Analyse moléculaire de l'interaction entre peupliers et Melampsora spp. par des approches génomiques et fonctionnelles

Cécile Lorrain

▶ **To cite this version:**

Cécile Lorrain. Analyse moléculaire de l'interaction entre peupliers et Melampsora spp. par des approches génomiques et fonctionnelles. Biologie végétale. Université de Lorraine, 2018. Français. NNT : 2018LORR0016 . tel-01920199

HAL Id: tel-01920199

https://theses.hal.science/tel-01920199

Submitted on 13 Nov 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

AVERTISSEMENT

Ce document est le fruit d'un long travail approuvé par le jury de soutenance et mis à disposition de l'ensemble de la communauté universitaire élargie.

Il est soumis à la propriété intellectuelle de l'auteur. Ceci implique une obligation de citation et de référencement lors de l'utilisation de ce document.

D'autre part, toute contrefaçon, plagiat, reproduction illicite encourt une poursuite pénale.

Contact : ddoc-theses-contact@univ-lorraine.fr

LIENS

Code de la Propriété Intellectuelle. articles L 122. 4
Code de la Propriété Intellectuelle. articles L 335.2- L 335.10
http://www.cfcopies.com/V2/leg/leg_droi.php
http://www.culture.gouv.fr/culture/infos-pratiques/droits/protection.htm

Université de Lorraine, Collegium Sciences et Technologies
Ecole Doctorale «Ressources, Procédés, Produits, Environnement»

Thèse

Présentée et soutenue publiquement pour l'obtention du grade de

Docteur de l'Université de Lorraine

spécialité Biologie Végétale et Forestière par

Cécile Lorrain

Analyse moléculaire de l'interaction entre peupliers et *Melampsora* spp. par des approches génomiques et fonctionnelles

Soutenance publique le 28 mars 2018

Jury

Elodie Gaulin, Maître de Conférence, Université de Toulouse	Rapportrice
Thomas Kroj, Directeur de Recherche, INRA Montpelier	Rapporteur
Mélanie Morel-Rouhier, Professeur, Université de Lorraine	Examinatrice
Jessica Soyer, Chargée de Recherche, INRA Grignon	Examinatrice
Sébastien Duplessis, Directeur de Recherche, INRA Grand Est Nancy	Directeur de thèse
Arnaud Hecker, Maître de Conférence, Université de Lorraine	Directeur de thèse

UMR 1136 INRA / Université de Lorraine, Interactions Arbres - Microorganismes
INRA Centre Grand Est - Nancy – 54280 Champenoux
Faculté des Sciences et Techniques – 54500 Vandoeurvre-lès-Nancy

Remerciements :

Pour commencer je souhaite remercier les membres de mon jury de thèse **Thomas Kroj**, **Elodie Gaulin**, **Jessica Soyer** et **Mélanie Morel-Rouhier** qui ont accepté d'évaluer mes travaux.
Je remercie également **Benjamin Petre, Claire Veneault-Fourrey** et **Isabelle Fudal**, les membres de mon comité de suivi qui m'ont aidé tout au long de ma thèse. Ben, merci pour tout ce que tu m'as apporté, c'est toi qui m'as donné envie de continuer à travailler sur la rouille du peuplier pendant ma thèse.

Je remercie tout particulièrement mes deux directeurs de thèse **Sébastien Duplessis** et **Arnaud Hecker**. Merci pour tout ce que vous m'avez appris. Seb, trop à dire…j'avoue, #facilitantpower. Merci pour tout ce que tu m'as appris, pour toutes les opportunités dont tu m'as fait profiter. Arnaud, merci pour ta patience, ton soutien et surtout d'avoir supporté mes étourderies… Je mesure la chance que j'ai eu d'avoir des encadrants comme vous deux et vous exprime toute ma reconnaissance.

Un grand merci à l'équipe écogénomique pour l'accueil, au sein du labo :

Je remercie **Francis Martin** notamment pour m'avoir donné l'opportunité de participer à ECFG13. Merci à **Claire** et **Annegret** pour leurs conseils, leur bonne humeur et leurs rires. Merci à **Marc**, **Stéphane**, **Aurélie**, **Claude** et **Elena.**

Un grand merci à **Manue** que j'ai harcelé plus d'une fois avec les TE, les transporteurs avec le pipeline TRANU, et j'en passe ! Merci pour ton aide indispensable !

Merci **Christine** pour ton aide avec les peupliers transformés et tes huiles essentielles ! Merci pour ton invitation chez toi. Merci **Patrice** pour t'être occupé des peupliers en serre S2. Merci à **Jean-Louis**, **Cyril**, **Béa** et **Frédéric**. Merci à **Nicole**, **Agnès** et **Marie-Claude** vous êtes toujours là pour nous ré-expliquer mille fois comment rédiger nos ordres de mission et nos commandes.

Un ENORME merci **Clémence ma PFF** ! Je n'arrive pas à coire que ça fait déjà 3 ans que tu es venue faire ton stage…Team PFF always !
Maïra et **Lauralie** merci pour votre soutien infaillible. Maïra désolée de t'avoir rendue malade en voiture…je n'oublierai pas de sitôt notre exapade dans les montagnes Suisses !
Laurette merci pour ta bonne humeur toujours la personne qui sait tout ce qu'il se passe au labo.
François, tout ça à cause d'un repiquage de calles de peupliers.
Clémence, **Véronica** et **Lauralie**, merci d'avoir repris le Journal club…ce petit cadeau empoisonné.
Feng thank you for your smile, it was always nice to talk with you.
Merci à ceux qui sont déjà partis **Clément**, **Yohann**, **Yannick**, **Cora** et **Igor**.

Je remercie tous les membres de l'équipe Redox :

Nicolas Rouhier merci de m'avoir accueillit dans l'équipe, **Eric, Mélanie, Rodnay, Jean-Pierre**, **Jeremy** un grand merci pour votre accueil.

Merci à **Jean-Michel**, **Tiffaine** et **Raphael** pour leur aide avec la FPLC.
Merci à **Thomas**, **Flavien**, **Nico**, **Johnathan**, **Elena** et **Mélanie** pour votre présence, votre aide, vos parties de foot dans le couloir et j'en passe. Remplir des boites de cones était toujours un moment sympa avec vous !
Un vrai plaisir de « passer au 3 », toujours dans la bonne humeur et c'est totalement par hasard qu'à chaque fois que j'étais dans les murs il y avait un pot ^^

Je remercie l'équipe écologie des pathogènes forestiers :

Merci à **Pascal Frey** pour son accueil. Merci à **Stéphane** pour son aide avec le *dN/dS*, **Fabien** et **Bénédicte** qui ont accompagné mes premiers pas dans un laboratoire…il y a 5 ans déjà.
Un grand merci à **Jeremy** pour les peupliers…et encore désolée de m'être tranformée en folle des ciseaux…Merci à **Agathe**, **Marie**, **Axelle, Antoine, Michael** et tous ceux que j'oublie.

Je remercie les laboratoires du Canada :

Je tiens à remercier tout particulièremet **Hugo Germain** et toute son équipe pour m'avoir accueillit pendant deux mois à l'UQTR.

Un grand merci à **Cristof** et **Michelle Ramptisch** pour l'accueil chaleureux chez eux, à Morden.

Merci à **Philippe Tanguay** pour l'accueil à Québec.

Je remercie les structuralistes :

Merci à **André Padilla** et **Karine de Guillen** grâce à qui nous avons mis en place notre collaboration à Effectome.

Merci à **Pascale Tsan** pour avoir ré-ouvert le dossier P4.

Pour finir je remercie tout particulièrement :

Toute ma famille pour son soutien inconditionnel.
Mes **parents** qui m'ont toujours encouragé dans mes études, maman mon modèle et papa avec nos débats du dimanche qui fatiguaient tout le monde. Mes **sœurs** qui m'ont supporté en coloc' chacune leur tour, ces moments restent gravés parmi mes souvenirs les plus heureux…avec nos ballades à cheval ^^
Merci, **Martine** et **Gérard, Eric** et **Brigitte** toujours là pour moi.
Merci, à mon parrain **Thierry** qui m'a toujours encouragé.

Liste des figures :

Figure 1 : Représentation schématique du système immunitaire des plantes

Figure 2 : Récepteurs PRR caractérisés et/ou identifiés avec leurs ligands

Figure 3 : Mécanismes de reconnaissance des effecteurs par les récepteurs NLR conduisant à l'ETI.

Figure 4 : Réponse systémique acquise, mémoire transgénérationnelle et signaux associés

Figure 5 : Comparaison entre les concepts du modèle en zigzag et le modèle d'invasion.

Figure 6: Structures de sécrétion des effecteurs des pathogènes filamenteux.

Figure 7 : Les trois étapes de la biologie des effecteurs

Figure 8 : Cycles de vie hétéroïques et macrocycliques de deux rouilles du peuplier *M. larici-populina* et *M. allii-populina*.

Figure 9: Ré-annotation des éléments transposables dans le génome version 2.0 de *M. larici-populina*

Figure 10 : Homologues partagés parmi 16 espèces fongiques.

Figure 11 : Prédiction et analyse du sécrètome de *M. larici-populina* v2.0

Figure 12 : Catégories KOG annotées dans le sécrétome de *M. larici-populina* 2.0

Figure 13 : Environnement génique des génomes et sécrètome de *M. larici-populina* 2.0.

Figure 14 : Position de différentes catégories de gènes annotés de *M. larici-populina* 2.0 sur les 18 groupes de liaisons.

Figure 15 : Homologues partagés entre *M. allii-populina* et 22 espèces fongiques.

Figure 16 : Divergence moléculaire entre les 22 génomes fongiques analysés par MCL.

Figure 17 : Prédiction et analyse du sécrètome de *M. allii-populina*

Figure 18 : Catégories KOG annotées dans le sécrétome de *M. allii-populina*

Figure 19 : Environnement génique des génomes et sécrètome de *M. allii-populina*.

Figure 20 : Distribution des ratios *dN/dS* pour les orthologues identifiés entre *M. allii-populina* et *M. larici-populina*.

Figure 21 : Démarche expérimentale ayant conduit à la caractérisation fonctionnelle et structurale de deux effecteurs candidats de *M. larici-populina* MLPCTP1 et MLP124017.

Figure 22 : MLPCTP1-GFP cible les chloroplastes de transformants stables d'Arabidopsis.

Figure 23 : MLPCTP1-GFP cible les chloroplastes de transformants stables de peupliers 717-1B4.

Figure 24 : Production, purification de MLP124017 et analyse des structures secondaires

Figure 25 : Expression de MLPCTP1 chez *E. coli* et *P. pastoris*.

Figure 26 : Le processus d'infection complexe de l'hôte écidien de *P. striiformis* f. sp. *tritici*.

Figure 27 : Structures d'infection des urédiniospores et des basidiospores de Pucciniales : deux mécanismes de pénétration différents.
Figure 28 : Densité en TE sur les 18 groupes de liaison du génome de *M. larici-populina*.
Figure 29 : Exemples d'approches *in vitro* applicables à la caractérisation des effecteurs.

Liste des tableaux :

Tableau 1 : Ressources génomique des Pucciniales
Tableau 2: Statistiques de réassemblage de la version 2.0 du génome de *M. larici-populina* (Mlp) 98AG31 genome versus la version 1.0
Tableau 3: Classification REPET des séquences consensus identifiées chez *M. larici-populina* 2.0 versus les consensus identifiés dans la version 1.0.
Tableau 4: Statistiques de réassemblage du génome de *M. allii-populina* 12AY07 en comparaison avec *M. larici-populina* 2.0
Tableau 5: Classification REPET des séquences consensus « full-length fragments » identifiées chez *M. alii-populina*.
Tableau S1 : Description des Transporteurs annotées dans le génome de *M. larici-populina*.
Tableau S2 : Description des CAZYmes annotées dans la version 2.0 du génome de *M. larici-populina* versus la version 1.0.
Tableau S3 : Description des Protéases annotées dans le génome de *M. allii-populina*.
Tableau S4 : Sécrètome de *M. larici-populina* 2.0 (disponible sur demande)
Tableau S6 : Familles mutligéniques en expansion chez *M. allii-populina* en comparaison avec 22 génomes fongiques.
Tableau S7 : Description des CAZYmes annotées dans le génome de *M. allii-populina*.
Tableau S8 : Description des Transporteurs annotées dans le génome de *M. allii-populina*.
Tableau S9 : Description des Protéases annotées dans le génome de *M. allii-populina*.
Tableau S10 : Sécrètome de *M. allii-populina* (disponible sur demande)

Table des matières

CHAPITRE I : Introduction, contexte et objectifs de la thèse 13

Partie 1: Introduction générale 14

1. Modes de vie des parasites des plantes et types d'interaction 14
2. Les interactions moléculaires plante-microorganisme : l'immunité des plantes 15
3. Les Effecteurs : concepts et définitions 22
4. Génomique des interactions plante-microorganisme 24
5. L'interaction peuplier-*Melampsora larici-populina* 28

Partie 2: Biologie, génomique et analyse fonctionnelle d'effecteurs des Pucciniales 31

Article n°1 Biology of rust fungi: advances in understanding obligate biotroph parasitism 32

Partie 3: Prédiction et identification d'effecteurs candidats chez *M. larici-populina*..**Erreur ! Signet non défini.**

Article n°2: Effector-mining in the poplar rust fungus *M. larici-populina* secretome 74

Partie 4: Contexte et objectifs de la thèse 89

CHAPITRE II : Analyse transcriptomique du cycle de vie de Melampsora larici-populina : comparaison des profils d'expression génique lors de l'infection du mélèze et du peuplier, les hôtes télien et écidien de l'agent de la rouille du peuplier 93

Partie 1 : Introduction 94

Partie 2 : Article n°3: The rust fungus *Melampsora larici-populina* expresses a conserved genetic program and distinct sets of secreted proteins during infection of its two host plants: larch and poplar 96

CHAPITRE III : Génomique des rouilles du peuplier 134

Partie 1 : Introduction 135

Partie 2 : Matériel et Méthodes 142

Partie 3 : Contribution à l'annotation de la version 2 du génome de *Melampsora larici-populina*..... 146

Partie 4: Contribution à l'annotation du génome de *Melampsora allii-populina*..... 159

Partie 5 : Conclusions..... 169

CHAPITRE IV : Analyse fonctionnelle d'effecteurs candidats de M. larici-populina .. 174

Partie 1 : Introduction..... 175

Partie 2 : Article n°4 NMR structural characterization of MLP124266 and MLP124017, two candidate effector proteins produced by the leaf rust pathogen Melampsora larici-populina 180

Partie 3: Article n°5 Rust fungal effectors mimic host transit peptides to translocate into chloroplasts..... 218

Partie 4 : Informations complémentaires..... 246

1. Matériel et Méthodes..... 246

2. Résultats..... 248

2.1 Expression des effecteurs CTP1 et 124017 de *M. larici-populina* dans le peuplier et *A. thaliana* 248

2.2. Expression de protéines recombinantes pour les effecteurs candidats MLPCTP1 et MLP124017 en vue de leur caractérisation structurale. 250

3. Discussion..... 254

CHAPITRE V : Discussion générale..... 260

Partie 1 : Questions et stratégie de recherche pour l'étude de l'interaction peuplier-rouille du peuplier 261

Partie 2 : Les apports de la génomique à l'étude de la biologie des Pucciniales 262

1. Apports de la transcriptomique à l'étude de l'hétéroécie chez les Pucciniales..... 263

2. Revisiter l'annotation des génomes afin d'affiner les répertoires de gènes clés associées à la biotrophie 269

3. L'architecture des génomes de *M. larici-populina* et *M. allii-populina* : invasion par les éléments transposables (TE) 273

4. Vers la génomique comparative des Pucciniales 275

Partie 3 : La biologie des effecteurs dans un pathosystème non-modèle : avancées et limites 277

1. Vers la caractérisation fonctionnelle des effecteurs des Pucciniales 277

2. La biologie structurale appliquée aux effecteurs 278

3. Les effecteurs chloroplastiques 280

4. La caractérisation fonctionnelle des effecteurs: la partie émergée de l'iceberg 281

Conclusions et Perpectives 296

Références 301

Annexes 334

CHAPITRE I

Introduction, contexte et objectifs de la thèse

Partie 1: Introduction générale

Dans la nature, les plantes sont en contact permanent avec des microorganismes présents dans leur environnement et les interactions qui en découlent peuvent être bénéfiques, neutres ou néfastes pour la plante. Certains microorganismes jouent un rôle déterminant dans la vigueur et le bon état sanitaire des plantes, notamment en engageant des associations mutualistes apportant des bénéfices à la plante. Au contraire, d'autres microorganismes se développent au détriment des plantes et peuvent engendrer des maladies. Ces microorganismes phytopathogènes, qu'ils soient des champignons, des oomycètes, des bactéries ou des virus, peuvent entraîner des pertes économiques conséquentes sur les plantes de grandes cultures.

1. Modes de vie des parasites des plantes et types d'interaction

Les organismes phytopathogènes peuvent être divisés en trois grandes catégories en fonction de leur stratégie d'infection : les nécrotrophes, les hémibiotrophes et les biotrophes (Kemen & Jones, 2012). Les nécrotrophes puisent leurs nutriments et leur énergie des cellules mortes de leurs hôtes (Horbach *et al.,* 2011). Ces derniers vont donc tuer leurs hôtes afin de se développer. Ce sont des pathogènes à large spectre d'hôtes qui sécrètent un répertoire important d'enzymes dégradant les parois végétales ainsi que des toxines. Les biotrophes dérivent leurs nutriments et leur énergie des tissus vivants de leurs hôtes (Staples, 2000). Ainsi, ils établissent une interaction durable avec leurs hôtes qu'ils maintiennent en vie une fois l'infection établie. La plupart des biotrophes sont des parasites obligatoires ce qui signifie qu'ils ne peuvent pas vivre en dehors de leur hôte. Les parasites biotrophes obligatoires possèdent des structures d'infection spécialisées, sécrètent peu d'enzymes de dégradation et de métabolites secondaires et ils ne causent pas directement la mort de l'hôte qu'ils colonisent (Kemen & Jones, 2012). Les hémibiotrophes quant à eux sont intermédiaires des précédents modes trophiques. En effet, ces parasites colonisent leurs hôtes en deux phases : une phase asymptomatique de croissance biotrophe suivie d'une phase nécrotrophe au cours de laquelle les symptômes sont exprimés (Lee & Rose, 2010). La longueur de la phase de croissance biotrophe peut varier en fonction du pathogène (Goswami & Kistler, 2004; Teixeira *et al.,* 2014). Ces trois grands modes de vie des parasites des plantes sont retrouvés aussi bien chez les bactéries, les champignons, les oomycètes et chez les insectes.

Des interactions dites hôtes et non-hôtes peuvent être distinguées entre plantes et parasites (Lepoivre, 2003). L'interaction non-hôte est la plus répandue et désigne l'incompatibilité fondamentale entre une plante et un parasite. Par opposition, l'interaction hôte existe lorsque certains biotypes (ou souches ; pathovars) d'une espèce parasite peuvent coloniser certains

biotypes (ou cultivars) de plantes et former des interactions dites « compatibles » ou « incompatibles » (Lepoivre, 2003). Une interaction compatible aboutie à la colonisation d'un cultivar de plante sensible par un parasite alors qualifié de virulent. Lors de l'interaction compatible la colonisation peut varier de manière quantitative en fonction de la sensibilité de la plante (sensibilité totale ou résistance partielle) et de l'agressivité du parasite. Une interaction incompatible désigne la résistance totale d'un cultivar d'une plante à une souche parasite désignée alors comme avirulent (Lepoivre, 2003).

Le spectre des organismes associés aux plantes s'étend bien au-delà des parasites. De nombreuses espèces de bactéries ou de champignons établissent des interactions mutualistes avec leurs hôtes, promouvant la croissance des deux partenaires de l'interaction avec un échange bidirectionnel de nutriments (Selosse, 2000). Par exemple, des nutriments azotés peuvent être apportés à la plante par des bactéries ou des champignons symbiotiques mutualistes et en retour la plante hôte apporte des ressources carbonées au microorganisme symbiote (Martin *et al.,* 2017). Dans ce type d'associations mutualistes les microorganismes colonisent le système racinaire de leurs plantes hôtes après avoir établi un dialogue moléculaire complexe et coordonné pour activer une variété de modifications anatomiques et développementales chez l'hôte et ainsi former des structures d'échanges dédiées à l'interaction (Martin *et al.*, 2017).

2. Les interactions moléculaires plante-microorganisme : l'immunité des plantes

Depuis l'Antiquité, les maladies des plantes étaient considérées comme la manifestation de « la colère des dieux » (Large, 1950). Ce n'est qu'après la deuxième moitié du XIXème siècle qu'Anton de Bary montre l'implication d'agents pathogènes dans les maladies des plantes (Large, 1950). Régulièrement, les agents pathogènes ont ravagé les cultures avec des conséquences importantes sur les sociétés humaines, comme lors la grande famine d'Irlande causée par l'oomycète *Phytophthora infestans*, l'agent du mildiou de la pomme de terre (Peterson, 1992 ; Viennot-Bourgin, 1981). Ces événements ont mené à l'émergence de la phytopathologie, un champ disciplinaire des sciences principalement centré sur la gestion des maladies des plantes cultivées par l'amélioration des plantes et la recherche de cultivars résistants (Dangl e*t al.,* 2013). Depuis les années 1980s et l'application de la biologie moléculaire à cette discipline, l'étude des interactions moléculaires plante-microorganisme a permis une meilleure compréhension des mécanismes sous-jacents aux maladies des plantes. Ainsi, près d'un demi-siècle après l'émergence du concept d'existence de gènes de résistance aux parasites chez les plantes (Flor, 1971), les premiers gènes de résistance ont pu être identifiés chez les plantes (Dangl *et al.,* 2013). Depuis les années 2000s, la

« révolution » génomique qui a été rendue possible par le développement des outils de séquençages des acides nucléiques et d'analyses bioinformatiques, a révélé la complexité et la diversité des gènes et des fonctions impliquées dans les mécanismes moléculaires des interactions entre plantes et microorganismes (Dodds & Rathjen, 2010 ; Jones *et al.*, 2016 ; Pritchard & Birch, 2014 ; Cui *et al.*, 2015).

En 2006, les chercheurs Jones et Dangl ont proposé un modèle établissant un cadre général pour la compréhension des mécanismes moléculaires sous-jacents à l'immunité chez les plantes (Jones & Dangl, 2006). Ce modèle dit « en zig-zag » repose sur plusieurs étapes successives a permis de concilier les avancées des chercheurs dans le domaine tout en intégrant un aspect de coévolution des interactions entre les plantes et les microorganismes pathogènes. Il fait apparaître deux niveaux de reconnaissance des pathogènes. Le premier niveau est la reconnaissance de molécules conservées chez les pathogènes, appelés « motifs moléculaires associés aux pathogènes/microbes » (*pathogen-/microbe-associated molecular patterns* en anglais ; PAMPs ou MAMPs étant les acronymes établis dans le domaine) qui permet le déclenchement de « l'immunité médiée par la détection de motifs moléculaires » (*pattern-triggered immunity* en anglais ; PTI étant l'acronyme établi dans le domaine). Le second niveau des défenses immunitaires des plantes reconnait et répond à des facteurs de la pathogénèse appelés « effecteurs » qui dans ce modèle jouent le rôle de suppresseurs du PTI (Jones & Dangl, 2006). Ce niveau d'immunité est appelé « immunité médiée par la détection d'effecteurs » (*effector-triggered immunity* en anglais ; ETI étant l'acronyme établi dans le domaine). L'action de contournement des mécanismes de résistance de la plante par l'expression d'effecteurs par le pathogène est appelée « sensibilité médiée par l'action d'effecteurs » (*effector-triggered susceptibility* en anglais ; ETS étant l'acronyme établi dans le domaine). L'ETS se place entre la PTI et l'ETI. Ce modèle intègre également la notion de coévolution entre les deux protagonistes de l'interaction avec le concept de « course à l'armement ». Dans ce cadre, l'ETI est suivie de nouveau d'une phase ETS avec un nouvel effecteur, et de nouveau une phase ETI dans le cas où ce nouvel effecteur est reconnu par une protéine de résistance de la plante. Ce modèle en zig-zag est relativement simple à appréhender. Il est intégratif et il a permis lors de sa publication de regrouper différents concepts d'interactions moléculaires plante-microorganisme.

Une autre représentation régulièrement donnée de ce modèle intégratif de l'immunité chez les plantes se fait par la représentation des différents niveaux de détection des parasites au niveau de la cellule de la plante. Deux exemples fréquemment repris par la communauté sont ceux de Doods & Rathgen (2010) et de Dangl et collaborateurs (2013), qui permet de

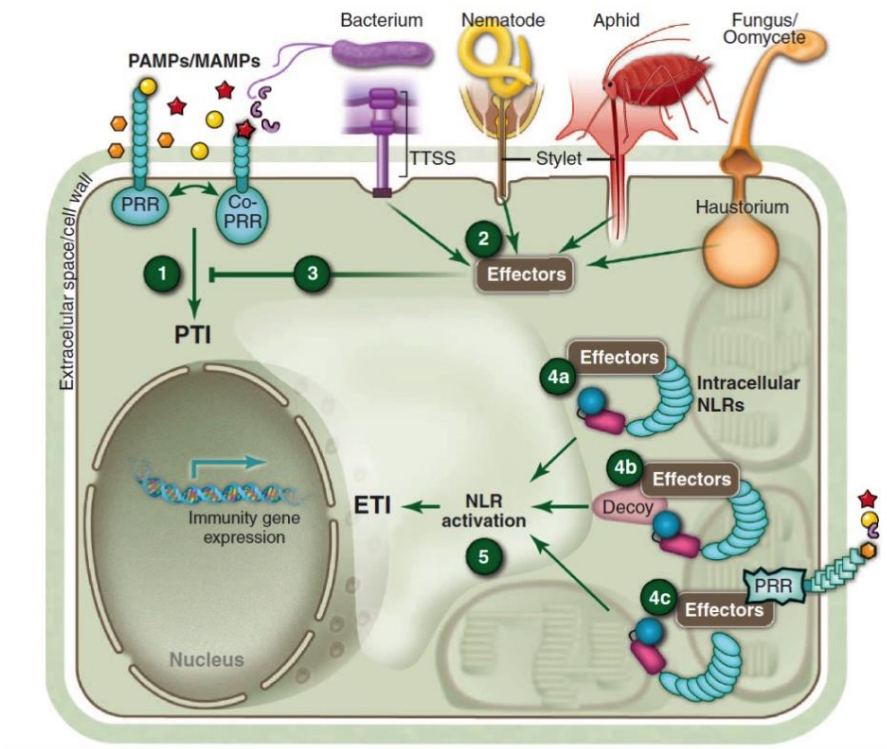


Figure 1 : Représentation schématique du système immunitaire des plantes

Les agents pathogènes expriment des PAMP (*pathogen-associated molecular patterns* en anglais) et MAMP (*microbe-associated molecular patterns* en anglais) lorsqu'ils colonisent les plantes (les couleurs sont codées selon les agents pathogènes). Les plantes les perçoivent via les PRR (*Pathogen recognition receptor* en anglais) extracellulaires et initient l'immunité (PTI *pattern-triggered immunity* en anglais; étape 1). Les pathogènes délivrent alors des effecteurs de la virulence, à la fois dans l'apoplaste des cellules végétales et à l'intérieur des cellules de la plante (étape 2). Ces effecteurs sont adressés à des emplacements subcellulaires spécifiques où ils peuvent réprimer la PTI et permettre l'établissement de la virulence (étape 3). Des récepteurs NLR (*Nod-like receptor* en anglais) intracellulaires peuvent percevoir les effecteurs de différentes manières : par l'interaction directe entre l'effecteur et le récepteur (étape 4a); par la détection de l'interaction entre l'effecteur et une protéine leurre qui mime structurellement une cible de l'effecteur (étape 4b) ; ou en détectant une modification d'une cible de l'effecteur dans la cellule hôte comme le domaine cytosolique d'une PRR (étape 4c). *Dangl et al., 2013*.

décrire le système immunitaire des plantes avec deux niveaux de récepteurs, l'un à la surface, et l'autre à l'intérieur de la cellule végétale (Figure 1 ; Dangl *et al.,* 2013). Ce modèle gomme ainsi l'aspect successif des phases PTI et ETI décrites dans le modèle de 2006. Le premier niveau du système immunitaire des plantes est constitué de récepteurs extracellulaires reconnaissant les motifs moléculaires (*pattern-recognition receptors* en anglais ; PRR étant l'acronyme établi dans le domaine) (Boutrot & Zipfel, 2017). Les PRR sont des protéines modulaires composées typiquement (mais pas exclusivement) de domaines kinases et de domaines répétés riches en leucines qui sont activées par la reconnaissance des motifs moléculaires conservés des parasites. L'activation des PRR conduits à la PTI et la mise en place d'un ensemble de réponses physiologiques (par exemple, entrée de calcium dans la cellule, génération massive et transitoire d'espèces réactives de l'oxygène et activation de

protéines kinases ; Figure 2) conduisant à la limitation de la colonisation de l'organisme infectant le plante. Les parasites utilisent leur répertoire d'effecteurs pour contourner ou interférer avec les réponses PRR-dépendantes (Dangl *et al.,* 2013).

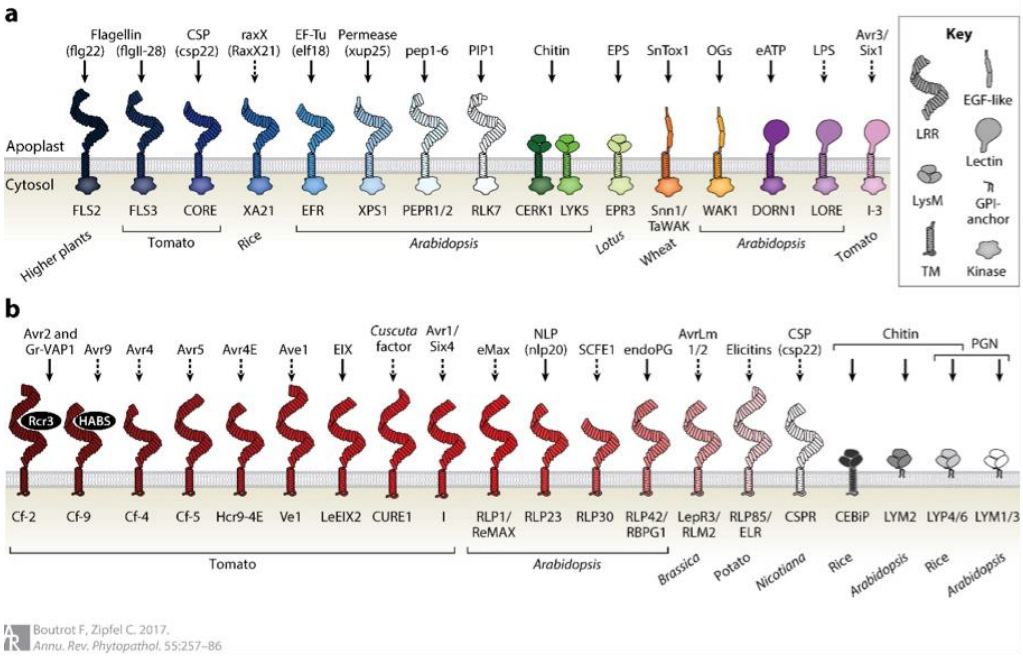


Figure 2 : Récepteurs PRR caractérisés et/ou identifiés avec leurs ligands

(a) Récepteurs RK (*receptor kinase* en anglais), (b) Récepteurs RLP (*receptor-like proteins* en anglais) ; les flèches noires indiquent les démonstrations d'interaction directes ; les flèches en pointillés indique que l'interaction directe n'a pas été montrée. Abréviations: EGF, *epidermal growth factor* ; EIX, *ethylene-inducing xylanase* ; EPS, *extracellular polysaccharides* ; GPI, *glycophosphatidylinositol* ; LPS, *lipopolysaccharide* ; LRR, *leucine-rich repeat* ; OGs, *oligogalacturonides* ; PGN, *peptidoglycan ;* TM, *transmembrane. Boutrot & Zipfel, 2017*

Ces répertoires d'effecteurs ont été décrits chez de nombreuses espèces d'organismes associés aux plantes comme les bactéries et leurs effecteurs délivrés dans la cellule par le système de sécrétion de type III (*Type Three Secretion System*, TTSS ; (Block & Alfano, 2011) ; les oomycètes et champignons pathogènes dont l'haustorium est considéré comme le siège de sécrétion des effecteurs, bien que le mécanisme contrôlant passage de la matrice apoplastique entourant ce type cellulaire (matrice extrahaustoriale) au cytoplasme de l'hôte ne soit pas encore résolu (Petre & Kamoun, 2014); les nématodes et pucerons qui sécrètent des effecteurs par leurs glandes salivaires qui sont ensuite délivrés dans le cytoplasme de l'hôte via un stylet (Jaouannet *et al.,* 2014 ; Mitchum *et al.*, 2013). Une fois sécrétés dans la cellule hôte, ces effecteurs peuvent être reconnus par les plantes via leurs récepteurs intracellulaires qui forment le deuxième niveau du système immunitaire des plantes (Figure 1 ; Dangl *et al.,* 2013). Jusqu'à présent, la plupart des gènes de résistance appartiennent à une famille multigénique très diversifiée et polymorphe chez les plantes qui codent des récepteurs intracellulaires modulaires de type *Nucleotide-binding leucine rich repeat* (NBLRR) aussi

appelé *NOD-like receptors* (NLR) (Dodds & Rathjen, 2010). Les récepteurs NLR sont activés par une reconnaissance spécifique d'un ligand donné (par exemple : la reconnaissance directe d'effecteurs, ou la reconnaissance d'une cible modifiée par un effecteur ; Figure 3 ; Cui *et al.,* 2015).

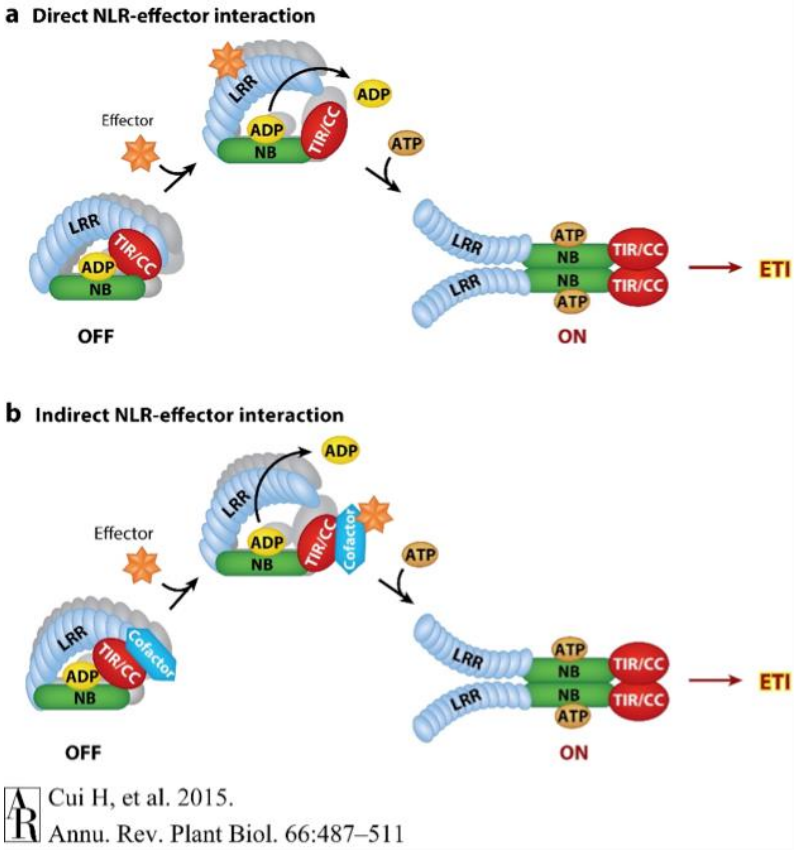


Figure 3 : Mécanismes de reconnaissance des effecteurs par les récepteurs NLR conduisant à l'ETI.

La reconnaissance spécifique d'un effecteur dans la cellule hôte se fait de façon directe ou indirecte. Un récepteur NLR est maintenu en état OFF en l'absence d'effecteur par inhibition intramoléculaire due au domaine superhélice en N-terminal (*coiled-coil*, en anglais ; CC). Pour un récepteur TIR (*Toll-interleukin 1* en anglais) c'est le repliement de son domaine C-terminal riche en leucines répétées (*leucine-rich-repreat* en anglais ; LRR) qui maintient l'état OFF. L'état OFF peut également être maintenu de façon intermoléculaire par les régions N-terminales de dimères de NLR (*nod-like receptor* en anglais). Les NLR peuvent reconnaître des effecteurs directement (panel a) ou indirectement par la reconnaissance de modification d'une cible d'un effecteur (panel b). La reconnaissance directe ou indirecte d'un effecteur entraîna l'activation du complexe NLR par l'échange d'un ADP (adénosine diphosphate) en ATP (adénosine triphosphate) dans le domaine NB (*nucleotide-binding domain* en anglais) et une série de changements conformationnels. Le passage à l'état ON du NLR initie la voie de signalisation de l'ETI (*Effector-triggred immunity* en anglais). *Cui et al., 2015*

L'activation des récepteurs NLR entraine la réponse ETI qui active la voie de signalisation et de régulation des réponses physiologiques de défense permettant la limitation de la prolifération pathogène. Les réponses de défense incluent la production de molécules antimicrobiennes (par exemple, les protéines associées à la pathogénèse dite *pathogenicity-*

related ou protéines PR et des peptides antimicrobiens comme les thaumatines, les chitinases ou des métabolites secondaires comme les phytoalexines), la sécrétion de molécules toxiques, le dépôt de callose pour limiter physiquement la progression des parasites. Le niveau de réponse observé lors de l'ETI peut être plus impotant que lors de la PTI, notamment avec la mise en place d'une réponse de type hypersensible (*hypersensitive response*, HR).Si les modèles conceptualisés ci-avant se situent à l'échelle de la cellule, certaines réponses de défense sont propre à un type cellulaire donné ou bien dépassent l'échelle cellulaire (Figure 4). Les hormones végétales, notamment l'acide salicilique (SA), l'acide jasmonique (JA) ou l'éthylène, jouent également des rôles clés et coopératifs ou antagonistes dans l'activation des réponses associées à l'ETI (Beckers & Spoel, 2006). Il a été montré récemment que le JA joue un rôle dans l'établissement de l'ETI par un mécanisme impliquant deux récepteurs de SA appelés NPR3 et NPR4 qui interagissent avec les répresseurs de transcription JAZ aux stades précoces de l'ETI activant ainsi l'expression de la voie de biosynthèse du JA (Liu *et al.,* 2016). Les défenses immunitaires mises en place par les plantes consistent donc en une réponse complexe et multiple associée à des modifications rapides des voies de signalisation au sein des cellules individuelles pour obtenir une réponse plus globale à un niveau systémique.

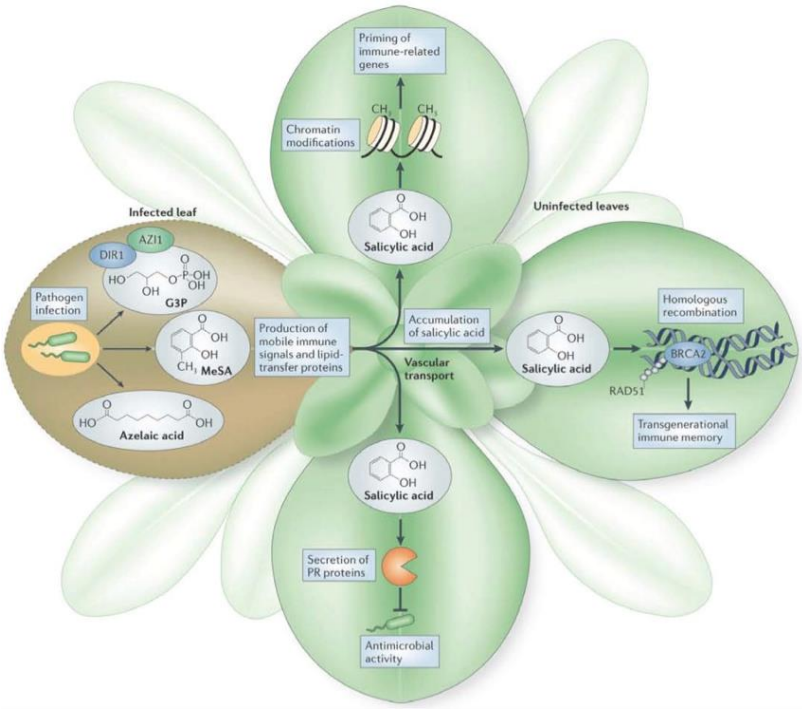


Figure 4 : Réponse systémique acquise, mémoire transgénérationnelle et signaux associés

L'infection localisée provoquée par le parasites induit l'accumulation de molécules signales mobiles comme l'acide azélaiques, les salcylates de méthyle et le glycérol-3-phosphate, ainsi que la protéine DIR1 (*defective induced resistance 1*, en anlgais). Les signaux mobiles sont transportés par le système vasculaire dans les tissus non-infectés. L'acide salicylique stimule l'expression de protéines PR et induit des modifications épigénétiques de l'ADN permettant de transmettre une mémoire immunitaire aux générations suivantes. *Spoel & Dong, 2012*

Si les modèles décrits plus haut sont très largement acceptés et repris par la communauté scientifique pour replacer les principaux éléments de la réponse immunitaire face des parasites chez les plantes, de nombreux cas n'entrent pas dans ce schéma et sont autant d'exceptions qui appellent la mise en place de nouveaux modèles, ou au moins à souligner les cas qui sortent du cadre général. Par exemple, si le modèle général s'applique parfaitement aux pathogènes biotrophes, il en est tout autre pour les pathogènes nécrotrophes. Aussi, certains PAMPs peuvent entrainer des HR (par exemple : la glycoprotéine CBEL chez *Phytophthora parasitica* var *nicotianae,* Khatib *et al.,* 2004; la famille des protéines PWL chez *Magnaporthe oryzae*, Khang *et al.,* 2010; Valent & Khang, 2010) ou des récepteurs de type PRR peuvent participer à la reconnaissance d'effecteurs dans le cadre de l'ETI (par exemple : l'effecteur bactérien Ax21 reconnu par le récetpeur PRR Xa21 chez le riz ; Lee *et al.*, 2009). Ainsi des articles font régulièrement état du cadre trop strict du modèle courant (Figure 5 ; par exemple : Cook *et al.,* 2015 ; Pritchard & Birch, 2014 ; Thomma *et al.,* 2011 détaillent les différents cas de figures sortant du cadre).

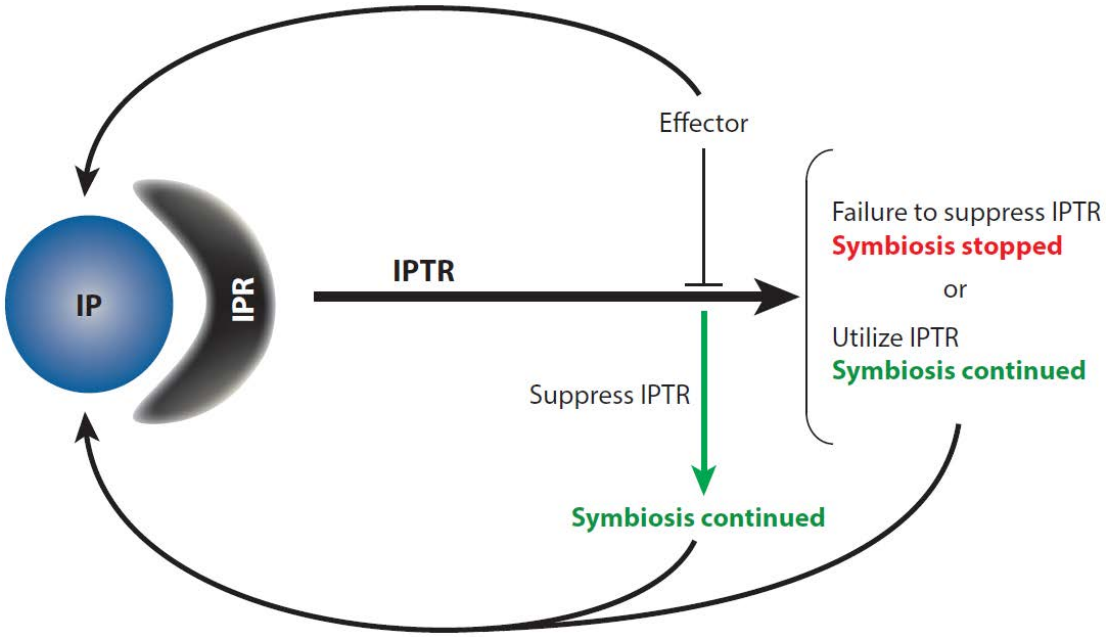


Figure 5: Description du modèle d'invasion.

Le modèle d'invasion permet de montrer que la séparation des rôles et des réponses des molécules immunitaires est floue. Dans ce modèle les ligands sont appelés IP (*invasion patterns*, en anglais) qui sont perçus par les récepteurs IPR (*invasion pattern receptor*, en anglais) conduisant à la réponse immunitaire IPTR (*IP-triggered response,* en anglais). La réponse IPTR peut être supprimée (ex. biotrophes) ou utilisiée (ex. nécrotrophes) pour promouvoir la colonisation. *Cook et al., 2015*

De même, il a pu être montré que des micro-organismes mutualistes biotrophes peuvent utiliser des mécanismes comparables à ceux des microorganismes pathogènes lors de l'établissement d'une interaction compatible avec leur hôte (Kloppholz *et al.,* 2011 ; Plett *et al.,* 2014) Ainsi près de 10 ans après la proposition du modèle en zigzag proposé par Jones &

Dangl (2006), des chercheurs proposent des modèles alternatifs plus intégratifs et permettant de prendre en considération tous les cas de figures lors de l'établissement d'une symbiose (*sensu stricto*, association mutualiste ou pathogène), tel que le modèle d'invasion (Figure 5 ; *Invasion Model* en anglais ; Cook *et al.,* 2015). Ce modèle est présenté comme un système général de surveillance d'une invasion chez les plantes. La PTI et l'ETI sont considérées sous un même angle avec des effecteurs et des PAMPs considérés dans un ensemble plus large de molécules appelées des motifs d'invasion (*Invasion Patterns* en anglais ; IP) qui peuvent être reconnues par des récepteurs d'IP, appelés IPR déclenchant alors une réponse appelée IPTR pour *IP-triggered response* (Figure 5 ; Cook *et al.,* 2015). Ce type de modèle qui a l'avantage d'être plus intégratif n'a pas encore supplanté les modèles précédents toujours en vogue pour l'ensemble de la communauté des phytopathologistes moléculaires.

3. Les Effecteurs : concepts et définitions

Les organismes parasites des plantes ont développé des stratégies pour éviter les différents niveaux de reconnaissance de l'hôte, comme modifier la physiologie et la structure des cellules hôtes pour détourner les nutriments nécessaires à leur développement et réaliser leur cycle de vie (Gaulin, 2017). Les pathogènes sécrètent un arsenal de molécules appelées « effecteurs » dont les fonctions variées visent à promouvoir le succès de l'infection (Hogenhout *et al.,* 2009 ; Win *et al.,* 2012). La notion d'effecteur a été utilisée pour la première fois dans le domaine des interactions plante-microorganisme avec la découverte de protéines sécrétées par les bactéries gram-négatives dans les cellules hôtes (Abramovitch *et al.,* 2006 ; McCann & Guttman, 2008). Ces effecteurs protéiques ont d'abords été décrits comme facteurs d'avirulence pour leur propriété de déclencheurs de réponses immunitaires spécifiques (Block & Alfano, 2011 ; Zhou & Chai, 2008). Ce qui a pu un temps générer une vision biaisée des effecteurs en les considérant essentiellement comme des immunosuppresseurs. Bien que de nombreux effecteurs bactériens et de pathogènes filamenteux sont en effet impliqués dans la suppression des réponses immunitaires, beaucoup d'effecteurs ont montré des fonctions bien plus diverses telles que des fonctions enzymatiques (par exemple : AvrPphB et AvrRpt2 des protéases à cystéines de *Pseudomonas syringae* spp), inhibiteur de kinase AvrPto, de *P. syringae*, acétytransférase HopZ1a de *P. syringae*, activateur de transcription chez *Xanthomonas campestris* (*transcription activator like* en anglais ; TAL ; voir synthèse (Deslandes & Rivas, 2012). La définition actuelle décrit les effecteurs comme des molécules produites par un microorganisme qui vont exercer un effet sur les cellules de la plante hôtes (Gaulin, 2017). Ce concept rappelle celui du « phénotype étendu » décrit par Richard Dawkins dans un livre éponyme et qui décrit des gènes dont l'effet va au-delà de la cellule dans laquelle il est exprimé (Dawkins, 1982).

La plupart des effecteurs décrits sont des protéines cependant, des effecteurs non protéiques ont également été décrits tels que des métabolites, des toxines (Amselem *et al.,* 2011; Arias *et al.,* 2012 ; Collemare & Lebrun, 2011) ou des petits ARN interférents (Weiberg *et al.*, 2013). Parmi les effecteurs protéiques, on peut distinguer deux classes : les effecteurs apoplastiques et les effecteurs cytoplasmiques. Les effecteurs apoplastiques sont sécrétés et fonctionnent dans l'apoplaste et la plupart ont été décrits comme des inhibiteurs des protéases de l'hôte (De Wit, 2016 ; Song *et al.,* 2009). Les effecteurs cytoplasmiques sont quant à eux transloqués dans le cytoplasme de la cellule hôte où ils peuvent cibler divers compartiments cellulaires et moduler un vaste spectre de fonctions cellulaires (Figure 6 ; Petre & Kamoun, 2014 ; Rovenich *et al.,* 2014). Les effecteurs protéiques étant plus principalement ciblés dans ce manuscrit, le terme général d'effecteur sera utilisé pour référer aux effecteurs protéiques.

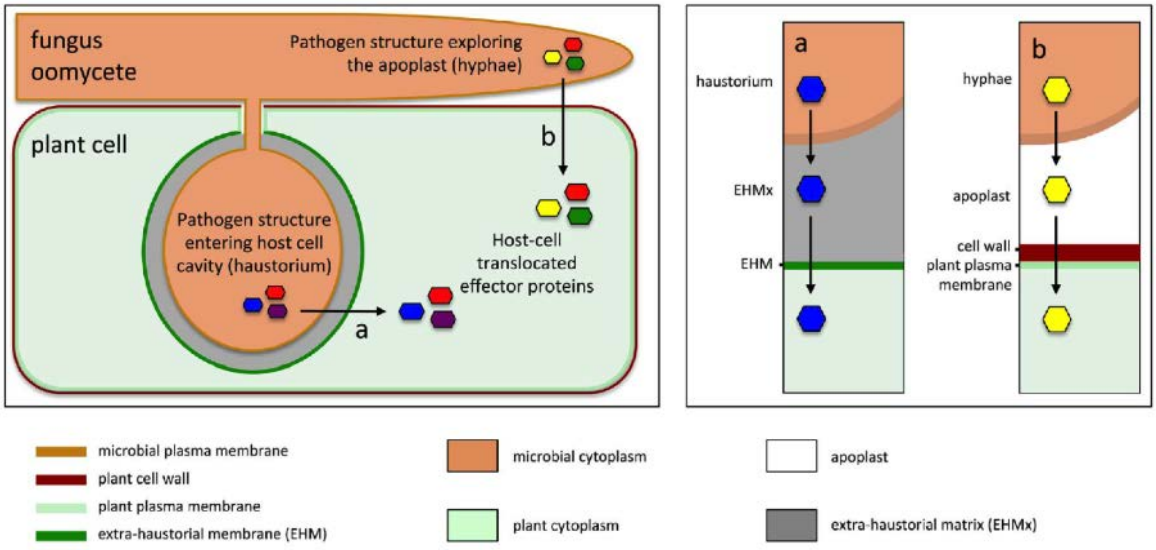


Figure 6: Structures de sécrétion des effecteurs des pathogènes filamenteux.

Panel de gauche : Les champignons et oomycètes pathogènes biotrophes possèdent des structures d'infection différenciés comme les hyphes extracellulaires, les hyphes d'invasion et les haustoria. Ces derniers pénètrent la cavité cellulaire et invaginent la membrane végétale. Haustoria (a) et hyphes (b) sécrètent des effecteurs qui sont transloqués dans le cytoplasme de la cellule hôte par un mécanisme encore inconnu. Panel de droite : Les effecteurs sécrétés depuis les haustoria (a) et les hyphes (b) traversent différentes inferfaces biologiques (matrice extra-haustoriale [EHMx]/ membrane extra-haustoriale [EHM] pour les effecteurs sécrétés par les haustoria et l'apoplaste/paroi cellulaire/membrane plasmique pour les effecteurs sécrétés par les hyphes). *Petre & Kamoun, 2014.*

Les effecteurs de pathogènes ont évolué pour interférer avec les mécanismes cellulaires et physiologiques de la cellule hôte à l'avantage du parasite (McCann & Guttman, 2008). Certains effecteurs identifiés surtout chez les bactéries présentent une activité enzymatique de type protéases, kinases ou ubiquitinases (Deslandes & Rivas, 2012). Chez le champignon pathogène *Ustilago maydis* responsable du charbon du maïs, un effecteur présentant une

activité de type chorismate mutases (Cmu) semble mimer l'activité de certaines enzymes pour affecter les procédés métaboliques liés à la production de métabolites secondaires en lien avec l'immunité (Djamei *et al.,* 2011 ; Tanaka *et al.*, 2014).D'autres effecteurs interagissent avec une protéine cible pour altérer sa fonction tels que les effecteurs inhibiteurs de protéases (Dong *et al.,* 2014). Les effecteurs cytoplasmiques sont souvent décrits comme ciblant des protéines végétales pour affecter leur stabilité (par exemple, en promouvant leur accumulation ou leur dégradation dans la cellule hôte) ou leur activité (par exemple, en inhibant leur activité enzymatique ou leur activité de transduction de signal ; (Win *et al.,* 2012).Certains effecteurs peuvent également se lier à l'ADN et moduler l'expression de gènes, comme par exemple les effecteurs *Transcription activator-like* (TAL) des bactéries *Xanthomonas* spp. (Boch *et al.,* 2009).

Au cours de la dernière décennie, l'étude de la biologie des effecteurs a reposé sur un pipeline d'analyse qui peut se découper en trois principales étapes (Figure 7): des effecteurs candidats sont tout d'abord identifiés par des approches de génomique et de transcriptomique (*effector-mining*) pour établir des répertoires d'effecteurs candidats ; des cribles à haut-débit (*effectoromics*) sont appliqués aux candidats pour confirmer des effecteurs (aussi dits effecteurs *bona fide*) ou mettre la priorité sur des candidats prometteurs ; enfin, les quelques effecteurs candidats mis en évidence par les cribles effectoromiques sont analysés en détails dans le but de comprendre leurs fonctions dans les cellules hôtes et comment ils participent au succès de l'infection du parasite (Petre *et al.,* 2014; **article 1**). Les cribles d'effectoromique peuvent concerner la localisation des effecteurs dans la cellule végétale, l'impact sur la réponse immunitaire (suppression ou induction de la réponse PTI/ETI) ou encore l'identification de cibles potentielles.

