+00	1.00e+00
Red stratum vs PAR (no concatenation)	16	2.29673	-417.742	-388.587	58.31	**1.01e-06**	**4.02e-06**
Green stratum vs PAR (no concatenation)	3	0	-332.85	-333.992	-2.28	1.00e+00	1.00e+00

(D) In the Southern cluster of *Microbotryum lychnidis-dioicae*

Test	d.f.	Selection coefficient	ln likelihood null model	ln likelihood alternative model	Chi2 statistics	p-values	Adjusted p-values
Red stratum vs PAR	1	4.0771	-348.271	-341.431	13.68	**2.17e-04**	**8.67e-04**
Green stratum vs PAR	1	1.1329	-339.258	-342.942	-7.37	1.00e+00	1.00e+00
Red stratum vs PAR (no concatenation)	16	1.48588	-410.369	-392.762	32.65	**8.22e-03**	**2.46e-02**
Green stratum vs PAR (no concatenation)	3	0.639097	-338.134	-342.151	-8.03	1.00e+00	1.00e+00

S.2 Absence of antagonistic selection in *Microbotryum lychnidis-dioicae*

1 **Little evidence of antagonistic selection in the evolutionary strata of fungal**

2 **mating-type chromosomes (*Microbotryum lychnidis-dioicae*)**

3 Anna Liza Bazzicalupo*[1,2], Fantin Carpentier†[2], Sarah Perin Otto‡ and Tatiana Giraud† §

4 * Department of Botany, 3200-6270 University Blvd., University of British Columbia, Vancouver,

5 BC V6T 1Z4, Canada

6 † Ecologie Systématique Evolution, Univ. Paris-Sud, CNRS, AgroParisTech, Université Paris-

7 Saclay, 91400 Orsay, France

8 ‡ Department of Zoology & Biodiversity Research Centre, 6270 University Blvd., University of

9 British Columbia, Vancouver, BC V6T 1Z4, Canada

10 § Corresponding author: tatiana.giraud@u-psud.fr; Phone: +33 6 34644514

11 [1] Present address: Department of Microbiology and Immunology, Montana State University, Leon

12 Johnson Hall, Bozeman, MT 59715, United States of America

13 [2] These authors equally contributed to this work

17

18

19
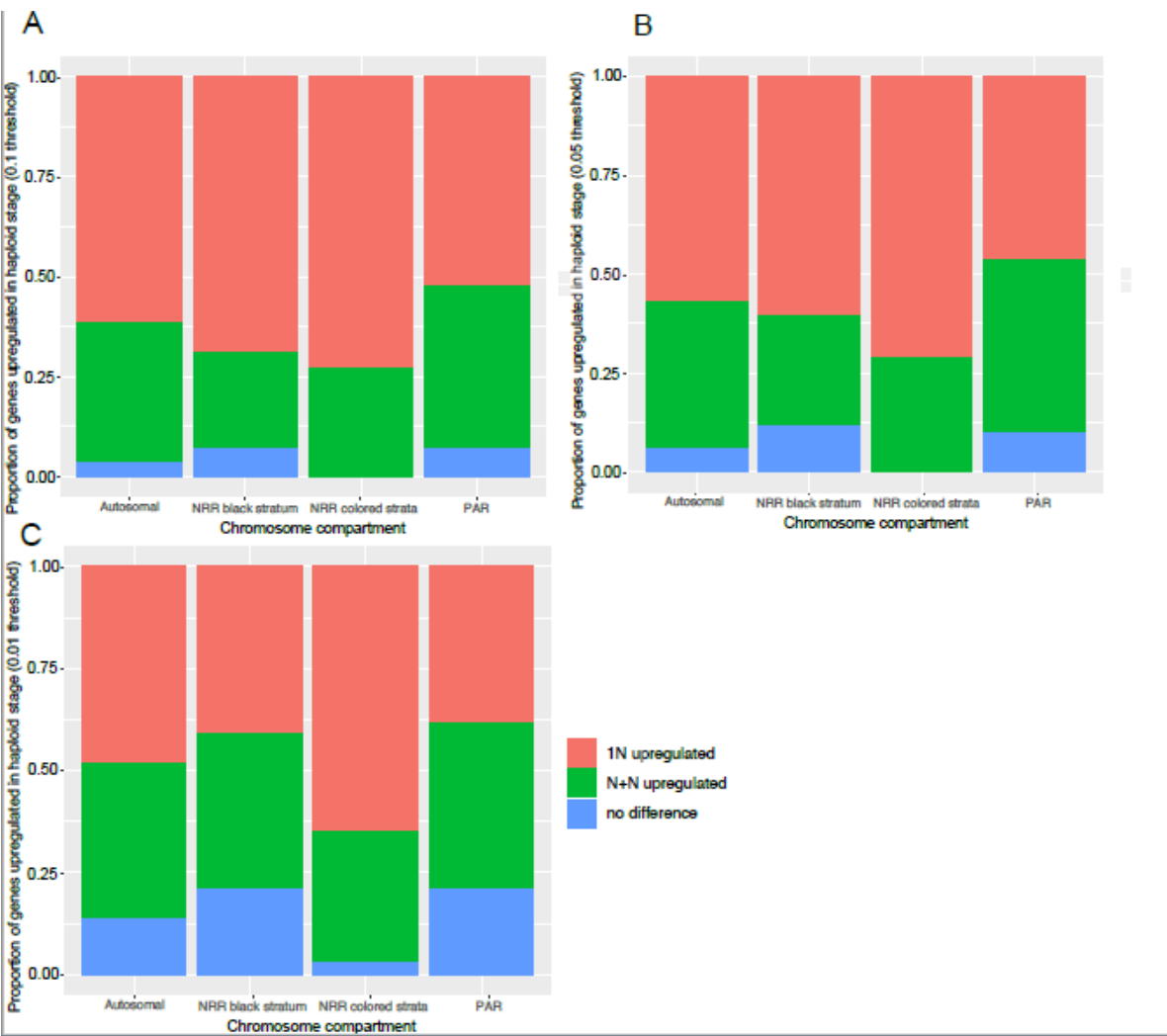
20
21
22
23 **Supplementary Figure 1:** Proportions of *Microbotryum lychnidis-dioicae* genes upregulated in
24 at least one haploid stage/mating type (in red, 1N upregulated), upregulated at the dikaryotic
25 stage (in green, N+N upregulated) or showing no differential expression (in blue). Expression
26 level was considered significantly different at the 0.1 (A), 0.05 (B) or 0.01 (C) threshold. Genes
27 are separated according to their genomic compartment: autosomes, pseudoautosomal regions
28 (PARs) of the mating-type chromosome, non-recombining region (NRR) of the mating-type
29 chromosome, and into the black versus color evolutionary strata (blue, purple, black, orange, red
30 and green).
31
32
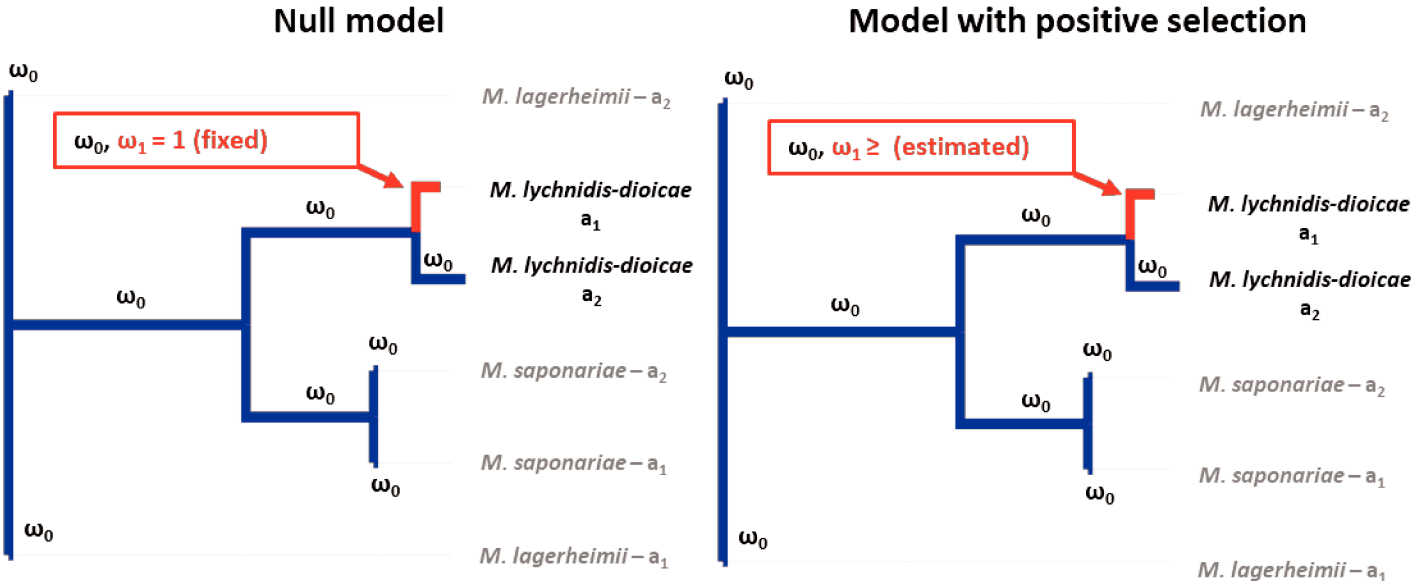
Site class	Null model		Model with positive selection	
	Background	Foreground	Background	Foreground
0	$0 < \omega_0 < 1$	$0 < \omega_0 < 1$	$0 < \omega_0 < 1$	$0 < \omega_0 < 1$
1	$\omega_0 = 1$	$\omega_0 = 1$	$\omega_0 = 1$	$\omega_0 = 1$
2a	$0 < \omega_0 < 1$	$\omega_1 = 1$	$0 < \omega_0 < 1$	$\omega_1 \geq 1$
2b	$\omega_0 = 1$	$\omega_1 = 1$	$\omega_0 = 1$	$\omega_1 \geq 1$

Supplementary Figure S2: Illustration of the positive selection "test 2" from codeml. We performed branch-site tests of positive selection in *Microbotryum lychnidis-dioicae*, by performing a ratio test between the maximum likelihood values of two models of sequence evolution with different values of $\omega = d_N/d_S$ (the ratio between the number of non-synonymous over synonymous substitutions). Each model assumes a single ω value (ω_0) for the background branches (in blue) and two ω values (ω_0 and ω_1) for the foreground branch (the focal branch in which we test the occurrence of positive selection, in red, the sequence association to either the a_1 or a_2 mating type in *M. lychnidis-dioicae*). The null model (left) allows three classes of sites (the classes 1 and 2b being equivalent in this model) while the model with positive selection (right) allows a fourth site class with ω estimates being allowed to be higher than one.

S.3   Convergent evolution of recombination suppression

Supplementary Information

Multiple convergent supergene evolution events in mating-type chromosomes

Branco et al.

Supplementary Information Includes:
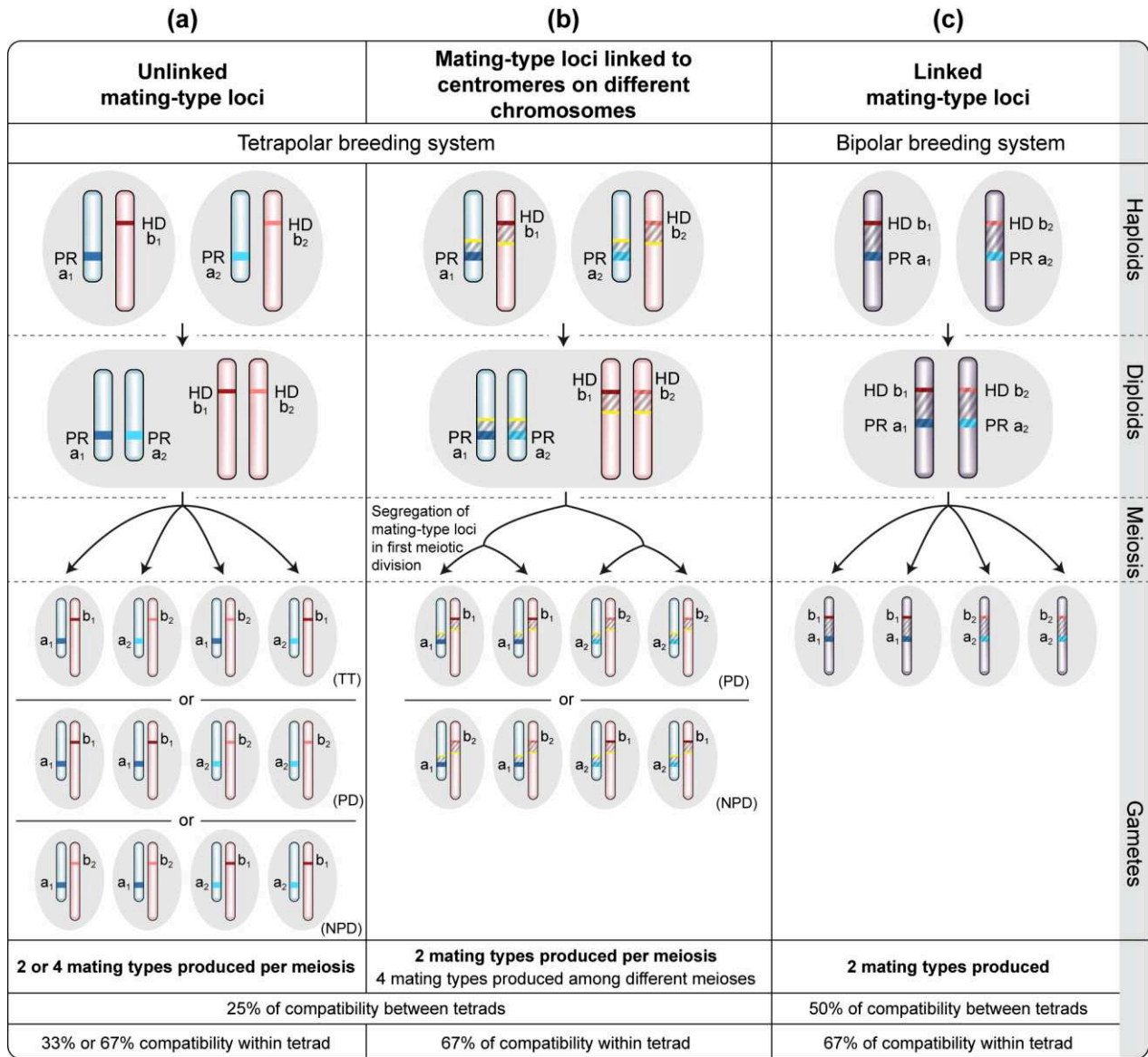
Supplementary Figure 1. Odds of compatibility among gametes of a diploid individual in basidiomycete fungi. In basidiomycetes gametes are fully compatible only if they carry different alleles at both mating-type loci, i.e., the *PR* (including pheromone receptor and pheromone genes, with a1 and a2 alleles) and *HD* (including homeodomain genes, with b1 and b2 alleles) loci. **(a)** when the *PR* and *HD* mating-type loci are completely unlinked (with no mating-type loci or centromere linkage and loci located on different chromosomes, depicted in blue and red, respectively), the percentage of compatibility of a given gamete with the other gametes produced by the same diploid individual is 25% across multiple meioses (a given gamete is compatible with one of every four gametes), and the percentage within tetrad is 33% (a given gamete is compatible with one of the other three gametes in the tetrad) or 67% (a given gamete is compatible with two of the three remaining gametes in the tetrad) depending on segregation of the mating type alleles. The different types of gametes produced are tetratypes (TT), parental ditypes (PD) or non-parental ditypes (NPD), that result from allele segregation and whether crossing-over occurs between one of the two loci and the centromere; **(b)** when the *PR* and *HD* mating-type genes are linked to the centromeres of different chromosomes (blue and red, respectively), the percentage of compatibility of a given gamete with the other gametes produced by the same diploid individual is 25% across multiple meioses but 67% within a tetrad (a given gamete is compatible with two of the three other gametes in the tetrad) due to the segregation of variation occurring only at meiosis I for both mating type loci. The different types of gametes produced are parental ditypes (PD) or non-parental ditypes (NPD), which depend on segregation; **(c)** when the *HD* and *PR* loci are fully linked to each other on the same chromosome, the percentage of compatibility of a given gamete among the other gametes produced by the same diploid individual is 50% across multiple meioses (a given gamete is compatible with one of every two gametes), and 67% within a single meiotic tetrad.

2
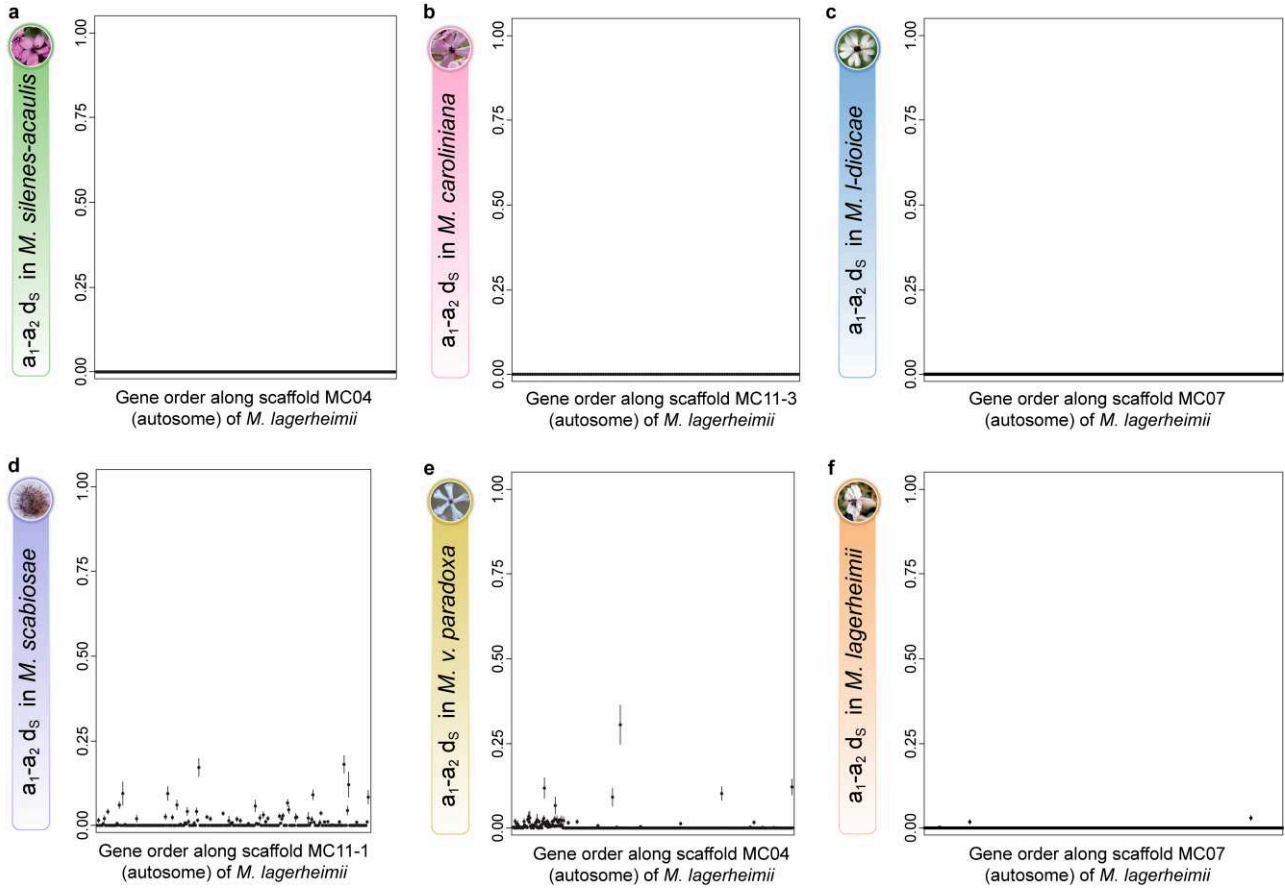
Supplementary Figure 2. Per-gene synonymous divergence and standard error ($d_S \pm SE$) between alleles in the a_1 versus a_2 haploid genomes of an autosome per species, following the gene order of the homologous *Microbotryum lagerheimii* autosomes. Standard errors were estimated with the yn00 program of the PAML package. We selected a representative, well assembled, autosome for each species (number of genes N and contig lengths L are indicated in brackets): **(a)** *M. silenes-acaulis* (N=539 genes, L=1821 kb); **(b)** *M. v. caroliniana* (N=120 genes, L=481 kb); **(c)** *M. lychnidis-dioicae* (N=499 genes, L= 1585 kb); **(d)** *M. scabiosae* (N=235 genes, L= 1030 kb); The relatively high d_S results from the sequencing of haploid genomes from different individuals; **(e)** *M. v. paradoxa* (N=533 genes, L= 1822 kb); The relatively high d_S in one part of autosomes likely results from an outcrossing event followed by a selfing event; **(f)** *M. lagerheimii* (N=547 genes, L= 1585 kb).
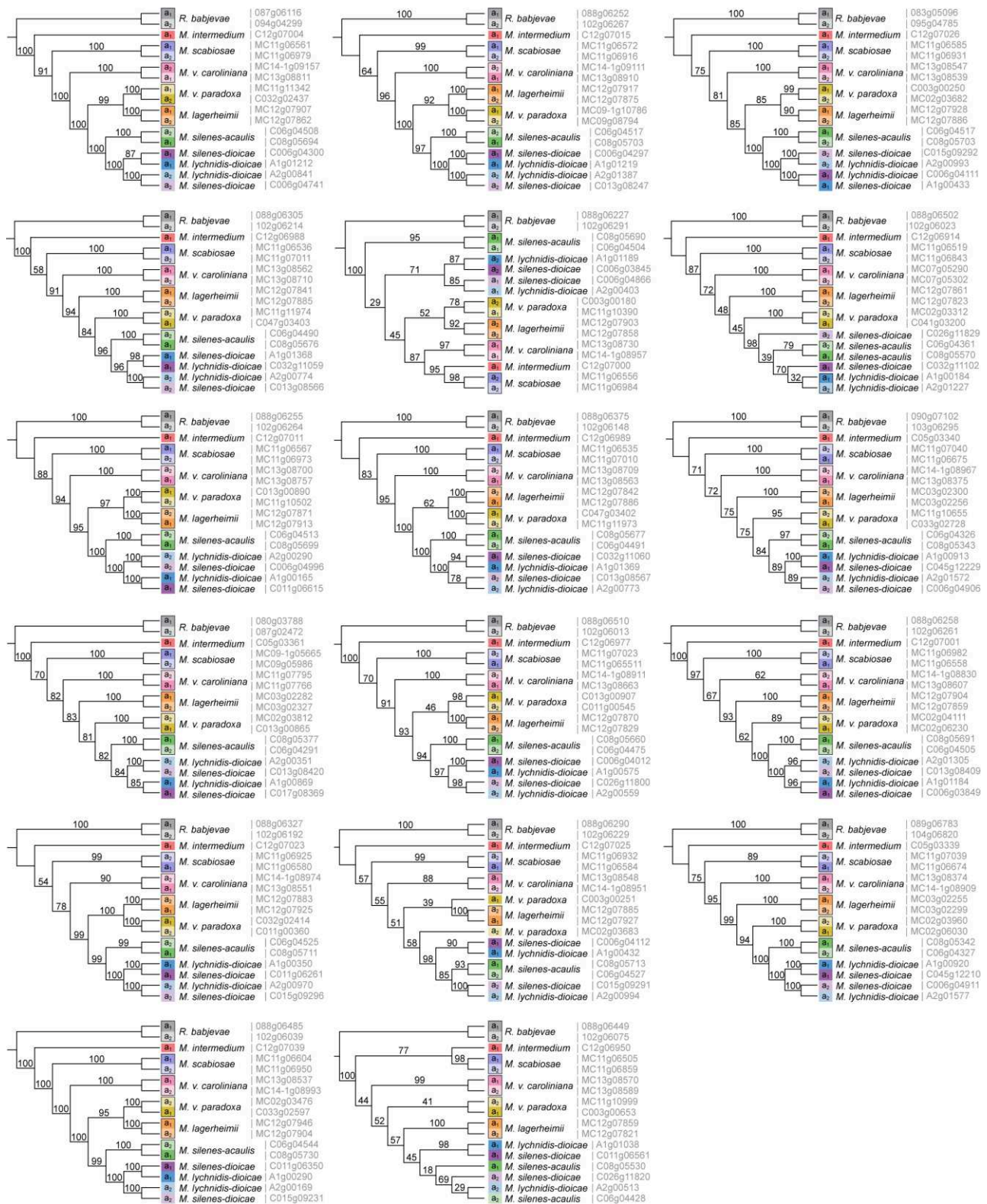
Supplementary Figure 3. Individual genealogies for the 17 genes used for dating the linkage of mating-type loci. The individual genealogies of these 17 genes ancestrally located between the *HD* and *PR* loci illustrate that trans-specific polymorphism in this genomic region occurs only between *Microbotryum lychnidis-dioicae* and *M. silenes-dioicae*, supporting that linkage between the two mating-type loci occurred independently in the other species.
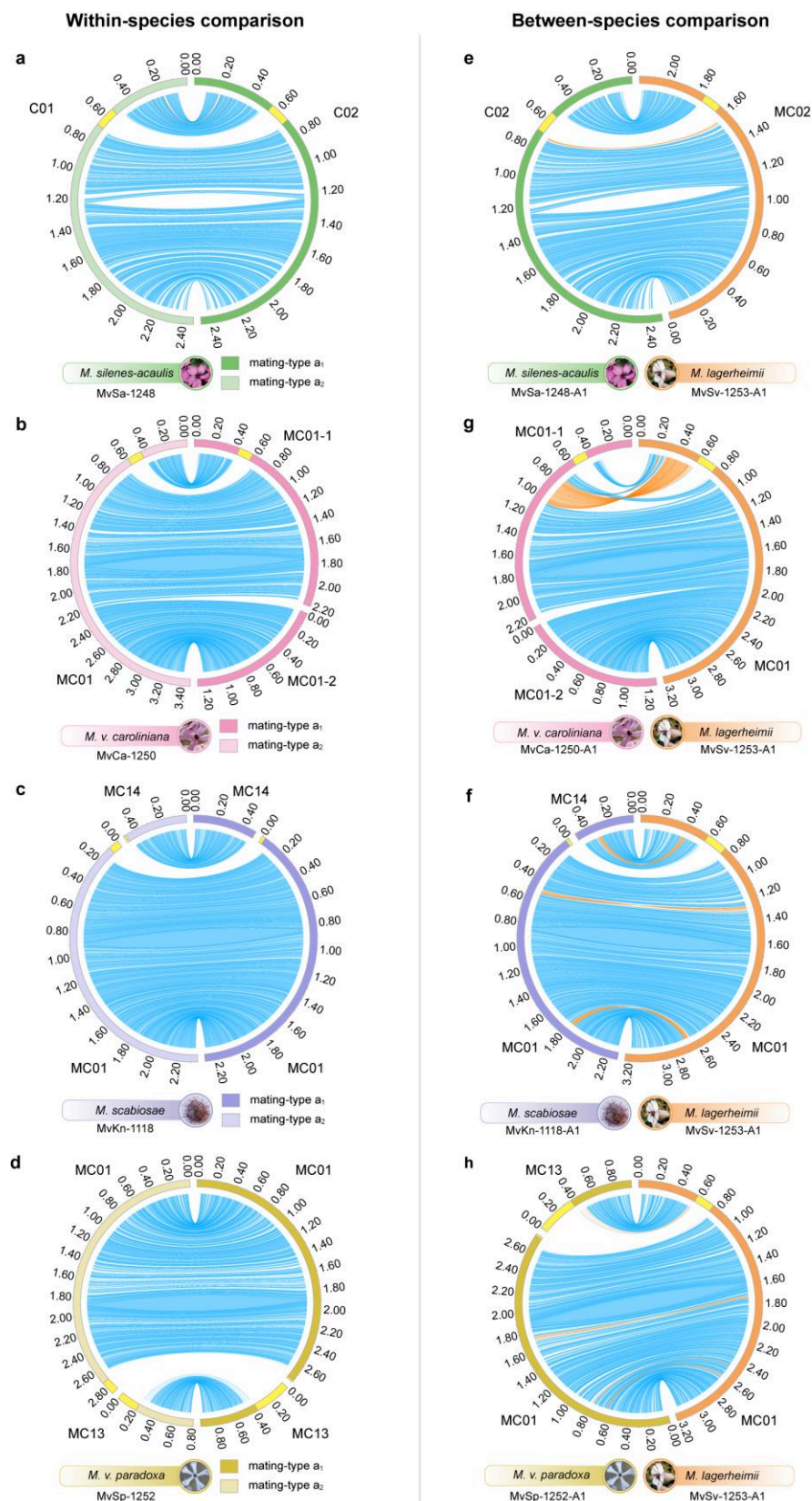
Supplementary Figure 4. Comparison of gene order between autosomes. Blue and orange ribbons connect alleles within-species or single-copy orthologs between-species (i.e. between each focal species and *M. lagerheimii*). The link size is proportional to gene length and orange ribbons represent inversions. Contig size scale is indicated in Megabases. **(a)** to **(h)** Comparisons between the best assembled autosomes within (left) and between (right) species. Yellow regions on the outer track indicate centromeric regions as based on the presence of putative centromere-specific repeats.

i — *M. lagerheimii* MvSv-1253 — mating-type a₁, mating-type a₂

j — *M. silenes-dioicae* MvSd-1303 — mating-type a₁, mating-type a₂

k — *M. silenes-acaulis* MvSa-1248 — mating-type a₁, mating-type a₂

l — *M. v. caroliniana* MvCa-1250 — mating-type a₁, mating-type a₂

m — *M. scabiosae* MvKn-1118 — mating-type a₁, mating-type a₂

n — *M. v. paradoxa* MvSp-1252 — mating-type a₁, mating-type a₂

key to tracks:
Genes
Transposable elements
de-novo detected centromeric repeats
Centromeric repeats previously detected in *M. lychnidis-dioicae*

Supplementary Figure 4 (continued). (i) to (n) Within-species comparisons between *Microbotryum* autosomal contigs (a₁ versus a₂ genomes) with all contigs larger than 40 kb represented (except for *M. silenes-acaulis* where only contigs larger than N90 length are plotted due to genome assembly fragmentation). The internal tracks indicate the following features: 1) predicted genes that do not match transposable elements, 2) predicted transposable elements, 3) *M. lagerheimii de-novo* detected centromeric repeats (using the novel method described in the material and methods), and 4) centromeric repeats previously detected in *M. lychnidis-dioicae*[1].
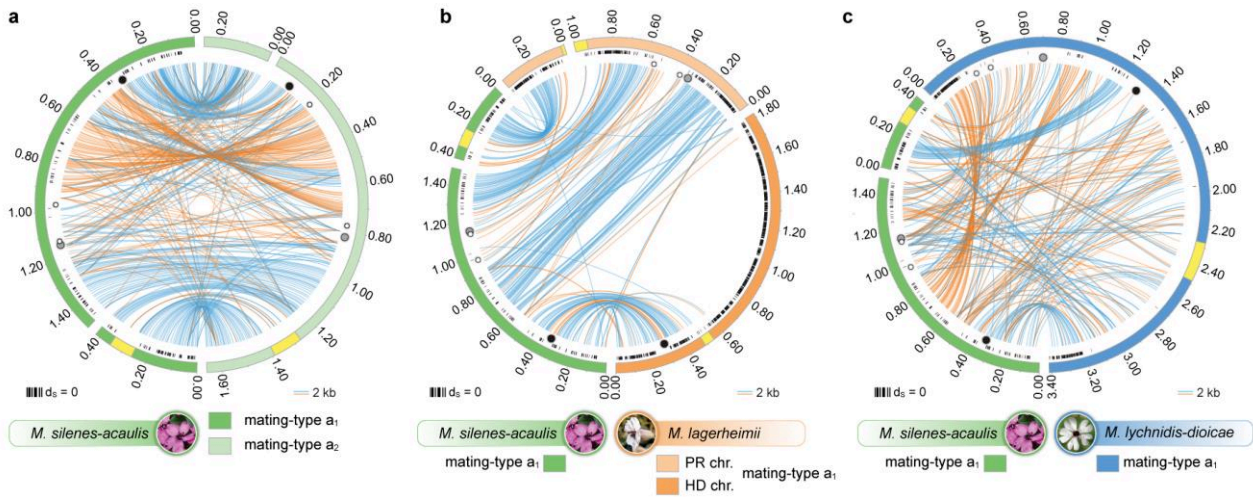
Supplementary Figure 5. Comparison of gene order between mating-type chromosomes in *Microbotryum silenes-acaulis*. The *HD*, *PR* and pheromone genes are indicated by black, grey and white small circles, respectively. Blue and orange lines link single-copy orthologs, the latter corresponding to inversions. Yellow regions on the outer track indicate the putative centromere-specific repeats[1] and black marks along the chromosome/contig track indicate genes that have no synonymous substitutions between a_1 and a_2 alleles within species (d_S=0). Links show contiguous BLASTN similarity across a length (in kb) indicated in each panel. The contig size scale is indicated in Megabases. **(a)** *M. silenes-acaulis* a_2 (left) versus a_1 (right) mating-type chromosomes. The a_1 and a_2 mating-type chromosomes of *M. silenes-acaulis* were assembled in two contigs each. Contig comparison showed they constituted a single chromosome, whose assembly was broken in different locations. The PARs on both edges of the mating-type chromosomes are collinear while the non-recombining region displays a large inversion without significant gene shuffling; **(b)** Comparison between the a_1 mating-type chromosomes of *M. silenes-acaulis* (left) and *M. lagerheimii* (right), taken as a proxy for ancestral gene order[2]. The *M. silenes-acaulis* mating-type chromosome corresponds to the whole *M. lagerheimii PR* chromosome and the small arm of the *M. lagerheimii HD* chromosome. The corresponding mating-type chromosomes in the two species are highly collinear, indicating very recent recombination suppression; **(c)** Comparison between the a_1 mating-type chromosomes of *M. silenes-acaulis* (left) and *M. lychnidis-dioicae* (right). Mating-type chromosomes in the two species appear to be entirely homologous but shared a single PAR (the collinear region close to the *HD* locus). The complete ancestral *PR* chromosome became linked to the same arm of the ancestral *HD* chromosome as in *M. lychnidis-dioicae*, except that the opposite edge of the ancestral *PR* chromosome was juxtaposed to the *HD* chromosome arm. The extremity of the ancestral *PR* chromosome that became a PAR in *M. lychnidis-dioicae* was thus found in the middle of the *M. silenes-acaulis* mating-type chromosomes, and vice-versa, while the two species shared the other PAR (see Fig. 2 for additional details).
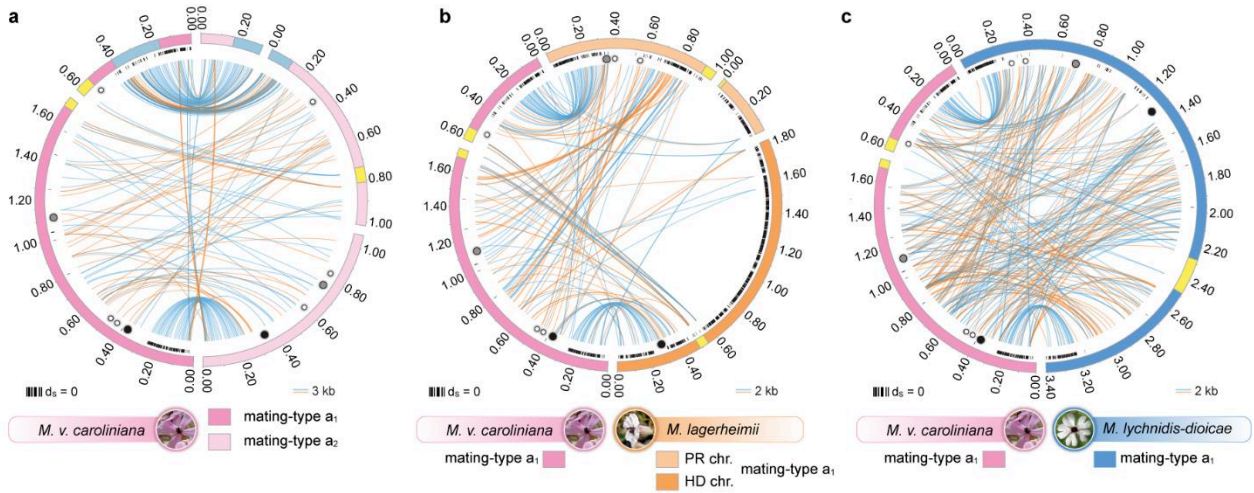
Supplementary Figure 6. Comparison of gene order between mating-type chromosomes in *M. v. caroliniana*. (a) *M. v. caroliniana* a_2 (left) versus a_1 (right) mating-type chromosomes. The a_1 and a_2 mating-type chromosomes of *M. v. caroliniana* were assembled in two contigs each that constitute a single chromosome with assemblies broken in different locations. The two PARs on both edges of the a_1 and a_2 mating-type chromosomes were highly collinear with a_1 - a_2 synonymous substitutions (d_S) close to 0, at the extremities. Conversely, the non-recombining region was large and highly rearranged between the a_1 and a_2 mating-type chromosomes and exhibited non-zero d_S values. The locations of the genes colored in light blue in Fig. 3b are also indicated as light blue boxes on the outer track. **(b)** Comparison between the a_1 mating-type chromosomes of *M. v. caroliniana* (left) and *M. lagerheimii* (right), taken as proxy for the ancestral state[2]. Contrary to the other species where the complete ancestral *PR* chromosome was incorporated as part of the extant mating-type chromosomes, only one arm of each of the *M. lagerheimii PR* and *HD* chromosomes matched the extant *M. v. caroliniana* mating-type chromosome. **(c)** Comparison between a_1 mating-type chromosomes of *M. v. caroliniana* (left) and *M. lychnidis-dioicae* (right). The mating-type chromosome of the two species were mostly homologous and shared two PARs (collinear regions). Contig size scale is indicated in Megabases. All other symbols and features are as described in Supplementary Figure 5.
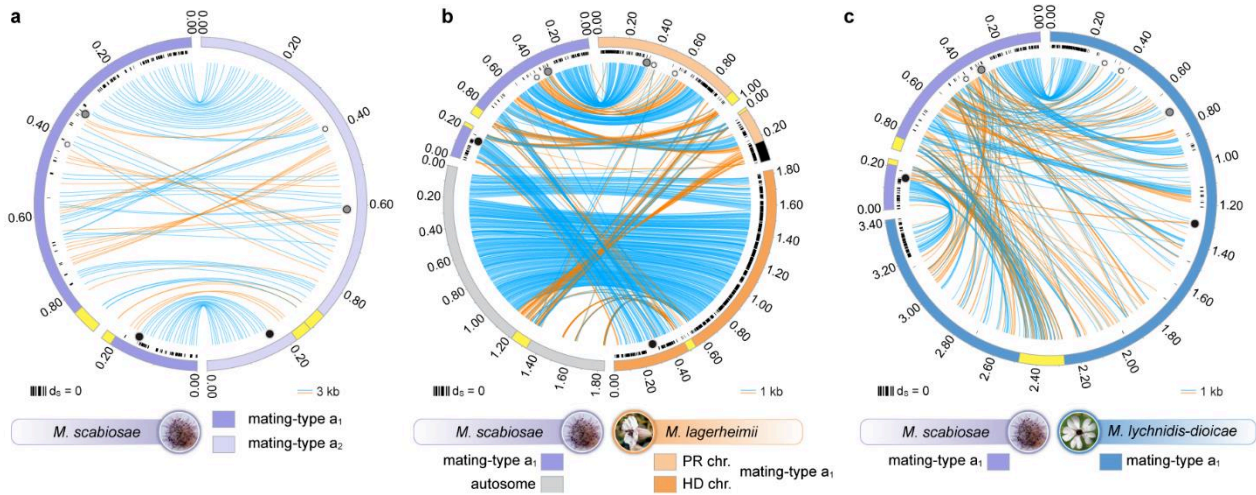
Supplementary Figure 7. Comparison of gene order between mating-type chromosomes in *M. scabiosae*. (a) *M. scabiosae* a_2 (left) versus a_1 (right) mating-type chromosomes. The a_1 and a_2 mating-type chromosomes of *M. scabiosae* were assembled into one and two contigs, respectively, thus constituting a single chromosome. Whereas the PARs on both edges of mating-type chromosomes were collinear and included many zero d_S values, the large non-recombining region had non-zero dS values between a_1 and a_2 mating-types. **(b)** Comparison between the a_1 mating-type chromosome and an autosome of *M. scabiosae* (left) versus a_1 mating-type chromosomes of *M. lagerheimii* (right). The *M. scabiosae* mating-type chromosome corresponds to the *M. lagerheimii PR* chromosome plus the small arm of the *HD* chromosome, resembling patterns in other species. However, the edge of the *M. lagerheimii PR* chromosome (black box on the outer track) corresponded to the center of an *M. scabiosae* autosome suggesting rearrangement within a chromosome arm rather than rearrangement at putative centromeres. **(c)** Comparison between the a_1 mating-type chromosomes of *M. scabiosae* (left) and *M. lychnidis-dioicae* (right). Contig size scale is indicated in Megabases. Symbols and features are as described in Supplementary Figure 5.
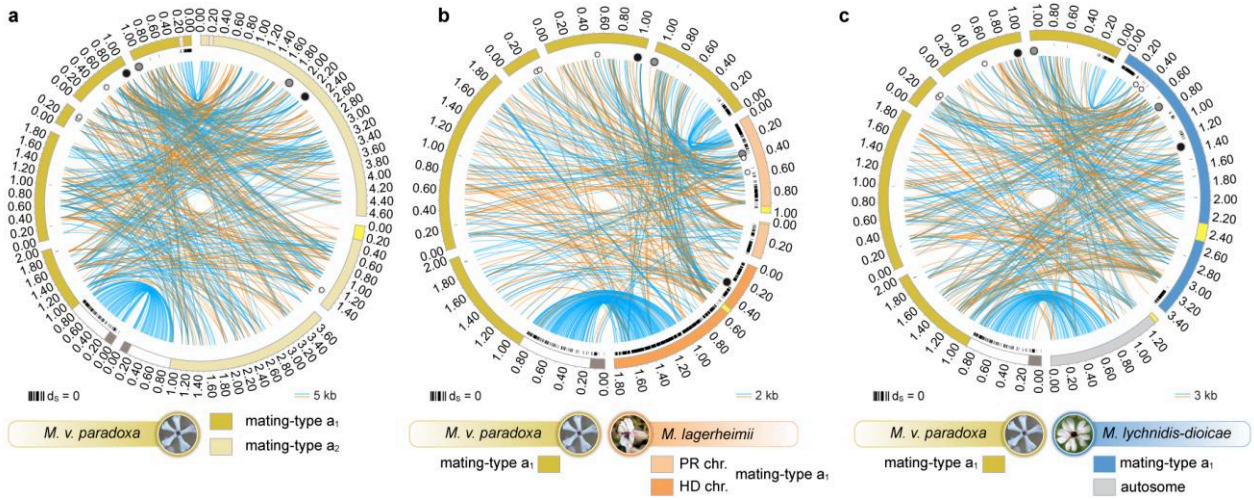
Supplementary Figure 8. Comparison of gene order between mating-type chromosomes in *M. v. paradoxa*. (a) *M. v. paradoxa* a_1 (left) versus a_2 (right) mating-type chromosomes. Some contigs within the non-recombining regions could not be oriented with certainty given the high degree of rearrangements. The locations of the genes colored in pink and white in Fig. 3d are also indicated as pink or white boxes on the outer track. **(b)** Comparison between the a_1 mating-type chromosomes of *M. v. paradoxa* (left) and *M. lagerheimii* (right). Unlike all the other *Microbotryum* species analyzed, where only one arm of the ancestral *HD* chromosome became integrated in the mating-type chromosome, here the whole *HD* and *PR* ancestral chromosomes fused to form the *M. v. paradoxa* mating-type chromosome. An extremity of the *M. v. paradoxa* mating type chromosome (grey region in the outer track) corresponds to rearrangements of regions in the middle of the *M. lagerheimii* mating type chromosomes (see Fig. 2e), supporting complete recombination suppression up to the edge of the *M. v. paradoxa* mating-type chromosome. **(c)** Comparison between the a_1 mating-type chromosome of *M. v. paradoxa* (left) versus a_1 mating-type chromosome and an autosome of *M. lychnidis-dioicae* (right). The mating-type chromosomes in both species were mostly homologous and shared one PAR. The other *M. v. paradoxa* PAR corresponded to the MC12 *M. lychnidis-dioicae* autosome. The grey extremity of the *M. v. paradoxa* mating type chromosome corresponds to rearrangements of regions in the middle of the *M. lychnidis-dioicae* mating type chromosome (Fig. 2e), further supporting complete recombination suppression up to the edge of the *M. v. paradoxa* mating type chromosome. Contig size scale is indicated in Megabases. All other symbols and features are as described in Supplementary Figure 5.
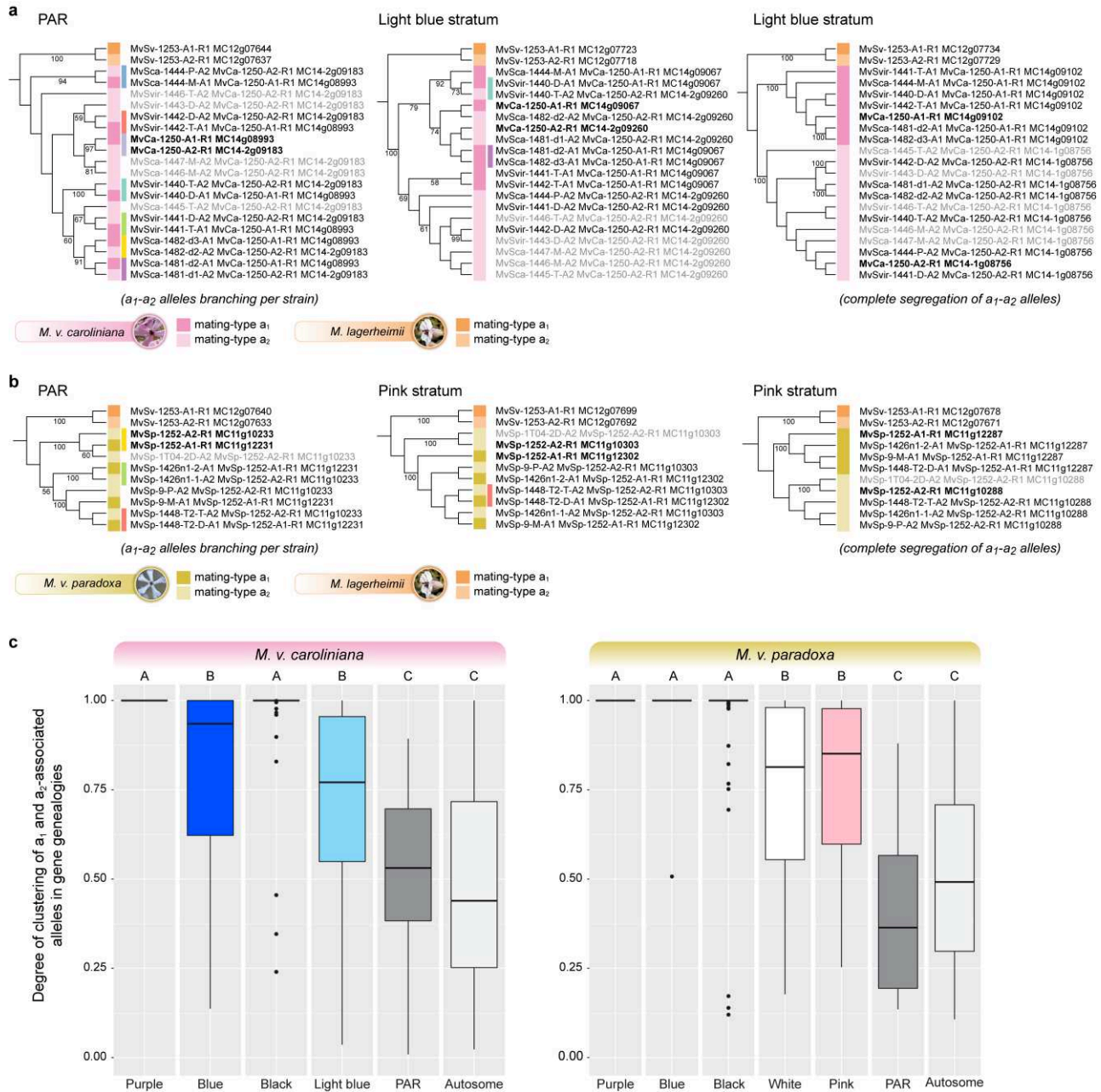
Supplementary Figure 9. Clustering of a$_1$ versus a$_2$-associated alleles in 11 strains of *Microbotryum violaceum caroliniana* and 5 strains of *M. v. paradoxa*. Examples of genes genealogies with different levels of clustering of a$_1$ versus a$_2$-associated alleles in *M. v. caroliniana* **(a)** and in *M. v. paradoxa* **(b)**; grey-colored sequence names represent strains for which only one mating type was sequenced; sequences in bold are from the reference genomes; a$_1$ and a$_2$-associated alleles of the same strain are colored only when clustered in the tree, different diploids having different colors. **(c)** Distribution of the degree of clustering of a$_1$ versus a$_2$-associated alleles in *M. v. caroliniana* (left) and *M. v. paradoxa* (right), in different genomic regions (an autosome as well as the various evolutionary strata and the pseudo-autosomal regions (PARs) in mating-type chromosomes). An index of 1 means that a$_1$ and a$_2$-associated alleles are fully separated in gene genealogies (see Material and Methods). Different capital letters at the top indicate significantly different means (Student's t-tests, Supplementary Table 4). Number of genes analyzed in *M. v. caroliniana* in the different regions: autosome 529, blue stratum 22, purple stratum 10, black stratum 70, light blue stratum 41, PARs 56. Number of genes analyzed in *M. v. paradoxa* in the different regions: autosome 490, blue stratum 11, purple stratum 11, black stratum 150, pink stratum 24, white stratum 206, PARs 19. The boxplots represent the median (center line), the 25th percentile and 75th percentiles (box bounds), the 5th percentile and 95th percentiles (whiskers), and points being the outliers from these 95th and 5th percentiles.

11
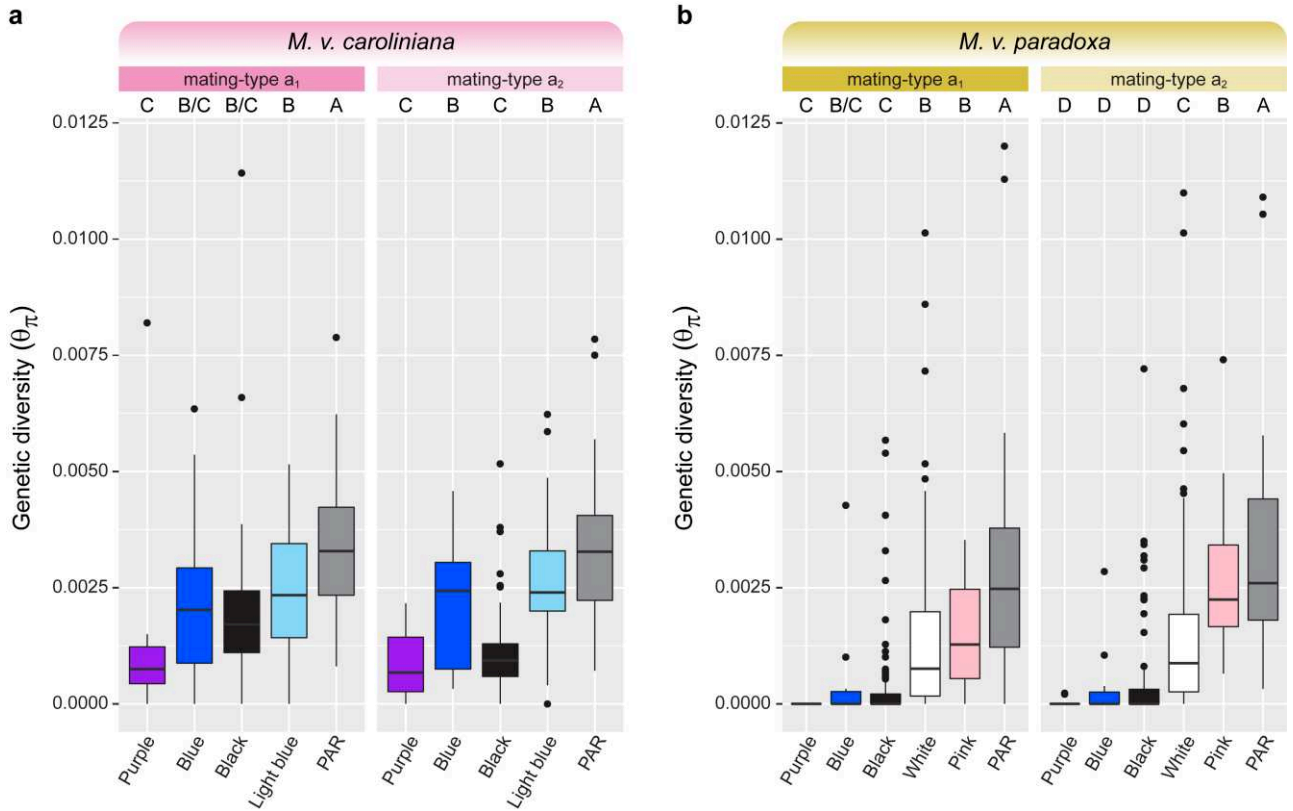
Supplementary Figure 10. Genetic diversity ($\theta\pi$) in (a) 11 strains of *Microbotryum violaceum caroliniana* and (b) 5 strains of *M. v. paradoxa*. Genetic diversity was computed per species and per mating type, in the different genomic regions of mating-type chromosomes (a_1 and a_2): the old shared evolutionary strata (blue and purple), the young species-specific evolutionary strata (black, white, pink and light blue) and the pseudo-autosomal regions (PARs). Different capital letters at the top indicate significantly different means within species for a given mating type (Student's t-tests, Supplementary Table 5). Number of genes analyzed in *M. v. caroliniana* in the different regions: blue stratum 22, purple stratum 10, black stratum 70, light blue stratum 41, PARs 59. Number of genes analyzed in *M. v. paradoxa* in the different regions: blue stratum 11, purple stratum 11, black stratum 150, pink stratum 24, white stratum 206, PARs 19. The boxplots represent the median (center line), the 25th percentile and 75th percentiles (box bounds), the 5th percentile and 95th percentiles (whiskers), the points being the outliers from these 95th and 5th percentiles.

Supplementary Table 1. Statistics on the genomes and mating-type chromosomes of the *Microbotryum* species analyzed in this study, and in the different genomic partitions of the mating-type chromosomes: recombining regions (RR), non-recombining regions (NRR), pseudo-autosomal regions (PAR) for **(a)** *M. lagerheimii* (with unlinked mating type loci; statistics therefore are given for the *HD* and *PR* mating-type chromosomes), **(b)** the remaining species with linked *PR* and *HD* loci, and thus a single mating-type chromosome, and (c) whole genome assembly statistics.

(a)

		M. lagerheimii	
No. of mating-type chromosomes			2
No. of contigs for the	*HD*	a_1 chromosome	1
		a_2 chromosome	2
	PR	a_1 chromosome	2
		a_2 chromosome	2
Size (bp) of the	*HD*	a_1 chromosome	1,823,320
		a_2 chromosome	1,828,681
	PR	a_1 chromosome	1,368,423
		a_2 chromosome	1,300,106
Size (bp / %) of the RR on the	*HD*	a_1 chromosome	1,567,057 / 85.95
		a_2 chromosome	1,577,456 / 86.26
	PR	a_1 chromosome	681,586 / 49.81
		a_2 chromosome	659,051 / 50.69
Size (bp / %) of the NRR on the	*HD*	a_1 chromosome	256,263 / 14.05
		a_2 chromosome	251,225 / 13.74
	PR	a_1 chromosome	686,837 / 50.19
		a_2 chromosome	641,055 / 49.31

(b)

		M. scabiosae	*M. silenes-acaulis*	*M. v. caroliniana*	*M. v. paradoxa*
No. of mating-type chromosomes		1	1	1	1
No. of contigs for the mating-type chromosome	a_1	1	2	2	4
	a_2	2	2	3	3
Size (bp) of the mating-type chromosome	a_1	1,129,805	1,909,079	2,353,735	5,963,545
	a_2	1,163,263	1,924,428	2,385,087	9,963,855
Size (bp / %) of the PAR of the mating-type chromosome	a_1	411,060 / 36.4	996,791 / 52.2	407,308 / 17.3	140,099 / 2.18
	a_2	424,491 / 36.5	1,039,328 / 54.0	395,275 / 16.7	146,725 / 1.47
Size (bp / %) of the NRR of the mating-type chromosome	a_1	718,744 / 63.6	912,288 / 47.8	1,701,704 / 72.3	6,294,107 / 97.82
	a_2	738,771 / 63.5	885,100 / 45.9	1,734,165 / 72.7	9,817,130 / 98.53

Supplementary Table 1 (continued)**.**

(c)

Genome	Contig number	Length (bp)	N50 (bp)	Percent masked for repeats	Reference
M. intermedium	24	23,461,035	1,644,950	0.0625	Branco et al. 2017
M. lagerheimii a_1	42	25,806,628	1,584,903	0.0743	Branco et al. 2017
M. lagerheimii a_2	37	25,667,824	1,585,144	0.0764	Branco et al. 2017
M. lychnidis-dioicae a_1	48	29,901,156	1,736,850	0.1778	Branco et al. 2017
M. lychnidis-dioicae a2	37	30,318,316	1,730,088	0.1719	Branco et al. 2017
M. silenes-acaulis a_1	77	29,349,019	1,490,411	0.1410	This study
M. silenes-acaulis a_2	89	29,858,089	1,555,341	0.1454	This study
M. scabiosae a_1	123	24,532,870	1,157,998	0.0811	This study
M. scabiosae a_2	147	25,410,563	1,198,341	0.0797	This study
M. silenes-dioicae a_1	144	33,593,023	936,137	0.2302	Branco et al. 2017
M. silenes-dioicae a_2	128	33,958,966	1,321,747	0.2331	Branco et al. 2017
M. v. caroliniana a_1	131	28,837,994	1,517,213	0.1433	This study
M. v. caroliniana a_2	137	28,906,851	1,311,234	0.1419	This study
M. v. paradoxa a_1	225	40,892,155	1,035,691	0.2212	This study
M. v. paradoxa a_2	156	39,900,209	1,869,174	0.2196	This study
Rhodotorula babjevae a1	29	21,772,635	1,453,553	0.0025	Branco et al. 2017
Rhodotorula babjevae a2	34	21,676,322	1,321,968	0.0038	Branco et al. 2017

Supplementary Table 2. Comparisons of synonymous substitution (d$_S$) means across genes between evolutionary strata (referred by the colors used in Fig. 3) and the pseudo-autosomal regions (PARs) for *Microbotryum silenes-acaulis*, *M. v. caroliniana*, *M. v. paradoxa* and *M. scabiosae*. Number of genes analyzed in *M. silenes-acaulis* in the different genomic regions: blue and purple strata pooled 20, black stratum 67, pink stratum 24, PARs 118. Number of genes analyzed in *M. scabiosae* in the different regions: blue and purple strata pooled 21, black stratum 66, PARs 69. Number of genes analyzed in *M. v. caroliniana* in the different regions: blue and purple strata pooled 13, black stratum 46, light blue stratum 29, PARs 48. Number of genes analyzed in *M. v. paradoxa* in the different regions: blue and purple strata pooled 8, black stratum 69, pink stratum 17, white stratum 140, PARs 11. **(a)** d$_S$ (mean and standard error) across genes between the large putative strata. **(b)** Analyses of variance (ANOVA) per species. **(c)** Pairwise mean comparisons per species and genomic region (Student's t tests), with statistically significant differences indicated in bold and with an asterisk. The d$_S$ distributions significantly deviated from normality (Shapiro-Wilk tests, W= 0.10, P<00001 for *M. silenes-acaulis*, W= 0.18, P<00001 for *M. v. caroliniana*, W= 0.41, P<00001, for *M. v. paradoxa,* W= 0.18, P<00001 for *M. scabiosae*) and variances were significantly different among genomic regions for some species (Levene tests, F-ratio= 1.00, d.f.=3, P=0.41 for *M. silenes-acaulis*, F-ratio= 2.80, d.f.=4, P=0.028 for *M. v. caroliniana*, F-ratio= 67.09, d.f.=4, P<0.0001 for *M. v. paradoxa,* F-ratio= 4.23, d.f.=4, P=0.01 for *M. scabiosae*). **(d)** Kruskal-Wallis non-parametric tests**,** also showing significant differences among genomic regions.