4. Génomique des interactions plante-microorganisme

Depuis les années 2000s, l'entrée de la biologie dans l'ère de la génomique a conduit à une transition depuis le séquençage de génomes uniques d'espèces eucaryotes modèles vers un séquençage systématique de différents individus au sein de différents taxons dans un règne donné. Depuis le premier pathogène filamenteux séquencé en 2003 (*Neurospora crassa* ; Galagan *et al.,* 2003), l'amélioration de la vitesse et de la profondeur de séquençage des nouvelles technologies mises en œuvre, a fait de la génomique un domaine de recherche établi et un front de science très actif.

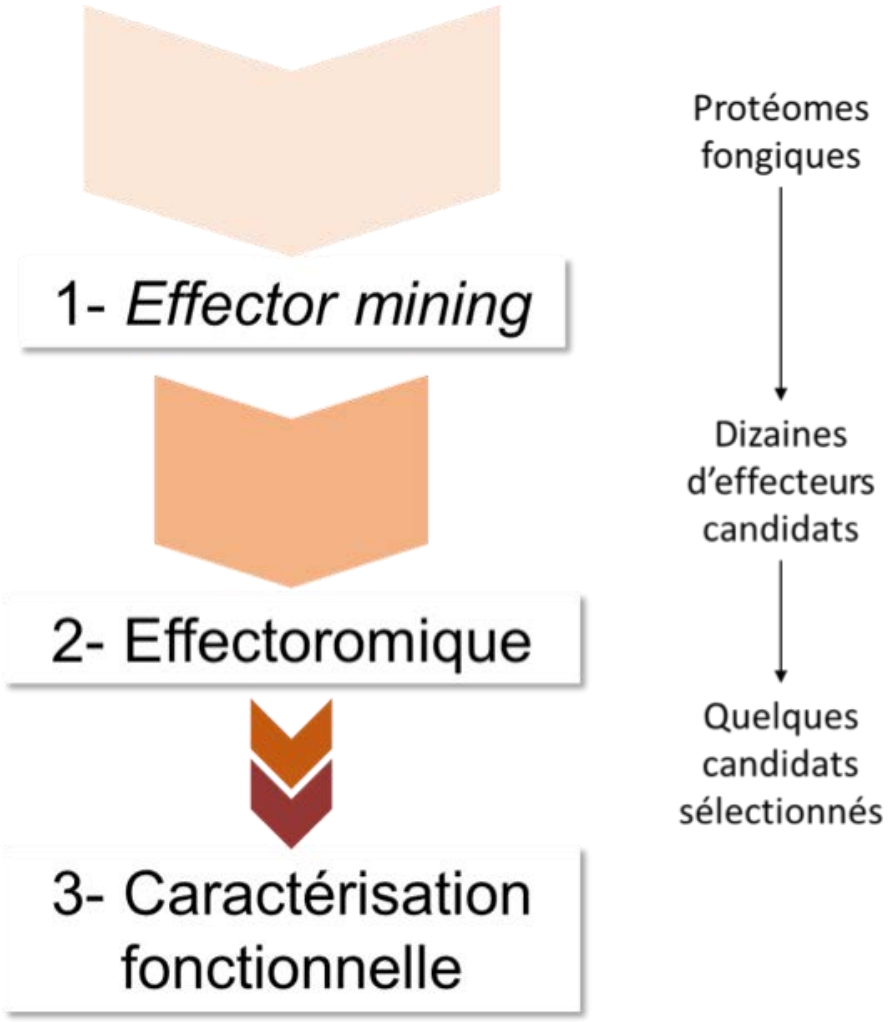


Figure 7 : Les trois étapes de la biologie des effecteurs

La plupart des études sur les effecteurs des pathogènes filamenteux utilisent le même pipeline d'indentification des effecteurs candidats par des approches globales de génomique et transcriptomique (*effecteor-mining*, en anglais). Les effecteurs candidats sont criblés pour leurs propriétés d'avirulence/virulence et/ou leurs cibles cellulaires et moléculaires par effectoromique. Enfin quelques candidats mis en évidence sont étudiés plus en détails pour comprendre leur fonction dans le succès de l'infection.

Des initiatives génomiques à grande échelle se multiplient comme par exemple le projet « 1000 génomes fongiques » rassemblant des consortia de recherche internationaux en collaboration avec le Joint Genome Institute (JGI) du département de l'énergie américain (https://jgi.doe.gov/). De plus, les progrès continus de la bioinformatique ont favorisé cette révolution. Des génomes de haute qualité peuvent maintenant être séquencés et assemblés et des projets visant l'ensemble des informations génétiques portées par l'ADN de type *Encyclopedia of DNA elements* (ENCODE) ont permis des avancées majeures dans la

cartographie de génomes d'espèces modèles (Hug *et al.,* 2016). Cependant, l'obtention de génomes de référence complets n'est pas des plus faciles. En effet, des problèmes d'assemblage et d'annotation sont souvent rencontrés lorsque les génomes présentent une grande complexité (éléments répétés, fort polymorphisme entre haplotypes). Le coût pour séquencer un génome a très largement diminué avec l'émergence des dernières générations des technologies de séquençage. Ainsi, certaines technologies peuvent générer beaucoup de fragments courts avec une bonne fidélité, alors que d'autres permettent de séquencer de très longs fragments avec un taux d'erreur encore élevé. Les limites majeures à l'étude des génomes eucaryotes sont en général, l'isolation et l'intégrité de l'ADN utilisé pour réaliser le séquençage et l'assemblage des fragments obtenus pour des génomes complexes riches en éléments répétés (Aime, *et al.,* 2017).

L'analyse des génomes de microorganismes filamenteux a montré que la plupart présentent un génome de taille modeste comprise entre 30 et 60 Mb (Raffaele & Kamoun, 2012). Certains possèdent des génomes de plus petites tailles comme par exemple certains Pucciniomycètes ou Ustilaginomycètes (par exemple *Mixia osmundae*, Toome *et al.,* 2014 ou *Ustilago maydis ;* Kamper *et al.,* 2006). D'autres champignons présentent au contraire des génomes de taille bien plus importante comme les Ascomycètes *Tuber melanosporum* (125 Mb ; Martin et al., 2010) ou *Blumeria graminis* (~120 Mb ; Spanu *et al.,* 2010) ou encore les génomes des Pucciniales (**cf. partie III article 1**). Les génomes des champignons associés aux plantes sont marqués par des répertoires de gènes spécifiques liés à l'interaction avec l'hôte (Duplessis *et al.,* 2014; Raffaele & Kamoun, 2012). Des gènes codant des enzymes impliquées dans la décomposition des parois végétales (par exemple les *carbohydrate-active enzymes* ou CAZymes), les transporteurs de nutriments et d'eau, ou les effecteurs candidats sont autant de catégories cellulaires essentielles au succès de l'infection et que l'on retrouve en nombre variable dans les génomes des champignons phytopathogènes. En effet, le répertoire de gènes codants les CAZymes peut être associé au mode de vie du parasite (mais aussi chez des champignons non pathogènes saprotrophes). Les biotrophes obligatoires par exemple ou les mutualistes possèdent un répertoire réduit de gènes codant les CAZymes. Cette caractéristique pourrait permettre de minimiser le risque de déclenchement de réactions de défense chez l'hôte lors de la colonisation (Duplessis *et al.,* 2014).

Les effecteurs candidats peuvent être identifiés dans les génomes des micro-organismes pathogènes filamenteux en ciblant des caractéristiques communément partagées. Ainsi les gènes codant des effecteurs sont souvent spécifiques de l'espèce, ils codent des protéines sécrétées qui portent le plus souvent un peptide signal nécessaire à la sécrétion (Sperschneider *et al.,* 2017). La spécificité des gènes des effecteurs à certains rangs

taxonomiques pourrait refléter l'évolution avec leur(s) hôte(s). Par ailleurs, le core-génome conservé chez des espèces de rang taxonomique plus ou moins proche, coderait des fonctions cellulaires et physiologiques basales et essentielles (Duplessis *et al.,* 2011). Les recherches systématiques d'effecteurs candidats au sein des sécrétomes prédits dans les génomes d'organismes phytopathogènes ont révélé des centaines, voire des milliers, de protéines sécrétées propres à chaque génome. Les effecteurs candidats ainsi sélectionnés sont ensuite généralement sélectionnés sur la base de leur profil d'expression au cours de l'infection à travers des approches de transcriptomique (cf. **partie 3, article 2**). Chez les oomycètes, des familles d'effecteurs candidats ont aussi pu être identifiés sur la base de motifs particuliers conservés en aval du peptide signal (par exemple les familles d'effecteurs RxLR ou CRN ; Bozkurt *et al.,* 2012). La fonction de ces motifs est encore inconnue. Certaines publications très discutées au sein de la communauté scientifique avaient incriminé le motif RXLR dans le processus de translocation dans le cytoplasme des cellules hôtes, cependant il a récemment été montré que le motif RxLR de l'effecteur Avr3a de *P. infestans* était clivé avant sa sécrétion (Hofmann, 2017 ; Wawra *et al.,* 2017). Le séquençage de génomes de pathogènes filamenteux a donc permis de réaliser des progrès très rapides dans l'édification de répertoires d'effecteurs candidats spécifiques à de nombreuses espèces phytopathogènes et d'ouvrir des perspectives pour l'étude de leur rôle dans les processus d'infection.

Des analyses de génomique comparative ont mis en lumière les forces évolutives agissant sur les répertoires de gènes codant des effecteurs (Guttman *et al.,* 2014). L'étude comparative des gènes codant les effecteurs des oomycètes *P. infestans*, *Phytophthora ramorum* et *Hyaloperonospora arabidopsis* a montré la présence de familles de gènes paralogues d'effecteurs à RxLR sous pression de sélection positive (Rinke *et al.,* 2013 ; Win *et al.,* 2007). Ce type de sélection indique qu'une pression de sélection est exercée sur les gènes étudiés et elle peut être estimée sur la base du ratio entre les substitutions non-synonymes et les substitutions synonymes (*dN/dS* ou *pN/pS*) observées entre deux séquences codantes paralogues ou orthologues (Kryazhimskiy & Plotkin, 2008).D'autres études de polymorphisme intra- et inter-spécifique ont montré que les signatures de sélection positive sont retrouvées principalement dans les gènes codant les effecteurs (ou effecteurs candidats ; Guttman *et al.,* 2014). Enfin, le re-séquençage d'isolats et l'analyse de leur polymorphisme intrinsèque combiné au développement d'outils statistiques permettent de réaliser des analyses de cartographie par *Genome wide associations* (GWAS). Cet outil est aussi puissant pour détecter des variants génétiques associés à la pathogénicité des microorganismes (Bartoli & Roux, 2017).

Chez les pathogènes filamenteux oomycètes et champignons, les sécrétomes prédits peuvent se montrer très importants au sein des génomes .Ces derniers sont également parfois marqués par une invasion d'éléments transposables et la présence de larges familles multigéniques (dont des gènes d'effecteurs candidats). Ces caractéristiques expliquent la taille importante de ces génomes par rapport aux autres microorganismes filamenteux non associés aux plantes, ce qui dénoterait une plasticité génomique plus conséquente permettant l'évolution des gènes en lien avec les mécanismes d'interaction (Raffaele & Kamoun, 2012). Une association entre l'environnement génomique des gènes codant des effecteurs candidats et des régions riches en éléments transposables a été mise en évidence dans les génomes de différents microorganismes phytopathogènes comme l'ascomycète *Leptospheria maculans* et l'oomycète *P. infestans* par exemple, suggérant ainsi que l'activité des éléments transposables pourrait jouer un rôle crucial dans la diversification des répertoires d'effecteurs candidats (Faino *et al.,* 2016 ; Haas *et al.,* 2009 ; Rouxel *et al.,* 2011).

5. L'interaction peuplier-*Melampsora larici-populina*

Le peuplier est une espèce endémique en France, faisant partie des écosystèmes ripariaux. Il est aussi utilisé en populiculture pour la production de bois. Le peuplier fut le premier arbre dont le génome a été séquencé, c'est également l'un des arbres les plus facilement manipulables en laboratoire, ce qui en fait une espèce modèle de référence en biologie forestière (Tuskan *et al.,* 2006). Le genre *Populus* regroupe les peupliers et trembles et appartient à la famille des Salicacées. Le genre *Populus* regroupe 29 espèces de peupliers auxquels il faut ajouter des hybrides naturels ou résultant de croisements. Les peupliers sont des arbres dioïques à croissance rapide, avec un feuillage caduque à feuilles simples. Ils poussent sur terrain humide, typiquement dans des ripisylves d'Europe, d'Amérique de Nord et d'Asie. En volume, le peuplier est la deuxième essence feuillue récoltée en France, derrière le chêne. En général des hybrides interaméricains (*Populus thricocarpa x Populus deltoides*) ou euraméricains (*P. deltoides x Populus nigra*) sont cultivés dans les peupleraies. Les grumes sont récoltées à 15-20 ans pour la production de bois d'emballage alimentaire et de bois énergie. En France, la principale limitation à la populiculture vient des épidémies annuelles de rouilles foliaires causées par les champignons du genre *Melampsora*.

Melampsora larici-populina, un des agents de la rouille du peuplier (Basidiomycète, Pucciniomycète) appartient à l'ordre des Pucciniales (cf. **partie 2, article 1**) qui représente le plus large groupe de champignons phytopathogènes, avec plus de 8000 espèces décrites (Aime *et al.,* 2014). Les Pucciniales sont des biotrophes obligatoires qui sont inféodés à leurs hôtes pour réaliser leur cycle de vie. A lui seul, le genre *Melampsora* compte plus d'une

cinquantaine d'espèces dont 17 sont pathogènes des peupliers (genre *Populus* ; Vialle *et al.,* 2011). Sur les neuf espèces présentes en Europe (dont *Melampsora allii-populina* et *Melampsora medusae* f. sp. *deltoidae*), *M. larici-populina*, est celle qui cause les dégâts les plus importants dans les peupleraies françaises (Pinon & Frey, 2005). *M. larici-populina* possède un cycle de vie hétéroïque (c'est-à-dire qu'il réalise son cycle de vie sur deux hôtes) et macrocyclique (produisant cinq types de spores différents), typique des Pucciniales. La reproduction sexuée est effectuée sur l'hôte dit « écidien », le mélèze au début du printemps. La multiplication asexuée se produit durant la période estivale sur un second hôte dit « télien », le peuplier. Les symptômes occasionnés sur ces deux hôtes, sont très différents. Ainsi, l'infection des aiguilles de mélèze par *M. larici-populina* est marquée par des symptômes limités dans le temps sans impacter le développement ou la croissance de l'hôte. Au contraire, l'infection des peupliers par *M. larici-populina* a des conséquences considérables sur ces derniers notamment par le détournement des photosynthétats des feuilles infectées au profit du champignon. De plus, la multiplication intense du champignon durant la période estivale provoque une défoliation partielle ou totale des arbres ce qui diminue la mise en réserve des substances carbonées et azotées en automne. Les attaques répétées peuvent engendrer des pertes de croissance allant jusqu'à 60% et un affaiblissement général des arbres, pouvant causer leur mort (Pinon & Frey, 2005). A ce jour, toutes les résistances complètes (interaction incompatible) à la maladie de la rouille causée par ce champignon sélectionnées par les améliorateurs et déployées dans les peupleraies ont été contournées par *M. larici-populina*. Le statut pérenne des arbres implique que ces résistances complètes doivent se maintenir pendant des années dans des plantations monoclonales. Cela crée donc une forte pression de sélection et favorise les contournements des résistances et l'émergence de nouvelles virulences dans les populations du pathogène fongique. Ainsi, un évènement majeur de contournement de la résistance R7 des peupliers cultivés en France et en Europe du Nord depuis 1982, par *M. larici-populina* a eu lieu en 1994 (Persoons *et al.,* 2017). Les populations de *M. larici-populina* portant la virulence correspondante (Vir7) se sont rapidement dispersés sur les hôtes sensibles. Une approche de génétique des populations a révélé que le contournement de la résistance R7 a permis l'émergence d'un groupe génétique unique et homogène de *M. larici-populina* qui a supplanté la population initiale présente avant le contournement (Persoons *et al.,* 2017).

Les progrès réalisés à travers l'étude des génomes et transcriptomes du peuplier et de *M. larici-populina*, placent ce pathosystème comme un modèle avancé pour l'étude des interactions moléculaires entre arbres et parasites (Duplessis *et al.,* 2011a ; Duplessis *et al.,* 2011b ; Hacquard *et al.,* 2010 ; 2012 ; 2013 ; Joly *et al.,* 2010; Pernaci e*t al.,* 2014 ; Persoons *et al.,* 2017 ; Petre *et al.,* 2012 ; Tuskan *et al.,* 2006). C'est aussi un modèle avancé

pour l'étude des rouilles en général. Toutefois, le couple établi entre une plante pérenne d'une part et un champignon biotrophe obligatoire d'autre part n'est pas un modèle au sens où on pourrait l'entendre pour des aspects de dissection génétique de mécanismes moléculaires des interactions comme cela peut être le cas dans un couple comme *Arabidopsis thaliana* et *Pseudomonas syringae*. Depuis 2011, les annotations et prédictions *in sillico* du sécrétome de *M. larici-populina* ont contribué à mettre en évidence un large répertoire de gènes codant des protéines sécrétées (« secreted proteins », SP). *M. larici-populina* possède plus d'un millier de gènes SP de fonctions inconnues qui pourraient correspondre à des effecteurs (Duplessis *et al.,* 2011a ; Hacquard *et al.,* 2012). A ce jour, les mécanismes moléculaires sous-jacents à la colonisation des cellules foliaires de peuplier par *M. larici-populina* restent à élucider, notamment la compréhension du rôle potentiel des effecteurs candidats mis à jour.

Pour conclure cette introduction, deux articles de synthèse sont présentés dans les Parties 2 et 3 qui font suite. L'un correspond à un article de synthèse en préparation qui fait l'état des avancées récentes sur l'étude des Pucciniales responsables des maladies de rouilles chez les plantes avec un focus plus particulier sur les analyses fonctionnelles mises en place chez différentes espèces causant des maladies de rouille des plantes. L'autre est une courte synthèse qui porte sur la prédiction du sécrétome chez *M. larici-populina*, une approche clef pour l'identification d'effecteurs candidats chez ce champignon pathogène.

Partie 2: Biologie, génomique et analyse fonctionnelle d'effecteurs des Pucciniales

Cette partie correspond à un article de synthèse en préparation (et non finalisé dans sa version présentée dans la thèse) rédigé en anglais à l'invitation de la revue internationale *The New Phytologist* dans la série des *Tansley Reviews* et pour lequel je suis première auteure. Dans cet article de synthèse intitulé « **Advances in understanding obligate biotrophy in rust fungi** », nous présentons des éléments très généraux de la biologie des Pucciniales (taxonomie, maladies causées, cycles de vie, évolution) puis les avancées les plus récentes réalisées en génomique d'une part et plus particulièrement sur l'identification et l'analyse d'effecteurs candidats. La partie portant sur la génomique reste volontairement modérée, une synthèse récente ayant été publiée sur ce sujet par un de mes directeurs de thèse et à laquelle nous nous référons (Aime *et al.,* 2017). Après un travail exhaustif de fouille bibliographique et d'analyse des données publiées, j'ai personnellement pris en charge la rédaction de la partie principale de l'article qui porte sur la génomique et l'analyse fonctionnelle d'effecteurs chez les rouilles. Cette partie sert donc d'introduction très générale aux chapitres qui font suite dans cette thèse.

Article n°1 : Advances in understanding obligate biotrophy in rust fungi

Cécile Lorrain, Karen Conclaves, Hugo Germain, Arnaud Hecker et Sébastien Duplessis

Article de synthèse invité en préparation dont la soumission dans la revue *The New Phythologist* est prévue en février 2018

Advances in understanding obligate biotrophy in rust fungi

Cécile Lorrain[1], Karen Gonçalves[2], Hugo Germain[2], Arnaud Hecker[1], Sébastien Duplessis[1]

[1] UMR 1136 INRA/Université de Lorraine Interactions Arbres/Microorganismes, INRA Centre Grand Est - Nancy, 54280 Champenoux, France
[2] Department of Chemistry, Biochemistry and Physics, Université du Quebec à Trois-Rivières, Trois-Rivières, QC, Canada G9A 5H7.

The New Phytologist, Tansley Review

Corresponding author
DUPLESSIS Sébastien
UMR 1136 Interactions Arbres/Micro-organismes
INRA Centre Grand Est - Nancy
54280 Champenoux
France
Tel, +33 383 39 40 13
Fax, +33 383 39 40 69
sebastien.duplessis@inra.fr

Abstract

Rust fungi are the largest group of plant pathogens and one of the most threatening for major crops in modern agriculture. These obligate biotrophs can only grow and reproduce on their living host plants, which makes fundamental progress on understanding their biology very complex. Despite their economical importance, many aspects of their biology remain obscure. The dawn of genomics has opened avenues for identifying candidate genes underlying pathogenesis mechanisms in rust fungi. Despite the fact that rust fungi also possess among the largest and most repetitive genomes in fungi, transcriptomics and predictive tools have supported the identification of key genetic determinants necessary to set a successful infection. Thousands of genes encoding secreted proteins highly expressed during host infection have been reported in different rust species and represent a tremendous potential for rust candidate effectors. In the past three years, the systematic application of effectoromic screens relying essentially on heterologous plant systems has moved fast the field ahead, pointing in the direction of priority effectors for functional studies and unravelling new avirulence genes. These advances on understanding rust effector biology are as many promises to inform and speed up rust diseases management.

Key-words: Pucciniales, fungal genomics, rust effectors, effectoromic screens, heterologous systems, virulence, pathogenesis.

I - Introduction

Obligate biotrophy is widespread across plant pathogens (Kemen & Jones, 2012). Plant pathogenic fungal biotrophs extend to powdery mildews, smuts and rusts and share common features with biotrophic oomycetes (Kemen & Jones, 2012; Duplessis et al., 2013). As obligate biotrophs, rust fungi can only feed, grow and reproduce on their living host(s) and they differentiate specific infection structures called haustoria that are necessary to set intimate interactions inside infected host tissues (Voegele & Mendgen, 2011; Kemen et al., 2015). Rust fungi co-evolved with their host plants with which they exhibit a high specificity of interaction (Aime et al., 2017). The dependence on the host implies an inability to grow on artificial media, rendering difficult their manipulation under laboratory conditions (Williams, 1984). Also, the full completion of a rust lifecycle was never achieved outside a host.

Understanding the basis of obligate biotrophy and how evolution has shaped this particular trophic mode have not yet been resolved. Recent progresses under the angle of genomics have provided some cues about missing functions or over- and under-represented gene categories that may support biotrophic features (Spanu, 2012). However, comparison of general genomic features like genome size, repeat content or number of predicted genes did not reveal a single evolutionary trajectory towards this trophic mode, rather supported convergent evolution (Kemen et al., 2015; Duplessis et al., 2013). A common feature is the presence of a large secretome with over one to two thousands predicted genes encoding secreted proteins in rust fungi (Lo Presti et al., 2015; Aime et al., 2017). Considering the current models in molecular phytopathology (Jones & Dangl 2006; Dodds & Rathgen 2009; Cook et al., 2015), these secreted proteins likely contain effectors that are essential to establish a successful infection. With lists of thousands of putative candidates in rust fungi, the field is now moving towards the comprehension of the functional role of effectors in the manipulation of host immunity and physiology (Figueroa et al., 2016).

In this review, we present an overview of the recent progress made in the dissection of molecular processes underlying biotrophy in rust fungi. We first report on the peculiar life cycle of rust fungi and the current knowledge of the biology of infection. We then illustrate how the genomic era has been a tremendous lever to foster our understanding of processes at play during plant infection by rust fungi. We show the recent pivotal change operated in rust research with the systematic study of pathogenicity determinants represented by effectors. Indeed, in the past five years, we observed an acceleration of functional studies in rust fungi. We particularly survey the most recent progresses before to address the next challenges to come in the study of rust pathogens.

II - Rust fungi: a diverse and serious threat to agriculture

The Pucciniales (rust fungi) represent one of the largest fungal orders with more than 8,000 species described so far, with up to 10 families and 250 new species described in the past decade (Toome-Heller, 2016). They have the ability to infect a wide variety of host plants from ferns to monocots and gymnosperms to angiosperms, which suggests an ancestral adaptation to the biotrophic lifestyle (Aime *et al.*, 2014). Worldwide, rust fungi are among the most serious threats for crops cultivation (e.g. wheat, soybean or coffee) (Pennisi, 2010; Dean *et al.*, 2012; Fischer *et al.*, 2012), as well as for trees including species widely used for wood production and bioenergy (e.g. poplar, eucalypt or pines) (Aime *et al.,* 2006; Sniezko *et al.,* 2012). To fend off rust diseases, the plant breeding industry has primarily put effort in the development of resistant plants. However, complete resistances have been overcome by rust fungi through the years (Ellis *et al.,* 2014). Among wheat rust fungi, the stem rust fungus *Puccinia graminis* f. sp. *tritici* has been a major threat, and although it has devastated wheat culture thoughout human history, it has been mostly contained for decades in most production areas in the world (Ellis *et al.*,,2014; Figueroa *et al.,* 2017). The emergence in Uganda in 1998 of the Ug99 races of the wheat stem rust fungus, virulent against major wheat resistance genes has put wheat cultivation at risk (Singh *et al.,*. 2011; Singh *et al.,* 2015).

The recent literature provides examples of loss estimates directly inputable to rust fungi. Overall rust fungi are causing important damages to wheat production totalling between 4 to 5 billions US$ of global annual losses (Figueroa *et al.,* 2017). The wheat yellow rust disease caused by *Puccinia striiformis* f. sp. *tritici* is affecting wheat production from 5% to 50% worldwide depending on the year with accounted annual losses of US$979 millions (Chen *et al.*, 2014; Beddow *et al.*, 2015). The wheat leaf rust *Puccinia triticina* is the most common and widely distributed wheat rust fungus and can lead to significant losses in number and weight of kernels (Bolton *et al.,* 2008). Since the 90s the Asian soybean rust fungus *Phakopsora pachyrhizi* has spread with important outbreaks in South America and with the current demand for soybean, this rust disease has caused important losses and high costs in fungicides, which were estimated in 2003 at two US billion for a single year (Goellner *et al.*, 2017), and with increasing costs in Brazil only of over US$ 2 billions yearly between 2003 and 2014 (Godoy *et al.*, 2016). Another remarkable rust pathogen threatening agriculture is the coffee leaf rust causal agent *Hemileia vastarix* that damages coffee cultivation across Africa, Asia and South America leading to US$ 2-3 billion losses per year (Talhinas *et al.*, 2017). Rust fungi are among pathogens that regularly threaten agriculture and they are carefully kept under surveillance for their serious ability to jeopardize some crop production worldwide (Pennisi, 2010; Fischer *et al.,* 2012; Singh *et al.,* 2011; Bueno-Sancho *et al.,* 2017).

III- The different facets of rust life cycles and their unresolved evolution

Rust fungi display complex life cycles that often require alternation on two unrelated host plants, and some have life cycles lacking a few stages (Aime *et al.*, 2017). Rust fungi are host specific pathogens. Some species infect a single host (e.g. *Melampsora lini* on flax, Lawrence et al. 2007), whereas others infect two hosts at different stages and eventually different plant species within a taxonomical clade (e.g. *Melampsora allii-populina* infects *Aliium*, *Arum*, *Muscari* and *Populus* spp.; Vialle *et al.,* 2011).

Many rust fungi display a life cycle with five different spore stages (pycniospores, aeciospores, urediniospores, teliospores and basidiospores) and alternation on two host plants (Figure 1). Some rust fungi can be macrocyclic and autoecious, and they produce the five spore stages on a unique host (e.g. the flax rust fungus *M. lini*, Lawrence *et al.,* 2007). During the life cycle, the sequential order of stages is typical of macrocyclic rust fungi. The most reduced known rust life cycles are termed microcyclic and they are composed of only two types of spores: basidiospores and teliospores or pycniospores and teliospores (Figure 1). Demicyclic rust fungi lack the urediniospore stage even though they can be autoecious or heteroecious. For instance, *Gymnosporangium juniper-virginiane* horn shaped teliospores grow on juniper in spring. Teliospores germinate to form basidia and basidiospores. Basidiospores germinate on apple tree to from pycnia and pycniospores. After fertilization, aecia produce aeciospores to infect juniper tree where teliospores are then formed (Mims, 1977; Pearson *et al.,* 1977, Figure 1). Hemicyclic rust fungi exhibit only teliospores and urediniospores and are autoecious (Aime *et al.,* 2017, Anikster *et al.,* 2004, Figure 1). The complete life cycle of some rust fungi remains unknown. For instance, the alternate hosts of rust species of utmost importance like the soybean rust fungus *P. pachyrhizi* or the coffee rust *H. vastatrix* remain unknown (Slaminko *et al.,* 2008; Talhinhas *et al.,* 2017).

In a typical heteroecious macrocyclic life cycle, meiosis takes place in short-lived basidia germinating from teliospores. Haploid basidiospores infect the aecial host within which they differentiate various fungal structures, like protoaecia and pycnia. Haploid pycniospores and receptive hyphae are produced within pycnial nectar droplets. Fertilization of receptive hyphae with pycniospores of opposite mating types occurs inside droplets. After plasmogamy, dikaryotic aecia differentiate inside the host and aeciospores are released and dispersed by the wind. Aeciospores infect the telial host in which uredinia and urediniospores are produced. From there a vegetative cycle occurs repeatedly on the telial host for an extended period of time that can go on for several weeks or months during summer. In early autumn, uredinia differentiate into telia that go through an overwintering period during which karyogamy will

occur, giving rise to diploid dormant teliospores (examples of typical heteroecious cycles can be found in Leonard & Szabo, 2005; Hacquard *et al.,* 2011; Aime *et al.,* 2014).

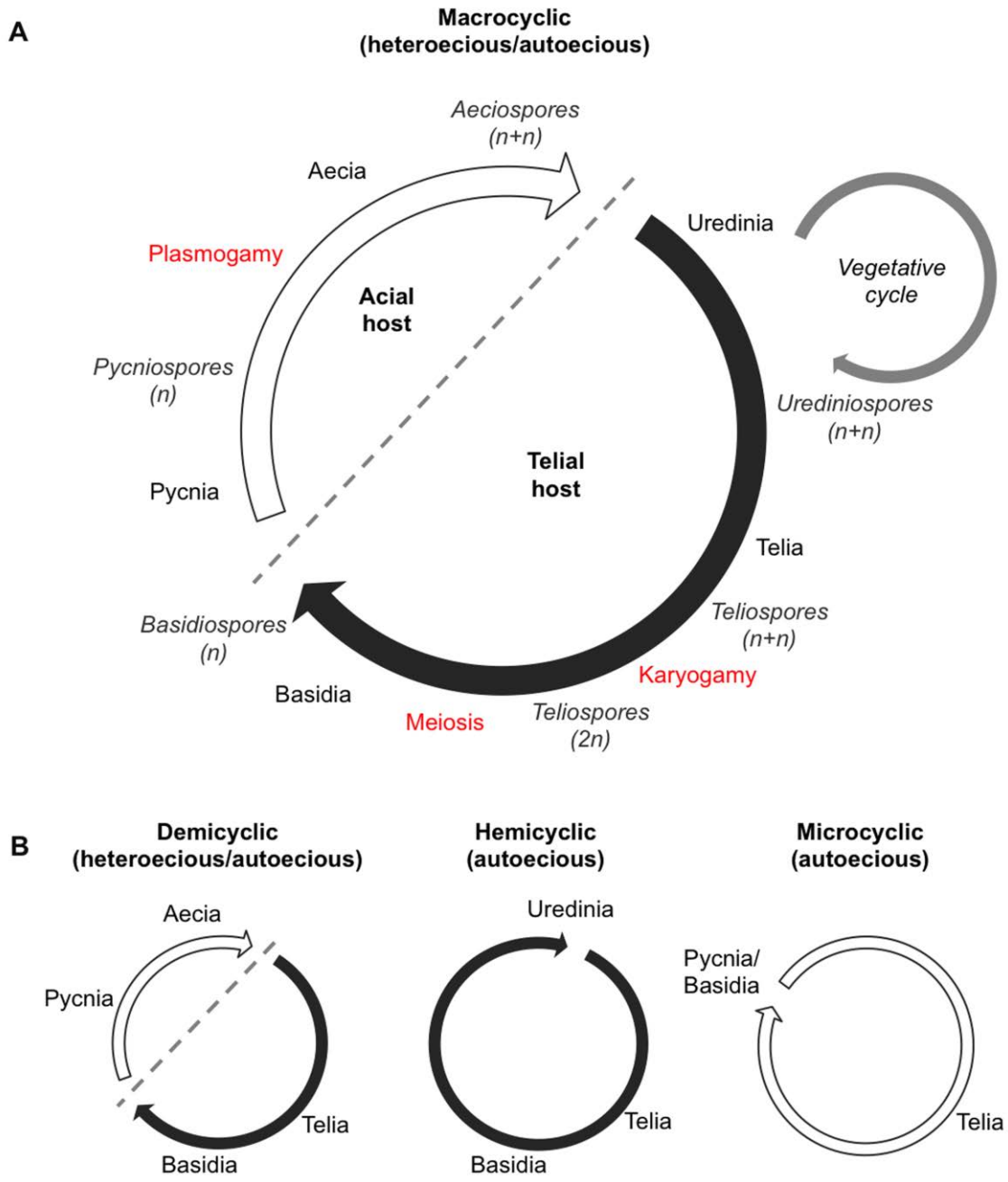


Figure 1: life cycles of rust fungi.
(A) Macrocyclic life cycle. Macrocyclic rust fungi can be autoecious or heteroecious -i.e. life cycle realised on one or two host plants, respectively- and produce five spore stages. Haploid basidiospores (n) infect the aecial host, pycnia contain pycniospores (n) of different mating types. After fertilization and plasmogamy, dikaryotic aecia release aeciospores (n+n) that infect the telial host. Dikaryotic urediniospores (n+n) are then produced repeatedly during vegetative cycle through the summer season. In autumn, teliospores (n+n) are formed in telia. Karyogamy occurs in telia and meiosis initiates. After winter, basidia (2n) differentiate and produce basidiospores (n). (B) Derived rust life cycles. Demicyclic rust fungi can be heteroecious or autoecious and do not form uredinia. Hemicyclic rust fungi are autoecious with absence of pycnial and aecial stages. Microcyclic life cycle is reduced to telia and basidia or pycnia according to species.

The diversified life cycles of rust fungi have raised questions about their evolution and particularly co-evolution with their hosts. As obligate parasites, rust fungi were thought to have extensively coevolved with their hosts, suggesting that ancestral rust fungi may have infected ferns, then gymnosperms and then angiosperms, following the plant diversification (Cunningham, 1931; Leppik, 1953, 1965; Savile, 1976). However phylogenetic approaches have shown that fern rust fungi were not ancestrals and that some rust fungi infecting angiosperms belonged to an ancestral family (Sjamsuridza *et al.*, 1999; Aime *et al.*, 2006). So far, the basal rust fungus *Caeoma torreyae* is found on gymnosperm suggesting that rust fungi are more recent than first thought (Aime *et al.*, 2006). Recent data based on phylogenetic comparisons between rust fungi and their host plants support a diversification through host jumps during coevolution with their hosts (McTaggart *et al.*, 2016; Aime *et al.*, 2017). Heteroecious rust fungi alternate between unrelated host plants such as gymnosperms and angiosperms or monocots and dicots, however these hosts must share common ecosystems. This indicates that alternate hosts of rust fungi with unresolved life cycles may be found in the same ecosystem than the current known host, or that this host has become extinct or replaced from this ecosystem during its evolution.

IV- The biology of rust infection

The continuous progress and evolution of microscopy techniques during the last century has led to precise descriptions of both the host infection process and the various stages of rust life cycles in a couple of model rust fungi (Littlefield & Heath 1979; Bushnell & Roelfs 1984). The genetic dissection of the flax-flax rust pathosystem by Harold H. Flor has enabled the formulation of the gene-for-gene concept and has pioneered identification of resistance genes in plants and for a better understanding of the plant immune system (Flor 1971; Dodds & Rathgen 2009). It is however only in the past two decades or so that the first resistance and avirulence genes were cloned in rust pathosystems (Ellis et al., 2007). By the end of the twentieth century, pioneer molecular studies of rust biotrophy have opened a new era for understanding mechanisms underlying the infection process. The following section briefly relates these early molecular researches.

1. Early stages of infection: host penetration and haustorium formation

In heteroecious rust fungi, basidiospores infect the aecial host and aeciospores and urediniospores infect the telial host by different penetration mechanisms. Monokaryotic basidiospores germinate and directly penetrate epidermal cells after differentiating an appressorium (Hahn, 2000). Little is known about the infection process in the aecial host. The dikaryotic aeciospores and urediniospores both enter the host leaf through stomata openings after the differentiation of an appressorium (Voegele, 2009). The role of the appressorium in

telial host infection is not clearly determined. It does not seem to play an active role for penetration through the stomata, although a slight turgor pressure (0.35MPa) maybe observed (Terhune et al., 1993). Different transcriptomic analyses of appressoria issued from urediniospores highlighted genes related to metabolism and cell cycle, indicating that this structure play an active role in the penetration process (Hu et al., 2007; Stone et al., 2012; Talhinhas et al., 2014). Urediniospores from the soybean rust fungus directly penetrate their host through the epidermis like basidiospores of heteroecious fungi forming an appressorium with a high turgor pressure of 5.13 MPa (Loehrer et al., 2014). This observation is similar to other fungal plant pathogens like the rice blast fungus *Magnaporthe oryzae* (Ryder & Talbot, 2015). However, *P. pachyrhizi* basidiospores are hyaline and devoid of melanine, indicating a different penetration modality. Almost all physiological studies of host infection by rust fungi focused on urediniospores (Voegele, 2009).

The haustorium represents a distinctive feature of obligate biotrophs. After going through the stomata, the penetration hypha forms a substomatal vesicle from which an infection hypha differentiates (Voegele, 2006). A haustorium mother cell starts forming a penetration peg that secretes cell wall degrading enzymes to penetrate the host cell wall in order to differentiate a haustorium within the host cell cavity (Hahn, 2000; Garnica et al., 2014). The haustorium is an invagination within the host cell plasmalemma, which remains intact. Different to other obligate biotrophs, rust fungal haustoria are marked by the presence of a dense neckband that separates the extrahaustorial matrix surrounding the haustorial structure from the apoplasm (Garnica et al., 2014). The formation of this specialized infection structures induces structural changes inside the infected host cell (Voegele & Mendgen 2003). It is also an intense site of expression of secreted proteins.

2. Nutrients uptake from the host

Rust fungi completely depend on the host plant to survive as it represents their only nutrients source (Staples, 2000). Known nutrient uptake components of rust haustoria remain limited to a few sugar and amino acid transporters (Voegele & Mendgen 2011; Struck et al., 2015). Rust fungi are deficient for some essential pathways (e.g. nitrate and sulphur assimilation), illustrating their dependence on nutrient uptake from their hosts (Duplessis et al., 2011; Kemen et al., 2015). The first discovered amino acid permeases were AAT1p, AAT2p and AAT3p from *U. fabae* that showed a specific or a preferential expression in haustoria (Figure 2, Hahn & Mendgen 1997; Hahn et al., 1997; Struck 2015). Homologs of AAT genes up-regulated in haustoria were later identified in other rust fungi (Hacquard et al., 2010; Garnica et al. 2013).

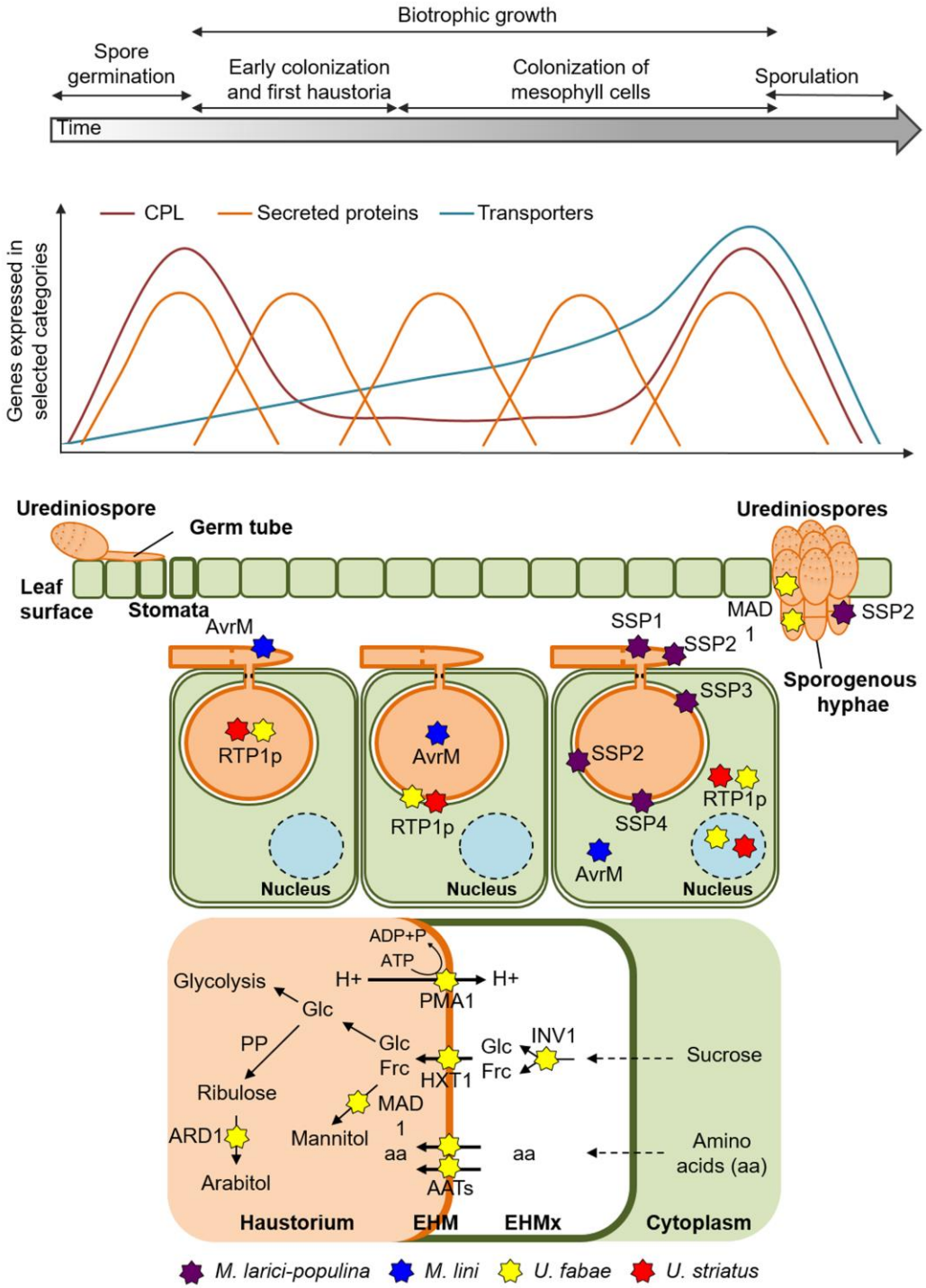


Figure 2. Schematic representation of telial host infection by rust urediniospores

The upper panel describes the different steps of rust infection on the telial host from spore germination to sporulation. The medium panel represents expression profiles of selected gene categories during the infection process (according to time-course transcriptomics in Duplessis et al. 2011b and Dobon et al. 2016). These profiles have been simplified according to the global average expression. Orange curves represent coordinated waves of expression for secreted protein genes; red curve represents the main expression profile for CAZymes, proteases and lipases (CPL) encoding genes; blue curve represents expression for transporter encoding genes. The lower panel shows a schematic representation of the interaction between rust infection hyphae and haustoria and plant host cells. Urediniospore germinate on leaf surface and penetrate plant tissue through stomata. Infection hyphae grow between host cells and an haustorium is formed inside the host cell cavity. After several days of biotrophic growth, new

Figure 2. Interaction between rust fungi and host cell
urediniospores are produced and released on the leaf surface. Immunolocalization of candidate effectors is presented. *Melampsora lini* AvrM is localized at the periphery of fungal hyphae, in the haustorium, in the extra-haustorial matrix and in flax cell cytosol. Rust-transferred proteins (RTP1p) from *Uromyces fabae* and *Uromyces striatus* are localized in the haustorium, in the extra-haustorial matrix and in the nucleus and the cytosol at later stages of colonization. SSP1, SSP2, SSP3 and SSP4 correspond to *Melampsora larici-populina* small-secreted proteins (SSPs) Mlp123523, Mlp123227, Mlp123932 and Mlp37347 respectively. *M. larici-populina* SSPs are localized at the periphery of the haustorium (SSP4, SSP3) and of infection hyphae colonizing plant mesophyll (SSP1), and both at the periphery of haustoria and sporogenous hyphae (SSP2). Close-up of haustoria-host cell interface shows enzymes and transporters of rust fungi characterized in *U. fabae*. PMA1 is a H+ ATPase that is active as the first haustoria formed and it localizes at the haustorial membrane. PMA1 sets up electrochemical gradient essential for nutrient uptake. ARD1 is a NADP+ dependent D-arabitol dehydrogenase localized in the lumen of haustoria. HXT1 represents the proton-motive force driven hexose (D-glucose and D-fructose) transporter characterized in the haustorial membrane of the bean rust fungus. MAD1 is a mannitol dehydrogenase localized in haustoria, in sporogenous hyphae and in spores. AATs represent the three amino acid transporters (AAT1, AAT2, AAT3) of *U. fabae*. They all localize at the hautorial membrane.

The oligopeptide transporter (OPT) gene family is expanded in rust fungi compared with other basidiomycetes and OPT genes were shown to be highly expressed during host infection and in haustoria of rust fungi (Duplessis et al., 2011a; Garnica et al. 2013). Recently, OPT genes were associated to virulence in the smut fungus *U. maydis* suggesting that extracellular peptides may be an important source of nutrient for the fungus (Lanver et al., 2018). So far, no functional characterization of OPT has been conducted in rust fungi. Host oligopeptides could represent as well an important source of nitrogen and sulphur for rust fungi during biotrophic growth.

Studies conducted on *U. fabae* isolated haustoria identified *PMA1p* encoding a H+-ATPase that was functionally characterized in yeast. PMA1p activity establishes the proton electrochemical gradient that provides energy for transport processes and nutrient uptake (Figure 2, Struck et al., 1996; Struck et al., 1998). The important role of proton gradient for the absorption of sugar was later shown for the *U. fabae* hexose transporter HXT1p localized at the haustorium membrane and its role as a proton-motive D-glucose and D-fructose transporter was demonstrated in yeast (Figure 2, Voegele et al., 2001). Plants present low levels of monomeric hexoses in leaves and sucrose is the main sugar in this organ (Lohaus et al., 2001; Weber & Roitsch 2000). The *U. fabae* invertase INV1p expressed during infection and localized in the extrahaustorial matrix was reported to be necessary for the breakdown of sucrose in D-glucose and D-fructose, the substrates for HXT1p (Figure 2; Voegele et al., 2006). Recently, the invertase PsINV was investigated in the wheat rust *P. striiformis* sp. *tritici* showing a high efficiency for sucrose hydrolysis as well as an increased expression during wheat infection (Chang et al., 2017). Other genes related to metabolism such as the

glucokinase GLK1p, the mannitol dehydrogenase 1 (MAD1p) and the NADH+ dependent D-arabitol dehydrogenase ARD1p from the bean rust *U. fabae* have provided valuable informations about carbon storage and utilisation in this rust fungus (Figure 2; Link et al., 2005; Voegele et al.,2005; see Voegele & Mendgen, 2011 for details). Homologs of these different genes were also found in other rust fungi in which they showed high level of expression during infection of their respective hosts, supporting conserved sugar uptake mechanisms in rust fungi (Hacquard et al., 2010; Duplessis et al., 2011; Garnica et al., 2013; Zheng et al., 2013). No sucrose transporter has been identified in rust fungi (Duplessis et al., 2011a), indicating that hexose transport maybe central in haustoria. No recent functional analysis of uptake mechanisms has been conducted in rust fungi except for *P. striiformis* f. sp. *tritici*.

3. Infection control centre: the haustorium as a multifaceted structure

The haustorium is not only a specialized machinery for nutrient diversion, it is also an intense site of expression of secreted proteins, including effectors (Duplessis et al., 2012; Garnica et al., 2014). In rust fungi, secreted effector proteins were shown to reside inside host cell cytoplasm, however how effectors translocate from the haustorium to the host cell remains unknown (Petre & Kamoun 2014). Until recently, only a handful of effectors were described in rust fungi (Petre et al., 2014). The rust transferred proteins RTP1 from *Uromyces* spp. were the first fungal proteins reported translocated into their host cells (Kemen et al., 2005). The proteins were localized inside host cells, then in the host nucleus, suggesting a specific traffic towards its final localization (Figure 2). The *M. lini* effector AvrM has also been detected inside flax cells by immunogold transmission electron microscopy (Figure 2; Rafiqi et al., 2010). In both examples, trafficking from the fungal haustoria to the host cell was temporally regulated. RTP1 is part of a conserved multigene family found across Pucciniales, including basal species like *H. vastatrix* (Fernandez et al., 2012; Pretsch et al., 2013; Link et al., 2014). Detailed analysis of RTP1 indicates a dual function as a protease inhibitor in vitro and as a filament-forming protein in the extrahaustorial matrix that could act as a stabilizing factor for the fungal structure (Pretsch et al., 2013; Kemen et al., 2013). Immunolocalization of two RTP homologs in the poplar rust *M. larici-populina* has revealed a dual localization at the periphery of haustoria and infection hyphae (Figure 2; Hacquard et al., 2012). Two other poplar rust secreted proteins were detected *in planta* by immunolocalization: a homolog of the flax rust avirulence effector AvrL567 was found at the periphery of poplar rust haustoria but not beyond in the host cells; and SSP15 which is localized both at the periphery of haustoria and of paraphises producing new spores in uredinia (Figure 2; Hacquard et al., 2012). All observations gathered for rust proteins indicate complex and diverse patterns of localization within infected host tissues and their precise functions during infection remains to be clarified (Petre et al., 2014).

V- Rusts in the genomics era: the ever-expanding list of candidate effector genes

Up to 20 genomes totalling 11 different rust species have been reported within less than 10 years (Table 1; see Aime *et al.*, 2017 for details). Beyond the minimal informations delivered by these studies regarding the predicted genes and the genome architecture and organisation; genome-wide comparative analyses, transcriptome and population studies or prediction of secreted proteins to identify rust effector proteins have considerably extended our understanding of rust fungal biology (Table 1).

Table 1: Effector-mining in rust fungi

Pucciniales species	Genomic data	Transcriptomic data	Secretome	Candidate Secreted Effector Proteins (CSEPs)
Austropuccinia psidii (myrtle rust)	Incomplete draft genome (Tan et al. 2014)	-	-	-
Cronartium quercuum f. sp. *fusiforme* (fusifrome pine rust)	Not published*	-	666 SSPs (Pendleton et al. 2014)	132 CSEPs families (Pendleton et al. 2014)
Cronartium ribicola (white pine blister rust)	-	RNAseq (ribes infected leaves and white pine needles, Liu et al. 2015)	734 SPs (Liu et al. 2015)	540 CSEPs detected at early needle infection stage (Liu et al. 2015)
Hemileia vastatrix (coffe rust)	Incomplete Draft genome (Cristancho et al. 2014)	RNAseq (infected leaves, Fernandez et al. 2012; urediniospores and appressoria, Talhinas et al. 2014; different infection stages, Maia et al. 2016)	483 SPs (Cristancho et al. 2014)	382 CSEPs (Fernandez et al. 2012); 516 CSEPs highly expressed *in planta* (Talhinas et al. 2014); 60 CSEPs (Maia et al. 2016)

Melampsora larici-populina (poplar rust)	Reference genome (Duplessis et al. 2011a)	oligonucleotide arrays (microdissection of infected poplar leaves, Hacquard et al. 2010; infected leaves, urediniospores, Duplessis et al. 2011a; 2011b; telia, Hacquard et al 2013); RNAseq (infected leaves, Petre et al. 2012; basidia and infected larch needles; Lorrain et al. 2018a)	1,184 SSPs (Hacquard et al. 2012); 1549 SPs (Saunders et al. 2012); 2,168 SPs (Lorrain et al. 2015)	19 CSEPs highly up-regulated transcripts in palisade mesophyll (Hacquard et al. 2010); 509 CSEPs genes expressed *in planta* (Duplessis et al. 2011a); 270 CSEP specifically expressed *in planta* (Duplessis et al. 2011b); 19 CSEPs early expressed *in planta* (Petre et al. 2012); 245 CSEPs preferentially expressed in larch and basidia, 103 CSEPs preferentially expressed in poplar and urediniospores, 98 CSEPs only expressed in planta (larch & poplar) and 688 CSEPs commonly expressed in larch, poplar, basidia and urediniospores

Melampsora lini (flax rust)	Reference genome (Nemri et al. 2014)	cDNA library (isolated haustoria, Cantanzariti et al. 2006); RNAseq (infected leaves, urediniospores, Nemri et al. 2014)	1,085 SPs (Nemri et al. 2014)	21 HESPs expressed in isolated haustoria (Cantanzariti et al. 2006); 725 CSEPs (Nemri et al. 2014)
Phakopsora pachyrhizi (soybean rust)	Draft genome (Loehrer et al. 2014)	RNAseq (isolated haustoria, Link et al. 2014b; infected leaves, Tremblay et al. 2012; 2013)	851 SPs (de Carvalho et al. 2016)	156 CSEPs (Link et al. 2014); 176 CSEP (Tremblay et al. 2012); 35 CSEPs expressed in haustoria (Kunjeti et al. 2016)
Puccinia coronata f. sp. *avenae* (oat crown rust)	Draft genome (Miller et al. 2017)	RNAseq (urediniospores, infected leaves, isolated haustoria, Miller et al. 2017)	1,532 and 1,548 SPs of 2 isolates (Miller at al. 2017)	529 and 549 CSEPs of 2 isolates (Miller et al.2017)
Puccinia graminis f. sp. *tritici*	Multiple reference genomes (Duplessis et al. 2011;Upadhyaya et al. 2015)	DNA arrays (infected leaves, urediniospores, Duplessis et al. 2011a) RNAseq (isolated haustoria and urediniospores, Upadhyaya et al. 2015)	1,106 SSPs (Duplessis et al. 2011a); 1,852 SPs (Saunders et al. 2012); 1,459 SPs (Zheng et al. 2013); 1,924 SPs (Upadhyaya et al. 2015)	442 CSEPs expressed *in planta* (Duplessis et al. 2011a); 336 CESPs (Upadhyaya et al. 2014); 520 CSEPs expressed in haustoria (Upadhyaya et al. 2015)
Puccinia sorghi (maize rust)	Reference genome (Rochi et al. 2017)	-	1,599 SPs (Rochi et al. 2017)	-

Puccinia striiformis f. sp. *tritici (wheat yellow rust)*	Multiple reference genomes (Cantu et al. 2011; Cantu et al. 2013; Zheng et al. 2013; Cuomo et al. 2017; Kiran et al. 2017; Schwessinger et al. 2017)	DNA arrays (Huang et al. 2008) ESTs (isolated haustoria, Yin et al. 2009); RNAseq (infected leaves, urediniospores, isolated haustoria, Cantu et al. 2013; Garnica et al. 2013)	15 SPs (Yin et al. 2009); 2,999 SPs (Cantu et al. 2013); 1,299 SPs (Garnica et al. 2013); 687, 727, and 720 SPs of 3 isolates (Kiran et al. 2017)	933 CSEPs (Cantu et al. 2013); 437 CSEPs (Garnica et al. 2013); 518, 554, and 562 CSEPs of 3 isolates (Kiran et al. 2017)
Puccinia triticina (wheat rust)	Multiple reference genomes (Kiran et al. 2016; Cuomo et al. 2017)	RNAseq (infected leaves, Bruce et al. 2014; urediniospores, infected leaves, isolated haustoria, pycnia and aecia Cuomo et al. 2017)	753 SPs (Xu et al. 2011);1,450 SPs (Bruce et al. 2014)	543 CSEPS (Bruce et al. 2014);1,358 predicted CESPs with 138 highly expressed during wheat infection (Cuomo et al. 2017)
Uromyces appendiculatus (common bean rust)	Link et al 2014a	RNAseq (isolated haustoria, Link et al. 2014a)	413 SPs (Link et al. 2014a)	24 CSEPs expressed in infected beans (Cooper et al. 2016)
Uromyces fabae (broad bean rust)	Link et al 2014b	RNAseq (urediniospores, Link et al. 2014b)	599 SPs (Link et al. 2014b)	-

* Genomes available at the JGI website: https://genome.jgi.doe.gov/programs/fungi/index.jsf
SP secreted protein; SSP small-secreted protein; CSEP candidate secreted effector proteins

1. Sequenced rust fungal genomes

Rust fungi have large genomes in a range from 60Mb to more than 300Mb (Aime *et al.*, 2017) but flow cytometry estimates the largest rust genome at over 2Gb (Ramos *et al.*, 2015; Tavares *et al.*, 2014). Rust genomes are sequenced from dikaryotic spores and they present a remarkable level of heterozygosity (Persoons *et al.*, 2014; Upadhyaya *et al.*, 2014; Cuomo *et al.*, 2017). Repeat elements range from 18% to 75% in rust genomes, which represents a serious challenge to rust genome assembly. This is particularly true for rust genomes sequenced only with short read technologies (see Aime *et al.*, 2017 for details). Rust fungi possess large numbers of genes (15,000 to 20,000 per genome) and about half of them lack functional annotation by comparison to international databases. Many genes are also specific to Pucciniales and may represent innovations in rust biology (Aime *et al.*, 2017). The lack of complete and high-quality rust genome assemblies and standardized practices in the community have somehow restricted possibilities for comparative studies. However, new sequencing technologies producing long-reads and improved assembly tools can support better assembly of large rust genomes. The genomes of *P. striiformis* f. sp. *tritici* 104E and two isolates of the oat crown rust *Puccinia coronata* f. sp. *avenae* were recently sequenced with long reads sequencing approaches providing de novo assemblies of greater quality (Schwessinger *et al.*, 2017; Miller *et al.*, 2017). It was possible to phase the haplotypes for a significant portion of these dikaryotic genomes (50 to 92%). The genome size of these two rust species were in the small range of estimated rust genome sizes (83 to 105 Mb), and it remains to be determined how assembly will perform with genomes in the range of 1Gb or more with >70% repetitive content. Determining the complete and phased genome assembly is important to highlight intrinsic differences between the two rust nuclei and a better knowledge of rust genomes architecture may help to understand the impact of transposition on rust virulence and evolution.