(a)

Stratum	Mean d$_S$ ± SE			
	M. silenes-acaulis	*M. v. caroliniana*	*M. scabiosae*	*M. v. paradoxa*
Blue and purple strata	0.241±0.149	0.092±0.069	0.090±0.044	0.299±0.365
Black stratum	0.003±0.001	0.102±0.050	0.048±0.013	0.099±0.001
Light blue stratum		0.005±0.001		
Pink stratum				0.012±0.002
White stratum				0.005±0.000
PARs	0.003±0.001	<0.0001	0.051±0.046	<0.0001

(b)

Species	d.f.	Sum of squares	F-ratio	*P-value*
M. silenes-acaulis	2	1.112	10.94	**<0.0001***
M. v. caroliniana	3	0.323	2.3659	0.0739
M. scabiosae	2	0.030	0.20	0.8182
M. v. paradoxa	4	0.860	36.84	**<0.0001***

Supplementary Table 2. Continued.

(c)

M. silenes-acaulis					
Stratum 1	Stratum 2	Mean difference	Standard error	Student's t	P-value
Blue and purple	**Black**	**0.2385**	**0.0556**	**4.28**	**<0.0001***
Blue and purple	**PAR**	**0.2378**	**0.0523**	**4.54**	**<0.0001***
PAR	Black	0.0004	0.0348	0.01	0.9900

M. v. caroliniana					
Stratum 1	Stratum 2	Mean difference	Standard error	Student's t	P-value
Black	**PAR**	**0.1025**	**0.0440**	**2.33**	**0.0215***
Black	Light blue	0.0973	0.0506	1.92	0.0567
Blue and purple	PAR	0.0916	0.0667	1.37	0.1720
Blue and purple	Light blue	0.0865	0.0712	1.21	0.2271
Black	Blue and purple	0.0108	0.0670	0.16	0.8720
Light blue	PAR	0.0052	0.0502	0.10	0.9178

M. scabiosae (degrees of freedom = 153)					
Stratum 1	Stratum 2	Mean difference	Standard error	Student's t	P-value
Blue and purple	Black	0.04216	0.0687	0.61	0.5405
Blue and purple	PAR	0.03906	0.0684	0.57	0.5686
PAR	Black	0.00310	0.0472	0.06	0.9478

M. v. paradoxa (degrees of freedom = 240)					
Stratum 1	Stratum 2	Mean difference	Standard error	Student's t	P-value
Blue and purple	**PAR**	**0.2690**	**0.0355**	**7.58**	**<0.0001***
Blue and purple	**White**	**0.2635**	**0.0277**	**9.49**	**<0.0001***
Blue and purple	**Pink**	**0.2569**	**0.0327**	**7.84**	**<0.0001***
Blue and purple	**Black**	**0.1696**	**0.0285**	**5.94**	**<0.0001***
Black	**PAR**	**0.0994**	**0.0248**	**4.01**	**<0.0001***
Black	**White**	**0.0939**	**0.0112**	**8.35**	**<0.0001***
Black	**Pink**	**0.0873**	**0.0207**	**4.22**	**<0.0001***
Pink	PAR	0.0121	0.0296	0.41	0.6594
Pink	White	0.0066	0.0196	0.34	0.7684
White	PAR	0.0055	0.0239	0.23	0.3679

(d)

Species	d.f.	Chi2	P-value
M. silenes-acaulis	2	5.69	P=0.0581
M. v. caroliniana	3	65.54	**P<0.0001***
M. scabiosae	2	22.22	**P<0.0001***
M. v. paradoxa	4	129.26	, **P<0.0001***

Supplementary Table 3. Strains used for polymorphism analyses: strain ID, *Microbotryum* species and its abbreviation, host species of collection, location of collection, mating type sequenced and coverage of sequencing.

strain ID	*Microbotryum* species and abbreviation	Host species	Location of collection	GPS coordinates	Mating types	Sequencing coverage
1T04	*M. violaceum paradoxa*	*S. paradoxa*	Volpaie, near Lamole, Italy	N42 26.298' E13 34.584	a2	20x
1426	*M. violaceum paradoxa*	*S. paradoxa*	Volpaie, near Lamole, Italy	N42 26.298' E13 34.584	a1 & a2	38x 26x
1448	*M. violaceum paradoxa*	*S. paradoxa*	Volpaie, near Lamole, Italy	N42 26.298' E13 34.584	a1 & a2	50x 42x
9	*M. violaceum paradoxa*	*S. paradoxa*	Volpaie, near Lamole, Italy	43°32'35.7"N 11°21'35.1"E"	a1 & a2	38x 34x
1440	*M. violaceum caroliniana*	*S. caroliniana*	Cliftons Pond, North Carolina	35°59'58.2"N 78°20'51.2"W	a1 & a2	68x 46x
1441	*M. violaceum caroliniana*	*S. virginica*	Route 8, Floyd County, Virginia, US	36°50'35.8"N 80°19'00.1"W	a1 & a2	56x 60x
1442	*M. violaceum caroliniana*	*S. virginica*	Charlottesville Reservoir, Virginia, US	38°01'35.5"N 78°33'31.8"W	a1 & a2	48x 56x
1443	*M. violaceum caroliniana*	*S. virginica*	Charlottesville Reservoir, Virginia, US	38°01'35.5"N 78°33'31.8"W	a2	38x
1444	*M. violaceum caroliniana*	*S. caroliniana*	Blue Ridge Parkway, US	37°44'46.5"N 79°17'59.7"W	a1 & a2	66x 46x
1445	*M. violaceum caroliniana*	*S. caroliniana*	Gilbert Creek, Kentucky, US	37°58'28.9"N 84°50'46.0"W	a2	50x
1446	*M. violaceum caroliniana*	*S. virginica*	Old Garth Road, Charlottesville, Virginia, US	38°03'38.6"N 78°31'57.5"W	a2	56x
1446	*M. violaceum caroliniana*	*S. virginica*	Old Garth Road, Charlottesville, Virginia, US	38°03'38.6"N 78°31'57.5"W	a2	50x
1447	*M. violaceum caroliniana*	*S. virginica*	Sugar Hollow, near Charlottesville, Virginia, US	38°08'16.8"N 78°44'23.3"W	a2	54x
1481	*M. violaceum caroliniana*	*S. caroliniana*	Site 2 Virginia near Briary gap route 257, US	38°25'44.8"N 79°02'09.9"W	a1 & a2	52x 40x
1482	*M. violaceum caroliniana*	*S. caroliniana*	Site 2 Virginia near Briary gap route 257, US	38°25'44.8"N 79°02'09.9"W	a1 & a2	38x

Supplementary Table 4. Comparisons of the degree of clustering of a_1 versus a_2-associated alleles in multiple genomes of *Microbotryum violaceum caroliniana* (a) and *M. v. paradoxa* (b) using Student's t tests, pairwise among the different genomic regions of mating-type chromosomes: the old shared evolutionary strata (blue and purple), the young species-specific evolutionary strata (black, white, pink and light blue) and the pseudo-autosomal regions (PARs). ANOVAs were significant for *M. v. caroliniana* (Sum of squares 3.68, d.f. 5, F-ratio 50.86, P<0.0001) and for *M. v. paradoxa* (Sum of squares 30.95, d.f. 6, F-ratio 91.09, P<0.0001). The index distributions significantly deviated from normality (Shapiro-Wilk tests, W= 0.92 for *M. v. caroliniana*, W= 0.89 for *M. v. paradoxa*; P<00001 for both) and variances were significantly different among genomic regions (Levene tests, F-ratio= 28.31, d.f.=5, P<0.0001 for *M. v. caroliniana*; F-ratio= 41.27, d.f.=6, P<0.0001 for *M. v. paradoxa*). However, non-parametric tests also showed significant differences among genomic regions (Kruskal-Wallis tests, Chi2= 211.35, d.f.=5, P<0.0001 for *M. v. caroliniana*; Chi2= 389.97, d.f.=6, P<0.0001 for *M. v. paradoxa*). Number of genes analyzed in *M. v. caroliniana* in the different regions: autosome 529, blue stratum 22, purple stratum 10, black stratum 70, light blue stratum 41, PARs 56. Number of genes analyzed in *M. v. paradoxa* in the different regions: autosome 490, blue stratum 11, purple stratum 11, black stratum 150, pink stratum 24, white stratum 206, PARs 19.

(a) *M. v. caroliniana*				
Stratum 1	Stratum 2	Difference	Student's t	*P* -value
purple	**Autosome**	**0.5168507**	**6.02**	**<.0001***
black	**Autosome**	**0.4833078**	**14.12**	**<.0001***
purple	**PAR**	**0.4744464**	**5.14**	**<.0001***
black	**PAR**	**0.4409036**	**-9.14**	**<.0001***
purple	**Light blue**	**0.3092683**	**3.26**	**0.0012***
blue	**Autosome**	**0.2921234**	**4.99**	**<.0001***
black	**Light blue**	**0.2757254**	**-5.21**	**<.0001***
blue	**PAR**	**0.2497192**	**-3.69**	**0.0002***
purple	**blue**	**0.2247273**	**2.19**	**0.0289***
Light blue	**Autosome**	**0.2075824**	**4.76**	**<.0001***
black	**blue**	**0.1911844**	**-2.91**	**0.0038***
Light blue	**PAR**	**0.1651781**	**-2.99**	**0.0029***
blue	Light blue	0.0845410	-1.19	0.2349
PAR	Autosome	0.0424042	1.12	0.2625
purple	black	0.0335429	0.37	0.7124

Supplementary Table 4. Continued.

(b) *M. v. paradoxa*				
Stratum 1	Stratum 2	Difference	Student's t	*P* -value
purple	**PAR**	**0.5935789**	**6.58**	**<.0001***
black	**PAR**	**0.5684323**	**-9.8**	**<.0001***
blue	**PAR**	**0.5487608**	**-6.09**	**<.0001***
purple	**Autosome**	**0.4886918**	**6.73**	**<.0001***
black	**Autosome**	**0.4635452**	**20.87**	**<.0001***
blue	**Autosome**	**0.4438737**	**6.12**	**<.0001***
pink	**PAR**	**0.3590373**	**4.91**	**<.0001***
white	**PAR**	**0.3400984**	**5.96**	**<.0001***
pink	**Autosome**	**0.2541502**	**5.11**	**<.0001***
purple	**white**	**0.2534806**	**-3.44**	**0.0006***
white	**Autosome**	**0.2352113**	**11.90**	**<.0001***
purple	**pink**	**0.2345417**	**2.71**	**0.0069***
black	**white**	**0.2283339**	**-8.94**	**<.0001***
black	**pink**	**0.2093950**	**-4.01**	**<.0001***
blue	**white**	**0.2086624**	**-2.83**	**0.0047***
blue	**pink**	**0.1897235**	**-2.19**	**0.0288***
Autosome	PAR	0.1048871	-1.88	0.0598
purple	blue	0.0448182	0.44	0.6588
purple	black	0.0251467	0.34	0.7352
black	blue	0.0196715	-0.26	0.7914
pink	white	0.0189389	-0.37	0.7122

Supplementary Table 5. Comparisons of the genetic diversity ($\theta\pi$) among genomic regions in multiple genomes of (a) *Microbotryum violaceum caroliniana* a_1 (b) *M. v. caroliniana* a_2 (c) *M. v. paradoxa* a_1 and (d) *M. v. paradoxa* a_2 using Student's t tests, comparing the different genomic regions of mating-type chromosomes (a_1 and a_2): the old shared evolutionary strata (blue and purple), the young species-specific evolutionary strata (black, white, pink and light blue) and the pseudo-autosomal regions (PARs). ANOVAs were significant for *M. v. caroliniana* a_1 (sum of squares 0.000075, d.f. 4, F-ratio 8.37, P<0.0001), *M. v. caroliniana* a_2 (sum of squares 0.000181, d.f. 4, F-ratio 29.01, P<0.0001), *M. v. paradoxa* a_1 (sum of squares 0.00021, d.f. 5, F-ratio 21.57, P<0.0001) and *M. v. paradoxa* a_2 (sum of squares 0.00028, d.f. 5, F-ratio 24.47, P<0.0001). The genetic diversity ($\theta\pi$) distributions significantly deviated from normality (Shapiro-Wilk tests, W= 0.90, P<0.00001 for *M. v. caroliniana* a_1, W= 0.93, P<0.0001 for *M. v. caroliniana* a_2, W= 0.65, P<0.00001 for *M. v. paradoxa* a_1, = 0.69, P<0.00001 for *M. v. paradoxa* a_2) and variances were significantly different among genomic regions (Levene tests, F-ratio=7.96, d.f.=4, P<0.0001 for *M. v. caroliniana* a_1; F-ratio=37.78, d.f.=4, P<0.0001 for *M. v. caroliniana* a_2; F-ratio=43.31, d.f.=5, P<0.0001 for *M. v. paradoxa* a_2). However, non-parametric tests also showed significant differences among genomic regions (Kruskal-Wallis tests, Chi2= 48.32, d.f.=4, P<0.0001 for *M. v. caroliniana* a_1; Chi2= 82.97, d.f.=4, P<0.0001 for *M. v. caroliniana* a_2; Chi2= 130.32, d.f.=5, P<0.0001 for *M. v. paradoxa* a_1; Chi2= 149.58, d.f.=5, P<0.0001 for *M. v. paradoxa* a_2). Number of genes analyzed in *M. v. caroliniana* in the different regions: blue stratum 22, purple stratum 10, black stratum 70, light blue stratum 41, PARs 59. Number of genes analyzed in *M. v. paradoxa* in the different regions: blue stratum 11, purple stratum 11, black stratum 150, pink stratum 24, white stratum 206, PARs 19.

(a) *M. v. caroliniana a_1*				
Stratum 1	Stratum 2	Difference	Student's t	*P*-value
PAR	**purple**	**0.0019131**	**-3.74**	**0.0002***
PAR	**black**	**0.0013819**	**5.23**	**<.0001***
PAR	**blue**	**0.0010881**	**2.91**	**0.0040***
Light blue	**purple**	**0.0010654**	**-2.02**	**0.0448***
PAR	**Light blue**	**0.0008476**	**2.79**	**0.0058***
blue	purple	0.0008249	-1.45	0.1498
Light blue	black	0.0005343	1.81	0.0709
black	purple	0.0005311	-1.05	0.2949
blue	black	0.0002938	0.80	0.4226
Light blue	blue	0.0002405	0.61	0.5437

(b) *M. v. caroliniana a_2*				
Stratum 1	Stratum 2	Difference	Student's t	*P*-value
PAR	**purple**	**0.0024554**	**-5.75**	**<.0001***
PAR	**black**	**0.0021985**	**9.96**	**<.0001***
Light blue	**purple**	**0.0017719**	**-4.02**	**<.0001***
Light blue	**black**	**0.0015151**	**6.17**	**<.0001***
blue	**purple**	**0.0013512**	**-2.84**	**0.0050***
PAR	**blue**	**0.0011042**	**3.54**	**0.0005***
blue	**black**	**0.0010943**	**3.58**	**0.0004***
PAR	**Light blue**	**0.0006834**	**2.69**	**0.0077***
Light blue	blue	0.0004208	1.27	0.2038
black	purple	0.0002568	-0.61	0.5436

Supplementary Table 5. Continued.

(c) *M. v. paradoxa a₁*				
Stratum 1	Stratum 2	Difference	Student's t	*P* -value
PAR	**purple**	**0.0032774**	**-6.35**	**<.0001***
PAR	**black**	**0.0030036**	**9.01**	**<.0001***
PAR	**blue**	**0.0027362**	**5.68**	**<.0001***
PAR	**white**	**0.0020227**	**-6.23**	**<.0001***
PAR	**pink**	**0.0017545**	**-2.24**	**<.0001***
pink	**purple**	**0.0015229**	**-4.72**	**0.0029***
white	**purple**	**0.0012547**	**2.95**	**0.0039***
pink	**black**	**0.0012491**	**6.85**	**<.0001***
pink	blue	0.0009818	**4.02**	0.0541
white	**black**	**0.0009809**	**6.52**	**<.0001***
white	blue	0.0007136	2.13	0.0994
blue	purple	0.0005411	-0.60	0.3639
black	purple	0.0002738	-0.68	0.5305
pink	white	0.0002682	-3.74	0.3736
blue	black	0.0002674	0.13	0.5402

(d) *M. v. paradoxa a₂*				
Stratum 1	Stratum 2	Difference	Student's t	*P* -value
PAR	**purple**	**0.0035103**	**-6.20**	**<.0001***
PAR	**black**	**0.0032006**	**8.83**	**<.0001***
PAR	**blue**	**0.0031398**	**5.17**	**<.0001***
pink	**purple**	**0.0025074**	**-2.99**	**<.0001***
pink	**black**	**0.0021978**	**4.07**	**<.0001***
PAR	**white**	**0.0021792**	**-6.04**	**<.0001***
pink	**blue**	**0.0021370**	**1.93**	**<.0001***
white	**purple**	**0.0013311**	**2.90**	**0.0034***
pink	**white**	**0.0011764**	**-0.89**	**0.0002***
white	**black**	**0.0010214**	**6.55**	**<.0001***
PAR	**pink**	**0.0010028**	**-4.09**	**0.0257***
white	**blue**	**0.0009606**	**1.65**	**0.0339***
blue	purple	0.0003705	**-0.91**	0.5516
black	purple	0.0003097	**-0.63**	0.4970
blue	black	0.0000608	**0.61**	0.8939

Supplementary Table 6. Analyses of variance (ANOVA) testing the effect of species, genomic region (autosomes, PARs, and the different evolutionary strata) and mating-type (a₁ vs. a₂). (a) Transposable element proportions in six species, two mating types and 10 genomic regions (an autosome, pseudo-autosomal regions, and the 12 evolutionary strata across species that could be delimited in the current state and used for TE count). The TE content distribution significantly deviated from normality (Shapiro-Wilk tests, W= 0.74 P<0001) and variances were significantly different among genomic regions (Levene test, F-ratio= 4.55, d.f.=11, P<0.0001). However, a non-parametric test also showed significant differences among genomic regions (Kruskal-Wallis test, Chi2= 27.25, d.f.=11, P=0.004). **(b) Gene loss proportions** in six species, two mating types and eight genomic regions (an autosome, pseudo-autosomal regions, and the seven evolutionary strata across species that could be delimited in the current state and for which gene losses could be assessed). Statistically significant results are indicated in bold and with an asterisk. The gene loss distribution significantly deviated from normality (Shapiro-Wilk tests, W= 0.77, P<0001) and variances were significantly different among genomic regions (Levene test, F-ratio= 11.70, d.f.=7, P<0.0001). However, a non-parametric test also showed significant differences among genomic regions (Kruskal-Wallis test, Chi2= 40.55, d.f.=7, P<0.0001).

(a)

Source	DF	Sum of squares	F ratio	P-value
Genomic regions	11	0.0533	94000	**<0.0001***
Species	5	0.0084	3.271	**0.0170***
Mating type	1	<0.001	0.078	0.7822

(b)

Source	DF	Sum of squares	F ratio	P-value
Genomic regions	6	1.3700	22.397	**<0.0001***
Species	5	0.2346	4.602	**0.0027***
Mating type	1	0.0010	0.099	0.7552

Supplementary Note 1: Script for the computation of the clustering index in gene genealogies

```
##
## Rscript for paper 'Multiple convergent supergene evolution events in mating-type chromosomes'
## Branco et al., In revision.
##

require(ape)
## this function takes as input the nodal distance matrix of a tree and returns
## an array obtained by successively sampling (randomly in the case of multiple identical distances)
## the minimum distance between A1s and A2s.
minvals<-function(m) {
        A1<-grep("A1",colnames(m))
        A2<-grep("A2",colnames(m))
        M<-melt(m[A1,A2])
        cpt<-0
        val<-array()
        while(nrow(M)>1) {
                cpt<-cpt+1
                MiNis<-which(M[,3]==min(M[,3]))
                if (length(MiNis)==1) mini<-MiNis
                else mini<-sample(MiNis,1)
                val[cpt]<-M[mini,3]
                names<-c(as.character(M[mini,1]), as.character(M[mini,2]))
                M<-M[!is.element(M[,1],names)&!is.element(M[,2],names),]
        }
        return(val)
}




##for a given tree 'tr', this function
##returns the index described in the M&M.
ComputeScore<-function(tr) {
        tr<-drop.tip(tr, tr$tip.label[grep("^MvSv-", tr$tip.label)]) ##remove outgroup
        tr<-compute.brlen(tr,1) ##transform patristic to nodal distance (we only use topology for the
test)
        tr$edge.length[is.element(tr$edge[,1],which(as.numeric(tr$node.label)<30)+Ntip(tr))&tr$edge[,2
]>Ntip(tr)]<-0 ##Suppress badly supported nodes (bootstrap < 30)
        tr<-di2multi(tr)
        tr<-compute.brlen(tr,1)
        mat<-cophenetic(tr) ##compute pairwise distance matrix

        meandist1<-mean(replicate(10,mean(minvals(mat)))) #replicate 10 times the computation of the
observed score with mincals (see M&M)
        meanperm<-array()
        for (i in 1:1000) {
                #reshuffling of the matrix
                new<-sample(colnames(mat))
                colnames(mat)<-new
                rownames(mat)<-new
                mat<-mat[new,new]
                #computation of the score after reshuffling
                meanperm[i]<-mean(minvals(mat))
        }
        res<-sum(meandist1>=meanperm)/1000 ##non parametric test: proportion of the reshuffling giving
smaller value of the computed score
        res
}
```

Supplementary References

1   Badouin, H., Hood, M. E., Gouzy, J., Aguileta, G., Siguenza, S., Perlin, M. H., Cuomo, C. A., Fairhead, C., Branca, A. & Giraud, T. Chaos of rearrangements in the mating-type chromosomes of the anther-smut fungus *Microbotryum lychnidis-dioicae*. *Genetics* **200**, 1275-1284 (2015).
2   Branco, S., Badouin, H., Rodríguez de la Vega, R., Gouzy, J., Carpentier, F., Aguileta, G., Siguenza, S., Brandenburg, J., Coelho, M., Hood, M. & Giraud, T. Evolutionary strata on young mating-type chromosomes despite lack of sexual antagonism. *Proceedings of the National Academy of Sciences of the United States of America* **114**, 7067-7072 (2017).

S.4 Convergent recombination cessation between mating-type genes and centromeres

Supplemental Material

Convergent recombination cessation between mating-type genes and centromeres in selfing anther-smut fungi

Fantin Carpentier*†[1], Ricardo C. Rodríguez de la Vega†[1], Sara Branco[1,2], Alodie Snirc[1], Marco A. Coelho[3,4], Michael E. Hood#[5], Tatiana Giraud#[1]

Table of contents

key to tracks:
- $d_S = 0$
- Transposable elements (TE)
- Genes
- Centromeric repeats

Links:
- one-to-one orthologs in the same orientation
- one-to-one orthologs in inverted orientation

Supplemental Fig. S1. Interspecific gene order comparison of PR and HD mating-type chromosomes of *Microbotryum lagerheimii* and *M. saponariae*. The outer track represents contigs, staggered every 200 kilobases. The HD, PR and pheromone genes are indicated by blue, dark purple and small light-purple circles, respectively. Blue and orange lines link single-copy orthologs, the latter corresponding to inversions. The link width is proportional to the corresponding gene length. Yellow regions on the contig track indicate centromeres, *i.e.* regions with low gene density, high TE density and enriched in tandem repeats (pink marks). The black marks along the contigs track indicate genes that have no synonymous substitutions between mating types within individuals ($d_S=0$). Green marks indicate the transposable elements (TEs) and grey marks non-TE genes. Pink tracks indicate the position of *de novo* detected tandem repeats. **(A)** Comparison of the *M. saponariae* (left, red) and *M. lagerheimii* (right, orange) b_1 HD chromosomes. **(B)** Comparison of the *M. saponariae* (left, light red) and *M. lagerheimii* (right, light orange) b_2 HD chromosomes. **(C)** Comparison of the *M. saponariae* b_1 (left, red) and *M. lagerheimii* b_2 (right, light orange) HD chromosomes. **(D)** Comparison of the *M. saponariae* b_2 (left, light red) and *M. lagerheimii* b_1 (right, orange) HD chromosomes. **(E)** Comparison of the *M. saponariae* (left, red) and *M. lagerheimii* (right, orange) a_1 PR mating-type chromosomes. **(F)** Comparison of the *M. saponariae* (left, light red) and *M. lagerheimii* (right, light orange) PR a_2 mating-type chromosomes. **(G)** Comparison of the *M. saponariae* a_1 (left, red) and *M. lagerheimii* a_2 (right, light orange) PR mating-type chromosomes. **(H)** Comparison of the *M. saponariae* a_2 (left, light red) and *M. lagerheimii* a_1 (right, orange) PR mating-type chromosomes.

Supplemental Fig. S2

A

HD b$_2$

HD b$_1$

M. intermedium   *M. lagerheimii*

HD b$_2$ mating-type chr.   HD b$_1$ mating-type chr.

B

HD b$_2$

HD b$_2$

M. intermedium   *M. lagerheimii*

HD b$_2$ mating-type chr.   HD b$_2$ mating-type chr.

C

PR a$_2$

PR a$_1$

M. intermedium   *M. lagerheimii*

PR a$_2$ mating-type chr.   PR a$_1$ mating-type chr.

D

PR a$_2$

PR a$_2$

M. intermedium   *M. lagerheimii*

PR a$_2$ mating-type chr.   PR a$_2$ mating-type chr.

key to tracks:
- ■ $d_S = 0$
- ■ Transposable elements (TE)
- ■ Genes

Links:
- — one-to-one orthologs in the same orientation
- — one-to-one orthologs in inverted orientation

4

Supplemental Fig. S2. Comparison of gene order between the mating-type chromosomes of *Microbotryum lagerheimii* and *M. intermedium*. The outer track represents contigs, staggered every 200 kilobases. The HD, PR and pheromone genes are indicated by blue, dark purple and small light-purple circles, respectively. Blue and orange lines link single-copy orthologs, the latter corresponding to inversions. The link width is proportional to the corresponding gene length. Yellow regions on the contig track indicate centromeres. The black marks along the contigs track indicate genes with no synonymous substitutions between mating-type alleles within individuals ($d_S=0$). Because only one haploid genome was available for *M. intermedium*, no d_S values were computed. Green marks indicate transposable elements (TEs) and grey marks non-TE genes. **(A)** Comparison of the *M. intermedium* b_2 (left, pink) and *M. lagerheimii* b_1 (right, orange) HD chromosomes. **(B)** Comparison of the *M. intermedium* b_2 (left, pink) and *M. lagerheimii* b_2 (right, light orange) HD chromosomes. **(C)** Comparison of the *M. intermedium* a_2 (left, pink) and *M. lagerheimii* a_1 (right, orange) PR chromosomes. **(D)** Comparison of *M. intermedium* a_2 (left, pink) and *M. lagerheimii* a_2 (right, light orange) PR chromosomes.

Supplemental Fig. S3

key to tracks:
- ■ $d_S = 0$
- ■ Transposable elements (TE)
- ■ Genes

Links:
- ■ one-to-one orthologs in the same orientation
- ■ one-to-one orthologs in inverted orientation
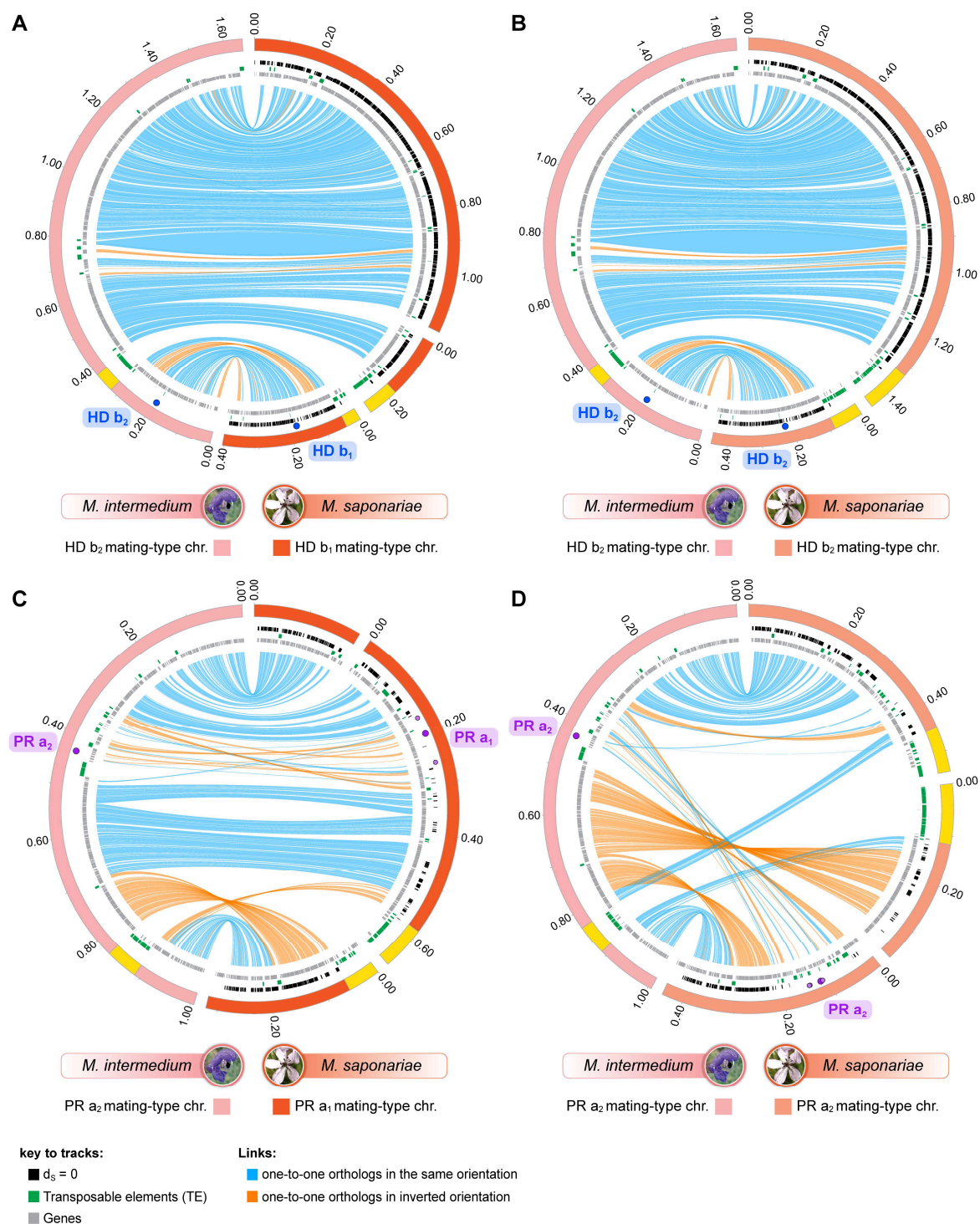
Supplemental Fig. S3. Comparison of gene order between the mating-type chromosomes of *Microbotryum saponariae* and *M. intermedium*. The outer track represents contigs, staggered every 200 kilobases. The HD, PR and pheromone genes are indicated by blue, dark purple and small light-purple circles, respectively. Blue and orange lines link single-copy orthologs, the latter corresponding to inversions. The link width is proportional to the corresponding gene length. Yellow regions on the contig track indicate centromeres. The black marks along the contigs track indicate genes with no synonymous substitutions between mating-type alleles within individuals ($d_S=0$). Because only one haploid genome was available for *M. intermedium*, no d_S values were computed. Green marks indicate transposable elements (TEs) and grey marks non-TE genes. **(A)** Comparison of the *M. intermedium* b_2 (left, pink) and *M. saponariae* b_1 (right, red) HD chromosomes. **(B)** Comparison of the *M. intermedium* (left, pink) and *M. saponariae* (right, light red) b_2 HD chromosomes. **(C)** Comparison of the *M. intermedium* a_2 (left, pink) and *M. saponariae* a_1 (right, red) PR chromosomes. **(D)** Comparison of the *M. intermedium* (left, pink) and *M. saponariae* (right, light red) a_2 PR chromosomes.
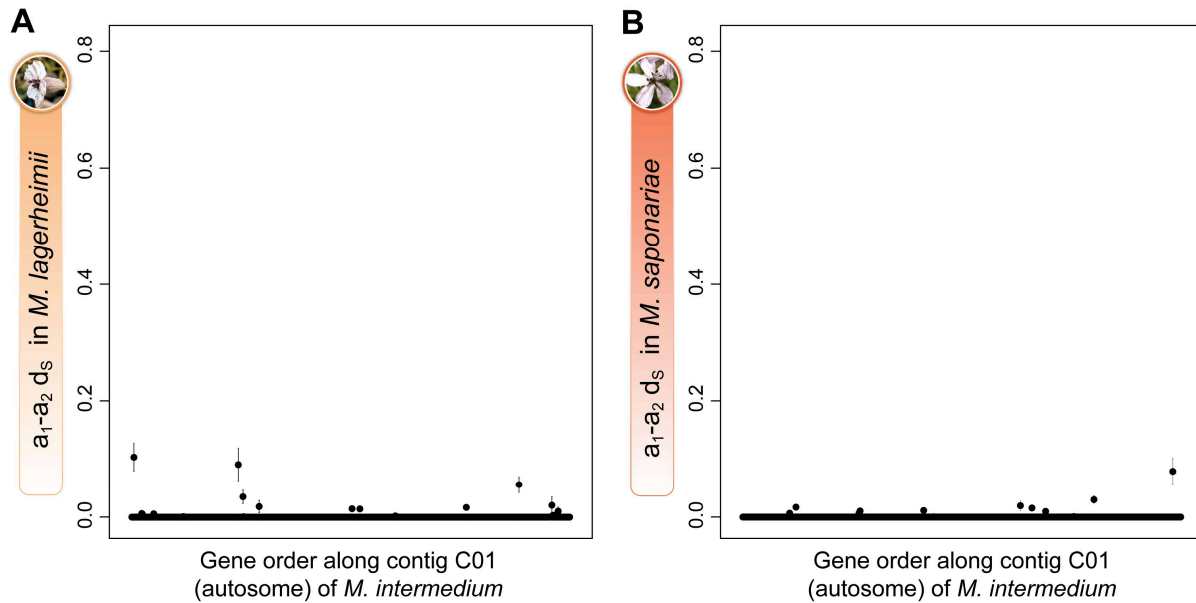
Supplemental Fig. S4. Per-gene synonymous divergence between mating types and its respective standard error (dS ± SE) between autosomal alleles of (A) *Microbotryum lagerheimii* and (B) *M. saponariae*, along the ancestral gene order of a *M. intermedium* autosome. Synonymous divergence is plotted against the genomic coordinates of an autosome of *M. intermedium* for all single-copy genes shared by this autosome, as a proxy for ancestral gene order. Almost all the autosomal genes show a null synonymous divergence between mating types within individuals, as expected in highly selfing organisms such as anther-smut fungi.

Supplemental Fig. S5
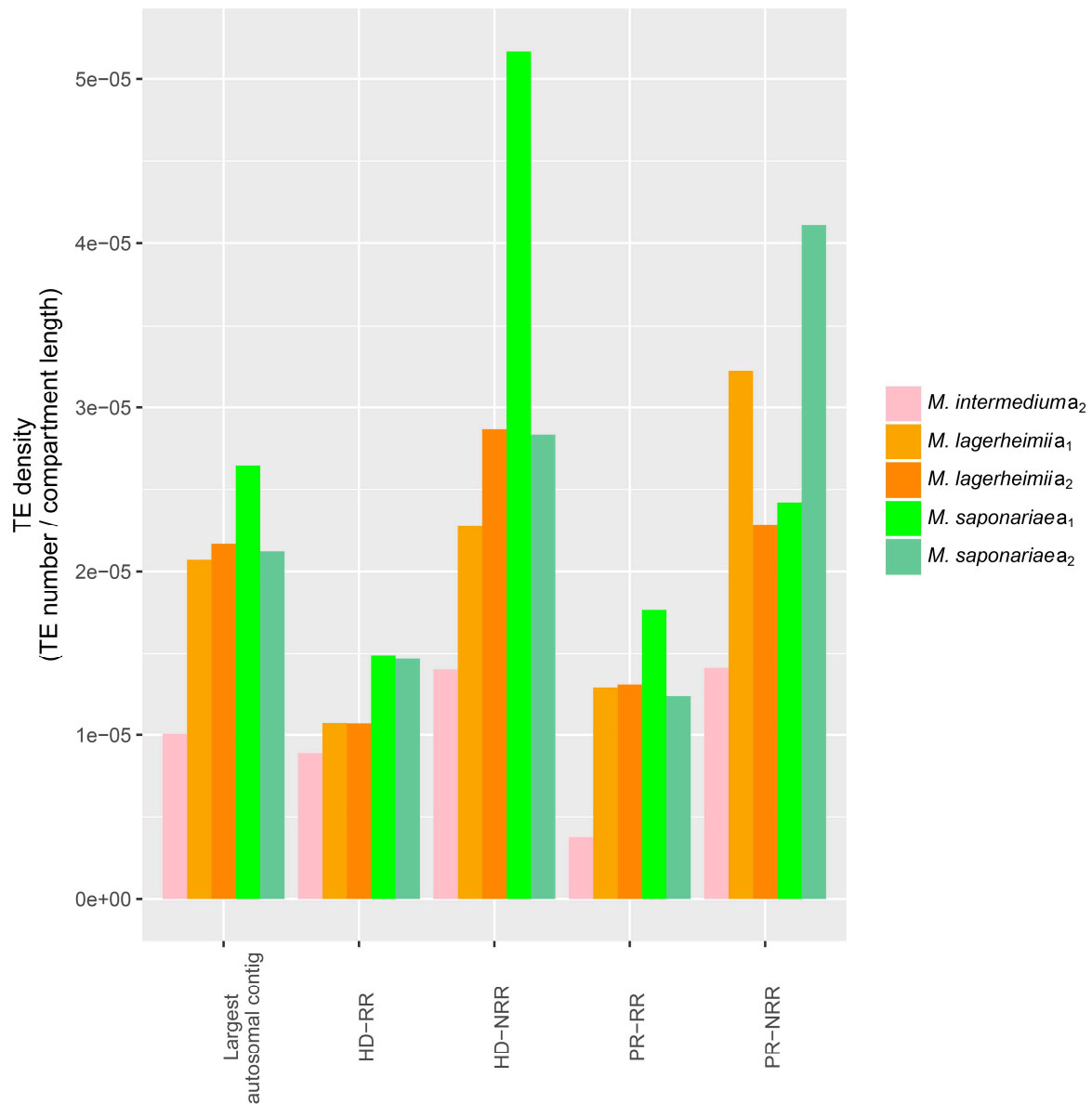
Supplemental Fig. S5. Density of transposable elements (TEs) in *Microbotryum* species. The TE density (number of TEs detected divided by the compartment length) is plotted per genomic compartment, i.e. for the largest autosomal contig, the recombining region, RR, the non-recombining region, NRR, of either the PR or HD mating-type chromosomes, for each available haploid genome of the three studied species with mating types segregating independently: *Microbotryum intermedium* (a_2 genome), *M. lagerheimii* and *M. saponariae* (a_1 and a_2 genomes).

Supplemental Fig. S6A

A
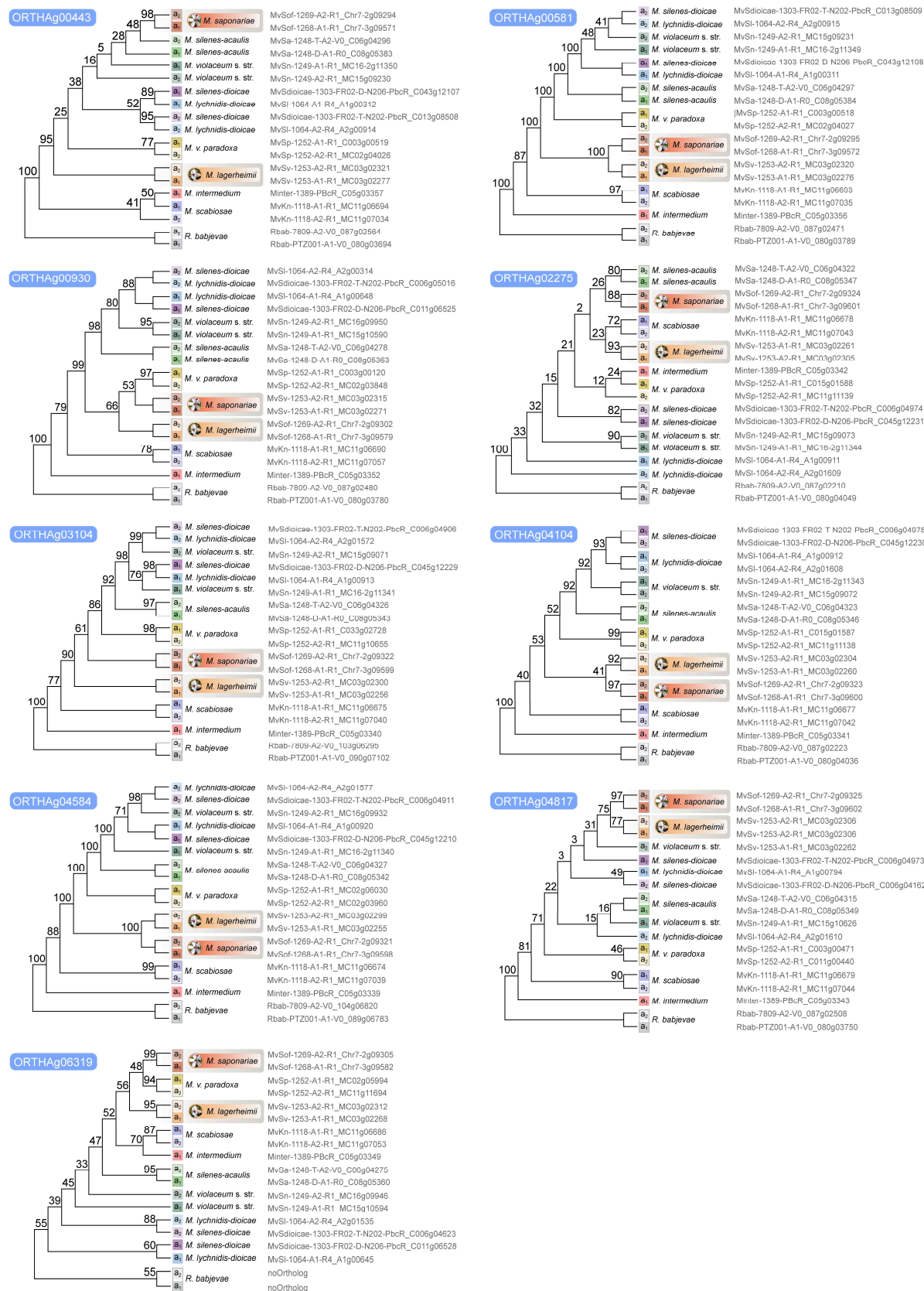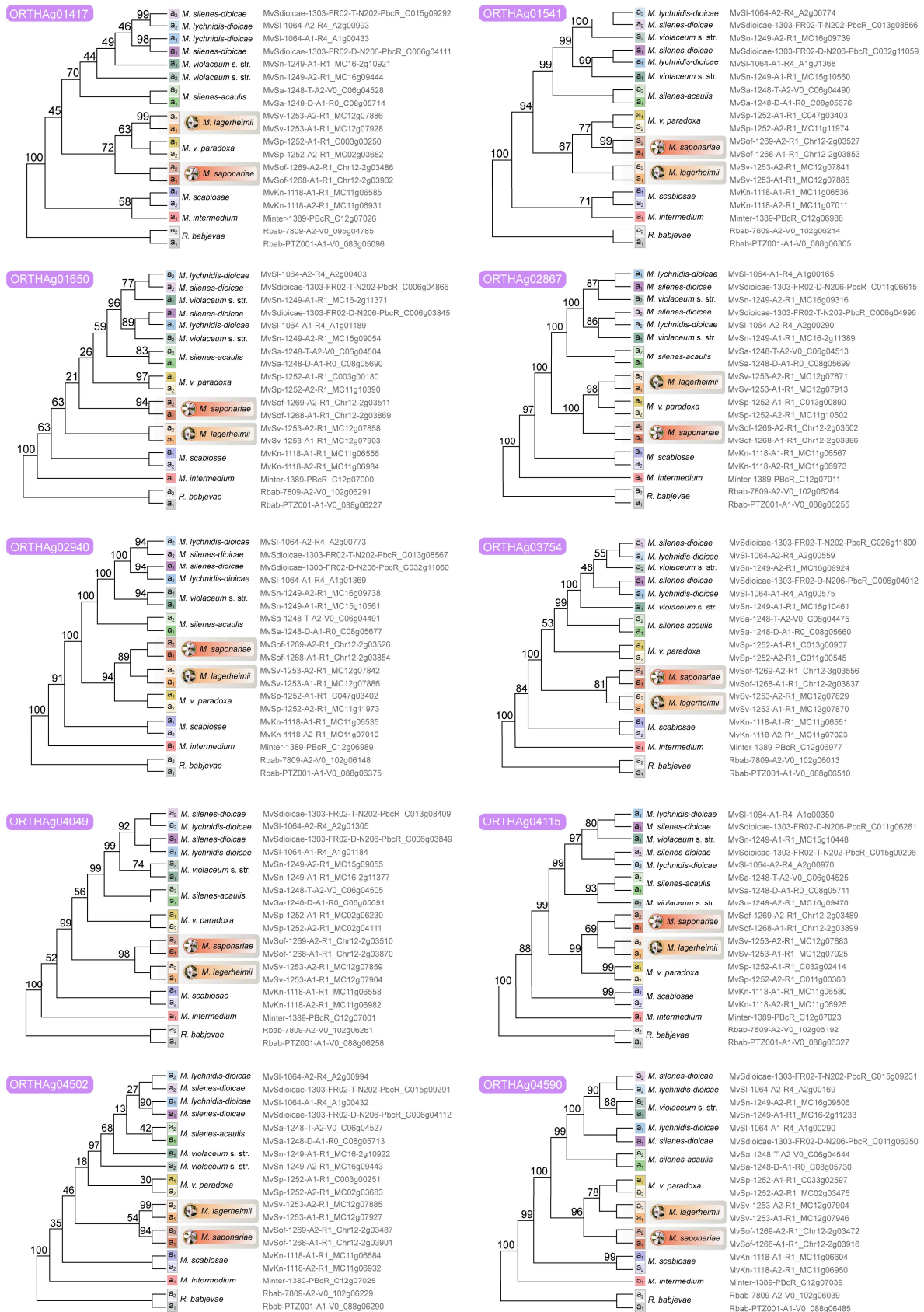
Supplemental Fig. S6. Individual genealogies for the 19 of the genes used for dating the linkage between mating-type loci and centromeres. The individual genealogies of these genes located between the centromeres and the HD-proximal or PR-proximal strata containing the mating-type loci illustrate the lack of trans-specific polymorphism in this genomic region between *Microbotryum lagerheimii* and *M. saponariae*, supporting that complete recombination cessation between the mating-type loci and the centromeres occurred independently in the two species. **(A)** Gene genealogies of nine genes located between the HD-proximal stratum and the centromere. **(B)** Gene genealogies of ten genes located between PR-proximal and the centromere.

Supplemental Table S1. Statistics on the genome assemblies of the *Microbotryum saponariae* genomes analysed in this study

Sample	Accession Numbers	# Contigs	Length of the smallest contig (bp)	Length of the largest contig (bp)	N50 (bp)	L50 (# contigs)	N90 (bp)	L90 (# contigs)	Mean Length (bp)	Median length (bp)	Assembly size (bp)
					Assembly statistics						
Microbotryum saponariae from *Saponaria officinalis* (1268) a_1	GCA_900015975	161	3,858	2,799,476	945,666	10	195,491	32	183,285.323	20,702	29,508,937
Microbotryum saponariae from *Saponaria officinalis* (1269) a_2	GCA_900015475	72	9,639	2,733,711	1,443,984	9	441,166	22	397,966.3472	35,381.5	28,653,577

Supplemental Table S2. Statistics of the mating-type chromosomes of the *Microbotryum saponariae* genomes analysed in this study, and in the different genomic partitions of the mating-type chromosomes: recombining regions (RR), non-recombining regions (NRR), pseudo-autosomal regions (PAR), homeodomain gene (HD) mating-type chromosome and pheromone receptor gene (PR) mating-type chromosome.

Genomics regions	Statistics
Number of mating-type chromosomes (haploid)	2
Number of contigs for the a1 HD mating-type chromosome	3
Number of contigs for the a2 HD mating-type chromosome	2
Number of contigs for the a1 PR mating-type chromosome	3
Number of contigs for the a2 PR mating-type chromosome	3
Size of the a1 HD mating-type chromosome (bp)	1,836,444
Size of the a2 HD mating-type chromosome (bp)	1,903,821
Size of the a1 PR mating-type chromosome (bp)	1,179,386
Size of the a2 PR mating-type chromosome (bp)	1,346,385
Size (percentage) of the RR on the a1 HD mating-type chromosome (bp)	85.19% (1,564,501 bp)
Size (percentage) of the NRR on the a1 HD mating-type chromosome (bp)	7.59% (139,334 bp)
Size (percentage) of the RR on the a2 HD mating-type chromosome (bp)	81.37% (1,549,232 bp)
Size (percentage) of the NRR on the a2 HD mating-type chromosome (bp)	7.34% (139,779 bp)
Size(percentage) of the RR on the a1 PR mating-type chromosome (bp)	46.55% (548,957 bp)
Size (percentage) of the NRR on the a1 PR mating-type chromosome (bp)	41.69% (491,719 bp)
Size (percentage) of the RR on the a2 PR mating-type chromosome (bp)	39.33% (529,587 bp)
Size (percentage) of the NRR on the a2 PR mating-type chromosome (bp)	45.47% (612,166 bp)
Size (percentage) of the centromere on the a1 HD mating-type chromosome (bp)	7.22% (132,606 bp)
Size (percentage) of the centromere on the a2 HD mating-type chromosome (bp)	11.28% (214,807 bp)
Size(percentage) of the centromere on the a1 PR mating-type chromosome (bp)	11.76% (138,706 bp)
Size (percentage) of the centromere on the a2 PR mating-type chromosome (bp)	15.2% (204,628 bp)

Supplemental File S1. Centromeric repeats *de novo* detected (provided as a separate file).

Supplemental methods

Segregation analyses

For mating-type segregation analysis we used *M. lagerheimii* collected on *Lychnis flos-jovis* in Valle Pesio, Italy (GPS 44.188400, 7.670650). *Microbotryum* fungi undergo meiosis immediately following spore germination, and the septate basidium allows for the micromanipulation and isolation of post-meiotic yeast-like cells as linear tetrads (Hood et al. 2015). Haploid cells were isolated from opposite poles of meiosis I across replicate meioses from the same diploid parent and characterized for mating type segregation by PCR amplification of allele-specific markers. For the PR locus, primers 660 and 588 were used that discriminate between a_1 and a_2 alleles based on allele-specific amplification (Devier et al. 2009; Hood et al. 2015). For the HD locus, the primer pair HD4-F5 (5' CCATCGAGCTCCTTTTACCC) and HD-R1 (5' TCTAGGCAGCTCTTGCTC) was designed to produce PCR products of allele-specific size due to insertion/deletion mutations. Under centromere linkage, variation at heterozygous loci should segregate at the first meiotic division and thus always differing between meiotic products from separated at meiosis I; a crossing-over recombination between the locus and the centromere would result in different alleles between two meiotic products separated at meiosis I only fifty percent of the time. The observed proportion of instances where haploid product separated at meiosis I carried alternate alleles for both the PR and HD locus was calculated for 78 meioses, the corresponding 95% confidence interval was calculated using Vassarstats®.

DNA extraction and sequencing

We isolated haploid cells of opposite mating types from single tetrads using micromanipulation as described previously (Hood et al. 2015) for *M. saponariae* parasitizing *S. officinalis* (cell 1268, PRAT 47, a1 b1, and cell 1269, PRAT 48, a2 b2) collected near Chiusa di Pesio, (GPS coordinates 44.31713297, 7.622967437 on July 8th, 2012). DNA was extracted with the QIAGEN Genomic-tip 100/G (ref. 10243; Courtaboeuf, France) and Genomic DNA Buffer Set (ref. 19060) following manufacturer instructions and using a Carver hydraulic press (reference 3968, Wabash, IN, USA) for breaking cell walls. Haploid genomes were sequenced using the P6/C4 Pacific Biosciences SMRT technology (UCSD IGM Genomics Facility La Jolla, CA, USA).

Assembly and annotation

Assemblies of the genomes were generated with the wgs-8.2 version of the PBcR assembler (Koren et al. 2012) with the following parameters: genomeSize=30000000, assembleCoverage=50. Assemblies were polished with quiver software (https://github.com/PacificBiosciences/GenomicConsensus). A summary of raw data and assembly statistics is reported in Table S1 and S2. Contigs were aligned with optical maps of the two mating-type chromosomes obtained previously (Hood et al. 2015), with MapSolver software (OpGen), allowing generating oriented a_1 and a_2 pseudomolecules for each mating type chromosome. Mating-type chromosomes were identified by finding the contigs carrying the PR and HD mating-type genes using BLAST (Altschul et al. 1990).

Species tree and species incorporated in the study

To study the evolution of suppressed recombination in a phylogenetic context, we reconstructed the relationships between the *Microbotryum* species for which high-quality genomes were available (Fig. 2), with linked or unlinked mating-type loci: in addition to the newly sequenced *M. saponariae* genomes of opposite mating types, we used the high-quality a_1 and a_2 genomes of seven available *Microbotryum* species (*M. lychnidis-dioicae*, *M. silenes-dioicae*, *M. violaceum sensu stricto*, *M. lagerheimii*, *M. silenes-acaulis*, *M. scabiosae*, *M. violaceum paradoxa*), a high-quality a_2 genome of *M. intermedium*, and a high-quality outgroup genome of the red yeast *Rhodosporidium babjevae*.

The previously published genome assemblies of the species used in this study are available at the GenBank under the following accession numbers: GCA_900015445 for *M. lychnidis-dioicae* 1064 a_2; GCA_900015465 for *M. lychnidis-dioicae* 1064 a_1; GCA_900013405 for *M. lagerheimii* 1253 a_2; GCA_900015505 for *M. lagerheimii* 1253 a_1; GCA_900015455 for *M. violaceum s. str.* 1249 a_2; GCA_900015425 for *M. violaceum s. str.* 1249 a_1; GCA_900096595 for *M. intermedium*; SAMN09670553 for *M. silenes-dioicae* 1303 a_2; and GCA_900120095 for *M. silenes-dioicae* 1303 a_1; GCA_900015485 *M. violaceum paradoxa* 1252 a_2; GCA_900015495 for *M. v. paradoxa* 1252 a_1; PRJEB12080 and ERZ250708 for *M. scabiosae* 1118 a_2; PRJEB12080 and ERZ250707 for *M. scabiosae* 1118 a_1; GCA_900014955 for *M. v. caroliniana* 1250 a_2; GCA_900014965 for *M. v. caroliniana* 1250 a_1; SAMN09670554 for *M. silenes-acaulis* 1248 a_1; SAMN09670555 for *M. silenes-acaulis* 1248 a_2.

We used the translated gene models for the nine *Microbotryum* species and the outgroup *Rhodotorula babjevae* to obtain orthologous groups with orthAgogue (Ekseth et al. 2014) based on blastp+ 2.2.30 followed by Markov clustering (Van Dongen 2000). We aligned the protein sequences of 780 fully conserved single-copy genes with MAFFT v7.388 (Katoh and Standley 2013) and obtained the codon-based CDS alignments with TranslatorX (Abascal et al. 2010). We used RAxML 8.2.7 (Stamatakis 2006) to obtain maximum likelihood gene trees for all 780 fully conserved single-copy genes and a species tree with the concatenated alignment of 447,405 codons under the

18

GTRGAMMA substitution model. We estimated the branch support values by bootstrapping the species tree based on the concatenated alignment and by estimating the relative internode and tree certainty scores based on the frequency of conflicting bipartitions for each branch in the species tree among the fully conserved single-copy genes (Salichos et al. 2014).

We removed all transposable elements (see the *de novo* TE identification method) from the gene dataset used in all analysis, assuming a gene to be a TE if the gene sequence shares more than 50% of TE sequence.

We identified alleles as orthologous groups with a single sequence in each haploid genome for a given species. We used TranslatorX (Abascal et al. 2010) with the MUSCLE aligner v3.8.31 (Edgar 2004) to align nucleotide sequences of predicted genes. We estimated synonymous divergence (d_S) and its standard error with the yn00 program of the PAML package (Yang 2007) and we plotted d_S along the chromosomes using the ggplot2 R package (Wickham 2016). The gene assignment to the PR- or HD-proximal stratum was made using previously gene assignments (Branco et al. 2017, 2018), and by considering genes previously unassigned to a stratum and that are now identified as being in-between genes in the PR-proximal stratum. The latter genes were not assigned to the HD- or PR-proximal strata because the gene assignment was made based on d_S plot using the *M. lagerheimii* gene order, which was slightly different from the *M. intermedium* gene order used in this study.

Figures

We prepared the figures 3, S1, S2 and S3 using Circos (Krzywinski 2009). We linked alleles in Circos plots comparing contigs belonging to alternative mating-types of a given species, and the ortholog genes in Circos plots comparing contigs from distinct species.

Date estimates for recombination cessation

We used the allele codon-based alignments within each species based on the rationale that the divergence between alleles associated to the a_1 versus a_2 mating types is dependent on the time since recombination cessation. We used 9 orthologous groups (8,525 aligned codons) for dating the recombination cessation between the HD-proximal stratum and the centromere, and 10 orthologous (10,200 aligned codons) groups for dating recombination cessation between the PR-proximal stratum and the centromere (these were the genes for which a_1 versus a_2 alleles were available in all species studied and located between the ancient mating-type strata and the centromere and are indicated with red arrows on Figure 4). Divergence times were estimated using BEAST v2.4.0 (Drummond and Rambaut 2007), with the xml inputs being generated using BEAUTi (Drummond et al. 2012), and setting the following parameters (others left as default values): unlinked substitution (HKY+G with empirical frequencies for each codon position) and clock models, Yule process to model speciation, and 5,000,000 MCMC generations sampled every 1,000. For all runs, we used a single calibration prior at 0.42 million years, corresponding to the divergence between *M. lychnidis-dioicae* and *M. silenes-dioicae* (Gladieux et al. 2011), with a normal distribution and a sigma of 0.04. Time trees were annotated with BEAST's TreeAnnotator tool setting burnin to discard 10% of the trees. The divergence between *M. lychnidis-dioicae* and *M. silenes-dioicae* have been estimated earlier to 0.42 million years through the same approach, using another calibration point (Gladieux et al. 2011).

Gene genealogies

Gene genealogies were inferred for codon-based alignments of genes using RAxML (Stamatakis 2006) version 8.2.7, assuming the GTRGAMMA model and rapid bootstrap (options: -f a and -# 100). We analysed all single-copy genes for which we had both alleles in all species and were located between the HD-proximal or PR-proximal strata and the corresponding centromeres.

Transposable element identification, annotation and detection

Transposable elements were identified and annotated *de novo* in the *Microbotryum* high-quality genome assemblies, using both LTR-harvest (defaults parameters; Ellinghaus et al. 2008); and RepeatModeler (defaults parameters; Smit and Hubley 2015). Transposable element sequences were clustered per family to get a consensus sequence per annotation, using *usearch* (centroid method, id=0.7; Edgar 2010). These consensus sequences were compiled to form a *Microbotryum* transposable elements database (Hartmann et al. 2018), and were annotated with blast using Repbase (Bao et al. 2015) database (20.05) that we used as library in RepeatMasker (Smit et al. 2015) to retrieve their genomic location in all genomes.

De novo detection of centromeric repeats

Centromeric regions are poor in gene and rich in transposable and repetitive elements. As several regions fulfilled these criteria in each contig, we identified *de novo* centromeric-specific repeats (Melters et al. 2013) using Tandem-Repeat Finder (TRF v. 4.07b; Benson 1999) on assembled Illumina reads of the very same strains as those sequenced using the Pacific Bioscience technology. The mate-pair reads used to detect centromeric repeats in *M. lagerheimii* are available in the sequence read archive, with the accession number SRR7047936; the mate-pair reads used to detect centromeric repeats in *M. saponariae* were previously published (Fortuna et al. 2016). For both *M. lagerheimii* and *M. saponariae* genomes, we performed the assemblies as follows: we randomly chose 500,000 Illumina reads that we assembled with PRICE (v1.2; Ruby et al. 2013) using a random set of 1,000,000 reads as seed file, and using the following command line arguments: -mpp inputFile_R1 inputFile_R2 650 90 -picf 20000 seedFile 500 2 25 -nc 10 -mpi 85 -MPI 95 – tpi 85 -TPI 95 -logf logfile -o outputFile. PRICE works by round of assembly: in the first round, it maps randomly picked reads onto contigs (provided by the seedFile), assembles the reads that did not mapped, and then extends the contig with the unmapped assembled sequences. For the second and following rounds, PRICE considers the extended contigs as the reference to restart the process of picking, mapping reads, assembling the unmapped reads and extending the reference contigs. We analysed the presence of tandem repeats in each of the 10 assembly cycle output, using the following parameters in a TRF wrapper perl script (Melters et al. 2013): match=1, mismatch=1, indel=2, probability of match=80, probability of indel=5, min score=200, max period=2000. We performed these steps 15 times, picking randomly 500,000 input reads and 1,000,000 reads for the seed file. The repeats detected in the Illumina genomes were blasted against the corresponding high-quality genomes of *M. lagerheimii* and *M. saponariae*. We defined the centromeric regions of the mating-type chromosomes by identifying at the largest TE-rich, gene-poor regions containing the greatest density of tandem repeats. The delimitations of the centromeric regions using this method were congruent with those using BLAST of the *M. lychnidis-dioicae* centromeric repeats identified previously *in M. lychnidis-*

dioicae (Badouin et al. 2015). To identify centromeric repeats in *M. intermedium*, we therefore blasted the centromeric repeats identified in *M. lychnidis-dioicae* (Badouin et al. 2015), as no Illumina reads were available for this species. FASTA files containing the *de novo* identified centromeric repeats are provided in Supplemental File S1.

References

Abascal F, Zardoya R, Telford MJ. 2010. TranslatorX: Multiple alignment of nucleotide sequences guided by amino acid translations. *Nucleic Acids Res* **38**(Web Server Issue): W7-13.

Altschul SF, Gish W, Pennsylvania T, Park U. 1990. Basic Local Alignment Search Tool. : 403–410.

Badouin H, Hood ME, Gouzy J, Aguileta G, Siguenza S, Perlin MH, Cuomo CA, Fairhead C, Branca A, Giraud T. 2015. Chaos of rearrangements in the mating-type chromosomes of the anther-smut fungus *Microbotryum lychnidis-dioicae*. *Genetics* **200**(4): 1275–1284.

Bao W, Kojima KK, Kohany O. 2015. Repbase Update, a database of repetitive elements in eukaryotic genomes. *Mob DNA* **6**(1): 4–9.

Benson G. 1999. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res* **27**(2): 573–580.

Branco S, Badouin H, Rodríguez de la Vega RC, Gouzy J, Carpentier F, Aguileta G, Siguenza S, Brandenburg J-T, Coelho MA, Hood ME, et al. 2017. Evolutionary strata on young mating-type chromosomes despite the lack of sexual antagonism. *Proc Natl Acad Sci* **114**(27): 7067–7072.

Branco S, Carpentier F, Rodríguez de la Vega RC, Badouin H, Snirc A, Le Prieur S, Coelho MA, de Vienne DM, Hartmann FE, Begerow D, et al. 2018. Multiple convergent supergene evolution events in mating-type chromosomes. *Nat Commun* **9**(1): 2000.

Devier B, Aguileta G, Hood ME, Giraud T. 2009. Ancient trans-specific polymorphism at pheromone receptor genes in basidiomycetes. *Genetics* **181**(1): 209–223.

Drummond AJ, Rambaut A. 2007. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol* **7**(1): 1–8.

Drummond AJ, Suchard MA, Xie D, Rambaut A. 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol Biol Evol* **29**(8): 1969–1973.

Edgar RC. 2004. MUSCLE: Multiple sequence alignment with high accuracy and high throughput.

Nucleic Acids Res **32**(5): 1792–1797.

Edgar RC. 2010. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**(19): 2460–2461.

Ekseth OK, Kuiper M, Mironov V. 2014. OrthAgogue: an agile tool for the rapid prediction of orthology relations. *Bioinformatics* **30**(5): 734–736.

Ellinghaus D, Kurtz S, Willhoeft U. 2008. LTRharvest, an efficient and flexible software for de novo detection of LTR retrotransposons. *BMC Bioinformatics* **9**(18): 1–14.

Fortuna TM, Snirc A, Badouin H, Gouzy J, Siguenza S, Esquerre D, Le Prieur S, Shykoff JA, Giraud T. 2016. Polymorphic microsatellite markers for the tetrapolar anther-smut fungus *Microbotryum saponariae* based on genome sequencing. *PLoS One* **11**(11): e0165656.

Gladieux P, Vercken E, Fontaine MC, Hood ME, Jonot O, Couloux A, Giraud T. 2011. Maintenance of fungal pathogen species that are specialized to different hosts: allopatric divergence and introgression through secondary contact. *Mol Biol Evol* **28**(1): 459–471.

Hartmann FE, Rodríguez de la Vega RC, Brandenburg J-T, Carpentier F, Giraud T, Van De Peer Y. 2018. Gene presence–absence polymorphism in castrating anther-smut fungi: recent gene gains and phylogeographic structure. *Genome Biol Evol* **10**(5): 1298–1314.

Hood ME, Scott M, Hwang M. 2015. Breaking linkage between mating compatibility factors: tetrapolarity in *Microbotryum*. *Evolution (N Y)* **69**(10): 2561–2572.

Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* **30**(4): 772–780.

Koren S, Schatz MC, Walenz BP, Martin J, Howard JT, Ganapathy G, Wang Z, Rasko DA, McCombie RW, Jarvis ED, et al. 2012. Hybrid error correction and de novo assembly of single-molecule sequencing reads. *Nat Biotechnol* **30**(7): 693–700.

Krzywinski M et al. 2009. Circos: an information aesthetic for comparative genomics. *Genome Res* **19**(604): 1639–1645.

Melters DP, Bradnam KR, Young HA, Telis N, May MR, Ruby JG, Sebra R, Peluso P, Eid J, Rank

D, et al. 2013. Comparative analysis of tandem repeats from hundreds of species reveals unique insights into centromere evolution. *Genome Biol* **14**(1): R10.

Ruby JG, Bellare P, DeRisi JL. 2013. PRICE: Software for the targeted assembly of components of (meta) genomic sequence data. *G3 Genes | Genomes | Genet* **3**(5): 865–880.

Salichos L, Stamatakis A, Rokas A. 2014. Novel information theory-based measures for quantifying incongruence among phylogenetic trees. *Mol Biol Evol* **31**(5): 1261–1271.

Smit AF, Hubley R, Green P. 2015. RepeatMasker Open-4.0. *Repeat Masker Website*.

Smit AF, Hubley RR. 2015. RepeatModeler Open-1.0. *Repeat Masker Website*.

Stamatakis A. 2006. RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22**(21): 2688–2690.

Van Dongen SM. 2000. Graph clustering by flow simulation. PhD thesis. (https://dspace.library.uu.nl/handle/1874/848).

Wickham H. 2016. *ggplot2: elegant graphics for data analysis*. Springer.

Yang Z. 2007. PAML 4: Phylogenetic analysis by maximum likelihood. *Mol Biol Evol* **24**(8): 1586–1591.

S.5 Differential gene expression is associated with degeneration and not sexual antagonism in mating-type chromosomes of anther-smut fungi

Supporting Information Legends

Table S1. Number of single-copy genes with alleles in both a_1 and a_2 haploid genomes, with 70% protein sequence identity detection using reciprocal best BLASTp hits, before and after filtering out genes with transposable element (TE)-related functions.

Filtering removed 192 paralogous genes within each haploid genome and genes with TE-related functions, including 1750 and 1819 from a_1 and a_2 haploid genome, respectively.

Table S2. Identification of differentially expressed (DE) genes with either a_1-biased or a_2-biased expression (i.e., higher expression in a_1 or a_2, respectively) under various $Log2(a_1/a_2)$ criteria.

Table S3. Numbers and percentages of genes with differential expression (DE) within genomic compartments, including autosomes pseudo-autosomal regions (PARs), youngest and oldest evolutionary strata.

Chi-squared test was used to assess whether DE genes were non-randomly distributed between autosomes versus the genomic compartments on mating type chromosomes (MAT). *P* values <0.05 are in bold. NA: not applicable. The youngest strata only had one DE gene, so statistic comparisons could not be performed for these strata.

Table S4. Wilcoxon rank sum test statistics for comparisons of mean non-synonymous mutation rate (*dN*, A) and synonymous mutation rate (*dS*, B) of differentially expressed genes (DE) versus non-differentially expressed genes (non-DE) within genomic compartments, including autosomes pseudo-autosomal regions (PARs), youngest and oldest evolutionary strata.

P values <0.05 are in bold. NA: not applicable. NA: not applicable, as the youngest strata only had one DE gene, statistical comparisons could not be performed for this compartment.

Table S5. Wilcoxon rank sum test statistics for comparisons of divergence of alleles in *Microbotryum lychnidis-dioicae* from orthologs in *M. lagerheimii*.

This test assessed whether the allele with lower expression in *M. lychnidis-dioicae* was more divergent from orthologs in *M. lagerheimii* than the alleles with higher expression in *M. lychnidis-dioicae*, considering non-synonymous mutation rate (*dN*), synonymous mutation rate (*dS*), and the ratio (*dN/dS*) within genomic compartments, including autosomes pseudo-autosomal regions (PARs), youngest and oldest evolutionary strata. We calculated these

36 substitution rates for a_1 and a_2 alleles between these two species separately. NA: not applicable,

37 as the youngest strata only had one DE gene, statistical comparisons could not be performed for

38 this compartment.

39

40 **Table S6. Wilcoxon rank sum test statistics for comparisons of unoriented transposable**

41 **elements (TEs) insertion differences between alleles (within 20kb up and downstream) of**

42 **differentially expressed (DE) genes versus non-differentially expressed (non-DE) genes**

43 **within genomic compartments, including autosomes pseudo-autosomal regions (PARs),**

44 **youngest and oldest evolutionary strata.**

45 P-values <0.05 are in bold. NA: not applicable, as the youngest strata only had one DE gene,

46 statistical comparisons could not be performed for this compartment.

47

48 **Table S7. Two proportion Z test for comparisons of unoriented protein length difference**

49 **between alleles of differentially expressed (DE) genes versus non-differentially expressed**

50 **(non-DE) genes within genomic compartments, including autosomes, pseudo-autosomal**

51 **regions (PARs), youngest and oldest evolutionary strata.**

52 *P* values <0.05 are in bold. NA: not applicable, as proportions * sample size was less than five

53 for the PARs and youngest strata, statistical comparisons could not be performed for these

54 compartments.

55

56 **Table S8. Wilcoxon rank sum test statistics for comparisons of unoriented differences in**

57 **intron content between alleles of differentially expressed (DE) versus non-differentially**

58 **expressed (non-DE) genes within genomic compartments, including autosomes pseudo-**

59 **autosomal regions (PARs), youngest and oldest evolutionary strata.**

60 *P* values <0.05 are in bold. NA: not applicable, as the youngest strata only had one DE gene,

61 statistical comparisons could not be performed for this compartment.

62

63 **Table S9. Wilcoxon rank-sum test statistics for comparisons of unoriented differences in**

64 **overall GC content (GC0), and third codon position (GC3) between alleles of differentially**

65 **expressed (DE) versus non-differentially expressed (non-DE) genes within genomic**

66 **compartments, including autosomes pseudo-autosomal regions (PARs), youngest and oldest**

67 **evolutionary strata.**

68 (A) Overall GC content. (B) Third codon position GC content. *P* values <0.05 are in bold. NA:

69 not applicable, as the youngest strata only had one DE gene, statistical comparisons could not be

70 performed for this compartment.

71

72 **Fig. S1. Boxplot of gene expression in terms of Log$_2$TPM (transcripts per millions) for genes**

73 **with a$_1$-biased expression (in red), a$_2$-biased expression (in blue) and non-biased expression**

74 **(in grey) on mating-type chromosomes (MAT), as well as autosomes (Chr01-Chr18 and**

75 **other contigs).**

76

77 **Fig. S2. Boxplot of synonymous mutation rate (*dS*) for differentially expressed (DE) and**

78 **non-differentially expressed genes (Non-DE) of *Microbotryum lychnidis-dioicae*.**

79 Wilcoxon rank sum tests for comparisons of mean non-synonymous mutation rate (*dN*) of

80 differentially expressed genes (DE) versus non-differentially expressed genes (non-DE) within

81 genomic compartments: '***': *P* < 0.001, other comparisons were not significant. Genomic

82 compartments include autosomes, pseudo-autosomal regions (PARs), youngest and oldest

83 evolutionary strata.

84

85 **Fig. S3. Boxplot of differentially (a$_1$-biased in red, a$_2$-biased in blue) and non-differentially**

86 **expressed genes (not-biased in grey) and the sequence divergence between alleles of**

87 ***Microbotryum lychnidis-dioicae*.**

88 Wilcoxon rank sum tests for comparisons of genes with higher allele expression in the a$_1$ and a$_2$

89 haploid mating type genomes separately to non-differentially expressed genes for the mean non-

90 synonymous mutation rate (*dN*) (**A**), synonymous mutation rate (*dS*) (**B**); NS: not significant,

91 '***': *P* < 0.001, '**': *P* < 0.01, '*': *P* < 0.5, '.': *P* < 0.1, NS: not significant. As *dN* and *dS* of

92 almost all genes in autosome and PAR are zero, and there is only one DE gene on the youngest

93 strata, so no statistic test can be performed in these regions. Sample size for each genomic

94 compartment is listed either above or inside boxplot accordingly. Genomic compartments

95 include autosomes, pseudo-autosomal regions (PARs), youngest and oldest evolutionary strata.

96

97 **Fig. S4. Boxplot of differentially expressed (DE) and non-differentially expressed genes**

98 **(non-DE) and gene evolutionary rate *dN/dS* of *Microbotryum lychnidis-dioicae*.**

99 Wilcoxon rank sum tests for comparisons of evolutionary rate *dN/dS* of differentially expressed

100 genes (DE) versus non-differentially expressed genes (non-DE) within genomic compartments;

101 NS: not significant. As *dN/dS* of almost all genes in autosome and PAR are zero, and there is

102　only one DE gene on the youngest strata, so no statistic test can be performed in these regions.

103　Genomic compartments include autosomes, pseudo-autosomal regions (PARs), youngest and

104　oldest evolutionary strata.

105

106　**Fig. S5. Boxplot of sequence divergence, non-synonymous mutation rate *dN* (A),**

107　**synonymous mutation rate *dS* (B), between *Microbotryum lychinidis-dioicae* and *M.***

108　***lagerheimii*.**

109　Alleles of differentially expressed genes with hypothesized higher (red - lower expressed allele)

110　and lower (blue - higher expressed allele) substitution rates and hypothesized equal (grey; non-

111　differentially expressed genes) mutation rates were pooled from a_1 and a_2 genomes and assessed

112　for divergences from orthologs in *M. lagerheimii*. Genomic compartments include autosomes,

113　pseudo-autosomal regions (PARs), youngest and oldest evolutionary strata.

114

115　**Fig. S6. Dot plot of oriented differences of transposable element (TE) insertions and**

116　**differential gene expression between alleles of *Microbotryum lychnidis-dioicae*.**

117　TE insertions are shown for sliding-window intervals from upstream to downstream of genes,

118　were differences between alleles were calculated as the TE number for the allele with lower

119　expression minus the TE number for the higher expressed allele; a positive value thus represented

120　an excess of TEs in the lower expressed allele. Sliding window intervals are shown as **A**:

121　upstream 20kb to 10kb, **B**: upstream 15kb to 5kb, **C**: upstream 10kb to gene, **D**: upstream 5kb to

122　downstream 5kb, **E**: gene to downstream 10kb, **F**: downstream 5kb to 15kb, and **G**: downstream

123　10kb to 20kb.

124

125　**Fig. S7. Comparisons of differentially expressed (DE) and non-differentially expressed**

126　**(non-DE) genes between mating types of *Microbotryum lychnidis-dioicae* for differences**

127　**between alleles in third codon position GC content (GC3) within genomic compartments.**

128　Wilcoxon rank sum test statistics for comparisons of unoriented differences in GC content of

129　third codon position (GC3) between alleles of differentially expressed (DE) versus non-

130　differentially expressed (non-DE) genes within genomic compartments; ***: $P < 0.001$, NS: non

131　significant. Genomic compartments include autosomes, pseudo-autosomal regions (PARs),

132　youngest and oldest evolutionary strata. The notation "*a*" indicates that the youngest evolutionary

133　strata contained only one DE gene, precluding comparisons to non-DE genes within this

134　compartment.

135

Fig. S8. Comparison of proportion (A) and number (B) of differentially expressed (DE) genes detected between a_1 and a_2 haploid mating type genomes of *Microbotryum lychnidis-dioicae*.

Differentially expressed (DE) genes on mating-type chromosome (MAT) chromosomes and autosomes (auto), at various percentage protein sequence identities used as threshold for identification of alleles for genes in a_1 and a_2 haploid genomes. **: $P < 0.01$ using Chi-square test. All other comparisons of DE genes on mating-type chromosome and autosomes are not significant.

Fig. S9. Pairwise correlation of raw counts from RNAseq data between replicates for haploid a_1 cell culture (A) and haploid a_2 cell culture (B).

Fig. S10. Multidimensional scaling (MDS) plot of RNAseq libraries of *Microbotryum lychnidis-dioicae*.

Water denotes water agar (i.e. low nutrients) culture condition.

Fig. S11. The ratio index of coding sequence and protein sequence of *Microbotryum lychnidis-dioicae*.

The ratio of predicted coding sequence divided by the predicted protein sequence multiplied by three for a_1 and a_2 alleles, among four genomic compartments.

Table S1. Number of single-copy genes with alleles in both a_1 and a_2 haploid genomes, with 70% protein sequence identity detection using reciprocal best BLASTp hits, before and after filtering out genes with transposable element (TE)-related functions. Filtering removed 192 paralogous genes within each haploid genome and genes with TE-related functions, including 1750 and 1819 from a_1 and a_2 haploid genome, respectively.

Number of single-copy genes	Mating-type chromosome	Autosomes
Before filtering out genes of TE-related functions	434	10,018
After filtering out genes of TE-related functions	371	9,025

Table S2. Identification of differentially expressed (DE) genes with either a_1-biased or a_2-biased expression (i.e., higher expression in a_1 or a_2, respectively) under various $\text{Log2}(a_1/a_2)$ criteria.

DE criteria	a_1 bias	a_2 bias
$\text{Log2}(a_1/a_2) > 0$	392	203
$\text{Log2}(a_1/a_2) > 1$	286	139
$\text{Log2}(a_1/a_2) > 2$	112	48
$\text{Log2}(a_1/a_2) > 3$	65	29

Table S3. Numbers and percentages of genes with differential expression (DE) within genomic compartments, including autosomes pseudo-autosomal regions (PARs), youngest evolutionary strata (previously identified red and green strata; Branco et al. 2017) and oldest evolutionary strata (blue, purple, orange and black strata; Branco et al. 2017). Chi-squared test was used to assess whether DE genes were non-randomly distributed between autosomes versus the genomic compartments on mating type chromosomes (MAT). *P* values <0.05 are in bold. NA: not applicable. The young strata only had one DE gene, so statistic comparisons could not be performed for these strata.

	Higher expression in a_1				Higher expression in a_2			
	Autosomes	MAT			Autosomes	MAT		
		PARs	Youngest strata	Oldest Strata		PARs	Youngest strata	Oldest Strata
DE gene number	351	4	0	36	156	8	1	38
Total number	8207	114	29	198	8207	114	29	198
Percentage	4.28%	3.51%	0.00%	18.18%	1.90%	7.02%	3.45%	19.19%
X^2 test (X^2)	NA	0.02	1.24	68.91	NA	11.60	0.04	210.95
X^2 test (*P* value)	NA	0.88	0.41	**9.99E-05**	NA	**6.60E-04**	1	**9.99E-05**

Table S4. Wilcoxon rank sum test statistics for comparisons of mean non-synonymous mutation rate (*dN*, A) and synonymous mutation rate (*dS*, B) of differentially expressed genes (DE) versus non-differentially expressed genes (non-DE) within genomic compartments, including autosomes pseudo-autosomal regions (PARs), youngest evolutionary strata (previously identified red and green strata; Branco et al. 2017) and oldest evolutionary strata (blue, purple, orange and black strata; [1]. *P* values <0.05 are in bold. NA: not applicable. NA: not applicable, as the youngest strata only had one DE gene, statistical comparisons could not be performed for this compartment.

(A)

dN	Autosome	PARs	Youngest strata	Oldest Strata
Mean of DE genes	0	0	0.013	0.053
Mean of non-DE genes	0	0	0.006	0.033
Number of DE genes	117	4	1	43
Number of non-DE genes	5742	75	26	106
Wilcoxon W	334269	146	NA	1433
P-value	0.449	0.94	NA	**< 0.001**

(B)

dS	Autosome	PARs	Youngest strata	Oldest Strata
Mean of DE genes	0	0	0.013	0.027
Mean of non-DE genes	0	0	0.003	0.015
Number of DE genes	117	4	1	43
Number of non-DE genes	5742	75	26	106
Wilcoxon W	333918	116	NA	1422
P-value	0.404	0.455	NA	**< 0.001**

Table S5. Wilcoxon rank sum test statistics for comparisons of divergence of alleles in *Microbotryum lychnidis-dioicae* from orthologs in *M. lagerheimii*. This test assessed whether the allele with lower expression in *M. lychnidis-dioicae* was more divergent from orthologs in *M. lagerheimii* than the alleles with higher expression in *M. lychnidis-dioicae*, considering non-synonymous mutation rate (*dN*), synonymous mutation rate (*dS*), and the ratio (*dN/dS*) within genomic compartments, including autosomes pseudo-autosomal regions (PARs), youngest evolutionary strata (previously identified red and green strata; Branco et al. 2017) and oldest evolutionary strata (blue, purple, orange and black strata; [1]. We calculated these substitution rates for a_1 and a_2 alleles between these two species separately. NA: not applicable, as the youngest strata only had one DE gene, statistical comparisons could not be performed for this compartment.

| Substitution type | Autosome | | PARs | | Youngest strata | | Oldest Strata | |
| | High vs low mutation rates | | High vs low mutation rates | | High vs low mutation rates | | High vs low mutation rates | |
	W	*P*-value	W	*P*-value	W	*P*-value	W	*P*-value
dN	8066.0	0.999	18.5	1.000	NA	NA	1262.5	0.934
dS	8060.5	0.995	19.0	0.936	NA	NA	1167.0	0.909
dN/dS	8057.0	0.991	17.0	0.936	NA	NA	1247.5	0.989

Table S6. Wilcoxon rank sum test statistics for comparisons of unoriented transposable elements (TEs) insertion differences between alleles (within 20kb up and downstream) of differentially expressed (DE) genes versus non-differentially expressed (non-DE) genes within genomic compartments, including autosomes pseudo-autosomal regions (PARs), youngest evolutionary strata (previously identified red and green strata; Branco et al. 2017) and oldest evolutionary strata (blue, purple, orange and black strata; Branco et al. 2017). *P*-values <0.05 are in bold. NA: not applicable, as the youngest strata only had one DE gene, statistical comparisons could not be performed for this compartment.

Transposable elements (TEs)	Autosome	PARs	Youngest strata	Oldest Strata
Difference of DE genes	0.312	0.333	1.000	4.567
Differences of non-DE genes	0.107	0.078	2.143	3.903
Number of DE genes	507	12	1	74
Number of non-DE genes	7700	102	28	124
Wilcoxon W	31387793	546	NA	4062
P-value	**<0.001**	0.192	NA	0.173

Table S7. Two proportion Z test for comparisons of unoriented protein length difference between alleles of differentially expressed (DE) genes versus non-differentially expressed (non-DE) genes within genomic compartments, including autosomes pseudo-autosomal regions (PARs), youngest evolutionary strata (previously identified red and green strata; Branco et al. 2017) and oldest evolutionary strata (blue, purple, orange and black strata; Branco et al. 2017). *P* values <0.05 are in bold. NA: not applicable, as proportions * sample size was less than 5 for the PARs and youngest strata, statistical comparisons could not be performed for these compartments.

Protein Length	Autosomes	PARs	Youngest strata	Oldest Strata
Proportion unequal length DE genes	0.037	0.000	0.000	0.757
Proportion unequal length non-DE genes	0.012	0.020	0.357	0.605
Number of DE genes	507	12	1	74.000
Number of non-DE genes	7700	102	28	124
Z	4.640	NA	NA	2.186
P-value	**0.0002**	NA	NA	**0.029**

Table S8. Wilcoxon rank sum test statistics for comparisons of unoriented differences in intron content between alleles of differentially expressed (DE) versus non-differentially expressed (non-DE) genes within genomic compartments, including autosomes pseudo-autosomal regions (PARs), youngest evolutionary strata (previously identified red and green strata; Branco et al. 2017) and oldest evolutionary strata (blue, purple, orange and black strata; Branco et al. 2017). *P* values <0.05 are in bold. NA: not applicable, as the young strata only had one DE gene, statistical comparisons could not be performed for this compartment.

Intron	Autosomes	PARs	Youngest strata	Oldest Strata
Difference between DE alleles	0.007	0.001	0.005	0.169
Difference between non-DE alleles	0.002	0.006	0.02	0.101
Number of DE genes	507	12	1	74
Number of non-DE genes	7700	102	28	124
Wilcoxon W	1920124	605	NA	3205
P-value	**0.033**	0.888	NA	**0.001**

Table S9. Wilcoxon rank-sum test statistics for comparisons of unoriented differences in overall GC content (GC0), and third codon position (GC3) between alleles of differentially expressed (DE) versus non-differentially expressed (non-DE) genes within genomic compartments, including autosomes pseudo-autosomal regions (PARs), youngest evolutionary strata (previously identified red and green strata; Branco et al. 2017) and oldest evolutionary strata (blue, purple, orange and black strata; Branco et al. 2017). (A) Overall GC content. (B) Third codon position GC content. *P* values <0.05 are in bold. NA: not applicable, as the youngst strata only had one DE gene, statistical comparisons could not be performed for this compartment.

(A)

GC0	Autosomes	PAR	Youngest strata	Oldest Strata
Difference between DE alleles	0.000124	0.00065	0.001373	0.01027
Difference between non-DE alleles	0.000081	0.000054	0.00171	0.0067
Number of DE genes	507	12	1	74
Number of non-DE genes	7700	102	28	124
Wilcoxon W	1907831	578	NA	3010
P-value	**<0.001**	0.318	NA	**<0.001**

(B)

GC3	Autosomes	PAR	Youngest strata	Oldest Strata
Difference between DE alleles	0.000344	0	0.004464	0.02063
Difference between non-DE alleles	0.000081	0.000148	0.00423	0.01382
Number of DE genes	507	12	1	74
Number of non-DE genes	7700	102	28	124
Wilcoxon W	1903874	594	NA	3168
P-value	**<0.001**	0.549	NA	**0.001**

References:

1. Branco S, Badouin H, Rodríguez RC, Vega D, Gouzy J, Carpentier F. Evolutionary strata on young mating-type chromosomes despite the lack of sexual antagonism. Proc Natl Acad Sci U S A. 2017;114: 7367–7072. doi:10.1073/pnas.1701658114

Fig. S1

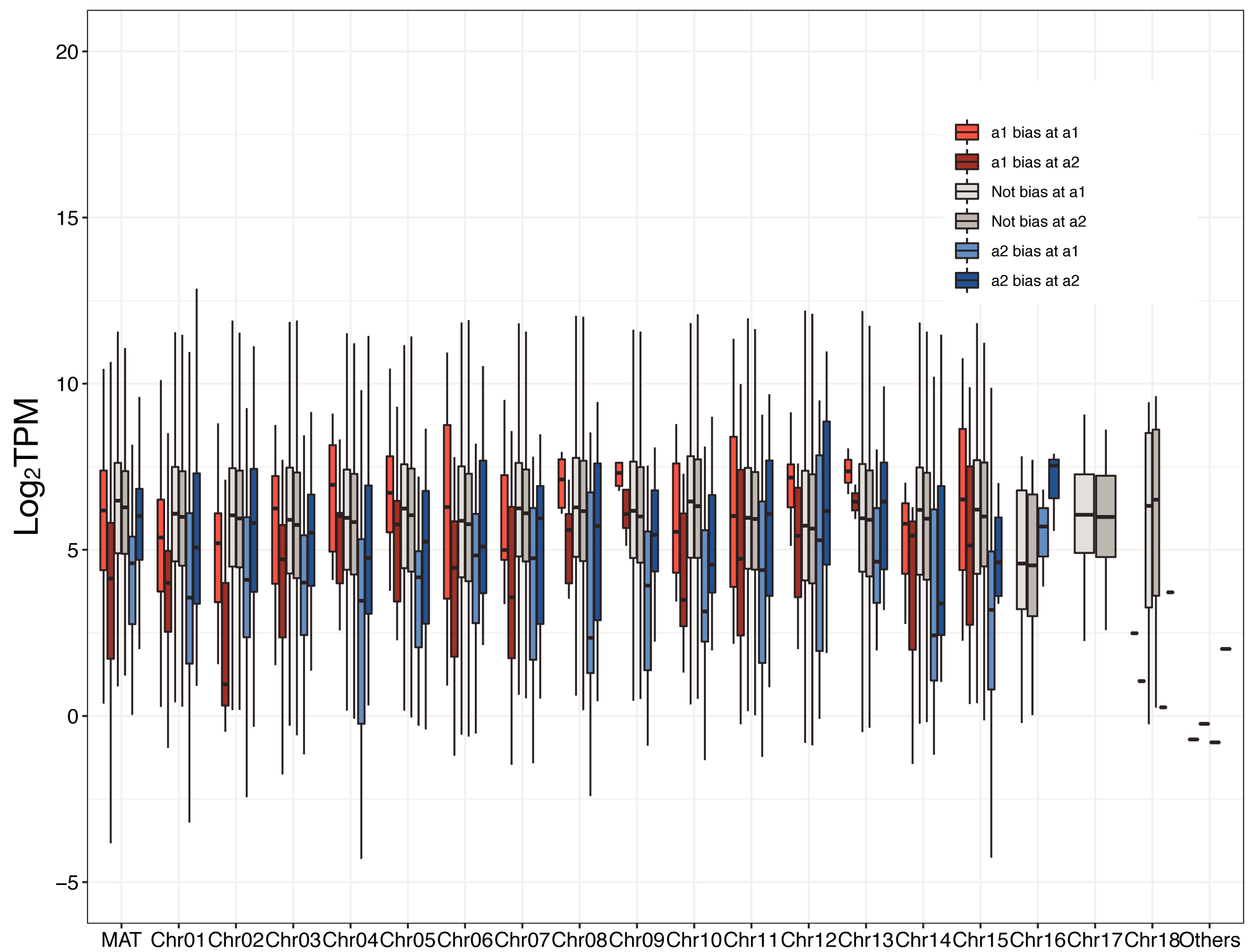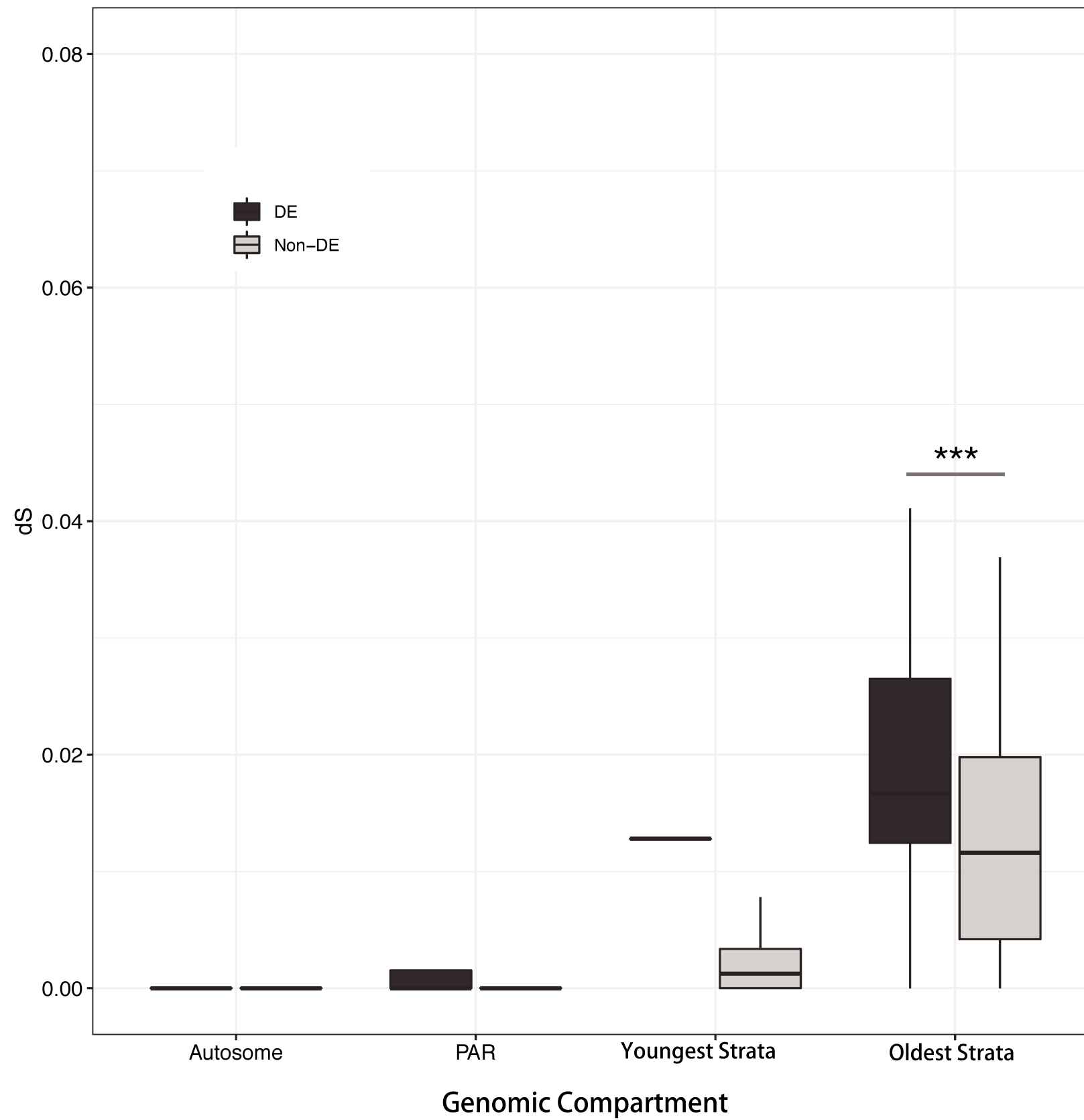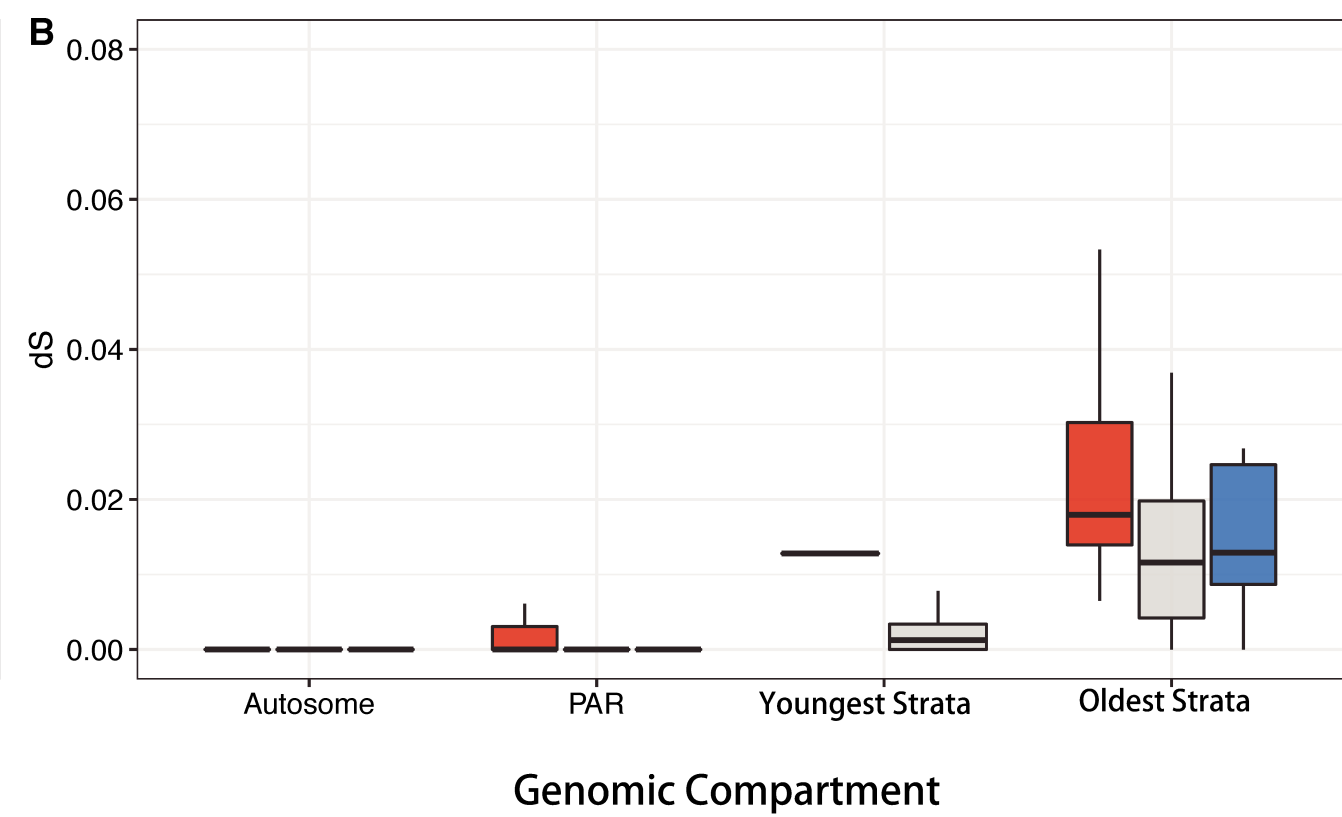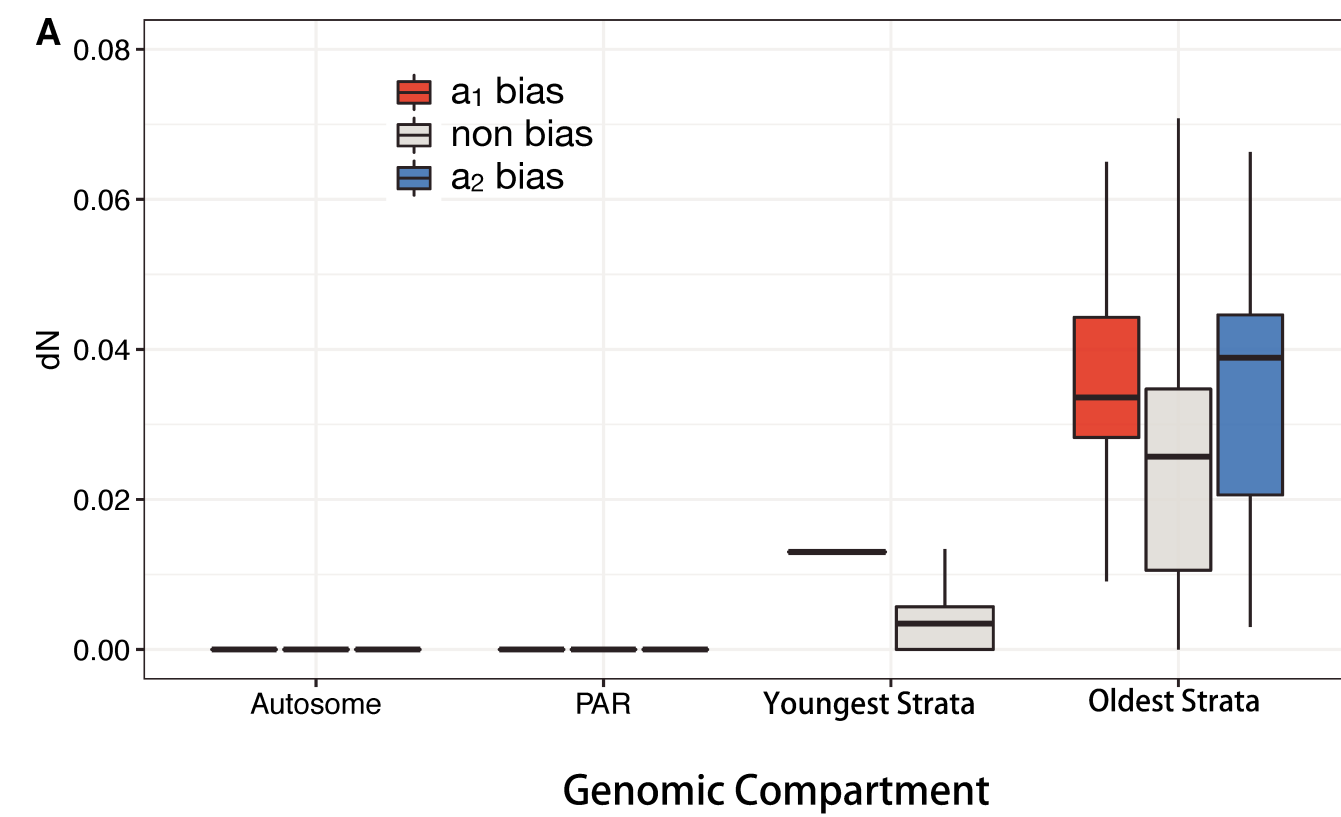**Fig. S2**