2. Prediction of candidate effectors in rust secretomes

Early after the conceptualization of plant pathogen effectors and their importance in the infection process, it was established that secreted proteins of unknown function can represent candidate effectors (Win *et al.,* 2012). Therefore, secretome prediction has been systematically applied to rust genomes with the gold rush promise of identifying effectors. Some features like a small size, a high content in cysteine residues or taxonomical specificity have been used to identifiy candidate effectors (Lorrain *et al.*, 2015; Sperschneider *et al.*, 2017a). Based on such criteria, long lists of rust candidate effectors have been identified (Table 1; Pendleton *et al.*, 2014; Duplessis *et al.*, 2011; Nemri *et al.*, 2014; Saunders *et al.*, 2012; Cantu *et al.*, 2013). Their numbers can vary depending on the precise criteria used to flag them

as candidates, but secreted protein encoding genes are typically in the range of one to two thousands per genome (ca. 10% of the coding space; Table 1). Both orphan and gene families are found among Pucciniales-specific candidate effectors. Expanded gene families of unknown function are a common feature of rust fungal genomes, among which are several secreted protein families (Aime *et al.*, 2017). Recently, machine-learning prediction tools have been established to help the identification of candidate effectors within fungal secretomes, the prediction of their localization *in planta* or to discriminate between apoplastic and cytoplasmic effectors (see details in Table 2; Sperschneider *et al.*, 2015; Sperschneider *et al.*, 2017b; Sperschneider *et al.*, 2017c). These tools are valuable because they are not based on *a priori* criteria and they can be used to revisit sets of candidate effector predicted in rust genomes (Table 2). As for other machine-learning methods, the publication of more *bona fide* effectors should improve future prediction, provided that rust effectors do have conserved features to truly support their prediction.

Table 2: Prediction and analysis tools dedicated to effector-mining in fungal genomes

Program/ Pipeline	Description	Website	Reference
EffectorP	Machine learning tool to predict effector candidates in fungal secretomes	http://effectorp.csiro.au/	Sperschneider et al. 2016
LOCALIZER	Machine learning prediction tool for subcellular localization dedicated to fungal effector candidates and plant proteins	http://localizer.csiro.au/	Sperschneider et al. 2017a
ApoplastP	Machine learning prediction tool to separate apoplastic fungal effectors over cytosplasmic effectors	http://apoplastp.csiro.au/	Sperschneider et al. 2017b
MCL clustering	Hierachical clustering of secreted proteins to priorotize candidate effectors	-	Saunders et al. 2012; Cantu et al. 2013; Nemri et al. 2014

3. Transcriptomics of rust infection: pointing the needle in the haystack

A specific expression *in planta* is a critical feature to define pathogenic effectors. Thus, transcriptomics has been an approach of choice to reduce large sets of rust candidate effectors to a smaller number of priority targets for functional study (Table 1; Duplessis *et al.*, 2012; Duplessis *et al.*, 2014). Pioneer transcriptomic studies used cDNA libraries to identify haustorially expressed secreted proteins from isolated haustoria of *U. fabae* and *M. lini* (Hahn & Mendgen 1997; Catanzariti *et al.*, 2006). RNA-sequencing (RNAseq) has been instrumental to obtain insights about rust fungi without reference genomes, e.g. *H. vastatrix*, *Gymnosporangium* spp., *P. pachyrhizi*, *Uromyces appendiculatus*, (Table 1; Loeher *et al.*, 2014; Link *et al.*, 2014a; Cristancho *et al.*, 2014; Tao *et al.*, 2017). Transcriptomics of host-infection and of isolated haustoria have helped prioritization of candidate effectors in different rust species (Duplessis *et al.*, 2014; Aime *et al.*, 2017). Many genes expressed at late stages of colonization encode carbohydrate active enzymes, proteases, lipases and transporters (Figure 2; Duplessis *et al.*, 2011b; Hacquard *et al.*, 2011; Hacquard *et al.*, 2012). These findings are consistent with nutrient uptake by sugar and amino acids transporters after haustoria were established in the plant tissue (Voegele *et al.*, 2011; Struck *et al.*, 2015). Different sets of candidate effectors are expressed sequentially in coordinated waves of expression from early to late infection stages (i.e. from spore germination to sporulation; Figure 2; Dobon *et al.*, 2016; Duplessis *et al.*, 2011b; Hacquard *et al.*, 2011; Hacquard *et al.*, 2012; Bruce *et al.*, 2013; Cantu *et al.*, 2013; Fernandez *et al.*, 2012; Tremblay *et al.*, 2013; Rutter *et al.*, 2017). Transcriptome of *M. larici-populina* microdissected uredinial units have shown that dozens of small secreted protein genes are among the most highly expressed genes in fungal infection structures whereas spores are already released on the leaf surface (Hacquard *et al.*, 2010). Such coordinated expression patterns indicate that rust fungi may release and deliver early and late effectors with different roles in the establishment and in the maintenance of biotrophy. Despite the valuable knowledge gained by studying the infection process in the main host of rust fungi, gene expression at other stages of the rust life cycle is overlooked. Particularly, there is scarce information about infection of alternate hosts and sexual stages. Only a few studies considered gene expression on the alternate host so far (Xu *et al.*, 2011; Liu *et al.*, 2015; Cuomo *et al.*, 2017; Lorrain *et al.*, 2018a). In *M. larici-populina*, the comparison of expression profiles in the telial host (poplar) and in the aecial host (larch) revealed concomitant expression of sets of candidate effectors in both hosts as well as in each host specifically (Lorrain *et al.*, 2018a). Such an approach is helpful to narrow down the list of effectors related to the main host where disease makes most of the damage.

4. Population genomics to unravel effectors under selection and avirulence genes

Population genomics and the study of evolutionary processes can help the identification of avirulence genes among candidate effectors (Moller & Stukenbrock 2017). The pressure exerted by components of the host plant immune system can be detected by traces of positive selection in effector genes under surveillance. Signatures of rapid evolution and polymorphism have been extensively searched in filamentous pathogens, including rust fungi (Raffaele & Kamoun 2012; Stukenbrock 2013; Persoons *et al.*, 2014; Upadhyaya *et al.,* 2014; Figure 2). Three *M. lini* avirulence genes AvrL567, AvrP4 and Avr123 were found under positive selection at the intraspecific and/or interspecific level (Dodds *et al.*, 2004; Barret *et al.*, 2009). Genome-wide search for single nucleotide polymorphisms or insertion/deletions rust fungal isolates has enabled the detection of avirulence candidates among secreted protein genes (Cantu *et al.*, 2013; Persoons *et al.*, 2014; Upadhyaya *et al.,* 2014; Bruce *et al.*, 2014; Wu *et al.*, 2017; Cuomo *et al.*, 2017; Bueno-Sancho *et al.*, 2017). The sequencing of many isolates is now possible at low costs and high-density genetic maps can be constructed through the resequencing of progenies or by restriction site detection sequencing (RAD-seq) in rust fungi (Yuan *et al.,* 2017; Pernaci *et al.,* 2014; Anderson *et al.*, 2016). Two recent studies have used genomics and resequencing approaches to identify AvrSr35 and Avr50, two *P. graminis* f. sp. *tritici* avirulence effector genes, demonstrating the impact of genomics on the field (see details in section VI; Chen *et al.,* 2017; Salcedo *et al.,* 2017). In the coming years, more key pathogenesis determinants will undoubtedly be unravelled by population genomics and genome-wide association studies.

VI- Functional characterization of rust effectors

Most rust fungi that are major danger for agriculture now have lists of candidate effectors: some may play a critical role in setting a successful infection, some can be targets of host resistance proteins (avirulence protein) and by identifying their targets inside the host cell, effectors may reveal host susceptibility. Investigating candidate effectors can foster selection of more durable resistances to rust diseases. Still, the number to scrutinize is very large and as obligate biotrophs, rust fungi remain difficult to manipulate. As a consequence high-to medium-throughput pipelines have been developed to achieve functional characterization of candidate rust effectors and to help formulate hypotheses about their functions in plant cells. Effectoromics screens aim to unravel cellular and molecular targets of given candidates and if/how they play a role in pathogenesis (Alfano *et al.,* 2009). Heterologous plant systems are methods of choice to investigate non-model candidate effector functions (Petre *et al.,* 2014; Lorrain *et al.*, 2018b). The figure 3 presents the panel of approaches used to identify and unravel function of candidate effectors in the plant cell. In the following section, we describe the different strategies and approaches applied in the past five years to screen effectors in rust fungi. We report these studies by rust species according to the importance of the community-

driven efforts and distinguish screening approaches and work on specific genes. Table 3 summarizes the number of candidates assayed in each rust species.

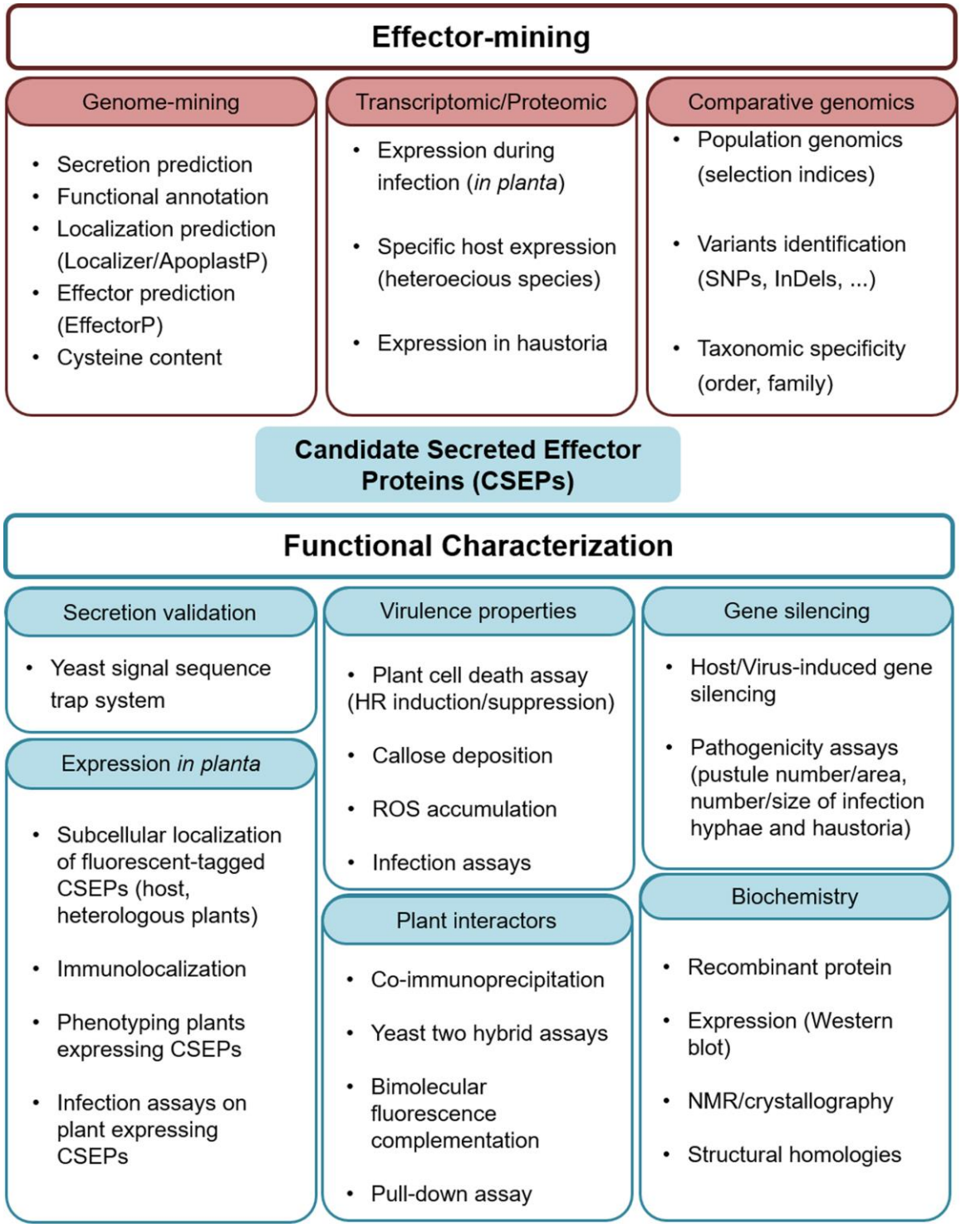


Figure 3: Common approaches applied in effector biology of rust fungi.
The panel *Effector-mining* shows approaches applied to rust genomes and transcriptomes to identify candidate secreted effector proteins (CSEPs). The panel *Functional characterization* details the different approaches and methods deployed in the community to scrutinize role and functions of rust fungal CSEPs.

1. Effectoromics of the yellow rust fungus *P. striiformis*

P. striiformis f. sp. *tritici* is the rust species that attracted most attention from the community. Both medium-screens and dedicated efforts have been employed for functional characterization of candidate effectors and pathogenesis related functions (e.g. signalling, a key cellular category in plant-microbe interactions; Hamel *et al.,* 2012). Multiple efforts from different research groups have used different pipelines and approaches. The host-induced gene silencing (HIGS) approach established to silence *Puccinia* spp. genes using the *Barley stripe mosaic virus* (BSMV) on wheat (Yin *et al.,* 2011) has been instrumental to assess rust genes in wheat-rust interactions (Tang *et al.* 2017).

Two medium-throughput screens have focused on *P. striiformis* f. sp. *tritici* candidate effectors. Based on expression profiling in haustoria (Cantu *et al.,* 2011; Yin *et al.,* 2015), Ramachandran et al. (2016) scrutinized 20 Pucciniacae candidates -including nine from *P. striiformis* f. sp. *tritici*- for their ability to suppress cell death. Seven were able to suppress a hypersensitive response (HR) triggered in *N. benthamiana*. The plant cell death suppression property was confirmed in wheat for one candidate through delivery by *Pseudomonas fluorescens* (Ramachandran *et al.*, 2016). Sixteen mature candidate effectors were transiently expressed in *N. benthamiana* in fusion with GFP to determine their subcellular localization and to identify their putative interactors by co-immunoprecititation and mass spectrometry (co-IP/MS) (Petre *et al.,* 2016a). Nine were informative in term of localization or for interacting with plant proteins. One candidate showed accumulation in processing bodies and interaction with enhancer of mRNA decapping protein. The interaction between the wheat enhancer of mRNA decapping protein 4 (TaEDC4) and the yellow rust effector was confirmed by coIP in *N. benthamiana* (Petre *et al.,* 2016a). Altogether these two screens led to the identification of 16 candidate of interest with virulence activities, a specific subcellular localization and/or a putative plant protein interactor (Table 3). A small-throughput screen for immunity suppression was set for six candidate effectors, identifying PEC6 a candidate with avirulence properties (Liu *et al.,* 2016). PEC6 suppressed plant immune responses in *N. benthamiana*, enhanced bacterial growth in *Arabidopsis* and showed interaction with an adenosine kinase by yeast-two-hybrid and BiFC (Liu *et al.,* 2016). Fungal infection was also reduced on wheat when PEC6 was silenced by HIGS (Liu *et al.,* 2016).

Several studies investigated yellow rust candidate effectors or signalling proteins in separate efforts. PstSCR1, a small cysteine-rich candidate effector, induces plant cell death when expressed with its signal peptide or injected as a purified protein *in N. benthamiana*, but there is no information about its potential role in *P. striiformis* f. sp. *tritici* (Dagvadorj *et al.,* 2017). Remarkably, Pr. Kang‘s group (NWAFU, Yangling, China) has established a pipeline of

diversified molecular approaches to unravel the role of *P. striiformis* f. sp. *tritici* proteins during interaction with wheat and in total, more than ten yellow rust genes have been studied in the past five years (Tang *et al.,* 2015; Tang *et al.,* 2017). Selected candidate genes are most often assayed by HIGS to validate their involvement in pathogenicity through quantitative estimation of infection-related traits. Depending on given candidates, different complementary approaches were applied such as subcellular localization in heterologous systems or in wheat protoplasts (Cheng *et al.,* 2015; Cheng *et al.,* 2016a; Wang *et al.,* 2017; Zhu *et al.,* 2017), complementation assays on yeast or fungal pathogen (Guo *et al.,* 2011; Liu *et al.,* 2014; Liu, Guan *et al.,* 2016; Zhu *et al.,* 2017a), yeast invertase secretion assay (Cheng *et al.,* 2016a; Liu, Guan *et al.,* 2016; Wang *et al.,* 2017), plant cell death BAX-triggered suppression assays (Cheng *et al.,* 2016a; 2016b; Liu, Guan *et al.,* 2016) and overexpression in fission yeast coupled with the monitoring of stress responses (Cheng *et al.,* 2015; Tang *et al.,* 2015; Cheng *et al.,* 2016b; Jiao *et al.,* 2017; Zhu *et al.,* 2017a; 2017b). Among the proteins characterized trough the aforementioned pipeline, six are protein kinases or regulators of kinases in fungal transduction pathways (Guo *et al.,* 2011; Jiao *et al.,* 2017; Cheng *et al.,* 2015; Cheng *et al.,* 2016b; Qi, Zhu *et al.,* 2017) and the others correspond to a mitochondrial adenine nucleotide tranclocase antiporter (PsANT; Tang *et al.,* 2015), a superoxide dismutase (Liu, Guan *et al.,* 2016), an isocitrate lyase required for spore germination and a transcription factor. Two studied proteins are candidate effectors (Cheng *et al.,* 2016a; Wang *et al.,* 2017). PSTha5a23 is a specific *P. striiformis* f. sp. *tritici* candidate highly expressed during interaction with wheat. It was localized in the cytoplasm of wheat protoplasts, supressed cell death in *N. benthamiana* and increased the number of pustules when overexpressed in wheat (Cheng *et al.,* 2016a). The candidate effector PNPi was found to interact in wheat with the homolog of the key defense regulator NPR1 where it competes with the transcription factor TGA2.2 that is known to activate *Pathogenesis-related* genes expression (Wang *et al.,* 2017). The functional pipeline applied in Pr. Kang's group to characterize molecular determinants of *P. striiformis* f. sp. *tritici* has proven useful and efficient in delivering new insights on pathogenesis of the wheat yellow rust fungus.

2. Effectoromics of the wheat stem rust fungus *P. graminis* f. sp. *tritici*

A high-throughput effectoromic screen based on BSMV-mediated HIGS was applied to a large set of Pucciniaceae candidate effectors, which were conserved in the three wheat rust fungi *P. graminis* f. sp. *tritici*, *P. striiformis* f. sp. *tritici* and *P. triticina* (Yin *et al.,* 2015). Among 86 silenced genes, 10 were required for full infection by *P. graminis* f. sp. *tritici* including secreted proteins, metabolic and transport-related genes (Yin *et al.,* 2015). Reduced development of *P. striiformis* f. sp. *tritici* and *P. triticina* was also observed through transient silencing for four and three of those genes, respectively, indicating that screens performed at the level of

taxonomical families may reveal crucial functions for infection shared between rust species (Yin *et al.,* 2015). A tryptophan 2-monooxygenase (Pgt-IaaM) gene involved in auxin synthesis is required for full virulence of *P. graminis*. f. sp. *tritici* and displayed pleiotropic auxin-related phenotypes when expressed in arabidopsis (Yin *et al*., 2014). This study demonstrates the modulation of the plant hormone balance by a rust fungus. In their assay, Ramachandran et al. (2016) also screened plant cell death suppression properties of 11 wheat stem rust candidate effectors and two were able to suppress HR in *N. benthamiana*. Another effectoromic screen was established to deliver rust candidate effectors in wheat through a specific type III secretion system (*Pseudomonas fluorescens* effector to host analyzer strain EtHAn) (Upadhyaya *et al.,* 2014). A total of 25 *P. graminis* f. sp. *tritici* candidates were monitored for plant cell death induction and one was able to induce a genotype-specific HR suggesting it maybe an avirulence protein (Upadhyaya *et al.,* 2014). Four candidate effectors predicted by LOCALIZER (cf. section V.2) to accumulate in nuclei or in chloroplasts of plant cells were validated in *N. benthamiana* (Sperschneider *et al.,* 2017).

The recent identification and validation of the wheat stem rust avirulence genes *Avr35* and *Avr50* is a perfect example of the power of combining genomic and functional approaches. In one case, *AvrSr35* has been identified in rust fungal mutants obtained by induced mutagenesis of an avirulent fungal isolate. *AvrSr35* encodes a small secreted protein of 578 amino acids and the interaction with the wheat resistance gene *Sr35* was verified by transient co-expression in *N. benthamiana* as well as by infiltration of the purified protein in wheat leaves (Salcedo *et al.,* 2017). AvrSr35 and Sr35 colocalize with the endoplasmic reticulum and interact in plant cells, however, whether the interaction is direct or indirect remains to be determined. In the second example, sequencing the genomes of an avirulent *AvrSr50* isolate and a spontaneous virulent mutant derived from this isolate identified a 2.5Mb region showing loss-of-heterozygosity (Chen *et al.,* 2017). Forty one genes encoding haustorially expressed secreted proteins present in this region were transiently co-expressed without their signal peptide with the wheat resistance gene *Sr50*, and a single candidate triggered a HR (Chen *et al.,* 2017). AvrSr50 showed a nucleocytolocalisation in *N. benthamiana* and exhibited a cell death suppression activity indicating it may have a role in suppressing defense response (Chen *et al.,* 2017). These back-to-back papers demonstrate the great promises of combining functional assays with (population)-genomics and transcriptomics to identify new rust avirulence genes.

3. Effectoromics of the wheat leaf rust *P. triticina*

Progress has been made in the study of wheat leaf rust pathogenicity through methodological developments. Complementation of an *Ustilago maydis* MAPK knockout mutant with a

P. triticina MAPK restored virulence, demonstrating that complementation of basidiomycete mutants could be an effective way of studying rust functions (Hu *et al.,* 2007). The knockdown of *P. triticina* signalling genes encoding an adenylate cyclase, a MAP kinase and a calcineurin through BSMV-mediated HIGS resulted in suppressed disease phenotypes (Panwar *et al.,* 2013). Such a host-delivered RNA interference system could be effective through generations in transgenic wheat lines, conferring protection against rust disease (Panwar *et al.,* 2017). A co-bombardement assay of selected secreted proteins with a β-glucoronidase (GUS) expressing vector has been developed to detect potential avirulence candidates. In this system, a reduction in GUS expression may result from a HR caused by the recognition of the secreted protein by the immune system. Two secreted proteins were flagged as potential avirulence candidates using this biolistic assay (Segovia *et al.,* 2016).

4. Effectoromics of the Asian soybean rust fungus *P. pachyrhizi*

One of the largest and most advanced effectoromic screens performed in rust fungi was recently reported in *P. pachyrhizi* (Qi *et al.*, 2017). The selection of 82 candidate effectors (PpECs) was based on haustoria RNAseq (Link *et al.*, 2014). Qi and colleagues cloned the 82 PpECs in different vectors to i) scrutinize their subcellular localization in *N. benthamiana*, ii) assay induction of an immune response in pepper and tomato, and iii) assay cell death suppression in *N. benthamiana*, arabidopsis and yeast. Among the 82 PpECs, 17 suppress plant immunity and 30 PpECs target different cell compartments such as nucleus, nucleolus and cytoplasm (Qi *et al.*, 2017). The secretion of one candidate, PpEC23, was confirmed in yeast and it was further shown to suppress the immune response in soybean. Interaction of PpEC23 with the soybean transcription repressor GmSPL12I was shown by yeast-two-hybrid and pull-down. PpEC23 interacts with itself in the cytoplasm and it localizes in the nucleus when it interacts with GmSPL12I (Qi *et al.*, 2016). With this large effectoromic screen, the authors can propose a shortlist of 17 *P. pachyrhizi* priority candidates that can now be specifically addressed in dedicated efforts (Qi *et al.*, 2017). This clearly illustrates the power of systematic and methodic large screens applied to rust fungi.

Parallel screening efforts were made to identify soybean rust fungus candidate effectors. For instance, six candidates selected by transcriptomics showed suppression of plant cell death triggered by *Pseudomonas* syringae pv. *tomato* DC3000 and accumulation in nucleus and cytosol after transient expression in *N. benthamiana* (de Carvalho *et al.,* 2016). In another study, two out of six selected candidates were found to increase *P. infestans* infection in *N. benthamiana*, suggesting role in the virulence of the soybean rust fungus (Kunjeti *et al*., 2016).

5. Effectoromics of the poplar rust fungus *M. larici-populina*

Twenty candidate effectors of the poplar rust fungus have been investigated in *N. benthamiana* to identify their cellular localization and their putative plant partners (Petre *et al.,* 2015). Eight candidates showed either a specific localization when transiently expressed without their signal peptide in fusion with GFP or a specific interaction with plant proteins by co-IP/MS. The candidate Mlp124017 was shown to interact with the poplar TOPLESS-related proteins 4 by co-IP when co-expressed in *N. benthamiana* (Petre *et al.,* 2015). This study also revealed a candidate effector targeting chloroplasts (chloroplast-targeted protein 1; MlpCTP1). A transit peptide in MlpCTP1 is necessary and sufficient to accumulate in chloroplasts (Petre *et al.,* 2016b). This gene is part of a Melampsoracae family and other members were shown to also possess a transit peptide and to target chloroplasts in plant cells. Sixteen other *M. larici-populina* candidates were assayed in *A. thaliana* through stable expression of GFP fusion to determine their subcellular localization (Germain *et al.,* 2017). Three candidates showed an accumulation in plasmodesmata, chloroplasts and cytoplasmic bodies and eleven promoted *Hyaloperonospora arabidopsis* growth *in planta* while five promoted *P. syringae* growth *in planta*. Three candidates promoted both oomycete and bacterial growth and may be considered as virulence factors (Germain *et al.,* 2017). Investigations conducted in two different heterologous systems highlights 11 poplar-rust candidate effectors with specific localization and/or plant protein interactors (Table 3).

Table 3: Effectoromic screens conducted on rust fungi

Reference	Approaches	Screened CSEPs	Highlighted CSEPs
Hemileia vastatrix			
Maia et al. 2017	Plant cell death suppression in coffee leaves	30	1
Melampsora larici-populina			
Petre et al. 2015; 2016	Localization and Co-IP/MS of CSEPs–fluorescent tag in *N. benthamiana*	20	8
Germain et al. 2017	Infection assays and localization of CSEPs–fluorescent tag in *Arabidopsis*	16	3
Melampsora lini			
Zhang et al. 2016	Recombinant protein production and structural analysis	4	4
Phakopsora pachyrihzi			

Qi et al. 2017	Localization and plant cell death suppression in *N. benthamiana, Arabidopsis* and yeast, immune induction in tomato and pepper	82	17
de Carvalho et al. 2016	Localization and plant cell death suppression in *N. benthamiana*	6	6
Kunjeti et al. 2016	Plant cell death suppression in *N. benthamiana*	6	2
Puccinia graminis f. sp. *tritici*			
Yin et al. 2015	HIGS to select proteins necessary for pathogenicity	76	10
Upadhyaya et al. 2014	HIGS to select proteins necessary for pathogenicity	25	1
Ramachandran et al. 2016	Plant cell death suppression in *N. benthamiana*	11	2
Puccinia striiformis f. sp. *tritici*			
Petre et al. 2016b	Localization and Co-IP/MS* of CSEPs–fluorescent tag in *N. benthamiana*	16	9
Kang and collaborators * (Northwest A&F University)	HIGS to select proteins necessary for pathogenicity	11	2
Ramachandran et al. 2017	Plant cell death suppression in *N. benthamiana*	9	7
Liu, Pederson et al. 2016	Plant cell death suppression in *N. benthamiana* and infection assays in *Arabidopsis*	6	1
Uromyces appendiculatus			
Cooper & Campbell, 2017	HIGS to select proteins necessary for pathogenicity	5	3

* Guo et al. 2011; Liu et al. 2014; Cheng et al. 2015; Tang et al. 2015; Cheng et al. 2016a; Cheng et al. 2016b; Liu, Guan et al. 2016; Jiao et al. 2017; Wang et al. 2017; Zhu et al. 2017; Qi, Zhu et al. 2017
Co-IP/MS co-immonuprecipitation coupled with mass spectrometry; HIGS host-induced gene silencing.

7. Effectoromics of the coffee rust fungus *H. vastarix*

Hundreds of candidate effectors have been identified in *H. vastatrix* by genomic and transcriptomic approaches (Fernandez *et al.,* 2012; Cristancho *et al.,* 2014; Talhinhas *et al.,* 2014). Similar to effectoromic screens conducted in *P. graminis* f. sp. *tritici*, a medium-throughput screen has been established to deliver candidates in coffee leaves through *Pseudomonas syringae* pv. *garcae* (Maia *et al.,* 2017). The impact of 30 *H. vastatrix* candidate

effectors delivered into coffee leaves on the symptoms caused by *P. syringae* pv. *garcae* infection was assessed to identify avirulence candidates. HvEC-016 reduced bacterial growth and suppressed plant cell death in coffee genotypes carrying a specific resistance gene (Maia *et al.,* 2017). This elegant approach successfully applied to a complex rust pathosystem opens promising perspectives for the identification of avirulence genes.

8. *Uromyces* spp. effectoromics

U. fabae has been a pioneer rust species for the analysis of the infection process, notably through the description of haustorially expressed genes (Hahn & Mendgen, 1997). The secretion of about a hundred *U. fabae* candidate secreted proteins has been validated using the yeast signal sequence trap system (Link & Voegele, 2008). Most functions studied in this fungus are discussed in section IV. A small-throughput screen has been established to assay five *Uromyces appendiculatus* proteins through *bean pod mottle virus*-mediated HIGS. Silencing of a trehalose phosphatase, a chitinase-like protein and a glycoside hydrolase prevented *U. appendiculatus* infection suggesting they might play an important role in fungal development and plant infection (Cooper *et al.,* 2016; Cooper & Campbell, 2017).

VII- Putting rusts to sleep: Pucciniales research outlooks

Genomics has been a fantastic lever to foster progress in the field of rust research (Duplessis et al. 2016). Despite the complexity of rust genomes, latest genomic reports are promising and the systematic use of long fragment sequencing techniques will foster the resolution of their assembly. It remains to be determined how these methods will do for rust fungi with genome size expected to be over 1Gb. The complexity of very large rust genomes may stand as an impediment for completion of genomic projects for agricultural flagship species. On the other hand, many reports have proven that transcriptomics can be a promising alternative in absence of a reference genome (see Link *et al.,* 2014 and Qi *et al.,* 2017 for *P. pachyrhizi* or Fernandez *et al.,* 2012 and Maia *et al.,* 2017 for *H. vastatrix*). The large repertoires of small secreted proteins presenting typical features of effectors in rust genomes is impressive with hundreds to thousands of candidates per species. Expression of these genes in successive waves indicates that cohorts of effectors may act in a concerted way for the fungus to progress unbeknown to the host plant and to maintain the biotrophic state. Whether effectors act along together or as complexes within the host cell will have to de determined. The many facets of the complex life cycles of rust fungi should be more systematically explored to have a better understanding of heteroecism. The recent transcriptomic studies conducted during alternate host infection for a few rust species provide a deeper knowledge of rust pathogenesis. In

heteroecious rust fungi of major crops, disease management can benefit from a better knowledge of reproductive mechanisms on the alternate host.

By providing an exhaustive survey of efforts deployed to search for effector proteins among hundreds of candidates, our review illustrates how the field of rust research has moved fast. The multiplicity of the rust pathosystems and the systematic application of cellular and molecular screens based on heterologous plant systems are many additions that pave the way towards a better understanding of the pathogenesis process. The success gained for the many rust fungi assayed so far indicates that similar efforts can be applied to any major rust fungus impacting agriculture or environement. The same tools can be employed to identify and validate airulence genes that are usefull informations to make progess in host resistance management, and the recent identification of the avirulence genes *AvrSr35* and *AvrSr50* of the wheat stem rust fungus are perfect illustrations (Moscou & van Esse, 2017; Chen *et al.,* 2017; Salcedo *et al.,* 2017).

Tools to properly address the molecular bases of rust-plant interaction are still cruelly lacking. For years, attempts to genetically transform rust fungi have been made with limited successes (Djulic *et al.,* 2011; Lawrence *et al.,* 2010). Gene silencing is the best proxy at the moment and it is widely and successfully applied to *P. striiformis* f. sp. *tritici* (Panwar *et al.,* 2017; Tang *et al.,* 2017). Current efforts are ongoing to achieve genetic transformation for some Pucciniaceae (Dr. B. Schwessinger, ANU, Australia; Prs. J. Guo and Z. Kang, NWAFU, China; personal communication), which would be the best way to address critical functions during interaction with the host plant (Figure 4). Development of standard routines in genomics or in the design of functional pipelines is also needed in the community to allow comparative studies. The recent focus done on effector genes has somehow clouded other categories related to infection. If understanding how te pathogen is able to re-wire the plant immune system to avoid recognition and establish in the host is important, it is as crucial to determine how it is able to re-route nutrients and sustain its growth and spore multiplication.

Progress on rust effector biology and pathogenesis in rust fungi can inform the selection of new resistances in the host plant or new means to control or contain disease in the field (Figure 4). Recent reviews have largely discussed how molecular research on effectors can translate into resistance management (Moscou & van Esse, 2017; Boch *et al.,* 2014; Nelson *et al.,* 2018). We should keep in mind that the gap between fundamental and applied research remains large and the fast progress made on all fronts in understanding obligate biotrophy support the fight against these fascinating plant pathogens that are rust fungi.

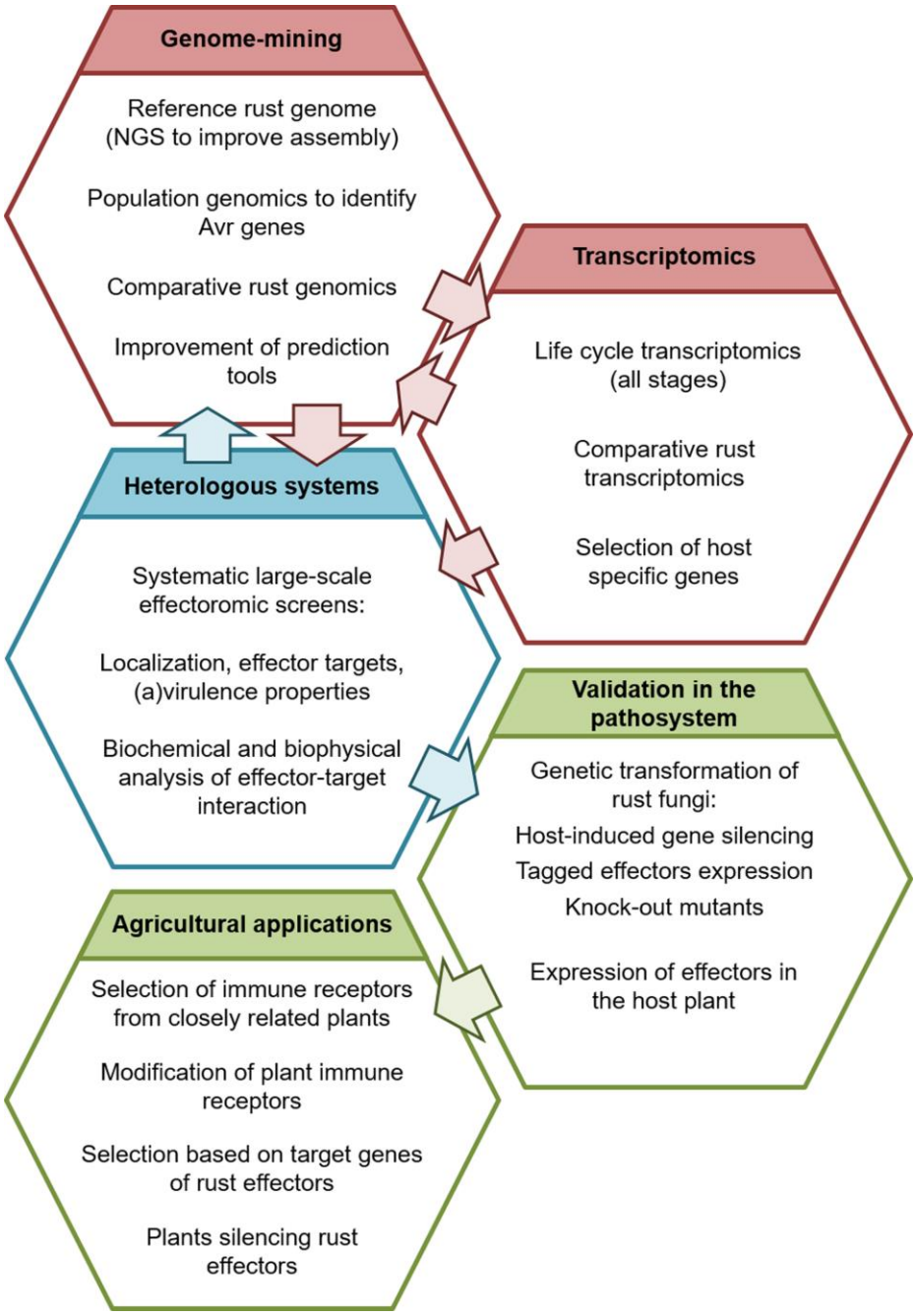


Figure 4: Perspectives for rust fungal effector biology.

Schematic illustration of perspectives for rust effector biology. Perspectives related to genomic and transcriptomic approaches are indicated in red; to heterologous systems are represented in blue and in green for the pathosystem. Better assembly of reference genomes, access to populational information and life cycle-informed transcriptomics will support an improved mining for CSEPs and avirulence genes identification in rust fungi; comparative rust genomics will enable better identification of core and lineage-specific CSEPs and improvement of prediction tools will help prioritization of CSEPs. Effectoromic screens pinpoint priority CSEPs with particular localization, plant target and avirulence properties. The main challenge of rust effector biology is to validate candidate effectors in the pathosystem through transformation of rust fungi and their hosts. Some approaches like host-induced gene silencing and expression of CSEPs in the host plant are readily available but so far the obtention of knock-out mutants remains a major drawback. The main output for rust effector research is translation to field disease management with selection of new resistances or less susceptible plants, some possible future applications in agriculture.

Acknowledgements

The authors would like to thank Aimee Orsini for her patience and for correcting English in our manuscript. CL and SD are supported by the French 'Investissements d'Avenir' program (ANR-11-LABX-0002-01, Lab of Excellence ARBRE). CL is supported by a young scientist grant (CJS) from INRA. SD and HG benefited from a CFQCU French-Quebec Cooperative project established between the Université de Lorraine (France) and the Université du Québec à Trois-Rivières (Canada).

LITTERATURE CITED

Aime, M.C., MacTaggart A.R, Mondo, S.J., & Duplessis, S. (2017). Phylogenetics and phylogenomics of rust fungi. *Advances in fungal Genetics*.100: 267-307.

Aime, M., Toome, M., & McLaughlin, D. (2014). The Pucciniomycotina. *In* eds *McLaughlin D., Spatafora J.W.*, The Mycota VII Part A. Springer-Verlag, 271–294.

Aime, M. C., (2006). Toward resolving family-level relationships in rust fungi (Uredinales). *Mycoscience*, 47(3), 112–122.

Aime, M. C., Matheny, P. B., Henk, D. A., Frieders, E. M., Nilsson, R. H., Piepenbring, M., et al. (2006). An overview of the higher level classification of Pucciniomycotina based on combined analyses of nuclear large and small subunit rDNA sequences. Mycologia, 98(6), 896–905.

Alfano, J. R. (2009). Roadmap for future research on plant pathogen effectors. *Molecular Plant Pathology*, 10(6), 805–813.

Anderson, C., Khan, M. A., Catanzariti, A.-M., Jack, C. A., Nemri, A., Lawrence, G. J., Jones, D. A. (2016). Genome analysis and avirulence gene cloning using a high-density RADseq linkage map of the flax rust fungus, *Melampsora lini. BMC Genomics*, 17, 667.

Anikster, Y., Szabo, L. J., Eilam, T., Manisterski, J., Koike, S. T., & Busnell, W. R. 2004. Morphology, life cycle biology and DNA sequence analysis of rust fungi on garlic and chives from California. Phytopathology, 9: 569-577.

Barrett, L. G., Thrall, P. H., Dodds, P. N., van der Merwe, M., Linde, C. C., Lawrence, G. J., & Burdon, J. J. 2009. Diversity and Evolution of Effector Loci in Natural Populations of the Plant Pathogen *Melampsora lini*. *Molecular Biology and Evolution*, 26(11), 2499–2513.

Beddow, J. M., Pardey, P. G., Chai, Y., Hurley, T. M., Kriticos, D. J., Braun, H.-J., & Yonow, T. 2015. Research investment implications of shifts in the global geography of wheat stripe rust. *Nature Plants*, 1, 15132.

Boch, J., Bonas, U., & Lahaye, T. 2014. TAL effectors-pathogen strategies and plant resistance engineering. *New Phytologist*, 204(4):823-32

Bolton, M. D., Kolmer, J. A., & Garvin, D. F. 2008. Wheat leaf rust caused by *Puccinia triticina*. *Molecular Plant Pathology*, 9(5), 563–575.

Bruce, M., Neugebauer, K., Joly, D., Migeon, P., Cuomo, C., Wang, S., & Fellers, J. 2014. Using transcription of six *Puccinia triticina* races to identify the effective secretome during infection of wheat. *Frontiers in Plant Science*, 4, 520.

Bueno-Sancho, V., Persoons, A., Hubbard, A., Cabrera-Quio, L. E., Lewis, C. M., Corredor-Moreno, P., Bunting, D.C.E., Ali, S., Cheng, S., Hodson, D. P., Madariaga-Burrow, R., Bryson, R., Thomas, J., Holdgate, S., & Saunders, D. G. O. (2017). Pathogenomic Analysis of Wheat Yellow Rust Lineages Detects Seasonal Variation and Host Specificity. *Genome Biology and Evolution*, 9(12), 3282–3296

Cantu, D., Segovia, V., MacLean, D., Bayles, R., Chen, X., Kamoun, S., Dubcovsky, J., Saunders, D.G.O., & Uauy, C. 2013. Genome analyses of the wheat yellow (stripe) rust pathogen *Puccinia striiformis* f. sp. *tritici* reveal polymorphic and haustorial expressed secreted proteins as candidate effectors. *BMC Genomics*, 14: 270.

Cantu, D., Govindarajulu, M., Kozik, A., Wang, M., Chen, X., Kojima, K.K., Jurka, J., Michelmore, R.W., Dubcovsky, J., & Harris, S. 2011. Next generation sequencing provides rapid access to the genome of *Puccinia striiformis* f. sp. *tritici*, the causal agent of wheat stripe rust. *PLoS ONE*, 6(8): e24230.

Chang, Q., Liu, J., Lin, X., Hu, S., Yang, Y., Li, D., Chen, L., Huai, B., Huang, L., Voegele, R.T., & Kang, Z. 2017. A unique invertase is important for sugar absorption of an obligate biotrophic pathogen during infection. *New Phytologist*, 215(4), 1548–1561.

Chen, X. & Kang, Z. 2017. Stripe Rust, 719.

Chen J., Upadhyaya N.M., Ortiz D., Sperschneider J., Li F., Bouton C., Breen S., Dong C., Xu B., Zhang X., Mago R., Newell K., Xia X., Bernoux M., Taylor J.M., Steffenson B., Jin Y., Zhang P., Kanyuka K., Figueroa M., Ellis J.G., Park R.F., & Dodds P.N. 2017 Loss of AvrSr50 by somatic exchange in stem rust leads to virulence for Sr50 resistance in wheat. *Science,* 358:1607-1610.

Cheng Y., Wu K., Yao J., Li S., Wang X., Huang L., & Kang Z. 2016. PSTha5a23, a candidate effector from the obligate biotrophic pathogen *Puccinia striiformis* f. sp. *tritici*, is involved in plant defense suppression and rust pathogenicity. *Environmental Microbiology*, 19(5), 1717-1729.

Cheng, Y., Wang, W., Yao, J., Huang, L., Voegele, R. T., Wang, X., & Kang, Z. 2016. Two distinct Ras genes from *Puccinia striiformis* exhibit differential roles in rust pathogenicity and cell death. *Environmental Microbiology*, 18(11), 3910–3922.

Cheng Y., Wang X., Yao J., Voegele R.T., Zhang Y., Wang W., Huang L., & Kang Z. 2015 Characterization of protein kinase PsSRPKL, a novel pathogenicity factor in the wheat stripe rust fungus. *Environmental Microbiology*, 17(8), 2601–2617.

Chen, W., Wellings, C., Chen, X., Kang, Z., & Liu, T. 2014. Wheat stripe (yellow) rust caused by *Puccinia striiformis* f. sp. *tritici*. *Molecular Plant Pathology*, 15(5), 433–446.

Chen, Z., Nunes, M. A., Silva, M. C., & Rodrigues, C. J. 2004. Appressorium turgor pressure of *Colletotrichum kahawae* might have a role in coffee cuticle penetration. *Mycologia*, 96(6), 1199–1208.

Cook, D. E., Mesarich, C. H., & Thomma, B. P. H. J. 2015. Understanding plant immunity as a surveillance system to detect invasion. *Annual Review of Phytopathology*, 53, 541‐563.

Cooper, B., & Campbell, K. B. 2017. Protection Against Common Bean Rust Conferred by a Gene-Silencing Method. *Phytopathology*, 107(8), 920–927.

Cooper, B., Campbell, K. B., Beard, H. S., Garrett, W. M., & Islam, N. 2016. Putative Rust Fungal Effector Proteins in Infected Bean and Soybean Leaves. *Phytopathology*, 106(5), 491–499.

Cristancho, M. A., Botero-Rozo, D. O., Giraldo, W., Tabima, J., Riaño-Pachón, D. M., Escobar, C., Rozo, Y., Rivera, L.F., Duran, A., Restrepo, S., Eilam, T., Anikster, Y., & Gaitan, A.L. 2014. Annotation of a hybrid partial genome of the coffee rust (*Hemileia vastatrix*) contributes to the gene repertoire catalog of the Pucciniales. *Frontiers in Plant Science*, 5, 594.

Cuomo, C.A., Bakkeren, G., Khalil, H.B., Panwar, V., Joly, D., Linning, R., Sakthikumar, S., Song, X., Adiconis, X., Fan, L., Goldberg, J.M., Levin, J.Z., Young, S., Zeng, Q., Anikster, Y., Bruce, M., Wang, M., Yin, C., McCallum, B., Szabo, L.J., Hulbert, S., Chen, X., & Fellers, J.P. 2017. Comparative analysis highlights variable genome content of wheat rusts and divergence of the mating loci. *G3*, 7: 361-376.

Cummins, G., & Hiratsuka, Y. 2003. Illustrated genera of rust fungi (APS, St Paul). pp.225.

Cunningham, G. H. 1931. The rust fungi of New Zealand (Dunnedin).pp.261.

Dagvadorj, B., Ozketen, A. C., Andac, A., Duggan, C., Bozkurt, T. O., & Akkaya, M. S. (2017). A *Puccinia striiformis* f. sp. *tritici* secreted protein activates plant immunity at the cell surface. *Scientific Reports*, 7(1), 1141.

Dean R., van Kan J. A. L., Pretorius Z. A., Hammond-Kosack K. E., di Pietro A., Spanu P. D., Rudd J. J., Dickman M., Kahmann R., Ellis J. & Foster G. D. 2012. The Top 10 fungal pathogens in molecular plant pathology. *Molecular Plant Pathology*, 13: 414–430.

de Carvalho M.C. da C.G., Nascimento L.C., Darben L.M., Polizel-Podanosqui A.M., Lopes-Caitar V.S., Qi M., Rocha C.S., Carazzolle M.F., Kuwahara M.K., Pereira G.A.G., Abdelnoor R.V., Whitham S.A., & Marcelino-Guimaraes F.C. (2016). Prediction of the in planta *Phakopsora pachyrhizi* secretome and potential effector families. *Molecular Plant Pathology*, 18(2017),12405.

Djulic, A., Schmid, A., Lenz, H., Sharma, P., Koch, C., Wirsel, S. G. R., & Voegele, R. T. (2011). Transient transformation of the obligate biotrophic rust fungus Uromyces fabae using biolistics. *Fungal Biology*, 115(7), 633‐642.

Dobon, A., Bunting, D. C. E., Cabrera-Quio, L. E., Uauy, C., & Saunders, D. G. O. (2016). The host-pathogen interaction between wheat and yellow rust induces temporally coordinated waves of gene expression. *BMC Genomics*, 17(1), 380.

Dodds, P. N., & Rathjen, J. P. 2010. Plant immunity: towards an integrated view of plant-pathogen interactions. *Nature Reviews Genetics*, 11(8).

Dodds, P. N., Lawrence, G. J., Catanzariti, A. M., Ayliffe, M. A., & Ellis, J. G. 2004. The *Melampsora* lini Avr567 avirulence genes are expressed in haustoria and their products are recognized inside plant cell. *Plant Cell*, 16.

Duplessis, S., Bakkeren, G., & Hamelin, R. 2014. Advancing knowledge on biology of rust fungi through genomics. *Advances in Botanic Research*, Vol. 70, p. 173‐209.

Duplessis, S., Bakkeren, G., Joly, D.L., eds. 2016 Genomics research on non-model plant pathogens: delivering novel insights into rust fungus biology. *Lausanne: Frontiers Media*. doi: 10.3389/978-2-88919-814-6.

Duplessis, S., Spanu, P. D., & Schirawski, J. 2013. Biotrophic fungi (powdery mildews, Rusts and Smuts). *In Ecological genomics of the fungi, Plant-interacting fungi section, ed R. Martin,* Wiley-Blackwell, 149-168.

Duplessis, S., Joly, D. J., & Dodds, P. N. 2012. Rust effectors. *Effectors in Plant-Microbes Interactions*, eds F. Martin and S. Kamoun (Oxford: Wiley-Blackwell), 155–193.

Duplessis, S., Cuomo, C. A., Lin, Y, C., Aerts, A., Tisserant, E., Veneault- Fourrey, C., Joly, D. L., Hacquard, S., Amselem, J., Cantarel, B., Chiu, R., et al. 2011a. Obligate biotrophy features unravelled by the genomic analysis of rust fungi. *Proceedings of the National Academy of Sciences of the United States of America*. 108: 9166-9171.

Duplessis, S., Hacquard, S., Delaruelle, C., Tisserant, E., Frey, P., Martin, F., Kohler, A. (2011b). *Melampsora larici-populina* transcript profiling during germination and timecourse infection of poplar leaves reveals dynamic expression patterns associated with virulence and biotrophy. *Molecular Plant-Microbe Interactions: MPMI* 24: 808-818.

Ellis, J.G.,Lagudah, E.S., Spielmeyer, W., Dodds, P.N. 2014 The past, present and future of breeding rust resistant wheat. *Frontiers in Plant Science*. Doi:10.3389.fpls.2014.00641

Ellis, J. G., Dodds, P. N., & Lawrence, G. J. 2007. Flax rust resistance gene specificity is based on direct resistance-avirulence protein interactions. *Annual Review of Phytopathology*, 45, 289–306.

Fernandez, D., Tisserant, E., Talhinhas, P., Azinheira, H., Vieira, A., Petitot, A.-S., Loureiro, A., Poulain, J., Da Silva, C., Silva, M. D.C., & Duplessis, S. 2012. 454-pyrosequencing of Coffea arabica leaves infected by the rust fungus Hemileia vastatrix reveals in planta-expressed pathogen-secreted proteins and plant functions in a late compatible plant-rust interaction. *Molecular Plant Pathology*, 13(1), 17–37.

Figueroa, M., Hammond-Kosack, K.E., Solomon, P.S. 2017. A review of wheat diseases—a field perspective. *Mol. Plant Pathol*.