Fig. S3
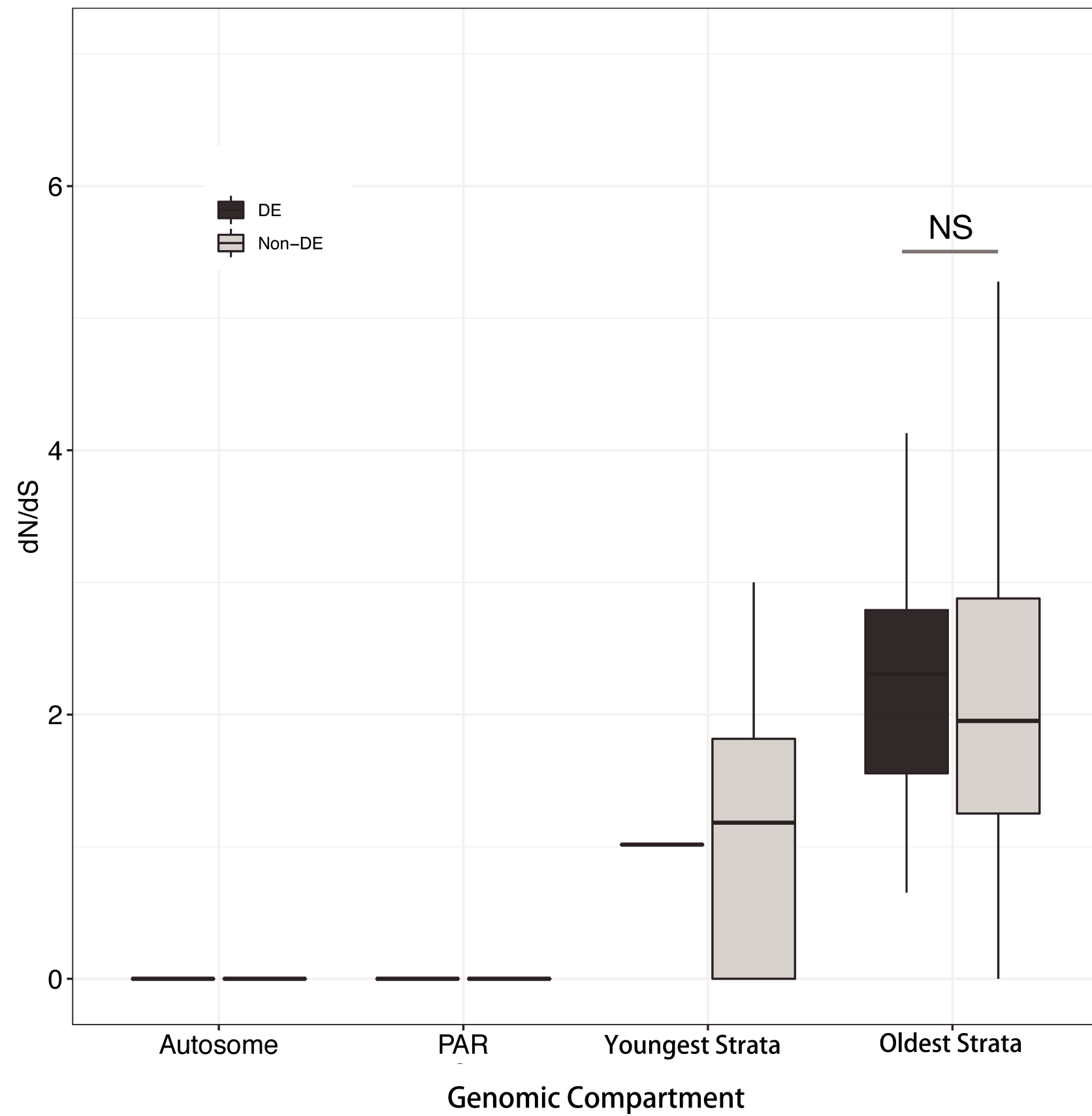
Fig. S4
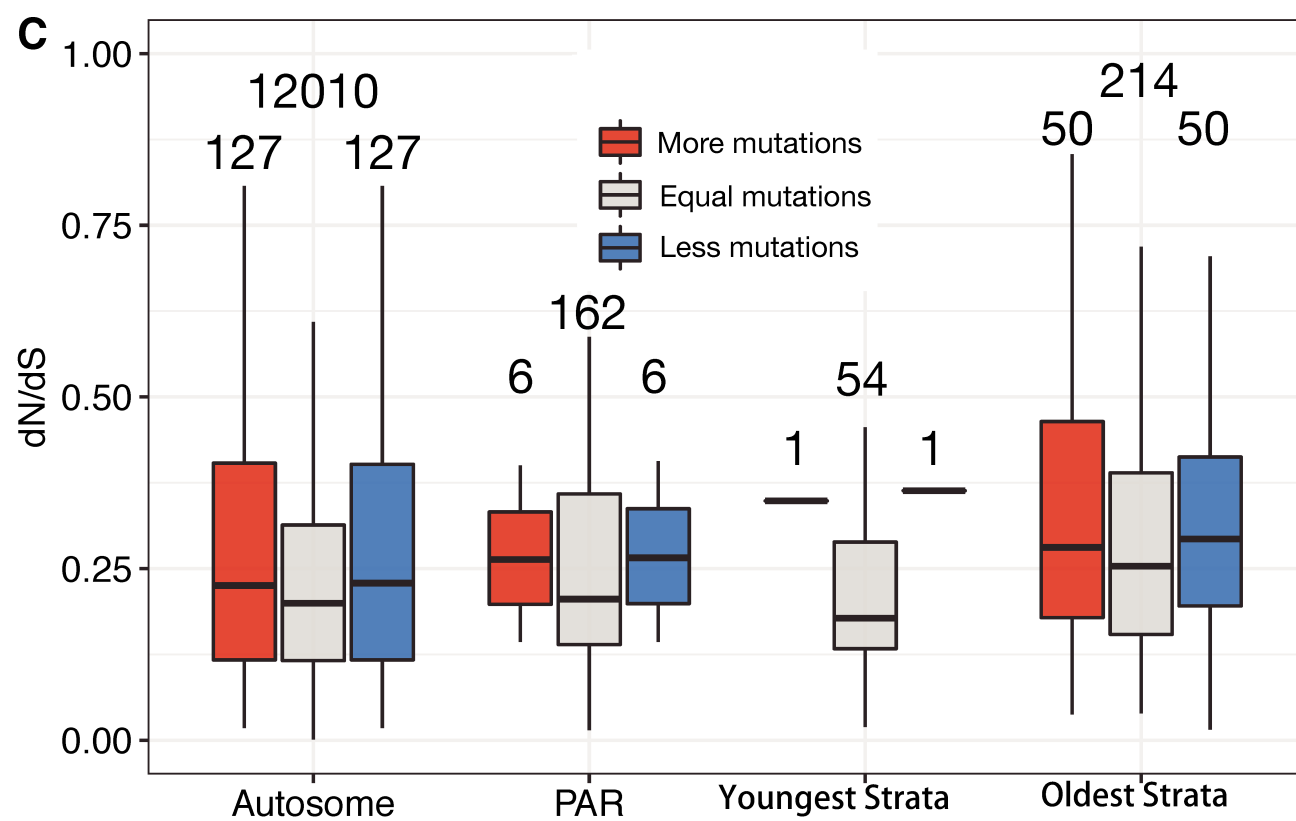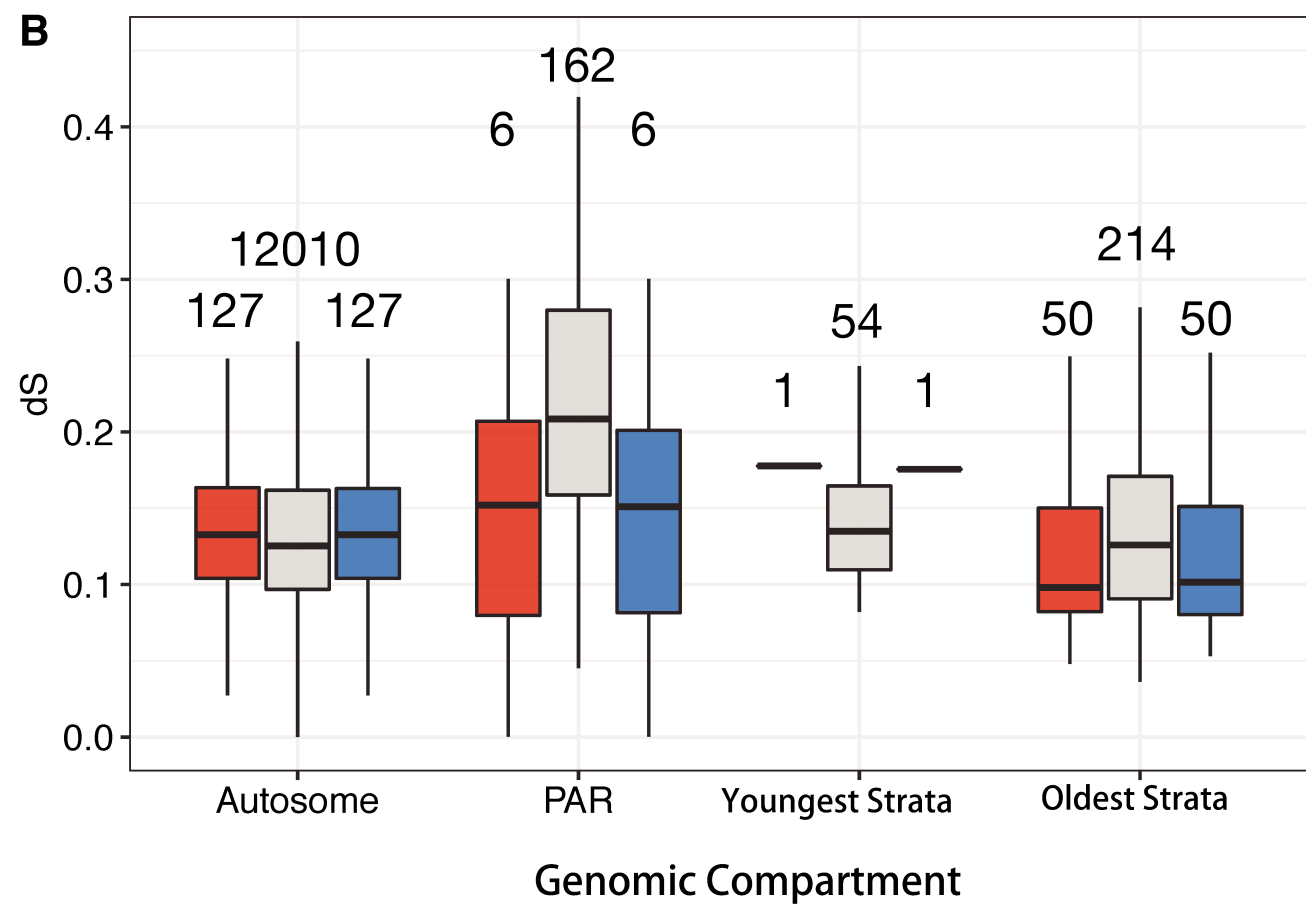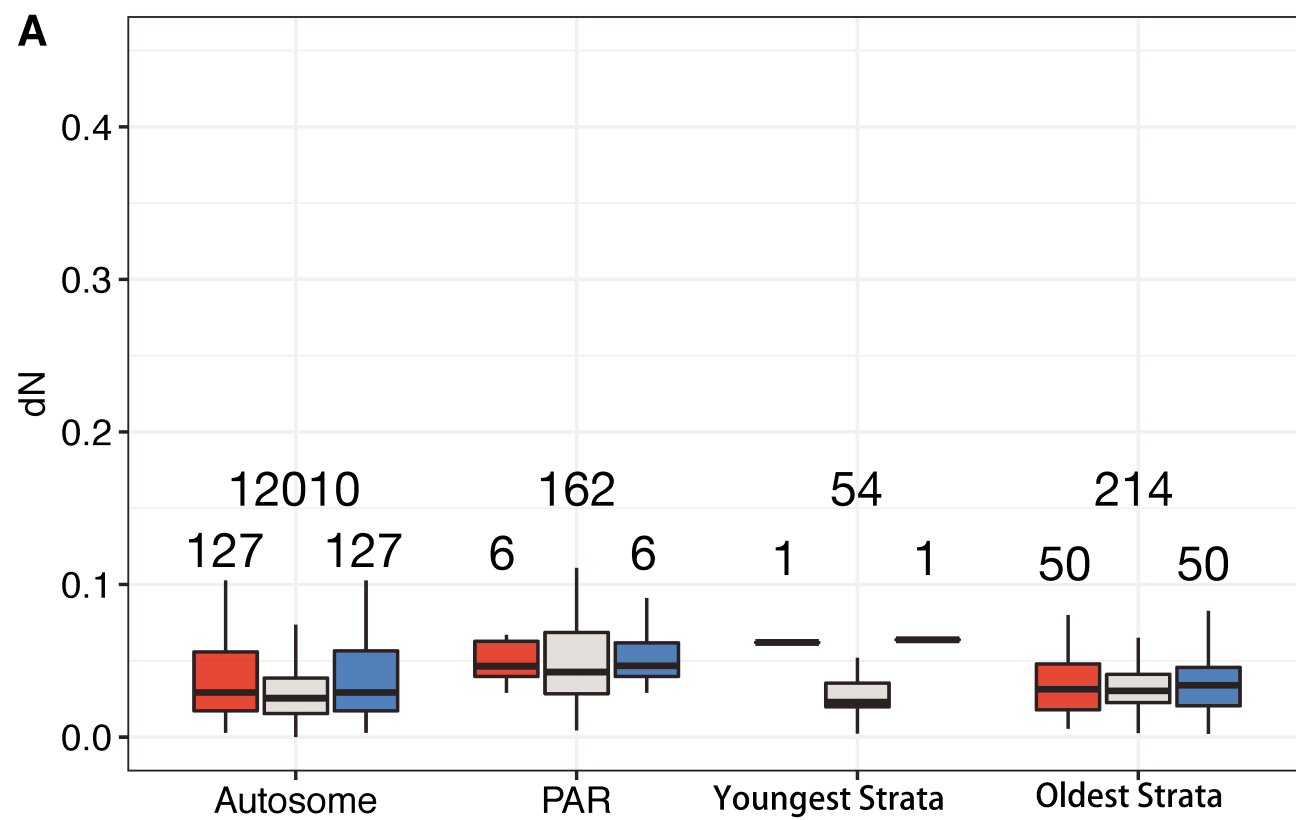
Fig. S5
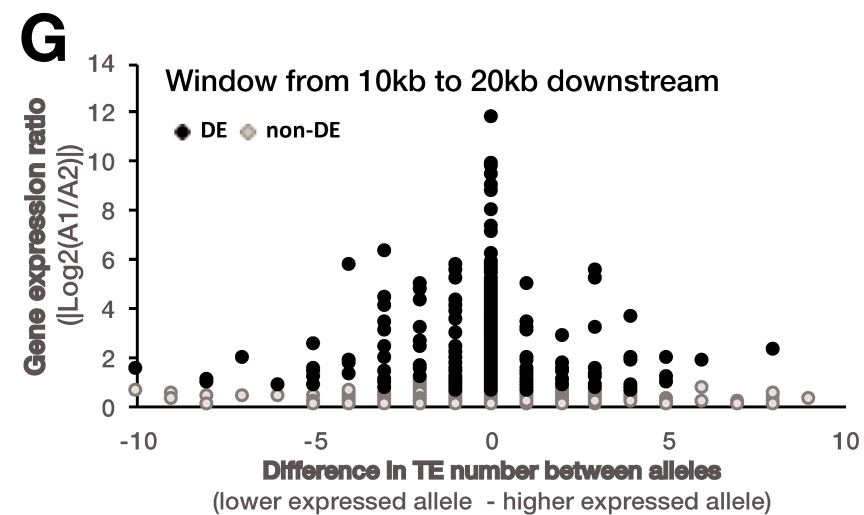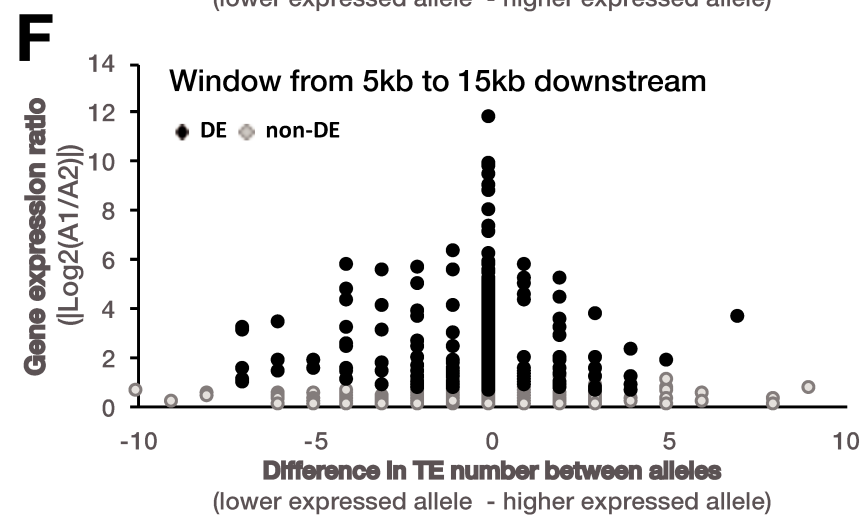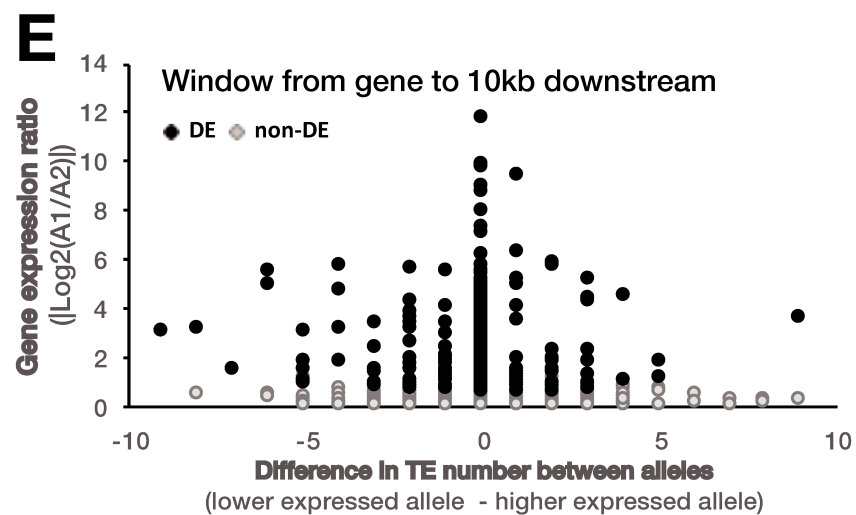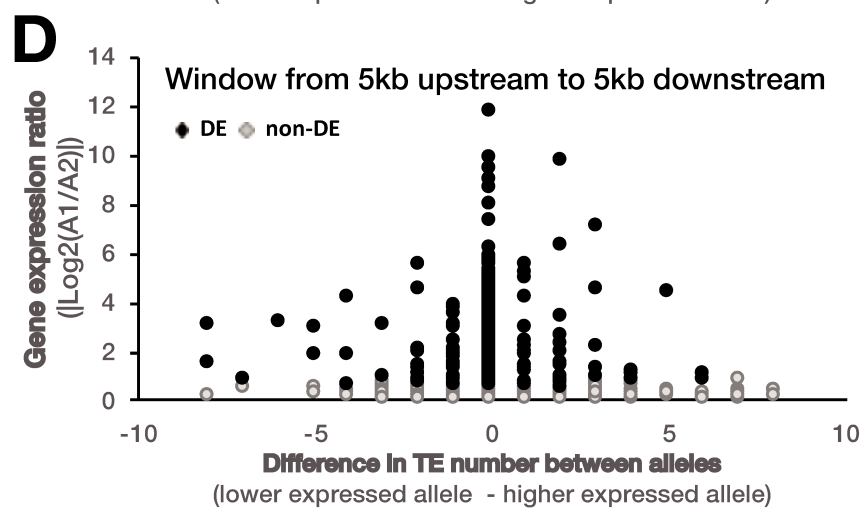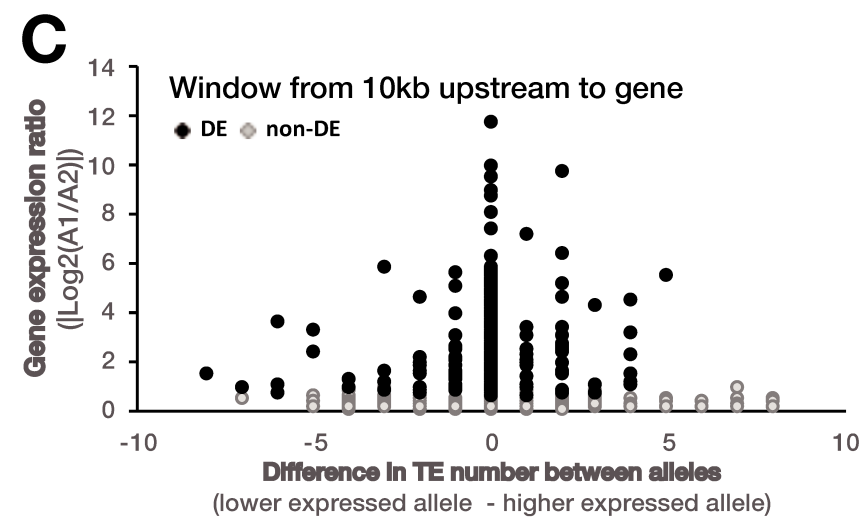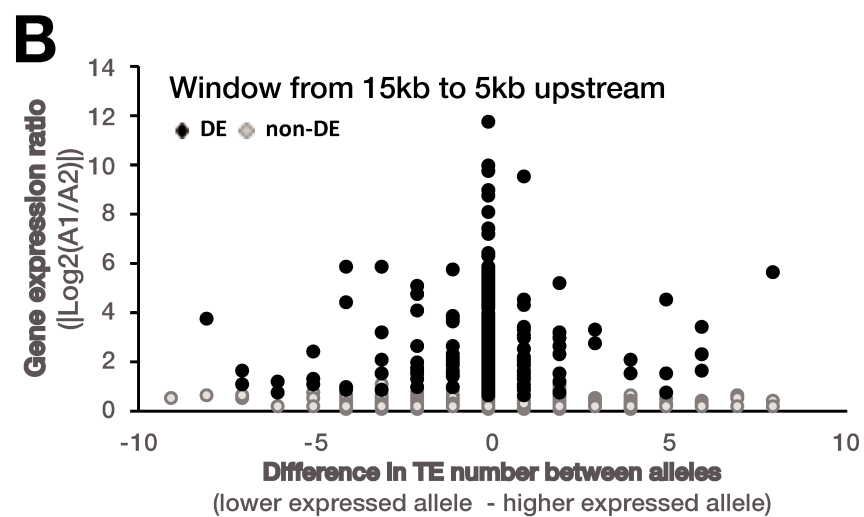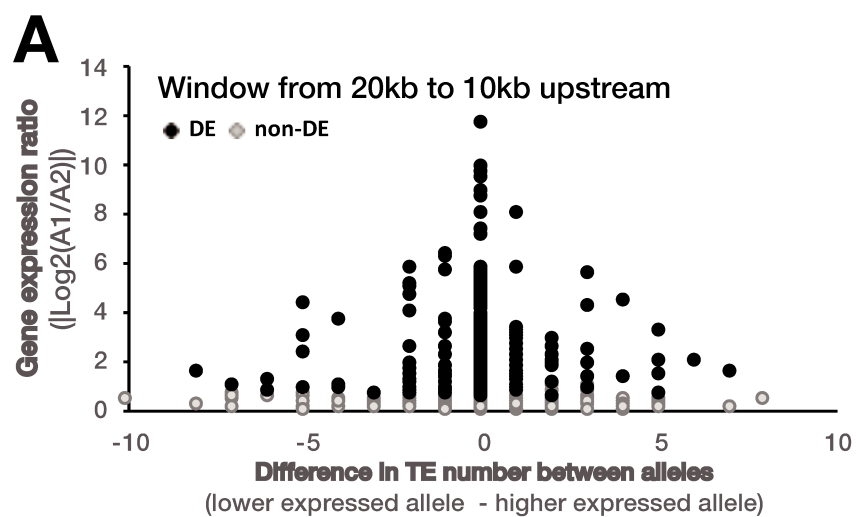
Fig. S6
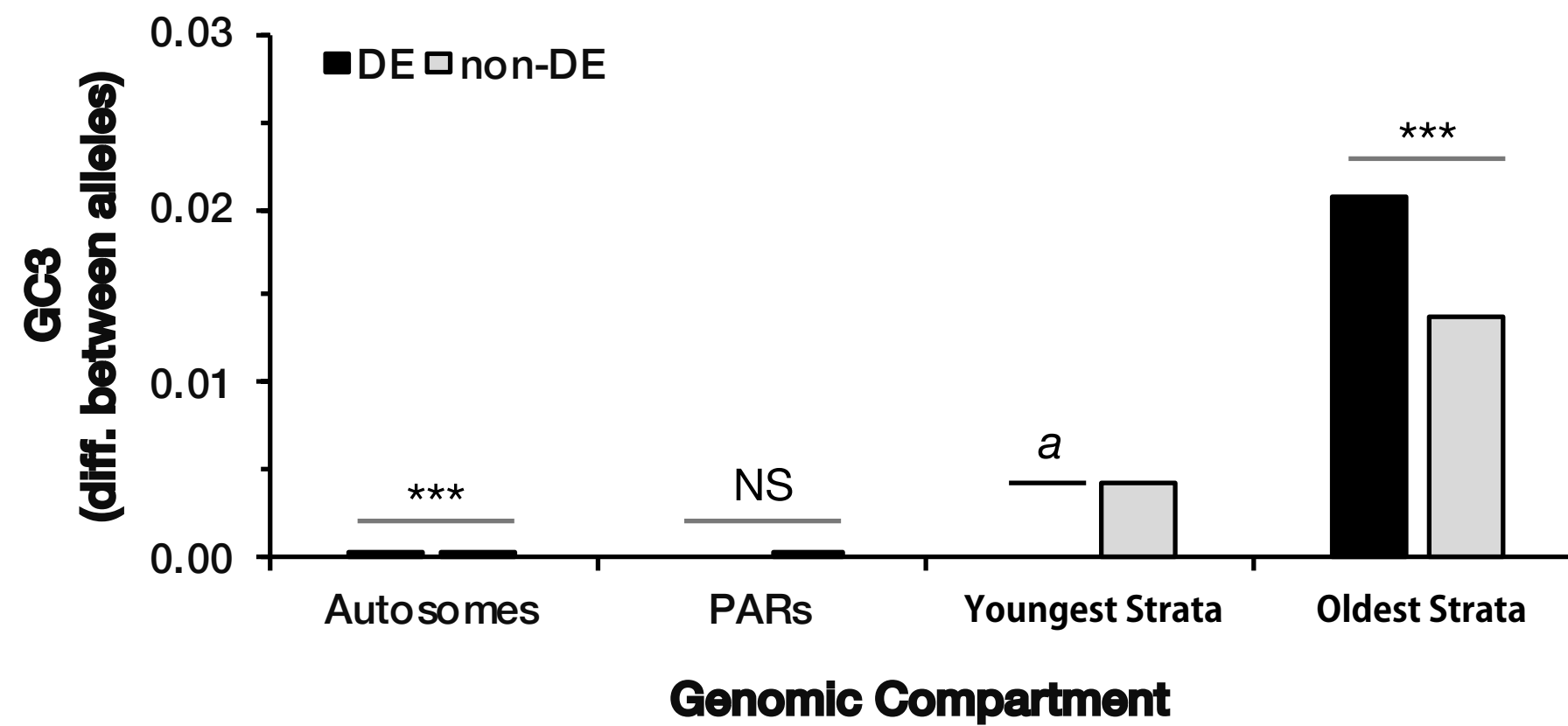
Fig. S7
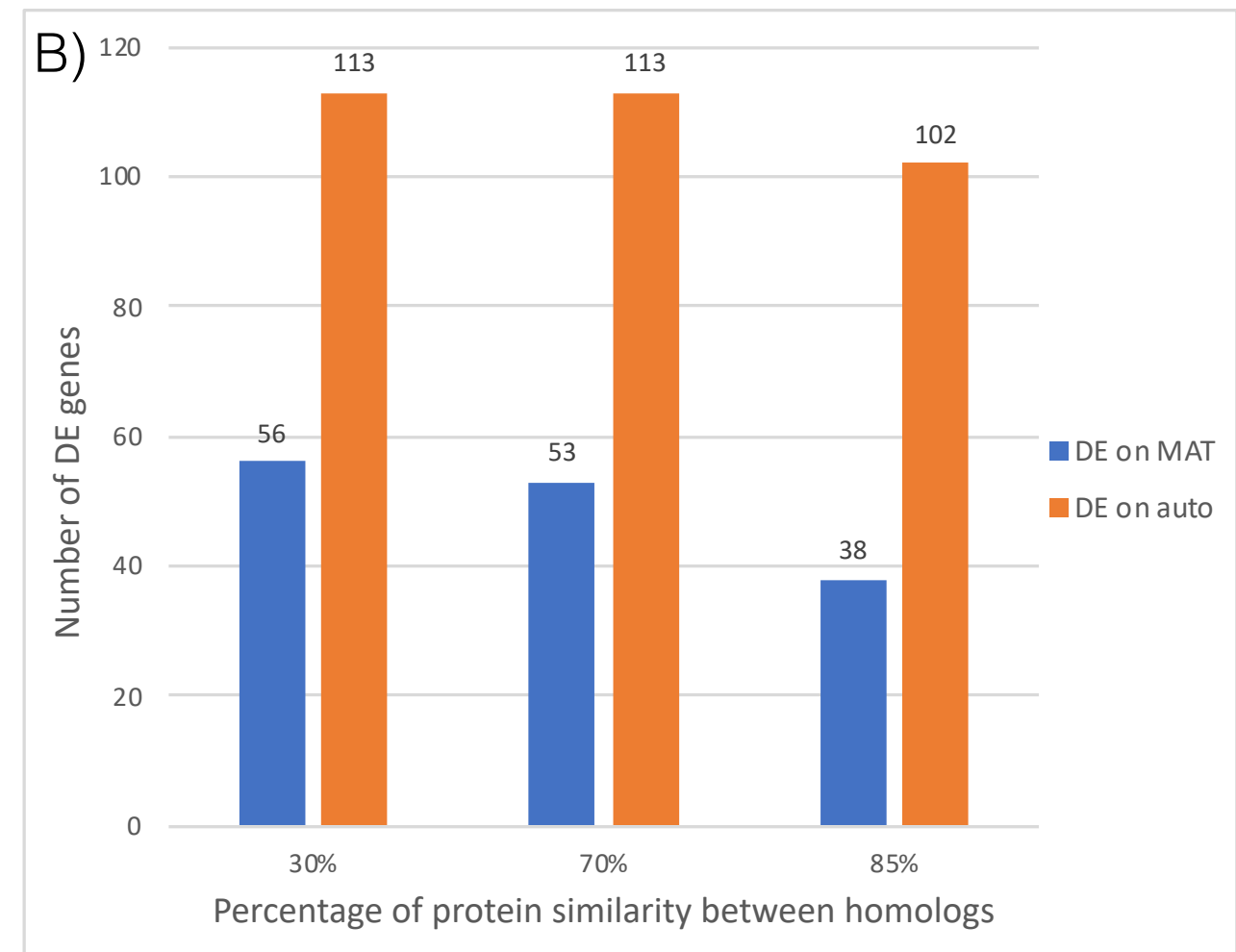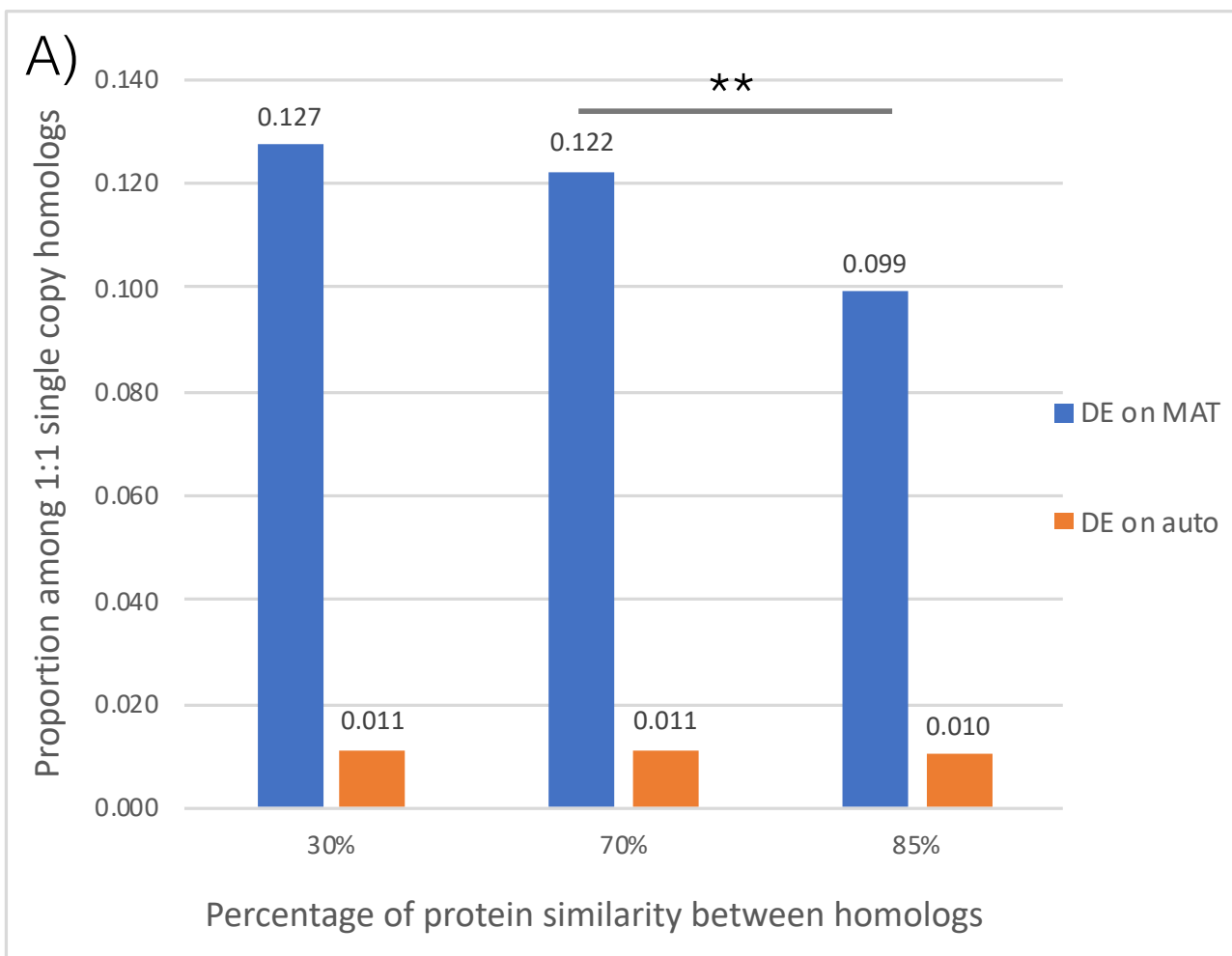
Fig. S8
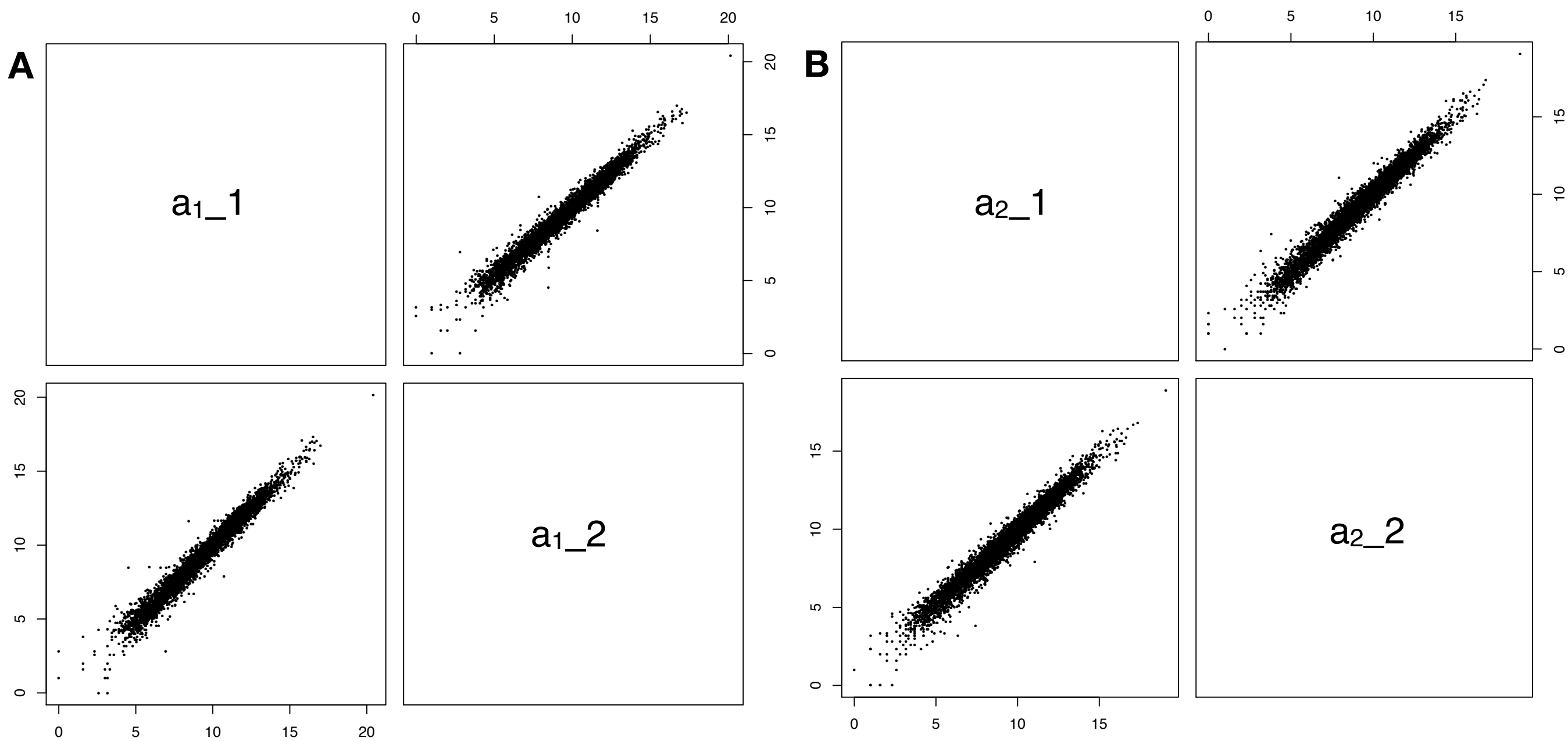
Fig. S9
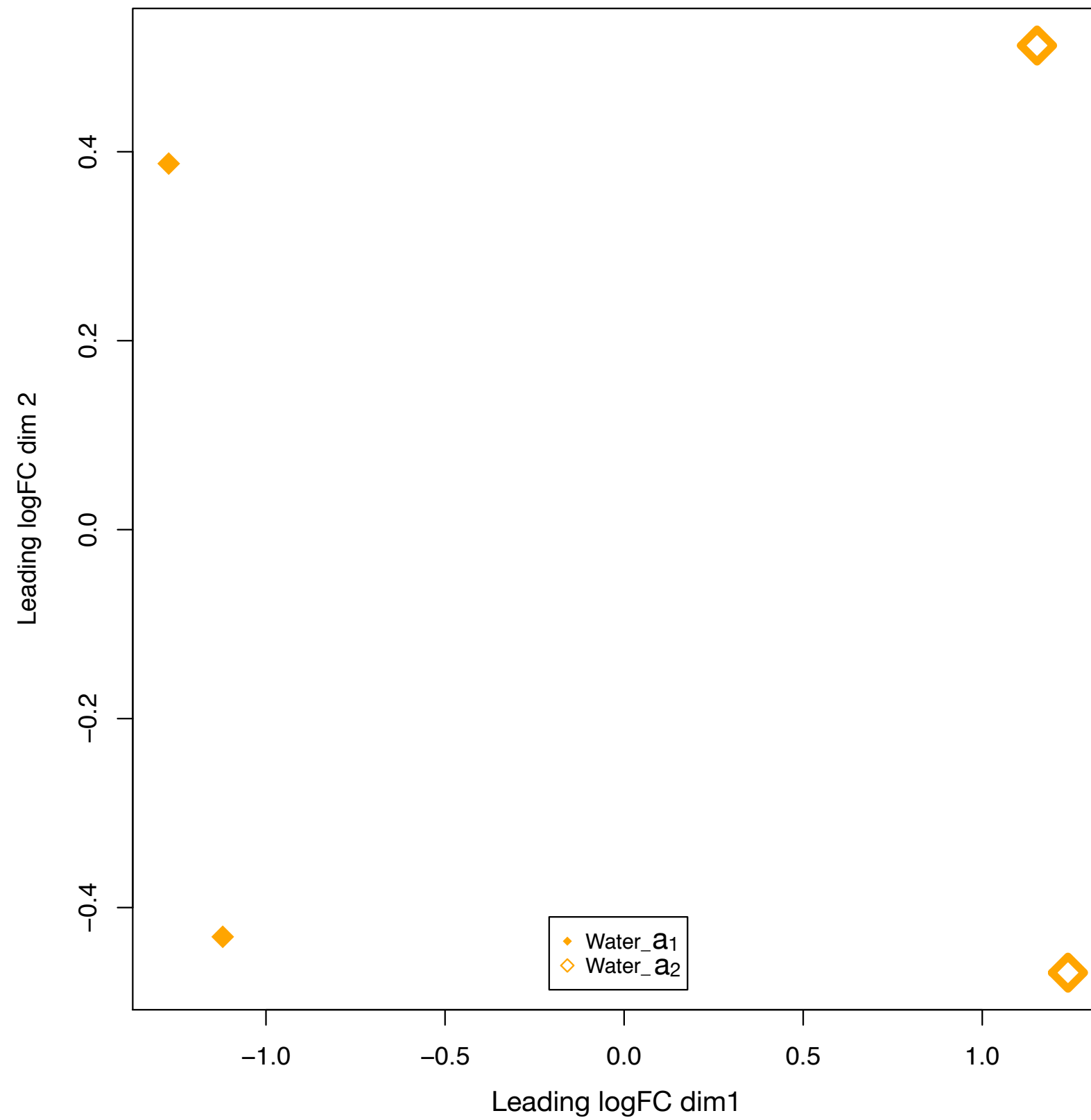
Fig. S10
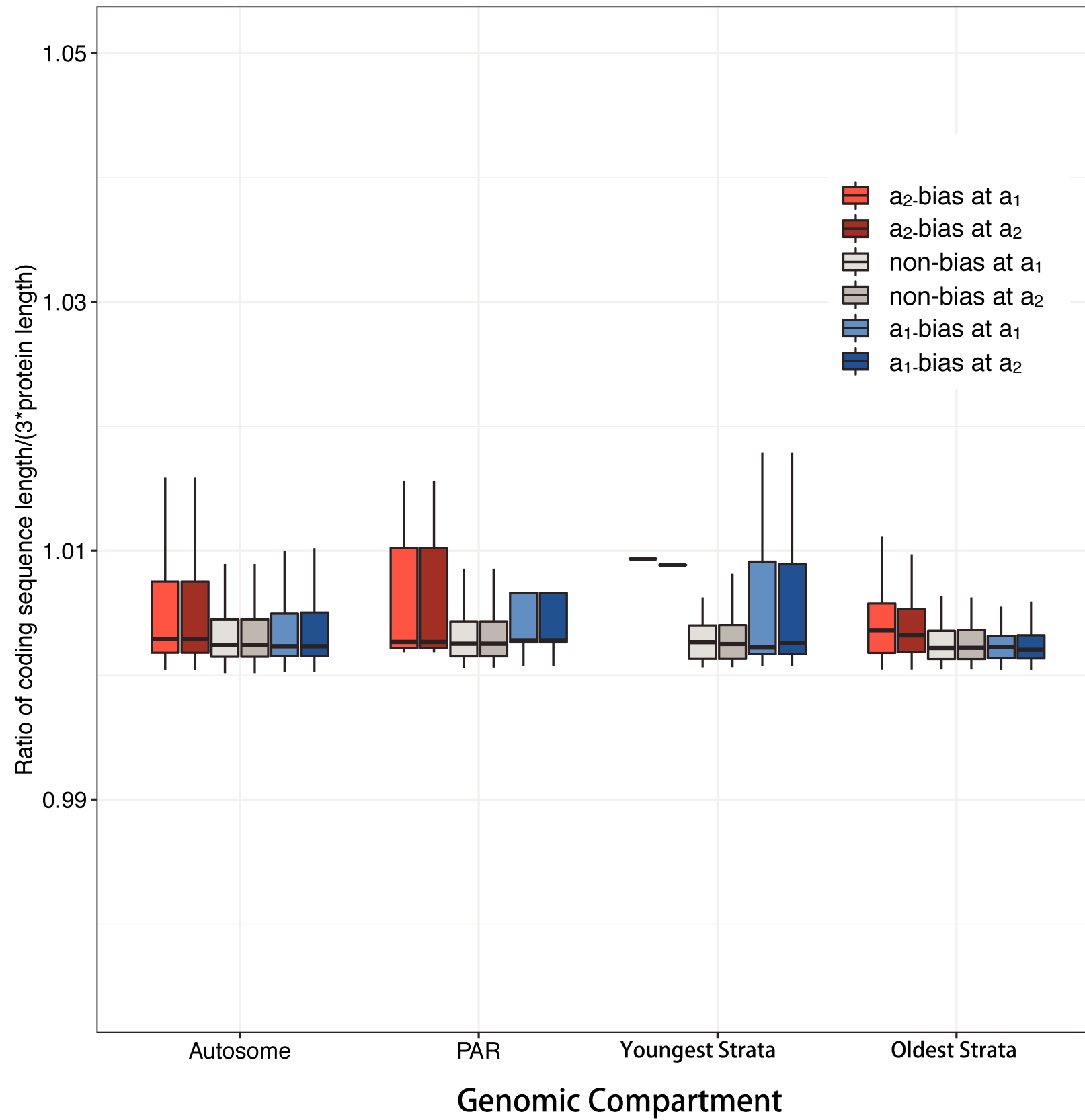
Fig. S11

Annexes

A. Article: Gene presence-absence polymorphism in castrating anther-smut fungi: recent gains and phylogeographic structure

Gene Presence–Absence Polymorphism in Castrating Anther-Smut Fungi: Recent Gene Gains and Phylogeographic Structure

Fanny E. Hartmann*, Ricardo C. Rodríguez de la Vega, Jean-Tristan Brandenburg, Fantin Carpentier, and Tatiana Giraud

Department Génétique et Ecologie Evolutives, Ecologie Systématique Evolution, Bâtiment 360, Univ. Paris-Sud, AgroParisTech, CNRS, Université Paris-Saclay, Orsay, France

*Corresponding author: E-mail: fanny.hartmann@u-psud.fr.

Abstract

Gene presence–absence polymorphisms segregating within species are a significant source of genetic variation but have been little investigated to date in natural populations. In plant pathogens, the gain or loss of genes encoding proteins interacting directly with the host, such as secreted proteins, probably plays an important role in coevolution and local adaptation. We investigated gene presence–absence polymorphism in populations of two closely related species of castrating anther-smut fungi, *Microbotryum lychnidis-dioicae* (MvSl) and *M. silenes-dioicae* (MvSd), from across Europe, on the basis of Illumina genome sequencing data and high-quality genome references. We observed presence–absence polymorphism for 186 autosomal genes (2% of all genes) in MvSl, and only 51 autosomal genes in MvSd. Distinct genes displayed presence–absence polymorphism in the two species. Genes displaying presence–absence polymorphism were frequently located in subtelomeric and centromeric regions and close to repetitive elements, and comparison with outgroups indicated that most were present in a single species, being recently acquired through duplications in multiple-gene families. Gene presence–absence polymorphism in MvSl showed a phylogeographic structure corresponding to clusters detected based on SNPs. In addition, gene absence alleles were rare within species and skewed toward low-frequency variants. These findings are consistent with a deleterious or neutral effect for most gene presence–absence polymorphism. Some of the observed gene loss and gain events may however be adaptive, as suggested by the putative functions of the corresponding encoded proteins (e.g., secreted proteins) or their localization within previously identified selective sweeps. The adaptive roles in plant and anther-smut fungi interactions of candidate genes however need to be experimentally tested in future studies.

Key words: copy number variation, pathogen, population structure, adaptive variation.

Introduction

Gene presence–absence polymorphisms segregating within species are probably a widespread and important source of genetic variation (Conrad et al. 2006, 2010). Several gene copy number changes have been shown to be adaptive, due to their effects on gene expression and phenotypic diversity (Henrichsen et al. 2009; Orozco et al. 2009). However, the proportion and type of genes affected by presence–absence polymorphism in natural populations have been little explored, and such studies are only now being facilitated by

new sequencing technologies (Schrider and Hahn 2010; Albalat and Cañestro 2016).

Several fungal plant pathogen species have small, compact genomes (Möller and Stukenbrock 2017), rendering them highly suitable for the fine mapping of intraspecific differences in gene content, as low repeat content and high-quality genome assemblies facilitate structural variant discovery (Guan and Sung 2016). Furthermore, some of the best-documented examples of adaptive evolution through gene loss or gene gain have been identified in fungal

pathogen–host plant interactions (Fouché et al. 2018). During host infection, fungal pathogens secrete proteins (i.e., effectors) that may be recognized by the host in a gene-for-gene relationship and trigger plant defense responses (Jones and Dangl 2006). Losses or gains of genes encoding proteins interacting with the host defense system have a strong impact on fitness (Presti et al. 2015). In agricultural ecosystems, the loss of entire effector genes, leading to the emergence of virulence, has been documented in several crop pathogens (Orbach et al. 2000; Schürch et al. 2004; Gout et al. 2006, 2007; Hartmann et al. 2017). The acquisition of new effector genes through horizontal gene transfer or duplication events is also associated with the emergence of virulence (Friesen et al. 2006; Khang et al. 2008; Jonge et al. 2012). The widespread use of new resistant host genotypes carrying specific resistance genes probably imposes a strong selective pressure on fungal pathogen populations in agricultural crops, promoting loss of the cognate effector gene or the retention of recently gained genes with potentially beneficial effects in terms of virulence, according to the "arms race" model of evolution (Brown and Tellier 2011). Comparisons of multiple genomes within agricultural pathogen species have shown that gene content variation can affect hundreds of genes, including a high proportion of effector genes and genes located in rapidly evolving genomic regions, in particular regions rich in repetitive elements (Gout et al. 2006; Yoshida et al. 2009, 2016; Plissonneau et al. 2016; Hartmann and Croll 2017; Plissonneau et al. 2018).

Despite its importance for coevolution, we know very little about the extent of gene presence–absence polymorphism in fungal plant pathogens in natural ecosystems. Several factors may drive specific dynamics of gene content change in natural populations. For example, wild plant parasites undergo frequent population extinction–recolonization events, with major consequences for infection success and population genetic structure (Tack and Laine 2014). High environmental heterogeneity and seed or spore banks also affect pathogen population structure and the process of coevolution in natural conditions (Laine and Tellier 2008; Koopmann et al. 2017). Host ecotype and environmental factors also have a strong impact on local pathogen populations, driving local adaptation (Laine et al. 2014; Stam et al. 2017). In natural conditions, the coevolutionary dynamics of plant–parasite interactions appear to follow essentially a "trench warfare" model, favoring balanced polymorphism at loci involved in host–parasite coevolution (Brown and Tellier 2011). In addition, the virulence of pathogen populations in wild pathosystems is usually a quantitative trait (Alexander and Antonovics 1995; Thrall et al. 2001; Koskela et al. 2002; Laine et al. 2011), suggesting that presence–absence variations of effector gene may be a less frequent mechanism of coevolution than in crops, in which virulence is more often a binary trait, with resistant varieties completely impeding infection (Möller and Stukenbrock 2017).

Microbotryum lychnidis-dioicae (MvSl) and *M. silenes-dioicae* (MvSd) are two closely related species of castrating anther-smut fungi. These highly specialized plant pathogens parasitize the white campion *Silene latifolia* and the red campion *S. dioica*, respectively (Le Gac et al. 2007; Refrégier et al. 2008). Virulence is mostly a quantitative trait, and no gene-for-gene relationship has been identified in anther-smut fungi (Alexander et al. 1993; Alexander and Antonovics 1995; Biere and Antonovics 1996; Chung et al. 2012). Populations of MvSl are genetically more diverse than those of MvSd, and form distinct genetic clusters, corresponding to the footprints of southern glacial refugia. By contrast, MvSd displays little population differentiation across Europe (Vercken et al. 2010; Badouin et al. 2017). The strong costructure between the anther-smut pathogen MvSl and its host, together with cross-inoculation experiments, are consistent with plant local adaptation (Kaltz et al. 1999; Feurtey et al. 2016). MvSl has undergone numerous selective sweeps, consistent with rapid coevolution, affecting almost 17% of the genome, whereas only 2% of the genome is affected in MvSd (Badouin et al. 2017). These two fungal species have total haploid genome size of ∼30 Mb; hyphae in plants are dikaryotic and teliospores are diploid. Both species have dimorphic mating-type chromosomes containing large regions without recombination, but with extensive rearrangements and gene losses due to permanent heterozygosity (Fontanillas et al. 2014; Badouin 2015; Branco et al. 2017). By contrast, the autosomes display low levels of heterozygosity, due to high rates of selfing and the rarity of outcrossing events (Giraud et al. 2008; Vercken et al. 2010; Badouin et al. 2017).

We investigated homozygous gene presence–absence polymorphisms on autosomes of the two sister anther-smut species MvSl and MvSd, using available genome sequence data and high-quality reference genome assemblies. The genes displaying presence–absence polymorphism were mostly recently acquired, in a single species, through duplications in multiple-gene families and their absence alleles segregated at low frequencies. The phylogeographic structure of presence–absence polymorphism in MvSl corresponded to the previously identified genetic clusters based on SNPs. Altogether, these findings suggest that most gene presence–absence polymorphism is neutral. Nevertheless, the putative functions of some genes affected by presence–absence polymorphism (e.g., secreted proteins) or their localization within previously identified selective sweeps suggest that some gene loss or gain events may be adaptive. However, functional validation in future studies will be needed to test adaptive roles in plant and anther-smut fungi interactions.

Materials and Methods

Genome Data Used for Gene Presence–Absence Polymorphism Calling

We analyzed the genome sequences of 39 MvSl isolates and 19 MvSd isolates collected from across Western Europe,

previously obtained with Illumina paired-end sequencing technology, with a mean coverage of $100\times$ (Whittle et al. 2015; Badouin et al. 2017). In total, 14 strains were sequenced in these previous studies as haploids of the a_1 or a_2 mating type, whereas the remaining 47 strains were sequenced as diploids. We downloaded raw data publicly available from the NCBI Short Read Archive (SRA) under the BioProject IDs PRJNA295022 and PRJNA269361 (see summary in supplementary table S1, Supplementary Material online).

We studied gene presence–absence polymorphism in MvSl and MvSd, using the high-quality reference genomes of the MvSl-1064 strain and the MvSd-1303 strain, previously obtained and annotated for gene models (see summary in supplementary table S2, Supplementary Material online; Branco et al. 2017). We focused on autosomes, which are highly homozygous in these species (Badouin 2015). We therefore used only one of the two available haploid genomes for each strain for further analyses (the haploid genomes of the a_1 mating type). We used the MvSl-1064-A1-R4 assembly from the European Nucleotide Archive (ENA), accession number ERS1013679, and the MvSd-1303-D assembly, accession number ERS1436592 (Branco et al. 2017).

Read Mapping

We mapped Illumina reads against the reference genomes of each species. We trimmed Illumina raw reads of the MvSl and MvSd genomes for sequence quality and removed adapter sequences with the Cutadapt v1.12 software (Martin 2011). We used the options: -q 10, 10; –minimum-length 50; -a AGATCGGAAGAGCACACGTCTGAACTCCAGTCAC; -A AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTAGATCTCGGTGGTCGCCGTATT. We aligned MvSl trimmed reads against the MvSl-1064-A1-R4 genome and MvSd trimmed reads against the MvSd-1303-D a_1 genome. We used the short-read aligner bowtie2 v2.1.0 (Langmead et al. 2009) for the mapping of trimmed reads with the following software options: –very-sensitive-local –phred33 -X 1000. We removed PCR duplicates with the MarkDuplicates tool of Picard tools version 1.88 (http://broadinstitute.github.io/picard). We improved alignment accuracy in indel regions, by locally realigning the mapped reads with the RealignerTargetCreator and IndelRealigner tools of the Genome Analysis Toolkit (GATK) version 3.7 (McKenna et al. 2010). Mean genome-wide mapping coverage of MvSl strains ranged from $54\times$ to $350\times$ and $72\times$ to $156\times$ for the MvSl-1064-A1-R4 and MvSd-1303-D reference genomes, respectively. As differences in coverage between samples may lead to a bias in the specificity and sensitivity of copy number variation detection (Guan and Sung 2016), we normalized the coverage of all isolates to a single value ($30\times$), by sampling reads at random from alignment files with the samtools v1.3.1 software (Li et al. 2009).

Gene Presence–Absence Polymorphism Calling

We used a combination of read depth-based and split read-based methods to call homozygous missing fragments in the genomes analyzed, relative to the reference genomes of the MvSl-1064 and MvSd-1303 strains. Mating-type chromosomes contain an extensive region without recombination that is permanently sheltered in a heterozygous state, which led to genomic decay and in particular the loss of hundreds of genes (Fontanillas et al. 2014; Badouin 2015). We therefore focused here exclusively on autosomes. We first used the algorithm implemented in CNVnator v0.3, which uses all mapped reads and performs a statistical analysis of read coverage in bins along the genome sequence (Abyzov et al. 2011). We set the bin size at 100 bp and retained only deletion calls, that is, missing fragments, fulfilling the following criteria: P value < 0.05, length > 500 bp, q0 < 0.4 and normalized read coverage < 0.3. As a second method, we used the Pindel v0.5.7 software, which uses information about paired reads for which only one end can be mapped. Based on the anchor point of the mapped read, the insert size and the direction of the unmapped read, Pindel predicts missing fragments by breaking the unmapped read into fragments and mapping them separately (Ye et al. 2009). We converted .bam alignment files into Pindel input format with the sam2pindel tool of Pindel. We ran Pindel with default options and used the Pindel pindel2vcf tool to convert output files into variant calling format. We retained only homozygous deletion calls of >500 bp, supported by at least 15 reads (allele depth >15). The calling of missing fragments is more challenging in short contigs, due to frequent poor-quality assembly and high repeat content. We therefore restricted our analyses to the largest contigs (minimum size of 147 kb), which summed covered 90% of the length of the autosomal genome of the MvSl-1064 and MvSd-1303 strains. To detect genes present in the reference genomes that were absent from the Illumina genomes, we retained the missing fragments called with each method that covered at least one gene, over $>90\%$ of its length. We assessed the overlaps between missing fragments and gene models masked for repeats, with the bedtools "intersect" command (Quinlan and Hall 2010). We considered gene absence events identified by both the read depth-based method and the split read-based method as single gene absence events. We determined the total length of the genomic region affected by missing fragments for each strain by summing the lengths of missing fragments detected by the two methods. To avoid counting twice the lengths of overlapping missing fragments detected by the two methods, we took into account the reunion of overlapping missing fragments.

Quality Control of Detected Gene Presence–Absence Polymorphism

To assess the overall quality of gene presence–absence polymorphism calls, we performed several quality control

steps. First, we checked the gene absence events called with our pipeline using Illumina sequencing of the very same strain (Illumina-resequenced MvSl-1064 strain; Badouin et al. 2017) against the reference high-quality genome of the MvSl-1064 strain (Branco et al. 2017). Second, we used a comparison of high-quality genome assemblies of two MvSl strains, MvSl-1064 (used as a reference for gene presence–absence calling) and a newly sequenced strain, MvSl-1318 (see section "Analyses of strain-specific genes using de novo genome assemblies" below). We compared the number of gene absence events called with our pipeline by using the Illumina-resequenced MvSl-1064 strain (Badouin et al. 2017) against the MvSl-1318 reference genome with the number of genes found to be specific to the MvSl-1318 reference genome (i.e., lacking from the MvSl-1064 reference genome) in the comparison of the two high-quality genome assemblies. We studied global synteny between genome assemblies using the nucmer command from MUMmer package v3.1 (Kurtz et al. 2004). To identify gene sequences of the MvSl-1318 strain that were absent in the MvSl-1064 strain, we performed BLASTn analyses in the genome of the MvSl-1064 strain using the sequences of predicted genes of the MvSl-1318 strain as a query. Genes were considered as absent if no BLAST hit was found on the orthologous contig with a minimum identity of 90%, a bit score value of 100 and a length of at least 90% of the gene length. We used the Integrative Genomics Viewer (IGV) browser to visually inspect read mapping in genomic regions with detected gene absence events (Robinson et al. 2011; Thorvaldsdóttir et al. 2013). Finally, we validated the presence–absence polymorphism inferred in silico for two genes using polymerase chain reactions (PCR) on DNAs from the same strains whose genomes were studied. Primers were designed in the gene coding sequences using Primer 3.0 (Rozen and Skaletsky 2000). For the gene *MvSl-1064-A1-R4_MC01-1g01717* in the MvSl-1064 reference genome, we performed PCR in nine MvSl strains using 5′ GATGTCGATGCGCTCTTTGT 3′ and 3′ CGTCATCAGTGTGCCCTTTT 5′ as forward and reverse primers, respectively. For the gene *MvSdioicae_1303_FR02_D_N206_PbcR_C014g07422* in the MvSd-1303 reference genome, we performed PCR in nine MvSd strains using 5′ GACATCAGGCACCACTCACA 3′ and 3′ ATCCA CCCGTCAAATTCGCA 5′ as forward and reverse primers, respectively. PCR reactions were conducted in a 30 µl volume containing 5–10 ng genomic DNA, 0.4 mM each of forward and backward primers, 0.25 mM dNTP, 0.75 U Taq polymerase (DreamTaq, Thermo Fisher, Inc.), PCR buffer. PCR products were amplified for 30 cycles. The resulting amplicons were examined on 1.5% agarose gels. We also tested for significant differences in gene absence calls between haploid and diploid strains within MvSl and MvSd populations.

Gene Ontology Analysis and Functional Annotation

We assigned genes to functional categories using InterproScan v5.24-63.0 (Zdobnov and Apweiler 2001) that provided information for protein family (Pfam), gene ontology (GO), and pathways. We used SignalP v4.1 (Petersen et al. 2011) and TMHMM v2.0 (Krogh et al. 2001) to predict putative secreted proteins. We defined secreted proteins as proteins with a predicted signal peptide but no predicted transmembrane helices. Putative small secreted proteins were defined as secreted proteins with a size inferior to 300 amino acids and unknown functional domain. We filtered out from the set of genes affected by presence–absence polymorphism all genes with a transposable element activity related domain, that is, reverse transcriptase (PF00078, PF07727), integrase (PF00665, PF13976), Helitron helicase-like domain (PF14214), and GAG-polypeptide of long terminal repeat (LTR) copia-type (PF14223). GO enrichment analysis of genes affected by presence–absence polymorphism was performed with the hypergeometric test implemented in the R package GOstats. GO enrichment P values were adjusted for multiple testing with the Bonferroni correction and Benjamini and Hochberg (1995) correction using the p.adjust() R function.

Annotation of Repetitive Elements

Annotation of repetitive elements was performed using a set of repeat sequences (available upon request) detected de novo in a large set of 41 genomes of the *Microbotryum* genus. Briefly, repeat motifs were detected using a combination of two tools, LTRharvest (Ellinghaus et al. 2008) that performs high-quality annotation of LTR retrotransposons and RepeatModeler (Smit and Hubley 2008) that uses results from three other programs, RECON, RepeatScout, and Tandem Repeats Finder. Detected repeat motifs were clustered based on sequence similarities (min. 90% of similarity) and the best representative sequences per cluster, that is, the centroid sequences, were kept. A set of 3599 repeat sequences was built after removing redundant sequences among centroid sequences, and keeping only transposable element sequences present in at least seven genomes. We annotated genome assemblies using the set of 3599 repeat sequences as library in RepeatMasker v4.0.3 (Smit et al. 2013). We masked gene models for transposable elements in all the genomes. Telomeric and centromeric repeats were identified by performing a BLASTn search of telomeric and centromeric motifs previously identified in the MvSl-1064 genome (Badouin 2015). Subtelomeres were defined as regions within 100 kb of the contig extremities containing telomeric repeats. Centromeres were defined as regions within 100 kb of the contig extremities containing centromeric repeats.

Expression Data

To assess expression levels of genes affected by presence–absence polymorphisms, we used RNAseq data collected

using the MvSl-1064 strain (i.e., the same strain as the MvSl reference genome) both using in vitro cultures and in vivo late infection (Perlin et al. 2015). We trimmed raw reads using cutadapt v1.12 (Martin 2011) and mapped trimmed reads to the MvSl-1064 genome using TopHat v2.1.1 (Trapnell et al. 2009). We used the featureCounts program from the Subread v1.5.3 package (Liao et al. 2014) to calculate absolute transcript abundance (i.e., number of mapped reads for each transcript). For each tested condition, we computed reads per kilobase of transcript per million mapped reads (RPKM) values using a customized R script.

Comparative Genomics Analyses

For comparative genomics analyses, we used three additional genomes from previously sequenced *Microbotryum* species (Branco et al. 2017): *M. violaceum s. str.* collected on *S. nutans* (MvSn ERS1013671), *M. lagerheimii* collected on *S. vulgaris* (MvSv ERS1013677), and *M. intermedium* collected on *Salvia pratensis* (Mint ERS1324257). To investigate gene orthology relationships between species, we compared the predicted protein sequences of the genome assemblies of MvSl-1064, MvSd-1303, MvSn, MvSv, and Mint, with BLASTp+ v2.30. The output was processed in orthAgogue (Ekseth et al. 2014) that uses Markov clustering to obtain orthologous groups. To study the synteny between the MvSl-1064 and MvSd-1303 genomes, we used the nucmer command from MUMmer package v3.1 (Kurtz et al. 2004) and performed BLASTn+ v2.30 analysis on repeat masked genomic sequences. BLAST hits were stringently filtered for a minimum alignment length of 500 bp, a minimum identity of 80%, and a bit score value of 80.

Population Structure and Statistical Analyses

Principal component analyses were computed in the R package {ADEGENET} (dudi.pca function; Jombart 2008; Jombart and Ahmed 2011). Neighbor-joining trees were performed with the R package {Ape} (Paradis et al. 2004). These analyses were performed based on unique events of gene presence–absence polymorphism, that is, using only a single gene per missing fragments, to be conservative regarding the number of gene gain or loss events. To compare the population structure based on gene presence–absence polymorphism with the population structure based on SNPs, we called genome-wide SNPs for MvSl strains against the MvSl-1064-A1-R4 genome and for MvSd strains against the MvSd-1303-D a_1 genome. For SNP calling, we used GATK version 3.7 (McKenna et al. 2010). We ran HaplotypeCaller on each strain individually using a diploid mode. We performed joint variant calls using GenotypeGVCFs on a merged gvcf variant file. SNP call filtering for quality was performed using VariantFiltration and following the GATK Good Practices for hard-filtering of variants (QUAL < 250; QD < 2; MQ < 30.0; −10.5 > MQRankSum > 10.5; −5 > ReadPosRankSum > 5;

FS > 60; SOR > 3). We kept only biallelic SNPs with a high genotyping rate (> 90%) on autosomes. In total, we detected 216, 878 biallelic SNPs in MvSl and 33, 694 biallelic SNPs in MvSd. The variant call format files are available upon request. Pairwise F_{ST} were calculated for SNPs and gene presence–absence polymorphisms between genetic clusters using the R package {hierfstats} (Goudet 2005) implementing Yang's algorithm (Yang 1998). To test the significance of the F_{ST} value obtained between the MvSd populations that we found differentiated based on gene presence–absence polymorphism, we performed a permutation test with 1000 bootstraps, assigning randomly cluster identity to each strain, keeping the cluster sizes constant, and computing mean F_{ST} values for each permutation. We performed Kruskal–Wallis rank sum test and Pearson's chi-squared test in the R software v3.3.3. Plots were performed using the software Circos v0.67-7 (Krzywinski et al. 2009) and the R package {ggplot2} (Wickham 2009).

Analyses of Strain-Specific Genes Using De Novo Genome Assemblies

For identifying genes absent from the reference genome but present in the genomes of other MvSl strains, we performed de novo genome assemblies. As MvSl showed a population structure with four genetic clusters (Badouin et al. 2017), we performed de novo assemblies for three MvSl strains belonging to the genetic clusters different from that of the reference MvSl-1064 genome. We assembled the Illumina genomes of the MvSl-1005 and the MvSl-140-01 strains that belonged to the MvSl Northwestern and MvSl Southwester genetic clusters, respectively (Badouin et al. 2017), and that showed high read depth (233× and 89×, respectively). The assemblies of the MvSl-1005 and the MvSl-140-01 Illumina genomes were performed in SpAdes v3.11.0 (Bankevich et al. 2012) with the following options: –only-assembler –careful -cov-cutoff auto using the whole set of trimmed Illumina reads of each genome. KmerGenie (Chikhi and Medvedev 2014) was used to estimate the best k-mer length for genome de novo assembly. The best predicted k values were 61 and 87 for the MvSl-1005 genome and the MvSl-140-01 genome, respectively. These assemblies had lower quality than the MvSl-1064 genome assembly (see summary in supplementary table S2, Supplementary Material online) and they are available upon request. In order to obtain a high-quality genome assembly of a second MvSl strain, we sequenced separately the two haploid genomes (corresponding to the a_1 and a_2 mating types) of the diploid MvSl-1318 strain (collected on *S. latifolia*, Olomouc, Czech Republic, Coord. GPS: 49.569172 and 17.287057, previously shown to belong to the MvSl Eastern genetic cluster), with P6/C4 Pacific Biosciences SMRT technology. The sequencing and assembly of the MvSl-1318 genomes were generated as described in (Branco et al. 2017). Briefly, we assembled the genome using the wgs-8.2

version of the PBcR assembler and polished the assembly with the quiver software (https://github.com/PacificBiosciences/GenomicConsensus). The gene models were predicted with EuGene using a training data set as described in (Branco et al. 2017). The assembly of the haploid genome of a$_1$ mating type (MvSl-1318-D) and the assembly of the haploid genome of a$_2$ mating type (MvSl-1318-T) are available at GenBank under the BioProject accession number PRJNA437556 (BioSample IDs SAMN08667584 and SAMN08667587, respectively). The raw reads are available at NCBI SRA under the BioProject accession number PRJNA437556 (BioSample IDs SRR6825078 and SRR6825079 for MvSl-1318-D and MvSl-1318-T, respectively). We only used the a$_1$ haploid genome for our analyses. Gene models were masked for repeats.

To detect strain-specific genes in the assembled genomes of the MvSl-1318, MvSl-140-01, and MvSl-1005 strains, we applied the same procedure as the one used to detect genes present in the reference genomes and absent from the Illumina genomes, for consistency. We mapped Illumina sequencing data of the MvSl-1064 strain (Badouin et al. 2017) against the three genome assemblies and called gene presence–absence polymorphism with our gene presence–absence detection pipeline. We restricted our analyses to the largest contigs, which covered 90% of the autosomal genome length in the three genomes (minimum size of 317,619, 11,561, and 11,985 bp in the MvSl-1318 strain, MvSl-140-01 strain, and MvSl-1005 strain, respectively). We identified autosomal contigs in the three genome assemblies by global synteny analyses with the genome assemblies of the MvSl-1064 strain using the nucmer command from MUMmer package (Kurtz et al. 2004). As the MvSl-140-01 and MvSl-1005 strains were sequenced as diploid, we excluded any contigs with large synteny with the mating-type chromosomes a$_1$ and a$_2$. The MvSl-1064-A2-R4 assembly (the haploid genomes of a$_2$ mating type) was accessed from ENA at accession number ERS459551. To analyze population structure of MvSl-1318-specific genes, we called gene presence–absence polymorphism with our pipeline in all MvSl strains against the MvSl-1318 assembly and selected missing fragments affecting the MvSl-1318-specific genes. Mean genome-wide mapping coverage of MvSl strains ranged from 53× to 345× on the MvSl-1318-D reference genome.

Results

Gene Presence–Absence Event Calling and Quality Control

We called homozygous gene presence–absence polymorphisms by comparing the available high-coverage Illumina genomes of 38 MvSl strains and 19 MvSd strains sampled across Europe (supplementary table S1, Supplementary Material online; Whittle et al. 2015; Badouin et al. 2017) with high-quality reference genomes for these two species: MvSl-1064 and MvSd-1303 (supplementary table S2,

Supplementary Material online; Badouin 2015; Branco et al. 2017). Read mapping identified genes present in the reference genomes that were absent from the Illumina genomes. We called missing fragments of >500 bp by two different methods, to increase the sensitivity of copy number variation calls. The read depth-based method involved scanning the genome for local variation in read depth and calls missing fragments in genomic regions with low read-depth values (Abyzov et al. 2011). The split read-based method is based on the identification of read pairs for which only one read is mapped onto the genome (supplementary fig. S1, Supplementary Material online; Ye et al. 2009; Lin et al. 2015). We focused on missing fragments affecting >90% of gene lengths on autosomes, after masking gene models for repeated elements. Merging the missing fragments detected by the two methods resulted in the calling of 810 missing fragments in MvSl and 99 in MvSd. Most missing fragments were small (median length of 2.0 kb in MvSl and 3.1 kb in MvSd; supplementary fig. S2A, Supplementary Material online). In both MvSl and MvSd, there was little overlap between methods concerning the missing fragments detected (18.7% and 12.5%, respectively). This was expected, as the two methods were based on different read mapping approaches. The overlap between missing fragments detected by the two methods covered a mean of 68 kb per individual in MvSl, and 19 kb in MvSd (supplementary fig. S2B, Supplementary Material online). Most missing fragments affected a single gene (supplementary fig. S2C, Supplementary Material online). In MvSl, 473 missing fragments affected a single gene and 38 affected groups of at least five genes. In MvSd, 48 missing fragments affected a single gene and one affected a group of five genes. In total, we detected 953 gene absence events in MvSl, and 161 in MvSd. The number of missing genes detected per strain ranged from 2 to 50, with a median value of 24 genes in MvSl, and 1 to 15 with a median value of eight genes in MvSd (supplementary fig. S2D, Supplementary Material online).

Several quality control checks indicated that the inferences for the missing fragments were reliable. First, no missing fragments were detected following genome resequencing of the MvSl-1064 strain (Badouin et al. 2017), the source of the reference genome sequence. Second, we found a large overlap between the missing fragments detected with our pipelines and those detected by comparing high-quality genome assemblies for two MvSl strains. Out of 15 genes detected to be missing in one strain with our pipeline, four genes were detected present using BLAST analyses on de novo assemblies (supplementary text S1 and supplementary fig. S3, Supplementary Material online); however, BLAST analyses can match paralogs. The visual inspection of read mapping in genomic regions with detected gene absence events using a genome browser suggested a rate of false positive of ∼15% for all analyzed genomes. Finally, PCR tests validated in silico inferences for two genes affected by presence–absence
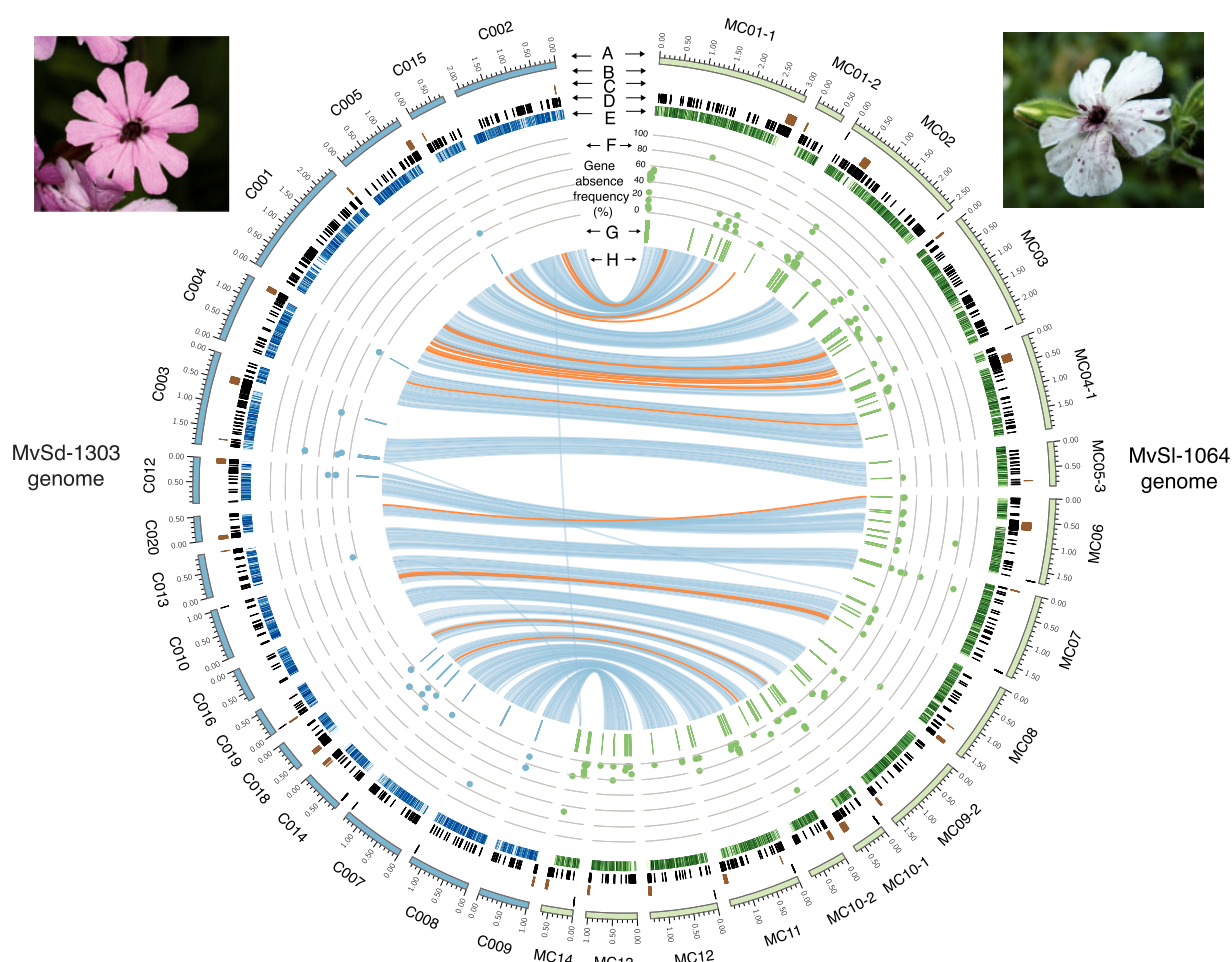
Fig. 1.—Genome-wide location of gene presence–absence polymorphism events and their frequencies in *Microbotryum lychnidis-dioicae* (MvSl) and *Microbotryum silenes-dioicae* (MvSd) strains. The tracks represent the following features: (*A*) Contigs > 500 kb in length from the MvSd-1303 (left, blue tracks) and MvSl-1064 (right, green tracks) reference genomes; (*B*) Location of subtelomeric repeats; (*C*) Location of centromeric repeats; (*D*) Location of transposable elements; (*E*) Gene density in 10-kb nonoverlapping windows; the color gradient shows gene density differences from 0% (light) to 100% (dark); (*F*) Location of genes displaying presence–absence polymorphism (*x* axis) and frequency of the gene absence allele (*y* axis, from 0% to 100%) in MvSd (left, blue dots) and MvSl strains (right, green dots); (*G*) Location of fragments identified as missing in MvSd (left, blue track) and MvSl strains (right, green track); (*H*) Links representing collinearity of genomic regions >10 kb between the MvSl-1064 and MvSd-1303 reference genomes, with the orange links corresponding to inversions. The images show spores from MvSd in the anthers of *Silene dioica* (left) and spores from MvSl in the anthers of *S. latifolia* (right).

polymorphism. The gene *MvSl-1064-A1-R4_MC01-1g01717* could not be amplified in five MvSl strains that were predicted to lack the gene in silico. The gene *MvSdioicae_1303_FR02_D_N206_PbcR_C014g07422* could not be amplified in three MvSd strains that were predicted to lack the gene in silico.

Contrasting Gene Presence–Absence Polymorphism Patterns between Species

We found that different genes and different numbers of genes displayed presence–absence polymorphism in MvSl and MvSd. We identified 186 autosomal genes displaying presence–absence polymorphism in the set of 38 MvSl strains

studied, corresponding to 2% of the total gene content of MvSl-1064 autosomes (fig. 1, right panel; supplementary table S3, Supplementary Material online). We identified 51 autosomal genes displaying presence–absence polymorphism in the set of 19 MvSd strains, corresponding to 0.6% of the total gene content of MvSd-1303 autosomes (fig. 1, left panel; supplementary table S3, Supplementary Material online). Fewer genes displayed presence–absence polymorphism in MvSd than in MvSl, consistent with the lower SNP and microsatellite diversity previously reported in MvSd (Vercken et al. 2010; Badouin et al. 2017). In agreement with this diversity difference, we identified 216, 878 biallelic SNPs in MvSl and 33, 694 biallelic SNPs in MvSd by mapping MvSl and MvSd Illumina genomes on their respective reference genome.
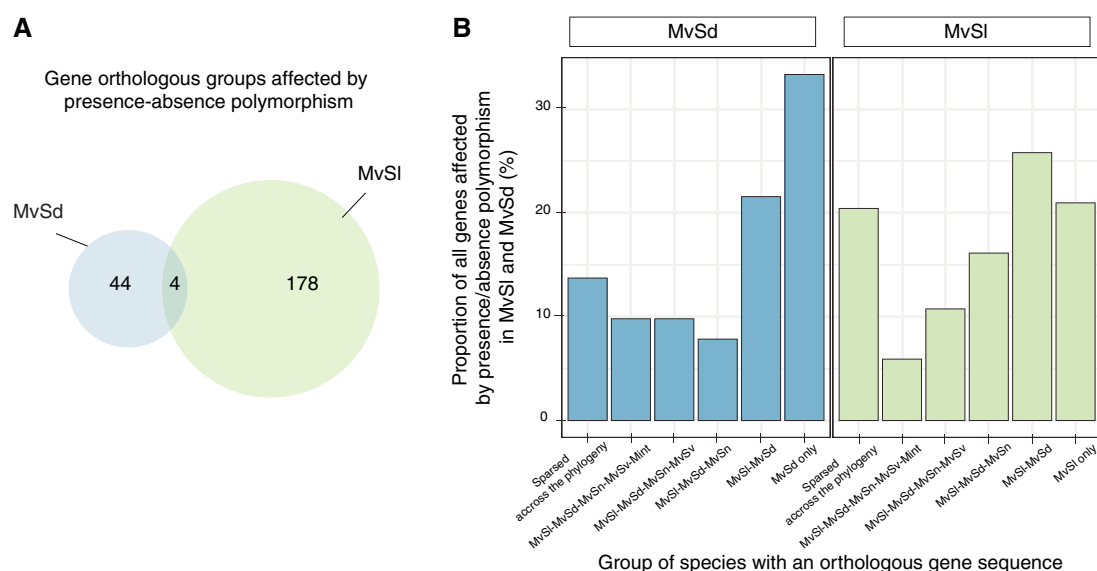
Fig. 2.—Evolutionary history of genes displaying presence–absence polymorphism. (*A*) Venn diagram of shared and specific orthologous gene group numbers displaying presence–absence polymorphism in *Microbotryum lychnidis-dioicae* (MvSl) and *Microbotryum silenes-dioicae* (MvSd). (*B*) Groups of *Microbotryum* species with orthologous gene sequences, for the genes displaying presence–absence polymorphism in MvSl and MvSd. We searched for orthologous groups in the sister species MvSl and MvSd and three closely related *Microbotryum* species: *M. violaceum s. str.* (MvSn), *M. lagerheimii* (MvSv), and *M. intermedium* (Mint).

Gene absence alleles segregated at low frequency in both species, with a median frequency of 5% in MvSl and 10% in MvSd (fig. 1*E*; supplementary table S3, Supplementary Material online).

We searched for orthology relationships between the genes identified as displaying presence–absence polymorphism on MvSl-1064 and MvSd-1303 autosomes. The two closely related species shared 7, 869 groups of orthologous autosomal genes, containing 83.3% and 78.2% of the total gene content of MvSl-1064 and MvSd-1303 autosomes, respectively. In total, 59% of the genes displaying presence–absence polymorphism in MvSl and 49% of those in MvSd belonged to orthologous groups present in both species (fig. 2*B*). However, only four orthologous groups were affected by presence–absence polymorphism in both species (fig. 2*A*). Gene presence–absence polymorphism affected different genomic regions in the two species (fig. 1*G*).

Functional Characteristics, Evolutionary History, and Genomic Environment of the Genes Displaying Presence–Absence Polymorphism

Most of the genes displaying presence–absence polymorphism encoded proteins of unknown function, to a greater extent than genome average. Indeed, a total of 86% and 88% of the genes displaying presence–absence polymorphism encoded proteins of unknown function compared to 44% and 45% for all autosomal genes in MvSl and MvSd, respectively. Based on the available expression data for MvSl-1064 in three types of conditions, poor and rich in vitro

medium and *in planta* (Perlin et al, 2015), we showed that the genes displaying presence–absence polymorphism in MvSl had lower levels of expression than other genes (comparisons of RPKM values; Kruskal–Wallis rank sum tests P value < 2.2e-16 in all conditions tested). The genes displaying presence–absence polymorphism in MvSl included a high proportion of genes not expressed *in planta* (RPKM < 1; 37%, vs 10% for other genes; one-tailed Fisher's exact test P value < 2.2e-16). In both species, the genes displaying presence–absence polymorphism were also significantly shorter than other genes (difference in median gene length in MvSl strains $= 928$ bp; Kruskal–Wallis rank sum test P value < 2.2e-16; difference in median gene length in MvSd strains $= 852$ bp; Kruskal–Wallis rank sum test P value $= 1.5$e-07). Finally, the genes displaying presence–absence polymorphism included a higher proportion of genes from multiple-copy gene families (24% and 25% of all genes displaying presence–absence polymorphism in MvSl and MvSd, respectively) than other genes (7% and 9% in MvSl and MvSd, respectively).

The genes displaying presence–absence polymorphism and encoding proteins with predicted functions included a large proportion of genes associated with biological processes relating to DNA repair mechanisms and the cellular response to stress in MvSl and to glycoprotein metabolism and glycosylation in MvSd. Corresponding gene ontology terms were not significantly enriched after Bonferroni multiple testing correction, but were significantly enriched after Benjamini and Hochberg multiple testing correction (supplementary table S4, Supplementary Material online). We found no enrichment in any specific cellular compartment. In particular, genes

encoding secreted proteins, which frequently play a major role in the interactions of fungal pathogens with their hosts (Fouché et al. 2018), were not overrepresented among the genes displaying presence–absence polymorphism. Three genes encoding small secreted proteins were nevertheless absent from some MvSl and MvSd strains. In MvSl, genes encoding a secreted peptidase and a histidine phosphatase, both potentially involved in pathogenicity (Monod et al. 2002; Albataineh and Kadosh 2016; Vincent et al. 2016), also displayed presence–absence polymorphism (supplementary table S3, Supplementary Material online). Five genes encoding proteins involved in RNA interference activity (Piwi domain, Argonaute domain) were absent in one MvSd strain (supplementary table S3, Supplementary Material online). Some plant pathogens use the RNA interference machinery to prevent the expression of host immunity genes (Weiberg and Jin 2015).

We investigated the evolutionary history of genes displaying presence–absence polymorphism by searching for orthologs in three *Microbotryum* species closely related to the sister species MvSl and MvSd, using the available genomes of *M. violaceum* s. str. (MvSn), *M. lagerheimii* (MvSv), and *M. intermedium* (Mint) (Branco et al. 2017). Most of the genes displaying presence–absence polymorphism (46.7% in MvSl and 54.8% in MvSd, fig. 2B) were recently acquired, that is, were present only in one or both of the sister species MvSl and MvSd and not in the other *Microbotryum* species. Only 5.9% of the genes displaying presence–absence polymorphism in MvSl and 9.8% of those displaying presence–absence polymorphism in MvSd were present in all five *Microbotryum* species. This finding indicates that presence–absence polymorphism is rare in ancient genes, that were present in the common ancestor of the five *Microbotryum* species. Most of the genes displaying presence–absence polymorphism in MvSl and MvSd instead corresponded to species-specific genes, that is, genes recently gained in a single species, after their divergence from other species (fig. 2B).

The genes displaying presence–absence polymorphism in MvSl were found in subtelomeric regions (defined as regions within 100 kb of the contig extremities and containing telomeric repeats) more frequently than would be expected by chance (fig. 3A). A total of 10.7% of genes in subtelomeric regions were affected by presence–absence polymorphism compared to 2% in the whole genome. We checked by visual inspection that the gene absence events called in subtelomeric regions were not artifacts due to low mapping quality. The small number of subtelomeric repeats in the MvSd-1303 genome precluded testing for an enrichment in gene presence–absence polymorphism in subtelomeric regions in MvSd, but gene absence events were frequently found close to the ends of the contigs (i.e., within 100 kb of the contig extremities). The genes displaying presence–absence polymorphism in MvSl were also found closer to centromeric regions (defined as regions within 100 kb of the contig

extremities and containing centromeric repeats) than would be expected by chance (fig. 3B). The genes displaying presence–absence polymorphism were also significantly closer to transposable elements than other genes. In MvSl, 42% of the genes displaying presence–absence polymorphism were located within 5 kb of the nearest transposable element, against only 19% of other genes (fig. 3C). In MvSd, 69% of genes displaying presence–absence polymorphism were located within 5 kb of the nearest transposable element, against only 23% of other genes (fig. 3D). We found significant differences in the abundances of the transposable element families closest to the genes displaying presence–absence polymorphism and those closest to other genes in MvSl, but not in MvSd (fig. 3E and F). LTR transposons, the most frequent family of repeats in the MvSl-1064 reference genome (Hood 2005; Perlin et al. 2015), were significantly more frequently found closest to genes displaying presence–absence polymorphism than to other genes (one-tailed Fisher's exact test P value $= 0.0007345$ in MvSl; not significant in MvSd). Conversely, members of the Helitron family were less frequently found closest to genes displaying presence–absence polymorphism than to other genes in MvSl (one-tailed Fisher's exact test P value $= 0.0001393$; not significant in MvSd).

Population Structure of Gene Presence–Absence Events

We investigated the population structure of gene presence–absence polymorphism in each species. For MvSl, we identified four genetic clusters based on the three first components of a principal component analysis and a dendrogram (fig. 4A–D). These clusters were consistent with the population subdivision previously described based on genome-wide SNPs (fig. 4C) and corresponding to the phylogeographic footprints of glacial refugia (Badouin et al. 2017). The congruence of the genetic subdivisions based on SNPs and gene presence–absence polymorphism in MvSl resulted from the large proportion of cluster-specific presence or absence among the genes affected by presence–absence polymorphisms (64%; fig. 5A). Within MvSl, we found higher mean pairwise F_{ST} for SNPs (mean value $= 0.32$ for pairwise comparisons) than for gene presence–absence polymorphisms (mean value $= 0.15$). Within MvSd, we observed low levels of population structure (fig. 4E–H), consistent with previous findings for SNPs (fig. 4G) and microsatellites (Vercken et al. 2010; Badouin et al. 2017). However, a principal component analysis of gene presence–absence polymorphism suggested here the existence of genetic differentiation between strains across one east–west longitudinal gradient (fig. 4E and F). When calculating mean pairwise F_{ST} between the eight MvSd strains sampled in Eastern Europe and the eleven MvSd strains sampled in Western Europe (fig. 4E), we found higher differentiation for gene presence–absence (mean pairwise F_{ST} value $= 0.12$) than for SNPs (mean pairwise F_{ST} value $= 0.08$). The observed mean F_{ST} value calculated between
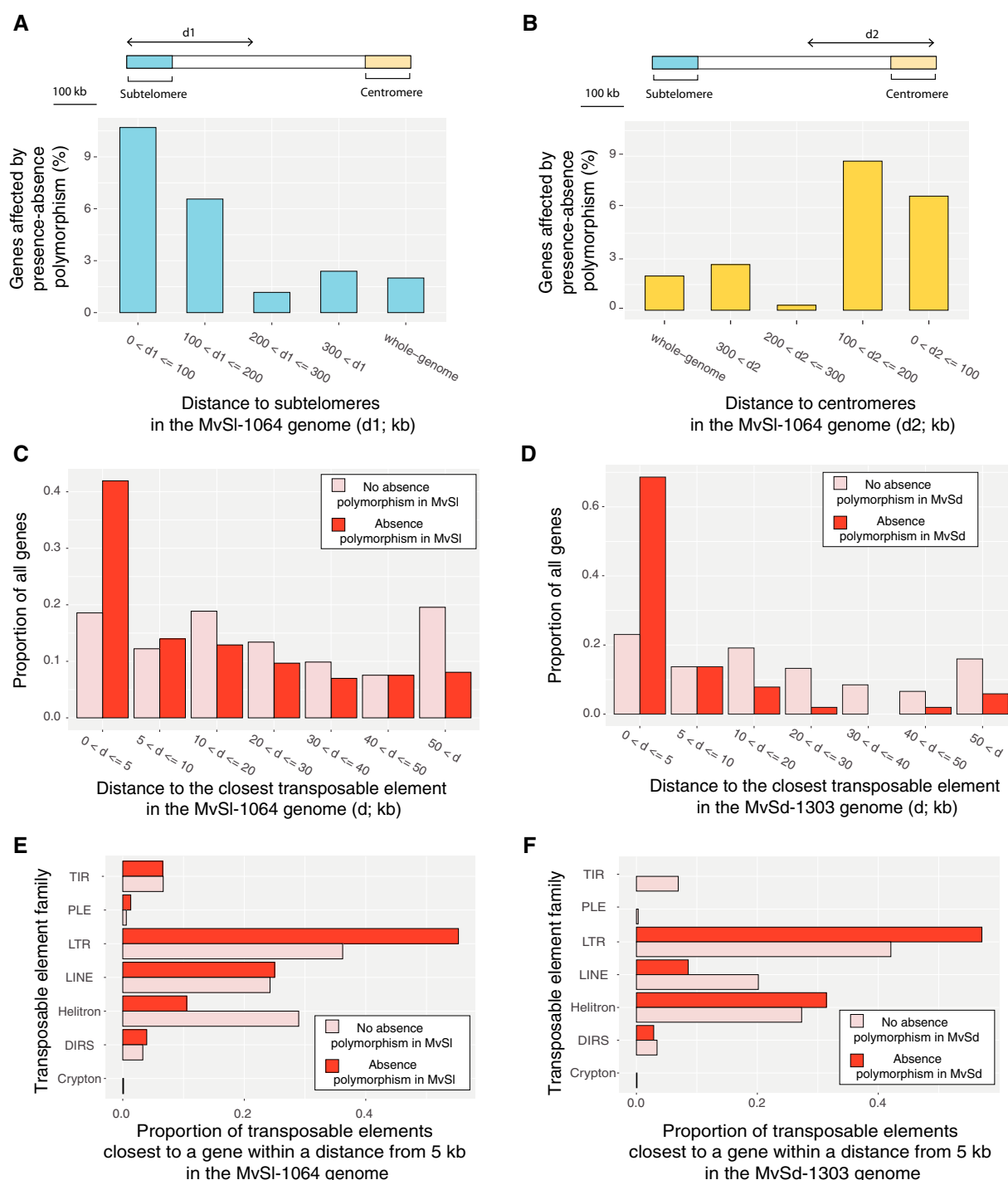
FIG. 3.—Genome features associated with gene presence–absence polymorphism in *Microbotryum lychnidis-dioicae* (MvSl) and *Microbotryum silenes-dioicae* (MvSd). (*A* and *B*) Proportion of genes displaying presence–absence polymorphism according to the distance to subtelomeres (*A*) and centromeres (*B*) in MvSl. Regions within 100 kb of contig extremities and containing subtelomeric repeats were considered to be subtelomeric. Regions within 100 kb of contig regions containing centromeric repeats were considered to be centromeric. Only contigs of >500 kb in length and containing subtelomeric motifs and centromeric repeats were included in the analysis (MC02, MC03, MC06, MC07, MC10-1, MC11, MC12, MC14). (*C* and *D*) Distance (kb) to the closest transposable element for genes displaying presence–absence polymorphism and genes not displaying presence–absence polymorphism in MvSl (*C*) and MvSd (*D*). Distances were classified into seven categories. (*E* and *F*) Family of the closest transposable element for genes displaying presence–absence polymorphism and genes not displaying presence–absence polymorphism in MvSl (*E*) and MvSd (*F*). Only transposable elements located <5 kb away from a gene were considered.
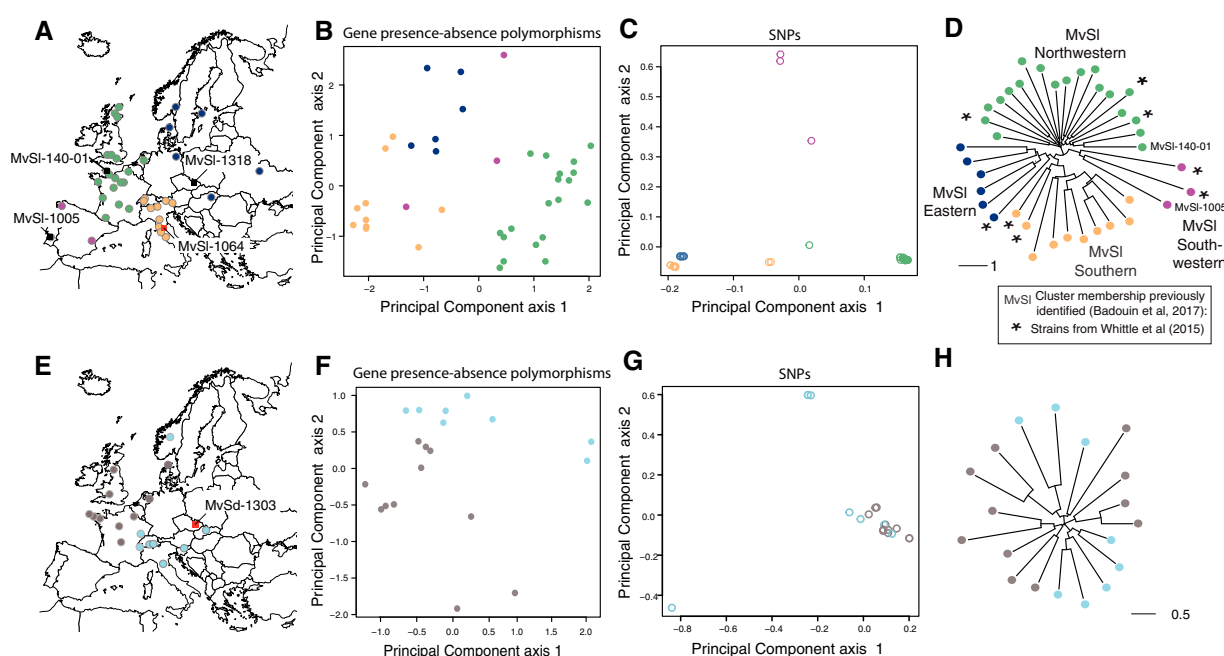
FIG. 4.—Population structure of gene presence–absence polymorphism in *Microbotryum lychnidis-dioicae* (MvSl) and *Microbotryum silenes-dioicae* (MvSd). (*A*) Geographic origin of the MvSl strains. Sampling location is indicated by circles for the MvSl strains sequenced with Illumina technology. The colors correspond to the genetic clusters identified by the dendrogram (D). The sampling locations for the reference genomes used for mapping are indicated by squares in red. The three MvSl strains that were de novo assembled (i.e., the MvSl-1318 strain originating from the MvSl Eastern cluster, the MvSl-140-1 strain originating from the MvSl Northwestern cluster, and the MvSl-1005 strain originating from the MvSl Southwestern cluster) are indicated by squares in black. (*B*) Principal component analysis of gene presence–absence polymorphism for MvSl, based on 124 unique gene presence–absence polymorphisms, that is, using only a single gene per missing fragment. Points correspond to strains and are colored according to the genetic clusters identified on the dendrogram (D). (*C*) Principal component analysis of 216, 878 genome-wide SNPs for MvSl. Points correspond to strains and are colored according to the genetic clusters identified on the dendrogram (D). (*D*) Neighbor-joining tree representing the genetic distance between the 38 MvSl strains on the basis of gene presence–absence polymorphism, based on 124 unique gene presence–absence polymorphisms. Points correspond to strains and are colored according to the genetic clusters identified in the dendrogram. The four genetic clusters (Northwestern, Southwestern, Eastern, and Southern) described by (Badouin et al. 2017) on the basis of single-nucleotide polymorphisms (SNPs) are indicated. The strains sequenced by (Whittle et al. 2015) are indicated by stars. The MvSl-140-1 strain originating from the MvSl Northwestern cluster and the MvSl-1005 strain originating from the MvSl Southwestern cluster that were de novo assembled are indicated. The scale bar indicates the number of differences between individuals. (*E*) Geographic origin of the MvSd strains. The sampling locations for MvSd strains sequenced with Illumina technology are indicated by circles. Colors indicate position on a geographic east–west gradient. The sampling location for the reference genome used for mapping is indicated by a square. (*F*) Principal component analysis of gene presence–absence polymorphism for MvSd, based on 35 unique gene presence–absence polymorphisms. Points correspond to strains, and are colored according to a geographic east–west gradient, as shown on the map (H). (*G*) Principal component analysis of genome-wide 33, 694 SNPs for MvSd. Points correspond to strains and are colored according to the genetic clusters identified on the dendrogram (D). (*H*) Neighbor-joining tree representing the genetic distance between the 19 MvSd strains on the basis of gene presence–absence polymorphism, based on 35 unique gene presence–absence polymorphisms. The scale bar indicates the number of differences between individuals. Points correspond to strains and are colored according to position on a geographic east–west gradient, as shown on the map.

these two groups for gene presence–absence polymorphisms was higher than expected by chance, lying within the 1% extreme values in 1000 permutations.

We investigated the possible presence of gene presence–absence polymorphisms in regions previously identified to have been affected by a recent selective sweep in the MvSl Northwestern genetic cluster (Badouin et al. 2017). The MvSl Northwestern genetic cluster displayed no enrichment of gene presence–absence polymorphisms in selective sweep regions (Pearson's chi-squared test $\chi^2 = 0.24379$, $P = 0.6215$). Nevertheless, 21 of the 93 genes absent from

at least one MvSl Northwestern strain were located within a selective sweep region, including three genes encoding secreted proteins (fig. 5B and C; supplementary table S3, Supplementary Material online). In selective sweep regions, seven gene absence were rare (frequency < 5%) and three gene absence events were almost fixed (frequency > 70%), as expected for selective sweeps (supplementary fig. S4, Supplementary Material online). Some of the gene presence–absence events, particularly those concerning genes encoding secreted proteins, may thus be adaptive, although functional studies are needed to confirm this.
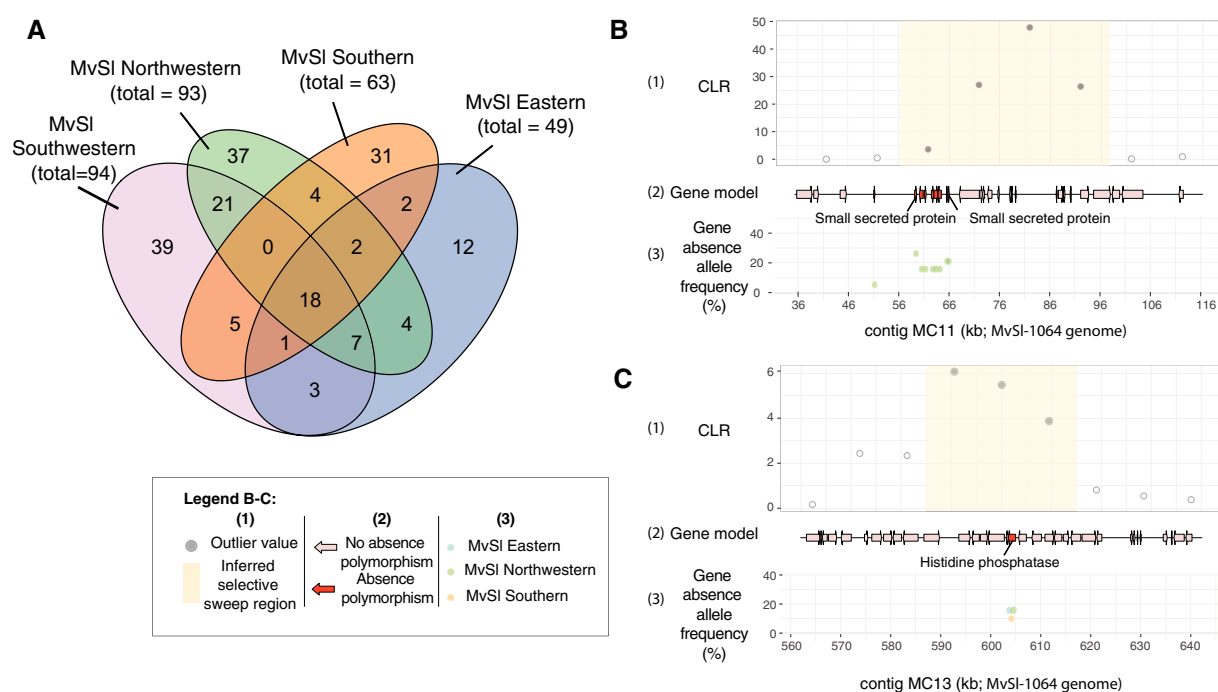
FIG. 5.—Gene presence–absence polymorphism within individual genetic clusters in *Microbotryum lychnidis-dioicae* (MvSl) and localization within previously identified selective sweep regions. (*A*) Venn diagram for shared and unique gene presence–absence polymorphism events in the four MvSl genetic clusters. Genetic clusters were defined on the basis of dendrograms (fig. 4*D*). (*B*) Localization of a gene presence–absence polymorphism within a selective sweep previously identified by (Badouin et al. 2017) in the Northwestern genetic cluster for MvSl. The region with coordinates 36–116 kb on the MvSl-1064 MC11 contig is represented, corresponding to the MvSlA1A2r3c_S01: 2964344-3005424 selective sweep region described by (Badouin et al. 2017); (*C*) Localization of a gene presence–absence polymorphism within a selective sweep previously identified by (Badouin et al. 2017) in the Northwestern genetic cluster for MvSl. The region with coordinates 560–640 kb on the MvSl-1064 MC13 contig is represented, corresponding to the MvSlA1A2r3c_S11: 445169-475073 selective sweep region described by (Badouin et al. 2017). For (*B*) and (*C*), the following are shown in each panel, from top to bottom: (1) The composite likelihood ratio (CLR) from a SweeD analysis with outlier values in light blue and the inferred selective sweep region in yellow. All data were retrieved from (Badouin et al. 2017); (2) Gene models for the MvSl-1064 strain, for genes displaying presence–absence polymorphism in the Northwestern cluster, shown in red; (3) Frequency of gene absence in the Northwestern genetic cluster (green), Eastern genetic cluster (blue), and Southern genetic cluster (orange) for MvS1.

For identifying genes absent from the MvSl-1064 genome reference but present in the genomes of other MvSl strains, we generated a high-quality genome assembly for the MvSl-1318 strain from the MvSl Eastern cluster and we assembled de novo the Illumina genomes of the MvSl-140-01 strain from the MvSl Northwestern cluster, and of the MvSl-1005 strain from the MvSl Southwestern cluster (fig. 4*A*; supplementary text S2, Supplementary Material online). Using our gene presence–absence detection pipeline, we showed that the MvSl-1318 genome contained 11 genes absent from the MvSl-1064 genome (fig. 6*A*). We identified nine genes in the MvSl-140-01 genome and 15 in the MvSl-1005 genome that were absent from the MvSl-1064 genome. The number of strain-specific genes identified in the three de novo assembled genomes (i.e., absent from the MvSl-1064 genome) was consistent with the number of genes present in the MvSl-1064 genome and absent from other MvSl strains (supplementary fig. S2*D*, Supplementary Material online). The population structure of the presence–absence polymorphism

of strain-specific genes in MvSl confirmed the MvSl phylogeographic structure corresponding to the plant local adaptation (fig. 6*C* and *D*).

Discussion

Gene Content Variation Affects Recently Gained Genes in Anther-Smut Fungi

Using a robust gene presence–absence calling procedure based on two different read mapping methods, we characterized the degree of gene content variation on autosomes of the fungal anther-smut species MvSl and MvSd. Autosomal chromosomes are highly homozygous and have a total assembly size of ~30 Mb in both species (Badouin 2015; Branco et al. 2017). The read depth-based and split read-based methods were complementary, as they were based on different read mapping approaches and detected different gene presence–absence polymorphism events. We detected
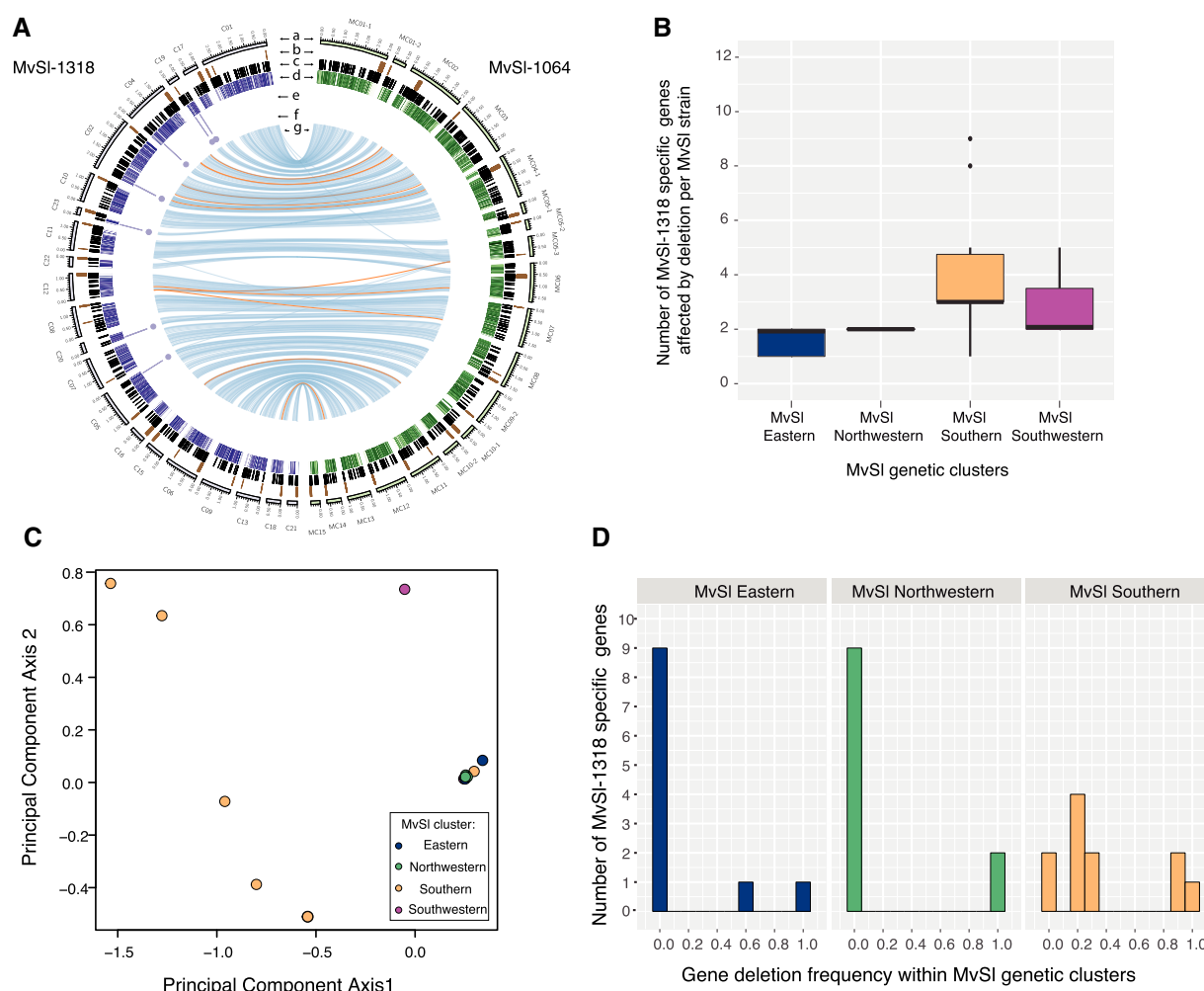
Fig. 6.—Identification of strain-specific genes in three *Microbotryum lychnidis-dioicae* (MvSl) de novo assembled genomes and effect of population structure on gene presence–absence polymorphism. (*A*) Comparison between high-quality genome assemblies of the MvSl strains MvSl-1064 and MvSl-1318. The different tracks are: (a) Contigs of the MvSl-1318 (left, purple tracks) and MvSl-1064 reference genomes (right, green tracks). Only contigs >320 kb were represented; (b) Location of centromeric repeats; (c) Location of transposable elements; (d) Gene density in 10-kb nonoverlapping windows (the gradient shows differences from 0% to 100%); (e) Location of detected missing fragments in MvSl-1064; (f) Location of missing genes in MvSl-1064, that is, MvSl-1318-specific genes; (g) Links representing collinearity of genomic regions >10 kb between the MvSl-1064 and MvSl-1318 reference genomes, with the orange links corresponding to inversions. (*B*) Distribution of the number of MvSl-1318-specific genes predicted to be absent per MvSl strain and per genetic cluster. (*C*) Principal component analysis of gene presence–absence polymorphism of MvSl-1318-specific genes in MvSl strains based on six unique gene presence–absence polymorphisms, that is, using only a single gene per missing fragments. Points correspond to strains and are colored according to the genetic clusters identified by the dendrogram (fig. 4*D*). (*D*) Distribution of gene absence frequency of MvSl-1318-specific genes in the four MvSl genetic clusters.

two to 50 missing genes per strain relative to the reference genome for the species concerned. We performed de novo genome assemblies for three additional MvSl strains to validate the degree of intraspecific gene content variation between two strains, as we detected about a dozen genes that were absent from the MvSl reference genome but present in other strains. Our results supported the view that one single reference genome underestimates the gene space of fungal species (Peter et al. 2018; Plissonneau et al. 2018). The level of gene presence–absence polymorphism in anther-smut pathogens and the sizes of the missing DNA fragments were consistent with findings for yeasts. In *Saccharomyces*

cerevisiae, gene content variation concerns a median of 31 genes per strain (Bergström et al. 2014). We found that gene presence–absence polymorphism contributed to the genetic variation of anther-smut fungi, consistent with the view that gene gains and losses are important sources of fungal genetic variation (Taylor et al. 2017; Fouché et al. 2018).

The functional characteristics of the genes displaying presence–absence polymorphism were similar in MvSl and MvSd: mostly genes of unknown function, with low expression levels, small sizes and enriched in genes from multiple-copy gene families. These functional characteristics suggest that the genes displaying presence–absence polymorphism do not

have housekeeping functions and have been gained recently. Our finding that a large proportion of the genes displaying presence–absence polymorphism belong to multiple-gene families suggests that these genes originate from recent duplication events in multiple-gene families. Following a duplication event, a gene is gained and one of the copies can diverge extensively and acquires a new function (Tautz and Domazet-Lošo 2011). Short length and low level of expression are also common features of genes that have recently arisen de novo, as reported for *Zymoseptoria tritici*, a fungal pathogen of wheat (Plissonneau et al. 2016; Hartmann and Croll 2017). A search for orthologs in outgroup species also suggested that most of the presence–absence polymorphisms observed probably resulted from recent gene gains in MvSl or MvSd. New genes may arise from noncoding DNA through the spontaneous evolution of an open reading frame and the gain of *cis*-regulatory elements (Carvunis et al. 2012; McLysaght and Guerzoni 2015), or through horizontal gene transfer (Marcet-Houben and Gabaldón 2010; Slot and Rokas 2011; Ropars et al. 2015), We found that gene presence–absence polymorphism was particularly prevalent in subtelomeric regions and close to repeats, as previously reported for yeasts and for the plant pathogens *M. oryzae* and *Z. tritici* (Gallone et al. 2016; Plissonneau et al. 2016; Steenwyk et al. 2016; Yoshida et al. 2016), consistent with the view that duplication and excision events may be driven by repeated elements. The high content of repeats found in subtelomeres and centromeres (fig. 1B–D) might at least partly explain the enrichment of gene presence–absence polymorphism events in these genomic regions.

Gene Presence–Absence Polymorphism and Host–Pathogen Coevolution in Natural Environments

Most of the gene presence–absence polymorphism detected is likely neutral. Indeed, previously identified selective sweep regions were not enriched in gene presence–absence events. In addition, gene absence alleles were skewed toward low-frequency variants in the two anther-smut fungal species. Neutral evolution for gene presence–absence polymorphism is also consistent with previous findings in other plant fungal pathogens. Whole-genome sequence comparisons between *Magnaporthe oryzae* strains specific to rice, millet, wheat, or oat revealed for instance that most of the genes displaying presence–absence polymorphism were genes whose functional domains were present multiple times in the genome, that is, genes likely displaying high levels of functional redundancy (Yoshida et al. 2016). In the wheat pathogen *Z. tritici*, mainly weak negative selection and neutrality have been shown to act on presence–absence polymorphism of recently gained genes, but also divergent selection in some cases (Hartmann and Croll 2017). In the anther-smut pathogens MvSl and MvSd, we also found that a few of the recently gained genes may be adaptive, with predicted functions potentially involved in host–pathogen interactions, such as secreted proteins, and/or being located within regions subject to recent selective sweeps. The enrichment in functions linked to cellular stress responses of genes displaying presence–absence polymorphism may also reflect adaptation to specific environments. The loss of genes encoding proteins involved in RNA interference activity might also be adaptive, as some plant pathogens use the RNA interference machinery to prevent the expression of host immunity genes (Weiberg and Jin 2015), but this has not been investigated in anther-smut fungi. We found that different genomic regions displayed gene presence–absence polymorphism in MvSl and MvSd specialized on different host plants, with differences in the affected secreted protein-encoding genes, in particular. The genes involved in the coevolution of anther-smut fungi with their host plants are unknown (Badouin et al. 2017), and the genes displaying presence–absence polymorphism within MvSl and MvSd populations and corresponding to secreted proteins and/or located within selective sweeps represent potentially interesting candidates for future exploration. Functional validation assays using gene transformation tools recently established for anther-smut fungi (Toh et al. 2016) should be performed in future studies to validate an adaptive role of the gene presence–absence events with relevant functions of genomic regions.

We identified only five genes encoding secreted proteins displaying presence–absence polymorphism in MvSl, and three in MvSd. These numbers correspond to lower percentages of total gene content than reported for several fungal crop pathogens, such as the rice BLAST pathogen *M. oryzae*, the wheat pathogens *Z. tritici* and *Stagonospora nodorum*, and the generalist crop pathogens *Verticillium dahliae* and *Fusarium oxysporum* (Yoshida et al. 2009; Ma et al. 2010; Jonge et al. 2013; Syme et al. 2013; Plissonneau et al. 2016). Many gene-for-gene relationships have been documented in crop-pathogen systems, several of which corresponded to recent losses or gains of genes in the fungus, to prevent host recognition by the plant defense system (Orbach et al. 2000; Gout et al. 2006, 2007; Jonge et al. 2012; Hartmann et al. 2017; Fouché et al. 2018). By contrast, no gene-for-gene relationship has been reported for anther-smut fungi, despite extensive studies on genetic variation in host resistance and fungal pathogenicity. Instead, the probability of infection is a quantitative trait (Alexander et al. 1993; Alexander and Antonovics 1995; Biere and Antonovics 1996; Chung et al. 2012), which likely corresponds to a control of host recognition by other genetic changes than gene loss or gain. Differences in coevolutionary dynamics between anthropized and natural pathosystems, such as the "arms race" and "trench warfare" evolution models (Brown and Tellier 2011) may also result in different mechanisms of adaptation to the host. More studies of this type are required, on noncrop pathogens in particular, to test hypotheses concerning the evolutionary causes of structural variation frequencies affecting genes involved in coevolutionary processes.

Diversity and Phylogeographic Structure

Gene presence–absence polymorphisms were more frequent in MvSl than in MvSd (2% vs 0.5% of all autosomal genes). The degree of gene presence–absence polymorphism in MvSd may be slightly underestimated, due to the lower reference genome quality for MvSd than for MvSl and the lower number of sequenced strains (supplementary table S2, Supplementary Material online). However, this is unlikely to account for the magnitude of the difference observed, which was consistent with those reported for SNPs and microsatellites (Vercken et al. 2010; Badouin et al. 2017).

The analysis of gene presence–absence polymorphism in MvSl revealed a population structure highly similar to that reported on the basis of SNPs and microsatellite data, with the same four genetic clusters corresponding to footprints of glacial refugia (Vercken et al. 2010; Badouin et al. 2017). The addition of MvSl strains (Whittle et al. 2015) compared to previous phylogeographic studies revealed very high levels of genetic diversity among the strains of the Southwestern cluster, as expected for strains derived from southern glacial refugia, which have been little explored to date (Vercken et al. 2010; Badouin et al. 2017). The correspondence between the structures revealed by analyses of neutral markers and those revealed by analyses of gene presence–absence polymorphism further reinforce the view that most of such polymorphism is neutral. In particular, gene gain or loss within gene families may be neutral due to functional redundancy (Albalat and Cañestro 2016). However, the host plant S. latifolia also displays the same genetic subdivision in Europe and local adaptation between clusters: plants from a given cluster are more resistant to MvSl strains from the same cluster (Feurtey et al. 2016). Some of the gene presence–absence polymorphism events that are congruent with these clusters may, therefore, also correspond to adaptive polymorphism resulting from coevolution with the host plant. Interestingly, this analysis of gene presence–absence polymorphism suggested the existence of a geographical population structure in MvSd, across an east–west gradient in Europe. Although this population structure had not been detected before, our analysis of genome-wide SNP data called against an MvSd reference genome partly recovered a similar population structure. However, PCAs and F_{ST} values indicated a stronger structure in gene presence–absence polymorphism than in SNPs in MvSd, which may reflect distinct population genetic properties of genetic markers in populations, such as mutation rate, or may result from an effect of selection on some of the gene gains/losses. The population structure based on gene presence–absence polymorphism in MvSd may thus reflect, at least partly, adaptive events, possibly in response to a population subdivision of the host in terms of resistance genes. Indeed, a population subdivision of the host plant species S. dioica separates the Western and Eastern populations (Hathaway et al. 2009; Rautenberg et al. 2010). However,

here again, functional validation studies are needed to test this hypothesis.

In conclusion, our findings show that gene presence–absence polymorphism contributes to intraspecific genetic variation in anther-smut fungi, although to a lower extent in terms of total gene content than previously reported for fungal crop pathogens. Few genes encoding secreted proteins showed presence–absence polymorphisms, as expected given the lack of gene-for-gene relationships in anther-smut fungi. Gene presence–absence polymorphism mostly affected recently gained genes, found in a single species, which were likely mostly neutral, with a few cases perhaps corresponding to innovations allowing adaptation to the host or interaction with the environment, which should be validated using functional studies in the near future.

Supplementary Material

Supplementary data are available at *Genome Biology and Evolution* online.

Acknowledgments

Literature cited

Abyzov A, Urban AE, Snyder M, Gerstein M. 2011. CNVnator: an approach to discover, genotype, and characterize typical and atypical CNVs from family and population genome sequencing. Genome Res. 21(6):974–984.

Albalat R, Cañestro C. 2016. Evolution by gene loss. Nat Rev Genet. 17(7):379–391.

Albataineh MT, Kadosh D. 2016. Regulatory roles of phosphorylation in model and pathogenic fungi. Med Mycol. 54(4):333–352.

Alexander HM, Antonovics J. 1995. Spread of anther-smut disease (Ustilago violacea) and character correlations in a genetically variable experimental population of Silene alba. J Ecol. 83(5):783–794.

Alexander HM, Antonovics J, Kelly AW. 1993. Genotypic variation in plant disease resistance–physiological resistance in relation to field disease transmission. J Ecol. 81(2):325–333.

Badouin H. 2015. Chaos of rearrangements in the mating-type chromosomes of the anther-smut fungus Microbotryum lychnidis-dioicae. Genetics 200(4):1275–1284.

Badouin H, et al. 2017. Widespread selective sweeps throughout the genome of model plant pathogenic fungi and identification of effector candidates. Mol Ecol. 26(7):2041–2062.

Bankevich A, et al. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol. 19(5):455–477.

Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc B Methodol. 57:289–300.

Bergström A, et al. 2014. A high-definition view of functional genetic variation from natural yeast genomes. Mol Biol Evol. 31(4):872–888.

Biere A, Antonovics J. 1996. Sex-specific costs of resistance to the fungal pathogen Ustilago violacea (Microbotryum violaceum) in Silene alba. Evolution 50(3):1098–1110.

Branco S, et al. 2017. Evolutionary strata on young mating-type chromosomes despite the lack of sexual antagonism. Proc Natl Acad Sci U S A. 114(27):7067–7072.

Brown JKM, Tellier A. 2011. Plant-parasite coevolution: bridging the gap between genetics and ecology. Annu Rev Phytopathol. 49(1):345–367.

Carvunis A-R, et al. 2012. Proto-genes and de novo gene birth. Nature 487(7407):370–374.

Chikhi R, Medvedev P. 2014. Informed and automated k-mer size selection for genome assembly. Bioinformatics 30(1):31–37.

Chung E, Petit E, Antonovics J, Pedersen AB, Hood ME. 2012. Variation in resistance to multiple pathogen species: anther smuts of Silene uniflora. Ecol Evol. 2(9):2304–2314.

Conrad DF, Andrews TD, Carter NP, Hurles ME, Pritchard JK. 2006. A high-resolution survey of deletion polymorphism in the human genome. Nat Genet. 38(1):75–81.

Conrad DF, et al. 2010. Origins and functional impact of copy number variation in the human genome. Nature 464(7289):704–712.

Ekseth OK, Kuiper M, Mironov V. 2014. orthAgogue: an agile tool for the rapid prediction of orthology relations. Bioinformatics 30(5):734–736.

Ellinghaus D, Kurtz S, Willhoeft U. 2008. LTRharvest, an efficient and flexible software for de novo detection of LTR retrotransposons. BMC Bioinformatics 9(1):18.

Feurtey A, et al. 2016. Strong phylogeographic co-structure between the anther-smut fungus and its white campion host. New Phytol. 212(3):668–679.

Fontanillas E, et al. 2014. Degeneration of the non-recombining regions in the mating-type chromosomes of the anther-smut fungi. Mol Biol Evol. 32(4):928–43.

Fouché S, Plissonneau C, Croll D. 2018. The birth and death of effectors in rapidly evolving filamentous pathogen genomes. Curr Opin Microbiol. 46:34–42.

Friesen TL, et al. 2006. Emergence of a new disease as a result of inter-specific virulence gene transfer. Nat Genet. 38(8):953–956.

Gallone B, et al. 2016. Domestication and divergence of Saccharomyces cerevisiae beer yeasts. Cell 166(6):1397–1410.e16.

Giraud T, Yockteng R, Lopez-Villavicencio M, Refregier G, Hood ME. 2008. Mating system of the anther smut fungus Microbotryum violaceum: selfing under heterothallism. Eukaryot Cell 7(5):765–775.

Goudet J. 2005. hierfstat, a package for r to compute and test hierarchical F-statistics. Mol Ecol Notes 5(1):184–186.

Gout L, et al. 2006. Lost in the middle of nowhere: the AvrLm1 avirulence gene of the Dothideomycete Leptosphaeria maculans. Mol Microbiol. 60(1):67–80.

Gout L, et al. 2007. Genome structure impacts molecular evolution at the AvrLm1 avirulence locus of the plant pathogen Leptosphaeria maculans. Environ Microbiol. 9(12):2978–2992.

Guan P, Sung W-K. 2016. Structural variation detection using next-generation sequencing data: a comparative technical review. Methods 102:36–49.

Hartmann FE, Croll D. 2017. Distinct trajectories of massive recent gene gains and losses in populations of a microbial eukaryotic pathogen. Mol Biol Evol. 34(11):2808–2822.

Hartmann FE, Sánchez-Vallet A, McDonald BA, Croll D. 2017. A fungal wheat pathogen evolved host specialization by extensive chromosomal rearrangements. ISME J. 11(5):1189–1204.

Hathaway L, Malm JU, Prentice HC. 2009. Geographically congruent large-scale patterns of plastid haplotype variation in the European herbs Silene dioica and S. latifolia (Caryophyllaceae). Bot J Linn Soc. 161(2):153–170.

Henrichsen CN, Chaignat E, Reymond A. 2009. Copy number variants, diseases and gene expression. Hum Mol Genet. 18(R1):R1–R8.

Hood ME. 2005. Repetitive DNA in the automictic fungus Microbotryum violaceum. Genetica 124(1):1–10.

Jombart T. 2008. adegenet: a R package for the multivariate analysis of genetic markers. Bioinformatics 24(11):1403–1405.

Jombart T, Ahmed I. 2011. adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. Bioinformatics 27(21):3070–3071.

Jones JDG, Dangl JL. 2006. The plant immune system. Nature 444(7117):323–329.

Jonge R, et al. 2012. Tomato immune receptor Ve1 recognizes effector of multiple fungal pathogens uncovered by genome and RNA sequencing. Proc Natl Acad Sci U S A. 109(13):5110–5115.

Jonge R, et al. 2013. Extensive chromosomal reshuffling drives evolution of virulence in an asexual pathogen. Genome Res. 23(8):1271–1282.

Kaltz O, Gandon S, Michalakis Y, Shykoff JA. 1999. Local maladaptation in the anther-smut fungus Microbotryum violaceum to its host plant Silene latifolia: evidence from a cross-inoculation experiment. Evol Int J Org Evol. 53(2):395–407.

Khang CH, Park S-Y, Lee Y-H, Valent B, Kang S. 2008. Genome organization and evolution of the AVR-Pita avirulence gene family in the Magnaporthe grisea species complex. Mol Plant Microbe Interact. 21(5):658–670.

Koopmann B, Müller J, Tellier A, Živković D. 2017. Fisher–Wright model with deterministic seed bank and selection. Theor Popul Biol. 114:29–39.

Koskela T, Puustinen S, Salonen V, Mutikainen P. 2002. Resistance and tolerance in a host plant-holoparasitic plant interaction: genetic variation and costs. Evol Int J Org Evol. 56:899–908.

Krogh A, Larsson B, von Heijne G, Sonnhammer ELL. 2001. Predicting transmembrane protein topology with a hidden markov model: application to complete genomes. J Mol Biol. 305(3):567–580.

Krzywinski M, et al. 2009. Circos: an information aesthetic for comparative genomics. Genome Res. 19(9):1639–1645.

Kurtz S, et al. 2004. Versatile and open software for comparing large genomes. Genome Biol. 5:R12.

Laine A-L, Burdon JJ, Dodds PN, Thrall PH. 2011. Spatial variation in disease resistance: from molecules to metapopulations. J Ecol. 99(1):96–112.

Laine A-L, Burdon JJ, Nemri A, Thrall PH. 2014. Host ecotype generates evolutionary and epidemiological divergence across a pathogen metapopulation. Proc R Soc B 281(1787):20140522.

Laine A-L, Tellier A. 2008. Heterogeneous selection promotes maintenance of polymorphism in host–parasite interactions. Oikos 117(9):1281–1288.

Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 10(3):R25.

Le Gac M, Hood ME, Fournier E, Giraud T. 2007. Phylogenetic evidence of host-specific cryptic species in the anther smut fungus. Evol Int J Org Evol. 61(1):15–26.

Li H, et al. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25(16):2078–2079.

Liao Y, Smyth GK, Shi W. 2014. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics 30(7):923–930.

Lin K, Smit S, Bonnema G, Sanchez-Perez G, de Ridder D. 2015. Making the difference: integrating structural variation detection tools. Brief Bioinform. 16(5):852–864.

Ma L-J, et al. 2010. Comparative genomics reveals mobile pathogenicity chromosomes in *Fusarium*. Nature 464(7287):367–373.

Marcet-Houben M, Gabaldón T. 2010. Acquisition of prokaryotic genes by fungal genomes. Trends Genet. 26(1):5–8.

Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal 17(1):10–12.

McKenna A, et al. 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 20(9):1297–1303.

McLysaght A, Guerzoni D. 2015. New genes from non-coding sequence: the role of de novo protein-coding genes in eukaryotic evolutionary innovation. Philos Trans R Soc B 370(1678):20140332.

Möller M, Stukenbrock EH. 2017. Evolution and genome architecture in fungal plant pathogens. Nat Rev Microbiol. 15(12):756–771.

Monod M, et al. 2002. Secreted proteases from pathogenic fungi. Int J Med Microbiol. 292(5-6):405–419.

Orbach MJ, Farrall L, Sweigard JA, Chumley FG, Valent B. 2000. A telomeric avirulence gene determines efficacy for the rice blast resistance gene *Pi-ta*. Plant Cell 12(11):2019–2032.

Orozco LD, et al. 2009. Copy number variation influences gene expression and metabolic traits in mice. Hum Mol Genet. 18(21):4118–4129.

Paradis E, Claude J, Strimmer K. 2004. APE: analyses of phylogenetics and evolution in R language. Bioinformatics 20(2):289–290.

Perlin MH, et al. 2015. Sex and parasites: genomic and transcriptomic analysis of *Microbotryum lychnidis-dioicae*, the biotrophic and plant-castrating anther smut fungus. BMC Genomics 16(1):461.

Peter J, et al. 2018. Genome evolution across 1, 011 *Saccharomyces cerevisiae* isolates. Nature 556:339–344.

Petersen TN, Brunak S, von Heijne G, Nielsen H. 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat Methods 8(10):785–786.

Plissonneau C, Hartmann FE, Croll D. 2018. Pangenome analyses of the wheat pathogen *Zymoseptoria tritici* reveal the structural basis of a highly plastic eukaryotic genome. BMC Biol. 16(1):5.

Plissonneau C, Stürchler A, Croll D. 2016. The evolution of orphan regions in genomes of a fungal pathogen of wheat. mBio 7(5):e01231-16.

Presti LL, et al. 2015. Fungal effectors and plant susceptibility. Annu Rev Plant Biol. 66(1):513–545.

Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26(6):841–842.

Rautenberg A, Hathaway L, Oxelman B, Prentice HC. 2010. Geographic and phylogenetic patterns in *Silene* section *Melandrium* (Caryophyllaceae) as inferred from chloroplast and nuclear DNA sequences. Mol Phylogenet Evol. 57(3):978–991.

Refrégier G, et al. 2008. Cophylogeny of the anther smut fungi and their caryophyllaceous hosts: prevalence of host shifts and importance of delimiting parasite species for inferring cospeciation. BMC Evol Biol. 8(1):100.

Robinson JT, et al. 2011. Integrative genomics viewer. Nat Biotechnol. 29(1):24–26.

Ropars J, et al. 2015. Adaptive horizontal gene transfers between multiple cheese-associated fungi. Curr Biol. 25(19):2562–2569.

Rozen S, Skaletsky H. 2000. Primer3 on the WWW for general users and for biologist programmers. Methods Mol Biol. 132:365–386.

Schrider DR, Hahn MW. 2010. Gene copy-number polymorphism in nature. Proc R Soc Lond B Biol Sci. 277(1698):3213–3221.

Schürch S, Linde CC, Knogge W, Jackson LF, McDonald BA. 2004. Molecular population genetic analysis differentiates two virulence mechanisms of the fungal avirulence gene *NIP1*. Mol Plant Microbe Interact. 17(10):1114–1125.

Slot JC, Rokas A. 2011. Horizontal transfer of a large and highly toxic secondary metabolic gene cluster between fungi. Curr Biol. 21(2):134–139.

Smit AFA, Hubley R. 2008. RepeatModeler Open-1.0 [cited 2017 Nov 30]. Available from: http://www.repeatmasker.org; last accessed November 30, 2017.

Smit AFA, Hubley R, Green P. 2013. RepeatMasker Open-4.0 [cited 2017 Nov 30]. Available from: http://www.repeatmasker.org; last accessed November 30, 2017.

Stam R, Scheikl D, Tellier A. 2017. The wild tomato species *Solanum chilense* shows variation in pathogen resistance between geographically distinct populations. PeerJ 5:e2910.

Steenwyk JL, Soghigian JS, Perfect JR, Gibbons JG. 2016. Copy number variation contributes to cryptic genetic variation in outbreak lineages of *Cryptococcus gattii* from the North American Pacific Northwest. BMC Genomics 17(1):700.