Figueroa, M., Upadhyaya, N. M., Sperschneider, J., Park, R. F., Szabo, L. J., Steffenson, B., Ellis, G.J., & Dodds, P. N. 2016. Changing the Game: Using Integrative Genomics to Probe Virulence Mechanisms of the Stem Rust Pathogen *Puccinia graminis* f. sp. *tritici*. *Frontiers in Plant Science*, 7, 205.

Flor, H. 1971. Current status of the gene-for-gene concept. *Annual Review of Phytopathology*, 9, 275‑296.

Fischer, M.C., Henck, D.A., Briggs, C.J., Brownstein, J.S., Madoff, L.C., McCraw S.L., Gurr, S.J. 2012. Emerging fungal threats to animal, plant and ecosystem health. *Nature* 484: 186-194.

Garnica, D. P., Nemri, A., Upadhyaya, N. M., Rathjen, J. P., & Dodds, P. N. 2014. The Ins and Outs of Rust Haustoria. *PLOS Pathogens*, 10(9), e1004329.

Garnica, D. P., Upadhyaya, N. M., Dodds, P. N., & Rathjen, J. P. 2013. Strategies for Wheat Stripe Rust Pathogenicity Identified by Transcriptome Sequencing. *PloS One*, 8(6), e67150.

Germain H., Joly D.L., Mireault C., Plourde M.B., Letanneur C., Stewart D., Morency M. J., Petre B., Duplessis S., & Seguin A. (2017) Infection assays in Arabidopsis reveal candidate effectors from the poplar rust fungus that promote susceptibility to bacteria and oomycete pathogens. *Molecular Plant Pathology*, 19: 191-200.

Godoy, C.V., Seixas, C.D.S., Soares, R.M., Marcelino-Guimaraes, F.C., Meyer, M.C., Costamilan, L.M. 2016. Asian soybean rust in Brazil: past, present and future. *Pesq. Agropec*. Bras. 5:407-421.

Goellner, K., Loehrer, M., Langenbach, C., Conrath, U., Koch, E., & Schaffrath, U. 2010. *Phakopsora pachyrhizi*, the causal agent of Asian soybean rust. *Molecular Plant Pathology*, 11(2):169-177.

Guo, J., Dai, X., Xu, J.-R., Wang, Y., Bai, P., Liu, F., Duan, Y., Zhang, H., Huang, L., & Kang, Z. (2011). Molecular Characterization of a Fus3/Kss1 Type MAPK from *Puccinia striiformis* f. sp. *tritici*, PsMAPK1. *PLOS ONE*, 6(7), e21895.

Hahn, M., 2000. The rust fungi: cytology, physiology and molecular biology of infection. *Fungal Pathology* (Kronstad J, ed), pp. 267–306. Kluwer Academic Publishers, Dordrecht.

Hahn, M., & Mendgen, K. 1997. Characterization of in plant-induced rust genes isolated from a haustorium-specific cDNA library. *Mol Plant Microbe Interact*, 10(4): 427-437.

Hahn, M., Neef, U., Struck, C., Göttfert, M., & Mendgen, K. 1997. A putative amino acid transporter is specifically expressed in haustoria of the rust fungus *Uromyces fabae*. *Molecular Plant-Microbe Interactions*, 10(4) : 438–445.

Hacquard S., Joly D.L., Lin Y.-C., Tisserant E., Feau N., Delaruelle C., Legue V., Kohler A., Tanguay P., Petre B., Frey P., Van de Peer Y., Rouzé P., Martin F., Hamelin R. C., & Duplessis S. 2012. A comprehensive analysis of genes encoding small secreted proteins identifies candidate effectors in *Melampsora larici-populina* (poplar leaf rust). *Mol. Plant-Microbe Interact. MPMI*, 25(3):279-293.

Hacquard, S., Petre, B., Frey, P., Hecker, A., Rouhier, N., & Duplessis, S. (2011). The poplar-poplar rust interaction: insights from genomics and transcriptomics. *Journal of Pathogens*, 2011, 716041.

Hacquard, S., Delaruelle, C., Legue, V., Tisserant, E., Kohler, A., Frey, P., Martin, F., & Duplessis, S 2010. Laser Capture microdissection of uredinia formed by *Melampsora larici-populina* revealed a transcriptional switch between biotrophy and sporulation. *Mol. Plant-Microbe Interact.MPMI*, 23 (10): 1275–86.

Hamel, L.-P., Nicole, M.-C., Duplessis, S., & Ellis, B. E. 2012. Mitogen-Activated Protein Kinase Signaling in Plant-Interacting Fungi: Distinct Messages from Conserved Messengers. *The Plant Cell*, 24(4): 1327–1351

Hoch, H. C., Staples, R. C., Whitehed, B., Comeau, J., & Wolf, E. D. 1987. Signaling for growth orientation and cell differentiation by surface topography in *Uromyces*. *Science*, 235, 1659–1663.

Hogenhout, S. A., Van der Hoorn, R. A. L., Terauchi, R., & Kamoun, S. 2009. Emerging concepts in effector biology of plant-associated organisms. *Molecular Plant-Microbe Interactions :* MPMI, 22(2): 115‑122.

Hu, G. G., Linning, R., Mccallum, B., Banks, T., Cloutier, S., Butterfield, Y., Liu, J., Kirkpatrick, R., Stott, J., Yang, G., Smailus, D., Jones, S., Marra, M., Schein, J., & Bakkeren, G. (2007). Generation of a wheat leaf rust, Puccinia triticina, EST database from stage-specific cDNA libraries. *Mol Plant Pathol,* 8.

Jiao, M., Yu, D., Tan, C., Guo, J., Lan, D., Han, E., Guo, J., Lan, D., Han, E., Qi, T., Voegele, R.T., Kang, Z., & Guo, J. (2017). Basidiomycete-specific PsCaMKL1 encoding a CaMK-like protein kinase is required for full virulence of *Puccinia striiformis* f. sp. *tritici*. *Environmental Microbiology*, 19(10):4177-4189.

Jones, J. D. G., & Dangl, J. L. 2006. The plant immune system. *Nature*, 444(7117), 323‑329.

Kemen A. C., Alger M. T., & Kemen E. 2015. Host-microbe and microbe-microbe interactions in the evolution of obligate plant parasitism. *New phytologist*. 206:1207-1228.

Kemen, E., Kemen, A., Ehlers, A., Voegele, R., & Mendgen, K. 2013. A novel structural effector from rust fungi is capable of fibril formation. *The Plant Journal*, 75(5), 767–780.

Kemen, E., and Jones, J. D. G. 2012. Obligate biotroph parasitism: can we link genomes to lifestyles? *Trends in Plant Science*, 17(8), 448‑457.

Kunjeti, S. G., Iyer, G., Johnson, E., Li, E., Broglie, K. E., Rauscher, G., & Rairdan, G. J. (2016). Identification of *Phakopsora pachyrhizi* Candidate Effectors with Virulence Activity in a Distantly Related Pathosystem. *Frontiers in Plant Science*, 7, 269.

Lanver, D., Müller, A.N., Happel, P., Schweizer, G., Haas, F.B., Franitza, M., Pellegrin, C., Reissmann, S., Altmüller, J., Rensing, S.A., Kahmann, R. (2018) The biotrophic development of *Ustilago maydis* studied by RNAseq analysis. *The Plant Cell* doi: 10.1105/tpc.17.00764

Lawrence, G. J., Dodds, P. N., & Ellis, J. G. (2010). Transformation of the flax rust fungus, *Melampsora lini*: selection via silencing of an avirulence gene. *The Plant journal : for cell and molecular biology*, 61(2), 364‑369

Lawrence, G. J., Dodds, P. N., & Ellis, J. G. 2007. Rust of flax and linseed caused by *Melampsora lini*. *Molecular Plant Pathology*, 8(4):349–364.

Loehrer, M., Botterweck, J., Jahnke, J., Mahlmann, D. M., Gaetgens, J., Oldiges, M., Horbach, R., Deising, H., & Schaffrath, U. 2014a. In vivo assessment by Mach-Zehnder double-beam interferometry of the invasive force exerted by the Asian soybean rust fungus (*Phakopsora pachyrhizi*). *The New Phytologist*, 203(2), 620–631.

Leppik, E. E. 1965. Some Viewpoints on the Phylogeny of Rust Fungi. V. Evolution of Biological Specialization. *Mycologia*, 57(1):6–22.

Link, T. I., & Voegele, R. T. 2008. Secreted proteins of Uromyces fabae: similarities and stage specificity. *Molecular Plant Pathology*, 9(1): 59–66.

Link, T. I., Lang, P., Scheffler, B. E., Duke, M. V., Graham, M. A., Cooper, B., Tucker, M. L., van de Mortel, M., Voegele, R.T., Mendgen, K., Baum, T. J., & Whitham, S. A. 2014a. The haustorial transcriptomes of Uromyces appendiculatus and *Phakopsora pachyrhizi* and their candidate effector families. *Molecular Plant Pathology,* 15(4): 379–393.

Link, T., Seibel, C., & Voegele, R. T. 2014b. Early insights into the genome sequence of *Uromyces fabae*. *Frontiers in Plant Science*, 5, 587.

Link, T., Lohaus, G., Heiser, I., Mendgen, K., Hahn, M., & Voegele, R. T. 2005. Characterization of a novel NADP(+)-dependent D-arabitol dehydrogenase from the plant pathogen *Uromyces fabae*. *Biochemical Journal*, 389(2), 289–295.

Littlefield, L. J., & Heath, M. C. 1979. Ultrasrtucure of Rust Fungi. *Academic Press*, 293.

Liu, J., Guan, T., Zheng, P., Chen, L., Yang, Y., Huai, B., Li, D., Chang, Q., Huang, L., & Kang, Z. (2016). An extracellular Zn-only superoxide dismutase from Puccinia striiformis confers enhanced resistance to host-derived oxidative stress. *Environmental Microbiology*, 18(11), 4118–4135.

Liu, J., Wang, Q.-L., Chang, Q., Han, L.-N., Pei, G.-L., Xue, Y.-Q., Jia, L.-M., Zhang, K., Duan, Y.-Y., & Kang, Z. (2014). Isocitrate lyase is required for urediniospore germination of *Puccinia striiformis* f. sp. *tritici*. *Molecular Biology Reports*, 41(12), 7797–7806.

Loehrer, M., Botterweck, J., Jahnke, J., Mahlmann, D. M., Gaetgens, J., Oldiges, M., Horbach, R., Deising, H., & Schaffrath, U. (2014a). In vivo assessment by Mach-Zehnder double-beam interferometry of the invasive force exerted by the Asian soybean rust fungus (Phakopsora pachyrhizi). The New Phytologist, 203(2), 620–631.

Loehrer, M., Vogel, A., Huettel, B., Reinhardt, R., Benes, V., Duplessis, S., Usadel, B., Schaffrath, U. (2014b). On the current status of Phakopsora pachyrhizi genome sequencing. Frontiers in Plant Science, 5, 377.

Lo Presti, L., Lanver, D., Schweizer, G., Tanaka, S., Liang L., Tollot, M., Zuccaro, A., Reissmann, S., Kahmann, R. 2015. Fungal effectors and plant susceptibility. *Annual Review of Plant Biology*, 66(1): 513–545.

Lorrain, C., Marchal, C., Hacquard, S., Delaruelle, C., Petrowski, J., Petre, B., Hecker, A., Frey, P., & Duplessis, S. (2018a). The rust fungus *Melampsora larici-populina* expresses a conserved genetic program and distinct sets of secreted protein genes during infection of its two host plants, larch and poplar. *Molecular Plant-Microbe Interactions : MPMI.*

Lorrain, C., Petre, B., & Duplessis, S. (2018b) Show me the way: rust effector targets in heterologous plant systems. *Current Opinion in Microbiology*, *accepted in press*.

Lorrain, C., Hecker, A., & Duplessis, S. (2015). Effector-Mining in the Poplar Rust Fungus Melampsora larici-populina Secretome. *Frontiers in Plant Science*, 6, 1051.

Maia, T., Badel, J. L., Marin-Ramirez, G., Rocha, C. de M., Fernandes, M. B., da Silva, J. C. F., de Azevedo-Junior, G.-H., & Brommonschenkel, S. H. 2017. The *Hemileia vastatrix* effector HvEC-016 suppresses bacterial blight symptoms in coffee genotypes with the SH 1 rust resistance gene. *The New Phytologist*, 213(3), 1315–1329.

McCallum, B. D., Fetch, T., Chong, J. 2007. Cereal rust control in Canada. *Aust. J. Agric. Res*., 58:639-647.

McTaggart, A. R., Shivas, R. G., van der Nest, M. A., Roux, J., Wingfield, B. D., & Wingfield, M. J. 2016. Host jumps shaped the diversity of extant rust fungi (Pucciniales). *The New Phytologist*, 209(3), 1149‐1158.

Miller, M. E., Zhang, Y., Omidvar, V., Sperschneider, J., Schwessinger, B., Raley, C., Raley, C., Palmer, J. M., Garnica, D., Upadhyaya, N., Rathjen, J., Taylor, J .M., Park, R. F., Dodds, P. N., Hirsch, C. D., Kianian, S. F., & Figueroa, M. 2017. De novo assembly and phasing of dikaryotic genomes from two isolates of *Puccinia coronata* f. sp. *avenae*, the causal agent of oat crown rust. *Preprint in BioRxiv.*

Mims, C. W. 1977. Ultrastructure of teliospore formation in the cedar-apple rust fungus *Gymnosporangium juniperi-virginianae*. *Canadian Journal of Botany*, 55(17): 2319-2329.

Moller, M., & Stukenbrock, E. H. 2017. Evolution and genome architecture in fungal plant pathogens. *Nature Reviews. Microbiology*, 15(12): 756–771

Moscou, M., & van Esse, P. 2017. The quest for durable resistance. *Science*, 358(6370): 1541-1542.

Nelson, R., Wiesner-Hanks, T., Wisser, R., & Balint-Kurti, P. 2018. Navigating complexity to breed disease-resistant crops. *Nature genetics reviews*, 19: 21.

Nemri, A., Saunders, D. G. O., Anderson, C., Upadhyaya, N. M., Win, J., Lawrence, G. J., Jones, D. A., Kamoun, S., Ellis, J. G., & Dodds, P. N. 2014. The genome sequence and effector complement of the flax rust pathogen *Melampsora lini. Frontiers in Plant Science*, 5, 98.

Panwar, V., & Bakkeren, G. 2017. Investigating gene function in cereal rust fungi by Plant-Mediated Virus-Induced Gene Silencing. *In Wheat Rust Diseases* (Humana Press, Vol. 1659). Periyannan S.

Pearson, R. C., Aldwinckle, H. S., & Seem, R. C. 1977. Teliospore germination and basidiospore formation in *Gymnosporangium juniperi-virginianae*: a regression model of temperature and time effects. *Canadian Journal of Botany*, 55(22), 2832–2837.

Pendleton, A. L., Smith, K. E., Feau, N., Martin, F. M., Grigoriev, I. V., Hamelin, R., Nelson, C. D., Burleigh, J. G., & Davis, J. M. 2014. Duplications and losses in gene families of rust pathogens highlight putative effectors. *Frontiers in Plant Science*, 5, 299.

Pennisi, E. 2010. Armed and dangerous. *Science*, 327(5967):804–805.

Periyannan, S. 2017. Wheat Rust Disease, Methods in Molecular Biology, vol 1659.

Perlin, M. 2012. Essential letters in the fungal alphabet: ABC and MFS transporters and their roles in survival and pthogenicity. *Advances in Genetics,* 85 : 211-249.

Pernaci, M., De Mita, S., Andrieux, A., Pétrowski, J., Halkett, F., Duplessis, S. & Frey, P. (2014). Genome-wide patterns of segregation and linkage disequilibrium: the construction of a linkage genetic map of the poplar rust fungus Melampsora larici-populina. *Frontiers in Plant Science*, 5:454.

Persoons, A., Morin, E., Delaruelle, C., Payen, T., Halkett, F., Frey, P., De Mita, S. & Duplessis, S. (2014). Patterns of genomic variation in the poplar rust fungus Melampsora larici-populina identify pathogenesis-related factors. *Frontiers in Plant Science,* 5: 450.

Petre B., Saunders D.G.O., Sklenar J., Lorrain C., Krasileva K.V., Win J., Duplessis S., & Kamoun S. (2016a) Heterologous Expression Screens in *Nicotiana benthamiana* Identify a Candidate Effector of the Wheat Yellow Rust Pathogen that Associates with Processing Bodies. *PloS One*, 11, e0149035.

Petre, B., Lorrain, C., Saunders, D. G. O., Win, J., Sklenar, J., Duplessis, S., & Kamoun, S. (2016b). Rust fungal effectors mimic host transit peptides to translocate into chloroplasts. *Cellular Microbiology*, 18(4), 453‐465.

Petre B., Saunders D.G.O., Sklenar J., Lorrain C., Win J., Duplessis S., & Kamoun S. (2015) Candidate Effector Proteins of the Rust Pathogen *Melampsora larici-populina* Target Diverse Plant Cell Compartments. *Mol. Plant-Microbe Interact*. 28:689–700

Petre B., Joly D.L., & Duplessis S. 2014. Effector proteins of rust fungi. *Front Plant Sci* 5, 416.

Petre B., & Kamoun S. 2014. How do filamentous pathogens deliver effector proteins into plant cells? *PLoS Biololgy*. 12, e1001801.

Pretsch, K., Kemen, A., Kemen, E., Geiger, M., Mendgen, K., & Voegele, R. 2013. The rust transferred proteins-a new family of effector proteins exhibiting protease inhibitor function. *Molecular Plant Pathology,* 14(1), 96–107.

Qi M., Grayczyk J.P., Seitz J.M., Lee Y., Link T.I., Choi D., Pedley K., Voegele R., Baum T., & Whitham S.A. 2017 Suppression or activation of immune responses by predicted secreted proteins of the soybean rust pathogen *Phakopsora pachyrhizi*. *Mol. Plant. Microbe Interact*. 31(1):163-174.

Qi M., Link T.I., Muller M., Hirschburger D., Pudake R.N., Pedley K.F., Braun E., Voegele R.T., Baum T.J., & Whitham S.A. 2016 A Small Cysteine-Rich Protein from the Asian Soybean Rust Fungus, Phakopsora pachyrhizi, Suppresses Plant Immunity. *PLoS Pathog*. 12, e1005827.

Raffaele, S., & Kamoun, S. 2012. Genome evolution in filamentous plant pathogens: why bigger can be better. *Nature Reviews Microbiology*, 10, 417.

Rafiqi, M., Gan, P. H. P., Ravensdale, M., Lawrence, G. J., Ellis, J. G., Jones, D. A., Hardham, A. R., Dodds, P. N. 2010. Internalization of Flax Rust Avirulence Proteins into Flax and Tobacco Cells Can Occur in the Absence of the Pathogen. *The Plant Cell*, 22(6), 2017.

Ramachandran, S. R., Yin, C., Kud, J., Tanaka, K., Mahoney, A. K., Xiao, F., & Hulbert, S. H. 2017. Effectors from Wheat Rust Fungi Suppress Multiple Plant Defense Responses. *Phytopathology*, 107(1), 75–83.

Ramos, A. P., Tavares, S., Tavares, D., Silva, M. D. C., Loureiro, J., & Talhinhas, P. 2015. Flow cytometry reveals that the rust fungus, *Uromyces bidentis* (Pucciniales), possesses the largest fungal genome reported--2489 Mbp. *Molecular Plant Pathology*, 16(9), 1006–1010.

Rutter, W.B., Salcedo, A., Akhunova, A., He, F., Wang, S., Liang, H., Bowden, R.L., & Akhunov, E. (2017) Divergent and convergent modes of interaction between wheat and *Puccinia graminis* f. sp. *tritici* isolates revealed by the comparative gene co-expression network and genome analyses. *BMC Genomics*, 18: 291.

Salcedo, A., Rutter, W., Wang, S., Akhunova, A., Bolus, S., Chao, S., Anderson, N., De Soto, M.F., Rouse, M., Szabo, L., Bowden, R.L., Dubcovsky, J., & Akhunov, E. 2017 Variation in the AvrSr35 gene determines Sr35 resistance against wheat stem rust race Ug99. *Science*, 358:1604-1606.

Savile, D. B. O. 1976. Evolution of the Rust Fungi (Uredinales) as Reflected by Their Ecological Problems. In M. K. Hecht, W. C. Steere, & B. Wallace (Eds.), *Evolutionary Biology*: Volume 9 (pp. 137–207).

Saunders D.G.O., Win J., Cano L.M., Szabo L.J., Kamoun S., & Raffaele S. 2012 Using hierarchical clustering of secreted protein families to classify and rank candidate effectors of rust fungi. *PloS One*, 7, e29847.

Schwessinger, B., Sperschneider, J., Cuddy, W., Miller, M., Garnica, D., Taylor, J., Dodds, P. N., Figueroa, M., Park, R., & Rathjen, J. 2017. A near complete haplotype-phased genome of the dikaryotic wheat stripe rust fungus *Puccinia striiformis* f. sp. *tritici* reveals high inter-haplome diversity. *bioRxiv*.

Segovia, V., Bruce, M., Shoup Rupp J.L., Huang, L., Bakkeren, G., Trick, H.N., Fellers, J. P. 2016. Two small secreted proteins from *Puccinia triticina* induce reduction of ß-glucoronidase transient expression in wheat isolines containing Lr9, Lr24 and Lr26. *Canadian Journal of Plant Pathology*. 38:91-106.

Singh, R. P., Hodson, D. P., Jin, Y., Lagudah, E. S., Ayliffe, M. A., Bhavani, S., Rouse, B. M., Pretorius, Z. A., Szabo, L. J., Huerta-Espino, J., Basnet, B. R., Lan, C., & Hovmoller, M. S. 2015. Emergence and Spread of New Races of Wheat Stem Rust Fungus: Continued Threat to Food Security and Prospects of Genetic Control. *Phytopathology*, 105(7), 872–884.

Singh, R.P., Hodson, D.P., Huerta-Espino, J., Jin, Y., Bhavani, S., Njau, P., Herrera-Foessel, S., Singh, P.K., Singh, S., Govindan, V. 2011. Races of the wheat stem rust fungus is a threat to world wheat production. *Annu. Rev. Phytopathol*. 49:465-481.

Sjamsuridzal, W., Nishida, H., Ogawa, H., Kakishima, M., & Sugiyama, J. 1999. Phylogenetic positions of rust fungi parasitic on ferns: Evidence from 18S rDNA sequence analysis. *Mycoscience*, 40(1), 21–27.

Slaminko, T. L., Miles, M. R., Frederick, R. D., Bonde, M. R., & Hartman, G. L. 2008. New Legume Hosts of *Phakopsora pachyrhizi* Based on Greenhouse Evaluations. *Plant Disease*, 92(5), 767–771.

Sniezko, R.A., Yanchuk, A.D., Kliejunas, J.T., Palmieri, K.M., Alexander, J.M., Frankel, S.J. (2012) Proceedings of the Fourth International Workshop on the Genetics of Host-Parasite Interactions in Forestry: Disease and Insect Resistance in Forest Trees. General Technical Report PSW-GTR-240. Albany, CA: U.S. Department of Agriculture, Forest Service, Pacific Southwest Research Station. 372 p

Spanu, P.D. (2012) The genomics of obligate (and non-obligate) biotrophs. *Annu. Rev. Phytopathol*. 50:91-109.

Spanu, P. D., Abbott, J. C., Amselem, J., Burgis, T. A., Soanes, D. M., Stuber, K., et al. 2010. Genome expansion and gene loss in powdery mildew fungi reveal trade-offs in extreme parasitism. *Science*, 330(6010), 1543‐1546.

Sperschneider J., Taylor J., Dodds P.N., & Duplessis S. 2017a. Computational methods for predicting effectors in rust pathogens. *Wheat Rust Disease - Methods and Protocols*, Springer in press

Sperschneider J., Catanzariti A.M., DeBoer K., Petre B., Gardiner D.M., Singh K.B., Dodds P.N., & Taylor J.M. 2017b LOCALIZER: subcellular localization prediction of both plant and effector proteins in the plant cell. *Scientifique Report*, 7,44598.

Sperschneider J., Dodds P.N., Singh K.B., & Taylor J.M. 2017c ApoplastP: prediction of effectors and plant proteins in the apoplast using machine learning. *bioRxiv*.

Sperschneider J., Gardiner D.M., Dodds P.N., Tini F., Covarelli L., Singh K.B., Manners, J.M., & Taylor J.M. 2016 EffectorP: predicting fungal effector proteins from secretomes using machine learning. *New Phytologist*. 210:743–761.

Sperschneider, J., Dodds, P. N., Gardiner, D. M., Manners, J. M., Singh, K. B. & Taylor, J. M. (2015). Advances and Challenges in Computational Prediction of Effectors from Plant Pathogenic Fungi. *PLoS Pathogens*, 11(5), e1004806.

Staples, R. C. (2000). Research on the rust fungi during the twentieth century. *Annual Review of Phytopathology*, 38(1), 49‐69.

Stewart, A., & Cromey, M. 2011. Identifying disease threats and management practices for bio-energy crops. *Current Opinion in Environmental Sustainability*, 3(1), 75–80.

Struck, C. 2015. Amino acid uptake in rust fungi. *Frontiers in Plant Science*, 6, 40.

Struck, C., Siebels, C., Rommel, O., Wernitz, M., & Hahn, M. 1998. The plasma membrane H(+)-ATPase from the biotrophic rust fungus *Uromyces fabae*: molecular characterization of the gene (PMA1) and functional expression of the enzyme in yeast. *Molecular Plant-Microbe Interactions*, 11(6), 458–465.

Struck, Hahn, & Mendgen. 1996. Plasma Membrane H+-ATPase Activity in Spores, Germ Tubes, and Haustoria of the Rust Fungus Uromyces viciae-fabae. *Fungal Genetics and Biology*, 20(1), 30–35.

Stukenbrock, E. H. 2013. Evolution, selection and isolation: a genomic view of speciation in fungal plant pathogens. *The New Phytologist*, 199(4), 895–907.

Talhinhas, P., Batista, D., Diniz, I., Vieira, A., Silva, D. N., Loureiro, A., Tavares, S., Pereira, A.-P., Azinheira, H.-G., Guerra-Guimaraes, L., Varzea, V., & Silva, M. do C. 2016. The coffee leaf rust pathogen Hemileia vastatrix: one and a half centuries around the tropics. *Molecular Plant Pathology*.

Talhinhas, P., Azinheira, H.G., Vieira, B., Loureiro, A., Tavares, S., Batista, D., Morin, E., Petitot, A.S., Paulo, O.S., Poulain, J., Da Silva, C., Duplessis, S., Silva, Mdo. C., & Fernandez, D. 2014. Overview of the functional virulent genome of the coffee leaf rust pathogen Hemileia vastatrix with an emphasis on early stages of infection. *Front Plant Sci,* 5: 88.

Tao, S.-Q., Cao, B., Tian, C.-M., & Liang, Y.-M. 2017. Comparative transcriptome analysis and identification of candidate effectors in two related rust species (*Gymnosporangium yamadae* and *Gymnosporangium asiaticum*). *BMC Genomics*, 18(1), 651.

Tang C.L., Wang X.J., Cheng Y.L., Liu M.J., Zhao M.X., Wei J. P., & Kang Z. S. 2015 New insights in the battle between wheat and *Puccinia striiformis*. *Front Agr Sci Eng*, 2:101–14

Tang, C., Xu, Q., Zhao, M., Wang, X., & Kang, Z. 2017. Understanding the lifestyles and pathogenicity mechanisms of obligate biotrophic fungi in wheat: The emerging genomics era. *The Crop Journal*.

Tang, C., Wei, J., Han, Q., Liu, R., Duan, X., Fu, Y., Huang, X., Wang, X., & Kang, Z. 2015. PsANT, the adenine nucleotide translocase of *Puccinia striiformis*, promotes cell death and fungal growth. *Scientific Reports*, 5, 11241.

Tavares, S., Ramos, A. P., Pires, A. S., Azinheira, H. G., Caldeirinha, P., Link, T., Abranches, R., Silva, M. D. C., Voegele, R. T., Loureiro,J., & Talhinhas, P. 2014. Genome size analyses of Pucciniales reveal the largest fungal genomes. *Frontiers in Plant Science*, 5, 422.

Terhune, B. T., Bojko, R. J., & Hoch, H. C. 1993. Deformation of Stomatal Guard Cell Lips and Microfabricated Artificial Topographies during Appressorium Formation by Uromyces. Experimental *Mycology*, 17(1), 70–78.

Toome-Heller, M. 2016. Latest developments in the research of rust fungi and their allies (Pucciniomycotina). In D.-W. Li (Ed.), Biology of microfungi (pp. 147–168). Cham: Springer International Publishing.

Tremblay, A., Hosseini, P., Li, S., Alkharouf, N. W., & Matthews, B. F. 2013. Analysis of Phakopsora pachyrhizi transcript abundance in critical pathways at four time-points during infection of a susceptible soybean cultivar using deep sequencing. *BMC Genomics*, 14(1), 614

Upadhyaya, N. M., Garnica, D. P., Karaoglu, H., Sperschneider, J., Nemri, A., Xu, B., Mago, E., Cuomo, C. A., Rathjen, J. P., Park, R. F., Ellis, J. G., Dodds, P. N. 2014a. Comparative genomics of Australian isolates of the wheat stem rust pathogen *Puccinia graminis* f. sp. *tritici* reveals extensive polymorphism in candidate effector genes. *Frontiers in Plant Science*, 5, 759.

Upadhyaya, N.M., Mago, R., Staskawicz, B.J., Ayliffe, M.A., Ellis, J.G., Dodds, P.N. 2014b. A bacterial type III secretion assay for delivery of fungal effector proteins into wheat. *Molecular Plant-microbe interactions*, 27: 255-264.

Ve, T., Williams, S. J., Catanzariti, A.-M., Rafiqi, M., Rahman, M., Ellis, J. G., Hardham, A. R., Jones, D. A., Anderson, P. A., Dodds, P. N., & Kobe, B. 2013. Structures of the flax-rust effector AvrM reveal insights into the molecular basis of plant-cell entry and effector-triggered immunity. *Proceedings of the National Academy of Sciences of the United States of America*, 110(43), 17594–17599.

Vialle, A., Frey, P., Hambleton, S., Bernier, L., & Hamelin, R. C. 2011. Poplar rust systematics and refinement of *Melampsora* species delineation. Fungal Diversity, 50(1), 227.

Voegele, R. T., & Mendgen, K. 2011. Nutrient uptake in rust fungi: how sweet is parasitic life? *Euphytica*, 179, 41–55.

Voegele, R. T., Hahn, M., & Mendgen, K. 2009. The Uredinales: Cytology, Biochemistry, and Molecular Biology. *In The Mycota* (A Comprehensive Treatise on Fungi as Experimental Systems for Basic and Applied Research) (Deising H.B. (eds) Plant Relationships, Vol. 5: 69–98

Voegele, R. T., Wirsel, S., Moll, U., Lechner, M., & Mendgen, K. 2006. Cloning and characterization of a novel invertase from the obligate biotroph *Uromyces fabae* and analysis of expression patterns of host and pathogen invertases in the course of infection. *Molecular Plant-Microbe Interactions*, 19(6), 625–634.

Voegele, R. T., Hahn, M., Lohaus, G., Link, T., Heiser, I., & Mendgen, K. 2005. Possible Roles for Mannitol and Mannitol Dehydrogenase in the Biotrophic Plant Pathogen Uromyces fabae. Plant Physiology, 137(1), 190–198.

Voegele, R. T., & Mendgen, K. 2003. Rust haustorium: Nutrient uptake and beyond. *New Phytol*, 159.

Voegele, R. T., Struck, C., Hahn, M., & Mendgen, K. 2001. The role of haustoria in sugar supply during infection of broad bean by the rust fungus *Uromyces fabae*. *Proc Natl Acad Sci USA*, 98.

Wang C.I.A., Guncar G., Forwood J.K., Teh T., Catanzariti A.M., Lawrence G.J., Loughlin F.E., Mackay J.P., Schirra H.J., Anderson P.A., Ellis J.G, Dodds P.N., & Kobe B. 2007. Crystal structures of flax rust avirulence proteins AvrL567-A and -D reveal details of the structural basis for flax disease resistance specificity. *Plant Cell*, 19:2898–2912.

Wang X., Yang B., Li K., Kang Z., Cantu D., & Dubcovsky J. (2016) A Conserved *Puccinia striiformis* Protein Interacts with Wheat NPR1 and Reduces Induction of Pathogenesis-Related Genes in Response to Pathogens. *Mol. Plant-Microbe Interact*., 29:977–989.

Weber, H., & Roitsch, T. (2000). Invertases and life beyond sucrose cleavage. *Trends in Plant Science*, 5(2), 47–48.

Wiese, J., Kleber, R., Hampp, R., & Nehls, U. (2000). Functional Characterization of the *Amanita muscaria* Monosaccharide Transporter, AmMst1. *Plant Biology*, 2(3), 278–282.

Williams, P. G. 1984. Obligate parasitism and axenic culture. In *The cereal rust volume* I origins, specificity, structure and physiology, Academic Press, 1, 399–430.

Wilson, R. A., & Talbot, N. J. 2009. Under pressure: investigating the biology of plant infection by *Magnaporthe oryzae*. *Nature Reviews Microbiology*, 7, 185.

Wu, J., Wang, Q., Liu, S., Huang, S., Mu, J., Zeng, Q., Huang, L., Han, D., & Kang, Z. 2017. Saturation Mapping of a Major Effect QTL for Stripe Rust Resistance on Wheat Chromosome 2B in Cultivar Napo 63 Using SNP Genotyping Arrays. *Frontiers in Plant Science*, 8, 653.

Xu, J., Linning, R., Fellers, J., Dickinson, M., Zhu, W., Antonov, I., Joly, D.L., Donaldson, M.E., Eilam, T., Anikster, Y., Banks, T., Munro, S., Mayo, M., Wynhoven, B., Ali, J., Moore, R., McCallum, B., Borodovsky, M., Saville, B., & Bakkeren, G. (2011) Gene discovery in EST sequences from the wheat leaf rust fungus *Puccinia triticina* sexual spores, asexual spores and haustoria, compared to other rust and corn smut fungi. *BMC Genomics*, 12: 161.

Yin, C., Downey, S. I., Klages-Mundt, N. L., Ramachandran, S., Chen, X., Szabo, L. J., Kigawa, T., Kamoun, S., & Hulbert, S. H. 2015. Identification of promising host-induced silencing targets among genes preferentially transcribed in haustoria of *Puccinia*. *BMC Genomics*, 16(1), 579.

Yin, C., Park, J.-J., Gang, D. R., & Hulbert, S. H. 2014. Characterization of a tryptophan 2-monooxygenase gene from *Puccinia graminis* f. sp. *tritici* involved in auxin biosynthesis and rust pathogenicity. *Molecular Plant-Microbe Interactions*, 27(3), 227–235.

Yuan, C., Wang, M., Skinner, D. Z., See, D. R., Xia, C., Guo, X., & Chen, X. 2018. Inheritance of Virulence, Construction of a Linkage Map, and Mapping Dominant Virulence Genes in *Puccinia striiformis* f. sp. *tritici* Through Characterization of a Sexual Population with Genotyping-by-Sequencing. *Phytopathology*, 108(1), 133–141.

Zale, J., Freshour, L., Agarwal, S., Sorochan, J., Ownley, B. H., Gwinn, K. D., & Castlebury, L. A. 2008. First Report of Rust on Switchgrass (*Panicum virgatum*) Caused by *Puccinia emaculata* in Tennessee. *Plant Disease*, 92(12), 1710–1710.

Zhang, X., Nguyen, N., Breen, S., Outram, M. A., Dodds, P. N., Kobe, B., Solomon, P. S., & Williams, S. J. 2017a. Production of small cysteine-rich effector proteins in *Escherichia coli* for structural and functional studies. *Molecular Plant Pathology*, 18(1), 141‐151.

Zhang X., Farah N., Rolston L., Ericsson D.J., Catanzariti A.M., Bernoux M., Ve T., Bendak K., Chen C., Mackay J.P., Lawrence G.J., Hardham A., Ellis J.G., Williams S.J., Dodds P.N., Jones D.A.,& Kobe B. 2017b. Crystal structure of the *Melampsora lini* effector AvrP reveals insights into a possible nuclear function and recognition by the flax disease resistance protein. *P. Mol. Plant Pathol*., 12597.

Zhu, X., Liu, W., Chu, X., Sun, Q., Tan, C., Yang, Q., Jiao, M., Guo, J., & Kang, Z. 2017a. The transcription factor PstSTE12 is required for virulence of *Puccinia striiformis f. sp. tritici. Molecular Plant Pathology*.

Zhu X, Qi T, Yang Q, et al. 2017b. Host-Induced Gene Silencing of the MAPKK Gene PsFUZ7 Confers Stable Resistance to Wheat Stripe Rust. *Plant physiology*.;175(4):1853-186

Article n°2: Effector-mining in the poplar rust fungus *M. larici-populina* secretome

Cécile Lorrain, Arnaud Hecker, Sébastien Duplessis

Article publié, le 15 décembre 2015 dans la revue *Frontiers in Plant Science*

Effector-mining in the poplar rust fungus *Melampsora larici-populina* secretome

Cécile Lorrain[1,2], Arnaud Hecker[1,2] and Sébastien Duplessis[1,2*]

[1] INRA, UMR 1136 Interactions Arbres/Microorganismes INRA/Université de Lorraine, Centre INRA Nancy Lorraine, 54280 Champenoux, France

[2] Université de Lorraine, UMR 1136 Interactions Arbres/Microorganismes Université de Lorraine/INRA, Faculté des Sciences et Technologies, 54506 Vandoeuvre-lès-Nancy, France

* Sébastien Duplessis, UMR 1136 Interactions Arbres/Microorganismes INRA/Université de Lorraine, Centre INRA Nancy Lorraine, 54280 Champenoux, France
duplessi@nancy.inra.fr

Abstract:
The poplar leaf rust fungus, *Melampsora larici-populina* has been established as a tree-microbe interaction model. Understanding the molecular mechanisms controlling infection by pathogens appears essential for durable management of tree plantations. In biotrophic plant-parasites, effectors are known to condition host cell colonization. Thus, investigation of candidate secreted effector proteins is a major goal in the poplar-poplar rust interaction. Unlike oomycetes, fungal effectors do not share conserved motifs and candidate prediction relies on a set of *a priori* criteria established from reported *bona fide* effectors. Secretome prediction, genome-wide analysis of gene families and transcriptomics of *M. larici-populina* have led to catalogues of more than a thousand secreted proteins. Automatized effector-mining pipelines hold great promise for rapid and systematic identification and prioritization of candidate secreted effector proteins for functional characterization. In this review, we report on and discuss the current status of the poplar rust fungus secretome and prediction of candidate effectors in this species.

Keywords: secretome, effector protein, poplar rust, prediction pipeline, manual annotation

Introduction

Filamentous plant pathogens share the common feature of manipulating immunity and physiology of their hosts by secreting molecules (Jones & Dangl, 2006 ; Kamoun, 2009 ; Win et al., 2012). Among these, secreted proteins and secondary metabolites can be defined as key players in the outcome and stability of host-parasite interactions with very diverse functions (Rovenich *et al*., 2014; Maclean *et al.,* 2014; Pusztahelyi *et al.,* 2015; Lo Presti *et al.,* 2015). Chemical effectors (i.e. secondary metabolites) are mostly secreted by necrotrophs and hemibiotrophs during their necrotrophic phase (Kemen *et al.,* 2015). Obligate biotrophs are organism that grow, feed and reproduce on living host tissues. They exhibit small or very reduced sets of genes encoding secondary metabolites and cell-wall degrading enzymes while they possess large repertoires of effector proteins (Duplessis *et al.,* 2014a; Kemen *et al.,* 2015). In the case of obligate biotrophs such as rust fungi, investigations have mostly focused on secreted proteins (SPs) of plant-associated organisms (i.e. the secretome) that could be candidate secreted effector proteins (CSEPs).

Rust fungi (Pucciniales, Basidiomycetes) are among the most studied fungal obligate biotrophs due to the important damage they cause on many cultivated plants (Dean *et al*., 2012). Rust fungi are physically associated with their host cells through the formation of specialized infection structures called haustoria, which are known as the sites of secretion of effector proteins (Rafiqi *et al*., 2012; Petre & Kamoun 2014). As a consequence of their biotrophic life style, almost all rust fungi cannot be grown on synthetic media and genetic transformation is usually not possible. Therefore very little is known about the molecular mechanisms underlying the colonization of host tissues by rust fungi and so far, only six rust fungi effector proteins have been reported (see Petre *et al*., 2014 for review). Next generation sequencing technologies have provided access to genomes or transcriptomes for several rust fungi, providing the basis for bioinformatic secretome predictions (Duplessis *et al.,* 2014b, Sperschneider *et al.,* 2015). So far, genomes of five rust fungi have been published: the poplar rust *Melampsora larici-populina,* the wheat stem rust *Puccinia graminis* f. sp*. tritici,* the wheat stripe rust *Puccinia striiformis* f. sp. *tritici,* the flax rust *Melampsora lini* and the coffee rust *Hemileia vastatrix* (Duplessis *et al.,* 2011a; Cantu *et al.,* 2011; 2013; Zheng *et al.,* 2013; Nemri *et al.,* 2014; Cristancho *et al.,* 2014). Secretomes of rust fungi have been established depending on the presence of predicted N-terminal signal peptides in proteins (Cantu *et al*., 2011; 2013; Duplessis *et al*., 2011a; Fernandez *et al*., 2012; Hacquard *et al*., 2012; Saunders *et al*., 2012; Garnica *et al*., 2013; Zheng *et al*., 2013; Bruce *et al*., 2013; Link *et al*., 2014; Nemri *et al*., 2014). Signal peptides can be determined by predictors available online (Emmanuelsson *et al.*, 2007). These predictions have revealed the presence of a plethora of SPs in rust fungal species. A recent comparison of genomic features in 84 plant-associated fungi has shown that

the proteomes of obligate biotrophs are enriched in SPs, most of which with unknown function (Lo Presti *et al*., 2015). This illustrates the importance of studying rust secretomes to identify potential CSEPs.

The poplar leaf rust fungus *M. larici-populina* causes annual epidemics and severe damage in Northern European poplar plantations. Investigations of the poplar-poplar rust pathosystem using "-omic" approaches has led to major progress in the description of this interaction (Hacquard *et al.,* 2011). The genome of *M. larici-populina* was one of the first rust genomes sequenced by an international research consortium (Duplessis, *et al*., 2011a). *In sillico* genome annotation and secretome prediction have been instrumental to unravel *M. larici-populina* SPs. Among the 16,399 predicted protein-coding genes reported in the poplar rust genome, 13.3% are predicted SPs (2168 SPs) of which 89.3% are of unknown function (Figure 1A). Other secreted proteins correspond to carbohydrate active enzymes (5.8%), lipases (2.3%), proteases (0.8%) and other functions (1.8%) (Figure 1A). Extensive genomic and transcriptomic studies have also positioned *M. larici-populina* as a model tree pathogen for molecular investigations (Joly *et al*., 2010; Hacquard *et al*., 2010; Duplessis *et al*., 2011a; Duplessis *et al*., 2011b; Hacquard *et al*., 2012; Petre *et al*., 2012; Hacquard *et al*., 2013; Pernaci *et al*., 2014; Persoons *et al*., 2014).

From *M. larici-populina* SPs to CSEPs: post-genomic strategies

Two independent studies have defined different pipelines to pinpoint priority poplar rust CSEPs from catalogues of predicted SPs (Figure 2; Hacquard *et al*., 2012; Saunders *et al*., 2012). In both studies, *M. larici-populina* secretome was predicted using the same prediction tools. SignalP2.1 was used to sort SPs from the proteome, TargetP1.1 to identify proteins likely retained inside fungal cells (e.g. in mitochondria; Emanuelsson *et al.,* 2000) and TMHMM to exclude proteins carrying transmembrane α-helix domains (Moller *et al.,* 2001; Figure 2). Considering that rust fungal genomes exhibit expanded lineage-specific multigene families compared to other Basidiomycetes, one study used the similarity-based Markov clustering TribeMCL program to group SPs in tribes to further investigate multigene families in *M. larici-populina* and *P. graminis* f. sp. *tritici* (Enright *et al*., 2002; Saunders *et al*., 2012). The second study also took advantage of TribeMCL clustering but added another level of annotation with expert curation of *M. larici-populina* SP genes. This led to the definition of SP gene families (Duplessis *et al*., 2011a; Hacquard *et al.,* 2012). It is worth noting that these studies used different parameters and fungal species to perform the TribeMCL clustering. By doing so, the initial repertoire of *M. larici-populina* SPs differed between the two studies (Figure 2).

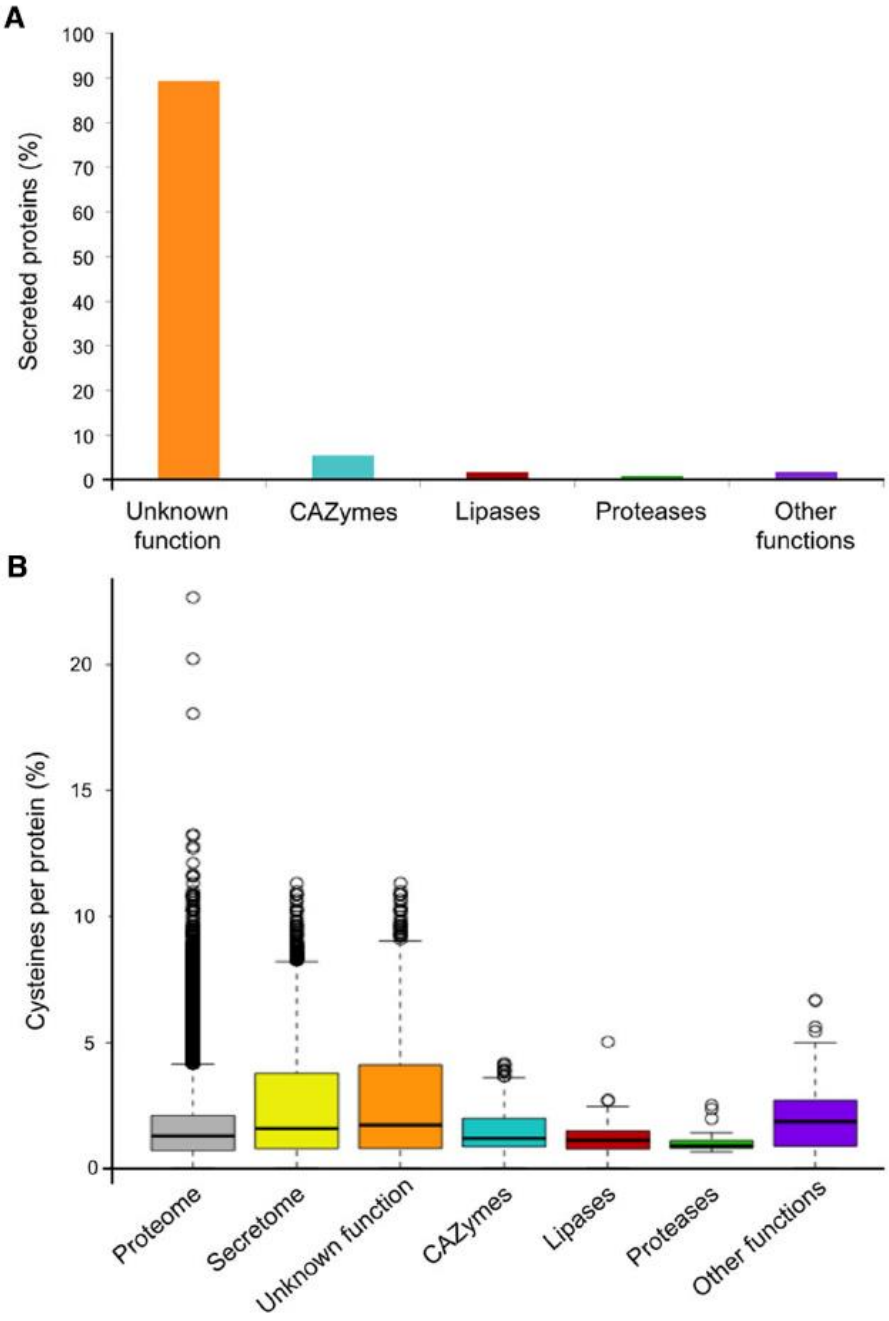


Figure 1: *M. larici-populina* predicted secretome and cysteine content.

(A) Composition of the M. larici-populina predicted secretome. Bars indicate the percentage of predicted secreted proteins (SPs): proteins of unknown function (orange), carbohydrate active enzymes (CAZymes; blue), lipases (red), proteases (green) and other functions (purple). The secretome prediction reported here has been updated based on the current version of the poplar rust genome available at the Joint Genome Institute Mycocosm (May 2015; http://genome.jgi-psf.org/programs/fungi). Annotation of SP functions was based on the expert curation reported in Duplessis et al., (2011a) and updated for CAZymes, Lipases and Proteases. (B) Cysteine content in *M. larici-populina* SPs. The percentage of cysteine residues per proteins are symbolized by box plots where the top and bottom of the boxes show the 25% and 75% quartiles, respectively, and the middle line indicates the median (50% quartile). The bottom whisker corresponds to the 1.5 interquartile range of the lower quartile and the top whisker indicates the 1.5 interquartile range of the upper quartile. White circles represent outliers.

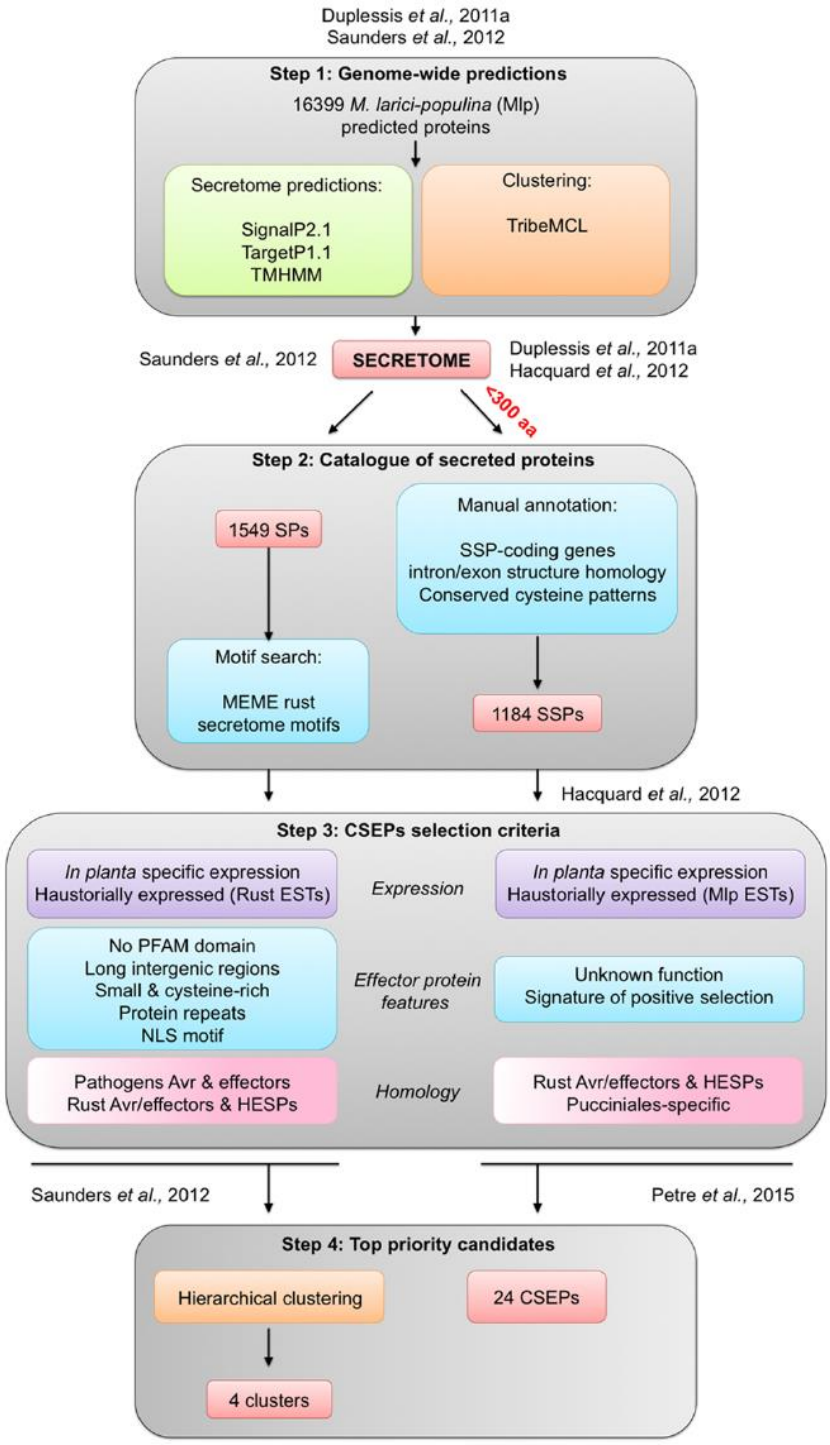


Figure 2: Overview of effector-mining pipelines applied to *M. larici-populina* secretome to prioritize candidate secreted effector proteins (CSEPs).

The pipelines can be divided in four main steps: step 1 *Genome-wide predictions* identifies the *M. larici-populina* secretome using prediction tools (green) and gene families clustering with TribeMCL (orange); step 2 *Catalogue of secreted proteins* establishes a set of secreted proteins; step 3 *CSEPs selection criteria* identifies CSEPs using different features and criteria; and step 4 *Top priority candidates* prioritizes CSEPs for further functional characterization. See text for details. Avr, avirulence protein; EST, expressed sequence tag; SP, secreted protein; SSP, small secreted protein; HESP, haustorially expressed secreted protein.

To face the plethora of predicted SPs in the poplar rust fungus and considering the high divergence and absence of conserved motifs in rust *bona fide* effectors, both effector-mining pipelines focused on *a priori* features of plant pathogen effectors (Hacquard *et al*., 2012; Saunders *et al*., 2012). Thus, criteria such as expression during infection or in purified haustoria were applied to prioritize candidates. Moreover, authors took advantage of other features such as the size of proteins, the content in cysteine residues, the presence of selection signatures (i.e. genes evolving under the pressure of host resistances), homology to known rust fungi effectors and/or previously reported haustorially expressed secreted proteins (HESPs), as well as organization in genes families taking into account specificity at a given taxonomical level (species, genera, family, order) (for review see Petre *et al.,* 2014).

Effector proteins are often described as small proteins (Stergiopoulos and de Wit, 2009; Tyler and Rouxel, 2012). Based on this observation, an arbitrary cut-off can be applied in CSEPs-mining studies to only focus on small secreted proteins (SSPs), although large effector proteins have also been reported (Petre *et al.,* 2014; Lo Presti *et al.,* 2015). The *Melampsora Genome Consortium* performed a manual curation of *M. larici-populina* SSP gene families (i.e. < 300 amino acids) taking advantage of expressed sequence tags (ESTs) from haustoria and rust-infected poplar leaves (Joly *et al*., 2010; Duplessis *et al.,* 2011a; Figure 2). Dedicated expert annotation led to the elimination and the addition of several genes encoding SSPs, and generated an additional 170 SSPs that were not predicted by automatic annotation (Duplessis *et al.,* 2011a). Manually annotated SSPs were particularly enriched in cysteine residues (on average 2.8% cysteines per protein compared to 1.6% in the whole proteome; Figure 1B). SSPs of unknown function are also clearly enriched in cysteine residues compared with annotated SPs (Figure 1B). The proportion of cysteine residues in effectors can indicate the presence of intra-molecular disulfide bridges that could help in stabilizing the protein structure in the inhospitable apoplastic environment (Stergiopoulos *et al*., 2013). For instance, the *Cladosporium fulvum* Avr2 fungal effector is a cysteine-rich protein, playing a role in apoplastic protease inhibition during the interaction with tomato leaf cells (Rooney *et al*., 2005). Although, it cannot be excluded that cysteine residues could play an important role in the structure of the proteins at their final destination in the host cell. Detailed analysis of SSP gene families taking into account intron/exon organization and cysteine codon positions have revealed conserved cysteine patterns (Hacquard *et al*., 2012). For instance, the largest SSP family composed of 111 members shares the conserved YxC//CxxY//YxC cysteine pattern (Duplessis *et al*., 2011a; Hacquard *et al*., 2012). This pattern is reminiscent of the Y/F/WxC motif reported in powdery mildew and wheat rust fungi (Godfrey *et al*., 2010). Interestingly, other obligate biotrophs also present large repertoires of SPs, such as the white rust oomycete *Albugo laibachii* in which

CHXC and CXHC cysteine-rich motifs are found (Kemen *et al.*, 2011). Both types of patterns were speculated to be potentially involved in delivery of effector into host cells.

In the study conducted by Hacquard *et al.* (2012), many features of effectors were considered in order to facilitate the selection of CSEPs of utmost interest (Figure 2). The following criteria were specifically considered: specific expression in infected host tissues, unknown function, homology to known rust effectors and HESPs, specificity to the Pucciniales order and signatures of positive selection (Hacquard *et al*., 2012). Effector-mining studies have frequently used evidence of expression during interaction with the host as a filter to identify critical CSEPs (Lo Presti *et al.,* 2015). The extensive knowledge of *M. larici-populina* SSP expression during host-infection and at other stages of its life cycle has been instrumental to further prioritize CSEPs (for detailed reviews see Duplessis *et al.,* 2012 and Duplessis *et al.,* 2014b). Time-course of poplar leaf infection by *M. larici-populina* has revealed dynamic patterns of SSP expression during early stages of infection, biotrophic growth and sporulation (Duplessis *et al.,* 2011b; Hacquard *et al.,* 2012). Expression in resting and germinating spores can be used to discriminate SSP genes specifically expressed *in planta*. The study performed by Hacquard *et al.* (2012) did not provide a defined list of CSEPs to be explored in priority but a comprehensive depiction of the complete repertoire of *M. larici-populina* SSP genes.

Manual curation of large fungal genomes such as rust fungi remains a time-consuming process and automatized pipelines could help to foster CSEP detection. The pipeline built by Saunders and collaborators was initially designed to scrutinize the secretomes of *M. larici-populina* and *P. graminis* f. sp. *tritici* (Saunders et al. 2012). It has been subsequently applied to mine the genomes of different fungi interacting with plants, including the wheat rust *P. striiformis* f.sp. *tritici* and the flax rust *M. lini* (Cantu *et al.,* 2013; Nemri *et al.,* 2014, Lin *et al.,* 2014). This *in sillico* pipeline computes Markov clustering generated tribes from available genome annotations (Saunders *et al.,* 2012). The pipeline considers three levels of information for SPs: functional annotation, detection of novel effector motifs and annotation of effector features (Figure 2). Most effectors do not have PFAM domains (Stergiopoulos & de Wit, 2009; Kamoun, 2007; Dodds *et al.,* 2009). The functional annotation step allows the selection of SPs with no conserved protein domain families (PFAM), with the exception of avirulence proteins that may have such domains. For instance, the Chitin Binding Module-like of Avr4 and the LysM domain of Ecp6 in *C. fulvum* both have PFAM annotations. In total, five PFAM domains were found in rust fungi and were considered for their obvious connection with pathogenicity (Saunders *et al.,* 2012). In a second step, the MEME tool was applied to detect *de novo* conserved patterns in rust SPs (Bailey *et al.,* 2009). Among identified motifs, five motifs containing one or two conserved cysteine residues with high positional constraints in SP tribes were highlighted (Figure 2; Saunders *et al.,* 2012). Interestingly, some motifs such as the motif 06

YxCxYxxCxW, were also identified by the manual annotation of *M. larici-populina* SSP families (Duplessis *et al.,* 2011a; Hacquard *et al.,* 2012). In a last step, usual features of effectors were scrutinized, including induction of expression during host infection, similarity to haustorial ESTs, small protein size (< 150 amino acids), content in cysteine residues and known effector motifs or repeats in protein. It has been reported that some effectors contain nuclear localization signals (NLS), suggestive of a potential nuclear localization in host cells (Liu *et al.,* 2011; Schornack *et al.,* 2010; Kanneganti *et al.,* 2007; Shen *et al.,* 2007). The presence of such NLS was also added to the features tested. It has been shown for different filamentous plant pathogens such as the fungus *Leptosphaeria maculans* and the oomycete *Phytophthora infestans* that effector genes reside in gene scarse regions marked by the presence of repeat elements such as transposable elements (Haas *et al.,* 2009; Rouxel *et al.,* 2011; for a review see Raffaele and Kamoun, 2012). This criterion was also taken into account in the pipeline by looking at the presence of long intergenic regions around SP genes. However, no significant link could be established between SSP genes and repeat-rich regions in the genomes *M. larici-populina* and *P. graminis* f. sp. *tritici* (Duplessis *et al.,* 2011a). Nonetheless, all these filters are informative and can be computed in a complex matrix comparing rust tribes in order to rank CSEP (Figure 2; Saunders *et al.,* 2012).

A priori criteria to prioritized CSEPs

Although typical features could be inferred from effector proteins, it is rather rare that a given effector will combine all these features at once. On the contrary, each feature will display a diverse distribution among SP families. Thereby, hierarchical clustering can be performed to enable ranking of tribes with the highest probability of containing CSEPs (Saunders *et al.,* 2012). In such a clustering approach, it is necessary to adjust the weight of each criterion. By doing so, Saunders and collaborators were able to derive four clusters with the most promising SP tribes that could correspond to CSEPs for further investigations. The largest tribe with 92 members in one of these clusters is specific to *M. larici-populina* and contains a large proportion of secreted proteins (73% with predicted signal peptide; Saunders *et al*., 2012). This tribe corresponds to the largest poplar rust SSP family (with 111 members) reported in Duplessis *et al*. (2011a). This SSP family is marked by the presence of highly conserved cysteine patterns highlighted in both studies. The difference in numbers of SSP members likely comes from the different levels of gene annotation considered for the *M. larici-populina* genome. The two studies identified tribes composed of SPs/SSPs and proteins without predicted signal peptides. In some tribes, SPs exhibited homology to HESPs, as well as known rust effector proteins (e.g. *M. lini* ArvM). It could be speculated that such proteins may be involved in haustoria functions (Saunders *et al.,* 2012). It could also reflect the evolution of

these families with gain or loss of the signal peptide towards the generation of new putative effector functions.

In a recent study, an effectoromic pipeline identified priority *M. larici-populina* CSEPs for expression in *Nicotiana benthamiana* as a heterologous system to study their localization in plant subcellular compartments and to identify potential plant interactors (Petre *et al.,* 2015a). The pipeline was applied to the catalogue of SSPs previously reported by Hacquard *et al.* (2012). Priority CSEPs were selected by giving a stronger weight to some of the typical criteria used in the two studies reported above. For instance, expression in haustoria, specific induction of expression during poplar infection, specificity to the Pucciniales and proteins of unknown functions were the most important features considered. These criteria were similar to those systematically applied for tribes ranking in Saunders *et al.* (2012). Redundant family members were also removed in order to focus on orphan and lineage-specific CSEPs, considering that pathogenicity mechanisms imply highly specific functions. Such stringent criteria led to a subset of 24 priority CSEPs from 1184 initial *M. larici-populina* SSPs (Petre *et al.,* 2015a). Among these, only three belong to priority clusters identified in the study by Saunders *et al.* (2012), in which an initial postulate was to focus on tribes and not on orphan genes. One supposed orphan SSP (CSEP 107772) finally proved to be member of a *M. larici-populina* multigene family (Petre *et al.,* 2015b). This particular example illustrates how automatic and dedicated pipelines strongly depend on the accuracy of genome annotation and parameters applied to gene family analysis tools, above any further selection criteria. Among the 24 selected CSEPs, 20 could be expressed in fusion with GFP in *N. benthamiana*, which further allows identification of specific localisation in plant cell compartments as well as potential plant interactors through coimmunoprecipitation and mass spectrometry. This study identified six *M. larici-populina* CSEPs with a specific localization pattern (i.e. nucleus, nucleolus, chloroplast, mitochondria and cytosolic bodies) and five that specifically associated with plant proteins representing potential interactors (Petre *et al.,* 2015a). The development of *in planta* assays in heterologous systems has provided the first step towards effector characterization for various pathogens (Bozkurt *et al.,* 2011; Caillaud *et al.,* 2012; Caillaud *et al.,* 2013; Petre et al., 2015a; Petre et al., 2015b). Another alternative already proven successful for different obligate biotrophic fungi, including rust fungi, is host-induced gene silencing (Panwar *et al.*, 2013; Spanu 2015). Knowledge of effector structure is useful to understand effector/plant protein interactions and to find structure homology within effectors (Maqbool *et al.,* 2015). *In vitro* structural biology of effectors is also an option of choice to further determine the role of effectors in rust fungi, pending fruitful production of recombinant CSEPs.

Pucciniales genomes are marked by the presence of very large catalogues of SPs. Among these only a fraction may be *bona fide* effectors. Although still imperfect, effector-mining pipelines are crucial to highlight top priority CSEPs in these fungal species. No robust ubiquitous effector motifs are found in rust fungal effectors, contrary to the RXLR motif of oomycetes effectors (Sperschneider *et al.,* 2015; Petre and Kamoun, 2014). As illustrated here, effector mining in the poplar rust fungus relies both on the quality of input data (i.e. gene annotation and gene families analysis) and on the several qualitative but also subjective criteria retained. Indeed, there is no absolute rule for ranking and weighting among the many effector features that could be taken into account. As illustrated here, knowledge about SP gene expression is a key criterion to select priority CSEPs. Considering the complex life cycle of heteroecious rust fungi like *M. larici-populina* (i.e. alternation on two different hosts and production of five different spores), a better insight of gene expression patterns along the whole life cycle can help to drastically reduce the catalogue of CSEPs. Pucciniales consist of more than 8000 species (Aime *et al.,* 2014). So far, less than 10 genomes have been published or sequenced with partial data available (see Duplessis *et al.,* 2014b for review). All rust fungal genomics and transcriptomics reports have shown that these species contain a high content of Pucciniales or species-specific genes. It is much likely that rust fungi possess highly specific effectors and increasing the amount of genomic data will surely help in focusing the efforts towards CSEPs identification. The development of dedicated bioinformatics tools for the prediction of fungal effectors is still much awaited (Sperschneider *et al.,* 2015).

Authors and contributors:

All the authors wrote and revised the manuscript.

Acknowledgments:

The authors warmly thank Emmanuelle Morin (INRA Nancy) for her continuous support with *M. larici-populina* genomics. Cécile Lorrain is supported by INRA, in the framework of a Contrat Jeune Scientifique and by the Labex ARBRE. Sébastien Duplessis acknowledges the support of the French ANR for the young scientist grant POPRUST (ANR-2010-JCJC-1709-01) and of the Région Lorraine. All authors benefit from the support of the ANR in the frame of a grant part of the “Investissements d’Avenir” program (ANR-11-LABX-0002-01, Lab of Excellence ARBRE).

Conflict of Interest Statement:

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References:

Aime, M. C., Merje, T., and McLaughlin, D J., (2014). 10 Pucciniomycotina. In Systematics and Evolution, edited by David J. McLaughlin and Joseph W. Spatafora, 7A: 271–94. *The Mycota.* Springer Berlin Heidelberg. http://dx.doi.org/10.1007/978-3-642-55318-9_10.

Bailey, T L., Boden, M., Buske, F A., Frith, M., Grant, C E., Clementi, *et al.*, (2009). MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Research* 37 (Web Server issue): W202–8. doi:10.1093/nar/gkp335.

Bozkurt T. O., Schornack S., Win J., Shindo T., Ilyas M., Oliva R., Cano L. M., Jones A. M., Huitema E., van der Hoorn R. A., Kamoun S. (2011) *Phytophthora infestans* effector AVRblb2 prevents secretion of a plant immune protease at the haustorial interface, *Processings of the National Academy of Science U.S.A.,* 108:20832-20837.

Caillaud M. C. Piquerez S. J. Fabro, G. Steinbrenner J., Ishaque N., Beynon J., Jones J. D. (2012) Subcellular localization of the Hpa RxLR effector repertoire identifies a tonoplast-associated protein HaRxL17 that confers enhanced plant susceptibility, *The Plant Journal*, 69(2), 252-265.

Caillaud M. C., Asai S., Rallapalli G., Piquerez S., Fabro G., Jones J. D. (2013) A downy mildew effector attenuates salicylic Acid-triggered immunity in Arabidopsis by interacting with the host mediator complex, *PLoS Biology*, 11(12).

Bruce, M., Neugebauer, K.A., Joly, D.L., Migeon, P., Cuomo, C. A., *et al*. (2013). Using transcription of six *Puccinia triticina* races to identify the effective secretome during infection of wheat. *Frontiers in Plant Science* 4: 520. doi:10.3389/fpls.2013.00520.

Cantu, D., Govindarajulu, M., Kozik, A., Wang, M., Chen, X., Kojima, K.K., *et al.* (2011). Next generation sequencing provides rapid access to the genome of *Puccinia striiformis f. sp. tritici*, the causal agent of wheat stripe rust. Edited by Steven Harris. *PLoS ONE* 6 (8): e24230. doi:10.1371/journal.pone.0024230.

Cantu, D., Segovia, V., MacLean, D., Bayles, R., Chen, X., Kamoun, S., *et al.*, (2013). Genome Analyses of the Wheat Yellow (stripe) Rust Pathogen *Puccinia Striiformis f. sp. tritici* Reveal Polymorphic and Haustorial Expressed Secreted Proteins as Candidate Effectors. *BMC Genomics* 14. doi:10.1186/1471-2164-14-270.

Cristancho, M. A., Botero-Rozo, D. O., Giraldo, W., Tabima, J., Riaño-Pachón, D. M., Escobar, C., *et al.,* (2014). Annotation of a hybrid partial genome of the coffee rust (*Hemileia vastatrix*) contributes to the gene repertoire catalog of the Pucciniales. *Frontiers in Plant Science*, 5, 594. doi:10.3389/fpls.2014.00594

Dean, R., Van Kan, J. A. L., Pretorius, Z. A., Hammond-Kosack, K. E., Di Pietro, A., *et al*. (2012). The top 10 fungal pathogens in molecular plant pathology. *Molecular Plant Pathology* 13 (4): 414–30. doi:10.1111/j.1364-3703.2011.00783.x.

Dodds, P.N., Rafiqi, M., Gan, P. H. P., Hardham, A. R., Jones, D. A., and Ellis, J. G. (2009). Effectors of biotrophic fungi and oomycetes: pathogenicity factors and triggers of host resistance. *New Phytologist* 183 (4): 993–1000. doi:10.1111/j.1469-8137.2009.02922.x.

Duplessis, S., Cuomo, C. A., Lin, Y., Aerts, A., Tisserant, E., Veneault-Fourrey, C., *et al*. (2011a). Obligate biotrophy features unraveled by the genomic analysis of rust fungi. *Proceedings of the National Academy of Sciences of the United States of America* 108 (22): 9166–71. doi:10.1073/pnas.1019315108.

Duplessis, S., Hacquard, S., Delaruelle, C., Tisserant, E., Frey, P., *et al*. (2011b). *Melampsora larici-populina* transcript profiling during germination and timecourse infection of poplar leaves reveals dynamic expression patterns associated with virulence and biotrophy. *Molecular Plant-Microbe Interactions : MPMI* 24 (7): 808–18. doi:10.1094/MPMI-01-11-0006.

Duplessis, S., Joly, D. J., and Dodds, P. N. (2012). “Rust effectors,” in Effectors in Plant- Microbes Interactions, eds F. Martin and S. Kamoun (Oxford: Wiley-Blackwell), 155–193.

Duplessis, S., Spanu, P. D., and Schirawski, J. (2014a). “Biotrophic fungi (powdery mildews, Rusts and Smuts),” in Ecological Genomics of the Fungi. Plant-Interacting Fungi Section, ed F. Martin (Hoboken, NJ: Wiley-Blackwell), 149–168.

Duplessis, S., Bakkeren, G., and Hamelin, R. (2014b). Advancing knowledge on biology of rust fungi through genomics. *Adv Bot. Res*. 70, 173–209. doi: 10.1016/B978-0-12-397940-7.00006-9

Emanuelsson, O., Brunak, S., von Heijne, G., and Nielsen, H. (2007). locating proteins in the cell using TargetP, SignalP and related tools. *Nature Protocols* 2 (4): 953–71. doi:10.1038/nprot.2007.131.

Emanuelsson, O., H. Nielsen, S. Brunak, and G. von Heijne. (2000). Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *Journal of Molecular Biology* 300 (4): 1005–16. doi:10.1006/jmbi.2000.3903.

Enright, A. J., S. Van Dongen, and C. A. Ouzounis. (2002). An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Research* 30 (7): 1575–84.

Fernandez, D., Tisserant, E., Talhinhas, P., Azinheira, H., Vieira, A., Petitot, A-S., Loureiro, A. *et al*. (2012). 454-pyrosequencing of coffea arabica leaves infected by the rust fungus *Hemileia vastatrix* reveals in planta-expressed pathogen-secreted proteins and plant functions in a late compatible plant–rust interaction. *Molecular Plant Pathology* 13 (1): 17–37. doi:10.1111/j.1364-3703.2011.00723.x.

Garnica, D.P., Upadhyaya, N. M., Dodds, P. N., and Rathjen. J. P. (2013). Strategies for wheat stripe rust pathogenicity identified by transcriptome sequencing. *PloS One* 8 (6). doi:10.1371/journal.pone.0067150.

Godfrey, D., Böhlenius, H., Pedersen, C., Zhang, Z., Emmersen, J. and Thordal-Christensen, H. (2010). Powdery mildew fungal effector candidates share N-terminal Y/F/WxC-motif. *BMC Genomics* 11: 317–317. doi:10.1186/1471-2164-11-317.

Haas, B.J., Kamoun, S., Zody, M.C., Jiang, R. H. Y., Handsaker, R. E., Cano, L. M., Grabherr, M. *et al*. (2009). Genome sequence and analysis of the irish potato famine pathogen *Phytophthora infestans*. *Nature* 461 (7262): 393–98. doi:10.1038/nature08358.

Hacquard, S., Delaruelle, C., Frey, P., Tisserant, E., Kohler, A., and Duplessis, S. (2013). Transcriptome analysis of poplar rust telia reveals overwintering adaptation and tightly coordinated karyogamy and meiosis processes. *Frontiers in Plant Science* 4: 456. doi:10.3389/fpls.2013.00456.

Hacquard, S., Petre, B., Frey, P., Hecker, A., Rouhier, N., & Duplessis, S. (2011). The Poplar-Poplar Rust Interaction: Insights from Genomics and Transcriptomics. *Journal of Pathogens*, 716041. doi:10.4061/2011/716041

Hacquard, S., Delaruelle, C., Legue, V., Tisserant, E., Kohler, A., Frey, P., *et al*. (2010). Laser Capture microdissection of uredinia formed by *Melampsora larici-populina* revealed a transcriptional switch between biotrophy and sporulation. *MPMI* 23 (10): 1275–86. doi:10.1094/MPMI-05-10-0111.

Hacquard, S., Joly, D. L., Lin, Y-C., Tisserant, E., Feau, N., Delaruelle C., Legue, V. *et al*. (2012). A comprehensive analysis of genes encoding small secreted proteins identifies candidate effectors in *Melampsora larici-populina* (poplar leaf rust). *MPMI* 25 (3): 279–93. doi:10.1094/MPMI-09-11-0238.

Joly, D. L., Feau, N., Tanguay, P. and Hamelin, R. C. (2010). Comparative analysis of secreted protein evolution using expressed sequence tags from four poplar leaf rusts (*Melampsora spp*.). *BMC Genomics* 11: 422–422. doi:10.1186/1471-2164-11-422.

Jones J. D. G., Dangl J. L. (2006). The Plant immune system. *Nature Reviews*. 444-16:323-329. doi:10.1038/nature05286.

Kamoun, S. (2007). Groovy times: filamentous pathogen effectors revealed. *Current Opinion in Plant Biology* 10 (4): 358–65. doi:10.1016/j.pbi.2007.04.017.

Kamoun, S. (2009). The secretome of plant-associated fungi and oomycetes. *In Plant Relationships*, edited by HolgerB. Deising, 5:173–80. The Mycota. Springer Berlin Heidelberg. http://dx.doi.org/10.1007/978-3-540-87407-2_9.

Kanneganti, T-D., Bai, X., Tsai, C-W., Win, J., Meulia, T., Goodin, M., *et al*. (2007). A functional genetic assay for nuclear trafficking in plants. *The Plant Journal : For Cell and Molecular Biology* 50 (1): 149–58. doi:10.1111/j.1365-313X.2007.03029.x.

Kemen A. C., Alger M. T., Kemen E. (2015). Host-microbe and microbe-microbe interactions in the evolution of obligate plant parasitism. *New phytologist*. 206:1207-1228. doi:0.1111/nph.13284

Lin K., Limpens E., Zhang Z., Ivanov S., Saunders D.G.O., *et al.,* (2014) Single Nucleus Genome Sequencing Reveals High Similarity among Nuclei of an Endomycorrhizal Fungus. *PLoS Genet* 10(1): e1004078. doi: 10.1371/journal.pgen.1004078.

Link, T. I., Lang, P., Scheffler, B. E., Duke, M. V., Graham, M. A., Cooper, B., Tucker, M. L. *et al*. (2014). The haustorial transcriptomes of *Uromyces appendiculatus* and *Phakopsora pachyrhizi* and their candidate effector families. *Molecular Plant Pathology* 15 (4): 379–93. doi:10.1111/mpp.12099.

Liu, T., Ye, W., Ru, Y., Yang, X., Gu, B., Tao, K., Lu, S., *et al*. (2011). Two host cytoplasmic effectors are required for pathogenesis of *Phytophthora sojae* by suppression of host defenses. *Plant Physiology* 155 (1). doi:10.1104/pp.110.166470.

Lo Presti, L., Lanver, D., Schweizer, G., Tanaka, S., Liang L., Tollot, M., *et al*. (2015). Fungal effectors and plant susceptibility. *Annual Review of Plant Biology* 66 (April): 513–45. doi:10.1146/annurev-arplant-043014-114623.

MacLean, A.M., Orlovskis Z., Kowitwanich K., Zdziarska A.M., Angenent G.C., *et al*. (2014) Phytoplasma Effector SAP54 Hijacks Plant Reproduction by Degrading MADS-box Proteins and Promotes Insect Colonization in a RAD23-Dependent Manner. *PLoS Biol* 12(4): e1001835. doi: 10.1371/journal.pbio.1001835

Maqbool, A., Saitoh, H., Franceshetti, M., Stevenson, C.E.M., Uremura, A., Kanzaki, H., Terauchi, R., Banfield, M.J. (2015) Structural basis of pathogen recognition by an integrated HMA domain in a plant NLR immune receptor. *eLife*. 4:e08709. DOI: 10.7554/eLife.08709.

Moller, S., Croning, M.D.R., Apweiler, R., (2001) Evaluation of methods for the prediction of membrane spanning regions. *Bioinformatics*, 17(7):646-653

Nemri, A., Saunders, D. G. O., Anderson, C., Upadhyaya, N. M., Win, J., Lawrence, G., *et al*. (2014). The genome sequence and effector complement of the flax rust pathogen Melampsora *lini*. *Frontiers in Plant Science* 5. doi:10.3389/fpls.2014.00098.

Okmen, B. and Doehlemann, G. (2014). Inside plant: biotrophic strategies to modulate host immunity and metabolism. *Current Opinion in Plant Biology* 20. doi: 10.1016/j.pbi. 2014.03.011.

Panwar, V., McCallum, B., Bakkeren, G. (2013) Host-induced gene silencing of wheat leaf rust fungus *Puccinia triticina* pathogenicity genes mediated by the Barley stripe mosaic virus. *Plant Mol Biol* 81(6):595-608. doi: 10.1007/s11103-013-0022-7.

Pernaci, M., De Mita, S., Andrieux, A., Pétrowski, J., Halkett, F., Duplessis, S. and Frey, P. (2014). Genome-wide patterns of segregation and linkage disequilibrium: the construction of a linkage genetic map of the poplar rust fungus *Melampsora larici-populina*. *Frontiers in Plant Science* 5: 454. doi:10.3389/fpls.2014.00454.

Persoons, A., Morin, E., Delaruelle, C., Payen, T., Halkett, F., Frey, P., De Mita, S. and Duplessis, S. (2014). Patterns of genomic variation in the poplar rust fungus *Melampsora larici-populina* identify pathogenesis-related factors. *Frontiers in Plant Science* 5: 450. doi:10.3389/fpls.2014.00450.

Petre, B., Joly, D. L. and Duplessis S. (2014). Effector proteins of rust fungi. *Frontiers in Plant Science* 5: 416. doi:10.3389/fpls.2014.00416.

Petre, B., and Kamoun, S. (2014). How do filamentous pathogens deliver effector proteins into plant cells? *PLoS Biol* 12 (2): e1001801. doi:10.1371/journal.pbio.1001801.

Petre, B., Morin, E., Tisserant, E., Hacquard, S., Da Silva, C., Poulain, J., Delaruelle, C. et al. (2012). RNA-seq of early-infected poplar leaves by the rust pathogen *Melampsora larici-populina* uncovers PTSULTR3;5, a fungal-induced host sulfate transporter. *PloS One* 7 (8). doi:10.1371/journal.pone.0044408.

Petre, B., Saunders, D.G.O., Sklenar, J., Lorrain, C., Win, J., Duplessis, S., and Kamoun S., (2015a). Candidate effector proteins of the rust pathogen *Melampsora larici-populina* target diverse plant cell compartments. *MPMI*, doi:10.1094/MPMI-01-15-0003-R.

Petre, B., Lorrain, C., Saunders, D.G.O., Win, J., Sklenar, J., Duplessis, S., and Kamoun S., (2015b). Rust fungal effectors mimic host transit peptides to translocate into chloroplasts. *Cellular Microbiology*, doi: 10.1111/cmi.12530

Pusztahelyi, T., Hold, I.J., Pósci, I. (2015). Secondary metabolites in fungus-plant interactions. *Frontiers in Plant Science*, doi: 10.3389/fpls.2015.00573.

Raffaele, S., and Kamoun S. (2012). Genome evolution in filamentous plant pathogens: why bigger can be better. *Nature Reviews*. *Microbiology* 10 (6): 417–30. doi:10.1038/nrmicro2790.

Rafiqi, M., Ellis, J. G., Ludowici, V.A., Hardham, A. R. and Dodds, P.N. (2012). Challenges and progress towards understanding the role of effectors in plant-fungal interactions. *Current Opinion in Plant Biology* 15 (4): 477–82. doi:10.1016/j.pbi.2012.05.003.

Rooney, H. C. E., Van't Klooster, J. W., van der Hoorn, R. A. L., Joosten, M. H. A. J., Jones, J. D. G. and de Wit, P. J. G. M. (2005). *Cladosporium* Avr2 inhibits tomato rcr3 protease required for cf-2-dependent disease resistance. *Science (New York, N.Y.)* 308 (5729): 1783–86. doi:10.1126/science.1111404.

Rouxel, T., Grandaubert, J., Hane, J. K., Hoede, C., van de Wouw, A. P., Couloux, A., *et al.,* (2011). Effector diversification within compartments of the Leptosphaeria maculans genome affected by Repeat-Induced Point mutations. *Nature Communications*, 2, 202–. doi:10.1038/ncomms1189

Rovenich, H., Boshoven, J. C. and Thomma, B. P. H. J. (2014). Filamentous pathogen effector functions: of pathogens, hosts and microbiomes. *Current Opinion in Plant Biology* 20 (August). doi:10.1016/j.pbi.2014.05.001.

Saunders, D. G. O., Win, J., Cano, L. M., Szabo, L. J., Kamoun, S. and Raffaele, S. (2012). Using hierarchical clustering of secreted protein families to classify and rank candidate effectors of rust fungi. *PloS One* 7 (1). doi:10.1371/journal.pone.0029847.

Schornack, S., van Damme, M., Bozkurt, T. O., Cano, L.M., Smoker, M., Thines, M., Gaulin, E. *et al*. (2010). Ancient class of translocated oomycete effectors targets the host nucleus. *PNAS*. 107 (40): 17421–26.

Shen, Q-H., Saijo, Y., Mauch, S., Biskup, C., Bieri, S., Keller, B. *et al*. (2007). Nuclear Activity of MLA Immune receptors links isolate-specific and basal disease-resistance responses. *Science (New York, N.Y.)* 315 (5815): 1098–1103. doi:10.1126/science.1136372.

Spanu, P.D. (2015) RNA-protein interactions in plant disease: hackers at the dinner table. *New Phytol* 207(4):991-995. doi: 10.1111/nph.13495.

Sperschneider, J., Dodds, P. N., Gardiner, D. M., Manners, J. M., Singh, K. B. and Taylor, J. M. (2015). Advances and Challenges in Computational Prediction of Effectors from Plant Pathogenic Fungi. *PLoS Pathogens*, 11(5), e1004806. doi:10.1371/journal.ppat.1004806

Stergiopoulos, I., Collemare, J., Mehrabi, R. and De Wit, P. J. G. M. (2013). Phytotoxic secondary metabolites and peptides produced by plant pathogenic dothideomycete fungi. *FEMS Microbiology Reviews* 37 (1). doi:10.1111/j.1574-6976.2012.00349.x.

Stergiopoulos, I., and De Wit, P. J. G. M. (2009). Fungal effector proteins. *Annual Review of Phytopathology* 47: 233–63. doi:10.1146/annurev.phyto.112408.132637.

Tyler, B. M. and Rouxel, T., (2012) Effectors of Fungi and Oomycetes: Their Virulence and Avirulence Functions and Translocation From Pathogen to Host Cells, in Molecular Plant Immunity (ed G. Sessa), Wiley-Blackwell, Oxford, UK. doi: 10.1002/9781118481431.ch7

Win, J., A. Chaparro-Garcia, K. Belhaj, D. G. O. Saunders, K. Yoshida, S. Dong, S. Schornack, *et al*. (2012). Effector biology of plant-associated organisms: concepts and perspectives. *Cold Spring Harbor Symposia on Quantitative Biology* 77: 235–47. doi:10.1101/sqb.2012.77.015933.

Zheng, W., Huang, L., Huang, J., Wang, X., Chen, X., Zhao, J., Guo, J. *et al*. (2013). High genome heterozygosity and endemic genetic recombination in the wheat stripe rust fungus. *Nature Communications* 4: 2673. doi:10.1038/ncomms3673.

Partie 4: Contexte et objectifs de la thèse

Lors de mon arrivée en thèse au sein de l'Unité Mixte de Recherche 1136 Interactions Arbres/Micro-organismes (UMR 1136 IAM), les avancées réalisées sur le plan moléculaire dans l'analyse de l'interaction entre le peuplier et *M. larici-populina* l'agent de la rouille foliaire du peuplier ont principalement porté sur l'infection de feuilles de peupliers, c'est à dire la phase de multiplication clonale du champignon qui correspond aux dégâts observés en plantations (Joly *et al.,* 2010 ; Duplessis *et al.,* 2011a ; 2011b ; Hacquard *et al.,* 2010, 2012 ; Petre *et al*., 2012). L'analyse du sécrétome à travers des approches de génomique et de transcriptomique a révélé des vagues d'expression successives de gènes codant des petites protéines sécrétées représentant de potentiels effecteurs candidats du champignon (Duplessis *et al.,* 2011b ; Hacquard *et al.,* 2012 ; Petre *et al.,* 2012). Des approches récentes de génomique ont été développées dans l'UMR 1136 IAM dans le cadre d'une collaboration avec le JGI et notamment au cœur de deux projets de thèses récents qui ont porté sur des analyses populationnelles et de cartographie génétique (Pernaci, 2015 ; Persoons, 2015). Le cycle de vie de *M. larici-populina* a pu être maitrisé en conditions semi-contrôlées, notamment pour la phase de reproduction sexuée chez le mélèze, qui n'était jusqu'alors pas accessible. Il a été possible d'établir une carte génétique du champignon à partir du re-séquençage par Illumina de 95 descendants d'un autocroisement sur mélèze de l'isolat 98AG31 dont le génome de référence est disponible au JGI (Pernaci *et al.,* non publié). Un nouvel assemblage des données de séquençage Sanger de *M. larici-populina* a été réalisé par le JGI et a été ancré sur la carte génétique à disposition, conduisant à une version 2 du génome du champignon ancrée sur 18 groupes de liaison (données non publiées, version 2 du génome disponible dans la base de données génomique *Mycocosm* du JGI). Dans le même temps, le JGI a réalisé en collaboration avec l'UMR 1136 IAM le séquençage du génome du champignon *Melampsora allii-populina*, un autre agent de la rouille foliaire sur peuplier, afin d'en réaliser l'analyse comparative avec le génome de *M. larici-populina* déjà à disposition. Par ailleurs, suite à l'identification d'effecteurs candidats de *M. larici-populina* l'UMR 1136 IAM a débuté leur analyse fonctionnelle dans le cadre du CJS du Dr. Benjamin Pêtre (2012-2014) et d'une collaboration établie avec le groupe de Sophien Kamoun au sein du *Sainsbury Laboratory* (TSL, Norwich, Royaume-Uni). J'ai d'ailleurs pu bénéficier d'un soutien du Labex ARBRE à Nancy pour effectuer un séjour de 6 mois dans le cadre de mon Master 2 en amont de ma thèse afin de participer à l'analyse fonctionnelle des effecteurs candidats aux côtés du Dr. Benjamin Pêtre. Une approche effectoromique a été développée au TSL afin d'exprimer des effecteurs candidats de rouille fusionnés à un marqueur fluorescent de manière transitoire par agroinfiltration dans des feuilles de *Nicotiana benthamiana*. A partir d'un panel initial de 20

candidats sélectionnés, cette approche a montré d'une part la localisation particulière de six effecteurs candidats dans les cellules de *N. benthamiana* par imagerie confocale *in vivo*. D'autre part, des protéines végétales cibles interagissant avec cinq effecteurs candidats ont pu être mise à jour par co-immunoprécipitation couplée à de la spectrométrie de masse (Petre *et al.,* 2015). Nous disposions ainsi d'une liste réduite d'effecteurs protéiques candidats prometteurs de *M. larici-populina* dont il restait à réaliser l'analyse structurale et fonctionnelle.

Le projet de thèse qui m'a été confié a porté sur deux axes découlant directement des travaux susmentionnés. D'une part, il s'est agit pour moi d'exploiter des données de génomique nouvellement acquises chez les deux *Melampsora* spp. responsables de maladies de rouille foliaire chez le peuplier. D'autre part, j'ai débuté l'analyse fonctionnelle de deux effecteurs candidats sélectionnés à l'issue de mon travail de Master.

Dans le premier axe de ma thèse, je me suis intéressée à l'analyse des données génomiques et transcriptomiques qui étaient à ma disposition à mon arrivée dans l'UMR 1136 IAM. Après avoir pris en main ces données génomiques, réalisé une nouvelle analyse du sécrétome avec les outils à disposition depuis la version du génome publiée en 2011 (Lorrain *et al.,* 2015) et m'être formée au langage R, j'ai réalisé une analyse du transcriptome de *M. larici-populina* au cours de la phase de reproduction sexuée du champignon et réalisé l'analyse transcriptomique comparative avec des données déjà publiées lors de l'infection chez le peuplier. Les résultats de ces analyses sont présentés sous la forme d'une publication dans le **Chapitre II**. Puis, après avoir suivi une formation dédiée à l'annotation des éléments transposables (URGI, INRA de Versailles), j'ai procédé à l'annotation des gènes et des éléments transposables dans le génome de *M. larici-populina*, puis dans le génome de *M. allii-populina* et j'ai réalisé l'analyse comparative entre les deux génomes et les autres génomes de Pucciniales à disposition dans les bases de données (**Chapitre III**).
Dans le second axe de ma thèse, j'ai débuté l'analyse fonctionnelle de deux effecteurs candidats jugés prioritaires suite aux travaux réalisés en collaboration avec le TSL : l'effecteur Mlp124017 dont l'interaction avec une protéine de la famille des TOPLESS-related proteins avait pu être montrée ; et l'effecteur renommé Chloroplast Targeted Protein 1 (CTP1) dont la localisation dans le chloroplaste avait été mise à jour. Le pipeline d'analyse fonctionnelle retenu et les travaux réalisées dans l'étude de ces deux effecteurs sont rapportés dans le **Chapitre IV**.

CHAPITRE II

Analyse transcriptomique du cycle de vie de *Melampsora larici-populina* : comparaison des profils d'expression génique lors de l'infection du mélèze et du peuplier, les hôtes télien et écidien de l'agent de la rouille du peuplier

Partie 1 : Introduction

Le programme génétique exprimé par *M. larici-populina* lors de l'infection son hôte télien, le peuplier, et correspondant à un cycle de reproduction asexuée a été largement décrit depuis la publication de son génome (Duplessis *et al.,* 2011a; Hacquard *et al.*, 2012; Joly *et al.,* 2010 ; Petre *et al.*, 2012). D'ailleurs, l'infection par des urédiniospores est la partie du cycle de vie des champignons causant les maladies de rouille qui est le plus largement étudié (Duplessis *et al.,* 2012 ; Duplessis *et al.,* 2014 ; Aime *et al.,* 2017). En effet, les dommages causés en plantation par les agents des rouilles, et notamment pour la rouille du peuplier, sont dues à la multiplication végétative et à la dispersion des urédiniospores lors de phases épidémiques (Pinon & Frey, 2005). Les analyses transcriptomiques réalisées chez *M. larici-populina* ont montré la présence de vagues d'expression successives d'effecteurs candidats, suggérant que le champignon régule finement son programme génétique au cours de l'infection du peuplier (Duplessis *et al*., 2011b; Hacquard *et al.,* 2012). Ces études ont aidé à prioriser des effecteurs candidats pour en faire la caractérisation fonctionnelle (cf. **Chapitre IV** ; Petre *et al.,* 2015 ; Germain *et al.,* 2016).

Cependant, en tant que champignon hétéroïque *M. larici-populina* effectue sa reproduction sexuée sur un hôte alternant, dit hôte écidien, le mélèze (*Larix* spp.) au cours du printemps (Hacquard *et al.,* 2011b). Jusqu'à présent, nous n'avions aucune connaissance du programme génétique associé à l'infection de l'hôte écidien, et c'est aussi le cas pour la plupart des Pucciniales (Aime *et al.*, 2017). Ainsi, il reste à élucider si *M. larici-populina* utilise des répertoires d'effecteurs similaires ou différents pour réaliser l'infection chez ses deux hôtes, le mélèze et le peuplier. Cette question est récurrente au sein de la communauté des mycologues, le caractère hétéroïque des Pucciniales étant propre à cet ordre de champignons. Notamment, les Pucciniales infectent le plus généralement des hôtes appartenant à des taxons de plantes éloignés phylogénétiquement posant ainsi des questions très intéressantes quant à leur co-évolution avec ces hôtes (Aime *et al.,* 2017 ; Duplessis *et al.,* 2014 ; McTaggart *et al.*, 2016 ; Schulze-Lefert & Panstruga, 2011). Un nombre très restreint d'études ont été dédiées à l'analyse de l'expression des gènes au cours des infections des deux hôtes chez les Pucciniales (Cuomo *et al.,* 2017 ; Liu *et al.*, 2015 ; Xu *et al.,* 2011) et aucune n'a décrit de manière exhaustive ces programmes d'expression.

La maitrise récente du cycle complet de *M. larici-populina* au sein de l'UMR 1136 IAM pour réaliser des analyses génétiques (Pernaci *et al.,* 2014), a aussi permis d'entrevoir la possibilité d'étudier l'expression des gènes à différents stades du cycle sexuée sur mélèze. Ainsi, au

cours de ma thèse, nous avons réalisé l'analyse du programme génétique exprimé par *M. larici-populina* dans des basides produites à partir de télies sur peuplier à la sortie de la dormance du champignon après l'hiver et au cours de l'infection des aiguilles de mélèze, plus particulièrement avant et après la fertilisation. Les ARN produit aux stades basides, pycnies et écidies du cycle de vie de *M. larici-populina* ont ainsi permis de réaliser une analyse de type *RNAsequencing* (RNAseq) avec la méthode de séquençage Illumina. Dans un second temps nous avons utilisé les données d'expression acquises antérieurement au cours du cycle végétatif de *M. larici-populina* (infection des feuilles de peuplier ; Duplessis *et al.,* 2011b) pour comparer les profils d'expressions sur les deux hôtes. Ce projet a permis de montrer l'existence d'un programme génétique partagé entre les deux hôtes ainsi que la présence de gènes préférentiellement exprimés sur peuplier ou sur mélèze, et notamment des groupes de gènes codant des petites protéines sécrétées qui représentent des effecteurs candidats potentiels chez chaque hôte (Lorrain *et al.,* 2018a).

Ce chapitre de thèse comprend un article de recherche intitulé « **The rust fungus *Melampsora larici-populina* expresses a conserved genetic program and distinct sets of secreted proteins during infection of its two host plants: larch and poplar** » accepté en janvier 2018 pour publication dans la revue *Molecular Plant-Microbe Interactions* (en accès First Look sur le site du journal : https://apsjournals.apsnet.org/toc/mpmi/0/ja). La version primaire de l'article est aussi disponible sous la forme d'une prépublication (*preprint*) sur le site BioRxiv (https://www.biorxiv.org/content/early/2017/12/06/229971). Pour ce travail, le matériel biologique avait été produit par Stéphane Hacquard, Jérémy Pétrowski, Sébastien Duplessis et Pascal Frey avant mon arrivée sur le projet. Les ARN ont été extraits par Christine Delaruelle et le séquençage RNAseq a été réalisé par la société Beckman Counter Genomics. J'ai procédé à l'ensemble des analyses de données en collaboration avec Clémence Marchal lors de son stage de Master 2 dans l'UMR 1136 IAM. J'ai aussi rédigé intégralement la première version de l'article.

Partie 2 : Article n°3: The rust fungus *Melampsora larici-populina* expresses a conserved genetic program and distinct sets of secreted proteins during infection of its two host plants: larch and poplar

Cécile Lorrain, Clémence Marchal, Stéphane Hacquard, Christine Delaruelle, Jérémy Pétrowski, Benjamin Petre, Arnaud Hecker, Pascal Frey, Sébastien Duplessis

Article accepté pour publication en janvier 2018 dans la revue *Molecular plant-microbe interactions*.

The rust fungus *Melampsora larici-populina* expresses a conserved genetic program and distinct sets of secreted protein genes during infection of its two host plants, larch and poplar

Cécile Lorrain[1], Clémence Marchal[1π], Stéphane Hacquard[1€], Christine Delaruelle[1], Jérémy Pétrowski[1], Benjamin Petre[1,2], Arnaud Hecker[1], Pascal Frey[1], Sébastien Duplessis[1*]

[1] INRA/Université de Lorraine, UMR 1136 Interactions Arbres/Microorganismes, INRA Centre Grand Est-Nancy, F-54280 Champenoux, France

[2] The Sainsbury Laboratory, Norwich Research Park, NR4 7UH, Norwich, United Kingdom

[π] Present address, John Innes Centre, Norwich Research Park, Norwich, United Kingdom

[€] Present address, Max Planck Institute for Plant Breeding Research, Cologne, Germany

* Author for correspondence

Sébastien DUPLESSIS

INRA - UMR 1136 Interactions Arbres/Microorganismes

INRA Centre Grand Est - Nancy, 54280 Champenoux, France

Tel, +33 383 39 40 13

Fax, +33 383 39 40 69

E-mail, sebastien.duplessis@inra.fr

ABSTRACT

Mechanisms required for broad spectrum or specific host colonization of plant parasites are poorly understood. As a perfect illustration, heteroecious rust fungi require two alternate host plants to complete their life cycles. *Melampsora larici-populina* infects two taxonomically unrelated plants, larch on which sexual reproduction is achieved and poplar on which clonal multiplication occurs leading to severe epidemics in plantations. We applied deep RNA sequencing to three key developmental stages of *M. larici-populina* infection on larch: basidia, pycnia and aecia; and we performed comparative transcriptomics of infection on poplar and larch hosts using available expression data. Secreted protein was the only significantly over-represented category among differentially expressed *M. larici-populina* genes between the basidial, the pycnial and the aecial stages, highlighting their probable involvement in the infection process. Comparison of fungal transcriptomes in larch and poplar revealed a majority of rust genes commonly expressed on the two hosts and a fraction exhibiting host-specific expression. More particularly, gene families encoding small secreted proteins presented striking expression profiles that highlight probable candidate effectors specialized on each host. Our results bring valuable new information about the biological cycle of rust fungi and identify genes that may contribute to host specificity.

Key-words: Pucciniales, Rust fungi, Heteroecism, Basidiospores, Pycniospores, Aeciospores, comparative transcriptomics, effectors.

INTRODUCTION

Rust fungi (Pucciniales) are obligate biotrophs that establish parasitic associations with their host plants. Pucciniales depict one of the largest orders of plant pathogenic fungi with more than 8,000 species infecting a wide range of hosts within ferns, gymnosperms and angiosperms (Aime *et al.,* 2014; Cummins & Hiratsuka, 2003). A remarkable feature of some rust fungi is heteroecism – they infect two unrelated host plants to complete their life cycle. To do so, they produce different spore forms and the life cycle is qualified as macrocyclic. For instance, the poplar rust fungus *Melampsora larici-populina* has a heteroecious and macrocyclic life cycle (Hacquard *et al.,* 2011a; Vialle *et al.,* 2011). It infects two unrelated host plants: Larch (*Larix* spp., conifer, aecial host) and Poplar (*Populus* spp., dicot, telial host). Infection of larch needles is achieved by *M. larici-populina* haploid basidiospores and results in the formation of pycnia and pycniospores (Pinon & Frey 2005; Hacquard *et al.,* 2011a). After fertilization and production of aecia, airbone aecispores infect poplar leaves leading to the production of urediniospores within a week (Harder 1984; Voegele *et al.,* 2009). Asexual multiplication of urediniospores causes severe epidemics in poplar plantations all over summer (Pinon & Frey 2005). In autumn, uredinia differentiate into telia, in which karyogamy occurs and diploid teliospores overwinter on dead poplar leaves (Hacquard *et al.,* 2013). When favourable conditions are met the next spring, basidiospores are produced and the cycle repeats again. The molecular bases of these tightly controlled developmental stages and the genetic programmes underlying host alternation in rust fungi remain poorly understood (Bakkeren *et al.,* 2016; Liu *et al.,* 2015; Xu *et al.,* 2011).

During infection, pathogens use effector proteins to modulate the cellular and the immune processes of the host plant (Win *et al.,* 2012; Lo Presti *et al.,* 2015). Effectors are critical to ensure successful establishment of the pathogen in host tissues (Lo Presti *et al.,* 2015). In filamentous plant pathogens (fungi and oomycetes), effector proteins are often small proteins that possess a signal peptide for secretion (Sperschneider et al. 2015). In rust fungi, genome and transcriptome analyses predicted hundreds of secreted proteins (SPs) representing candidate effectors expressed during host infection (Duplessis *et al.,* 2012; Duplessis *et al.,* 2014). Effectors are specifically adapted to manipulate host plants and they evolve through the pressure of host recognition receptors (Rovenich *et al.,* 2014).

The genetic program used by *M. larici-populina* to colonize poplar leaves has been extensively described (Joly *et al.,* 2010; Duplessis *et al.,* 2011b; Hacquard *et al.*, 2010, 2012; Petre *et al,.* 2012); mostly because of epidemics occurring in poplar plantations (Pinon & Frey 2005).

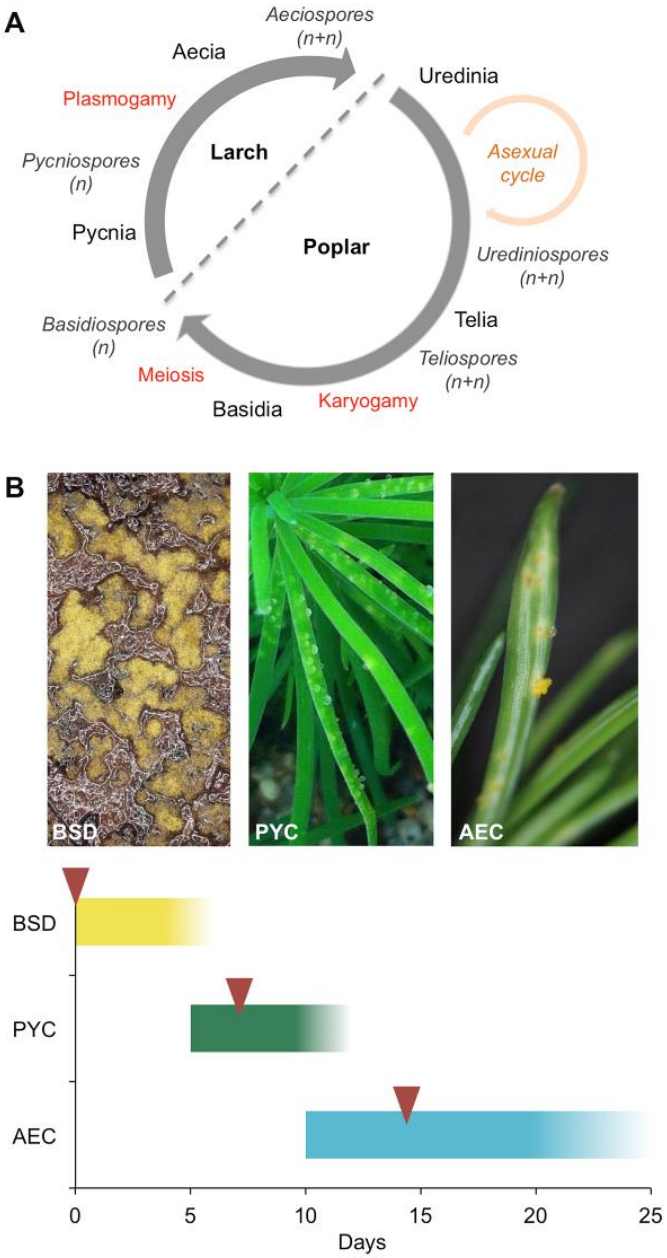


Figure 1: *Melampsora larici-populina* life cycle and experimental design for stage selection
(A) *M. larici-populina* heteroecious and macrocyclic life cycle. Haploid basidiospores (n) infect the aecial host (*Larix* spp.), (0) Pycnia are formed on larch needles producing nectar droplets containing pycniospores (n) with different mating types for fertilization. After plasmogamy, dikaryotic aecia establish. Aeciospores (n+n) infect leaves of the telial host (*Populus* spp.) and produce dikaryotic urediniospores (n+n). Asexual cycle with continuous production of urediniospores takes place during the summer season. In autumn, telia differentiate and produce teliospores (n+n) in which, karyogamy occurs and meiosis begins before dormancy. After winter, meiosis continues, basidia germinate and produce haploid basidiospores (n). (B) Pictures from left to right: basidia and basidiospores on a dead poplar leaf, pycnia nectar droplets on larch needles, and aecia and aeciospores on larch needles. Color bars indicate the three larch-infection related stages sampled for RNA sequencing. BSD refers to the basidial stage and the yellow bar represents the time between basidiospores inoculation and pycnia formation. PYC corresponds to the pycnial stage (visible as nectar droplets) and the green bar represents the period of presence of visible PYC on larch needles. AEC refers to the aecial stage and the blue bar indicates the presence of AEC on larch needles. Red arrows represent the time of collection of the three biological stages.

Transcriptome analyses revealed waves of expression for candidate effector genes at different time-points on poplar, suggesting that the fungus tightly controls its genetic program during infection (Duplessis *et al.,* 2011b; Hacquard *et al.,* 2012; Petre *et al.,* 2012). These studies helped to prioritize candidate effectors for functional analysis (Petre *et al.,* 2015; Petre *et al.,* 2016a; Germain *et al.,* 2017). However, it is not known whether similar or different sets of effectors are acting during the infection of the aecial and the telial host of *M. larici-populina*. This has been a general and recurrent question related to the capacity of heteroecious rust fungi to infect different hosts belonging to unrelated plant taxa (Schulze-Lefert & Panstruga 2011; Duplessis *et al.,* 2014). So far, a limited number of studies addressed gene expression in the different hosts of rust fungi (Cuomo *et al.,* 2017; Liu *et al.,* 2015; Xu *et al.,* 2011) and none clearly determined to which extent a portion of the gene complement may be unique to one host or another.

In this study, we took advantage of the recent ability to complete the life cycle of the poplar rust fungus under laboratory conditions (Pernaci *et al.,* 2014) to investigate the genetic programme used by the fungus during larch colonization. To this end, we used an Illumina RNA-sequencing (RNAseq) approach to assess fungal transcript abundance at three different stages of the life cycle: basidia, pycnia, and aecia. We also took advantage of previously published transcriptome data during poplar infection to compare *M. larici-populina* gene expression profiles on the two host plants. We identified genes preferentially expressed on larch or poplar, including secreted proteins that represent potential specific candidate effectors. We further scrutinized the expression features of pre-defined Small-Secreted Protein (SSP) gene families (Hacquard *et al.,* 2012) to investigate whether members from SSP families showed a host-specific expression.

RESULTS

Transcriptome of *M. larici-populina* larch infection related stages

To gain insights into the genetic program of *M. larici populina* during larch infection, we performed deep RNAseq at three distinct stages: the basidial stage collected on dead poplar leaves, the pycnial stage and the aecial stage collected on larch needles at 7 and 14 days post-inoculation (dpi), respectively (Fig. 1, Fig. S1). Large areas covered with apparent fungal structures were targeted in order to maximize the capture of fungal transcripts. An average of 114 million reads per sample and more than 1 billion reads in total were collected (Table S1) and mapped onto 16,399 *M. larici-populina* transcripts. Overall, 16,386 transcripts had at least one read assigned. The proportion of reads mapped ranged from 10.3 to 59.3% in replicate 2 of the pycnial stage and replicate 1 of the basidial stage, respectively (Table S1). This reflected

the different proportion of fungal transcripts in plant tissues. Each replicate reached a plateau in rarefaction curves independently of the stage (Fig. S2), indicating a sufficient sequencing depth to capture a full transcriptome at each stage. Principal component analysis of DESeq2 rlog-transformed normalized counts showed homogeneous biological replicates (Fig. S3A). The first axis (94% of the variance) separated the basidial stage from the aecial and the pycnial stages, whereas the second axis (3% of the variance) separated the aecial and the pycnial stage. One aecial replicate was apart from the others, probably because of the higher proportion of null counts observed for this replicate and the overall heterogeneity of the biological material at this stage compared to the basidial and the pycnial stages. Normalized counts distributed in a similar manner across all the samples (Fig. S3B). Average expression values were then considered for the basidial, pycnial and aecial stages. Overall, the capture of complete transcriptomes by the RNAseq approach is suitable for further comparative analysis between stages.