Syme RA, Hane JK, Friesen TL, Oliver RP. 2013. Resequencing and comparative genomics of *Stagonospora nodorum*: sectional gene absence and effector discovery. G3 (Bethesda) 3:959–969.

Tack AJM, Laine A-L. 2014. Ecological and evolutionary implications of spatial heterogeneity during the off-season for a wild plant pathogen. New Phytol. 202(1):297–308.

Tautz D, Domazet-Lošo T. 2011. The evolutionary origin of orphan genes. Nat Rev Genet. 12(10):692.

Taylor JW, et al. 2017. Sources of fungal genetic variation and associating It with phenotypic diversity. Microbiol Spectr 5(5): doi: 10.1128/microbiolspec.FUNK-0057-2016.

Thorvaldsdóttir H, Robinson JT, Mesirov JP. 2013. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Brief Bioinform. 14(2):178–192.

Thrall PH, Burdon JJ, Young A. 2001. Variation in resistance and virulence among demes of a plant host–pathogen metapopulation. J Ecol. 89(5):736–748.

Toh SS, Treves DS, Barati MT, Perlin MH. 2016. Reliable transformation system for *Microbotryum lychnidis dioicae* informed by genome and transcriptome project. Arch Microbiol. 198(8):813–825.

Trapnell C, Pachter L, Salzberg SL. 2009. TopHat: discovering splice junctions with RNA-Seq. Bioinformatics 25(9):1105–1111.

Vercken E, et al. 2010. Glacial refugia in pathogens: european genetic structure of anther smut pathogens on *Silene latifolia* and *Silene dioica*. PLoS Pathog. 6(12):e1001229.

Vincent D, et al. 2016. Editorial: how can secretomics help unravel the secrets of plant-microbe interactions? Front Plant Sci. 7:1777.

Weiberg A, Jin H. 2015. Small RNAs—the secret agents in the plant–pathogen interactions. Curr Opin Plant Biol. 26:87–94.

Whittle CA, Votintseva A, Ridout K, Filatov DA. 2015. Recent and massive expansion of the mating-type-specific region in the smut fungus *Microbotryum*. Genetics 199(3):809–816.

Wickham H. 2009. ggplot2: elegant graphics for data analysis. New York, NY: Springer Science & Business Media.

Yang R-C. 1998. Estimating hierarchical F-Statistics. Evolution 52(4):950–956.

Ye K, Schulz MH, Long Q, Apweiler R, Ning Z. 2009. Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. Bioinformatics 25(21):2865–2871.

Yoshida K, et al. 2009. Association genetics reveals three novel avirulence genes from the rice blast fungal pathogen *Magnaporthe oryzae*. Plant Cell 21(5):1573–1591.

Yoshida K, et al. 2016. Host specialization of the blast fungus *Magnaporthe oryzae* is associated with dynamic gain and loss of genes linked to transposable elements. BMC Genomics 17(1):370.

Zdobnov EM, Apweiler R. 2001. InterProScan – an integration platform for the signature-recognition methods in InterPro. Bioinformatics 17(9):847–848.

Associate editor: Yves Van De Peer

Supplementary Material

Gene presence-absence polymorphism in castrating anther-smut fungi: recent gene gains and phylogeographic structure

Fanny E. Hartmann[1]*, Ricardo C. Rodríguez de la Vega[1], Jean-Tristan Brandenburg[1], Fantin Carpentier[1], Tatiana Giraud[1]

[1] Ecologie Systématique Evolution, Bâtiment 360, Univ. Paris-Sud, AgroParisTech, CNRS, Université Paris-Saclay, 91400 Orsay, France

* Author for Correspondence: Fanny E. Hartmann, Ecologie Systématique Evolution, Bâtiment 360, Univ. Paris-Sud, AgroParisTech, CNRS, Université Paris-Saclay, 91400 Orsay, France, telephone number : 01 69 15 72 82, e-mail address: fanny.hartmann@u-psud.fr

Supplementary Text

Supplementary Text S1: Quality control procedure of detected gene presence-absence polymorphism using a comparison of high-quality genome assemblies of two *Microbotryum lychnidis-dioicae* (MvSl) strains.

We found a large overlap between the gene absence events detected using our gene presence-absence detection pipeline, and those detected using a comparison of high-quality genome assemblies of two MvSl strains. For this comparison, we assembled here the single-molecule real-time sequenced genome of the MvSl-1318 strain using the same method as for the reference MvSl-1064 genome (see Supplemental Material and Method in (Branco et al. 2017)). The two genome assemblies were highly comparable in total size, N50 and N90 values, number of gene models and percentage of repetitive elements (Supplementary Table S2). Global synteny analyses of the two assemblies showed that the autosomal contigs were highly collinear (Figure 6B). We mapped the MvSl-1064 genome resequencing Illumina data to the MvSl-1318 reference genome. Using our gene presence-absence detection pipeline, we detected 15 genes predicted in the MvSl-1318 reference genome to be missing in the MvSl-1064 genome both with the read depth-based and split read-based methods. We used the sequences of the 15 genes of MvSl-1318 genome as a query in blastn to check the presence of significant blast hits in the MvSl-1064 genome. Out of these 15 missing genes, we did not obtain any significant blast hits for nine of them, supporting the prediction of absence of these genes. For three genes, we obtained a significant single blast hit for each, and we considered these genes as false positive of our detection pipeline. For the three remaining genes, we obtained multiple blast hits, on both orthologous and non-orthologous contigs. We validated the lack of two of these genes by checking the read mapping in the genes regions the genome browser IGV (Robinson et al. 2011; Thorvaldsdóttir et al. 2013). The false positive rate of detected gene absence events with the read depth-based and split read-based

method was therefore 26% (4/15), close to the range of the false positive rate originally claimed for the methods (5-20%; (Abyzov et al. 2011)). Finally, we found no significant effect of the strain ploidy (strains being sequenced as either haploid or diploid), the source of genome data ((Badouin et al. 2017) or (Whittle et al. 2015)) or the number of mapped reads on the number of gene absence events called (Supplementary figure S3).

Supplementary Text S2: Detection of strain-specific genes and study of their population structure.

To further investigate gene presence-absence polymorphism within the species, we aimed at identifying the genes that were absent in the MvSl-1064 genome (used as reference for MvSl) but present in genomes of other MvSl strains, hereafter referred to as strain-specific genes. We first detected strain-specific genes in the MvSl-1318 strain originating from the MvSl Eastern cluster (Figure 4A). We sequenced and performed a high-quality *de novo* assembly of the MvSl-1318 genome. Assembly statistics of the MvSl-1318 genome were highly similar to those of the MvSl-1064 genome (N50= 1,357,587; Supplementary Table S2). Global synteny analyses of the two assemblies showed that autosomal contigs were highly collinear (Figure 6A). Using our gene presence-absence prediction pipeline and the comparison of the two genome assemblies, we identified 11 genes encoded in the MvSl-1318 genome predicted to be lacking in the MvSl-1064 genome (Figure 6A; for details see Supplementary Text S1). The 11 identified genes were referred hereafter to as MvSl-1318-specific genes. In addition, we assembled *de novo* two Illumina genomes, those of the MvSl-140-01 strain originating from the MvSl Northwestern cluster, and of the MvSl-1005 strain originating from the MvSl Southwestern cluster (Figure 4A). The quality of the two assemblies was lower than those of the MvSl-1318 and MvSl-1064 genomes as expected for Illumina sequencing data (N50= 48,604 kb for MvSl-140-1; N50= 47,824 kb for MvSl-1005; Supplementary Table S2). Using our gene presence-absence prediction pipeline, we identified nine and 15 genes predicted to be present in the MvSl-140-01 and MvSl-1005 genomes and absent in the MvSl-1064 genome. The lower quality of genome assemblies of the MvSl-140-01 and MvSl-1005 strains likely led us to under-estimate the number of strain-specific genes compared to the MvSl-1318 strain. However, the numbers of strain-specific genes identified in the three *de novo* assembled genomes were consistent with the number of genes of the MvSl-1064 genome

found to be absent in MvSl strains (Supplementary figure S2D). The majority of identified strain-specific genes (73%) was predicted to encode proteins without any known functional domain. One MvSl-1318-specific gene was predicted to encode a secreted protein with unknown functional domain. Strain-specific genes shared similar characteristics than genes affected by presence-absence polymorphism in MvSl (see Results section in the main text).

We studied the population structure of gene presence-absence polymorphism of MvSl-1318-specific genes in MvSl. Using our gene presence-absence prediction pipeline with the MvSl-1318 genome as a reference genome, we found that 9 out of 11 (82%) MvSl-1318-specific genes showed presence-absence polymorphisms in MvSl strains. Gene absence frequency was low in MvSl (median value of 8 %). On average three MvSl-1318-specific genes were absent per strain (Figure 6B-D). The MvSl Southern genetic cluster showed the highest number of deleted MvSl-1318-specific genes. This was expected as the MvSl-1318-specific genes are defined as gene absent in the MvSl-1064 strain that belongs to the Southern genetic cluster. A principal component analysis based on six unlinked gene presence-absence polymorphisms retrieve a population structure consistent with the previously identified four MvSl genetic clusters (Figure 6C; (Badouin et al. 2017)). The majority of MvSl-1318-specific genes were present in at least one strain of the four genetic clusters. One MvSl-1318-specific gene was present only in the Eastern cluster.

Supplementary References

Abyzov A, Urban AE, Snyder M, Gerstein M. 2011. CNVnator: an approach to discover, genotype, and characterize typical and atypical CNVs from family and population genome sequencing. Genome Res. 21:974–984. doi: 10.1101/gr.114876.110.

Badouin H et al. 2017. Widespread selective sweeps throughout the genome of model plant pathogenic fungi and identification of effector candidates. Mol. Ecol. 26:2041–2062. doi: 10.1111/mec.13976.

Branco S et al. 2017. Evolutionary strata on young mating-type chromosomes despite the lack of sexual antagonism. Proc. Natl. Acad. Sci. 114:7067–7072. doi: 10.1073/pnas.1701658114.

Robinson JT et al. 2011. Integrative genomics viewer. Nat. Biotechnol. 29:24–26. doi: 10.1038/nbt.1754.

Thorvaldsdóttir H, Robinson JT, Mesirov JP. 2013. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Brief. Bioinform. 14:178–192. doi: 10.1093/bib/bbs017.

Whittle CA, Votintseva A, Ridout K, Filatov DA. 2015. Recent and massive expansion of the mating-type-specific region in the smut fungus *Microbotryum*. Genetics. 199:809–816. doi: 10.1534/genetics.114.171702.

Supplementary Table S1: Whole-genome sequencing data of the *Microbotryum lychnidis-dioicae* (MvSl) populations and *Microbotryum silenes-dioicae* (MvSd) populations used in this study.
Information were retrieved from (Badouin et al. 2017) and (Whittle et al. 2015).

Sample ID	Fungal species	Host species	Country	BioProject.ID (1)	Accession.ID	Haploid
MvSd-1030	*M. silenes-dioicae*	*Silene dioica*	Knivholt, Frederikshavn, Denmark	PRJNA295022	SRS1072001	No
MvSd-1034	*M. silenes-dioicae*	*Silene dioica*	Roscoff, France	PRJNA295022	SRS1072002	No
MvSd-1056	*M. silenes-dioicae*	*Silene dioica*	Jedburgh, United Kingdom	PRJNA295022	SRS1072003	No
MvSd-1141	*M. silenes-dioicae*	*Silene dioica*	Italy	PRJNA295022	SRS1072020	Yes
MvSd-336-01	*M. silenes-dioicae*	*Silene dioica*	Morlaix, France	PRJNA295022	SRS1072005	No
MvSd-578-2	*M. silenes-dioicae*	*Silene dioica*	Plougasnou, France	PRJNA295022	SRS1072007	Yes
MvSd-637	*M. silenes-dioicae*	*Silene dioica*	Netherlands	PRJNA295022	SRS1072009	No
MvSd-69-05	*M. silenes-dioicae*	*Silene dioica*	Flafleralp Switzerland	PRJNA295022	SRS1072010	No
MvSd-707-1	*M. silenes-dioicae*	*Silene dioica*	Escault France	PRJNA295022	SRS1072011	No
MvSd-72	*M. silenes-dioicae*	*Silene dioica*	Flafleralp Switzerland	PRJNA295022	SRS1072013	Yes
MvSd-849-7	*M. silenes-dioicae*	*Silene dioica*	Å tajerska, Slovenia	PRJNA295022	SRS1072014	No
MvSd-851-6	*M. silenes-dioicae*	*Silene dioica*	Carpathian Mts., Slovakia	PRJNA295022	SRS1072015	No
MvSd-900-1	*M. silenes-dioicae*	*Silene dioica*	Finse, Norway	PRJNA295022	SRS1072016	No
MvSd-932-2	*M. silenes-dioicae*	*Silene dioica*	France	PRJNA295022	SRS1072017	No
MvSd-937-2	*M. silenes-dioicae*	*Silene dioica*	France	PRJNA295022	SRS1072018	No
MvSd-949-2	*M. silenes-dioicae*	*Silene dioica*	Bristol, United Kingdom	PRJNA295022	SRS1072019	No
MvSd-sp002	*M. silenes-dioicae*	*Silene dioica*	France	PRJNA295022	SRS1072021	No
MvSd-sp003	*M. silenes-dioicae*	*Silene dioica*	France	PRJNA295022	SRS1072022	Yes
MvSd-335-H3	*M. lychnidis-dioicae*	*Silene latifolia*	Vosges, France	PRJNA295022	SRS1072004	Yes
MvSl-00-10	*M. lychnidis-dioicae*	*Silene latifolia*	Italy	PRJNA295022	SRS1072024	Yes
MvSl-100-6	*M. lychnidis-dioicae*	*Silene latifolia*	Italy	PRJNA295022	SRS1072025	No
MvSl-1005	*M. lychnidis-dioicae*	*Silene latifolia*	Portugal	PRJNA295022	SRS1072026	No
MvSl-1040	*M. lychnidis-dioicae*	*Silene latifolia*	Golspie United Kingdom	PRJNA295022	SRS1072027	No
MvSl-1067	*M. lychnidis-dioicae*	*Silene latifolia*	Brittany, France	PRJNA295022	SRS1072030	No
MvSl-1069	*M. lychnidis-dioicae*	*Silene latifolia*	Cotswolds, England	PRJNA269361	SSR1695534	Yes
MvSl-1088	*M. lychnidis-dioicae*	*Silene latifolia*	Lausanne, Switzerland	PRJNA269361	SSR1695572	Yes
MvSl-1089	*M. lychnidis-dioicae*	*Silene latifolia*	Lausanne, Switzerland	PRJNA269361	SSR1695575	Yes
MvSl-1090	*M. lychnidis-dioicae*	*Silene latifolia*	Pyrenees, France	PRJNA269361	SSR1695537	Yes
MvSl-1103	*M. lychnidis-dioicae*	*Silene latifolia*	Woodstock, England	PRJNA269361	SSR1696452	Yes
MvSl-1134	*M. lychnidis-dioicae*	*Silene latifolia*	Ficheux, France	PRJNA295022	SRS1072031	No
MvSl-I00-3	*M. lychnidis-dioicae*	*Silene latifolia*	Italy	PRJNA295022	SRS1072064	Yes
MvSl-1140-3	*M. lychnidis-dioicae*	*Silene latifolia*	Muron, France	PRJNA295022	SRS1072032	No
MvSl-140-01	*M. lychnidis-dioicae*	*Silene latifolia*	Manche Le Pou, France	PRJNA295022	SRS1072033	No
MvSl-141-01	*M. lychnidis-dioicae*	*Silene latifolia*	St Albon, France	PRJNA295022	SRS1072034	No
MvSl-40-01	*M. lychnidis-dioicae*	*Silene latifolia*	Auffargis France	PRJNA295022	SRS1072035	Yes
MvSl-405	*M. lychnidis-dioicae*	*Silene latifolia*	Balsta, Sweden	PRJNA269361	SSR1694361	Yes
MvSl-443-2	*M. lychnidis-dioicae*	*Silene latifolia*	Kraghede Denmark	PRJNA295022	SRS1072037	No
MvSl-446-2	*M. lychnidis-dioicae*	*Silene latifolia*	Tokay, Hungary	PRJNA295022	SRS1072039	Yes
MvSl-462-3	*M. lychnidis-dioicae*	*Silene latifolia*	Charterhouse, United Kingdom	PRJNA295022	SRS1072041	No
MvSl-466-3	*M. lychnidis-dioicae*	*Silene latifolia*	Grignon France	PRJNA295022	SRS1072043	Yes
MvSl-576	*M. lychnidis-dioicae*	*Silene latifolia*	Teschow, Germany	PRJNA295022	SRS1072045	Yes
MvSl-641	*M. lychnidis-dioicae*	*Silene latifolia*	Netherlands	PRJNA295022	SRS1072047	Yes
MvSl-661	*M. lychnidis-dioicae*	*Silene latifolia*	Selommes, France	PRJNA295022	SRS1072050	No
MvSl-699	*M. lychnidis-dioicae*	*Silene latifolia*	Chernobyl, Ukraine	PRJNA295022	SRS1072052	No
MvSl-728-4	*M. lychnidis-dioicae*	*Silene latifolia*	Wiltshire, France	PRJNA295022	SRS1072054	No
MvSl-769	*M. lychnidis-dioicae*	*Silene latifolia*	Galicia, Spain	PRJNA269361	SSR1695550	Yes
MvSl-781	*M. lychnidis-dioicae*	*Silene latifolia*	Sesto Calende Italy	PRJNA295022	SRS1072055	No
MvSl-830-2	*M. lychnidis-dioicae*	*Silene latifolia*	Alps, Italy	PRJNA295022	SRS1072056	No
MvSl-856-2	*M. lychnidis-dioicae*	*Silene latifolia*	Grosley-sur-Risle, France	PRJNA295022	SRS1072057	No
MvSl-920	*M. lychnidis-dioicae*	*Silene latifolia*	Valencia, Spain	PRJNA269361	SSR1695576	Yes
MvSl-925	*M. lychnidis-dioicae*	*Silene latifolia*	Langoyene, Norway	PRJNA295022	SRS1072058	No
MvSl-933-1	*M. lychnidis-dioicae*	*Silene latifolia*	Plominhac, France	PRJNA295022	SRS1072059	No
MvSl-973	*M. lychnidis-dioicae*	*Silene latifolia*	Scotland, United Kindom	PRJNA295022	SRS1072060	No
MvSl-974	*M. lychnidis-dioicae*	*Silene latifolia*	Scotland, United Kindom	PRJNA295022	SRS1072061	No
MvSl-979-4	*M. lychnidis-dioicae*	*Silene latifolia*	Austria	PRJNA295022	SRS1072062	No
MvSl-980	*M. lychnidis-dioicae*	*Silene latifolia*	Austria	PRJNA295022	SRS1072063	No
MvSl-IOA	*M. lychnidis-dioicae*	*Silene latifolia*	Italy	PRJNA295022	SRS1072065	Yes
MvSl-1064	*M. lychnidis-dioicae*	*Silene latifolia*	Lamole, Italy	PRJNA295022	SRS1072029|ERS459551	Yes

(1) Related references :
BioProject PRJNA295022 (Badouin H, Gladieux P, Gouzy J, Siguenza S, Aguileta G, Snirc A, Le Prieur S, Jeziorski C, Branca A, Giraud T, Mol Ecol 26:2041–2062, 2017, doi:10.1111/mec.13976)
BioProject PRJNA269361 (Whittle CA, Votintseva A, Ridout K, Filatov DA, Genetics 199:809–816, 2015, doi: 10.1534/genetics.114.171702)

Supplementary Table S2: Characteristics of the genome assemblies of the *Microbotryum silenes-dioicae* (MvSd) strain MvSd-1303 and the four *Microbotryum lychnidis-dioicae* (MvSl) strains MvSl-1064, MvSl-1318, MvSl-1005, and MvSl-140-01.

Information on the genome assembly of the MvSl-1064 strain were retrieved from (Badouin et al. 2015) and (Branco et al. 2017). Information on the genome assembly of the MvSd-1303 strain were retrieved from (Branco et al. 2017). The strains MvSl-1318, MvSl-1005, and MvSl-140-01 were de novo assembled in this study (see Supplementary Material and Methods for details).

	Sample	MvSl-1064	MvSd-1303	MvSl-1318	MvSl-140-01	MvSl-1005
	Deposit database	European Nucleotide Archive (ENA)	European Nucleotide Archive (ENA)	GenBank	*Not submitted (available upon request)*	*Not submitted (available upon request)*
	Accession numbers	ERS1013679	ERS1436592	PRJNA437556	-	-
	# Contigs (contigs >1kb)	48	144	77	1'312	1'518
	Min length (bp) (contigs >1kb)	12'190	6'283	9'836	1'003	1'001
	Max length (bp)	3'412'169	2'425'171	2'966'845	263'144	263'534
whole genome assembly	N50 (bp)	1'736'850	936'137	1'357'587	48'604	47'824
	L50 (# Contigs)	6	12	8	137	139
	N90 (bp)	648'002	153'077	436'537	7'542	4'557
	L90 (# Contigs)	16	45	22	605	690
	Mean length(bp)	622'941	233'285	397'859	18'015	15'455
	Median length(bp)	44'679	32'003	41'929	6'325	3'715
	Assembly size (bp)	29'901'156	33'593'023	30'635'132	23'993'380	24'088'947
	# Gene models	12'266	13'443	12'451	9'419	9'317
	# Gene models masked for transposable elements	10'010	10'425	10'065		
	# Contigs	16	43	23	449	423
	Min length (bp)	588'402	146'758	317'619	11'561	11'985
90% of autosomes assembly	# Gene models	9'881	10'713	10'062	7'141	7'004
	# Gene models masked for transposable elements	8'523	8'982	8'663	-	-
	# Gene encoding putative secreted proteins masked for transposable elements	343	366	350	-	-

Supplementary Table S4: Significantly over-represented biological process gene ontology (GO) terms for genes affected by presence-absence polymorphism in *Microbotryum lychnidis-dioicae* (MvSl) and *Microbotryum silenes-dioicae* (MvSd).

The "Effective GO term count" column is indicating the number of genes with the given GO term. GO term enrichment pvalues corrected for multiple testing with Bonferroni correction and with Benjamini & Hochberg correction (Benjamini and Hochberg 1995) is presented.

Species	GO Term	Enrichment p-value (not corrected for mutiple testing)	Enrichment p-value (corrected with Bonferroni correction)	Enrichment p-value (corrected with Benjamini & Hochberg correction)	Odds Ratio	Expected GO term count	Effective GO term count	Total proteins per GO	Term
MvSl	GO:0006298	0.00018	0.00092	0.00092	167.62500	0.02226	2	10	mismatch repair
	GO:0006281	0.03090	0.15451	0.03978	9.92636	0.29154	2	131	DNA repair
	GO:0006974	0.03224	0.16121	0.03978	9.68939	0.29822	2	134	cellular response to DNA damage stimulus
	GO:0033554	0.03315	0.16574	0.03978	9.53731	0.30267	2	136	cellular response to stress
	GO:0006950	0.03978	0.19892	0.03978	8.58784	0.33383	2	150	response to stress
MvSd	GO:0009100	0.01369	0.09584	0.02012	152.15789	0.01374	1	20	glycoprotein metabolic process
	GO:0009101	0.01369	0.09584	0.02012	152.15789	0.01374	1	20	glycoprotein biosynthetic process
	GO:0006486	0.01369	0.09584	0.02012	152.15789	0.01374	1	20	protein glycosylation
	GO:0043413	0.01369	0.09584	0.02012	152.15789	0.01374	1	20	macromolecule glycosylation
	GO:0070085	0.01437	0.10061	0.02012	144.50000	0.01442	1	21	glycosylation
	GO:0044710	0.03003	0.21018	0.03503	Inf	0.34684	2	505	single-organism metabolic process
	GO:1901137	0.03877	0.27140	0.03877	50.96429	0.03915	1	57	carbohydrate derivative biosynthetic process
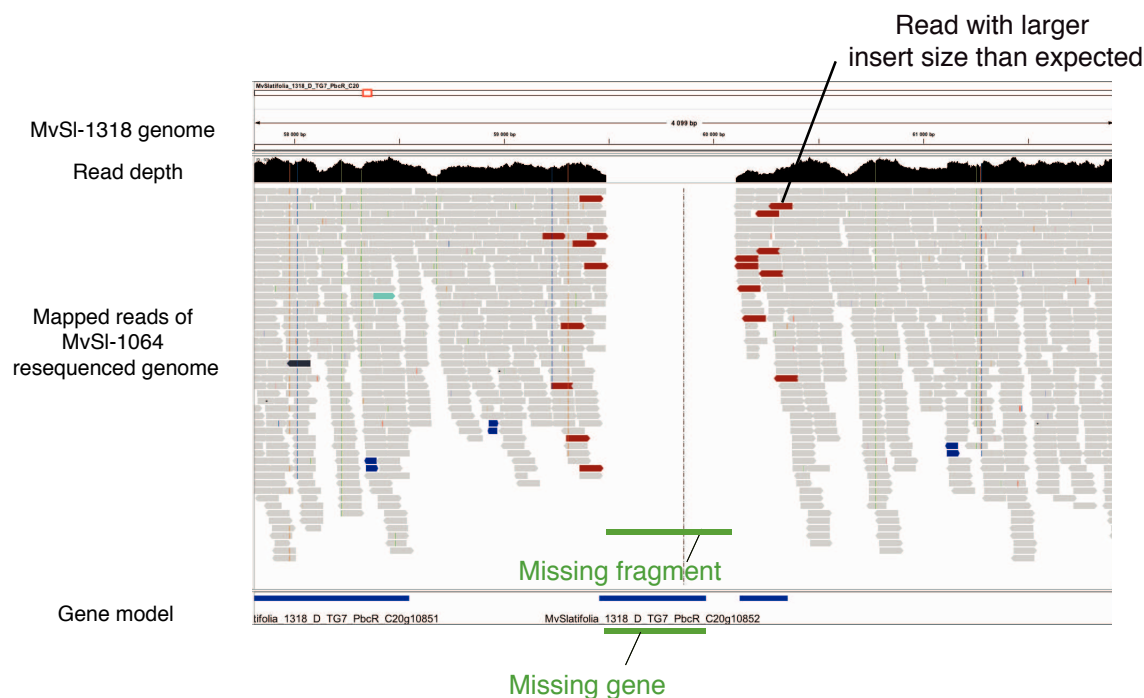
Supplementary figure S1: Detection of missing fragments based on the read depth-based and split read-based methods. Screen shots of mapped reads of the MvSl-1064 Illumina resequenced genome against the MvSl-1318 genome viewed in the Integrative Genomics Viewer (IGV; (Robinson et al. 2011); (Thorvaldsdóttir et al. 2013)). **A.** Example of a missing fragment detected based on the read depth-based method. The significant reduction in read coverage (upper track) led to the indicated missing fragment prediction in green. **B.** Example of a missing fragment detected based on the split read-based method. Red reads corresponded to reads with insert size larger than expected, led to the indicated missing fragment prediction in green.
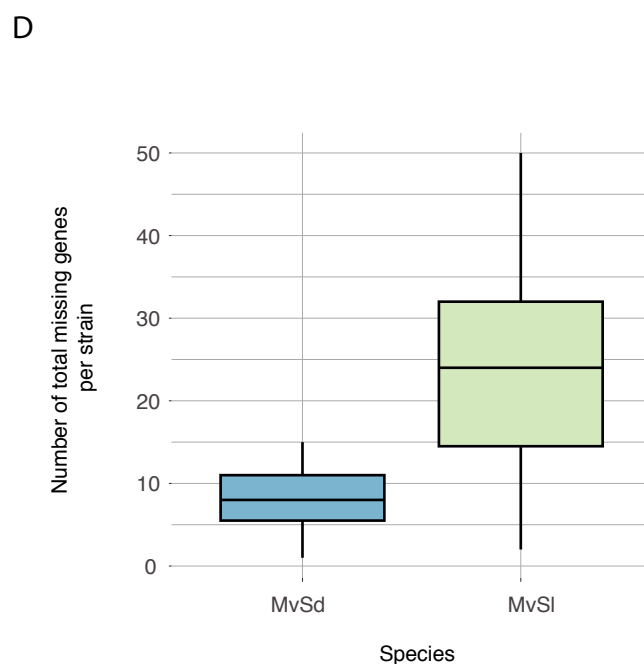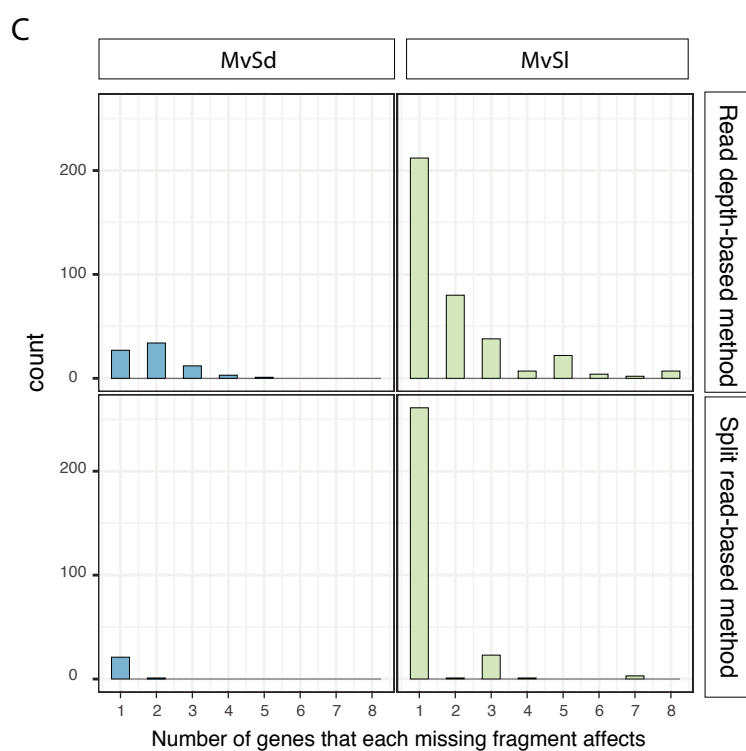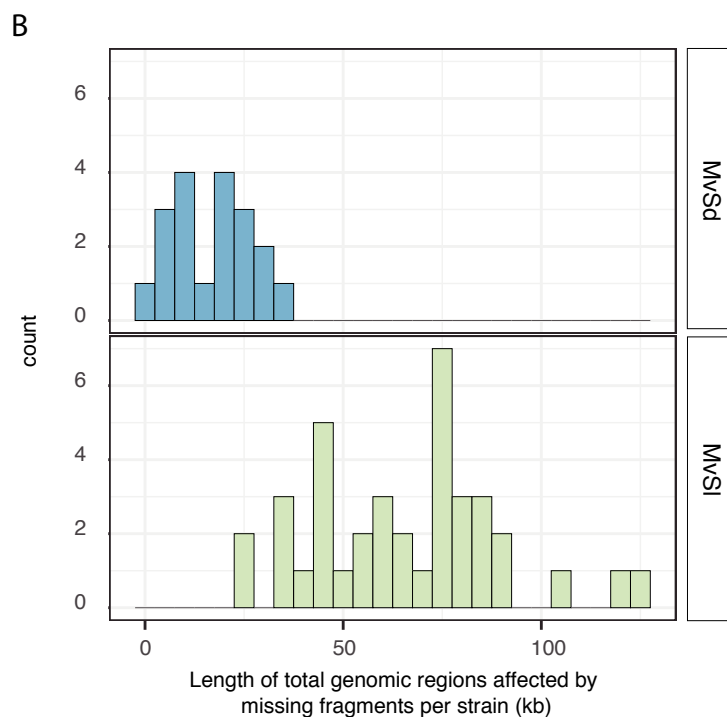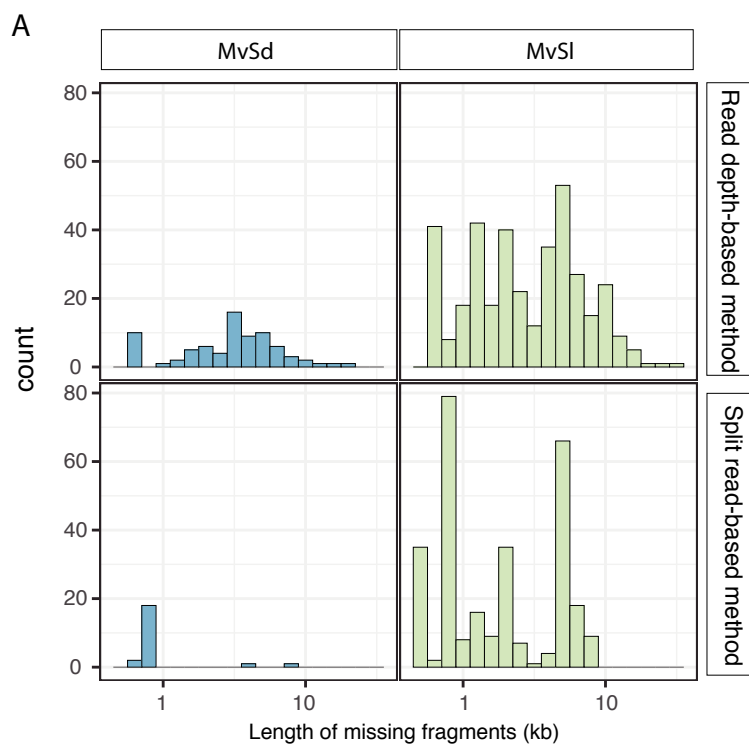
Supplementary figure S2: Characteristics of missing fragments detected based on the read depth-based and split read-based methods that affected more than 90% of a gene length. A. Distribution of the length of missing fragments per method and per species. **B.** Distribution of the total genomic regions affected by detected missing fragments for each strain and each species, considering the overlap between missing fragments detected by the two methods. **C.** Distribution of the number of genes that each missing fragment affects per method and per species. **D.** Distribution of the number of total missing genes for each strain per species.
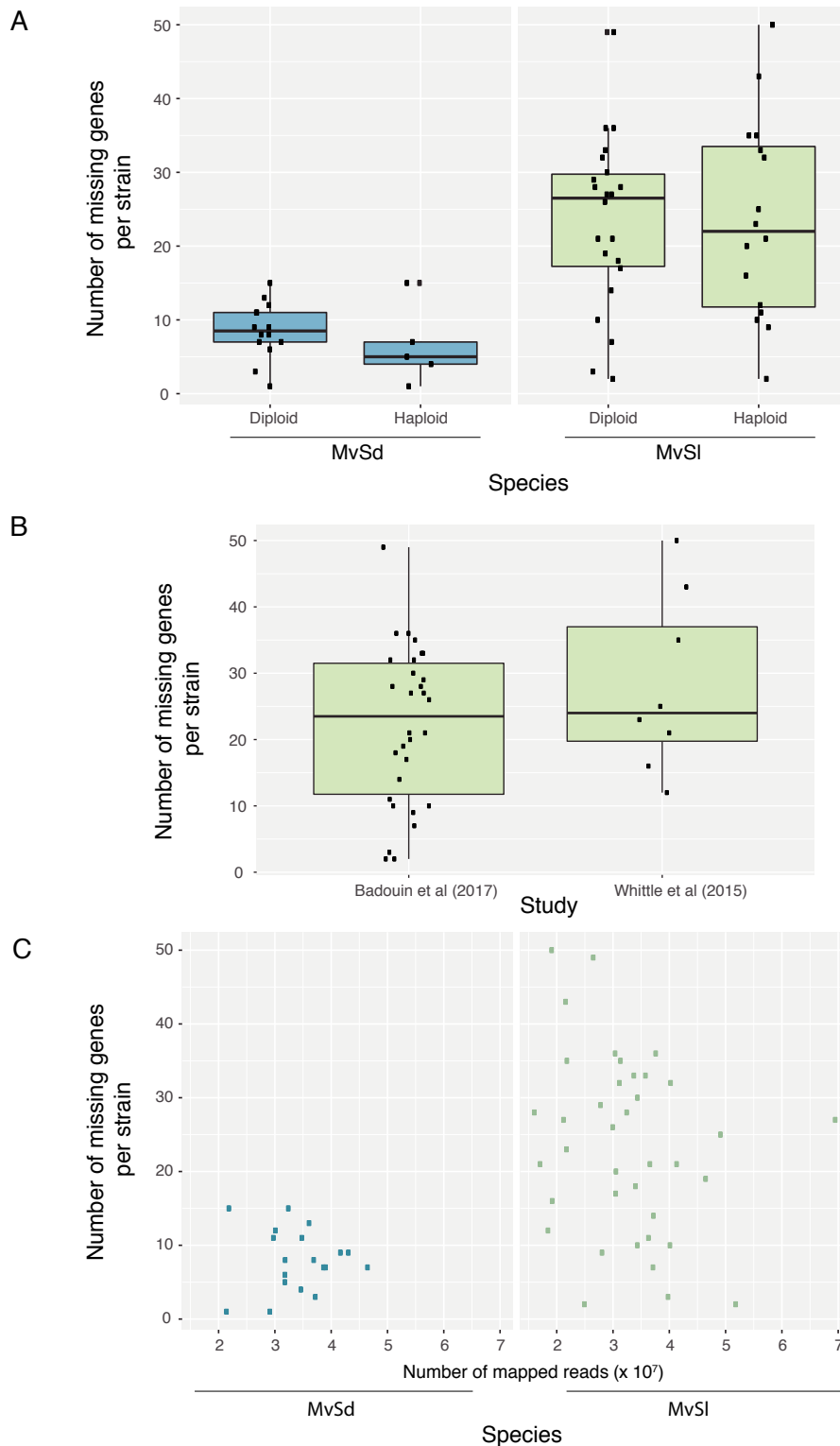
Supplementary figure S3: Effect of sample ploidy (haploid versus diploid), the genome data source (Badouin et al (2017) or Whittle et al (2015)) and the number of mapped reads on the number of missing genes detected per strain.
A. Number of missing genes in haploid and diploid strains in *Microbotryum lychnidis-dioicae* (MvSl) and *Microbotryum silenes-dioicae* (MvSd). Differences were not significant (Kruskal-Wallis rank sum test p-value = 0.9528 in MvSl strains; Kruskal-Wallis rank sum test p-value = 0.2269 in MvSd strains). **B.** Number of missing genes in MvSl strains resequenced by (Badouin et al. 2017) and MvSl strains resequenced by (Whittle et al. 2015). The difference was not significant (Kruskal-Wallis rank sum test p-value = 0.3424). **C.** Correlation between the number of missing genes detected per strain and the number of mapped reads (before random reads sampling procedure in samtools for coverage normalization among strains to 30x). We found no significant correlation (Pearson's correlation test, r = -0.2289905; p-value = 0.1667 in MvSl; Pearson's correlation test, r = -0.00525; p-value = 0.983 in MvSd).
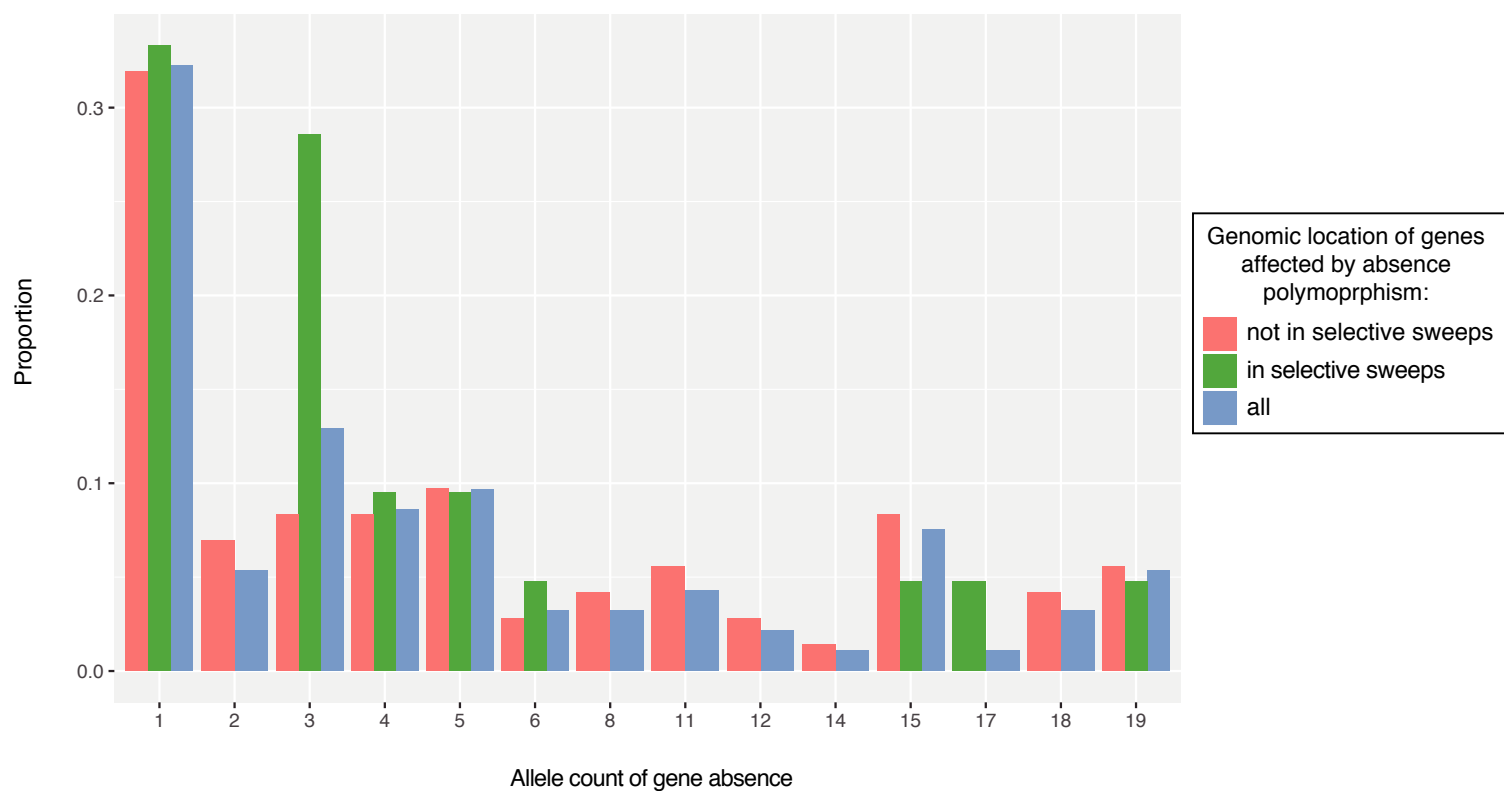
Supplementary figure S4: Distribution of allele counts for gene absence in MvSl isolates belonging to the Northwestern genetic cluster according to their genomic location, within a selective sweep region or not.

Supplementary Table S3: List of the genes affected by presence-absence polymorphism in *Microbotryum lychnidis-dioicae* (MvSl) and *Microbotryum silenes-dioicae* (MvSd).

Frequencies of the gene absence allele among all strains and functional annotation are shown. For the genes affected by presence-absence polymorphism in MvSl, the overlaps of genes with selective sweep regions identified in the Northwestern MvSl genetic cluster in (Badouin et al. 2017) are shown.

Species	gene id	Frequency of the gene absence allele (%)	Secretion prediction	InterPro annotations accession	InterPro annotations description	Gene ontology annotations	Overlap with a selective sweep region	Sweep region coordinates
MvSl	MvSl-1064-A1-R4_MC01-1g01697	47.4	Not secreted	NA	NA	NA	no	-
MvSl	MvSl-1064-A1-R4_MC01-1g01698	50	Not secreted	NA	NA	NA	no	-
MvSl	MvSl-1064-A1-R4_MC01-1g01699	47.4	Not secreted	NA	NA	NA	no	-
MvSl	MvSl-1064-A1-R4_MC01-1g01700	42.1	Not secreted	NA	NA	NA	no	-
MvSl	MvSl-1064-A1-R4_MC01-1g01701	26.3	Not secreted	NA	NA	NA	no	-
MvSl	MvSl-1064-A1-R4_MC01-1g01702	15.8	Not secreted	NA	NA	NA	no	-
MvSl	MvSl-1064-A1-R4_MC01-1g01703	7.9	Not secreted	NA	NA	NA	no	-
MvSl	MvSl-1064-A1-R4_MC01-1g01704	7.9	Not secreted	NA	NA	NA	no	-
MvSl	MvSl-1064-A1-R4_MC01-1g01714	5.3	Not secreted	IPR003140	Phospholipase/carboxylesterase/thioesterase	GO:0016787	no	-
MvSl	MvSl-1064-A1-R4_MC01-1g01715	44.7	Not secreted	NA	NA	NA	no	-
MvSl	MvSl-1064-A1-R4_MC01-1g01716	44.7	Not secreted	NA	NA	NA	no	-
MvSl	MvSl-1064-A1-R4_MC01-1g01717	44.7	Not secreted	NA	NA	NA	no	-
MvSl	MvSl-1064-A1-R4_MC01-1g01718	44.7	Not secreted	NA	NA	NA	no	-
MvSl	MvSl-1064-A1-R4_MC01-1g01719	44.7	Not secreted	NA	NA	NA	no	-
MvSl	MvSl-1064-A1-R4_MC01-1g01739	52.6	Not secreted	NA	NA	NA	no	-
MvSl	MvSl-1064-A1-R4_MC01-1g01740	55.3	Not secreted	NA	NA	NA	no	-
MvSl	MvSl-1064-A1-R4_MC01-1g01741	55.3	Not secreted	NA	NA	NA	no	-
MvSl	MvSl-1064-A1-R4_MC01-1g02329	84.2	Not secreted	NA	NA	NA	yes	MvSlA1A2r3c_S02:1357485-1418040
MvSl	MvSl-1064-A1-R4_MC01-1g02588	18.4	Not secreted	NA	NA	NA	no	-
MvSl	MvSl-1064-A1-R4_MC01-1g02594	2.6	Not secreted	NA	NA	NA	yes	MvSlA1A2r3c_S02:703363-744096
MvSl	MvSl-1064-A1-R4_MC01-1g02724	2.6	Not secreted	NA	NA	NA	no	-
MvSl	MvSl-1064-A1-R4_MC01-1g02725	13.2	Not secreted	NA	NA	NA	no	-
MvSl	MvSl-1064-A1-R4_MC01-1g02726	5.3	Not secreted	NA	NA	NA	no	-
MvSl	MvSl-1064-A1-R4_MC01-1g02804	23.7	Not secreted	NA	NA	NA	no	-
MvSl	MvSl-1064-A1-R4_MC01-1g02830	10.5	Not secreted	NA	NA	NA	no	-
MvSl	MvSl-1064-A1-R4_MC01-1g02859	2.6	Not secreted	NA	NA	NA	no	-
MvSl	MvSl-1064-A1-R4_MC01-2g03093	2.6	Not secreted	NA	NA	NA	no	-
MvSl	MvSl-1064-A1-R4_MC02g03349	7.9	Not secreted	NA	NA	NA	no	-
MvSl	MvSl-1064-A1-R4_MC02g03350	7.9	Not secreted	NA	NA	NA	no	-
MvSl	MvSl-1064-A1-R4_MC02g03351	7.9	Not secreted	NA	NA	NA	no	-
MvSl	MvSl-1064-A1-R4_MC02g03373	5.3	Not secreted	NA	NA	NA	no	-
MvSl	MvSl-1064-A1-R4_MC02g03410	52.6	Not secreted	NA	NA	NA	no	-
MvSl	MvSl-1064-A1-R4_MC02g03411	60.5	Not secreted	NA	NA	NA	no	-
MvSl	MvSl-1064-A1-R4_MC02g03448	7.9	Not secreted	NA	NA	NA	no	-
MvSl	MvSl-1064-A1-R4_MC02g03449	15.8	Not secreted	NA	NA	NA	no	-
MvSl	MvSl-1064-A1-R4_MC02g03451	21.1	Not secreted	NA	NA	NA	no	-
MvSl	MvSl-1064-A1-R4_MC02g03470	2.6	Not secreted	NA	NA	NA	no	-
MvSl	MvSl-1064-A1-R4_MC02g03494	7.9	Not secreted	NA	NA	NA	no	-
MvSl	MvSl-1064-A1-R4_MC02g03495	7.9	Not secreted	NA	NA	NA	no	-
MvSl	MvSl-1064-A1-R4_MC02g03496	7.9	Not secreted	NA	NA	NA	no	-
MvSl	MvSl-1064-A1-R4_MC02g03497	7.9	Not secreted	NA	NA	NA	no	-
MvSl	MvSl-1064-A1-R4_MC02g03498	7.9	Not secreted	NA	NA	NA	no	-
MvSl	MvSl-1064-A1-R4_MC02g03499	7.9	Not secreted	NA	NA	NA	no	-
MvSl	MvSl-1064-A1-R4_MC02g03500	7.9	Not secreted	NA	NA	NA	no	-
MvSl	MvSl-1064-A1-R4_MC02g03508	2.6	Not secreted	NA	NA	NA	no	-
MvSl	MvSl-1064-A1-R4_MC02g03509	2.6	Not secreted	NA	NA	NA	no	-
MvSl	MvSl-1064-A1-R4_MC02g03857	2.6	Not secreted	NA	NA	NA	no	-
MvSl	MvSl-1064-A1-R4_MC02g03873	7.9	Not secreted	NA	NA	NA	no	-
MvSl	MvSl-1064-A1-R4_MC02g04192	5.3	Not secreted	NA	NA	NA	no	-
MvSl	MvSl-1064-A1-R4_MC02g04193	5.3	Not secreted	IPR013955	Replication factor A, C-terminal		no	-
MvSl	MvSl-1064-A1-R4_MC02g04216	86.8	Not secreted	NA	NA	NA	no	-
MvSl	MvSl-1064-A1-R4_MC02g04225	13.2	Not secreted	NA	NA	NA	no	-
MvSl	MvSl-1064-A1-R4_MC02g04246	28.9	Not secreted	NA	NA	NA	no	-
MvSl	MvSl-1064-A1-R4_MC02g04247	28.9	Not secreted	NA	NA	NA	no	-
MvSl	MvSl-1064-A1-R4_MC03g04391	31.6	Not secreted	NA	NA	NA	no	-
MvSl	MvSl-1064-A1-R4_MC03g04392	2.6	Not secreted	NA	NA	NA	no	-
MvSl	MvSl-1064-A1-R4_MC03g04426	5.3	Not secreted	NA	NA	NA	no	-
MvSl	MvSl-1064-A1-R4_MC03g04428	13.2	Not secreted	NA	NA	NA	no	-
MvSl	MvSl-1064-A1-R4_MC03g04550	7.9	Not secreted	IPR006175	YjgF/YER057c/UK114 family		no	-
MvSl	MvSl-1064-A1-R4_MC03g04594	2.6	Not secreted	NA	NA	NA	no	-

MvSl	MvSl-1064-A1-R4_MC03g04666	2.6	Not secreted	NA	NA	NA	yes	MvSlA1A2r3c_S04:739414-800583
MvSl	MvSl-1064-A1-R4_MC03g04792	23.7	Not secreted	NA	NA	NA	no	-
MvSl	MvSl-1064-A1-R4_MC03g05038	2.6	Not secreted	NA	NA	NA	no	-
MvSl	MvSl-1064-A1-R4_MC03g05197	23.7	Not secreted	NA	NA	NA	yes	MvSlA1A2r3c_S04:2083932-2114999
MvSl	MvSl-1064-A1-R4_MC03g05217	5.3	Not secreted	IPR013860	Protein of unknown function DUF1752, fungi		no	-
MvSl	MvSl-1064-A1-R4_MC03g05230	2.6	Not secreted	NA	NA	NA	no	-
MvSl	MvSl-1064-A1-R4_MC03g05231	2.6	Not secreted	NA	NA	NA	no	-
MvSl	MvSl-1064-A1-R4_MC03g05232	2.6	Not secreted	NA	NA	NA	no	-
MvSl	MvSl-1064-A1-R4_MC04-1g05274	7.9	Not secreted	NA	NA	NA	no	-
MvSl	MvSl-1064-A1-R4_MC04-1g05284	5.3	Not secreted	NA	NA	NA	no	-
MvSl	MvSl-1064-A1-R4_MC04-1g05313	5.3	Not secreted	NA	NA	NA	no	-
MvSl	MvSl-1064-A1-R4_MC04-1g05314	5.3	Not secreted	NA	NA	NA	no	-
MvSl	MvSl-1064-A1-R4_MC04-1g05315	5.3	Not secreted	NA	NA	NA	no	-
MvSl	MvSl-1064-A1-R4_MC04-1g05510	5.3	Not secreted	NA	NA	NA	no	-
MvSl	MvSl-1064-A1-R4_MC04-1g05526	2.6	Not secreted	IPR024671	Autophagy-related protein 22-like		no	-
MvSl	MvSl-1064-A1-R4_MC05-3g06732	2.6	Not secreted	NA	NA	NA	no	-
MvSl	MvSl-1064-A1-R4_MC06g06889	5.3	Secreted protein	NA	NA	NA	no	-
MvSl	MvSl-1064-A1-R4_MC06g06892	2.6	Not secreted	NA	NA	NA	no	-
MvSl	MvSl-1064-A1-R4_MC06g07071	5.3	Not secreted	NA	NA	NA	no	-
MvSl	MvSl-1064-A1-R4_MC06g07089	2.6	Not secreted	NA	NA	NA	no	-
MvSl	MvSl-1064-A1-R4_MC06g07196	5.3	Not secreted	NA	NA	NA	no	-
MvSl	MvSl-1064-A1-R4_MC06g07197	5.3	Not secreted	NA	NA	NA	no	-
MvSl	MvSl-1064-A1-R4_MC06g07198	5.3	Not secreted	NA	NA	NA	no	-
MvSl	MvSl-1064-A1-R4_MC06g07322	76.3	Not secreted	NA	NA	NA	yes	MvSlA1A2r3c_S07:501146-512146
MvSl	MvSl-1064-A1-R4_MC06g07412	2.6	Not secreted	NA	NA	NA	no	-
MvSl	MvSl-1064-A1-R4_MC06g07413	2.6	Not secreted	NA	NA	NA	no	-
MvSl	MvSl-1064-A1-R4_MC06g07418	2.6	Not secreted	NA	NA	NA	no	-
MvSl	MvSl-1064-A1-R4_MC06g07545	2.6	Not secreted	NA	NA	NA	no	-
MvSl	MvSl-1064-A1-R4_MC06g07546	2.6	Not secreted	NA	NA	NA	no	-
MvSl	MvSl-1064-A1-R4_MC07g07674	42.1	Not secreted	NA	NA	NA	yes	MvSlA1A2r3c_S06:1415139-1426139
MvSl	MvSl-1064-A1-R4_MC07g07676	2.6	Not secreted	NA	NA	NA	no	-
MvSl	MvSl-1064-A1-R4_MC07g07692	21.1	Not secreted	NA	NA	NA	no	-
MvSl	MvSl-1064-A1-R4_MC07g07693	15.8	Secreted protein	IPR033121	Peptidase family A1 domain		no	-
MvSl	MvSl-1064-A1-R4_MC07g07703	86.8	Not secreted	NA	NA	NA	no	-
MvSl	MvSl-1064-A1-R4_MC07g08231	2.6	Not secreted	NA	NA	NA	no	-
MvSl	MvSl-1064-A1-R4_MC07g08249	2.6	Not secreted	NA	NA	NA	no	-
MvSl	MvSl-1064-A1-R4_MC08g08262	2.6	Not secreted	NA	NA	NA	yes	MvSlA1A2r3c_S08:7244-18244
MvSl	MvSl-1064-A1-R4_MC08g08263	2.6	Not secreted	NA	NA	NA	yes	MvSlA1A2r3c_S08:7244-18244
MvSl	MvSl-1064-A1-R4_MC08g08264	2.6	Not secreted	NA	NA	NA	yes	MvSlA1A2r3c_S08:7244-18244
MvSl	MvSl-1064-A1-R4_MC08g08310	2.6	Not secreted	NA	NA	NA	no	-
MvSl	MvSl-1064-A1-R4_MC08g08473	5.3	Not secreted	NA	NA	NA	no	-
MvSl	MvSl-1064-A1-R4_MC08g08860	60.5	Not secreted	NA	NA	NA	no	-
MvSl	MvSl-1064-A1-R4_MC09-2g08968	18.4	Not secreted	IPR015590	Aldehyde dehydrogenase domain	GO:0008152|GO:0016491|GO:0055114	no	-
MvSl	MvSl-1064-A1-R4_MC09-2g08969	2.6	Not secreted	NA	NA	NA	no	-
MvSl	MvSl-1064-A1-R4_MC09-2g09249	13.2	Not secreted	NA	NA	NA	no	-
MvSl	MvSl-1064-A1-R4_MC09-2g09281	15.8	Not secreted	NA	NA	NA	no	-
MvSl	MvSl-1064-A1-R4_MC09-2g09282	15.8	Not secreted	NA	NA	NA	no	-
MvSl	MvSl-1064-A1-R4_MC09-2g09360	2.6	Not secreted	IPR007696	DNA mismatch repair protein MutS, core	GO:0005524|GO:0006298|GO:0030983	no	-
MvSl	MvSl-1064-A1-R4_MC09-2g09361	2.6	Not secreted	IPR007696	DNA mismatch repair protein MutS, core	GO:0005524|GO:0006298|GO:0030983	no	-
MvSl	MvSl-1064-A1-R4_MC09-2g09361	2.6	Not secreted	IPR000432	DNA mismatch repair protein MutS, C-terminal	GO:0005524|GO:0006298|GO:0030983	no	-
MvSl	MvSl-1064-A1-R4_MC09-2g09361	2.6	Not secreted	IPR007861	DNA mismatch repair protein MutS, clamp	GO:0005524|GO:0006298|GO:0030983	no	-
MvSl	MvSl-1064-A1-R4_MC09-2g09362	2.6	Not secreted	NA	NA	NA	no	-
MvSl	MvSl-1064-A1-R4_MC09-2g09363	2.6	Not secreted	IPR007135	Autophagy-related protein 3		no	-
MvSl	MvSl-1064-A1-R4_MC09-2g09364	2.6	Not secreted	IPR013786	Acyl-CoA dehydrogenase/oxidase, N-terminal	GO:0016627|GO:0050660|GO:0055114	no	-
MvSl	MvSl-1064-A1-R4_MC09-2g09364	2.6	Not secreted	IPR009075	Acyl-CoA dehydrogenase/oxidase C-terminal	GO:0016627|GO:0055114	no	-
MvSl	MvSl-1064-A1-R4_MC09-2g09364	2.6	Not secreted	IPR006091	Acyl-CoA oxidase/dehydrogenase, central domain	GO:0016627|GO:0055114	no	-
MvSl	MvSl-1064-A1-R4_MC10-1g09588	5.3	Not secreted	NA	NA	NA	no	-
MvSl	MvSl-1064-A1-R4_MC10-1g09589	2.6	Not secreted	NA	NA	NA	no	-
MvSl	MvSl-1064-A1-R4_MC10-1g09591	2.6	Not secreted	NA	NA	NA	no	-
MvSl	MvSl-1064-A1-R4_MC10-1g09592	28.9	Not secreted	NA	NA	NA	no	-
MvSl	MvSl-1064-A1-R4_MC10-1g09593	28.9	Not secreted	NA	NA	NA	no	-
MvSl	MvSl-1064-A1-R4_MC10-1g09594	31.6	Not secreted	NA	NA	NA	no	-
MvSl	MvSl-1064-A1-R4_MC10-1g09596	5.3	Not secreted	NA	NA	NA	no	-
MvSl	MvSl-1064-A1-R4_MC10-1g09597	2.6	Not secreted	IPR000719	Protein kinase domain	GO:0004672|GO:0005524|GO:0006468	no	-
MvSl	MvSl-1064-A1-R4_MC10-1g09598	2.6	Not secreted	NA	NA	NA	no	-
MvSl	MvSl-1064-A1-R4_MC10-1g09599	2.6	Not secreted	NA	NA	NA	no	-
MvSl	MvSl-1064-A1-R4_MC10-1g09611	2.6	Not secreted	NA	NA	NA	yes	MvSlA1A2r3c_S15:633869-654126