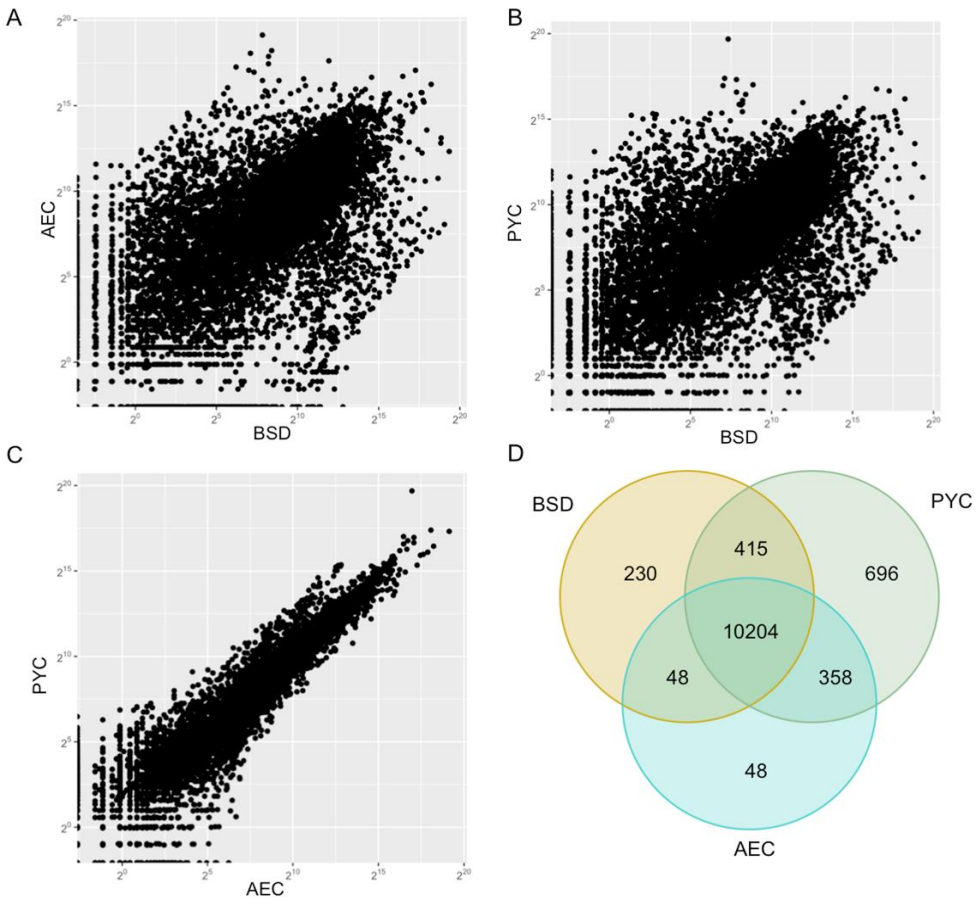


Figure 2: RNA-seq analysis of the basidial (BSD), pycnial (PYC) and aecial (AEC) stages of the rust fungus *Melampsora larici-populina*

(A) Scatter plot of rlog transformed sequencing counts of transcripts expressed in AEC *versus* BSD; (B) in PYC *versus* BSD and (C) in PYC *versus* AEC. (D) Venn diagram showing the number of genes detected by at least one read per sample in each condition basidial, pycnial and aecial stages.

Gene expression profiles correlate in the pycnial and the aecial stages

To further explore differences between the basidial stage and larch *in planta* stages, we analysed global gene expression profiles and differentially expressed genes (DEGs). Pycnial and aecial stages occurring in larch needles showed overall closer expression profiles, whereas they are more dispersed when compared with the basidial stage (Fig. 2A-C). In total, 10,204 transcripts (62% of the 16,399 predicted *M. larici-populina* transcripts) were commonly detected in the basidial, pycnial and aecial stages (Fig. 2D). Only 230, 696 and 48 transcripts were detected exclusively in the basidial, pycnial and aecial stages, respectively. Pairwise comparisons were performed between the three stages, and transcripts with normalized fold change and *padj*<0.05 in at least one paired comparison were deemed as significantly differentially expressed genes (DEGs; up- or down-regulation). In total, 7,743 transcripts were significantly differentially expressed in at least one stage compared to another (Table S2) and 7,031 and 7,079 transcripts were significantly differentially expressed between pycnial and basidial stages or aecial and basidial stages, respectively (Table 1, Table S3). Fewer DEGs (2,141) were detected between the aecial and the pycnial stages. We conclude that the closely related profiles recorded for the pycnial and aecial stages support the expression of a similar genetic program *in planta* in larch.

Table 1: Number of *Melampsora larici-populina* genes significantly differentially expressed at the basidial (BSD), pycnial (PYC) and aecial (AEC) stages.

Regulated genes[a]	PYC *vs.* BSD	AEC *vs.* BSD	AEC *vs.* PYC
Genes significantly differentially expressed (p <0.05)	7031	7079	2141
Up-regulated	3698	3503	1045
> 3-fold	2406	2280	90
> 10-fold	1526	1316	5
> 100-fold	522	442	0
> 1000-fold	105	67	0
Down-regulated	3333	3576	1096
< 3-fold	1872	2128	539
< 10-fold	849	1053	265
< 100-fold	319	429	27
< 1000-fold	57	126	0

[a] Down- and up-regulated genes were classified according to their fold-change ratio. Transcripts with a significant *p*-value (< 0.05) and more than a threefold change in transcript level were considered as significantly differentially expressed.

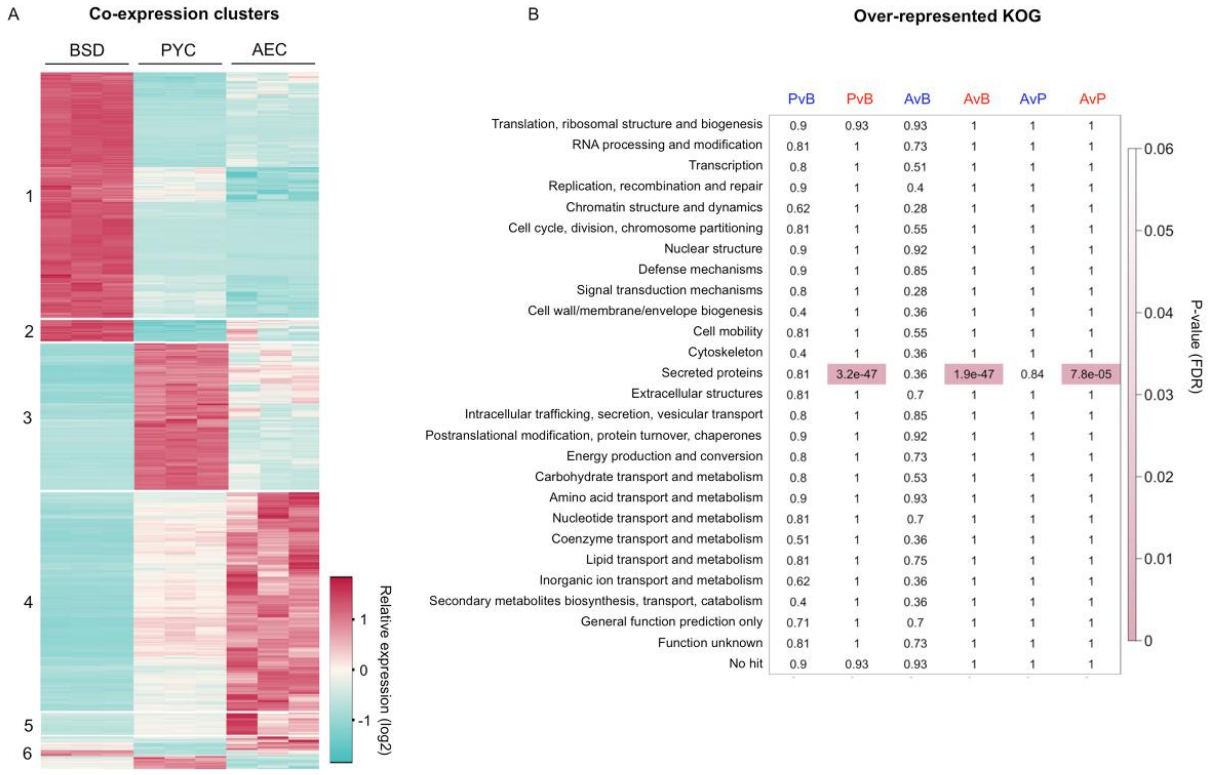


Figure 3: Co-expression analysis and KOG enrichment analysis of differentially expressed *Melampsora larici-populina* genes in the basidial and the larch infection stages

(A) Transcript profiling of 1,548 *M. larici-populina* significantly differentially expressed genes at the basidial (BSD), pycnial (PYC) and aecial (AEC) stages (DESeq2; *pajd* <0.05). Relative expression profiles are represented as log2 of normalized counts across three replicates of BSD, PYC and AEC, from over-expression (black) to under-expression (white). Using co-expression network analysis (WGCNA), genes were grouped into six clusters of co-expression. (B) FDR p-values matrix from over-represented KOG categories of 3 fold-change down-regulated (DR) genes and 3 fold-change up-regulated (UR) genes. Significantly enriched categories (FDR<0.05) are highlighted with an asterisk (*).

Secreted proteins is the only overrepresented category among DEGs detected on larch

To discriminate between sets of *M. larici-populina* genes expressed at the basidial stage and during larch infection, we conducted a co-expression network analysis (Langfelder & Horvath, 2008). This correlation-based network analysis identifies clusters of genes with expression profiles that are more linked to each other than they are with genes outside the cluster. Among 1,548 significantly DEGs (*padj*<0.05), we identified six clusters of transcripts sharing co-expression profiles (Fig. 3A, Fig. S4). The clusters positively correlated with the basidial stage were negatively correlated with the *in planta* pycnial and aecial stages (Fig. 3A, Fig. S4). Clusters 1 and 2 were significantly positively correlated with the basidial stage and represented 38.8 % of the DEGs (661). The cluster 3 was significantly positively correlated with the pycnial stage, representing 21.3% of the selected DEGs (330). Clusters 4 to 6 were significantly positively co-expressed at the aecial stage with 39,8% of DEGs (617; Fig. 3A). In total 61.2%

of the DEGs were significantly co-expressed *in planta* compared to the basidial stage. Clusters 3, 4 and 5 were significantly negatively correlated with the basidial stage, while positively correlated with *in planta* stages (Fig. 3A, Fig. S4).

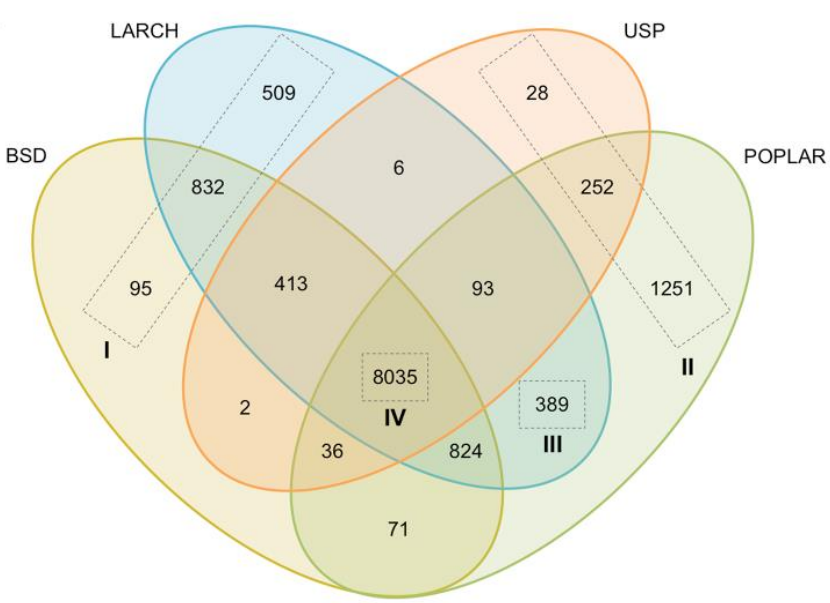
To further understand the transcriptome reprogramming occurring between basidial and larch *in planta* stages, we conducted an enrichment analysis of KOGs in pairwise comparisons of regulated genes. Transcripts regulated between two stages with at least a 3fold-change and a *padj*<0.05 were considered (Table 1). We also created a specific KOG-SP category for secreted proteins of unknown function. Among up-regulated genes in aecial versus pycnial (1045 DEGs), aecial versus basidial (3503 DEGs) and pycnial versus basidial (3698 DEGs) pairwise comparisons, the only significantly enriched category is KOG-SP (Fig. 3B). SPs represented 28.9%, 22.4% and 22.1% of aecial versus pycnial, aecial versus basidial and pycnial versus basidial up-regulated genes, respectively. SP encoding genes were over-represented in up-regulated genes of aecial and pycnial stages compared with the basidial stage. We also performed RT-qPCR on 22 randomly selected SP genes to confirm their RNAseq expression profiles (Fig. S5). A good correlation could be observed between RNAseq and RT-qPCR for pycnial and aecial stages whereas differences are noticeable for the basidial stage, which may relate to the intrinsic nature of the biological material (dead poplar leaves). Our results show that more SP genes are expressed during larch *in planta* stages compared to the basidial stage.

Comparison of *M. larici-populina* transcriptomes on larch and poplar highlights subsets of commonly and specifically expressed genes

To identify *M. larici-populina* genes specific to larch or poplar infection we compared transcriptome profiling of larch stages to previous expression studies performed on the telial host (Duplessis *et al.,* 2011b). We simplified our dataset to four explicit stages: the basidial stage as a larch-infecting spore stage, LARCH as the *in planta* stage on the aecial host, Urediniospores as a poplar-infecting spore stage and POPLAR as the *in planta* stage on the telial host. Expression data recorded for poplar-related stages were obtained with custom oligoarrays. Thus, direct calculation of fold-change was not possible between larch- and poplar-related stages due to the different technologies used. However, the transcriptomes were complete on both hosts (Duplessis *et al.,* 2011b and Fig. S2), allowing direct comparison to determine the presence of specifically expressed genes (at least count of 1 by RNAseq or above background value on oligoarrays). Overall, 14,883 transcripts were detected in poplar- or larch-related stages, and 8,035 transcripts are expressed in all four fungal stages (Fig. 4A). A total of 1,436 transcripts were only detected at the basidial stage and LARCH, and 1,531 transcripts were detected on Urediniospores and POPLAR. Interestingly, 389 genes were

specifically detected in LARCH and POPLAR, revealing a set of *in planta*-specific genes. In total, 2,967 (20%) genes are specifically expressed on larch or poplar *in planta* stages.

A



B

	I	II	III	IV	Expressed
CAZymes	35	11	8	310	454
Lipases	10	3	-	61	88
Proteases	20	8	8	304	380
Secreted Proteins	245	103	98	688	1645
Transporters	8	7	1	282	327
Other	1118	1399	274	6390	11989
Total	1436	1531	389	8035	14883

Figure 4: Comparative transcriptomic analysis between four stages of *Melampsora larici-populina* life cycle

(A) Venn diagram showing the number of genes expressed in BSD (basidial stage), LARCH (pycnial and aecial stages, i.e. during larch needle infection), USP (dormant and germinating urediniospores) and POPLAR (during poplar leaf infection). Group I refers to genes specifically expressed in larch-related stages; group II refers to genes specifically expressed in poplar-related stages; group III refers to the genes specifically expressed *in planta* and group IV refers to genes detected in all four stages. (B) Number of genes expressed in selected annotated gene categories related to host infection: carbohydrate-active enzymes (CAZymes), lipases, proteases, secreted proteins and transporters in selected groups I, II, III and IV. The total numbers of expressed genes for each selected annotated category is shown on the right. Total numbers of genes belonging to selected groups I, II, III and IV are shown below the table.

To determine whether *M. larici-populina* uses specific pathways or functions during colonization of larch and poplar, we surveyed expression profiles of gene categories. Firstly, no major difference could be determined on one host or another regarding the expression of KOG cellular categories (Fig. S6). Manual exploration of KEGG annotations revealed a similar result for metabolic pathways (Table S3). Secondly, of 6,660 genes of unknown function (excluding SP genes addressed below) with expression information, 58% were expressed at all stages, whereas 13% (863) and 14.5% (967) presented a specific expression pattern on larch and poplar, respectively (Fig. S7). We scrutinized cellular categories previously annotated in the *M. larici-populina* genome (Duplessis *et al.,* 2011a). Among those, 282 transporters, 310 CAZymes and 304 proteases were expressed at all four stages. Few genes in these surveyed categories were specifically detected in larch- and poplar-related stages. Likewise, only 8 transporters were detected only at the basidial stage and on LARCH, including three auxin efflux carriers and one oligopeptide transporter, which belong to expanded rust fungal gene families (Duplessis *et al.,* 2011a; Fig. 4B; Table S2). A total of 35 CAZyme genes were detected at the basidial and LARCH stages, including 22 glycoside hydrolases; and 20 proteases were detected at these stages, including 8 peptidase inhibitors. Even fewer genes were specifically detected in USP and POPLAR, with 7, 11 and 8 transporters, CAZymes and proteases, respectively. Overall, less than 10% of CAZymes, proteases, lipases or transporters are specifically expressed in polar or larch stages. Among mating-related genes, pheromone receptor genes were expressed in both hosts and one gene showed a very high expression in LARCH. Four pheromone precursor genes showed a very high expression in LARCH compared to other stages and were in the top 1% most highly expressed genes at the aecial stage (Table S2), whereas mating compatibility is established between pycniospores and receptive hyphae. In total, 60% of the specifically expressed genes on poplar and on larch are of unknown function.

Since SP genes were significantly over-represented in *M. larici-populina* genes expressed on larch (this study) and on poplar (Duplessis *et al.,* 2011b), we particularly considered their expression profiles during larch and poplar infection. Among 1,645 expressed SP genes, 688 (42%) were detected at all developmental stages (Fig. 4B). Transcripts specifically detected either on larch or poplar contained an important proportion of SP-encoding genes, accounting for 17 and 25% of the total number of genes specifically expressed at these stages, respectively. A lower proportion of SP genes were specifically detected on the telial host poplar, compared to the aecial host larch. Even if the proportion of SP genes expressed at all stages was important (688; 42% of 1,645 SP genes), it represented only 8.6% of the 8,035 transcripts expressed at all four stages. Moreover, of the top 10% genes preferentially highly expressed on larch or on poplar *in planta*, 18% are secreted proteins and among them, 14%

are SSPs (Fig. 5). In comparison, SPs and SSPs represent only 4 and 5% rescpectively of other expressed genes. Altogether, these results indicate an enrichment of SSPs among genes highly expressed during poplar or larch infection.

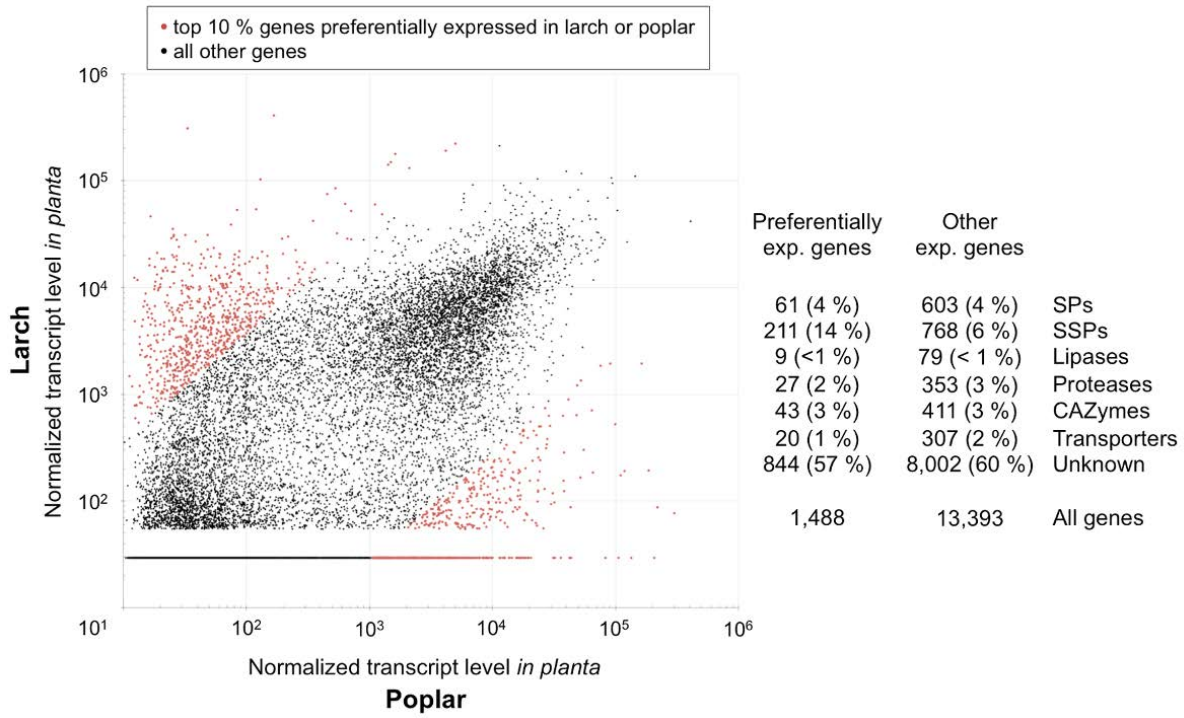


Figure 5: *Melampsora larici-populina* genes preferentially expressed *in planta* on larch and on poplar

Scatter plot representing the normalized gene expression level of 14,883 *M. larici-populina* genes expressed *in planta* on poplar (x-axis) and on larch (y-axis). The dark grey triangles represent the top 10% genes preferentially expressed on poplar (5%) and on larch (5%). Numbers and percentage of genes in secreted proteins (SPs), small secreted proteins (SSPs), lipases, proteases, CAZymes, transporters and genes of unkown function (Unknown) of the top 10% genes and of other expressed genes are detailled. The original expression data for rust-infected poplar leaves and rust-infected larch needles were quantile normalized to allow comparison.

Gene members of *M. larici-populina* SSP families present specific patterns of expression on larch and on poplar

We took the opportunity of the detailed analysis of SSP gene families in *M. larici-populina* (Duplessis *et al.,* 2011a; Hacquard *et al.,* 2012) to further explore their specific expression patterns on the two host plants. RNAseq and oligoarray expression data were quantile-normalized and compared to determine whether entire families or members of families are host-specific (Fig. 6). Overall, 205 and 351 SSP family genes were preferentially expressed *in planta* on larch and on poplar, respectively when the general expression profile of all SSP genes was considered (Fig. 6A). When focusing on singletons or families of SSPs grouping

two, three or more members, three-fourths of the genes were detected on the two hosts (Fig. 6B). The remaining fourth corresponded to genes expressed specifically in one or the other host (Basidia and LARCH or Urediniospore and POPLAR). We further considered large SSP families to identify family-specific expression patterns. Among 64 families of at least four members, 17 (26%) showed a preferential expression on one given host (9 and 8 SSP families preferentially expressed in larch or poplar *in planta* stages, respectively). For instance, the SSP family 1 (111 members) presented 24 members expressed specifically at the basidial stage and on LARCH and 4 only on Urediniospore and POPLAR, the remaining members showing a higher expression at the basidial stage and on LARCH (Fig. 6C). On the contrary, the SSP family 9 (11 members) showed a preferential expression on poplar, and only two members presented noticeable expression level on the other host. The SSP family 7 presented a striking profile with almost a half of the gene members expressed in the aecial host larch, whereas the other half was expressed in the telial host poplar. Some of these members presented a sequence identity of 75% and showed opposite expression profiles on the two host plants (e.g. SSP genes 86274 and 91075 expressed in larch and in poplar, respectively; data not shown). In smaller SSP gene families, the different situations exemplified in the Fig. 6C were also observed (Table S2). Expression profiles of *M. larici-populina* SSP gene families proved to be diverse with a 60% of genes expressed on the two host plants, 31% showing preferential expression on one host or the other and 9% of divided families (6 SSP families of more than three members).

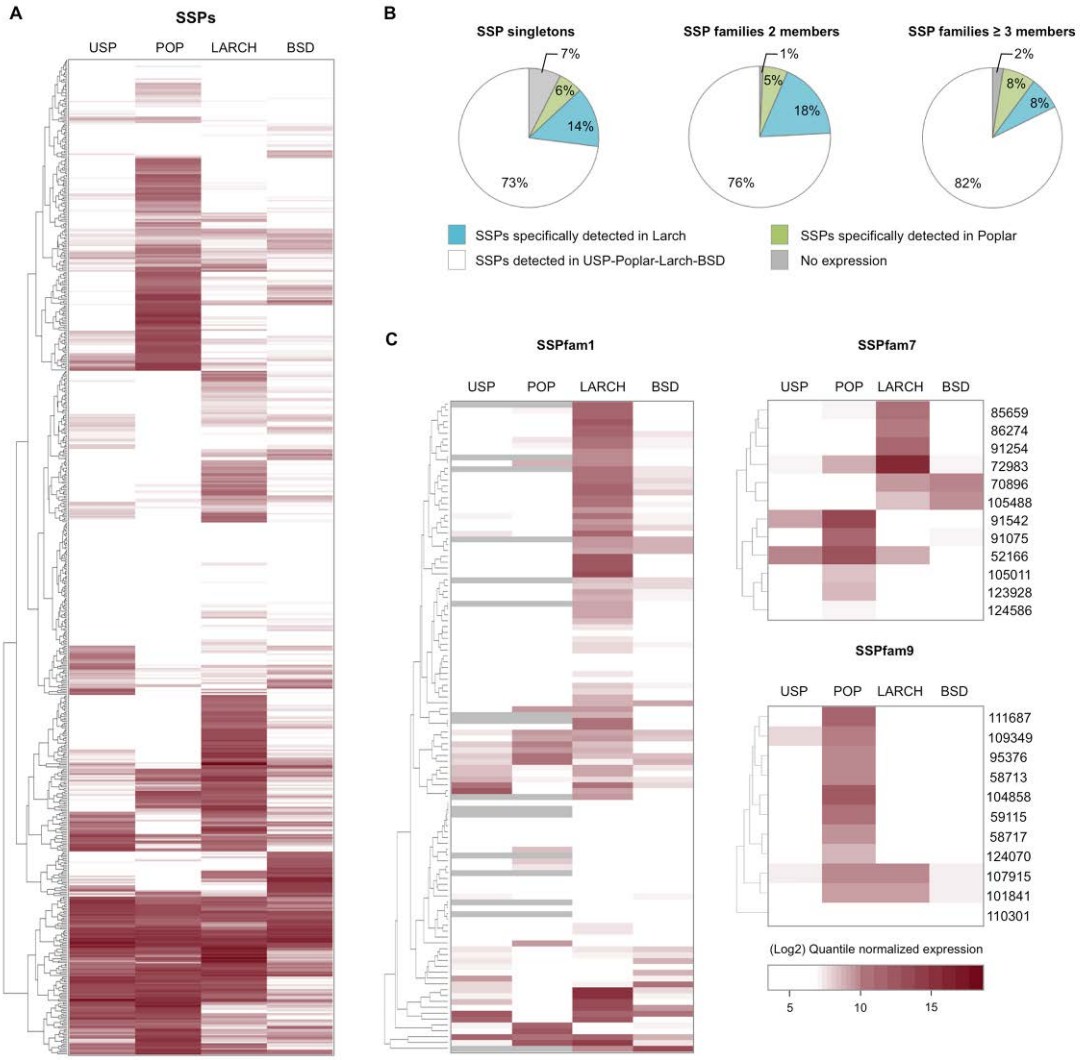


Figure 6: Transcript profiling of *Melampsora larici-populina* SSP genes on the aecial and the telial hosts

(A) Heatmap of expression profiles of *M. larici-populina* SSP transcripts. Expression levels (white to red) are shown as quantile normalized expression values (white corresponding to the lowest expression and dark red corresponding to the greatest expression). (B) Diagrams presenting the expression patterns for SSP singletons, SSP gene families with 2 members and SSP gene families with 3 or more members detected in larch- and poplar-related stages (i.e. basidial stage BSD and pycnial/aecial stages LARCH; and urediniospores USP and poplar infection POP, respectively). (C) Examples of expression profiles for selected SSP families: SSP family 1, preferentially expressed in larch; SSP family 7, expressed both in poplar and larch; SSP family 9, preferentially expressed in poplar. JGI protein ID numbers are indicated on the right of profiles for SSP families 7 and 9. Grey on the SSPfam1 expression profile indicates non-available information (no specific oligomers represented on the oligoarray).

DISCUSSION

In the present study, we report on gene expression profiling at three biological stages that are related to infection of larch, the aecial host of the poplar rust fungus *M. larici-populina*, taking advantage of an experimental set-up in semi-controlled laboratory conditions (Pernaci *et al.*, 2014). A full transcriptome was recovered at each of the stages through deep RNA-Sequencing allowing comparisons between stages. In total, 73% of *M. larici-populina* genes were expressed at the basidial, pycnial and aecial stages. This is similar to previous findings (76%) where stages related to infection of poplar, the telial host of the fungus were considered (Duplessis *et al.*, 2011a; Hacquard *et al.*, 2010, 2012). Altogether, 62% of the genes were commonly expressed in the basidial, pycnial and aecial stages, however the general profile was closer between the two larch-infection stages than with the basidial stage. This suggests an important reprogramming of the fungal transcriptome after the transition to host infection. Although the pycnial and aecial stages occur before and after fertilization, major *in planta* structures such as protoaecia are formed before the fertilization between pycniospores and receptive hyphae of compatible mating types (Harder, 1984; Voegele, Hahn, & Mendgen, 2009a), which may explain the similarity of the general expression profile at these stages. It is worth noticing that mating-related genes were among the most highly expressed genes detected at the pycnial stage. The analysis of the wheat leaf rust genome recently resolved the mating system and loci for a rust fungus (Cuomo *et al.,* 2017). Exploring expression of mating genes on the host on which sexual reproduction occurs would allow a better understanding of the underlying genetic processes.

Heteroecious rust fungi exhibit the most complex life cycle known in fungi. They involve infection of two taxonomically unrelated host plants and production of five different types of spores (Aime *et al.,* 2017). Transcriptomics of rust has been widely applied to asexual stages and spores responsible for epidemics and damages in plantations (Duplessis *et al.,* 2014). There is still scarce information about expression along the whole life cycle of rust fungi, and only a few studies looked at transcript profiles on alternate host plants. In *Puccinia triticina*, EST sequencing and RNA-Seq identified sets of genes expressed on the two host plants, meadow rue (*Thalictrum* spp.) and wheat, however at a limited scale (Xu *et al.,* 2011; Cuomo *et al.,* 2017). RNA-Seq was applied at a higher depth to *Cronartium ribicola* spores collected from the two host plants, *Pinus monticola* (aeciospores) and *Ribes nigrum* (urediniospores), and compared to two infection stages in western white pine (Liu *et al.,* 2015). Although no reference genome is available for *C. ribicola*, the RNA-Seq approach could identify a large proportion of *C. ribicola* genes expressed in all situations (61%) with only a few (4-5%) specific to each separate stages (Liu *et al.,* 2015).

In our study, the significant over-representation of SPs among regulated genes between stages was striking and mirrors previous findings on the poplar host (Duplessis *et al.,* 2011b; Hacquard *et al.,* 2010, 2012). This shows that the expression of SP genes is important for larch infection and central to biotrophy on both hosts. Many effectors of fungal pathogens are secreted proteins, including avirulence effectors identified in the flax rust fungus *Melampsora lini* (Tyler & Rouxel 2011; Sharpee & Dean 2016, Duplessis *et al.,* 2012). Rust fungi exhibit large secretomes composed of expanded gene families and SP/SSP genes specifically expressed *in planta* represent priority candidate effectors (Aime *et al.,* 2017; Petre *et al.,* 2014). Previous expression profiling reports of poplar infection by *M. larici-populina* have identified long lists of putative effectors that are key candidates for functional analysis (Hacquard *et al.,* 2010; Joly *et al.,* 2010; Duplessis *et al.,* 2011b; Hacquard *et al.,* 2012; Petre *et al.,* 2012; Hacquard *et al.,* 2013). No expression data was available before for infection of larch needles by *M. larici-populina*. Whether specific sets of effectors may be required for rust fungi to infect their different hosts was so far unaddressed (Shulze-Lefert & Panstruga 2011). Only a fraction of SSP genes was specifically expressed on poplar or on larch, in a similar fashion for singletons or multigene families and a larger portion is expressed on the two hosts. Our results confirm and largely extend the observations made in *P. triticina* or *C. ribicola* regarding specific gene expression (Xu *et al.,* 2011; Liu *et al.,* 2015; Cuomo *et al.,* 2017). Considering that rust effectors are likely to reside among these genes, our results could indicate that a considerable part of *M. larici-populina* effectors may share similar functions on the two hosts.

Some *M. larici-populina* SSP families showed a preferential expression on a single host, and more striking was the profile of families exhibiting members whose expression was host-specific. Families showing host-specific expression could reflect a diversification under the pressure of resistance genes or the evolution of a new effector function targeting a particular process not represented in the alternate host. Regarding gene families combining specificity of expression towards the two hosts, such as the *M. larici-populina* SSP7 family, we may observe adaptation of subsets of effectors i) with different functions in the cells of the two hosts, or ii) with a conserved function but different targets in the two plants, or iii) having the same target that evolved differently in the two plant lineages from a unique ancestral target. These results open many avenues for speculation about the evolution of SSP and candidate effectors families in rust fungi, but it remains to determine whether these particular families correspond to *bona fide* effectors. With the knowledge of preferential or specific expression on the two hosts, it is now possible to narrow down the list of candidate effectors that are only expressed in the main host where most of the damages are caused and prioritize future functional

investigations. Beside, the subset of candidate effectors expressed on the two host plants (i.e. core effectors) may be important in the rust general infection process and can represent useful targets for resistance breeding. These findings also open perspectives for the development of resistances in the alternate host in which the rust fungus accomplishes its sexual reproduction.

Beyond candidate effectors, other SP genes such as the CAZymes showed specific expression in poplar or in larch. Similar observations were made for *C. ribicola* infecting *Ribes* spp. or pine (Liu *et al.,* 2015). It is known that plant-cell wall varies in terms of composition, architecture and integrity across plant species (Kubicek *et al.,* 2014). CAZymes are essential enzymes for decomposing polysaccharides from plant cell walls and leaf or needle cuticles (Henrissat *et al.,* 2017). Rust fungal hyphae need to penetrate through the plant cuticle in the aecial host and to go through the plant cell wall to form haustoria within the host cell cavity in both hosts (Ragazzi *et al.,* 2005). Different small sets of CAZymes were expressed in poplar and larch, which could reflect the adaptation to the host polysaccharide composition in a conifer and a dicot. On the contrary, the survey of central metabolic pathways or cellular categories did not reveal major differences. *M. larici-populina* seems to express a conserved genetic program during the biotrophic growth in larch and poplar tissues suggesting a common strategy for the acquisition and utilisation of nutrients derived from the two hosts. A surprising feature of rust fungal genomes is the proportion of genes of unknown function with no homology out of the Pucciniales (Duplessis *et al.,* 2011a, Aime *et al.,* 2017). Among 6,660 *M. larici-populina* genes of unknown function with expression record, more than 10% are specifically expressed on poplar or on larch, whereas 58% are expressed on both. A better dissection of the infection process on the two hosts could help to draw hypotheses about processes not yet unravelled and in which these genes could be involved, e.g. during early colonization, biotrophic growth or sporulation. Moreover, systematic comparative transcriptomics in different heteroecious rust fungi, as well as with autoecious rust fungi, may provide new leads to unravel new functions related to biotrophy.

When the transcriptome is considered across the whole *M. larici-populina* life cycle, we could detect expression for 86% of the genes and a total of 2,047 genes were never detected. Although this relatively small number of genes may contain annotation errors, it is likely that some may be specific of sections of the rust biological life cycle not explored in this or in previous studies. Telia are overwintering structures adapted to drastic environmental conditions in which important biological processes such as karyogamy and meiosis occur (Mendgen, 1984). A small proportion of genes specifically expressed in telia were previously reported in *M. larici-populina* (Hacquard *et al.,* 2013), which could be explained by the fact that transcriptome profiling was carried out in early differentiating telia. This telia stage was not

included for comparison in the present study because it was obtained from environmental samples different from the reference strain 98AG31 (Hacquard *et al.,* 2013). When compared with the present study, only 117 genes could be detected as telia-specific (data not shown). More specific profiles can be expected at later stages of telia differentiation when karyogamy and meiosis occur or during entry and exit of winter dormancy. Moreover, the infection of the aecial host is far more complex than in the telial host with the establishment of distinct fungal cell types in a coordinated manner (Harder, 1984; Littlefield & Heath, 1979). Only three specific stages were selected for this work and specific expression of other genes may be expected during the course of the interaction with the aecial host, as previously illustrated in poplar (Duplessis *et al.,* 2011b; Hacquard *et al.,* 2010; Hacquard *et al.,* 2012).

Other fungal biotrophs possess the capacity to interact with multiple host species with various levels of specificity. Mycorrhizal fungi are biotrophs that engage in mutualistic associations with their host plants by forming specific infection structures (van der Heijden *et al.,* 2015). During the colonization process, they deliver symbiosis effectors to establish long lasting beneficial interactions (Martin *et al.,* 2016). Large-scale transcriptome studies have shown that ectomycorrhizal fungi also present specific expression of large sets of SSP genes of unknown function (Kohler *et al.,* 2015). Similar to our observations with *M. larici-populina*, mycorrhizal fungi express common and specific sets of SSP genes during their interaction with different host plants. This has been shown for ectomycorrhizal fungi in different interactions between *Suillus* spp. and *Pinus* spp. (Liao *et al.,* 2016) and *Laccaria bicolor* with poplar and Douglas fir (Plett *et al.,* 2015). The same was observed for the arbuscular mycorrhizal fungi *Rhizophagus irregularis* and *Gigaspora rosea* colonizing different host plants (Kamel *et al.,* 2017). The fungal root endophyte *Piriformospora indica* establishes different colonization strategies in a host-dependant manner, i.e. biotrophy in *Arabidopsis* and saprotrophy in barley (Lahrmann *et al.,* 2013). Transcriptome analysis of root colonization in these plants shows the expression of common and specific sets of SSP genes. Although the identified SSPs are different in these different fungal species, there is a common trend for their expression during infection, particularly during the establishment of biotrophic interactions, suggesting that they may participate in defining the level of specificity with different host plants.

MATERIAL & METHODS

Experimental setup and inoculation procedures

Specific *M. larici-populina* isolate 98AG31 stages were collected on the aecial host larch (*Larix decidua*) following the procedure reported in Pernaci *et al.* (2014). The Fig. 1 describes the experimental set up and the Fig. S1 details the overall procedure used to produce the biological material. The basidial stage was collected from dead poplar leaves and contained remains of

telia and freshly produced basidia and basidiospores. Large areas of dead leaves fully covered with yellow fungal material were targeted. The pycnial stage was collected from larch needles 7 dpi with basidiospores and are made up of different cell types such as *in planta* infection hyphae, flexuous hyphae and pycniospores. The aecial stage was collected from larch needles two weeks after inoculation and corresponded to freshly produced aecia and remaining pycnia. For both the pycnial and the aecial stages, sections of leaves fully covered with apparent fungal material were targeted and portions of leaves with no apparent fungus were removed. Three biological replicates were harvested at each stage; the basidial replicates came from different poplar leaves, while the pycnial and aecial replicates were produced from different boxes of inoculated larch seedlings. For each replicate (i.e. R1, R2, R3), the aecial and pycnial stages were harvested from different larch seedlings present in the same inoculated box, ensuring synchronization of these stages after a single initial basidiospore inoculation.

RNA isolation and Illumina RNA-Sequencing

Total RNA from the basidial stage samples were isolated with the RNeasy Plant Minikit (QIAGEN, Courtaboeuf, France) using RLC buffer with β-mercaptoethanol following manufacturer's recommendations. For each basidial replicate, several rounds of RNA isolation were performed using low inputs of material (10 to 20mg) to avoid RNA degradation from decayed leaf material, and RNA samples representing a total of 100mg of material were pooled together. Total RNA from the pycnial and the aecial stages were isolated from 100mg of infected larch needles using a protocol adapted to RNA isolation from conifers (Chang, Puryear, & Cairney, 1993). A DNase I (QIAgen) treatment was applied to all samples according to the manufacturer's recommendations to eliminate traces of genomic DNA. Electrophoretic RNA profiles and RNA quality were assessed with the Experion analyser using the Experion Standard-sens analysis kit (Bio-Rad, Marnes la Coquette, France) before sample shipment and in Beckman's facilities (Beckman Coulter Genomics, Grenoble, France) prior RNA-sequencing. The following RIN numbers were recorded before Illumina libraries preparation: 8.6, 8.3, 8.8, 6.9, 7, 6.8, 7.6, 8.6 and 8.3, corresponding to replicates R1, R2, R3 of basidial, pycnial and aecial samples, respectively. cDNA synthesis and RNA-Seq library construction were performed by Beckman Coulter Genomics using the Illumina TruSEQ RNA seq kit following the manufacturer's recommendations. Each library was quantified by qPCR and sequenced with the Illumina HiSeq2500 platform as paired-end 100nt reads. Data acquisition was performed in Beckman Coulter Genomics facilities according to standard Illumina's procedures. On average, 114 million paired-end reads were produced per sample (ranging from 95 to 129 million reads per sample, Table S1). The complete expression dataset is available at the NCBI Gene Expression Omnibus repository under the series number GSE106863.

Reference genome and annotation

The *M. larici-populina* isolate 98AG31 genome (version 1.0; US Department of Energy Joint Genome Institute; http://genome.jgi.doe.gov/Mellp1/Mellp1.home.html; Duplessis *et al.,* 2011a) was used as a reference. The RNAseq analysis was performed using the catalog of 16,399 *M. larici-populina* isolate 98AG31 reference transcripts. Automatic annotations such as Gene ontology (GO), Kyoto Eukaryotic Genes and Genomes (KEGG) and euKaryotic Orthologous Group (KOG) annotations were recovered from the reference genome website. The expert manual annotation of specific gene categories (e.g. carbohydrate-active enzymes, CAZymes; lipases; proteases; transporters) performed by the poplar rust genome consortium was used (Duplessis *et al.,* 2011a). The repertoire of SSPs previously described (Hacquard *et al.,* 2012) was extended to include all SPs of unknown function independently of their protein length as described in Lorrain *et al.* (2015).

Gene expression analysis

All reads per sample were mapped onto the 16,399 *M. larici-populina* transcripts using the RNAseq analysis tool of CLC Genomics Workbench 6.5.1 (QIAgen) with default parameters, except for the similarity fraction which was set at 90% across 90% of the read length. Reads showed an average per base Phred quality score higher than 30 over their 100 nt length in all samples and no trimming was applied before mapping. The number of unique reads mapped per transcript (counts) was recorded for each sample and the summary of mapping statistics is presented in Table S1. A saturation curve analysis was performed with the rarefy function of the VEGAN package (Oksanen *et al.,* 2017) to determine whether the sequencing depth of our data was sufficient. R packages DESeq2 (v1.10.1) and edgeR (3.12.0) were used to conduct differential expression analysis of *M. larici-populina* transcripts (Love *et al.,* 2014; Robinson *et al.*, 2010). Counts were normalized using the size factor method proposed by (Anders & Huber, 2010). Differentially expressed genes (DEGs) were then identified using both packages (padj<0.05) and the intercept was considered for further analyses with DESeq2 count values as the reference dataset. The R package WGCNA was used to generate clusters of co-expression (R script available on demand) among significantly DEGs (Langfelder & Horvath, 2008). For this purpose, only DEGs found between all three stages were considered.

KOG enrichment analysis

KOG enrichment analysis was performed as previously described in Hacquard *et al*. (2013). Secreted proteins of unknown function were extracted from the KOG categories “No Hit” and “Function unknown” to define a new KOG category that we termed “Secreted proteins”. Over-represented KOG categories in the basidial, the pycnial and the aecial stages up- or down-

regulated genes (3-fold change significantly DEGs) were determined as respect to the global gene distribution. Significance of over-represented KOG categories in each condition was estimated using the Fisher's Exact Test and False Discovery Rate (FDR) corrected by Benjamini-Hochberg Test (FDR < 0.05).

Comparison between *M. larici-populina* larch and poplar infection transcriptomes

Expression data for *M. larici-populina* dormant and germinating urediniospores, and for time-course infection of poplar leaves were recovered from a previous transcriptome analysis performed with whole-genome custom oligoarrays (Duplessis *et al.,* 2011b). Expression data were recovered for 14,883 *M. larici-populina* transcripts showing expression values higher than the background level. The expression sets were reduced for comparison purpose as follows: expression values in dormant and germinating urediniospores were grouped in a unique "Urediniospores" stage (highest expression value of the two stages retained) and expression values during the poplar leaf time-course infection at 24, 48, 96 and 168 hours post-inoculation (hpi) were pooled into a unique "Poplar" stage (highest expression value during the time-course retained). Larch infection expression datasets were similarly reduced into two sets: expression value in the basidial stage was considered as a unique larch-infecting spores set and expression values during larch needles infection (i.e. pycnial and aecial stages) were grouped into a unique "Larch" stage (highest expression value retained). We used a quantile normalization to estimate and compare transcript distributions within the oligoarrays and RNAseq datasets with the R package Limma (Ritchie *et al.,* 2015). Regarding the specific gene families comparison, some gene members or gene families had non-available expression records from oligoarrays. This was due to the oligonucleotide inability to discriminate between transcript species. These genes were thus not included in the comparative expression analysis.

RTqPCR

A random set of 22 SSP genes was selected to assess RNAseq expression profiles by RT-qPCR. Primer sequences and efficiency are presented in Table S4. Primer design, first-strand cDNA synthesis and PCR amplification were performed as previously described (Duplessis *et al.,* 2011b; Hacquard *et al.,* 2010). Transcript expression levels were normalized with *M. larici-populina* alpha-tubulin and elongation factor reference genes (Hacquard *et al.,* 2011b), considering the specific efficiency of each primer (Pfaffl, 2001). For each gene, in order to compare profiles measured by oligoarrays and RTqPCR, expression ratios were calculated between relative levels in each experimental condition and the average expression level across conditions.

ACKNOWLEDGEMENTS

The authors would like to acknowledge the 'Investissements d'Avenir' program (ANR-11-LABX-0002-01, Lab of Excellence ARBRE) and the Joint Genome Institute (Office of Science of the U.S. Department of Energy under contract no. DE-Ac02-05cH11231) for the sequencing and annotation of the genome of the poplar rust fungus *Melampsora larici-populina* and access made through the mycocosm portal. Emmanuelle Morin and Annegret Kohler at INRA Nancy are acknowledged for their bioinformatic support. Constructive remarks and suggestions from the editor and two anonymous reviewers, as well as from Benjamin Schwessinger are also acknowledged. Cecile Lorrain is supported by a young scientist grant (CJS) from INRA. Sébastien Duplessis, Arnaud Hecker and Pascal Frey planned and designed the research; Christine Delaruelle, Stéphane Hacquard and Jérémy Pétrowski performed experiments; Cécile Lorrain, Clémence Marchal, Benjamin Petre and Sébastien Duplessis analysed data; Cécile Lorrain, Clémence Marchal, Benjamin Petre and Sébastien Duplessis wrote the manuscript; all authors read and approved the final manuscript.

LITERATURE CITED

Aime, M., Toome, M., and McLaughlin, D. 2014. The Pucciniomycotina. In The Mycota VII Part A, (McLaughlin D, Spatafora JW), pp. 271–294.

Aime, M.C., MacTaggart A.R, Mondo, S.J., Duplessis, S. 2017. Phylogenetics and phylogenomics of rust fungi. Adv. Genet. 100: 267-307.

Anders, S., and Huber, W. 2010. Differential expression analysis for sequence count data. Genome Biol. 11: R106–R106.

Bakkeren, G., Joly, D.L., and Duplessis, S. 2016. Editorial: Genomics Research on Non-model Plant Pathogens: Delivering Novel Insights into Rust Fungus Biology. Front. Plant Sci. 7: 216.

Cantu, D., Segovia, V., MacLean, D., Bayles, R., Chen, X., Kamoun, S., Dubcovsky, J., Saunders, D.G.O., and Uauy, C. 2013. Genome analyses of the wheat yellow (stripe) rust pathogen *Puccinia striiformis* f. sp. *tritici* reveal polymorphic and haustorial expressed secreted proteins as candidate effectors. BMC Genomics 14: 270.

Chang, S., Puryear, J., and Cairney, J. 1993. A Simple and Efficient Method for Isolating RNA from Pine Tree. Plant Mol. Biol. Report. 11: 113-116.

Cummins, G., and Hiratsuka, Y. 2003. Illustrated genera of rust fungi. In Illustrated Genera of Rust Fungi, (APS, St Paul), pp. 1-225.

Cuomo, C.A., Bakkeren, G., Khalil, H.B., Panwar, V., Joly, D., Linning, R., Sakthikumar, S., Song, X., Adiconis, X., Fan, L., Goldberg, J.M., Levin, J.Z., Young, S., Zeng, Q., Anikster, Y., Bruce, M., Wang, M., Yin, C., McCallum, B., Szabo, L.J., Hulbert, S., Chen, X., Fellers, J.P. 2017. Comparative analysis highlights variable genome content of wheat rusts and divergence of the mating loci. G3 7: 361-376.

De Carvalho, M.C. da C.G., Nascimento, L.C., Darben, L.M., Polizel-Podanosqui, A.M., Lopes-Caitar, V.S., Qi, M., Rocha, C.S., Carazzolle, M.F., Kuwahara, M.K., Pereira, G.A.G., Abdelnoor, R.V., Whitham, S.A., Marcelino-Guimarães, F.C. 2016. Prediction of the in planta *Phakopsora pachyrhizi* secretome and potential effector families. Mol. Plant Pathol. 18: 363-377.

Dobon, A., Bunting, D.C.E., Cabrera-Quio, L.E., Uauy, C., and Saunders, D.G.O. 2016. The host-pathogen interaction between wheat and yellow rust induces temporally coordinated waves of gene expression. BMC Genomics 17: 380.

Dong, S., Remco, S., Cano, L.M., Song, S., Sklenar, J., Yoshida, K., Bozkurt, T.O., Oliva, R., Liu, Z., Tian, M., Win, J., Banfield, M.J., Jones, A.M.E., van der Hoorn, R.A.L., Kamoun, S. 2014. Effector specialization in a lineage of the Irish Potato famine pathogen. Science 343: 552-555.

Duplessis, S., Cuomo, C. A., Lin, Y, C., Aerts, A., Tisserant, E., Veneault- Fourrey, C., Joly, D. L., Hacquard, S., Amselem, J., Cantarel, B., Chiu, R., Coutinho, P., Feau, N., Field, M., Frey, P., Gelhaye, E., Goldberg, J., Grabherr, M., Kodira, C., Kohler, A., Hües, U., Lindquist, E. A., Lucas, S., Mago, R., Mauceli, E., Morin, E., Murat, C., Pangilinan, J. L., Park, R., Pearson, M., Quesneville, H., Rouhier, N., Sakthikumar, S., Schmutz, J., Selles, B., Shapiro, H., Tanguay, P., Tuskan, G.A., Henrissat, B., Van de Peer, Y., Rouzé, P., Ellis, J. G., Dodds, P. N., Schein, J. E., Zhong, S., Hamelin, R. C., Grigoriev, I. V., Szabo, L. J., and Martin, F. 2011a. Obligate biotrophy features unravelled by the genomic analysis of rust fungi. Proc. Natl. Acad. Sci. U.S.A. 108: 9166-9171.

Duplessis, S., Hacquard, S., Delaruelle, C., Tisserant, E., Frey, P., Martin, F., and Kohler, A. (2011b). *Melampsora larici-populina* transcript profiling during germination and timecourse infection of poplar leaves reveals dynamic expression patterns associated with virulence and biotrophy. Mol. Plant-Microbe Interact. 24: 808-818.

Duplessis, S., Joly, D.L., and Dodds, P.N. 2012. Rust Effectors. In Effectors in Plant-Microbe Interactions, (John Wiley & Sons Inc.), pp. 155-193.

Duplessis, S., Bakkeren, G., and Hamelin, R. 2014. Advancing knowledge on biology of rust fungi through genomics. Adv. Bot. Res. 70: 173-209.

Germain, H., Joly, D.L., Mireault, C., Plourde, M.B., Letanneur, C., Stewart, D., Morency, M.-J., Petre, B., Duplessis, S., and Seguin, A. 2017. Infection assays in *Arabidopsis* reveal candidate effectors from the poplar rust fungus that promote susceptibility to bacteria and oomycete pathogens. Mol. Plant Pathol. 19: 191-200.

Hacquard, S., Delaruelle, C., Legue, V., Tisserant, E., Kohler, A., Frey, P., Martin, F., and Duplessis, S. 2010. Laser capture microdissection of uredinia formed by *Melampsora larici-populina* revealed a transcriptional switch between biotrophy and sporulation. Mol. Plant-Microbe Interact. MPMI 23: 1275-1286.

Hacquard, S., Petre, B., Frey, P., Hecker, A., Rouhier, N., and Duplessis, S. 2011a. The poplar-poplar rust interaction: insights from genomics and transcriptomics. J. Pathog. 2011: 716041.

Hacquard, S., Veneault-Fourrey, C., Delaruelle, C., Frey, P., Martin, F., and Duplessis, S. 2011b. Validation of *Melampsora larici-populina* reference genes for in planta RT-quantitative PCR expression profiling during time-course infection of poplar leaves. Physiol. Mol. Plant Pathol. 75: 106-112.

Hacquard, S., Joly, D.L., Lin, Y.-C., Tisserant, E., Feau, N., Delaruelle, C., Legue, V., Kohler, A., Tanguay, P., Petre, B., Frey, P., Van de Peer, Y., Rouzé, P., Martin F., Hamelin, R.C., and Duplessis, S. 2012. A comprehensive analysis of genes encoding small secreted proteins identifies candidate effectors in *Melampsora larici-populina* (poplar leaf rust). Mol. Plant-Microbe Interact. MPMI 25, 279–293.

Hacquard, S., Delaruelle, C., Frey, P., Tisserant, E., Kohler, A., and Duplessis, S. 2013. Transcriptome analysis of poplar rust telia reveals overwintering adaptation and tightly coordinated karyogamy and meiosis processes. Front. Plant Sci. 4: 456.

Harder, D. 1984. Developmental ultrastructure of hyphae and spores. In The Cereal Rust, (Bushnell W. R., Roelfs A. P), pp. 322-361.

van der Heijden, M.G.A., Martin, F.M., Selosse, M.-A., and Sanders, I.R. 2015. Mycorrhizal ecology and evolution: the past, the present, and the future. New Phytol. 205: 1406-1423.

Henrissat, B., Surolia, A., and Stanley, P. 2015. A Genomic View of Glycobiology. In Essentials of Glycobiology, A. Varki, R.D. Cummings, J.D. Esko, et al., eds. (Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press) 2015-2017.

Huang, X., Chen, X., Coram, T., Wang, M., Kang, Z. 2011. Gene expression profiling of *Puccinia striiformis* f. sp. *tritici* during development reveals a highly dynamic transcriptome. J Genet Genomics 38: 357-71.

Jing, L., Guo, D., Hu, W., and Niu, X. 2017. The prediction of a pathogenesis-related secretome of *Puccinia helianthi* through high-throughput transcriptome analysis. BMC Bioinformatics 18: 166.

Joly, D.L., Feau, N., Tanguay, P., and Hamelin, R.C. 2010. Comparative analysis of secreted protein evolution using expressed sequence tags from four poplar leaf rusts (*Melampsora* spp.). BMC Genomics 11: 422.

Kamel, L., Tang, N., Malbreil, M., San Clemente, H., Le Marquer, M., Roux, C., and Frei dit Frey, N. 2017. The comparison of expressed candidate secreted proteins from two arbuscular mycorrhizal fungi unravels common and specific molecular tools to invade different host plants. Front. Plant Sci. 8: 124.
Kohler, A., Kuo, A., Nagy, L.G., Morin, E., Barry, K.W., Buscot, F., Canbäck, B., Choi, C., Cichocki, N., Clum, A., Colpaert, J., Copeland, A., Costa, M.D., Doré, J., Floudas, D., Gay, G., Girlanda, M., Henrissat, B., Herrmann, S., Hess, J., Högberg, N., Johansson, T., Khouja, H-R., LaButti, K., Lahrmann, U., Levasseur, A., Lindquist, E.A., Lipzen, A., Marmeisse, R., Martino, E., Murat, C., Ngan, C.Y., Nehls, U., Plett, J.M., Pringle, A., Ohm, R.A., Perotto, S., Peter, M., Riley, R., Rineau, F., Ruytinx, J., Salamov, A., Shah, F., Sun, H., Tarkka, M., Tritt, A., Veneault-Fourrey, C., Zuccaro, A., Mycorrhizal Genomics Initiative Consortium, Tunlid, A., Grigoriev, I.V., Hibbett, D.S., and Francis Martin. 2015. Convergent losses of decay mechanisms and rapid turnover of symbiosis genes in mycorrhizal mutualists. Nature Genet. 47: 410-415.
Langfelder, P., and Horvath, S. 2008. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics 9, 559.
Lahrmann, U., Ding, Y., Banhara, A., Rath, M., Hajirezaei, M. R., Döhlemann, S., von Wirén, N., Parniske, M., and Zuccaro, A. 2013. Host-related metabolic cues affect colonization strategies of a root endophyte. Proc Natl Acad Sci USA 110: 13965-13970.
Liao, H.-L., Chen, Y., and Vilgalys, R. 2016. Metatranscriptomic Study of Common and Host-Specific Patterns of Gene Expression between Pines and Their Symbiotic Ectomycorrhizal Fungi in the Genus Suillus. PLOS Genet. 12: e1006348.
Littlefield, L.J., and Heath, M.C. 1979. Ultrasrtucure of Rust Fungi. Academic Press, New-York.
Liu, J.-J., Sturrock, R.N., Sniezko, R.A., Williams, H., Benton, R., and Zamany, A. 2015. Transcriptome analysis of the white pine blister rust pathogen *Cronartium ribicola*: *de novo* assembly, expression profiling, and identification of candidate effectors. BMC Genomics 16: 678.
Lo Presti, L., Lanver, D., Schweizer, G., Tanaka, S., Liang, L., Tollot, M., Zuccaro, A., Reissmann, S., and Kahmann, R. 2015. Fungal effectors and plant susceptibility. Annu. Rev. Plant Biol. 66: 513-545.
Lorrain, C., Hecker, A., and Duplessis, S. 2015. Effector-Mining in the Poplar Rust Fungus *Melampsora larici-populina* Secretome. Front. Plant Sci. 6: 1051.
Love, M.I., Huber, W., and Anders, S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 15: 550.
Martin, F., Kohler, A., Murat, C., Veneault-Fourrey, C., and Hibbett, D.S. 2016. Unearthing the roots of ectomycorrhizal symbioses. Nat. Rev. Microbiol. 14: 760-773.
Mendgen, K. 1984. Development and physiology of teliospores. In The Cereal Rust, (Bushnell W. R., Roelfs A. P), pp. 362-384.
Oksanen, J., Blanchet, G.F., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., Minchin, P.R., O'Hara, R.B., Simpson, L., Solymos, P., Stevens, M.H.H.S., Szoecs, E., and Wagner, H. 2017. The vegan package. Community Ecol. Package.
Pernaci, M., De Mita, S., Andrieux, A., Petrowski, J., Halkett, F., Duplessis, S., and Frey, P. 2014. Genome-wide patterns of segregation and linkage disequilibrium: the construction f a linkage genetic map of the poplar rust fungus *Melampsora larici-populina*. Front. Plant Sci. 5: 454.
Petre, B., Morin, E., Tisserant, E., Hacquard, S., Da Silva, C., Poulain, J., Delaruelle, C., Mrtin, F., Rouhier, N., Kohler, A., and Duplessis, S. 2012. RNA-Seq of early-infected poplar leaves by the rust pathogen *Melampsora larici-populina* uncovers PtSultr3;5, a fungal-induced host sulfate transporter. PloS One 7: e44408.
Petre, B., Joly, D.L., and Duplessis, S. 2014. Effector proteins of rust fungi. Front Plant Sci 5: 416.
Petre, B., Saunders, D.G.O., Sklenar, J., Lorrain, C., Win, J., Duplessis, S., and Kamoun, S. 2015. Candidate effector proteins of the rust pathogen *Melampsora larici-populina* target diverse plant cell compartments. Mol. Plant-Microbe Interact. 28: 689-700.
Petre, B., Lorrain, C., Saunders, D.G.O., Win, J., Sklenar, J., Duplessis, S., and Kamoun, S. 2016. Rust fungal effectors mimic host transit peptides to translocate into chloroplasts. Cell. Microbiol. 18: 453-465.
Pfaffl, M.W. 2001. A new mathematical model for relative quantification in real-time RT-PCR. Nucleic Acids Res. 29: e45.

Pinon, J., and Frey, P. 2005. Interactions between poplar clones and *Melampsora* populations and their implications for breeding for durable resistance. In Rust Diseases of Willow and Poplar, (CAB International, Wallingford), pp. 139-154.
Plett, J.M., Tisserant, E., Brun, A., Morin, E., Grigoriev, I.V., Kuo, A., Martin, F., and Kohler, A. 2015. The mutualist *Laccaria bicolor* expresses a core gene regulon during the colonization of diverse host plants and a variable regulon to counteract host-specific defenses. Mol. Plant-Microbe Interact. 28: 261-273.
Ragazzi, A., Longo, N., Naldini, B., Moricca, S., and Dellavalle, I. 2005. Basidiospore-derived penetration by species of *Cronartium* and *Melampsora* an outline. In Rust Diseases of Willow and Poplar, (M.H Pei and A.R McCracken), pp. 161-174.
Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K. 2015. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 43: e47.
Robinson, M.D., McCarthy, D.J., and Smyth, G.K. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26: 139-140.
Rochi, L., Dieguez, M.J., Burguener, G., Darino, M.A., Pergolesi, M.F., Ingala, L.R., Cuyeu, A.R., Turjanski, A., Kreff, E.D., and Sacco, F. 2016. Characterization and comparative analysis of the genome of *Puccinia sorghi* Schwein, the causal agent of maize common rust. Fungal Genet. Biol. In press
Rovenich, H., Boshoven, J.C., and Thomma, B.P.H.J. 2014. Filamentous pathogen effector functions: of pathogens, hosts and microbiomes. Curr. Opin. Plant Biol. 20: 96-103.
Rutter, W.B., Salcedo, A., Akhunova, A., He, F., Wang, S., Liang, H., Bowden, R.L., and Akhunov, E. 2017. Divergent and convergent modes of interaction between wheat and *Puccinia graminis* f. sp. *tritici* isolates revealed by the comparative gene co-expression network and genome analyses. BMC Genomics 18: 291.
Schenk, P.M., Carvalhais, L.C., and Kazan, K. (2012). Unraveling plant-microbe interactions: can multi-species transcriptomics help? Trends Biotechnol. 30: 177-184.
Schulze-Lefert, P., and Panstruga, R. 2011. A molecular evolutionary concept connecting nonhost resistance, pathogen host range, and pathogen speciation. Trends Plant Sci. 16: 117-125.
Sharpee, W.C., and Dean, R.A. 2016. Form and function of fungal and oomycete effectors. Fungal Biol. Rev. 30: 62-73.
Sperschneider, J., Dodds, P.N., Gardiner, D.M., Manners, J.M., Singh, K.B., and Taylor, J.M. 2015. Advances and challenges in computational prediction of effectors from plant pathogenic fungi. PLoS Pathog. 11: e1004806.
Sperschneider, J., Dodds, P.N., Taylor, J.M., and Duplessis, S. 2017. Computational Methods for Predicting Effectors in Rust Pathogens. Methods Mol. Biol. Clifton NJ 1659: 73-83.
Talhinhas, P., Azinheira, H.G., Vieira, B., Loureiro, A., Tavares, S., Batista, D., Morin, E., Petitot, A.S., Paulo, O.S., Poulain, J., Da Silva, C., Duplessis, S., Silva, Mdo. C., and Fernandez, D. 2014. Overview of the functional virulent genome of the coffee leaf rust pathogen *Hemileia vastatrix* with an emphasis on early stages of infection. Front Plant Sci 5: 88.
Tan, K.-C., Ipcho, S.V.S., Trengove, R.D., Oliver, R.P., and Solomon, P.S. 2009. Assessing the impact of transcriptomics, proteomics and metabolomics on fungal phytopathology. Mol. Plant Pathol. 10: 703-715.
Tao, S.-Q., Cao, B., Tian, C.-M., and Liang, Y.M. 2017. Comparative transcriptome analysis and identification of candidate effectors in two related rust species (*Gymnosporangium yamadae* and *Gymnosporangium asiaticum*). BMC Genomics 18: 651.
Tyler, B.M., and Rouxel, T. 2012. Effectors of fungi and oomycetes: their virulence and avirulence functions and translocation from pathogen to host cells. In Molecular Plant Immunity, (Wiley-Blackwell), pp. 123-167.

Vialle, A., Frey, P., Hambleton, S., Bernier, L., and Hamelin, R. 2011. Poplar rust systematics and refinement of *Melampsora* species delineation. Fungal Div. 50: 227-248.

Voegele, R.T., Hahn, M., and Mendgen, K. 2009. The Uredinales: cytology, biochemistry, and molecular biology. In The Mycota V Plant Relationships, (Berlin: Springer), pp. 79-94.

Win, J., Chaparro-Garcia, A., Belhaj, K., Saunders, D.G., Yoshida, K., Dong, S., Schornack, S., Zipfel, C., Robatzek, S., Hogenhout, S.A., and Kamoun, S. 2012. Effector biology of plant-associated organisms: concepts and perspectives. Cold Spring Harb Symp. Quant. Biol.: 77.

Xu, J., Linning, R., Fellers, J., Dickinson, M., Zhu, W., Antonov, I., Joly, D.L., Donaldson, M.E., Eilam, T., Anikster, Y., Banks, T., Munro, S., Mayo, M., Wynhoven, B., Ali, J., Moore, R., McCallum, B., Borodovsky, M., Saville, B., and Bakkeren, G. 2011. Gene discovery in EST sequences from the wheat leaf rust fungus *Puccinia triticina* sexual spores, asexual spores and haustoria, compared to other rust and corn smut fungi. BMC Genomics 12: 161.

SUPPLEMENTAL TABLES

Table S1: General mapping information of RNA-sequencing data against *Melampsora larici-populina* genes.

Illumina reads of the three biological replicates of the basidial, pycnial and aecial stages were compared to the catalogue of predicted *M. larici-populina* transcripts. The initial number of reads, the number of reads mapped in pair, the total number of counted fragments and the number of unique matches are shown. The total number of reads across all nine data points is also indicated.

Sample	Number of reads	Number of reads mapped in pairs	Number of Counted fragments	Number of unique matches
Basidial stage 1	110 899 188	60 070 636 (54.17%)	66 678 884 (60.13%)	65 756 953 (59.29%)
Basidial stage 2	109 217 478	58 638 478 (53.69%)	64 834 906 (59.36%)	63 902 311 (58.51%)
Basidial stage 3	109 910 810	58 658 710 (53.37%)	64 917 365 (59.06%)	64 028 793 (58.26%)
Pycnial stage 1	134 211 440	12 322 320 (9.18%)	14 592 204 (10.87%)	14 244 025 (10.61%)
Pycnial stage 2	128 967 964	11 434 820 (8.87%)	13 600 883 (10.55%)	13 274 828 (10.29%)
Pycnial stage 3	117 849 960	11 825 990 (10.03%)	13 999 827 (11.88%)	13 674 207 (11.60%)
Aecial stage 1	112 726 392	19 068 802 (16.92%)	22 502 082 (19.96%)	21 906 385 (19.43%)
Aecial stage 2	110 931 202	17 322 076 (15.62%)	20 231 372 (18.24%)	19 644 706 (17.71%)
Aecial stage 3	95 589 730	14 338 304 (15%)	16 923 032 (17.70%)	16 610 790 (17.38%)
Total	1 030 304 164			

Table S2: *Melampsora larici-populina* transcripts annotation and expression information.

For each of the 16,399 *M. larici-populina* transcripts, the following annotation and expression information are provided: JGI ProteinID, protein size (amino acids), manually annotated categories (SP/SSP, secreted proteins/small secreted proteins; CAZyme, carbohydrate active enzymes; CytP450, cytochrome-P450), category information (details about subfamilies within a given category according to specific databases, i.e. CAZY for cazymes, MEROPS for proteases and TransportDB for transporters), SSP families with numbers of members (according to Hacquard et al. (2012)), gene ontology (GO) annotation, Kyoto encyclopedia of genes and genomes (KEGG) annotation, eukaryotic orthologous groups annotation (KOG; ID, description and class), expression in urediniospores (USP), poplar, basidial (BSD), pycnial (PYC) and aecial (AEC) stages (expression in USP and poplar corresponds to the highest expression level in dormant and germinated urediniospores and in a time-course infection of Beaupré leaves, respectively, according to Duplessis et al. (2011b) and larch corresponds to the highest expression level at the pycnial and the aecial stages), quantile-normalized expression values for USP, poplar, BSD and larch stages (USP_norm, POPLAR_norm, BSD_norm and LARCH_norm, respectively). In columns for USP and poplar, the red color indicates expression value below background. Automatic annotations for GO, KEGG and KOG databases were retrieved from the *M. larici-populina* genome webportal (https://genome.jgi.doe.gov/Mellp1/Mellp1.home.html) and manual annotations were defined according to Duplessis et al. 2011a.

TableS3: *Melampsora larici-populina* differentially expressed genes between the three larch infection-related stages.

The table presents the fold change expression levels between the three larch-infection basidial, (BSD), pycnial (PYC) and aecial (AEC) stages for genes found significantly differentially expressed between at least two stages with DESeq2 (adjusted *p*-value *padj* <0.05). JGI *M. larici-populina* proteinID, log2FoldChange, FoldChange and padj are detailed.

Table S4: RT-qPCR primers designed to determine expression profiles *of Melampsora larici-populina* SSP genes.

Specific 5' and 3' primers for 22 *M. larici-populina* SSP genes are detailed, along with the amplicon length (bp) and primers efficiency (%).

Protein ID	Length (bp)	Efficiency (%)	5' primer	3' primer
70394	152	83	TATCGCCACCCTCTTCACTC	GACAACCGGGATCAGACATT
71399	205	84	CCCATCTCGTTACCGGACTT	GCCAGGTTGGTCCTTGTATG
71402	212	94	GTATGCTCATGCCTGTGGTC	TGATGCCGTCCTTAATAGCC
72734	248	87	AACCGGCTCCTGCTAAGAAA	AGGGTAGGACGAAGTGCTGA
74496	163	100	CCTCAACGATTCACATCTCC	TATGGATTGCCAGGTTGGTC
93569	197	88	TACACCTCAATCGGCATCAC	CGGCGTCCAAAGAGTTCAG
101611	219	100	GATGAGAAGCCGGACACATT	ATCCCCATGCCTACATTCCT
114961	109	74	CCGACTTCGCTTGTCTTTGT	ACACCCCAAGCAGTAGCAG
114535	130	85	GGGAGAAACCAAAGGATTGTG	GTTGAGCAAACTAAGGATGGAGA
123584	130	101	CGTGCTGGACACTTGCTCT	TTGAACCACGGCTTTGATAGT
123735	227	94	AAATGTGGTGGCTGTTGTGA	GCTCTTGTGGAGACCCTTTG
124071	162	98	GCCAGCGTACTTCAACAACC	GACCCTCCTCCAAAGTCACA
124123	150	99	GCCAAGATGTTCGTTCTGGA	TACTATCAGGTGGAGGGGAATG
124174	127	94	AATGGGCCCAAATGGATA	TACCGTCACGGCAATTTGTA
124256	107	97	CCAAATGCAGCGTAGTGAA	ACCTACCAGTAGGAGCACAACC
124264	84	109	TCGTGGTGAGGAGGATTTCTG	CTACCACAAATACCACAACCTG
124266	102	107	ATTGCGAGGGGTGTTCAG	GAGAGCTCAATACACATACACCA
124272	120	102	GTCTCGAAGACGGTGACGAT	CGGCTCCAGTATCACAATCA
124497	143	93	AGCAATCTCTGTGGCATTCA	CATCCGGTTCTGTCGTAGGT
124499	139	93	TGACTCACAGTGTCAAGAGCAA	CGGTAGTGATAGGGGCAAAG
124509	68	93	ACTCACAGCTCCAAGAGCAAA	TGTGGACAGTAATCCCTTTCA
124518	188	98	GCCGTCTTCATGGTATTGGT	GGGCATTCTACATCAGTCCAG

SUPPLEMENTAL FIGURES

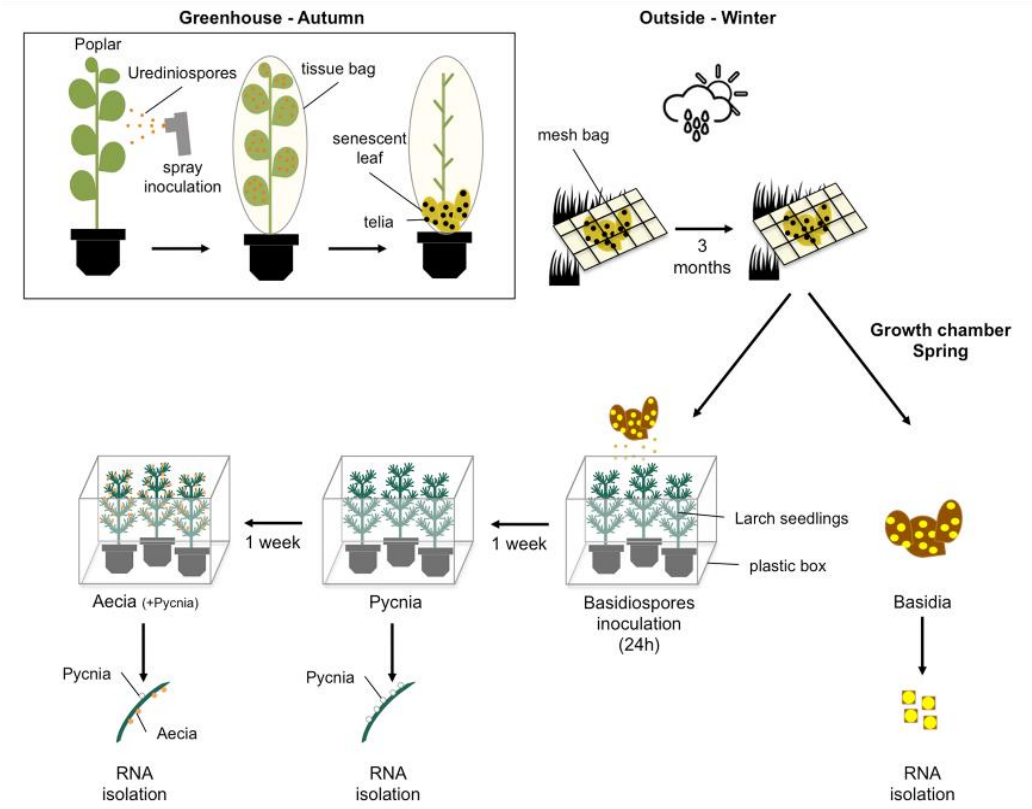


Figure S1. Experimental procedure for basidia production, inoculation of larch seedlings and obtention of pycnia and aecia of *Melampsora larici-populina* for RNA isolation according to Pernaci et al. 2014

Leaves of young poplar 'Robusta' cv. plants grown in greenhouses were inoculated with urediniospores of *M. larici-populina* isolate 98AG31 in September. After dehiscence, senescent infected leaves were let stand in tissue bags until telia developed. Leaves were then collected and placed outside in mesh bags to let teliospores overwinter under natural conditions. Leaves were used the following spring as primary source of basidiospore inoculum to infect young larch seedlings. From April to May, dead poplar leaves showing large black crust area corresponding to telia were placed in Petri dishes on moist paper at room temperature until yellow basidiospores can be seen at the surface of the leaves. A mixture of overwintered telia remains and freshly produced basidia and basidiospores were collected and snap frozen in liquid nitrogen. This corresponds to the basidial stage. *Larix decidua* seedlings were grown for three months in a greenhouse. Basidiospore-producing poplar leaves were placed over the larch seedlings in plastic boxes kept in the lab at room temperature in order to ensure infection of young soft needles. After one day, poplar leaves were removed to avoid continuous release of basidiospores over time. Larch needles showing high density of nectar droplets were collected and snap frozen in liquid nitrogen one week after inoculation. These samples contain different types of haploid fungal cells such as *in planta* infection hyphae and pycniospores (i.e. flexuous hyphae and pycniospores) and correspond to the pycnial stage. Boxes containing larch needles were kept closed most of the time and small flies (dark winged fungus gnats, Sciaridae, Diptera) naturally present in the substratum of the larch seedlings ensured fertilization in nectar droplets. Aecia and aeciospores were visible two weeks after inoculation on larch needles on the opposite side of pycnia. Needles showing a large number of aecia were collected and snap-frozen in liquid nitrogen (aecial stage). A few nectar droplets were still visible on some needles at this stage. Large areas covered with apparent fungal structures were targeted at the three stages in order to maximize the capture of fungal transcripts.

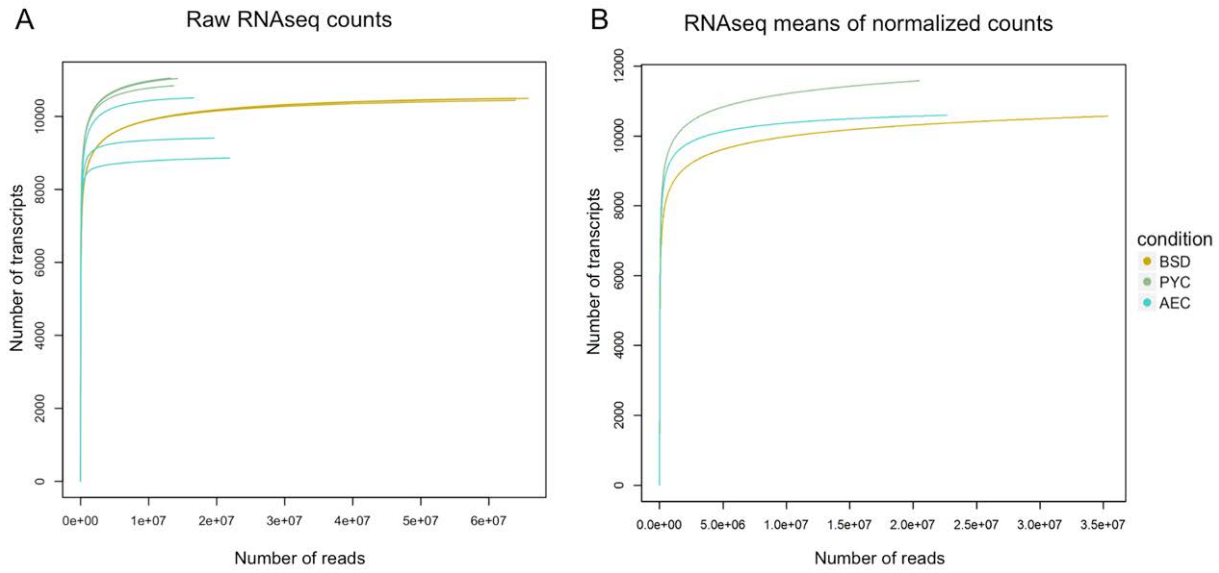


Figure S2: Sequencing saturation curves showing deep RNA sequencing of *Melampsora larici-populina* basidial, pycnial and aecial stages

The number of sequencing reads (x-axis) attributed to a given number of transcripts randomly selected from the total pool of sequencing reads (y-axis) is shown for the raw RNAseq counts of the three biological replicates of basidial (BSD), pycnial (PYC) and aecial (AEC) stages (A) and for the means of the normalized counts of the three biological replicates (B).

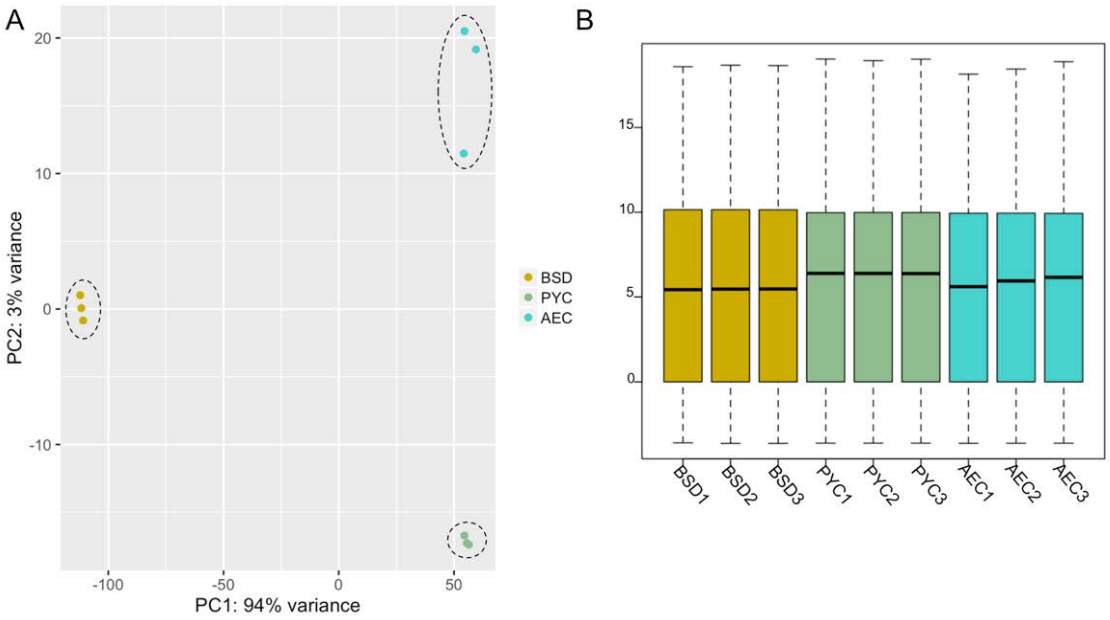


Figure S3: Comparison of RNAseq libraries between *Melampsora larici-populina* basidial (BSD), pycnial (PYC) and aecial (AEC) stages

(A) Principal component analysis (PCA) of *M. larici-populina* transcripts levels measured in BSD (gold), PYC (green) and AEC (blue) stages using RNA-sequencing. Reads detected per transcript (counts) were normalized using the size factor method proposed by Anders and Huber (2010) and rlog transformed before performing the PCA. The PCA plot places biological replicates along the two PC axes explaining 94% (x-axis) and 3% (y-axis) of the variance within samples. (B) Boxplots showing the distribution of sequencing counts for three biological replicates of basidial (gold), pycnial (green) and aecial (blue) stages. The top and bottom of boxes correspond to the 25 and 75 % quartiles, respectively. The middle line represents the median (50% quartile). Bottom whisker refers to the 1.5 interquartiles range of the lower quartile and top whisker indicates the 1.5 interquartile range of the upper quartile.

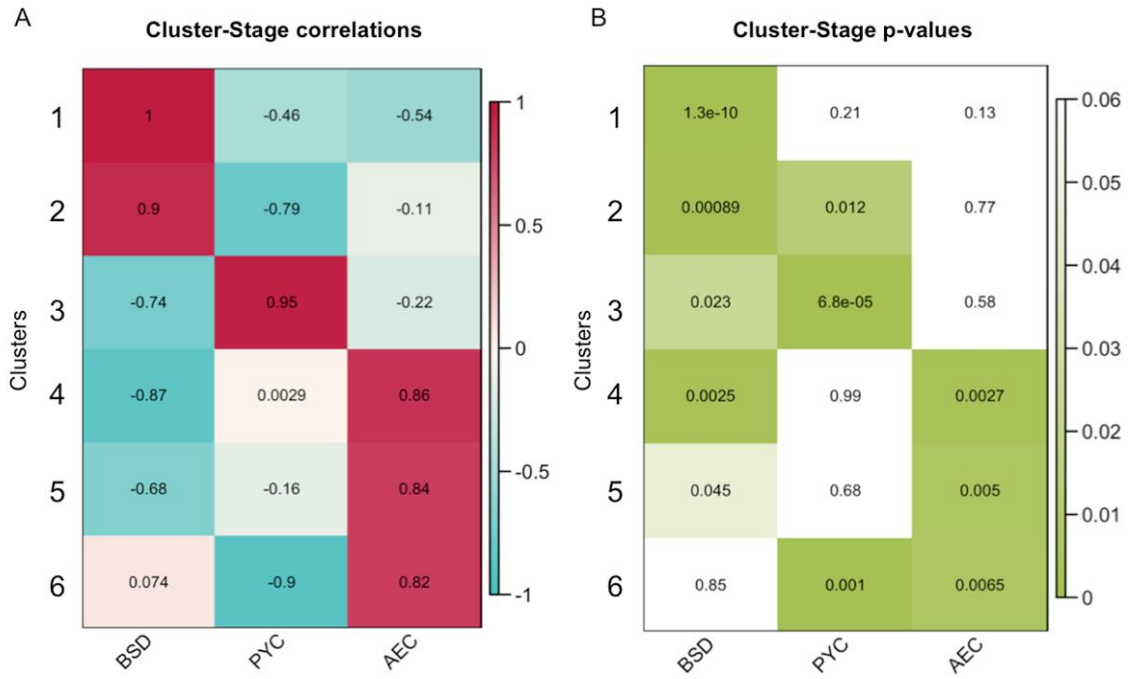


Figure S4: Expression patterns of six *Melampsora larici-populina* clusters of regulated genes
(A) Each row in the matrix corresponds to a cluster of co-expressed genes in each stage (BSD, PYC and AEC). Clusters are correlated from -1 (negative correlation in blue) to 1 (positive correlation in red). (B) *p*-values matrix of cluster-stage correlations with significant correlations (*p*-value<0.05) highlighted in green.

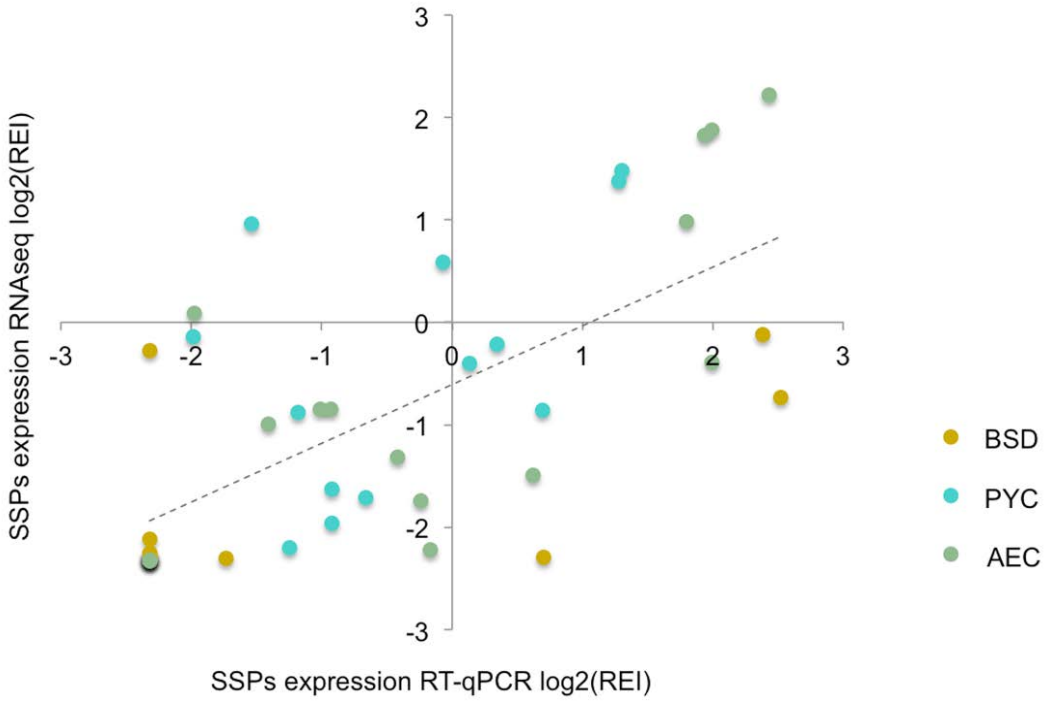


Figure S5: Comparison of SSP expression profiles monitored by RNAseq and RT-qPCR in basidial, pycnial and aecial stages

For each SSP, RT-qPCR expression levels were obtained from biological triplicates and normalized with *Mlp-a-TUB* and *Mlp-ELF1a* reference genes. For each gene, in order to compare profiles measured by RNAseq and RTqPCR, expression ratios were calculated between the relative expression level in each experimental condition and the average expression level across all conditions. Ratios were Log2-transformed for presentation.

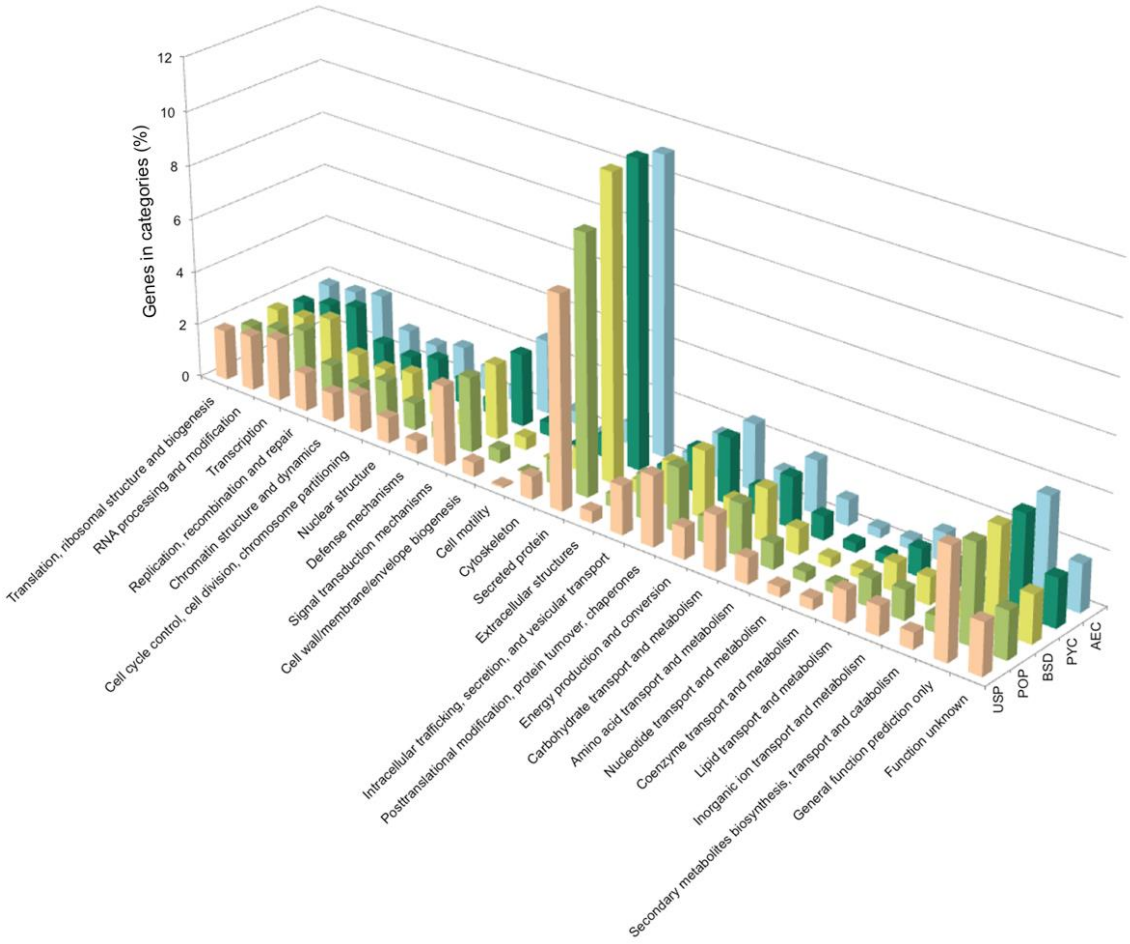


Figure S6: KOG categories distribution among expressed genes at different *Melampsora larici-populina* life cycle stages

Distribution of KOG categories among expressed genes detected at different stages of the *M. larici-populina* life cycle: USP, urediniospores; POP, infected poplar leaves; BSD, basidial stage; PYC, pycnial stage; AEC, aecial stage. The KOG category "No Hit" was the most abundant for all stages and was not included on graph for a better data representation. Note the inclusion of a specific category for "secreted protein (of unknown function as stated in the main text)". Distribution on the y-axis is provided as percentage values of the total number of genes detected at each stage. The lower proportion in the "secreted protein" category for USP and POP compared to other stages is due to non-available data on oligoarrays (i.e. gene families with no specific oligomers for given members).

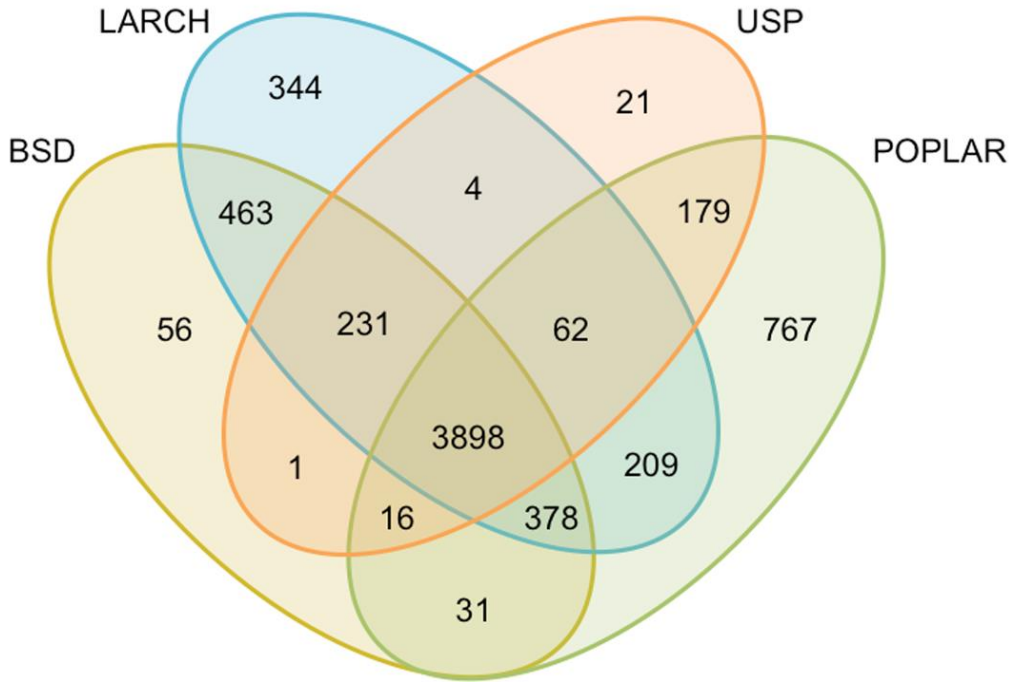


Figure S7: Comparison of expression profiles of genes of unknown function across *Melampsora larici-populina* life cycle

Venn diagram showing the number of genes of unknown function expressed in stages BSD (basidial stage), LARCH (pycnial and aecial stages, i.e. during larch needle infection), USP (dormant and germinating urediniospores) and POPLAR (during poplar leaf infection).

CHAPITRE III

Génomique des rouilles du peuplier

Partie 1 : Introduction

Malgré l'impact économique important des Pucciniales sur les productions végétales, il reste de nombreuses questions non-élucidées sur la biologie des Pucciniales et notamment les interactions avec leurs hôtes qui pourraient apporter un savoir indispensable pour le contrôle des maladies qu'elles provoquent en agriculture. Ces questions incluent : (i) Quand et comment les Pucciniales hétéroïques ont évolué, (ii) comment des Pucciniales phylogénétiquement proches ont évolué vers l'infection de plantes hôtes très éloignées phylogénétiquement, (iii), pourquoi et comment différents cycles de vies sont apparus chez les Pucciniales (microcyclique, macrocyclique, demi-cyclique, hémicyclique ; autoïque et hétéroïque) et, (iv) quels sont les mécanismes moléculaires impliqués dans le succès parasitaire de ces champignons pathogènes ? L'analyse des génomes des Pucciniales est un moyen de révéler les déterminants génétiques sous-jacents à ces différentes questions.

Les champignons et oomycètes possèdent des génomes de tailles vairiables. Jusqu'à maintenant, un nombre limité de génomes de grande taille ont été rapportés chez les champignons et parmi les plus notables, nous retrouvons l'Ascomycète *Tuber melanosporum* (la truffe noire du Périgord), des ascomycètes phytopathogènes responsables d'oïdium comme *Blumeria graminis*, *Golovinomyces orontii* ou *Erysiphe pisi* et les Pucciniales (Duplessis *et al.,* 2011a; Martin *et al.,* 2010; Spanu *et al.,* 2010; Wicker *et al.,* 2013). La teneur en éléments répétés comprenant notamment les éléments transposables, ainsi que le nombre de gènes font partie des facteurs pouvant expliquer cette variation importante de la taille des génomes fongiques (Kemen & Jones, 2012; Martin, 2014; Raffaele & Kamoun, 2012; Spanu & Panstruga, 2012). Il est intéressant de noter que parmi les génomes présentant les plus grandes tailles, on retrouve de nombreux champignons biotrophes et particulièrement biotrophes obligatoires (Pucciniales et Erysiphales).

Les génomes des champignons associés aux plantes sont également marqués par des répertoires de gènes spécifiques relatifs aux modes de vie mutualistes ou parasitaires établis avec leurs plantes hôtes. Les gènes codant des enzymes impliquées dans la dégradation des composants de la paroi végétale comme les CAZymes, dans la transduction du signal, dans le transport de nutriments ou encore les effecteurs sont essentiels pour le succès de la colonisation de l'hôte, la nutrition et la reproduction. La composition en CAZymes d'un génome fongique peut être un indicateur de son mode de vie. Les mutualistes ectomycorrhiziens ou les biotrophes obligatoires comme les Pucciniales présentent des répertoires en CAZymes réduits par rapport à des champignons nécrotrophes ou saprotrophes, indiquant une moindre capacité à altérer la paroi végétale, ce qui permettrait d'éviter de déclencher une réponse

immunitaire pendant l'étape de colonisation de l'hôte (Duplessis *et al.,* 2011a ; Martin *et al.,* 2008). On retrouve aussi un large répertoire de gènes codant des protéines prédites comme sécrétées (sécrétome), dont une partie est de petite taille (inférieure à 300 acides aminés). Au sein du sécrétome on peut distinguer des protéines présentant une homologie de fonction dans les bases de données comme des CAZymes, des lipases ou encore des protéases mais aussi des protéines de fonction inconnue qui représentent le répertoire d'effecteurs candidats ayant potentiellement un rôle dans le processus infectieux (Lowe & Howlett, 2012; Raffaele & Kamoun, 2012; Spanu & Panstruga, 2012). Ces génomes sont souvent marqués par une prépondérance d'éléments transposables ainsi parfois qu'un large nombre de familles multigéniques (Castanera *et al.,* 2016; Raffaele & Kamoun, 2012). L'association entre les gènes codant des familles d'effecteurs candidats et des régions riches en éléments transposables a été décrite dans les génomes de différents pathogènes filamenteux (par exemple *L. maculans* et *P. infestans*) suggérant que l'activité des transposons pourrait jouer un rôle central dans la diversification des répertoires d'effecteurs chez les pathogènes (Haas *et al.,* 2009; Rouxel *et al.,* 2011). En effet, les effecteurs de ces régions riches en éléments transposables peuvent présenter un polymorphisme élevé suggérant une adaptation au système immunitaire de l'hôte (Raffaele & Kamoun, 2012). Des résultats ont conduit à émettre l'hypothèse de l'existence de génomes à deux vitesses chez les pathogènes (Dong *et al*., 2015).

Les génomes de la rouille du peuplier *M. larici-populina* et de la rouille du blé *Puccinia striiformis* f. sp. *tritici* furent les premiers génomes de Pucciniales séquencés en 2011 dans le cadre de consortia de recherche internationaux et en collaboration avec le JGI (pour la rouille du peuplier) et le Broad Institute of MIT and Harvard (pour la rouille noire du blé) (Duplessis *et al.*, 2011a ; McDowell, 2011). L'organisation et la composition similaires de ces deux génomes ont mis en évidence les particularités des génomes des Pucciniales en comparaison avec d'autres champignons pathogènes parmi les Ascomycètes et d'autres Basidiomycètes. En effet, les Pucciniales présentent un génome de grande taille (101Mb pour *M. larici-populina*), comprenant un grand nombre d'éléments transposables (45% pour *M. larici-populina*) et un nombre important de gènes (16 399 pour *M. larici-populina*). La comparaison avec d'autres parasites biotrophes indiquent le partage de certaines de ces caractéristiques avec des différences notables, par exemple *B. graminis* présente un génome de très grande taille et un nombre conséquent d'éléments transposables avec un nombre de gènes limité, alors que *U. maydis* présente un génome de petite taille, pauvre en gènes et en éléments transposables (Duplessis *et al.,* 2014). Ces données indiquent différentes trajectoires évolutives chez ces champignons pathogènes des plantes. Depuis, plusieurs génomes de Pucciniales ont été séquencés, notamment les espèces appartenant aux familles des Pucciniaceae (4 espèces),

des Melampsoraceae (2 espèces), et les quatres espèces *Austropuccinia psidii*, *Hemileia vastatrix*, *Phakopsora pachyrhizi* et *Uromyces fabae* (Tableau 1). La qualité de ces génomes varie grandement en fonction de la stratégie de séquençage retenue (Aime *et al.,* 2017). En effet, la grande taille et la composition de ces génomes rend difficile leur assemblage et la qualité des annotations en gènes.

Tableau 1 : Ressources génomique des Pucciniales

Espèce	Taille du génome	Elements transposables et répétés (%)	Nombre de gènes	Référence
Melampsora larici-populina	101,1 Mb	45%	16 399	Duplessis et al. 2011
Melampsora lini	220 Mb†	45,9%	16 271	Nemri et al. 2014
Puccinia graminis f. sp. *tritici*	88,6-92 Mb	43,7%	17773 - 21 874	Duplessis et al. 2011 ; Upadhyaya et al.2015
Puccinia triticina	100.6-135.3 Mb	37,5 - 50,9%	26 384 – 27 678	Kiran et al. 2016 ; Cuomo et al. 2017
Puccinia striiformis f. sp. *tritici*	53-117,3 Mb†	17,8 – 48,9%	18 362 – 25 288	Cantu et al. 2011 ; Cantu et al. 2013 ; Zheng et al. 2013 ; Cuomo et al. 2017 ; Kiran et al. 2017 ;
Puccinia sorghi	102 Mb†	33%	21087	Rochi et al. 2016
Uromyces fabae	330-379 Mb†	n.a.	n.a.	Link et al. 2014
Hemileia vastatrix	333 Mb	74.4%	14 445	Cristancho et al. 2014
Austropuccinia psidii	103-145 Mb†	27%	> 19 000	Tan et al. 2014
Phakopsora pachyrhizi	850 Mb - 1 Gb†	n.a.	n.a.	Loehrer et al. 2014

† Taille estimée; n.a Donnée non renseignée

En 2014, dans le cadre de la thèse de Michaël Pernaci (UMR 1136 IAM), une carte génétique de *M. larici-populina* a été obtenue dans le but de cartographier certains des traits de vie lié à

la virulence et à l'agressivité du champignon (Pernaci, 2014). Pour ce faire, une population de 95 descendants issue de l'autofécondation de l'isolat de référence 98AG31 de *M. larici-populina* a été générée. Les génomes de ces 95 descendants ont ensuite été séquencés par Illumina, avec une profondeur de séquençage moyenne d'environ 6,9X. Ce résultat entre dans le cadre d'un projet *Community Science Programme* en partenariat avec le JGI. Les données de séquençage obtenues par la technologie Sanger pour le génome de *M. larici-populina* (Duplessis *et al.,* 2011a) ont été réassemblées avec des logiciels plus récents et plus performants. Il a alors été procédé à l'ancrage des scaffolds obtenus sur la carte génétique composée de 18 groupes de liaison (LGs) (Pernaci, 2014). La version 2 du génome de *M. larici-populina* est actuellement disponible dans la base de données génomiques Mycocosm du JGI (https://genome.jgi.doe.gov/programs/fungi/index.jsf). Dans cette nouvelle version, le génome de *M. larici-populina* comprend donc 18 chromosomes (LGs) pour une taille de près de 110Mb contre 101Mb dans sa première version. Il contient aussi 488 scaffolds de petite taille issus de l'assemblage qui n'ont pas pu être intégrés dans les LGs. En effet, on retrouve près de 5.8% de trous dans le génome et la phase des haplotypes n'est pas résolue (Tableau 2). Il est à noter que si on ne connaît pas le nombre attendu de chromosomes chez *M. larici-populina*, des travaux de génétique indiquent que 18 chromosomes seraient présents dans le génome haploïde de l'espère voisine *M. lini* (Boehm & Bushnell, 1992). Ceci laisserait supposer que le nombre fini de chromosomes a bien été atteint avec 18 LGs. Il est à noter qu'une carte génétique par RAD-Seq produite récemment chez *M. lini* n'a permis de résoudre que 27 LGs (Anderson *et al.,* 2016). L'annotation automatique du génome de *M. larici-populina* a été réalisée en utilisant les ESTs de la version 1 du génome ainsi que les données de RNAseq obtenues sur mélèze issues de l'article n°3 du **Chapitre II**. L'apport des données d'expression chez l'hôte alternant a permis d'étendre le catalogue de gènes prédits à 19 550 modèles de gènes. Dans le cadre de ma thèse, j'ai contribué à l'effort d'analyse de cette nouvelle version du génome en vue de sa publication en 2018. Sur la base des données mises à disposition par le JGI, je me suis attachée à réaliser l'annotation des éléments transposables ainsi que de certaines familles de gènes (dont les CAZYmes, protéases et transporteurs), et notamment du sécrétome dans la version 2.0 du génome de *M. larici-populina*. J'ai aussi procédé à l'analyse comparative des familles multigéniques chez les Pucciniales et chez d'autres champignons interagissant avec les plantes. Ce travail a été réalisé avec l'appui d'Emmanuelle Morin (IE Bioinformatique de l'UMR 1136 IAM). Les résultats issus de ces analyses sont détaillés dans la première partie de ce chapitre.

Tableau 2: Statistiques de réassemblage de la version 2.0 du génome de *Melampsora larici-populina* (Mlp) 98AG31 genome versus la version 1

Métrique	*Mlp* version 1*	*Mlp* version 2
Couverture de séquençage	6.9	7.47
Taille du génome, Mb	101.1	109.88
Scaffolds/Groupes de liaison (GL)	462	18 GL and 488 petits scaffolds
N50† par taille de scaffold, Mb	1.1	5.82
N50† par nombre de scaffolds	27	8
Taux de lacunes, %	3.4	5.8
Gènes codant des protéines	16399	19550
Taille moyenne de séquence codante, nt	1565	1468
Nombre moyen d'exon par séquence	4.92	4.37
Taille moyenne des exons, nt	247	254
Taille moyenne des introns, nt	118	108

* tire de Duplessis et al., 2011a
†La métrique N50 correspond au nombre N des plus larges scaffolds couvrant la moitié du génome total. La taille N50 montre la taille du plus petit scaffold parmi la métrique N50.