MvSl	*MvSl-1064-A1-R4_MC10-1g09638*	2.6	Not secreted	NA	NA	NA	no	-
MvSl	*MvSl-1064-A1-R4_MC10-1g09673*	2.6	Not secreted	NA	NA	NA	no	-
MvSl	*MvSl-1064-A1-R4_MC10-1g09809*	2.6	Not secreted	NA	NA	NA	no	-
MvSl	*MvSl-1064-A1-R4_MC10-1g09810*	2.6	Not secreted	NA	NA	NA	no	-
MvSl	*MvSl-1064-A1-R4_MC10-1g09824*	2.6	Not secreted	NA	NA	NA	no	-
MvSl	*MvSl-1064-A1-R4_MC10-1g09826*	2.6	Not secreted	NA	NA	NA	no	-
MvSl	*MvSl-1064-A1-R4_MC10-1g09827*	2.6	Not secreted	NA	NA	NA	no	-
MvSl	*MvSl-1064-A1-R4_MC10-1g09829*	2.6	Not secreted	NA	NA	NA	no	-
MvSl	*MvSl-1064-A1-R4_MC10-1g09830*	2.6	Not secreted	NA	NA	NA	no	-
MvSl	*MvSl-1064-A1-R4_MC10-2g09861*	13.2	Not secreted	NA	NA	NA	no	-
MvSl	*MvSl-1064-A1-R4_MC10-2g09919*	92.1	Not secreted	NA	NA	NA	no	-
MvSl	*MvSl-1064-A1-R4_MC10-2g10046*	2.6	Not secreted	IPR005135	Endonuclease/exonuclease/phosphatase		no	-
MvSl	*MvSl-1064-A1-R4_MC11g10180*	2.6	Not secreted	NA	NA	NA	no	-
MvSl	*MvSl-1064-A1-R4_MC11g10184*	15.8	Small secreted protein	NA	NA	NA	yes	MvSlA1A2r3c_S01:2964344-3005424
MvSl	*MvSl-1064-A1-R4_MC11g10185*	10.5	Not secreted	NA	NA	NA	yes	MvSlA1A2r3c_S01:2964344-3005424
MvSl	*MvSl-1064-A1-R4_MC11g10186*	10.5	Not secreted	NA	NA	NA	yes	MvSlA1A2r3c_S01:2964344-3005424
MvSl	*MvSl-1064-A1-R4_MC11g10187*	10.5	Not secreted	NA	NA	NA	yes	MvSlA1A2r3c_S01:2964344-3005424
MvSl	*MvSl-1064-A1-R4_MC11g10188*	10.5	Not secreted	NA	NA	NA	yes	MvSlA1A2r3c_S01:2964344-3005424
MvSl	*MvSl-1064-A1-R4_MC11g10189*	10.5	Not secreted	NA	NA	NA	yes	MvSlA1A2r3c_S01:2964344-3005424
MvSl	*MvSl-1064-A1-R4_MC11g10190*	13.2	Not secreted	NA	NA	NA	yes	MvSlA1A2r3c_S01:2964344-3005424
MvSl	*MvSl-1064-A1-R4_MC11g10191*	13.2	Small secreted protein	NA	NA	NA	yes	MvSlA1A2r3c_S01:2964344-3005424
MvSl	*MvSl-1064-A1-R4_MC11g10197*	2.6	Not secreted	NA	NA	NA	no	-
MvSl	*MvSl-1064-A1-R4_MC11g10198*	2.6	Not secreted	NA	NA	NA	no	-
MvSl	*MvSl-1064-A1-R4_MC11g10199*	2.6	Not secreted	NA	NA	NA	no	-
MvSl	*MvSl-1064-A1-R4_MC11g10233*	2.6	Not secreted	IPR000892	Ribosomal protein S26e	GO:0003735|GO:0005622|GO:0005840|GO:0006412	no	-
MvSl	*MvSl-1064-A1-R4_MC11g10262*	2.6	Not secreted	NA	NA	NA	no	-
MvSl	*MvSl-1064-A1-R4_MC11g10263*	2.6	Not secreted	NA	NA	NA	no	-
MvSl	*MvSl-1064-A1-R4_MC11g10284*	7.9	Not secreted	NA	NA	NA	no	-
MvSl	*MvSl-1064-A1-R4_MC11g10687*	5.3	Not secreted	NA	NA	NA	no	-
MvSl	*MvSl-1064-A1-R4_MC11g10722*	21.1	Not secreted	NA	NA	NA	no	-
MvSl	*MvSl-1064-A1-R4_MC11g10728*	5.3	Not secreted	NA	NA	NA	no	-
MvSl	*MvSl-1064-A1-R4_MC12g10787*	5.3	Not secreted	NA	NA	NA	no	-
MvSl	*MvSl-1064-A1-R4_MC12g10788*	2.6	Small secreted protein	NA	NA	NA	no	-
MvSl	*MvSl-1064-A1-R4_MC12g11127*	2.6	Not secreted	NA	NA	NA	yes	MvSlA1A2r3c_S09:935693-985491
MvSl	*MvSl-1064-A1-R4_MC13g11305*	13.2	Not secreted	NA	NA	NA	no	-
MvSl	*MvSl-1064-A1-R4_MC13g11325*	2.6	Not secreted	NA	NA	NA	no	-
MvSl	*MvSl-1064-A1-R4_MC13g11328*	2.6	Not secreted	NA	NA	NA	no	-
MvSl	*MvSl-1064-A1-R4_MC13g11331*	2.6	Not secreted	NA	NA	NA	no	-
MvSl	*MvSl-1064-A1-R4_MC13g11359*	2.6	Not secreted	NA	NA	NA	no	-
MvSl	*MvSl-1064-A1-R4_MC13g11366*	2.6	Not secreted	NA	NA	NA	no	-
MvSl	*MvSl-1064-A1-R4_MC13g11367*	2.6	Not secreted	NA	NA	NA	no	-
MvSl	*MvSl-1064-A1-R4_MC13g11374*	2.6	Not secreted	NA	NA	NA	no	-
MvSl	*MvSl-1064-A1-R4_MC13g11375*	2.6	Not secreted	NA	NA	NA	no	-
MvSl	*MvSl-1064-A1-R4_MC13g11394*	7.9	Not secreted	NA	NA	NA	no	-
MvSl	*MvSl-1064-A1-R4_MC13g11540*	2.6	Not secreted	NA	NA	NA	no	-
MvSl	*MvSl-1064-A1-R4_MC13g11541*	21.1	Secreted protein	IPR000560	Histidine phosphatase superfamily, clade-2	GO:0003993	yes	MvSlA1A2r3c_S11:445169-475073
MvSl	*MvSl-1064-A1-R4_MC13g11714*	2.6	Not secreted	NA	NA	NA	no	-
MvSl	*MvSl-1064-A1-R4_MC14g11780*	2.6	Not secreted	NA	NA	NA	no	-
MvSl	*MvSl-1064-A1-R4_MC14g11781*	2.6	Not secreted	NA	NA	NA	no	-
MvSl	*MvSl-1064-A1-R4_MC14g11782*	2.6	Not secreted	NA	NA	NA	no	-
MvSl	*MvSl-1064-A1-R4_MC14g11783*	2.6	Not secreted	NA	NA	NA	no	-
MvSl	*MvSl-1064-A1-R4_MC14g11800*	5.3	Not secreted	NA	NA	NA	no	-
MvSl	*MvSl-1064-A1-R4_MC14g11801*	5.3	Not secreted	NA	NA	NA	no	-
MvSl	*MvSl-1064-A1-R4_MC14g11802*	5.3	Not secreted	NA	NA	NA	no	-
MvSl	*MvSl-1064-A1-R4_MC14g11803*	5.3	Not secreted	NA	NA	NA	no	-
MvSl	*MvSl-1064-A1-R4_MC14g11804*	5.3	Not secreted	NA	NA	NA	no	-
MvSl	*MvSl-1064-A1-R4_MC14g11806*	2.6	Not secreted	NA	NA	NA	no	-
MvSl	*MvSl-1064-A1-R4_MC14g11809*	18.4	Not secreted	NA	NA	NA	no	-
MvSl	*MvSl-1064-A1-R4_MC14g11810*	13.2	Not secreted	NA	NA	NA	no	-
MvSl	*MvSl-1064-A1-R4_MC14g11819*	13.2	Not secreted	NA	NA	NA	no	-
MvSl	*MvSl-1064-A1-R4_MC14g11820*	5.3	Small secreted protein	NA	NA	NA	no	-
MvSl	*MvSl-1064-A1-R4_MC14g11950*	21.1	Not secreted	NA	NA	NA	no	-
MvSl	*MvSl-1064-A1-R4_MC14g11974*	68.4	Not secreted	NA	NA	NA	no	-
MvSd	*MvSdioicae_1303_FR02_D_N206_PbcR_C003g02008*	21.1	Not secreted	NA	NA	NA	*not studied*	-
MvSd	*MvSdioicae_1303_FR02_D_N206_PbcR_C003g02010*	21.1	Not secreted	NA	NA	NA	*not studied*	-

MvSd	*MvSdioicae_1303_FR02_D_N206_PbcR_C003g02503*	10.5	Not secreted	NA	NA	NA	*not studied*	-
MvSd	*MvSdioicae_1303_FR02_D_N206_PbcR_C003g02504*	10.5	Not secreted	NA	NA	NA	*not studied*	-
MvSd	*MvSdioicae_1303_FR02_D_N206_PbcR_C003g02535*	57.9	Not secreted	NA	NA	NA	*not studied*	-
MvSd	*MvSdioicae_1303_FR02_D_N206_PbcR_C003g02536*	57.9	Not secreted	NA	NA	NA	*not studied*	-
MvSd	*MvSdioicae_1303_FR02_D_N206_PbcR_C003g02537*	57.9	Not secreted	NA	NA	NA	*not studied*	-
MvSd	*MvSdioicae_1303_FR02_D_N206_PbcR_C003g02538*	15.8	Not secreted	NA	NA	NA	*not studied*	-
MvSd	*MvSdioicae_1303_FR02_D_N206_PbcR_C004g03137*	5.3	Not secreted	NA	NA	NA	*not studied*	-
MvSd	*MvSdioicae_1303_FR02_D_N206_PbcR_C004g03138*	5.3	Not secreted	NA	NA	NA	*not studied*	-
MvSd	*MvSdioicae_1303_FR02_D_N206_PbcR_C007g04688*	10.5	Not secreted	NA	NA	NA	*not studied*	-
MvSd	*MvSdioicae_1303_FR02_D_N206_PbcR_C008g05262*	73.7	Not secreted	NA	NA	NA	*not studied*	-
MvSd	*MvSdioicae_1303_FR02_D_N206_PbcR_C009g05417*	10.5	Not secreted	NA	NA	NA	*not studied*	-
MvSd	*MvSdioicae_1303_FR02_D_N206_PbcR_C009g05418*	26.3	Not secreted	NA	NA	NA	*not studied*	-
MvSd	*MvSdioicae_1303_FR02_D_N206_PbcR_C009g05419*	26.3	Not secreted	NA	NA	NA	*not studied*	-
MvSd	*MvSdioicae_1303_FR02_D_N206_PbcR_C012g06774*	15.8	Not secreted	NA	NA	NA	*not studied*	-
MvSd	*MvSdioicae_1303_FR02_D_N206_PbcR_C012g06780*	15.8	Not secreted	NA	NA	NA	*not studied*	-
MvSd	*MvSdioicae_1303_FR02_D_N206_PbcR_C012g06781*	26.3	Small secreted proteins	NA	NA	NA	*not studied*	-
MvSd	*MvSdioicae_1303_FR02_D_N206_PbcR_C013g07079*	5.3	Not secreted	NA	NA	NA	*not studied*	-
MvSd	*MvSdioicae_1303_FR02_D_N206_PbcR_C014g07422*	21.1	Small secreted proteins	NA	NA	NA	*not studied*	-
MvSd	*MvSdioicae_1303_FR02_D_N206_PbcR_C014g07423*	21.1	Not secreted	NA	NA	NA	*not studied*	-
MvSd	*MvSdioicae_1303_FR02_D_N206_PbcR_C014g07585*	5.3	Not secreted	NA	NA	NA	*not studied*	-
MvSd	*MvSdioicae_1303_FR02_D_N206_PbcR_C014g07586*	15.8	Not secreted	NA	NA	NA	*not studied*	-
MvSd	*MvSdioicae_1303_FR02_D_N206_PbcR_C014g07707*	5.3	Not secreted	NA	NA	NA	*not studied*	-
MvSd	*MvSdioicae_1303_FR02_D_N206_PbcR_C014g07708*	26.3	Not secreted	NA	NA	NA	*not studied*	-
MvSd	*MvSdioicae_1303_FR02_D_N206_PbcR_C015g07726*	21.1	Not secreted	NA	NA	NA	*not studied*	-
MvSd	*MvSdioicae_1303_FR02_D_N206_PbcR_C018g08795*	5.3	Not secreted	NA	NA	NA	*not studied*	-
MvSd	*MvSdioicae_1303_FR02_D_N206_PbcR_C018g08796*	5.3	Not secreted	NA	NA	NA	*not studied*	-
MvSd	*MvSdioicae_1303_FR02_D_N206_PbcR_C022g09526*	10.5	Not secreted	NA	NA	NA	*not studied*	-
MvSd	*MvSdioicae_1303_FR02_D_N206_PbcR_C022g09527*	42.1	Not secreted	IPR013955	Replication factor A, C-terminal	NA	*not studied*	-
MvSd	*MvSdioicae_1303_FR02_D_N206_PbcR_C022g09528*	26.3	Not secreted	NA	NA	NA	*not studied*	-
MvSd	*MvSdioicae_1303_FR02_D_N206_PbcR_C023g09600*	10.5	Not secreted	NA	NA	NA	*not studied*	-
MvSd	*MvSdioicae_1303_FR02_D_N206_PbcR_C023g09601*	10.5	Not secreted	NA	NA	NA	*not studied*	-
MvSd	*MvSdioicae_1303_FR02_D_N206_PbcR_C029g10640*	5.3	Not secreted	NA	NA	NA	*not studied*	-
MvSd	*MvSdioicae_1303_FR02_D_N206_PbcR_C033g11197*	5.3	Small secreted proteins	NA	NA	NA	*not studied*	-
MvSd	*MvSdioicae_1303_FR02_D_N206_PbcR_C038g11673*	5.3	Not secreted	NA	NA	NA	*not studied*	-
MvSd	*MvSdioicae_1303_FR02_D_N206_PbcR_C039g11781*	5.3	Not secreted	NA	NA	NA	*not studied*	-
MvSd	*MvSdioicae_1303_FR02_D_N206_PbcR_C041g11942*	5.3	Not secreted	NA	NA	NA	*not studied*	-
MvSd	*MvSdioicae_1303_FR02_D_N206_PbcR_C044g12163*	5.3	Not secreted	NA	NA	NA	*not studied*	-
MvSd	*MvSdioicae_1303_FR02_D_N206_PbcR_C044g12164*	5.3	Not secreted	NA	NA	NA	*not studied*	-
MvSd	*MvSdioicae_1303_FR02_D_N206_PbcR_C045g12191*	15.8	Not secreted	NA	NA	NA	*not studied*	-
MvSd	*MvSdioicae_1303_FR02_D_N206_PbcR_C045g12192*	15.8	Not secreted	NA	NA	NA	*not studied*	-
MvSd	*MvSdioicae_1303_FR02_D_N206_PbcR_C045g12193*	15.8	Not secreted	NA	NA	NA	*not studied*	-
MvSd	*MvSdioicae_1303_FR02_D_N206_PbcR_C045g12196*	15.8	Not secreted	NA	NA	NA	*not studied*	-
MvSd	*MvSdioicae_1303_FR02_D_N206_PbcR_C045g12197*	15.8	Not secreted	NA	NA	NA	*not studied*	-
MvSd	*MvSdioicae_1303_FR02_D_N206_PbcR_C045g12203*	5.3	Not secreted	NA	NA	NA	*not studied*	-
MvSd	*MvSdioicae_1303_FR02_D_N206_PbcR_C046g12271*	5.3	Not secreted	IPR004147	UbiB domain	NA	*not studied*	-
MvSd	*MvSdioicae_1303_FR02_D_N206_PbcR_C046g12272*	5.3	Not secreted	IPR023753	FAD/NAD(P)-binding domain	GO:0016491|GO:0055114	*not studied*	-
MvSd	*MvSdioicae_1303_FR02_D_N206_PbcR_C046g12273*	5.3	Not secreted	IPR014811	Argonaute, linker 1 domain	NA	*not studied*	-
MvSd	*MvSdioicae_1303_FR02_D_N206_PbcR_C046g12276*	5.3	Not secreted	IPR002685	Glycosyl transferase, family 15	GO:0000030|GO:0006486|GO:0016020	*not studied*	-
MvSd	*MvSdioicae_1303_FR02_D_N206_PbcR_C046g12277*	5.3	Not secreted	IPR001356	Homeobox domain	GO:0003677	*not studied*	-

B. Review: Genomic to understand adaptation, coevolution, host specialization and mating system – insights from castrating anther-smut fungi

1 **Understanding Adaptation, Coevolution, Host Specialization and Mating System by**

2 **combining Population and Comparative Genomics in Castrating Anther-Smut Fungi**

3

4 Fanny E. Hartmann[1], Ricardo C. Rodríguez de la Vega[1], Fantin Carpentier[1], Pierre Gladieux[2],

5 Amandine Cornille[3], Michael E. Hood[4], Tatiana Giraud[1]

6

7 [1] Ecologie Systématique Evolution, Bâtiment 360, Univ. Paris-Sud, AgroParisTech, CNRS, Université Paris-

8 Saclay, 91400 Orsay, France

9 [2] INRA, UMR BGPI, Bâtiment K; Campus International de Baillarguet, F-34398, Montpellier, France

10 [3] Génétique Quantitative et Evolution – Le Moulon, INRA ; Univ. Paris-Sud, CNRS, AgroParisTech, Université

11 Paris-Saclay, 91198 Gif-sur-Yvette, France

12 [4] Biology Department, Science Center Building, Amherst College, Rts 9 & 116, Amherst, MA USA 01002-5000

13

14

15 Corresponding author: fanny.hartmann@u-psud.fr.

16

17

18   **Abstract**

19

20   Anther-smut fungi constitute a powerful system to study host-pathogen specialization and

21   coevolution, with hundreds of *Microbotryum* species specialized on diverse Caryophyllaceae

22   plants, castrating their hosts through particular manipulation of hosts' reproductive organs

23   that facilitates disease transmission. *Microbotryum* fungi also have exceptional genomic traits,

24   including dimorphic mating-type chromosomes, that make this genus also an excellent model

25   for the evolution of mating systems and their influence on population-genetic structure and

26   adaptive potential. Important insights into the adaptation, coevolution, host specialization and

27   mating system evolution have been gained using anther-smut fungi, in particular with the

28   recent advent of genomic approaches. We argue and illustrate based on the *Microbotryum*

29   case studies that using a combination of genomic analyses is a powerful approach, where

30   comparative genomics, population genomics and transcriptomics data allow the integration of

31   different evolutionary perspectives and across timescales. We also highlight current

32   challenges and future studies that will contribute to advance our understanding of mechanisms

33   involved in adaptive processes in fungal pathogen populations.

34

37

38   **Introduction**

39

40   Pathogens thrive using living organisms as nutritional resources, which reduces their host

41   fitness. This leads to coevolutionary arms races, in which pathogens are selected for increased

42   abilities of host infection and exploitation, while hosts are selected for mechanisms of

43   resistance to particular diseases. Such coevolution occurs on short evolutionary scales, as a

44   never-ending process of adaptation and counter-adaptation (113). Across macro-evolutionary

45   scales, some pathogens also may undergo host shifts, forming new species by specialization

46   in combination with new hosts (42). Coevolution is a very different evolutionary process

47   from host specialization, despite the terms often being used interchangeably, and may involve

48   different genomic mechanisms and/or molecular interactions that have yet to be well resolved

49   (42).

50

51   An integrated understanding of the ecological and genetic/genomic mechanisms underlying

52   both coevolution and host specialization by pathogens is of fundamental importance. These

53   phenomena indeed represent cases of rapid adaptation, diversification and long-term species

54   interactions, shedding light on the processes generating and maintaining biodiversity and

55   ecosystem dynamics. Furthermore, knowledge on the genomic mechanisms involved in

56   coevolution and host specialization in fungal pathogens is important for controlling crop and

57   animal diseases and preventing emerging diseases that are a rising threat in domestic and wild

58   populations (52, 70). Fungi are the most important plant pathogens, causing dramatic crop

59   diseases, including many devastating diseased that are newly emergent following host shifts

60   (9, 45, 51).

61

3

62   Fungal pathogens also have to cope with their abiotic environment, such as temperature and

63   humidity (2, 35, 44, 47, 129). Understanding the mechanisms of adaptation to climatic

64   variables is thus similarly of fundamental and applied interest. Adaptation ability is however

65   impacted by genetic diversity and gene flow, which are themselves influenced by dispersal

66   rates and mating systems, that are therefore important life history traits to study for an

67   integrated understanding of evolution, adaptation, population subdivision and speciation (18,

68   60, 63).

69

70   From the advent of modern genetics a century ago, anther-smut fungi (*Microbotryum*

71   *violaceum* species complex, previously *Ustilago violacea*) have served as useful models for

72   the molecular controls of mating and adaptations to abiotic conditions (1, 17, 23, 29, 61, 62,

73   73, 93, 123, 124). With advances in population genetics and genomics, emphasis has grown

74   with regard to the natural diversity within this pathogen group, the dynamics of diseases, the

75   mating systems and genetic differentiation in relation to host plants in natural ecosystems (2,

76   27, 32, 38, 39, 58, 65, 67, 97, 105, 111, 131, 132). The anther-smut fungi belong to the

77   *Microbotryum* genus (basidiomycetes), which castrate plants of the Caryophyllaceae family,

78   replacing the pollen with their spores and aborting ovaries (Figure 1B). They constitute an

79   excellent model pathosystem, with hundreds of closely related fungal species specialized on

80   different host plants, resulting from numerous host shifts, with conspicuous symptoms, a rich

81   scientific history and occurring in natural ecosystems (Figure 1A) (83, 92, 97, 111, 118).

82   Furthermore, they are phylogenetically close to the rust fungi as damaging crop pathogens

83   (127). Most *Microbotryum* species are highly host-specific, but a few are more generalist,

84   parasitizing closely related host species (Figure 1A) (98, 105). Other *Microbotryum* species,

85   while distantly related, co-occur on the same host species, representing cases of convergence

86   (Figure 1A) (2, 97).

4

87

Host-pathogen coevolution in the *Microbotryum* model systems has been suggested based on patterns of plant local adaptation (43, 49) and congruent plant-pathogen genetic structure (49). *Microbotryum* species show little pre-zygotic isolation and increasing post-zygotic isolation strength with phylogenetic distance (28, 41, 98), which may allow gene flow among closely related species. Moreover, abiotic factors have been shown to play a role with the disease interactions in important ways (2), and *Microbotryum* fungi display an interesting mating system, with predominant automixis (i.e., intra-tetrad selfing), which has fostered multiple chromosomal rearrangements across the genus linking the mating-type loci controlling gamete compatibility (25, 26). A consideration of these features altogether allows the studying adaptation, coevolution, host specialization, differentiation and mating systems with unique power.

For tackling this complex suite of questions, comparative genomics and population genomics constitute highly relevant and complementary approaches, addressing different time scales of evolution. In contrast to life history traits, ecology and population structure, which have been extensively studied (7, 10, 22, 90), the genetic basis of interactions between *Microbotryum* fungi and their hosts is still little known; genomic approaches can elucidate the mechanisms and the functions involved in adaptation, coevolution, host specialization and speciation in this pathosystem. Analyses of gene expression between different stages of the life cycle can also inform on these processes. In particular, the pathogen's mating systems also influence adaptation, coevolution and host specialization (63, 66), especially in anther-smut fungi that are obligately completing the sexual cycle upon every disease transmission. Genomics can further help to understand the evolution of mating systems by studying the changes at the mating-type loci.

112

113   In this review, we discuss the recent insights into our understanding of adaptation,

114   coevolution and host specialization in anther-smut fungi gained from gene expression data

115   and comparative genomics (part 1) and from population genomics (parts 2 & 3). We then

116   discuss insights gained from genomics on mating system evolution (part 4). We illustrate that

117   the combination of multiple genomic approaches is needed for a full understanding of

118   evolution, as comparative genomics, population genomics and transcriptomics address

119   different timescales and have power for detecting different footprints of adaptive events.

120   Finally, future challenges to be addressed using genomics tools are discussed (part 5).

121

122   **1- COMPARATIVE GENOMICS AND TRANSCRIPTOMICS APPROACHES TO**

123   **UNDERSTAND ADAPTATION AND HOST SPECIALIZATION IN ANTHER-SMUT**

124   **FUNGI**

125   The sequencing of genomes and transcriptomes of *Microbotryum* species sheds light on

126   pathogenicity, adaptation and specialization mechanisms across long evolutionary timescales;

127   speciation events in castrating *Microbotryum* fungi have been dated from 0.4 to 11 MYA (26,

128   72) (Figure 2A). Phylogenomics enables obtaining an accurate understanding of the lineage

129   histories, and comparative genomics is highly suitable to identify genetic changes associated

130   with diversification at such large evolutionary scales.

131

132   **Genome architecture and identification of candidate genes involved in pathogenicity**

133   **using expression data**

134   One of the best studied anther-smut species is *M. lychnidis-dioicae*, parasitizing the white

135   campion *Silene latifolia* (Figure 1B). The diploid genomes of the Lamole *M. lychnidis-dioicae*

136   strain was the first eukaryote genome to be assembled with new sequencing technologies (16).

137 Comparative analysis of the Lamole strain of *M. lychnidis-dioicae* with other basidiomycetes

138 genomes revealed specific gene content features such as the absence of plant cell wall

139 degrading enzymes and expanded repertoires of major facilitator superfamily transporters,

140 secretory lipases, glycosyltransferases and enzymes that could manipulate host development

141 (104). Such features are likely related to the castrating and biotrophic lifestyle of anther-smut

142 fungi (104), where the fungus takes up a largely symptomless residence between the host cells

143 in the plant's growing points/meristems until the host initiates flower development.

144 Additionally, this pathogen has a remarkable ability to developmentally transform female host

145 plants to take on a male-like floral structure, with the growth of stamens that then bare spores

146 in place of pollen and the abortion of the ovary early in its development (13). Apart from the

147 accumulation of transposable elements (TEs) in the non-recombining regions of the mating-

148 type chromosomes, there was no genome compartmentalization into more or less repeat-rich

149 regions on autosomes (16, 104), in contrast to some other fungal pathogens with isochore

150 genomic architecture and localization of effector genes in repeat-rich regions (74).

151 Nevertheless, transposable elements were locally associated across the Lamole *M. lychnidis-*

152 *dioicae* genome with gene clusters of small secreted proteins and genes affected by within

153 species presence-absence polymorphism, suggesting a role of transposable elements in

154 genome rearrangements and duplications of genes putatively involved in host adaptation (80,

155 104). Although footprints typical of genome defense mechanisms against TEs, similar to

156 repeat-induced point mutation (RIP), were identified in anther-smut genomes, a massive

157 burst-like expansion of Gypsy-like retrotransposons in a *Microbotryum* strain suggested that

158 persistent transposable elements activity and expansion can occur (86, 87).

159

160 Transcriptomics conducted at several *in vitro* stages allow detecting genes upregulated in

161 certain conditions and thus likely involved in important functions at a given life stage.

162 Transcriptomic analyses using the Lamole *M. lychnidis-dioicae* strain enabled identifying

163 genes likely associated with nutrient uptake, the mating program and the dikaryotic switch

164 (54, 104, 125, 126, 136). *In silico* effector gene prediction combining *in planta* expression

165 data, sequence conservation and predicted localization, allowed identifying small secreted

166 proteins genes as candidate effectors, i.e. involved in pathogenicity, in *M. lychnidis-dioicae*,

167 *M. silenes-dioicae* and *M. violaceum var paradoxa* (20, 96). Eight genes in *M. silenes-dioicae*

168 and three genes in *M. violaceum var paradoxa* predicted to encode secreted proteins were

169 further confirmed to be secreted using yeast secretion trap (20, 96). Compared expression data

170 in male and female *S. latifolia* individuals during fungal infection revealed pathogen-mediated

171 changes in sex-biased gene expression and altered sexual dimorphism in the host (137).

172 Another transcriptome analysis of the early development stages of infected flowers detailed

173 changes in gene expression in *M. lychnidis-dioicae,* identifying gene categories likely to

174 manipulate the host development and reproductive system, such as potential effectors and

175 virulence factors (125). Further coupling experiments of host and pathogen gene expression

176 changes, and in further paired host-*Microbotryum* fungi, should help deciphering the major

177 components of the tight host-pathogen interactions described in the system.

178

179 **Comparative genomics studies within the *Microbotryum* genus**

180 Comparative genomics among *Microbotryum* fungi, and with other plant pathogens, has

181 provided insights into the specificity of castrating biotrophic pathogens growing

182 intracellularly (115) relative to other forms of parasitic nutritional ecology. Comparative

183 genomics among anther-smut fungi specialized on different hosts can help unravel the

184 genomic determinants of host specificity as well as the shared pathogenicity mechanisms.

185 Indeed, while substantial insights has been gained by the study of individual genomes of

186 *Microbotryum* species, whether features such as the conspicuous lack of cell-wall degrading

187   enzymes in the *M. lychnidis-dioicae* Lamole genome are common to the genus cannot be

188   known without a comparative genomics analysis that addresses both distantly and closely-

189   related species (77).

190

191   Early comparative studies focused on orthologous genes across single pass Sanger-sequenced

192   cDNA libraries, i.e. expressed sequence tags, from four *Microbotryum* species. The primary

193   focus was looking for signals of positive selection in terms of frequent amino-acid changes

194   (4). A subset of the genes evolving under positive selection between species was further

195   shown to be under strong purifying selection within two closely-related *Microbotryum*

196   species, *M. lychnidis-dioicae* and *M. silenes-dioicae*, suggesting that adaptive changes

197   concomitant with host shifts can be later fixed due to strong functional constraints within

198   species (69). Although the inferred function of some of the orthologous groups with signals of

199   positive selection could be associated with aspects of virulence or speciation, none of these

200   displayed features of effectors (such as secretory signals), likely because the expressed

201   sequence tags did not exhaustively cover the genomes. Indeed, only 53 clusters of orthologs

202   shared by at least three species and at least 300 nucleotides long could be retrieved (4).

203   Therefore, even though these analyses demonstrated the utility of comparative genomics to

204   identify candidate genes for diversifying selection in non-model organisms, the lack of whole

205   genome sequences prevented any insight about presence-absence polymorphisms or

206   substitutions both known to be important for adaptation to new hosts.

207

208   The number of high-quality genomes assemblies or shotgun sequencing from *Microbotryum*

209   species/strains has exploded recently, reaching nearly a hundred as by late 2018 (Table 1;

210   Figure 1A; (15, 16, 25, 26, 30, 56, 112, 134)). In comparative genomics, near-complete gene

211   lists can be clustered to obtain groups of homologous sequences that can be then used to build

212　phylogenetic profiles of gene content. Such comparisons allow the identification of gene

213　families that are species-specific and those that have been expanded or reduced in particular

214　lineages (8, 76). Species- or population-specific genes are either derived from within-group

215　innovation, a rather uncommon phenomenon (31), the result of differential losses or gene

216　duplications (78), or due to the non-vertical acquisition of gene-coding genome fragments, for

217　instance horizontal gene transfer (50). Expanded gene families require the escape from the

218　rampant pseudogenization (non-functionalization) of duplicated genes (99), whereas reduced

219　or complete losses of gene families is often related to ecological shifts (117), rendering the

220　product of those genes no longer needed for survival (5). Understanding these processes is

221　fundamental to the study of evolutionary ecology as they help to explain the genomic

222　architecture underlying the phenomenon of adaptive divergence. *In silico* annotations and

223　comparative analyses have identified hundreds of candidate effectors across multiple

224　*Microbotryum* species, enriched in gene families showing presence-absence polymorphism

225　across species (Figure 2B) (112), along with orthologous genes with landmarks of positive

226　selection between species and purifying selection within species (20), thus generalizing and

227　expanding previous findings. High-quality genome assemblies revealed little genomic

228　rearrangements in autosomes (26).

229

230　Studies of positive selection based on the comparisons of non-synonymous and synonymous

231　substitution rates (dN/dS; (135)) and on the comparisons of the proportions of non-

232　synonymous and synonymous polymorphisms within species and differences between species

233　(McDonald and Kreitman test; (100)) revealed no signature of diversifying selection between

234　sister *Microbotryum* species specialized on two closely related host species (15), but detected

235　a dozen of genes encoding secreted proteins with signs of positive selection between more

236　distantly related *Microbotryum* species specialized on more distant host species (Figure 2C)

10

237 (20). Future comparative genomics studies encompassing all currently sequenced genomes

238 will likely have high power to detect genes involved in host specialization by allowing further

239 disentangling the effects of pathogen and host phylogenetic distances. In particular,

240 combining population and comparative analyses should be very powerful to identify genes

241 under diversifying selection between species and purifying selection within species as well as

242 species-specific gene gains and losses. Building gene genealogies based on whole genomes

243 also allowed to resolve previously ambiguous relationships among some anther-smut species

244 (Figure 1A). The comparison of repeat contents and genomic rearrangements between

245 genomes will be a further key step to understand the role of genome dynamics in adaptive

246 processes in anther-smut fungi.

247

248
249 **2-POPULATION GENOMICS TO IDENTIFY ADAPTIVE GENETIC VARIATION**

250 **IN NATURAL PATHOGEN POPULATIONS**

251 Population genomics is a complementary approach to comparative genomics for

252 understanding adaptation in pathogen populations. Population genomics indeed address more

253 recent adaptive events, and on a broader range of evolutionary genetic phenomena, not only

254 gene gains/losses and recurrent changes in amino-acids. Selective sweeps can be detected

255 using population genomics, which can reveal positive selection on a single amino-acid change

256 or basepair substitutions in non-coding regions. Furthermore, population genomics can

257 address the questions of the genomic bases of host-pathogen coevolution and local adaptation,

258 that constitute more recent selection compared to the long-term selection underlying host

259 specialization, and possibly differential selection among geographically distant populations

260 (Figure 2A) (36, 75, 107). In contrast to major fungal-plant pathosystems, no gene-for-gene

261 relationship has been reported for anther-smut fungi. Instead, the probability of infection

262 shows quantitative variation (6, 7, 33), which suggests a rather complex genetic basis of co-

263   evolution and host local adaptation. Genome-wide population genomics approaches in anther-

264   smut fungi allowed identification of the complex genetic basis of recent adaptive events

265   through genome scans of selective sweeps and gene-presence absence polymorphism (3, 15,

266   80).

267

268   Selective sweep analyses allow one to identify loci that have recently been under positive

269   selection within populations and thus likely underlying coevolution and local adaptation,

270   whereas genes involved in host specialization are likely under purifying selection within

271   species after the initial adaptive events following host shifts. Analyses of whole genome

272   sequences of 53 genomes of the anther-smut sister species *M. lychnidis-dioicae* and *M.*

273   *silenes-dioicae* identified selective sweeps (Figure 2D) (15), likely resulting from dynamic

274   co-evolutionary arm race of the fungus with its hosts. The overlap between genes

275   differentially expressed *in planta* and *in vitro* and those lying within selective sweeps,

276   together with functional annotations, provided clues to genes and functions involved in plant-

277   pathogen interaction in the *Microbotryum-Silene* system. Candidate genes included glycoside

278   hydrolases, pectin lyases and an extracellular membrane protein with CFEM domain (15).

279   The pectin lyase function seems relevant in that *Microbotryum* fungi grow between cells of

280   the meristem (115), which is a pectin-occupied space. Extracellular membrane proteins with a

281   cysteine-rich CFEM domain are present in effectors in several fungal pathogens (95). This

282   study was also an opportunity to test for differences in intensity of coevolution between

283   anther smut fungi on different hosts. Interestingly, differences in the number and the location

284   of the selective sweeps were found between sister species. Footprints of positive selection

285   affected 17 % of the genome in *M. lychnidis-dioicae* and 1 % of the genome in *M. silenes-*

286   *dioicae* (15). Selective sweeps were scattered throughout the genomes. Linkage

287   disequilibrium was found to decay relatively slowly with physical distance along

288   chromosomes, as expected for selfing species, but still indicated effective recombination.

289   Polymorphism in each fungal species was negatively correlated with the recombination rates

290   along chromosomes, consistent with recurrent positive and/or background selection erasing

291   diversity on larger genomic regions when recombination is less frequent (15).

292

293   Population genomics can also contribute to our understanding of the impact of recent

294   anthopogenic factors on the genome and subsequent adaptation. Analyses of *M. lychnidis-*

295   *dioicae* genomes along a gradient of ionizing radiation levels around Chernobyl showed no

296   evidence of deleterious mutation accumulation in the form of non-synonymous substitutions

297   (3). Lower mean values of dN/dS were even found in Chernobyl compared to other areas of

298   the same eastern genetic cluster (3), which may be due to stronger selection in contaminated

299   areas against individuals bearing mildly deleterious mutations, i.e. stronger purifying

300   selection.

301

302   In addition to genome scans looking at signatures of positive selection, other population

303   genomic approaches make use of the genetic variation in pathogen populations to identify the

304   genomic architecture of local adaptation (19, 36, 107, 114). Population genomics enables the

305   unravelling the genomic bases of adaptation to abiotic conditions by searching for correlations

306   between local population allele frequencies and local environments (genetic-environment

307   association methods) (82). Such approaches can be used in anther-smut fungi along altitudinal

308   clines in Alpine populations on *Dianthus* or *Silene* hosts. Studies on the three species

309   parasitizing *S. vulgaris* in particular could be interesting as elevation and climate has been

310   shown to impact these anther-smut fungi (1, 2). Strong population structure as found in many

311   *Microbotryum* species at European scale (2, 15, 27, 55, 105) might be a challenge to the use

312   of such methods in particular, but these methods can be utilized at small geographical scales

313   and/or in species with less population subdivision.

314 Gene copy number variation segregating within species is also a widespread and an important

315 source of genetic variation and several examples of adaptive evolution through gene loss or

316 gene gain have been identified in agricultural fungal plant pathogens (57). Population

317 genomics allow to explore the extent and adaptive potential of such within-species variation.

318 Gene presence/absence polymorphism was found to contribute to the genetic variation in

319 populations of the two closely related species of castrating anther-smut fungi, *M. lychnidis-*

320 *dioicae* and *M. silenes-dioicae* (80). Genes displaying presence/absence polymorphism were

321 mostly recently acquired, in a single species, through duplications in multiple-gene families

322 and few genes predicted to encode secreted proteins were affected, suggesting defense against

323 host recognition by other genetic changes than gene loss or gain. Although most gene

324 presence/absence polymorphisms were likely neutral, the putative functions of some genes

325 affected by presence–absence polymorphism (e.g., secreted proteins) or their localization

326 within previously identified selective sweeps suggested that some gene loss or gain events

327 may be adaptive (80).

328

329 **3-INSIGHTS INTO THE DYNAMICS OF DIVERGENCE AND GENE FLOW FROM**

330 **POPULATION GENOMICS**

331 By providing a glimpse into intra-specific genetic diversity and its variation across the

332 genome, population genomics analyses are also highly useful to understand processes

333 underlying species divergence and phylogeography, quantifying rates of gene flow and its

334 heterogeneity along genomes, and providing accurate estimation of population size variations.

335 The occurrence of multiple *Microbotryum* sister species pairs in sympatry makes the system a

336 perfect model to study the dynamics of divergence and gene flow in fungal pathogen

337 populations.

338

Contrasted patterns of interspecific gene flow in the *Microbotryum* genus: a speciation continuum?

The two pathogens *M. lychnidis-dioicae* and *M. silenes-dioicae* and their respective sister host plants, *Silene latifolia* and *S. dioica*, are ubiquitous in Europe and their geographic distributions are largely overlapping, providing an ideal system for research on the formation and maintenance of species in sympatry. Microsatellite data from samples across Europe revealed rare disease transmission events between the host species and rare pathogen hybrids (72, 132). However, these approaches using a dozen microsatellite markers may lack power. Analyses of whole genome sequences of many pathogen samples that appeared of pure ancestry based upon the microsatellite data then revealed no evidence for admixture, indicating that introgression does not persist beyond one or two generations (15). In the laboratory, both fungal species can infect both host plants (40, 64, 98). Experimental crosses showed little premating isolation by assortative mating between the two pathogen species (28, 98, 132), even at sympatric sites (110), and a lack of post-mating barriers (41, 98). Hybrids were viable and fertile at least through the F2 generation in the greenhouse (40, 98, 131). F2 hybrids produced by selfed F1s had mostly returned to homozygosity, suggesting that genomic content derived from one of the two parental species had already begun to be purged (28, 40). This latter finding, combined with the fact that introgression does not appear to persist in nature, is consistent with strong genome-wide selection by the host plant and the scattering of genes involved in host specialization across the genome, as revealed in genome scans of selective sweeps (15). F_{ST} values were found near their maximum all along their genomes (Figure 3A).

Whereas strict host specialization is often the rule on *Silene* species (15, 97, 133), on *Dianthus* hosts in contrast population genetics approaches revealed four *Microbotryum*

364   lineages with broader and overlapping host specificities (Figure 1A) (97, 105). One

365   *Microbotryum* lineage was found only on *D. pavonius* while the others occurred spread across

366   several host species, some of them being shared among *Microbotryum* lineages. The sympatry

367   of *Microbotryum* lineages within populations, in particular in the Alps, led to hybridization

368   (105). The individuals with mixed ancestry based on clustering analyses of microsatellite data

369   suffered from significant meiotic sterility, which confirmed they were hybrids between

370   species (105). The larger host ranges of *Microbotryum* lineages on *Dianthus* hosts may be

371   explained by the recent divergence of their host plants. The *Dianthus* genus has indeed

372   undergone a recent radiation in Europe with morphologically diverse European *Dianthus*

373   species restricted to small geographically restricted ranges (130). The full extent and

374   evolutionary consequences of the hybridization on pathogen dynamics and evolution remains

375   to be explored. Along this line, the *Dianthus-Microbotryum* system may become, in the

376   coming years, a tractable model to investigate the impact of gene flow during divergence, and

377   whether selection due to local/host adaptation can make some genomic regions more or less

378   permeable to gene flow, which represents a current debate in evolutionary biology (37). These

379   questions could not be addressed so far based on the population genomics analyses of *M.*

380   *lychnidis-dioicae* and *M. silenes-dioicae* as no genomic introgression could be detected in

381   natural populations (15). In contrast, the hybrids detected in natural populations on *Dianthus*

382   hosts with significant sterility suggest the occurrence of introgressions (105). Other pairs of

383   *Microbotryum* species might also be suitable to address these questions of the impact and

384   heterogeneity of gene flow along the genome. For example, anther-smut fungi on the closely

385   related and sympatric native American species *S. virginica* and *S. caroliniana* (11, 12) could

386   not be separated into host-specialized species based on a few gene genealogies (58, 92, 111).

387   In this system, population genomics should allow elucidating whether anther-smut fungi on

388   these American *Silene* species show host differentiation or genome-wide gene flow, or

389   introgression only in genomic regions not involved in host specialization. Based on the few

390   genomes available so far (26), we find F_{ST} values between *Microbotryum* populations on the

391   two hosts, *S. virginica* and *S. caroliniana*, that are lower and more heterogeneous along the

392   genomes than between *M. lychnidis-dioicae* and *M. silenes-dioicae* referenced above (Figure

393   3B). These initial results suggests the occurrence of gene flow in some genomic regions. The

394   situation of anther-smut fungi on *S. vulgaris*, with three distant lineages with convergent

395   specialization on this same host species (2), would also be worth exploring using population

396   genomics to determine the extent of introgression and its genomic localization, and whether

397   the interspecific exchange of alleles has been deleterious or adaptive.

398

399   Another promising approach in anther-smut fungi for identifying genomic regions involved in

400   host adaptation will be to perform genome scans of differentiation between closely related

401   species or host races, if possible to avoid the potential pitfalls of such approaches (37). This

402   could contribute to our understanding of the role of gene flow in the early stages of

403   divergence and to identifying genomic regions less permeable to gene flow because of

404   selection for host adaptation and/or genetic incompatibities between lineages (24, 37). More

405   generally, such population genomics approaches would be valuable to use in plant pathogen

406   fungi.

407

408   **Phylogeography and demographic history inferences**

409   *Microbotryum lychnidis-dioicae* and *M. silenes-dioicae* also constituted case studies in

410   providing one of the most clear-cut examples of phylogeographic structure in pathogens,

411   thanks to a collection of samples whose density and geographical scale was unprecedented for

412   a disease association in natural populations. In *M. lychnidis-dioicae*, clustering analyses based

413   on microsatellite markers (133), as well as nuclear gene sequences (69, 72), revealed the

existence of three genetically distinct clusters, reflecting recolonization from well-recognized southern refugia after glaciation. Little admixture has been found between clusters based on microsatellites (49, 133), and this has later been confirmed by whole genome sequences (15). Indeed, SNPs (single nucleotide polymorphisms) revealed few shared polymorphisms and many fixed differences among the clusters, and pairwise F_{ST} values between them were high (0.56–0.74; Figures 3C and D), supporting low levels of inter-cluster gene flow (15). Whole genome sequences provided further insights into the age of divergence between the three *M. lychnidis-dioicae* lineages (Southern, Western and Eastern clusters), sequential size changes in the population size of derived lineages and also supported low levels of gene flow (15). Most notably, the pathogen genetic structure closely matched with the genetic structure of the host species *S. latifolia* with the same regionally defined Southern, Western and Eastern clusters, indicating that the anther-smut pathogen remained during the last glaciation in the same three distinct refugia as its host (i.e. in the Iberian, Italian and Balkan peninsulas) (49). The congruence of population structures between M. *lychnidis-dioicae* and its host appeared even stronger than what could be expected because of isolation by distance alone, suggesting that coevolution has played a significant role in the congruence of the population structures (49). Genome-wide gene presence-absence polymorphism recovered the same population structure (80). Inoculation experiments, indicating plant local adaptation for resistance to pathogens (49, 89, 91), were consistent with a contribution of adaptive factors to the observed congruence between pathogen and host population structures.

Microsatellite markers and genome-wide SNPs indicated that *M. silenes-dioicae* also exhibited a genetic structure, albeit with biogeographic patterns more difficult to interpret (15, 133) and very low F_{ST} values genome-wide (Figure 3E). Genome-wide gene presence/absence polymorphism revealed two different clusters with a more obvious east/west separation (80),

439 that may correspond to local adaptation of *S. dioica* clusters (81, 109), although this remains

440 to be assessed. This case study shows the power of various kinds of population genomic

441 studies to unravel weak and/or adaptive population subdivision.

442

443 **4-UNRAVELLING MATING SYSTEM AND GAMETE COMPATIBILITY**

444 **SYSTEMS USING BOTH COMPARATIVE GENOMICS AND POPULATION**

445 **GENOMICS**

446 The combination of comparative genomics and population genomics also can reveal

447 remarkable transitions in mating systems by elucidating the changes in genomic mechanisms

448 controlling mating compatibility. For the broad group of basidiomycete fungi, gamete

449 compatibility is controlled by two loci acting at the haploid stage, mating being successful

450 only between haploid cells carrying different alleles at both mating-type loci (34). The two

451 mating-type loci are i) the PR locus which encodes pheromone genes and a pheromone

452 receptor gene implicated in gamete recognition and fusion, and ii) the HD locus which

453 encodes homeodomain protein-coding genes allowing, after fusion, for the maintenance of the

454 dikaryon and hyphal growth (48, 94). Most basidiomycetes are outcrossing and have these

455 two loci unlinked, although some fungi in this group have linked mating-type loci (103).

456 Linkage of the two mating type loci is considered to be favored due to increased odds of

457 gamete compatibility under selfing when mating-type loci are linked (103). Interestingly,

458 most *Microbotryum* species are highly selfing and were long known to segregate only two

459 mating type phenotypes, but it remained uncertain whether this was due to mating-type loci

460 linkage or to the loss of role in mating-type determinism for one of the two mating-type loci,

461 as both cases occurred in basidiomycetes (85, 103). Comparative genomics of well-assembled

462 genomes allowed to resolve the complex genome architecture and long-term evolutionary

463 history of the repeat-rich and rearranged mating type chromosomes in anther-smut fungi, and

464   population genomics datasets were essential for identifying young events of recombination

465   suppression.

466

467   Population genomics confirmed high rates of selfing in all studied *Microbotryum* species, by

468   showing high levels of genome-wide homozygosity (15, 25, 26) and confirmed massive

469   recombination suppression on mating-type chromosomes (84, 134). High-quality genomes

470   assemblies allowed reconstructing the history of genomic events underlying the shift in

471   gamete compatibility system (25, 26). The long-read sequencing technology allowed

472   assembling the two repeat-rich mating-type chromosomes of the Lamole *M. lychnidis-dioicae*

473   strain, which confirmed linkage between the two mating-type loci HD and PR (Figure 4A)

474   (16). Genome comparisons between multiple *Microbotryum* species showed that the ancestral

475   state had unlinked mating-type loci on two distinct chromosomes, and that independent

476   rearrangements and chromosome fusions occurred in multiple species, convergently linking

477   the two mating-type loci by large regions without recombination (Figure 4B) (25, 26). This

478   shows that natural selection can repeatedly lead to similar phenotypes through multiple

479   different evolutionary pathways.

480

481   Following recombination suppression, a chaos of rearrangements occurred on mating-type

482   chromosomes (16), as well as TE and non-synonymous substitution accumulation (16, 54), as

483   is typical in non-recombining regions, and in particular on sex chromosomes (14). The high-

484   quality assemblies allowed the detailed characterization of extensive rearrangements and

485   repeat accumulations on the two mating-type chromosomes (16, 25, 26). Another important

486   characteristic feature of sex chromosomes was observed on the mating-type chromosomes of

487   multiple *Microbotryum* species, i.e., the stepwise extension of the regions with recombination

488   suppression. The progressive extension of the regions without recombination revealed a

489   pattern of clear "evolutionary strata", i.e., decreasing divergence between alleles on the

490   alternative mating-type chromosomes farther from the mating-type loci (Figure 4C).

491   Population genomics was essential for providing evidence of early events of recombination

492   suppression in several species, by showing the segregation of alleles according to their

493   associated mating-type, decreased levels of diversity as expected under lower population

494   effective sizes and that high divergence between alleles associated with the alternative mating

495   types was due to balancing selection on mating types rather than elevated substitution rates

496   (25, 26). Indeed, as soon as recombination ceased, alleles on the two mating-type

497   chromosomes diverged gradually with time (Figure 4D). Finding such evolutionary strata in

498   fungi, which lack male and female roles, challenged the classical view for the evolution of sex

499   chromosomes. Indeed stepwise recombination suppression in sex chromosomes was thought

500   to be due primarily to sexual antagonism, i.e., the selection to link genes with alleles

501   beneficial in one sex, and deleterious in the other, to the sex determining gene (21). The

502   finding of evolutionary strata in fungi without sexual antagonism indicates that alternative

503   hypotheses should be explored to explain the progressive spread of recombination

504   suppression, such as overdominance, epigenetic modifications associated with transposable

505   elements or neutral rearrangements (88, 108).

506

507

508   **5-CONCLUDING REMARKS AND PERSPECTIVES**

509   A thorough understanding of the major roles played by pathogens requires the integrative

510   study of both ongoing processes of coevolution and dynamics of specialization that impact the

511   emergence new diseases. Investigations of the anther-smut fungi utilizing comparative

512   approaches to genomics and gene expression profiles, combined with population-level

513   studies, illustrate the strength of combining different genomic approaches addressing different

514   scales of evolution (Figure 2). The availability of genomic data for multiple sister species and

515   multiple populations within species makes the anther-smut system quite exceptional for

516   identifying the genetic mechanisms involved in adaptation, coevolution, host specialization

517   and mating system at different evolutionary times (Table 1; Figure 1). Comparative genomics

518   has long been the predominant approach for studying adaptation in plant fungal pathogens

519   (46, 53, 59, 102, 107) and has provided important insights into the mechanisms of adaptation,

520   e.g., through horizontal gene transfers, gene gains/losses, hybridization or recurrent amino-

521   acid changes (71). Comparative genomics by definition does not consider population-level

522   variation, such that population genomics is a complementary approach for insights into

523   evolutionary processes acting at the local and regional scales. For example, several recent

524   studies have revealed gene presence/absence polymorphism within species (78, 80, 119, 120).

525   In addition, comparative genomics can only detect a specific type of positive selection,

526   involving frequent changes of amino-acids. Positive selection of a single amino-acid change

527   or of regulatory regions can only be detected by looking for selective sweeps using population

528   genomics. Some recent studies based on population genomics have in fact revealed important

529   aspects of adaptation in fungal plant pathogens, showing footprints of introgression, selective

530   sweeps and amino acid-changes (68, 79, 101, 116, 121, 122).

531

532   Furthermore, cross referencing candidate genes that are highlighted by multiple indications of

533   being subject to natural selection during parasitism as outlined here (e.g. genes found within a

534   selective sweep, upregulated in the plant and having experienced gene family expansion

535   compared to other fungal pathogens) can strengthen their putative roles as pathogen effector

536   that are central in the specificity of fungal-plant combinations. Functional studies can help

537   understanding the role of the candidate genes. Promising transformation protocol have been

538   developed in *M. lychnidis-dioicae* (128) and will likely facilitate the characterization of key

539　genes involved in the interaction between the anther-smut fungi and their Caryophyllaceae

540　host plants, an important challenge for the coming years. Transcriptomes and epigenomes of

541　multiple *Microbotryum* species and multiple strains within species will likely be

542　complementary to the current available genomic ressources to identify the role of regulatory

543　and epigenetic mechanisms in the adaptation of anther-smut fungi to their hosts and

544　environment, their divergence and their mating-type chromosome organisation, contributing

545　to further understanding the mechanisms involved in adaptive processes in fungal pathogen

546　populations. It will also be interesting to investigate similar levels of among and within

547　species sampling and genome sequencing using pathogens of different levels of obligate

548　parasitism, including facultative and opportunistic pathogens, as well as hemibiotrophy and

549　necrotrophy.

550

551

552

553　**ACKNOWLEDGMENTS**

558

559

560

REFERENCES :

562

563   1.   Abbate JL, Antonovics J. 2014. Elevational disease distribution in a natural plant–

564   pathogen system: insights from changes across host populations and climate. *Oikos*.

565   123(9):1126–36

566   2.   Abbate JL, Gladieux P, Hood ME, de Vienne DM, Antonovics J, et al. 2018. Co-

567   occurrence among three divergent plant-castrating fungi in the same *Silene* host

568   species. *Mol. Ecol.* 27(16):3357–70

569   3.   Aguileta G, Badouin H, Hood ME, Møller AP, Prieur SL, et al. 2016. Lower

570   prevalence but similar fitness in a parasitic fungus at higher radiation levels near

571   Chernobyl. *Mol. Ecol.* 25(14):3370–83

572   4.   Aguileta G, Lengelle J, Marthey S, Chiapello H, Rodolphe F, et al. 2010. Finding

573   candidate genes under positive selection in non-model species: examples of genes

574   involved in host specialization in pathogens. *Mol. Ecol.* 19(2):292–306

575   5.   Albalat R, Cañestro C. 2016. Evolution by gene loss. *Nat. Rev. Genet.* 17(7):379–91

576   6.   Alexander HM, Antonovics J. 1995. Spread of anther-smut disease (*Ustilago violacea*)

577   and character correlations in a genetically variable experimental population of *Silene*

578   *alba*. *J. Ecol.* 83(5):783–94

579   7.   Alexander HM, Thrall PH, Antonovics J, Jarosz AM, Oudemans PV. 1996. Population

580   dynamics and genetics of plant disease: a case study of anther- smut disease. *Ecology*.

581   77(4):990–96

582   8.   Ames RM, Money D, Ghatge VP, Whelan S, Lovell SC. 2012. Determining the

583   evolutionary history of gene families. *Bioinformatics*. 28(1):48–55

584 9.   Anderson PK, Cunningham AA, Patel NG, Morales FJ, Epstein PR, Daszak P. 2004.

585   Emerging infectious diseases of plants: pathogen pollution, climate change and

586   agrotechnology drivers. *Trends Ecol. Evol.* 19(10):535–44

587 10.   Antonovics J, Alexander HM. 1992. Epidemiology of anther-smut infection of *Silene*

588   *alba* (= *S. latifolia*) caused by *Ustilago violacea* : patterns of spore deposition in

589   experimental populations. *Proc R Soc Lond B*. 250(1328):157–63

590 11.   Antonovics J, Hood ME, Thrall PH, Abrams JY, Duthie GM. 2003. Herbarium studies

591   on the distribution of anther-smut fungus (*Microbotryum violaceum*) and *Silene* species

592   (Caryophyllaceae) in the eastern United States. *Am. J. Bot.* 90(10):1522–31

593 12.   Antonovics J, Stratton D, Thrall PH, Jarosz AM. 1996. An anther-smut disease

594   (*Ustilago violacea*) of fire-pink (*Silene virginica*): Its biology and relationship to the

595   anther-smut disease of white campion (*Silene alba*). *Am. Midl. Nat.* 135(1):130–43

596 13.   Audran J-C, Batcho M. 1981. Microsporogenesis and pollen grains in *Silene dioica* (L.)

597   Cl. and alterations in its anthers parasitised by *Ustilago violacea* (Pers.) Rouss. *Acta*

598   *Soc Bot Pol*, pp. 29–32

599 14.   Bachtrog D. 2013. Y-chromosome evolution: emerging insights into processes of Y-

600   chromosome degeneration. *Nat. Rev. Genet.* 14(2):113–24

601 15.   Badouin H, Gladieux P, Gouzy J, Siguenza S, Aguileta G, et al. 2017. Widespread

602   selective sweeps throughout the genome of model plant pathogenic fungi and

603   identification of effector candidates. *Mol. Ecol.* 26(7):2041–62

604 16.   Badouin H, Hood ME, Gouzy J, Aguileta G, Siguenza S, et al. 2015. Chaos of

605   rearrangements in the mating-type chromosomes of the anther-smut fungus

606   *Microbotryum lychnidis-dioicae*. *Genetics*. 200(4):1275–84

607 17.   Baker HG. 1947. Infection of species of *Melandrium* by *Ustilago violacea* (Pers.)

608   Fuckel and the transmission of the resultant disease. *Ann. Bot.* 11(43):333–48

609   18.   Barrett LG, Thrall PH, Burdon JJ, Linde CC. 2008. Life history determines genetic

610   structure and evolutionary potential of host–parasite interactions. *Trends Ecol. Evol.*

611   23(12):678–85

612   19.   Bartoli C, Roux F. 2017. Genome-wide association studies in plant pathosystems:

613   toward an ecological genomics approach. *Front. Plant Sci.* 8:763

614   20.   Beckerson WC, Rodríguez de la Vega RC, Hartmann FE, Giraud T, Perlin Michael H.

615   2018. Cause and effectors: Secretome comparison of members from the anther-smut

616   pathogen species complex, *Microbotryum violaceum*. *In prep.*

617   21.   Bergero R, Charlesworth D. 2009. The evolution of restricted recombination in sex

618   chromosomes. *Trends Ecol. Evol.* 24(2):94–102

619   22.   Bernasconi G, Antonovics J, Biere A, Charlesworth D, Delph LF, et al. 2009. *Silene* as

620   a model system in ecology and evolution. *Heredity.* 103(1):5–14

621   23.   Biere A, Honders SC. 1998. Anther smut transmission in *Silene latifolia* and *Silene*

622   *dioica*: impact of host traits, disease frequency, and host density. *Int. J. Plant Sci.*

623   159(2):228–35

624   24.   Bierne N, Welch J, Loire E, Bonhomme F, David P. 2011. The coupling hypothesis:

625   why genome scans may fail to map local adaptation genes. *Mol. Ecol.* 20(10):2044–72

626   25.   Branco S, Badouin H, Rodríguez de la Vega RC, Gouzy J, Carpentier F, et al. 2017.

627   Evolutionary strata on young mating-type chromosomes despite the lack of sexual

628   antagonism. *Proc. Natl. Acad. Sci.* 114(27):7067–72

629   26.   Branco S, Carpentier F, Rodríguez de la Vega RC, Badouin H, Snirc A, et al. 2018.

630   Multiple convergent supergene evolution events in mating-type chromosomes. *Nat.*

631   *Commun.* 9(1):2000

632 27. Bueker B, Eberlein C, Gladieux P, Schaefer A, Snirc A, et al. 2016. Distribution and

633   population structure of the anther smut *Microbotryum silenes-acaulis* parasitizing an

634   arctic–alpine plant. *Mol. Ecol.* 25(3):811–24

635 28. Büker B, Petit E, Begerow D, Hood ME. 2013. Experimental hybridization and

636   backcrossing reveal forces of reproductive isolation in *Microbotryum. BMC Evol. Biol.*

637   13(1):224

638 29. Carlsson U, Elmqvist T. 1992. Epidemiology of anther-smut disease (*Microbotryum*

639   *violaceum*) and numeric regulation of populations of *Silene dioica. Oecologia.*

640   90(4):509–17

641 30. Carpentier F, Rodríguez de la Vega RC, Branco S, Snirc A, Coelho MA, et al. 2018.

642   Convergent recombination cessation between mating-type genes and centromeres in

643   selfing anther-smut fungi. *In prep.*

644 31. Casola C. 2018. From de novo to "de nono": The majority of novel protein coding

645   genes identified with phylostratigraphy are old genes or recent duplicates. *Genome*

646   *Biol. Evol.* 10(11):2906–18

647 32. Cattrall ME, Baird ML, Garber ED. 1978. Genetics of *Ustilago violacea.* III. Crossing

648   over and nondisjunction. *Bot. Gaz.* 139(2):266–70

649 33. Chung E, Petit E, Antonovics J, Pedersen AB, Hood ME. 2012. Variation in resistance

650   to multiple pathogen species: anther smuts of *Silene uniflora. Ecol. Evol.* 2(9):2304–14

651 34. Coelho MA, Bakkeren G, Sun S, Hood ME, Giraud T. 2017. Fungal sex: The

652   Basidiomycota. *Microbiol. Spectr.* 5(3):FUNK-0046-2016

653 35. Cordier T, Robin C, Capdevielle X, Fabreguettes O, Desprez-Loustau M-L, Vacher C.

654   2012. The composition of phyllosphere fungal assemblages of European beech (*Fagus*

655   *sylvatica*) varies significantly along an elevation gradient. *New Phytol.* 196(2):510–19

656   36.   Croll D, McDonald BA. 2017. The genetic basis of local adaptation for pathogenic
657   fungi in agricultural ecosystems. *Mol. Ecol.* 26(7):2027–40

658   37.   Cruickshank TE, Hahn MW. 2014. Reanalysis suggests that genomic islands of
659   speciation are due to reduced diversity, not reduced gene flow. *Mol. Ecol.*
660   23(13):3133–57

661   38.   Day AW, Garber ED. 1988. 30 - *Ustilago violacea*, anther smut of the
662   Caryophyllaceae. In *Advances in Plant Pathology*, ed GS Sidhu. 6:457–82. Academic
663   Press

664   39.   Day AW, Jones JK. 1968. The production and characteristics of diploids in *Ustilago*
665   *violacea*. *Genet. Res.* 11(1):63–81

666   40.   de Vienne DM, Hood ME, Giraud T. 2009a. Phylogenetic determinants of potential
667   host shifts in fungal pathogens. *J. Evol. Biol.* 22(12):2532–41

668   41.   de Vienne DM, Refrégier G, Hood ME, Guigue A, Devier B, et al. 2009b. Hybrid
669   sterility and inviability in the parasitic fungal species complex *Microbotryum*. *J. Evol.*
670   *Biol.* 22(4):683–98

671   42.   de Vienne DM, Refrégier G, López-Villavicencio M, Tellier A, Hood ME, Giraud T.
672   2013. Cospeciation vs host-shift speciation: methods for testing, evidence from natural
673   associations and relation to coevolution. *New Phytol.* 198(2):347–85

674   43.   Delmotte F, Bucheli E, Shykoff JA. 1999. Host and parasite population structure in a
675   natural plant–pathogen system. *Heredity.* 82(3):300–308

676   44.   Desprez-Loustau M-L, Capron G, Dupuis F. 1998. Relating germination dynamics of
677   *Melampsora pinitorqua* teliospores to temperature and rainfall during overwintering.
678   *Eur. J. For. Pathol.* 28(5):335–47

679   45.   Desprez-Loustau M-L, Robin C, Buée M, Courtecuisse R, Garbaye J, et al. 2007. The
680   fungal dimension of biological invasions. *Trends Ecol. Evol.* 22(9):472–80

681   46.   Duplessis S, Cuomo CA, Lin Y-C, Aerts A, Tisserant E, et al. 2011. Obligate biotrophy

682   features unraveled by the genomic analysis of rust fungi. *Proc. Natl. Acad. Sci.*

683   108(22):9166–71

684   47.   Enjalbert J, Duan X, Leconte M, HovmøLler MS, De Vallavieille-Pope C. 2005.

685   Genetic evidence of local adaptation of wheat yellow rust (*Puccinia striiformis* f. sp.

686   *tritici*) within France. *Mol. Ecol.* 14(7):2065–73

687   48.   Feldbrügge M, Kämper J, Steinberg G, Kahmann R. 2004. Regulation of mating and

688   pathogenic development in *Ustilago maydis*. *Curr. Opin. Microbiol.* 7(6):666–72

689   49.   Feurtey A, Gladieux P, Hood ME, Snirc A, Cornille A, et al. 2016. Strong

690   phylogeographic co-structure between the anther-smut fungus and its white campion

691   host. *New Phytol.* 212(3):668–79

692   50.   Feurtey A, Stukenbrock EH. 2018. Interspecific gene exchange as a driver of adaptive

693   evolution in fungi. *Annu. Rev. Microbiol.* 72(1):377–98

694   51.   Fisher MC, Hawkins NJ, Sanglard D, Gurr SJ. 2018. Worldwide emergence of

695   resistance to antifungal drugs challenges human health and food security. *Science.*

696   360(6390):739–42

697   52.   Fisher MC, Henk DA, Briggs CJ, Brownstein JS, Madoff LC, et al. 2012. Emerging

698   fungal threats to animal, plant and ecosystem health. *Nature.* 484(7393):186–94

699   53.   Floudas D, Binder M, Riley R, Barry K, Blanchette RA, et al. 2012. The paleozoic

700   origin of enzymatic lignin decomposition reconstructed from 31 fungal genomes.

701   *Science.* 336(6089):1715–19

702   54.   Fontanillas E, Hood ME, Badouin H, Petit E, Barbe V, et al. 2014. Degeneration of the

703   non-recombining regions in the mating-type chromosomes of the anther-smut fungi.

704   *Mol. Biol. Evol.* 32(4):928–43

705   55.   Fortuna TM, Namias A, Snirc A, Branca A, Hood ME, et al. 2018. Multiple infections,

706   relatedness and virulence in the anther-smut fungus castrating *Saponaria* plants. *Mol.*

707   *Ecol.* doi:10.1111/mec.14911:

708   56.   Fortuna TM, Snirc A, Badouin H, Gouzy J, Siguenza S, et al. 2016. Polymorphic

709   microsatellite markers for the tetrapolar anther-smut fungus *Microbotryum saponariae*

710   based on genome sequencing. *PLOS ONE*. 11(11):e0165656

711   57.   Fouché S, Plissonneau C, Croll D. 2018. The birth and death of effectors in rapidly

712   evolving filamentous pathogen genomes. *Curr. Opin. Microbiol.* 46:34–42

713   58.   Freeman AB, Kellye Duong K, Shi T-L, Hughes CF, Perlin MH. 2002. Isolates of

714   *Microbotryum violaceum* from North American host species are phylogenetically

715   distinct from their European host-derived counterparts. *Mol. Phylogenet. Evol.*

716   23(2):158–70

717   59.   Friesen TL, Stukenbrock EH, Liu Z, Meinhardt S, Ling H, et al. 2006. Emergence of a

718   new disease as a result of interspecific virulence gene transfer. *Nat. Genet.* 38(8):953–

719   56

720   60.   Gandon S, Michalakis Y. 2002. Local adaptation, evolutionary potential and host–

721   parasite coevolution: interactions between migration, mutation, population size and

722   generation time. *J. Evol. Biol.* 15(3):451–62

723   61.   Garber ED, Day AW. 1985. Genetic mapping of a phytopathogenic Basidiomycete,

724   *Ustilago violacea*. *Bot. Gaz.* 146(4):449–59

725   62.   Garber ED, Ruddat M. 2002. Transmission genetics of *Microbotryum violaceum*

726   (*Ustilago violacea*): a case study. *Advances in Applied Microbiology*, pp. 107–127

727   63.   Gibson AK, Hood ME, Giraud T. 2012. Sibling competition arena: selfing and a

728   competition arena can combine to constitute a barrier to gene flow in sympatry.

729   *Evolution*. 66(6):1917–30

730 64. Gibson AK, Refrégier G, Hood ME, Giraud T. 2014. Performance of a hybrid fungal
731   pathogen on pure-species and hybrid host plants. *Int. J. Plant Sci.* 175(6):724–30

732 65. Giraud T. 2004. Patterns of within population dispersal and mating of the fungus
733   *Microbotryum violaceum* parasitising the plant *Silene latifolia*. *Heredity*. 93(6):559–65

734 66. Giraud T, Gladieux P, Gavrilets S. 2010. Linking the emergence of fungal plant
735   diseases with ecological speciation. *Trends Ecol. Evol.* 25(7):387–95

736 67. Giraud T, Jonot O, Shykoff JA. 2005. Selfing propensity under choice conditions in a
737   parasitic fungus, *Microbotryum violaceum*, and parameters influencing infection
738   success in artificial inoculations. *Int. J. Plant Sci.* 166(4):649–57

739 68. Gladieux P, Condon B, Ravel S, Soanes D, Maciel JLN, et al. 2018. Gene flow
740   between divergent cereal- and grass-specific lineages of the rice blast fungus
741   *Magnaporthe oryzae*. *mBio*. 9(1):e01219-17

742 69. Gladieux P, Devier B, Aguileta G, Cruaud C, Giraud T. 2013. Purifying selection after
743   episodes of recurrent adaptive diversification in fungal pathogens. *Infect. Genet. Evol.*
744   *J. Mol. Epidemiol. Evol. Genet. Infect. Dis.* 17:123–31

745 70. Gladieux P, Feurtey A, Hood ME, Snirc A, Clavel J, et al. 2015. The population
746   biology of fungal invasions. *Mol. Ecol.* 24(9):1969–86

747 71. Gladieux P, Ropars J, Badouin H, Branca A, Aguileta G, et al. 2014. Fungal
748   evolutionary genomics provides insight into the mechanisms of adaptive divergence in
749   eukaryotes. *Mol. Ecol.* 23(4):753–73

750 72. Gladieux P, Vercken E, Fontaine MC, Hood ME, Jonot O, et al. 2011. Maintenance of
751   fungal pathogen species that are specialized to different hosts: allopatric divergence
752   and introgression through secondary contact. *Mol. Biol. Evol.* 28(1):459–71

753 73. Goldschmidt, V. 1928. Vererbungsversuche mit den biologischen arten den
754   Antherenbrandes (*Ustilago violacea* Pers.). *Z. Bot.* 21(1):1–90

755   74.   Gout L, Fudal I, Kuhn M-L, Blaise F, Eckert M, et al. 2006. Lost in the middle of

756   nowhere: the *AvrLm1* avirulence gene of the Dothideomycete *Leptosphaeria maculans*.

757   *Mol. Microbiol.* 60(1):67–80

758   75.   Grünwald NJ, McDonald BA, Milgroom MG. 2016. Population genomics of fungal and

759   oomycete pathogens. *Annu. Rev. Phytopathol.* 54(1):323–46

760   76.   Han MV, Thomas GWC, Lugo-Martinez J, Hahn MW. 2013. Estimating gene gain and

761   loss rates in the presence of error in genome assembly and annotation using CAFE 3.

762   *Mol. Biol. Evol.* 30(8):1987–97

763   77.   Hardison RC. 2003. Comparative Genomics. *PLOS Biol.* 1(2):e58

764   78.   Hartmann FE, Croll D. 2017. Distinct trajectories of massive recent gene gains and

765   losses in populations of a microbial eukaryotic pathogen. *Mol. Biol. Evol.* 34(11):2808–

766   22

767   79.   Hartmann FE, McDonald BA, Croll D. 2018. Genome-wide evidence for divergent

768   selection between populations of a major agricultural pathogen. *Mol. Ecol.*

769   27(12):2725–41

770   80.   Hartmann FE, Rodríguez de la Vega RC, Brandenburg J-T, Carpentier F, Giraud T.

771   2018. Gene presence–absence polymorphism in castrating anther-smut fungi: recent

772   gene gains and phylogeographic structure. *Genome Biol. Evol.* 10(5):1298–1314

773   81.   Hathaway L, Malm JU, Prentice HC. 2009. Geographically congruent large-scale

774   patterns of plastid haplotype variation in the European herbs *Silene dioica* and *S.*

775   *latifolia* (Caryophyllaceae). *Bot. J. Linn. Soc.* 161(2):153–70

776   82.   Hoban S, Kelley JL, Lotterhos KE, Antolin MF, Bradburd G, et al. 2016. Finding the

777   genomic basis of local adaptation: Pitfalls, practical solutions, and future directions.

778   *Am. Nat.* 188(4):379–97

779   83. Hood ME, Mena-Alí JI, Gibson AK, Oxelman B, Giraud T, et al. 2010. Distribution of

780   the anther-smut pathogen *Microbotryum* on species of the Caryophyllaceae. *New*

781   *Phytol.* 187(1):217–29

782   84. Hood ME, Petit E, Giraud T. 2013. Extensive divergence between mating-type

783   chromosomes of the anther-smut fungus. *Genetics.* 193(1):309–15

784   85. Hood ME, Scott M, Hwang M. 2015. Breaking linkage between mating compatibility

785   factors: Tetrapolarity in *Microbotryum. Evolution.* 69(10):2561–72

786   86. Horns F, Petit E, Hood ME. 2017. Massive expansion of Gypsy-Like retrotransposons

787   in *Microbotryum* fungi. *Genome Biol. Evol.* 9(2):363–71

788   87. Horns F, Petit E, Yockteng R, Hood ME. 2012. Patterns of repeat-induced point

789   mutation in transposable elements of Basidiomycete fungi. *Genome Biol. Evol.*

790   4(3):240–47

791   88. Ironside JE. 2010. No amicable divorce? Challenging the notion that sexual antagonism

792   drives sex chromosome evolution. *BioEssays.* 32(8):718–26

793   89. Kaltz O, Gandon S, Michalakis Y, Shykoff JA. 1999. Local maladaptation in the

794   anther-smut fungus *Microbotryum violaceum* to its host plant *Silene latifolia*: evidence

795   from a cross-inoculation experiment. *Evol. Int. J. Org. Evol.* 53(2):395–407

796   90. Kaltz O, Shykoff JA. 2001. Male and female *Silene latifolia* plants differ in per-contact

797   risk of infection by a sexually transmitted disease. *J. Ecol.* 89(1):99–109

798   91. Kaltz O, Shykoff JA. 2002. Within- and among-population variation in infectivity,

799   latency and spore production in a host–pathogen system. *J. Evol. Biol.* 15(5):850–60

800   92. Kemler M, Göker M, Oberwinkler F, Begerow D. 2006. Implications of molecular

801   characters for the phylogeny of the Microbotryaceae (Basidiomycota:

802   Urediniomycetes). *BMC Evol. Biol.* 6:35

803   93.   Kneip H. 1919. Untersuchungen uber den Antherenbrand (*Ustilago violacea* Pers.): Ein

804   Beitrag zum Sexualitätproblem. *Zeitschrt Bot*. 11:275–84

805   94.   Kües U. 2000. Life history and developmental processes in the basidiomycete *Coprinus*

806   *cinereus*. *Microbiol Mol Biol Rev*. 64(2):316–53

807   95.   Kulkarni RD, Thon MR, Pan H, Dean RA. 2005. Novel G-protein-coupled receptor-

808   like proteins in the plant pathogenic fungus *Magnaporthe grisea*. *Genome Biol*.

809   6(3):R24

810   96.   Kuppireddy V, Uversky V, Toh S, Tsai M-C, Beckerson W, et al. 2017. Identification

811   and initial characterization of the effectors of an anther smut fungus and potential host

812   target proteins. *Int. J. Mol. Sci.* 18(11):2489

813   97.   Le Gac M, Hood ME, Fournier E, Giraud T. 2007a. Phylogenetic evidence of host-

814   specific cryptic species in the anther smut fungus. *Evol. Int. J. Org. Evol.* 61(1):15–26

815   98.   Le Gac M, Hood ME, Giraud T. 2007b. Evolution of reproductive isolation within a

816   parasitic fungal species complex. *Evolution*. 61(7):1781–87

817   99.   Levasseur A, Pontarotti P. 2011. The role of duplications in the evolution of genomes

818   highlights the need for evolutionary-based approaches in comparative genomics. *Biol.*

819   *Direct*. 6:11

820   100. McDonald JH, Kreitman M. 1991. Adaptive protein evolution at the *Adh* locus in

821   Drosophila. *Nature*. 351(6328):652–54

822   101. Mohd-Assaad N, McDonald BA, Croll D. 2018. Genome-wide detection of genes

823   under positive selection in worldwide populations of the barley scald pathogen.

824   *Genome Biol. Evol.* 10(5):1315–32

825   102. Nagy LG, Riley R, Tritt A, Adam C, Daum C, et al. 2016. Comparative genomics of

826   early-diverging mushroom-forming fungi provides insights into the origins of

827   lignocellulose decay capabilities. *Mol. Biol. Evol.* 33(4):959–70

828   103. Nieuwenhuis BPS, Billiard S, Vuilleumier S, Petit E, Hood ME, Giraud T. 2013.

829   Evolution of uni- and bifactorial sexual compatibility systems in fungi. *Heredity*.

830   111(6):445–55

831   104. Perlin MH, Amselem J, Fontanillas E, Toh SS, Chen Z, et al. 2015. Sex and parasites:

832   genomic and transcriptomic analysis of *Microbotryum lychnidis-dioicae*, the biotrophic

833   and plant-castrating anther smut fungus. *BMC Genomics*. 16:461

834   105. Petit E, Silver C, Cornille A, Gladieux P, Rosenthal L, et al. 2017. Co-occurrence and

835   hybridization of anther-smut pathogens specialized on *Dianthus* hosts. *Mol. Ecol.*

836   26(7):1877–90

837   106. Pfeifer B, Wittelsbürger U, Ramos-Onsins SE, Lercher MJ. 2014. PopGenome: an

838   efficient Swiss army knife for population genomic analyses in R. *Mol. Biol. Evol.*

839   31(7):1929–36

840   107. Plissonneau C, Benevenuto J, Mohd-Assaad N, Fouché S, Hartmann FE, Croll D. 2017.

841   Using population and comparative genomics to understand the genetic basis of

842   effector-driven fungal pathogen evolution. *Front. Plant Sci.* 8(119)

843   108. Ponnikas S, Sigeman H, Abbott JK, Hansson B. 2018. Why do sex chromosomes stop

844   recombining? *Trends Genet.* 34(7):492–503

845   109. Rautenberg A, Hathaway L, Oxelman B, Prentice HC. 2010. Geographic and

846   phylogenetic patterns in *Silene* section *Melandrium* (Caryophyllaceae) as inferred from

847   chloroplast and nuclear DNA sequences. *Mol. Phylogenet. Evol.* 57(3):978–91

848   110. Refrégier G, Hood ME, Giraud T. 2010. No evidence of reproductive character

849   displacement between two sister fungal species causing anther smut disease in *Silene*.

850   *Int. J. Plant Sci.* 171(8):847–59

851   111. Refrégier G, Le Gac M, Jabbour F, Widmer A, Shykoff JA, et al. 2008. Cophylogeny

852   of the anther smut fungi and their caryophyllaceous hosts: prevalence of host shifts and

853   importance of delimiting parasite species for inferring cospeciation. *BMC Evol. Biol.*

854   8:100

855   112. Rodríguez de la Vega RC, Rodriguez P, Hoitinga P, Aguileta G, Badouin H, Gouzy J,

856   Siguenza S, Le Prieur S, Esquerre D, Hood ME, Giraud T. 2018. Genomic mechanisms

857   of host specialization in plant pathogens: the case of the castrating anther-smut fungi.

858   *In prep.*

859   113. Salvaudon L, Giraud T, Shykoff JA. 2008. Genetic diversity in natural populations: a

860   fundamental component of plant–microbe interactions. *Curr. Opin. Plant Biol.*

861   11(2):135–43

862   114. Sánchez-Vallet A, Hartmann FE, Marcel TC, Croll D. 2018. Nature's genetic screens:

863   using genome-wide association studies for effector discovery. *Mol. Plant Pathol.*

864   19(1):3–6

865   115. Schäfer AM, Kemler M, Bauer R, Begerow D. 2010. The illustrated life cycle of

866   *Microbotryum* on the host plant *Silene latifolia*. *Botany*. 88(10):875–85

867   116. Schirrmann MK, Zoller S, Croll D, Stukenbrock EH, Leuchtmann A, Fior S. 2018.

868   Genomewide signatures of selection in *Epichloë* reveal candidate genes for host

869   specialization. *Mol. Ecol.* 27(15):3070–86

870   117. Sharma R, Mishra B, Runge F, Thines M. 2014. Gene loss rather than gene gain is

871   associated with a host jump from monocots to dicots in the smut fungus

872   *Melanopsichium pennsylvanicum*. *Genome Biol. Evol.* 6(8):2034–49

873   118. Sloan DB, Giraud T, Hood ME. 2008. Maximized virulence in a sterilizing pathogen:

874   the anther-smut fungus and its co-evolved hosts. *J. Evol. Biol.* 21(6):1544–54

875   119. Steenwyk J, Rokas A. 2017. Extensive copy number variation in fermentation-related

876   genes among *Saccharomyces cerevisiae* wine strains. *G3 Genes Genomes Genet.*

877   7(5):1475–85

878 120. Steenwyk JL, Soghigian JS, Perfect JR, Gibbons JG. 2016. Copy number variation

879 contributes to cryptic genetic variation in outbreak lineages of *Cryptococcus gattii* from

880 the North American Pacific Northwest. *BMC Genomics*. 17(1):700

881 121. Stukenbrock EH, Christiansen FB, Hansen TT, Dutheil JY, Schierup MH. 2012. Fusion

882 of two divergent fungal individuals led to the recent emergence of a unique widespread

883 pathogen species. *Proc. Natl. Acad. Sci.* 109(27):10954–59

884 122. Taylor JW, Branco S, Gao C, Hann-Soden C, Montoya L, et al. 2017. Sources of fungal

885 genetic variation and associating It with phenotypic diversity. *Microbiol. Spectr.* 5(5):

886 123. Thrall PH, Biere A, Antonovics J. 1993. Plant life-history and disease susceptibility--

887 The occurrence of *Ustilago violacea* on different species within the Caryophyllaceae. *J.*

888 *Ecol.* 81(3):489–98

889 124. Thrall PH, Burdon JJ. 2004. Host–pathogen life history interactions affect biological

890 control success. *Weed Technol.* 18(sp1):1269–74

891 125. Toh SS, Chen Z, Rouchka EC, Schultz DJ, Cuomo CA, Perlin MH. 2018. Pas de deux:

892 an intricate dance of anther smut and Its host. *G3 Genes Genomes Genet.*

893 g3.300318.2017

894 126. Toh SS, Chen Z, Schultz DJ, Cuomo CA, Perlin MH. 2017. Transcriptional analysis of

895 mating and pre-infection stages of the anther smut, *Microbotryum lychnidis-dioicae.*

896 *Microbiology*. 163(3):410–20

897 127. Toh SS, Perlin MH. 2016. Resurgence of less-studied smut fungi as models of

898 phytopathogenesis in the omics age. *Phytopathology*. 106(11):1244–54

899 128. Toh SS, Treves DS, Barati MT, Perlin MH. 2016. Reliable transformation system for

900 *Microbotryum lychnidis dioicae* informed by genome and transcriptome project. *Arch.*

901 *Microbiol.* 198(8):813–25

902   129. Vacher C, Vile D, Helion E, Piou D, Desprez-Loustau M-L. 2008. Distribution of
903   parasitic fungal species richness: influence of climate versus host species diversity.
904   *Divers. Distrib.* 14(5):786–98

905   130. Valente LM, Savolainen V, Vargas P. 2010. Unparalleled rates of species
906   diversification in Europe. *Proc. R. Soc. Lond. B Biol. Sci.* 277(1687):1489–96

907   131. van Putten WF, Biere A, Damme JMMV. 2003. Intraspecific competition and mating
908   between fungal strains of the anther smut *Microbotryum violaceum* from the host plants
909   *Silene latifolia* and *S. dioica*. *Evolution*. 57(4):766–76

910   132. van Putten WF, Biere A, van Damme JMM. 2005. Host-related genetic differentiation
911   in the anther smut fungus *Microbotryum violaceum* in sympatric, parapatric and
912   allopatric populations of two host species *Silene latifolia* and *S. dioica*. *J. Evol. Biol.*
913   18(1):203–12

914   133. Vercken E, Fontaine MC, Gladieux P, Hood ME, Jonot O, Giraud T. 2010. Glacial
915   refugia in pathogens: European genetic structure of anther smut pathogens on *Silene*
916   *latifolia* and *Silene dioica*. *PLOS Pathog.* 6(12):e1001229

917   134. Whittle CA, Votintseva A, Ridout K, Filatov DA. 2015. Recent and massive expansion
918   of the mating-type-specific region in the smut fungus *Microbotryum*. *Genetics*.
919   199(3):809–16

920   135. Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol.*
921   *Evol.* 24(8):1586–91

922   136. Yockteng R, Marthey S, Chiapello H, Gendrault A, Hood ME, et al. 2007. Expressed
923   sequences tags of the anther smut fungus, *Microbotryum violaceum*, identify mating
924   and pathogenicity genes. *BMC Genomics*. 8(1):272

925   137.   Zemp N, Tavares R, Widmer A. 2015. Fungal infection induces sex-specific

926   transcriptional changes and alters sexual dimorphism in the dioecious plant *Silene*

927   *latifolia*. *PLOS Genet.* 11(10):e1005536

928

929

930 **Tables and Figures**

931 **Table 1: Whole genome public resources in anther-smut fungi (*Microbotryum* genus).**

932

933

Figure 1: Specialization and co-phylogenies of anther-smut fungi and their corresponding host plants. *(A)* Cladograms representing relationships between species of anther-smut fungi (left, *Microbotryum* genus) and species of host plants (right, mainly Caryophyllaceae), that are a consensus from previous phylogenetic analyses (26, 30, 111). Availability of short reads or long reads-based genome assemblies or population genomic data for the species of anther-smut fungi as presented in Table 1 is indicated with a black square near the fungal cladogram leaves. Dashed lines indicate specialization of a fungal species on a host species, with pink lines for fungal species infecting different hosts, and orange links for fungal species infecting the same host. The sequenced strain *M. intermedium* was sampled on *Salvia pratensis*, although this fungal species is usually found on *Scabiosa* hosts. *(B)* Infected host plants by anther-smut fungi. Numbers refer to host species as in panel A. Spores of anther-smut fungi are visible in the anthers of the flowers (Photo credits: Michael E. Hood, Tatiana Giraud, Maarteen Strack van Schijndel).

Figure 2: Evolutionary processes in anther-smut fungi studied by comparative genomics and population genomics methods. *(A)* Schema highlighting differences in time scales between host specialization, species divergence, coevolution and local adaptation events in four host-specialized *Microbotryum* species. *(B)* Type of genomic variation investigated according to the evolutionary event time scales. *(C)* Examples of methods recently used to investigate various evolutionary events in anther smut fungi, focusing on between-species variation (1), between- and within-species variation (2), or only within-species variation (3). Information on gene annotation, gene expression and gene presence-absence polymorphism may be coupled to narrow down the number of candidate genes to be involved in host specialization, coevolution and local adaptation. (1) Study of gene content variation between whole genome shotgun assemblies of 19 *Microbotryum* species. Core and complementary

959 (species-specific) genomes were computed by sampling groups of 1 to 18 *Microbotryum*

960 species (112). Increase in size of the complementary genome with the number of sampled

961 genomes highlights the dynamic gene gain and loss within the *Microbotryum* genus. Genes

962 contained in the complementary genome are putative candidate genes for host specialization.

963 (2) Identification of genes under positive selection using polymorphism in one species and

964 divergence with an outgroup using McDonald–Kreitman tests. An excess of the ratio between

965 non-synonymous (D_N) and synonymous (D_S) substitutions between species compared to the

966 ratio between synonymous (P_S) and non-synonymous (P_N) polymorphisms within species is

967 indicative of positive selection within the focal species indicated by an asterisk. Examples are

968 shown for the orthologous group ORTHAg06728 and ORTHAg05587 (20). (3) Genome scan

969 to identify selective sweeps based on allele frequency spectrum in *M. lychnidis-dioicae*. A

970 selective sweep is characterized by an excess of rare variants. Composite likelihood ratio

971 (CLR) tests estimate the probability of the presence of a selective sweep taking into account

972 demographic history and genome-wide allele frequency spectrum (15).

973

974 **Figure 3: Distribution of divergence along genomes between species of host-specialized**

975 **anther-smut fungi (*Microbotryum* genus) based on F_{ST} genome scans.** F_{ST} distributions are

976 based on the genomes of five strains in each group for comparisons, except for strains on *S.*

977 *caroliniana* for which only three genomes were available. *(A)* Divergence distribution

978 between *M. silenes-dioicae* and *M. lychnidis-dioicae. (B)* Divergence distribution between

979 *Microbotryum* fungi on *S. caroliniana* and *S. virginica. (C)* (resp. *(D)*) Divergence

980 distribution between Southern and Northern (resp. Eastern) European genetic clusters of *M.*

981 *lychnidis-dioicae* parasitizing *S. latifolia. (E)* Divergence distribution between Eastern and

982 Western European genetic clusters of *M. silenes-dioicae* parasitizing *S. dioica.* In each panel,

983 from top to bottom: density curve of genome-wide per-gene F_{ST} values; chromosomal

984   distribution of per-gene F_{ST} values on the species largest chromosome (on 2 Mb for each to

985   ease comparisons); map showing the sampling location of sequenced strains (genomes used

986   for F_{ST} distributions are shown as squares); pictures of infected hosts by each host-specialized

987   species (Photo credits: Michael E. Hood). For each pairwise comparison, F_{ST} values were

988   calculated per gene for all genes present on autosomal contigs (i.e. not belonging to mating-

989   type chromosomes) and carrying at least 1 SNP using the PopGenome R package (106). Red

990   dashed lines correspond to median F_{ST} values. Genomic data were described in (15, 26, 134).

991   Mapping, SNP calling and F_{ST} calculations were described in (26, 80).

992

993   **Figure 4: Genomic rearrangements and evolutionary strata on mating-type**

994   **chromosomes of *Microbotryum lychnidis-dioicae* on *Silene latifolia*.** *(A)* Circos plot

995   allowing to retrieve the rearrangements events which occurred during the evolution of the *M.*

996   *lychnidis-dioicae* mating-type chromosome. The two mating-type chromosomes (PR and HD

997   mating-type chromosome) of *M. lagerheimii* are used as proxy for the ancestral state (25).

998   The outer tracks represent contigs with scale in Mb. The blue and orange lines link orthologs,

999   with inversions in orange. The blue and purple dots represent the HD and PR loci,

1000   respectively, and the yellow regions the centromeres. *(B)* Evolutionary scenario of the *M.*

1001   *lychnidis-dioicae* mating-type chromosome evolution from the two ancestral mating-type

1002   chromosomes through a centromere-to-telomere fusion event, which brought the PR and HD

1003   loci onto the same chromosome and allowed their linkage through a recombination

1004   suppression (dashed lines). *(C)* Demonstration of stepwise recombination suppression using

1005   per-gene synonymous divergence and their respective standard error (dS ± SE) between

1006   alleles from *M. lychnidis-dioicae* associated to the a_1 versus a_2 mating types along the mating-

1007   type chromosome gene order of *M. lagerheimii*, as a proxy for ancestral gene order.

1008   Evolutionary strata of different divergence levels (colored differently) shows that

43

1009   recombination suppression extended stepwise from the HR and PR mating-type loci. *(D)*

1010   Examples of two gene genealogies showing contrasted clustering levels of alleles at non-

1011   mating-type genes associated with the a_1 versus a_2 mating types (dark grey and light grey

1012   squares, respectively, at the tips of the gene genealogy). The left panel shows the gene

1013   genealogy of a gene belonging to the pseudo-autosomal region (or PAR), with no trans-

1014   specific polymorphism, i.e., intermingled alleles associated with a_1 and a_2 mating types. The

1015   right panel shows the gene genealogy of a gene belonging to the non-recombining region,

1016   with completely separated alleles associated with a_1 versus a_2 mating types of both *M.*

1017   *lychnidis-dioicae* and *M. silenes-dioicae,* and thus trans-specific polymorphism. The branch

1018   length scale is indicated at the bottom of each gene genealogy.

1019
1020

Table 1 : Whole genome public resources in anther-smut fungi (Microbotryum genus).

	Fungal species name	Host plant of sampling	Number of distinct genotypes	References	Public database	Project ID/ Strain ID / Accession ID*
Long read based assemblies	*M. intermedium*	*Salvia pratensis***	1	(25)	GenBank	PRJEB15277 : Microbotryum Intermedium Assembly (GCA_900096595)
	M. lagerheimii	*Silene vulgaris*	1	(25)	GenBank	PRJEB12080 : MvSv-1253-A1-R1 (GCA_900015505); MvSv-1253-A2-R1 (GCA_900013405)
	M. lychnidis-dioicae	*Silene latifolia*	2	(16) (80)	GenBank GenBank	PRJEB12080 : MvSl-1064-A1-R4 (GCA_900015465); MvSl-1064-A2-R4 (GCA_900015445) PRJNA437556 : MvSl-1318_A1 (GCA_003121365); MvSl-1318_A2 (GCA_003121355)
	M. violaceum sensu lato (M.v. caroliniana)	*Silene caroliniana*	1	(26)	GenBank	PRJEB12080: MvCa-1250-A1-R1 (GCA_900014965); MvCa-1250-A2-R1 (GCA_900014955)
	M. violaceum sensu lato (M.v. paradoxa)	*Silene paradoxa*	1	(26)	GenBank	PRJEB12080 : MvSp-1252-A1-R1 (GCA_900015495); MvSp-1252-A2-R1 (GCA_900015485)
	M. violaceum sensu stricto	*Silene nutans*	1	(25)	GenBank	PRJEB12080 : MvSn-1249-A1-R1 (GCA_900015425); MvSn-1249-A2-R1 (GCA_900015455)
	M. saponariae	*Saponaria officinalis*	1	(30)	GenBank	PRJEB12080 : MvSof-1268-A1-R1 (GCA_900015975); MvSof-1269-A2-R1 (GCA_900015475)
	M. scabiosae	*Knautia arvensis*	1	(26)	GenBank	PRJEB12080 : MvKn-1118-A1-R1 (GCA_900008855); MvKn-1118-A2-R1 (GCA_900015415)
	M. silenes-acaulis	*Silene acaulis*	1	(26)	GenBank	PRJNA437556: ASM366583v1 (GCA_003665835); ASM366582v1 (GCA_003665825)
	M. silenes-dioicae	*Silene dioica*	1	(25)	GenBank	PRJEB16741 : MsdSdi1 (GCA_900120095); PRJNA437556 : MsdSdi2 (ID requested)
Whole genome shotgun assemblies	*M. carthusianorum*	*Dianthus superbus*	1	(112)	GenBank	PRJNA437556 : MvDC3-001-A2-G1 (ID requested)
	M. coronariae	*Lychnis flos-cuculi*	1	(112)	GenBank	PRJNA437556 : MvLf-1062-A1-G1 (ID requested)
	M. lagerheimii	*Silene vulgaris*	1	(112)	GenBank	PRJNA437556 : MvSv1-300-38-G1 (ID requested)
	M. lychnidis-dioicae	*Silene latifolia*	1	(104) (112)	GenBank GenBank	PRJNA41281 : p1A1 Lamole (GCA_000166175) PRJNA437556 : MvSlA1A2r2 (ID requested)
	M. major	*Silene otites*	1	(112)	GenBank	PRJNA437556 : MvSo-338-G1 (ID requested)
	M. silenes-acaulis	*Silene acaulis*	1	(112)	GenBank	PRJNA437556 : MvSa-10-04-A1-G1 (ID requested)
	M. silenes-dioicae	*Silene dioica*	1	(112)	GenBank	PRJNA437556 : MvSd-IT02-32-2-17A-A2-1141 (ID requested)
	M. silenes-inflatae	*Silene vulgaris*	1	(112)	GenBank	PRJNA437556 : Sv2-78-06-G1 (ID requested)
	M. stellariae	*Myosoton aquaticum*	1	(112)	GenBank	PRJNA437556 : MvMa-946-A1-G1 (ID requested)
	M. superbum	*Dianthus pavonius*	1	(112)	GenBank	PRJNA437556 : MvDp-1065-A2-G1 (ID requested)
	M. superbum	*Dianthus seguieri*	1	(112)	GenBank	PRJNA437556 : MvDC1-001-A2-G1 (ID requested)
	M. violaceum sensu lato	*Silene sp.*	1	(112)	GenBank	PRJNA437556 : MvSc-a-1127-A2-G1 (ID requested)
	M. violaceum sensu lato (M.v. caroliniana)	*Silene caroliniana*	1	(112)	GenBank	PRJNA437556 : MvCa-1131-A1-G1 (ID requested)
	M. violaceum sensu lato (M.v. italica)	*Silene italica*	1	(112)	GenBank	PRJNA437556 : MvSi-1128-A1-G1 (ID requested)
	M. violaceum sensu lato (M.v. lemmonii)	*Silene lemmonii*	1	(112)	GenBank	PRJNA437556 : MvSlm-001-A2-G1 (ID requested)
	M. violaceum sensu lato (M.v. paradoxa)	*Silene paradoxa*	1	(112)	GenBank	PRJNA437556 : MvSp-880-A1-G1 (ID requested)
	M. violaceum sensu lato	*Silene pusilla*	1	(112)	GenBank	PRJNA437556 : MvSpu-866-A1-G1 (ID requested)
	M. violaceum sensu stricto	*Silene nutans*	1	(112)	GenBank	PRJNA437556 : MvSn-1014-A1-G1 (ID requested)
	M. violaceo-irregulare	*Silene vulgaris*	1	(112)	GenBank	PRJNA437556 : MvSv3-001-G1 (ID requested)
Sequence archive (reads)	*M. lychnidis-dioicae*	*Silene latifolia*	39	(134) (15)	NCBI Short Read Archive NCBI Short Read Archive	PRJNA269361 PRJNA295022
	M. silenes-dioicae	*Silene dioica*	19	(15)	NCBI Short Read Archive	PRJNA295022
	M. violaceum sensu lato (M.v. paradoxa)	*Silene paradoxa*	4	(26)	NCBI Short Read Archive	PRJEB16741
	M. violaceum sensu lato (M.v. caroliniana)	*Silene caroliniana; Silene virginica*	11	(26)	NCBI Short Read Archive	PRJEB16741
	M. saponariae	*Saponaria officinalis*	1	(56)	GenBank	PRJEB11435

*Information were retrieved on public databases on 22.11.18. For long read based assemblies, assemblies of the two mating type a1 and a2 are indicated if available.
ID requested indicate that the genomic data were submitted to the public database and are currently processed
**the sequenced strain was sampled on *Salvia pratensis*, although the fungal species is usually found on *Scabiosa* hosts
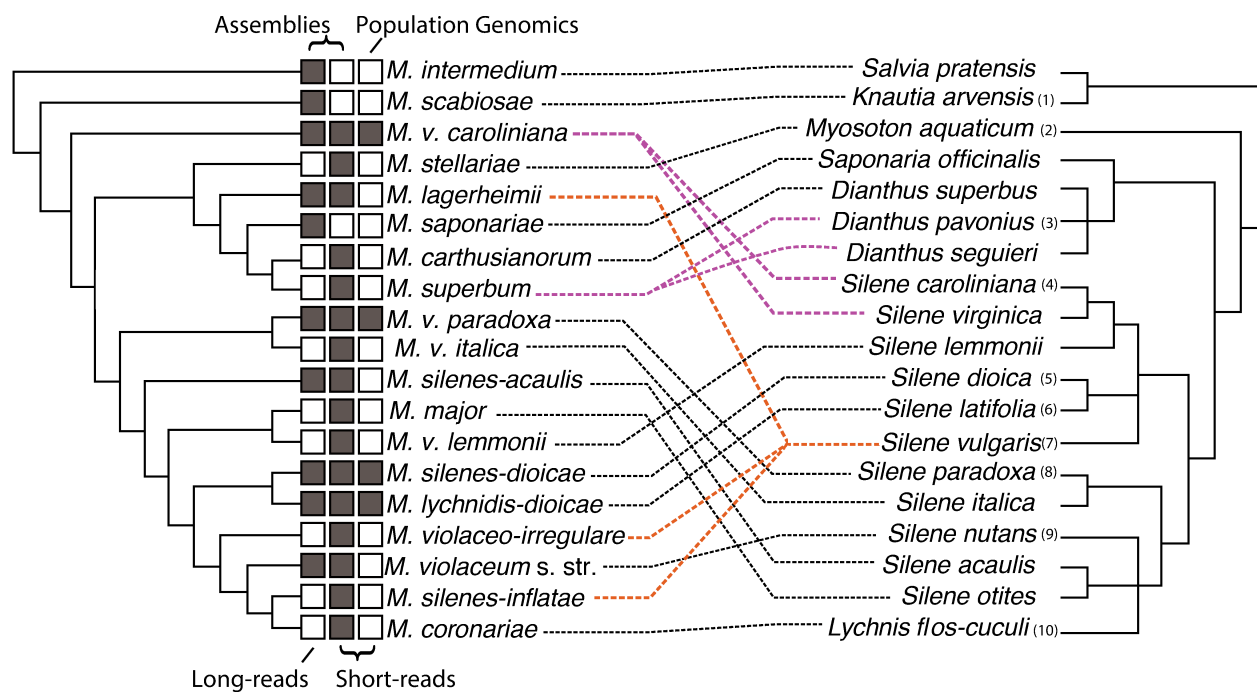
Figure 1

Figure 2

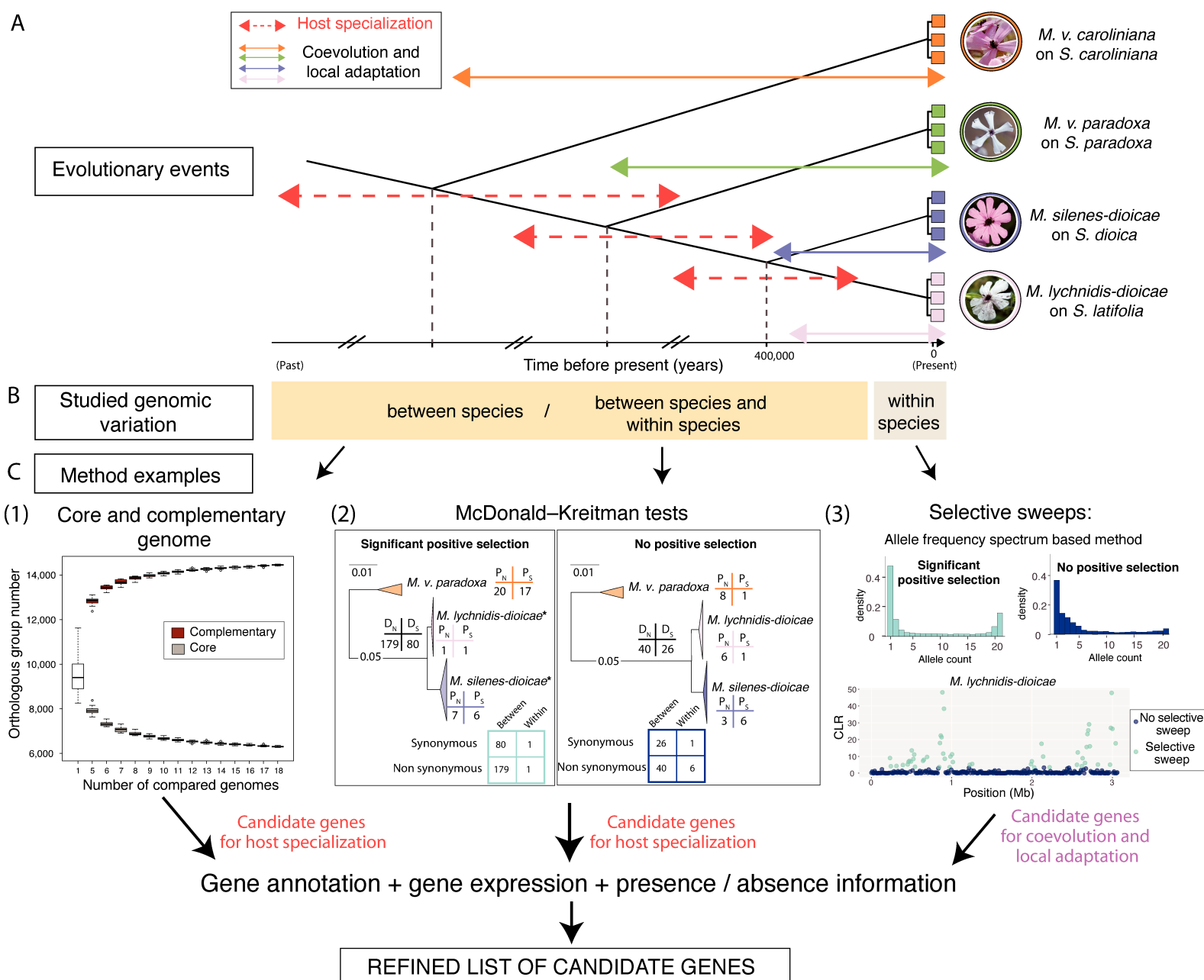**Figure 3**
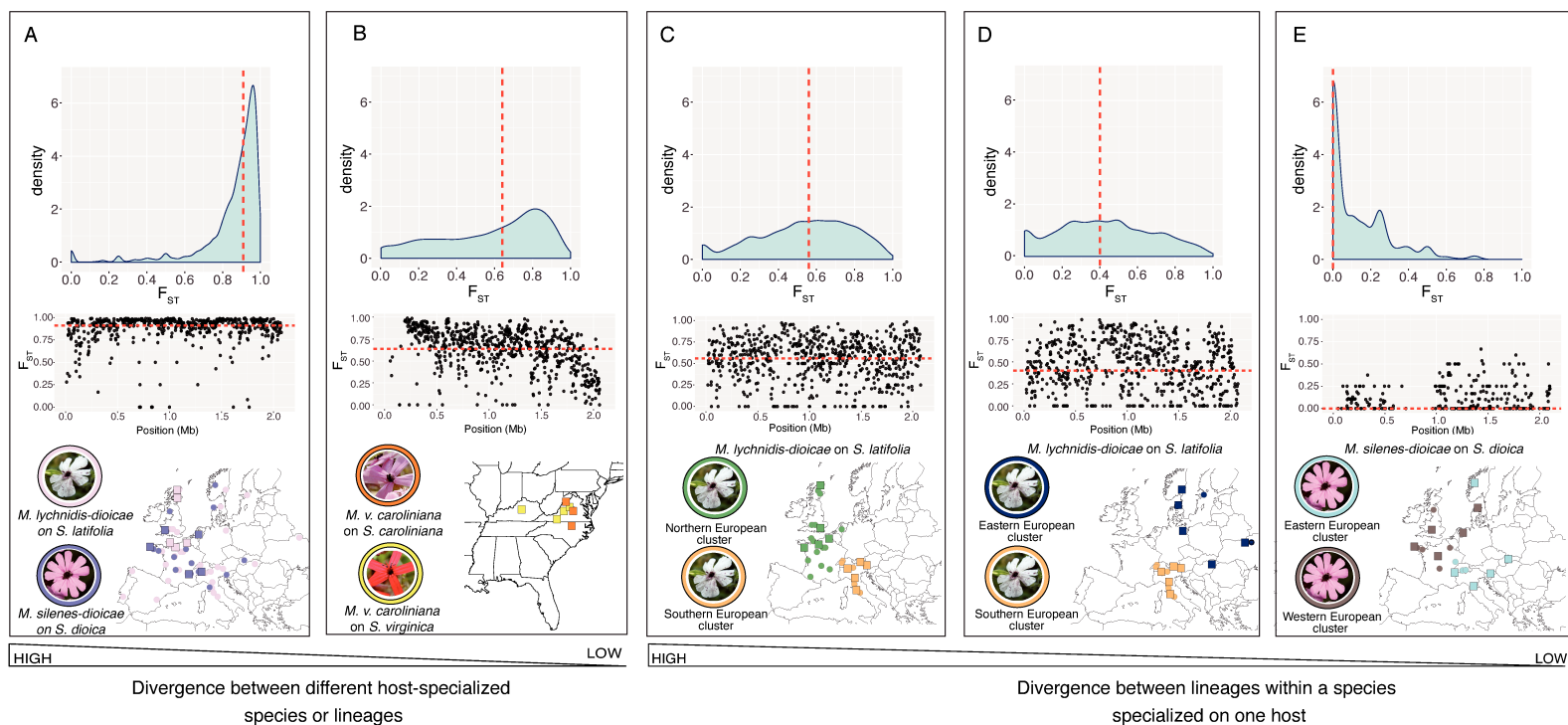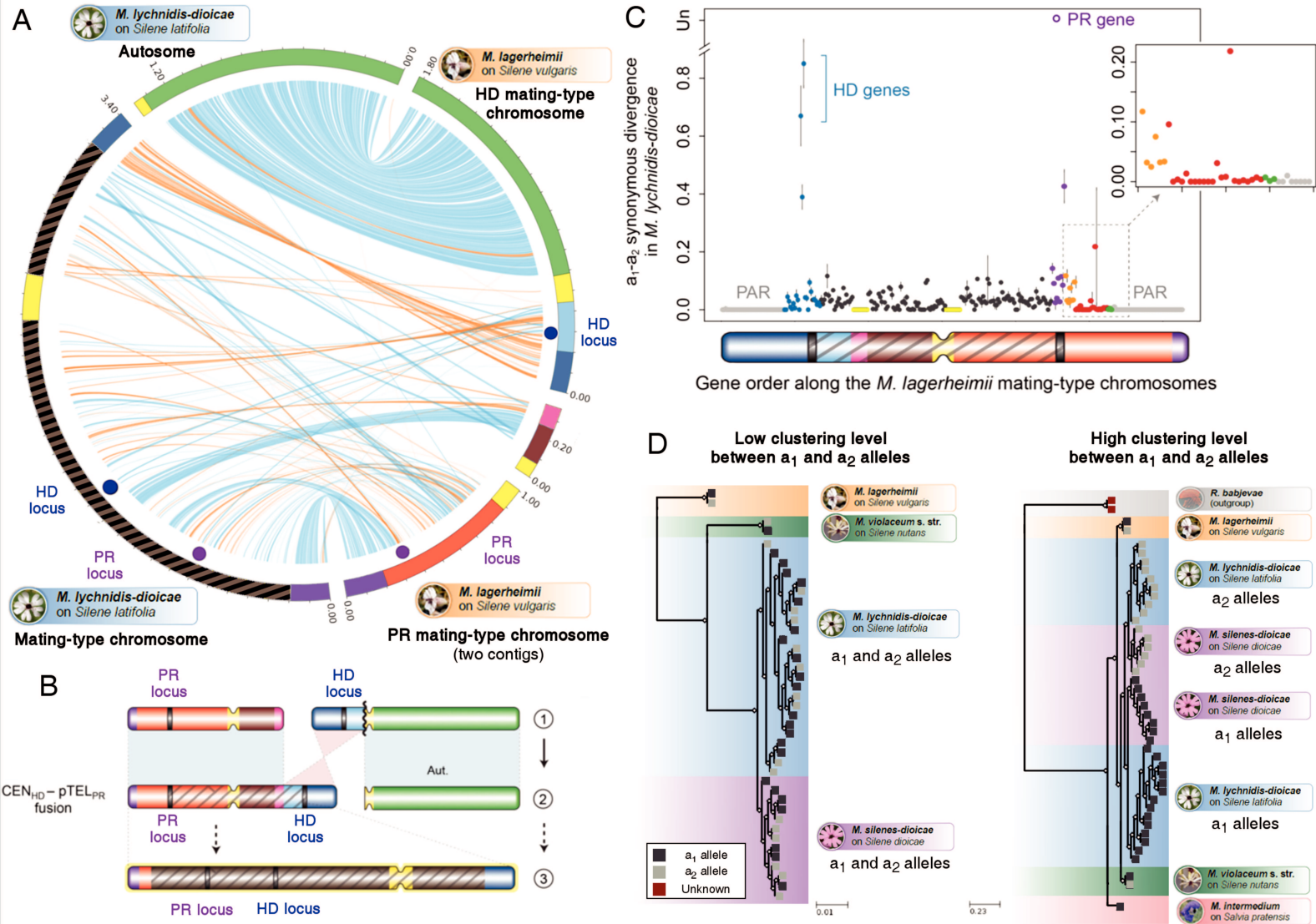
Figure 4

Title: Evolution of non-recombining region in mating-type chromosome from the fungal genus *Microbotryum*

Abstract. In sexual organisms, recombination suppression can evolve in specific genomic regions to protect beneficial allelic combinations, resulting in the transmission of multiple genes as a single locus, which is called a supergene. Supergenes determine complex phenotypes, such as gender in organisms with sex chromosomes. Some sex chromosomes display successive steps of recombination suppression known as "evolutionary strata", which are commonly thought to result from the successive linkage of sexually antagonistic genes (i.e. alleles beneficial to one sex but detrimental to the other) to the sex-determining region. There has however been little empirical evidence supporting this hypothesis. Fungi constitute interesting models for studying the evolutionary causes of recombination suppression in sex-related chromosomes, as they can display non-recombining mating-type chromosomes not associated with male/female functions. Here, we studied the evolution of recombination suppression on mating-type chromosomes in the Microbotryum plant-castrating fungi using comparative genomic approaches.

In *Microbotryum* fungi, mating occurs between gametes with distinct alleles at the two mating-type loci, as is typical of basidiomycete fungi. We showed that recombination suppression evolved multiple times independently to link the two mating-type loci from an ancestral state with mating-type loci on two distinct chromosomes. Recombination suppression either linked the mating-type genes to their respective centromere or linked mating-type loci after they were brought onto the same chromosome through genomic rearrangements that differed between species. Both types of linkage are beneficial under the intra-tetrad mating system of *Microbotryum* fungi as they increase the odds of gamete compatibility. Recombination suppression thus evolved multiple times through distinct evolutionary pathways and distinct genomic changes, which give insights about the repeatability and predictability of evolution. We also reported the existence of independent evolutionary strata on the mating-type chromosomes of several *Microbotryum* species, which questions the role of sexual antagonism in the stepwise extension of non-recombining regions because mating-types are not associated with male/female functions. Previous studies reported little phenotypic differences associated to mating-types, rending unlikely any antagonistic selection between mating types (*i.e.* "mating-type antagonism", with genes having alleles beneficial to one mating-type but detrimental to the other). The genes located in non-recombining regions on the mating-type chromosomes can be differentially expressed between mating types, but our analyses indicated that such differential expression was more likely to result from genomic degeneration than from mating-type antagonism. Deleterious mutations are indeed known to accumulate in non-recombining regions resulting in modifications of gene expression or of protein sequence. We concluded that antagonistic selection cannot explain the formation of evolutionary strata in *Microbotryum* fungi. Alternative mechanisms must be therefore be considered to explain the stepwise expansion of non-recombining regions, and they could also be important on sex chromosomes. This work thus prompts for future studies to identify further evolutionary strata not associated with male/female functions as well as to elucidate their evolutionary causes and consequences in terms of genomic degeneration.

Titre : Evolution des régions non-recombinantes sur les chromosomes de types sexuels chez les champignons du genre *Microbotryum*

Mots-clés : génomique comparative, évolution, chromosomes sexuels, champignons

Résumé. Chez les organismes sexués, des suppressions de recombinaison peuvent évoluer dans certaines régions génomiques pour conserver des combinaisons d'allèles bénéfiques, ce qui aboutit à la transmission de plusieurs gènes en un seul locus, alors appelé « supergène ». Les supergènes déterminent des phénotypes complexes, comme l'identité sexuelle chez les organismes qui ont des chromosomes sexuels. Sur certains chromosomes sexuels, la région sans recombinaison s'est étendue plusieurs fois successivement, produisant des « strates évolutives ». Il est communément admis que ces strates évolutives sont issues de liaisons successives de gènes sexuellement antagonistes (qui ont des allèles bénéfiques à un sexe mais délétère à l'autre) à la région qui détermine le sexe, mais peu de preuves empiriques soutiennent cette hypothèse. Les champignons constituent des modèles intéressants pour étudier les causes évolutives des suppressions de recombinaison parce qu'ils peuvent avoir des chromosomes de types sexuels non recombinants sans être associés à des fonctions mâle ou femelles. Dans cette thèse, nous avons étudié l'évolution de la suppression de recombinaison sur les chromosomes de type sexuel chez les champignons castrateurs de plantes du genre *Microbotryum*. Chez les champignons *Microbotryum*, les croisements ne sont possibles qu'entre des gamètes qui ont des allèles distincts aux deux locus de types sexuels. Nous avons montré que les suppressions de recombinaison ont évolué plusieurs fois indépendamment pour lier les deux locus de types sexuels, depuis l'état ancestral avec les locus de types sexuels situés sur deux chromosomes différents. La suppression de recombinaison a soit lié les locus de types sexuels à leur centromère respectif, ou a lié les locus de types sexuels entre eux après que des réarrangements chromosomiques, différents dans les différentes espèces, les aient amenés sur le même chromosome. Les deux sortes de suppression de recombinaison sont bénéfiques sous le mode de reproduction par auto-fécondation intra-tétrade de *Microbotryum*, parce qu'ils augmentent le taux de compatibilité entre gamètes. Les suppressions de recombinaison ont donc évolué plusieurs fois indépendamment *via* des chemins évolutifs et des changements génomiques différents, ce qui renseigne sur la répétabilité de l'évolution.

De plus, nous avons révélé l'existence de strates évolutives sur les chromosomes de type sexuels de plusieurs espèces de *Microbotryum*, ce qui remet en cause le rôle de l'antagonisme sexuel dans la formation de strates évolutives, les types sexuels n'étant pas associés à des fonctions mâles / femelles. Des études précédentes ont rapporté peu de différences phénotypiques associées aux types sexuels, ce qui rend peu probable qu'une sélection antagoniste existe entre types sexuels sur de nombreux gènes (l'existence de gènes avec des allèles bénéfiques à un type sexuel mais délétère à l'autre). Certains gènes situés dans les régions non-recombinantes des chromosomes de types sexuels étaient différentiellement exprimés entre types sexuels, mais nos analyses suggèrent qu'un tel différentiel d'expression peut être dû à la dégénérescence. En effet, des mutations délétères s'accumulent dans les régions non-recombinantes, ce qui peut modifier l'expression des gènes ou les séquences protéiques. Nous avons donc conclu que la sélection antagoniste ne peut pas expliquer la formation des strates évolutives chez les champignons *Microbotryum*. Par conséquent, des mécanismes alternatifs doivent être considérés pour expliquer l'extension progressive des régions non-recombinantes, et ces mécanismes pourraient aussi générer des strates évolutives sur les chromosomes sexuels. Ces travaux incitent de futures études à d'une part identifier d'autres strates évolutives qui ne sont pas associées à des fonctions mâles/femelles, et d'autre part à identifier leurs causes évolutives et leurs conséquences en termes de dégénérescence.