Dans le cadre du projet 1000 génomes fongiques, les chercheurs sont invités à travers l'option *Nominate New Species!* à proposer des espèces dans des familles taxonomiques pour lesquelles peu ou pas de génomes sont actuellement disponibles (https://genome.jgi.doe.gov/programs/fungi/1000fungalgenomes.jsf). Dans ce projet, les chercheurs extraient et envoient de l'ADN génomique de haut poids moléculaire au JGI qui en réalise le séquençage et l'annotation automatique avec la mise à disposition d'un portail d'accès au génome dans la base de données Mycocosm du JGI. Ainsi, suite à une proposition de Sébastien Duplessis et Pascal Frey de l'UMR 1136 IAM, le génome de *M. allii-populina* a été séquencé et mis en accès sur le site du JGI en 2016. A la base du projet de séquençage de cette espèce, il s'agissait de comparer les catalogues de gènes entre deux espèces de Melampsoraceae proches, hétéroïques infectant un même hôte télien le peuplier et deux hôtes écidiens distincts afin de tenter d'identifier les gènes qui seraient spécifiques à l'infection de

chacun des hôtes chez ces deux espèces. En effet, si *M. larici-populina* et *M. allii-populina* partagent le peuplier (*Populus* spp.) comme hôte télien, elles alternent respectivement sur le mélèze (*Larix* spp.) et *Allium* spp. (aïl, oignon, muscaris, etc), (Figure 4). Ces deux espèces phylogénétiquement proches partagent un même environnement et peuvent infecter simultanément un même arbre (Frey & Plnon, 1997). De plus, *M. allii-populina* étant moins agressif que *M. larici-populina*, les résistances des peupliers cultivés en France ont été sélectionnées contre *M. larici-populina* mais certaines de ces résistances empêchent également l'infection des peupliers par *M. allii-populina* (Frey *et al.,* 2005). Lors de la proposition de ce projet, la taille du génome avait été estimée par défaut comme comparable à celle de *M. larici-populina* autour de 100 Mb. Ce génome s'est avéré avoir une taille beaucoup plus conséquente avec 336Mb. L'accès à la technologie de séquençage PacBio au JGI et à de nouvelles procédures d'assemblage ont permis de générer des fragments de séquences plus longs et de réaliser un assemblage en 3 365 scaffolds seulement. Si ce chiffre peut paraître considérable au regard d'autres génomes de champignons, il reflète la complexité d'assemblage de grands génomes riches en éléments transposables. A titre de comparaison, pour la rouille du lin *M. lini*, le génome qui possède une taille de 220Mb et qui a été séquencé par la méthode Illumina, a été assemblé en 21 310 scaffolds (Nemri *et al.,* 2014). De plus, l'utilisation de l'approche PacBio pour *M. allii-populina* a permis d'identifier des portions haplotypiques du génome, ce qui permet de comparer les différences entre les deux noyaux de l'isolat dicaryotique séquencé. Un séquençage RNAseq à partir de transcrits issus de spores et de feuilles de peuplier infectées a permis de générer un support pour l'annotation des gènes. Un total de 23 089 gènes a ainsi été identifié. Dans le cadre de ma thèse, j'ai utilisé les données mise à disposition par le JGI pour réaliser l'annotation des éléments transposables ainsi que de certaines familles de gènes (dont les CAZYmes, protéases et transporteurs), et notamment du sécrétome dans le génome de *M. allii-populina*. J'ai aussi procédé à l'analyse comparative des familles multigéniques chez les Pucciniales et chez d'autres champignons interagissant avec les plantes et plus particulièrement entre *M. allii-populina* et *M. larici-populina*. Ce travail a aussi été réalisé avec l'appui d'Emmanuelle Morin. Les résultats issus de ces analyses sont détaillés dans la seconde partie de ce chapitre. Dans les génomes des Pucciniales, les éléments transposables (TE) représentent une part importante du génome (plus de 40% dans les génomes de *M. larici-populina* et de *Puccinia graminis* f. sp. *tritici* ; Duplessis *et al.,* 2011a). Afin d'en réaliser l'annotation experte dans les

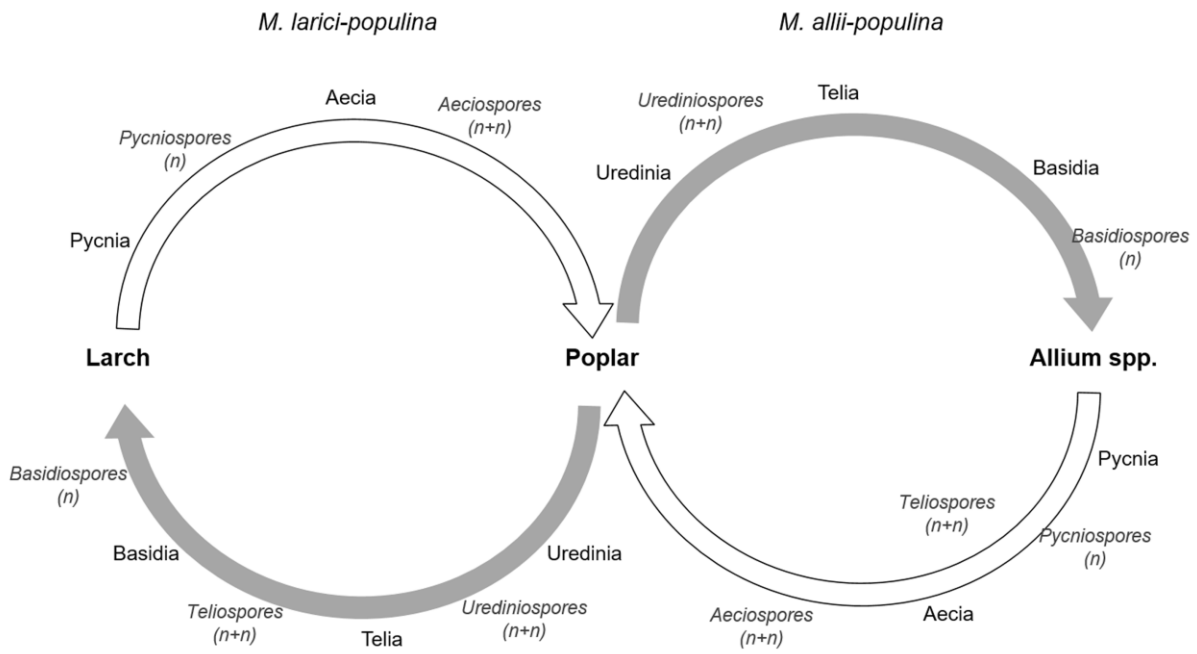


Figure 8 : Cycles de vie hétéroïques et macrocycliques de deux rouilles du peuplier *M. larici-populina* et *M. allii-populina*.

Au printemps, les basidiospores haploïdes (n) infectent les hôtes écidiens respectifs (*Larix* et *Allium* spp.) de *M. larici-populina* et *M. allii-populina*. Les pycnies sont formées sur les hôtes écidiens et produisent des pycniospores (n) avec différents types de compatibilité sexuelle pour la reproduction sexuée. Après la plasmogamie, les écidies dikaryotiques sont formées et produisent des écidiospores (n+n) qui pourront infecter l'hôte télien, le peuplier. L'infection du peuplier par les écidiospores conduit à la production d'urédiniospores (n+n) qui vont se disperser et se multiplier de façon clonale et répétitive pendant toute la période estivale. En automne, les urédies se différencient en télies produisant des téliospores (n+n) dans lesquelles se produisent la karyogamie et le début de la méiose avant la dormance hivernale. Au printemps suivant, la méiose se poursuit dans les basides qui germent et produisent des basidiospores (n) pour un nouveau cycle de vie.

deux génomes à disposition, j'ai suivi une formation à l'annotation des TE avec le pipeline REPET au sein de l'équipe de l'URGI de l'INRA de Versailles (http://urgi.versailles.inra.fr/index.php/urgi/Tools/REPET). Brièvement, ce pipeline permet d'identifier les TE par l'outil TEdenovo puis de réaliser l'annotation des consensus obtenus avec l'outil TEannot (Flutre *et al.,* 2011). Ici aussi, j'ai travaillé avec l'appui d'Emmanuelle Morin de l'UMR 1136 IAM ainsi qu'avec l'aide des formateurs rencontrés à l'URGI qui m'ont aidé à mener à bien cette annotation dans des génomes très riches en TEs.

Les deux parties qui suivent décrivent ma contribution à l'analyse des données de génomique à disposition pour les génomes de *M. larici-populina* et *M. allii-populina*. Ces dernières seront utilisées pour la rédaction de deux articles en préparation pour la publication de ces génomes en collaboration avec le JGI.

Partie 2 : Matériel et Méthodes

Génomes

Les génomes de *M. larici-populina* (https://genome.jgi.doe.gov/Mellp2_3/Mellp2_3.home.html) et *M. allii-populina* (https://genome.jgi.doe.gov/Melap1finSC_191/Melap1finSC_191.home.html) ont été séquencés, assemblés et annotés par le JGI et sont disponibles sur le site Mycocosm. Les annotations fonctionnelles basées sur les bases de données *Gene Ontology* (GO) ; *Eukaryotic Othologous Group* (KOG) ; *Kyoto Encyclopedia of Genes and Genomes* (KEGG), la base de données *Carbohydrate-Active enZYmes Database* (CAZY) et la base de données des peptidases MEROPS, sont disponibles également sur le site Mycocosm.

Annotation des éléments transposables

Afin d'estimer la proportion d'éléments transposables présents dans la nouvelle version du génome de *M. larici-populina* et dans le génome de *M. allii-populina* l'annotation des éléments transposables a été réalisée *de novo* avec le pipeline d'annotation REPET sur la version non masquée de l'assemblage v2 du génome *M. larici-populina* et dans le génome de *M. allii-populina* (Flutre *et al.*, 2011; Quesneville *et al.,* 2005). Ce même pipeline avait également été utilisé pour l'annotation des TE dans la première version du génome de *M. larici-populina*. La recherche *de novo* des séquences répétées au sein des 18 LG et des 488 petits scaffolds du génome de *M. larici-populina* et dans les 3637 scaffolds de *M. allii-populina* a été réalisée par BLASTER (Quesneville *et al.,* 2005 ; *High scoring pairs* (HSP) retenus : > 90% d'identité; taille comprise entre 100b et 20Kb et *e*-value < 10^{-300}). Les HSP identifiées ont été groupées par similarité de séquence avec GROUPER, RECON et PILER (Bao & Eddy, 2002; Edgar & Myers, 2005; Quesneville *et al.,* 2003). Des alignements multiples des 20 plus longs membres de chaque groupe de HSP d'au moins trois membres ont été utilisés pour déterminer la séquence consensus de chaque groupe avec les programmes MAP et MAFFT (Huang, 1994; Katoh *et al.,* 2002). Les séquences consensus ont été classées avec l'outil TEclassifier et la redondance éventuelle entre les consensus a été éliminée avec BLASTER et MATCHER. Les séquences consensus dont la catégorie n'a pas pu être identifiée sont placées dans la catégorie NoCat. L'annotation des TE identifiés par TEdenovo est réalisée sur la base des TE connus par le pipeline TEannot. Les trois méthodes BLASTER, CENSOR et REPEATMASKER ont été utilisées pour annoter les fragments de TE dans le génome de *M. larici-populina et de M. allii-populina* en se basant sur les séquences consensus issues du pipeline TEdenovo. Pour discriminer les TE des simples séquences répétées (*Simple Sequence Repeats*, SSR), les trois méthodes TRF, MREPS et REPEATMASKER ont été utilisées et les séquences redondantes TE/SSR ont été éliminées. Enfin, pour résoudre les

problèmes d'annotation des TE imbriqués, une procédure dite *long join*, qui permet de trouver et d'assembler les fragments de TE séparés par l'insertion récente d'un autre TE, a été utilisée.

Analyse MCL

Depuis la publication de la première version du génome de *M. larici-populina* (Duplessis *et al.,* 2011a), de nombreux génomes de champignons ont été séquencés, notamment à travers le projet de séquençage de 1000 génomes fongiques en collaboration avec le JGI. Afin d'identifier et de comparer les familles multigéniques dans la nouvelle version du génome de *M. larici-populina* et dans le génome de *M. allii-populina*, nous avons comparé à l'aide de l'outil BLASTP les protéomes de 16 et 22 génomes de champignons disponibles publiquement sur le site du JGI, pour *M. larici-populina* et *M. allii-populina* respectivement. Les 16 espèces fongiques se composent comme suit : cinq génomes de Pucciniales (dont deux Melampsoraceae spp.) *M. larici-populina* 98AG31 v2; *Melampsora lini* CH5; *Puccinia graminis* f. sp. *tritici* v2; *Puccinia striiformis* f. sp. *tritici* PST130; *Cronartium quercuum* f. sp. *fusiforme* G11; trois Pucciniomycètes: *Microbotryum violaceum*, *Mixia osmunde; Sporobolomyces roseus;* quatre Basidiomycètes: *Laccaria bicolor*, *Cryptococcus neoformans* var *neoformans*, *Piriformospora indica*, *Heterobasidion annosum*, *Ustilago maydis;* et quatre Ascomycètes: *Leptosphaeria maculans ; Magnaporthe oryzae ; Neurospora crassa* OR74A. Les 23 espèces fongiques se composent comme suit : six génomes de Pucciniales (dont trois Melampsoraceae spp.) *M. allii-populina* ; *M. larici-populina* 98AG31 v2; *M. lini* CH5; *P. graminis* f. sp. *tritici* v2.0; *P. striiformis* f. sp. *tritici* PST130; *C. quercuum* f. sp. *fusiforme* G11; les Pucciniomycètes: *R. graminis* ; *M. violaceum*, *M. osmunde; S. roseus;* les Basidiomycètes: *L. bicolor*, *C. neoformans* var *neoformans*, *P. indica*, *H. annosum*, *U. maydis;* et les Ascomycètes: *B. graminis* f. sp*. tritici* 96224*; S. populicola ; C. fulvum; L. maculans ; P. tritici-repentis; S. sclerotiorum ; M. oryzae ; N. crassa* OR74A. Les familles multigéniques ont été construites par similarité de séquences et groupées par le programme MCL (Enright *et al.,* 2002)

Annotation du sécrétome

Pour établir les catalogues des gènes codant des protéines sécrétées dans la version 2.0 du génome de *M. larici-populina* et dans le génome de *M. allii-populina*, un pipeline *in sillico* combinant plusieurs outils de prédiction disponibles au sein de l'UMR 1136 IAM a été utilisé. Ce pipeline reprend les éléments déjà abordés au chapitre I (Lorrain *et al.*, 2015 ; voir aussi Sperschneider *et al.,* 2017). Les sécrétomes ont été prédits parmi les 19550 gènes de la version 2 du génome de *M. larici-populina* et les 23089 gènes de *M. allii-populina* en combinant trois outils de détection de peptide signal sur les séquences protéiques déduites : TargetP, SignalP4.1 et Phobius (Emanuelsson *et al.,* 2007; Käll *et al.,* 2004; Petersen *et al.*, 2011).

Une protéine a été considérée comme sécrétée si au moins deux des outils de prédiction sur trois prédisaient la présence d'un peptide signal dans la région N-terminale de la séquence protéique. Afin d'éliminer les protéines transmembranaires ou les protéines retenues dans le réticulum endoplasmique, nous avons utilisé les outils TMHMM, Phobius et PS-SCAN pour prédire la présence de domaines transmembranaires et de motifs de rétention dans le réticulum endoplasmique (de Castro *et al.,* 2006; Moller *et al.,* 2001). Les séquences protéiques portant au moins un domaine transmembranaire ne se superposant pas au peptide signal prédit et/ou contenant un motif de rétention dans le réticulum endoplasmique ont été retirées du sécrétome. Les familles multigéniques dont au moins la moitié des membres sont prédits comme sécrétés ont été considérées comme des familles de protéines sécrétées. Ainsi, certains membres dont la sécrétion n'était prédite que par un seul des trois outils de prédiction et présentant la même structure génique (nombre et organisation des introns et exons) que ceux prédits sécrétés au sein d'une même famille ont été comptabilisés dans le sécrétome de *M. larici-populina* pour tout ce qui concerne les analyses comparatives avec d'autres génomes fongiques. Enfin, nous avons profité de la publication récente d'un nouvel outil de prédiction des effecteurs chez les champignons, EffectorP, pour prédire des effecteurs candidats dans le sécrétome de *M. larici-populina* (Sperschneider *et al.,* 2016). Nous avons également utilisé les programmes ApoplastP et LOCALIZER pour prédire la possible localisation des protéines sécrétées de *M. larici-populina* et de *M. allii-populina* (Sperschneider *et al.,* 2017b; Sperschneider *et al.,* 2017c). Enfin, les cartes de densité des gènes de *M. larici-populina* et *M. allii-populina* ont été réalisées avec la méthode développée par Saunders *et al.* (2014).

Annotation des CAZymes, Protéases et Transporters

Les gènes codant les CAzymes dans la nouvelle version du génome de *M. larici-populina* et dans le génome de *M. allii-populina* ont été annotées en collaboration avec Bernard Henrissat (UMR 7257 AFMB CNRS-Université Aix-Marseille) en utilisant la base de données CAZY (http://www.cazy.org; Cantarel *et al.,* 2009). Les CAZymes ont été classées en familles de glycoside hydrolases (GH), glucosyltransferases (GT), polysaccharide lyases (PL), carbohydrate esterases (CE), activités auxiliaires (AA) et carbohydrate-binding modules (CMB). Le répertoire de CAZymes de la version 2 du génome de *M. larici-populina* et dans le génome de *M. allii-populina* ont été comparé aux CAZymes répertoriées chez les biotrophes obligatoires et chez les 15 et 22 génomes fongiques utilisés pour l'analyse comparative MCL. Les gènes codant les protéases dans la version 2 du génome de *M. larici-populina* et dans le génome de *M. allii-populina* ont été identifiés et annotés sur la base des recherches d'homologie et des alignements réalisés avec la base de données MEROPS (https://www.ebi.ac.uk/merops/).

En tant que champignon biotrophe obligatoire, les nutriments nécessaires au développement de *M. larici-populina* et de *M. allii-populina* proviennent des hôtes qu'il infecte (cf. **Chapitre I**, article n°1). Les gènes codant des transporteurs dans la nouvelle version du génome ont été prédits *de novo* par recherche d'homologie (BLASTP) puis alignements du protéome contre la base de donnée *Transporter Classification Database* (TCDB ; http://www.tcdb.org/). Les assignements avec une e-value < 10^{-5} et comprenant un domaine PFAM correspondant à une sous-famille TCDB ont été annotés comme transporteurs. L'annotation des familles de transporteurs qui avaient été établie dans la version 1 du génome a aussi été utilisée pour valider les gènes déjà identifiés pour *M. larici-populina*.

Partie 3 : Contribution à l'annotation de la version 2 du génome de *Melampsora larici-populina*

La version 2.0 du génome de *M. larici-populina* a été ré-assemblée et ancrée sur 18 LGs et 488 scaffolds et contient 19 550 gènes soit 3151 gènes supplémentaires par rapport à la version 1.0 (Tableau 2). Sur les 16 399 modèles de gènes de l'annotation de la version 1, seuls 13 160 (80%) ont été transférés sur la version 2, ce qui indique que 2 339 gènes n'ont pas été conservés car ils ont été modifiés (séparés en deux gènes, fusionnés à un autre gène, étendu ou effacé. Notamment, parmi les 1 872 gènes qui avaient été manuellement annotés par le consortium en charge de l'annotation experte du génome, seuls 1 627 (87%) sont représentés de manière identique dans la version 2. Parmi les 19 550 gènes du génome v2 de *M. larici-populina*, 61% (12 112) ne présentent ni fonction ni domaine connus. Dans le détail, 4933 (25%) gènes présentent une annotation dans la base de donnée *Gene Ontology* (GO), 5463 (28%) présentent une annotation dans la base de données *euKaryotic Orthologous Groups* (KOG) et 7247 (37%) présentent une annotation dans la base de donnée Interpro ; représentant au total 7563 (39%) gènes avec au moins une annotation dans une de ces bases de données. Ce chiffre doit toutefois être modéré puisque certaines annotations dans ces bases restent floues et peu informative. On peut donc considérer que plus de 60% des gènes sont de fonction inconnue.

Dans cette partie sont présentés les résultats obtenus pour la ré-annotation du génome de *M. larici-populina* version 2.

1. Annotation des éléments transposables (TE)

La proportion de TE observée dans la version 2 du génome de *M. larici-populina* est estimée à 42% contre 45% dans la version 1.0 (Tableau 3). Dans la version 1 du génome, 55% des consensus n'ont été assignés à aucune classe de TE, 34,9% ont été annotés comme transposons à ADN (classe II) et 10% comme rétrotransposons (classe I ; Tableau 3; Figure 9). L'annotation *de novo* des TE dans la nouvelle version du génome montre désormais 3,3% de TE non assignés, 67,0% de transposons à ADN et 29,7% de rétrotransposons. Ainsi, la part des TE non annotés diminue drastiquement entre les deux versions du génome de *M. larici-populina* permettant ainsi l'amélioration de l'annotation des éléments transposables. On peut noter que les TE de classe II qui se transposent avec un mécanisme de type couper-coller restent majoritaire dans la version 2 du génome en comparaison avec les TE de classe I (transposition de type copier-coller). Comme dans la première version du génome les TE de la catégorie des TIR (classe II) sont les plus abondants dans la version 2, avec une

augmentation sensible de leur proportion puisqu'ils représentent 42,1% des consensus contre 30% dans la version 1. Ils sont suivis par la catégorie des LRT (classe I) qui représentent 25% des consensus contre 7,1% seulement dans la version 1. L'annotation des TE dans la nouvelle version du génome de *M. larici-populina* permet d'avoir une meilleure vision des classes et des catégories en présence. Si l'on considère l'ancrage sur 18 LG correspondant vraisemblablement aux chromosomes du champignon, la faible proportion de trous non comblés dans l'assemblage et la proportion de TE assignés (>95%), nous pouvons considérer que nous disposons d'une annotation quasi-définitive pour le génome de *M. larici-populina*.

Tableau 3: Classification REPET des séquences consensus identifiées chez *M. larici-populina* 2.0 versus les consensus identifiés dans la version 1.0.

TE consensus	***Mlp1.0***	**Pourcentage sur 2,020 consensus**		***Mlp2.0***	**Pourcentage sur 1,729 consensus**	
LTR	144	7.13%	Class I 202 (10.0%)	440	25.01%	Class I 513 (29.7%)
DIRS	11	0.54%		34	1.93%	
LINE	45	2.23%		20	1.14%	
SINE	2	0.10%		3	0.17%	
TRIM	0	0		9	0.51%	
TIR	604	29.9%	Class II 704 (34.9%)	740	42.07%	Class II 1159 (67.0%)
MITE	96	4.75%		219	12.45%	
Helitron	4	0.20%		20	1.14%	
Maverick	0	0		5	0.28%	
NoCat*	1,114	55.1%	-	57	3.3%	-

*NoCat représente les éléments transposables non identifiés par le pipeline REPET

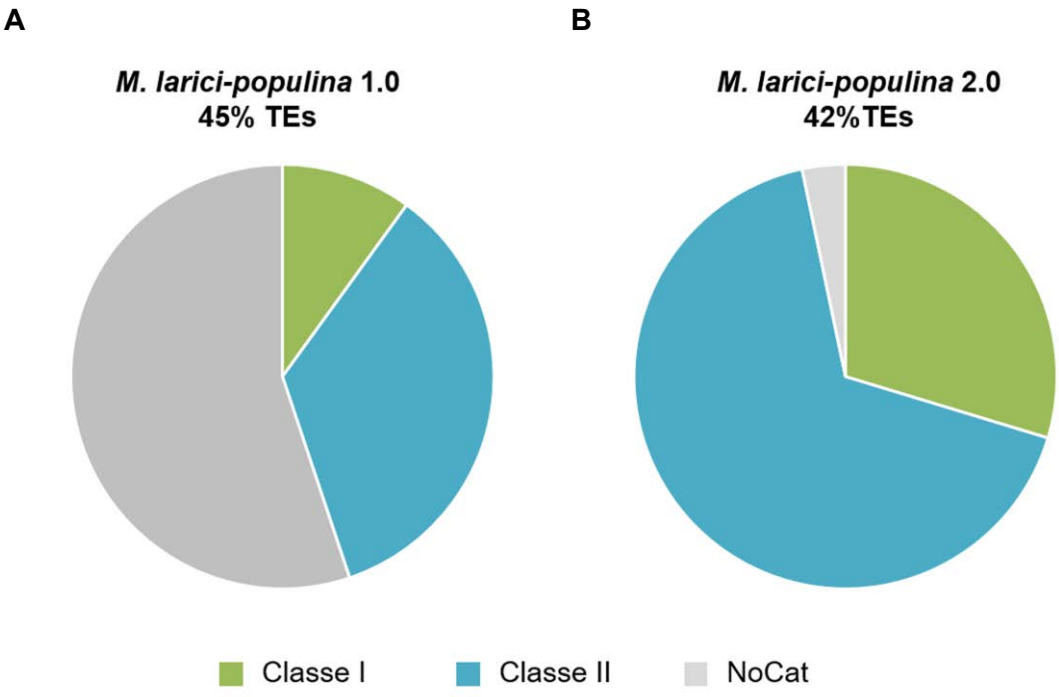


Figure 9: Ré-annotation des éléments transposables dans le génome version 2.0 de *M. larici-populina*
Les éléments transposables de *M. larici-populina* versions 1.0 et 2.0 ont été annotés avec le pipeline REPET (Flutre *et al.,* 2011). Les éléments transposables ont été identifiés par l'outil TEdenovo et annotés par l'outil TEannot. A. Classification des éléments transposables de *M. larici-populina* version 1.0 tirés de Duplessis *et al.* (2011a). B. Classification des éléments transposables ré-annotés dans la version 2.0 du génome de *M. larici-populina*. La Classe I (vert) représente les éléments transposable appartenant à la catégorie des rétrotransposons selon la classification de Wicker *et al.,* (2007). La Classe II (bleu) représente les éléments transposables appartenant à la catégorie des transposons à ADN. La catégorie NoCat (gris) représente les éléments transposables identifiés par TEdenovo et non catégorisés.

2. Analyse comparative des familles de gènes dans le génome de *M. larici-populina*, les génomes de Pucciniales et d'autres champignons

Avant de procéder à l'analyse comparative par MCL, les protéomes prédits de différents génomes sélectionnés ont été comparés à la base de données PFAM compilant les domaines généralement attribués aux protéines dont des TE déjà répertoriés dans les génomes eucaryotes (par exemple : PF01359 Transposase_1 ; Punta *et al.*, 2012). Nous avons pu constater que certains génomes contenaient des homologues de TE, dont la version 2 du génome de *M. larici-populina* (18 gènes). Ces derniers n'ont pas été considérés pour l'analyse comparative. Sur les 19532 gènes restant, 6038 sont des singletons (copie unique chez *M. larici-populina*) et 13494 appartiennent à des familles de 2 gènes ou plus. Parmi les singletons,

1848 sont uniques à *M. larici-populina*. La famille la plus large contient 301 gènes du génome de *M. larici-populina*, et elle présente des homologues en très large nombre (de 74 à 173 gènes par génome) au sein des Pucciniales. Au delà, seuls deux basidiomycètes pathogènes, *Microbotryum violaceum* (pathogène de plante) et *Cryptococcus neoformans* (pathogène de l'homme) possèdent un nombre réduit de copies (7 et 6 respectivement). Un total de 16 familles multigéniques de plus de 50 membres est retrouvé chez *M. larici-populina* représentant 1626 gènes. A l'exception d'une famille de plus de 50 gènes retrouvée en large nombre dans tous les génomes explorés, une très large majorité de ces familles est spécifique des Pucciniales ou de quelques génomes au sein des Pucciniales. En comparaison, la version 1 comprenait 5304 familles multigéniques avec au moins deux membres. Parmi celles-ci, 19 (comprenant un total de 1597 gènes) comptaient plus de 50 membres (Duplessis *et al.,* 2011a).

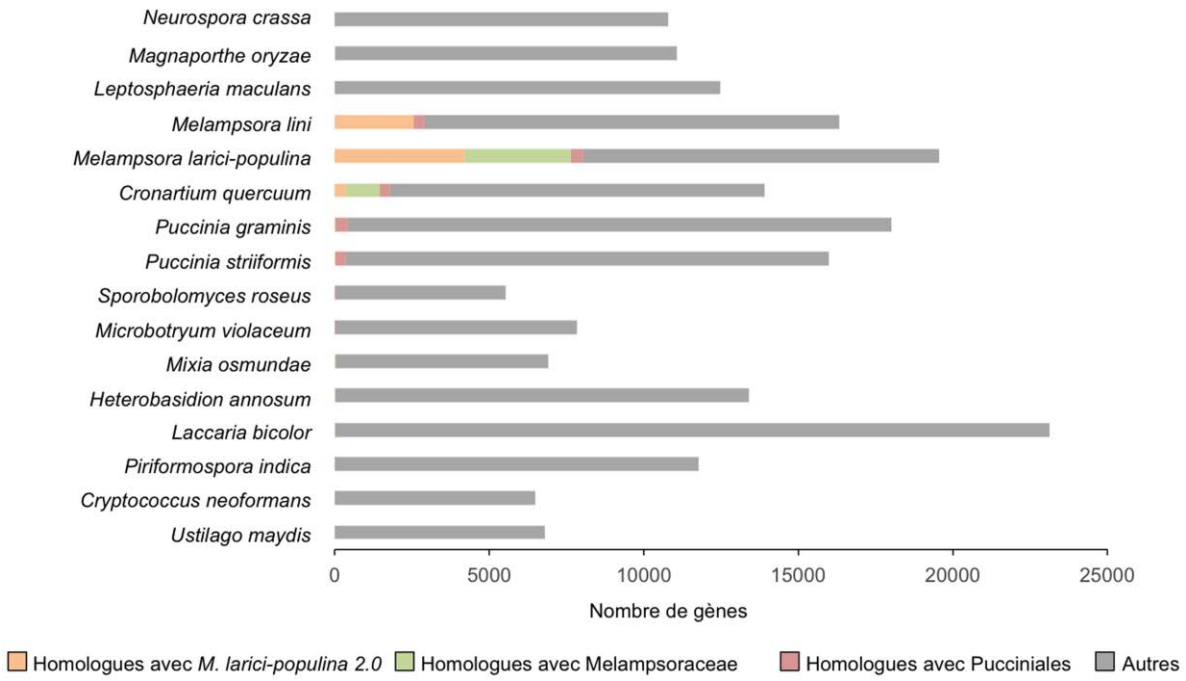


Figure 10 : Homologues partagés parmi 16 espèces fongiques.
Les axes des ordonnées et des abscisses indiquent respectivement le nombre de gènes parmi les génomes de 16 espèces fongiques sélectionnées pour l'analyse MCL (*Markov Clustering*). Les barres en orange représentent les homologues partagés avec *M. larici-populina* 2.0 ; en vert les homologues partagés parmi deux Melampsoraceae (*M. larici-populina* 2.0 et *M. lini*) ; en rouge les homologues partagés parmi six Pucciniales (*M. larici-populina* 2.0 ; *M. lini ; Cronartium quercuum ; Puccinia graminis* f. sp*. tritici ; Puccinia striifomis* f. sp *tritici*) et en gris les autres gènes de chaque génome.

Dans la première version du génome de *M. larici-populina*, 26% des familles multigéniques étaient identifiées comme spécifiques de l'espèce. La comparaison de la version 2 du génome de *M. larici-populina* avec 15 autres espèces fongiques dont *M. lini* de la famille des Melampsoracea, révèle que 78,5% (15331 gènes) des gènes possèdent au moins un orthologue dans une autre espèce. Cependant, 18,5% (3612 gènes) de ces gènes sont spécifiques des six Pucciniales considérées dans l'analyse (*M. larici-populina* 98AG31 v2; *Melampsora lini* CH5; *Puccinia graminis* f. sp. *tritici* v2.0; *Puccinia striiformis* f. sp. *tritici* PST130; *Cronartium quercuum* f. sp. *fusiforme* G11), 17,5% (3423 gènes) sont spécifiques des deux Melampsoraceae et 21,5% (4201 gènes) sont propre à *M. larici-populina* (Figure 10). Parmi ces gènes spécifiques aux Pucciniales, aux Melampsoraceae et à *M. larici-populina*, on trouve respectivement 3326 (92,1%) ; 3310 (96,6%) et 4162 (99%) gènes sans fonction connue et 437 (12%) ; 528 (15,4%) et 705 (16,7%) gènes codant des protéines prédites comme sécrétées. Ces résultats montrent que la majorité des gènes de la nouvelle version du génome de *M. larici-populina* présentent au moins un orthologue partagé avec une autre espèce fongique, parmi lesquels 38,5% sont spécifiques des Pucciniales et des Melampsoraceae. Dans l'ensemble, près de 42% des gènes sont propres à l'ordre des Pucciniales (ordre, famille ou espèce). Plus de 90% de ces gènes spécifiques aux Pucciniales, aux Melampsoraceae et à *M. larici-populina* sont de fonction inconnue, contre 61% pour le génome de *M. larici-populina* considéré dans son ensemble.

3. Prédiction du sécrétome

Les trois outils TargetP, SignalP4.1 et Phobius ont mis en évidence 2623 protéines portant un peptide signal dont 2197 ne contiennent ni domaine transmembranaire, ni motif de rétention dans le réticulum endoplasmique (Figure 11). Grâce à l'annotation des familles multigéniques dans la version 2 du génome de *M. larici-populina* décrite précédemment, nous avons pu mettre en évidence 563 familles de gènes SP de plus de deux membres. Nous avons vérifié la structure en introns/exons des gènes au sein de familles de SP et inclu dans le catalogue des SP les gènes codant pour des protéines prédites par un seul des trois outils utilisés mais dont la structure introns/exons est similaire aux autres SP au sein d'une même famille multigénique. Ainsi le sécrétome final de *M. larici-populina* comprend 2238 SP dont 64,6% font partie de familles de SP de plus de deux membres, 12% font partie de familles de deux membres et plus et 23,4% SP sont des singletons (Figure 11). Parmi ces 2238 SP, 64,2% (1439) sont des gènes codant des petites protéines sécrétées de moins de 300 acides aminés (*small secreted proteins* ; SSP). Enfin, pour rechercher de possibles effecteurs candidats parmi les 2238 SP , l'outil de prédiction EffectorP a permis de mettre en évidence 932 effecteurs candidats rassemblant des caractéristiques typiques des effecteurs fongiques et

ont donc été classés comme tel. Le sécrétome de *M. larici-populina* contient également 1380 SP prédites comme non-apoplastiques par l'outil ApoplastP et 832 SP sont prédites comme résidentes de l'apoplaste. Nous avons également utilisé l'outil LOCALIZER qui prédit 282 SP ciblant les chloroplastes, 139 ciblant les mitochondries et 485 ciblant les noyaux des cellules végétales. Les gènes codant des effecteurs ciblant les chloroplastes sont bien prédits comme tel (famille CTP ;Petre *et al.,* 2016 ; Cf. **Chapitre IV ; Annexe 1**).

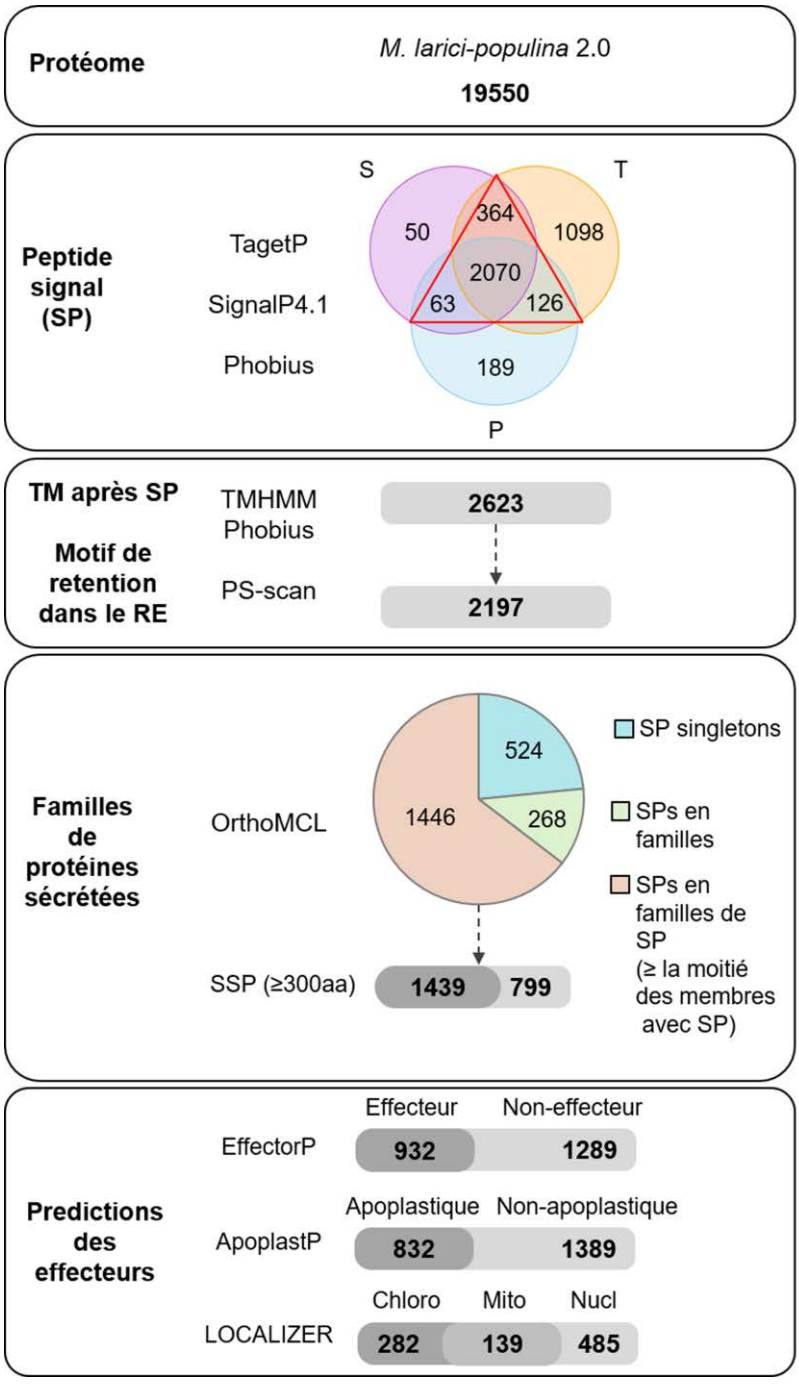


Figure 11 : Prédiction et analyse du sécrètome de *M. larici-populina* v2.0

Le sécrètome de *M. larici-populina* a été prédit en utilisant les trois outils de prédiction de peptide signal suivants : TargetP1.1 (T) ; SignalP4.1 (S) et Phobius (P). Une protéine donnée a été considérée comme prédite sécrétée « Secreted Protein » (SP) si au moins deux sur trois prédictions indiquent la présence d'un peptide signal. La présence dans la séquence protéique de chaque protéine de *M. larici-populina* 2.0 de domaines transmembranaires et de motifs de rétention dans le réticulum endoplasmique est prédite par les outils de prédiction TMHMM et PS-scan, respectivement. Les SP protéines avec plus d'un domaine transmembranaire prédit ou avec un domaine transmembranaire prédit en dehors des 60 premiers acides aminés ne sont pas considérées comme sécrétées. De même les protéines contenant un motif de rétention dans le réticulum endoplasmique prédit sont éliminées du sécrètome. Les familles multigéniques déterminées par analyse MCL qui contiennent au moins la moitié des membres prédits comme sécrétés sont considérées comme des familles multigéniques de protéines sécrétées (Familles de SP). Parmi ces familles, les protéines non prédites comme sécrétées par au moins deux des trois outils de prédiction

Figure 11. Prédiction et analyse du sécrètome de M. larici-populina *v2.0*
des protéines sécrétées mais montrant une structure introns/exons similaire aux autres membres sécrétés de la famille, sont ajoutées au sécrètome de *M. larici-populina* 2.0. Les petites protéines sécrétées de moins de 300 acides aminés sont identifiées comme SSP (« small secreted proteins »). Enfin les effecteurs candidats de *M. larici-populina* 2.0 sont prédits parmi le sécrètome avec l'outil de prédiction EffectorP et leur localisation intracellulaire *in planta* est prédite par les deux outils ApoplastP et LOCALIZER.

Dans le sécrétome de *M. larici-populina* de la version 1 du génome, 1184 SSP avaient été identifiées par combinaison d'outils de prédiction et d'annotation manuelle. Nous en retrouvons 84% (995 SSP) ré-annotées dans le génome version 2. Dans la version 1 du génome des homologues des effecteurs connus chez les Pucciniales étaient retrouvés. Ainsi, 24 homologues des Avrs de *M. lini* étaient présents dans la version 1 du génome : 1 SP homologues de AvrL567 ; 8 SP homologues de AvrM ; 13 SP homologues de AvrP4, 1 SP homologue de AvrP123 et chez *U. fabae et U. striatus* 1 SP homologue de RTP1. Dans la version 2 du génome nous retrouvons des gènes codant des SP homologues d'effecteurs caractérisés chez *M. lini* tels que : 1 SP homologue de AvrL567 ; 8 SP homologues de AvrM ; 5 SP homologues de AvrP4, 1 SP homologue de AvrP123 et 3 SP homologues de RTP1. A l'image du génome entier, la majorité des protéines sécrétées sont sans fonction (76,8% ; 1719 gènes) mais certaines SP présentent des fonctions prédites dans les bases de données GO, KOG, IPR ou annotées spécifiquement. Au total, 519 SP (23,2%) ont une fonction ou un domaine prédits. Parmi les fonctions prédites sur la base de données KOG, on retrouve 2010 SP de fonction inconnues et 228 avec diverses fonctions prédites (Figure 12). Les fonctions les plus représentées (plus de 1% du sécrétome) sont les fonctions associées aux « Modifications post-traductionnelles, renouvellement des protéines et chaperonnes » (46 SP) et au « Transport et métabolisme carboné » (34 SP). Les autres catégories KOG sont représentées à moins de 1% du sécrétome (Figure 12). Dans les catégories annotées plus en détail et impliquées dans le succés de l'infection on retrouve 158 CAZymes sécrétées, 11 expansines, 43 des protéases et on trouve 1 transporteur (*Mitochondrial carrier protein*). Il est également à noter que 12 SP portent un domaine CFEM (*Extracellular membrane protein*, en anglais) qui est un domaine spécifique des champignons et potentiellement impliqué dans l'infection (Kulkarni *et al.,* 2003). Parmi les effecteurs caractérisés chez *M. lini*, 14 sont transférés dans la version 2 du génome.

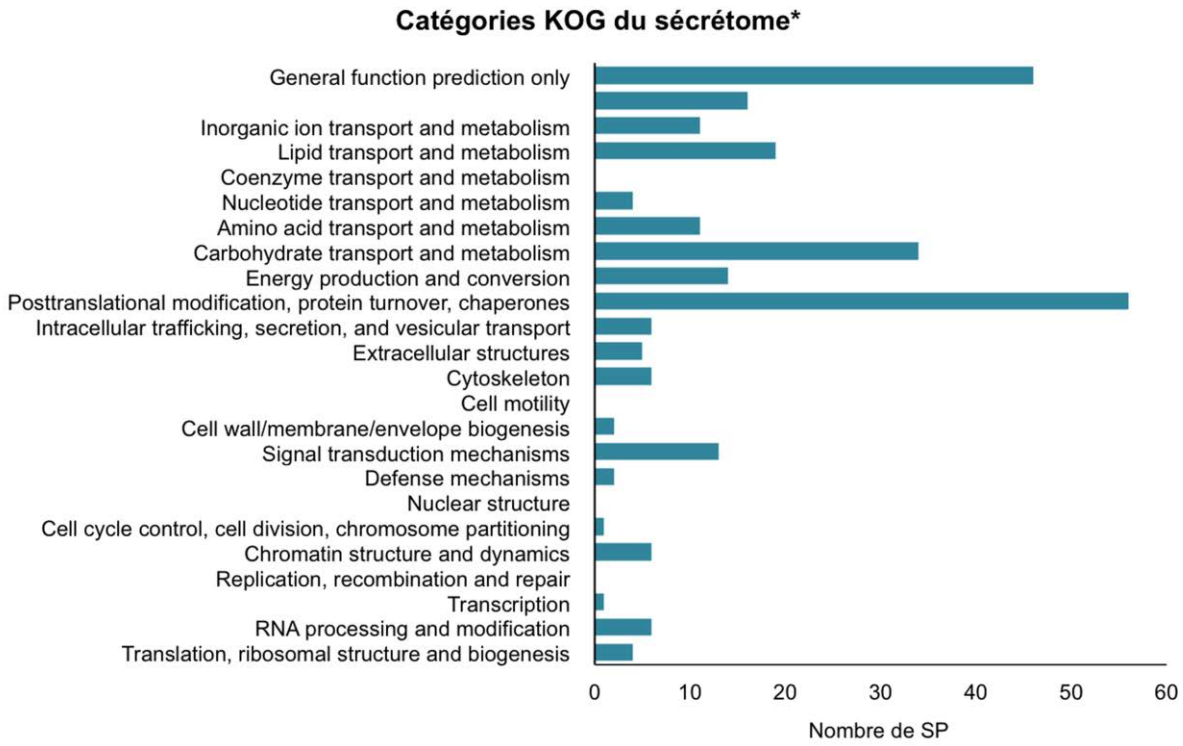


Figure 12 : Catégories KOG annotées dans le sécrétome de *M. larici-populina* 2.0
Les axes des abscisses et des ordonnées indiquent respectivement le nombre de protéines sécrétées (SP) et les catégories KOG. *Les SP sans fonction KOG ne sont pas affichées sur ce graphique et représentent 2010 SP.

Certains agents phytopathogènes filamenteux, comme par exemple l'oomycète *P. infestans*, présentent un génome à deux vitesses avec d'une part le génome conservé concentrant la majorité des gènes dans des régions pauvres en TE et avec un faible espacement intergénique, et d'autre part les gènes codant les effecteurs dans des régions riches en TE et faible en gènes avec un espace intergénique important (Raffaele & Kamoun, 2012). L'analyse de la version 1 du génome de *M. larici-populina* n'avait pas révélé d'association significative entre les gènes codant des SSP et les régions riches en éléments transposables (Duplessis *et al.,* 2011a). Ici, la nouvelle version du génome correspond à un nouvel assemblage qui a pu modifier les environnements proches de régions contenant des gènes SP. Afin de vérifier cela, nous avons cartographié les régions intergéniques du génome et du sécrétome. Contrairement à ce qui a été décrit chez l'oomycète *P. infestans* ou l'ascomycète *L. maculans*, les gènes codant le sécrétome de *M. larici-populina* présentent une distribution similaire aux autres gènes (Figure 13).

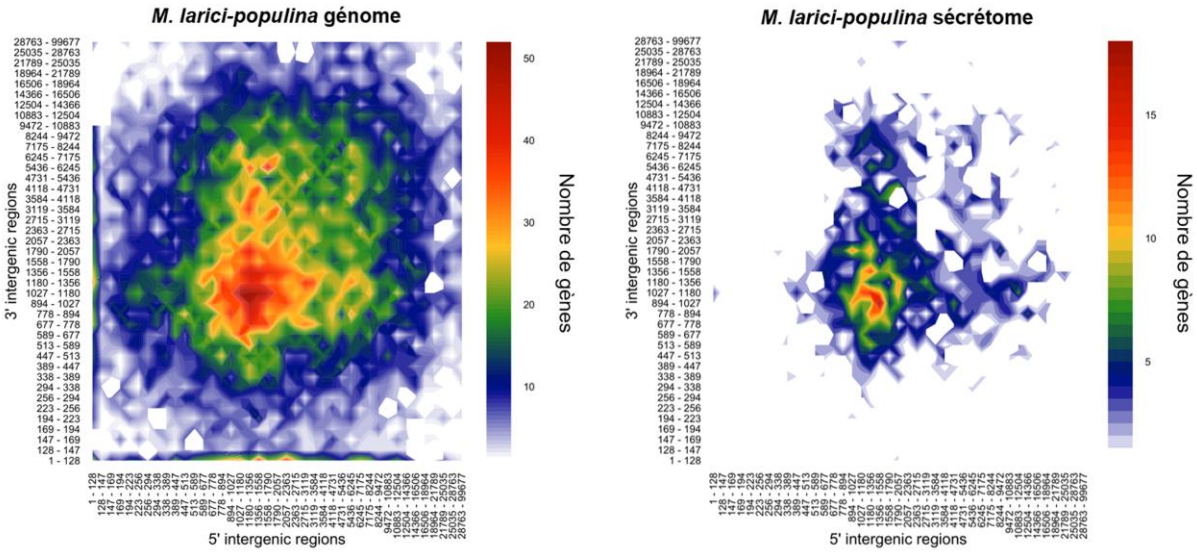


Figure 13 : Environnement génique des génome et sécrètome de *M. larici-populina* 2.0.
Les axes des abscisses et des ordonnées indiquent respectivement les régions intergéniques en 5' et 3' du génome de *M. larici-populina*. Les gènes de *M.larici-populina* 2.0 ont été triés selon des « fenêtres » à deux dimensions sur la base de la distance avec les régions intergéniques environnantes en 5' et 3'. A. « Heatmap » représentant la distribution des distances intergéniques en 5' et 3' pour tous les gènes de *M. larici-populina.* B. pour le sécrètome de *M. larici-populina* 2.0. Les échelles représentent le nombre de gènes par fenêtre.

Cette analyse du sécrétome de *M. larici-populina* 2.0 permet d'établir le nouveau catalogue d'effecteurs candidats prédits par les derniers algorithmes de prédiction génomique disponibles.

4. Annotation des familles de gènes codant des CAZymes, des protéases et des transporteurs

Les familles de gènes codant des enzymes hydrolytiques ciblant les polymères végétaux, tels que les protéases, les lipases et les enzymes dégradant les sucres comme les CAZymes sont fortement exprimées par les Pucciniales à différents stades de l'infection de leurs hôtes (Duplessis *et al.,* 2011b ; Lorrain *et al.,* 2018). Ces enzymes jouent un rôle vraisemblablement important dans la colonisation de l'hôte. Au total, 377 gènes codant des CAZymes ont été annotés dans la nouvelle version du génome de *M. larici-populina*, contre 315 dans la version 1.0 (Table S2). Une large majorité des CAZymes identifiées dans la version 1, sont conservées dans la version 2 du génome et les 62 nouvelles CAZymes correspondent à 11 CBM, 10 CE, 20 GH, 1 PL et 9 GT (Table S2). Un répertoire réduit en GH avait été mis en évidence dans la première version du génome de *M. larici-populina* (173 GH), désormais 193 GH sont présentes dans le génome, ce qui est comparable au répertoire du champignon saprotrophe *Coprinopsis cinerea* (188 GH). Comme observé dans la première version du génome, une expansion modérée est observée pour certaines familles de GH impliquées dans la dégradation de la

paroi végétale par clivage de la cellulose et hémicellulose (GH5 ; GH7, GH10, GH12, GH26, GH27 et GH47; Table S2). On note aussi un nombre très restreint de PL (7) et CBM (16) également impliquées dans la dégradation de la paroi végétale. Enfin le génome de *M. larici-populina* contient 20 protéines apparentées aux expansines végétales qui jouent un rôle dans la relaxation de la paroi (contre 9 identifiées dans la version 1.0 ; Table S2 ; Cosgrove, 2000). Au total, 345 protéases ont été annotées dans la nouvelle version du génome de *M. larici-populina*, contre 402 dans la version 1.0. La nouvelle annotation confirme les expansions observées dans la première version, notamment un arsenal important d'endo- et d'exo-protéases (A01 ; S8 ; S9 et S10 ; Table S3).
Au total, 347 transporteurs ont été annotés dans la version 2.0 du génome de *M. larici-populina* parmi lesquels 107 ne sont pas retrouvés dans les 356 transporteurs rapportés dans la version 1.0 (Table S1). Il est intéressant de noter que les trois familles de transporteurs parmi les plus abondants dans la version 1.0 du génome de *M. larici-populina* (50 transporteurs ABC, 88 transporteurs MSF et 22 transporteur OPT), se retrouvent diminuées de moitié voire plus dans la version 2 .0 (19 transporteurs ABC, 44 transporteurs MSF et 10 transporteur OPT). Les transporteurs de ces familles sont impliqués dans la réponse aux stress, le transport de sucres et le transport d'oligopeptides, respectivement. Parmi les 107 « nouveaux » transporteurs identifiés dans la nouvelle version du génome, 24 font partie de la famille *eukaryotic nuclear pore complex* (E-NPC), 22 de la famille des *H+/Na+-translocating NADH Dehydrogenase* (NDH) et 10 de la famille des *Nuclear mRNA Exporter* (mRNA-E ; Table S1). Ces résultats soulignent les différences entre les pipelines d'annotation automatiques comme celui que nous avons mis en place pour l'annotation des transporteurs de la version 2 du génome de *M. larici-populina*.

Nous avons cartographié les catégories annotées présentées ci-dessus sur les 18 LG du génome de *M. larici-populina* afin de déterminer si certaines régions du génome présentaient une densité particulière pour certains types de gènes (Figure 14). Les gènes codant des protéines sécrétées (SP/SSP), des CAZYmes, des protéases et des transporteurs sont répartis uniformément le long des 18 groupes de liaisons du génome et nous n'observons pas de point chaud particulier (Figure 14).

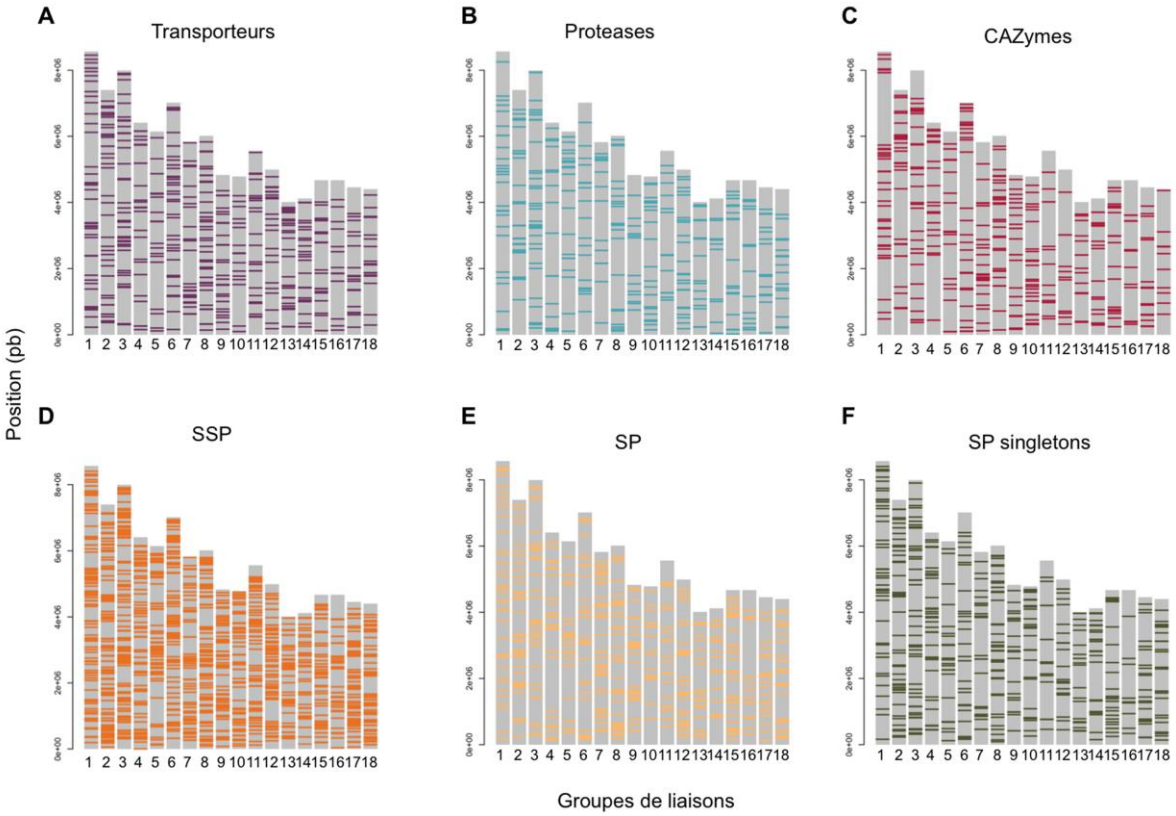


Figure 14 : Position de différentes catégories de gènes annotés de *M. larici-populina* 2.0 sur les 18 groupes de liaisons.

A. position des transporteurs prédits par alignement BLASTP du protéome de *M. larici-populina* 2.0 contre la base de données TCDB (« Transporter Classification Database ») avec une « e-value » seuil de $>10^{-5}$, les transporteurs prédits et présentant un domaine PFAM ont été considérés comme transporteurs de *M. larici-populina* 2.0 B. position des protéases de *M. larici-populina* ont été annotées sur la base de données MEROPS. C. position des CAZymes (« Carbohydrate active enzymes ») ont été annotées sur la base de données CAZY. D position des petites protéines sécrétées (SSP) de moins de 300 acides aminés. E. Position des protéines sécrétées (SP) de plus de 300 acides aminés. F. Position des SP et SSP singletons sur les 18 chromosomes de *M. larici-populina* 2.0.

TCDB CATEGORY	***#Mlp v1***	***#Mlp v2***
Total	356	347
ATP-binding Cassette (ABC) Superfamily	50	19
ACT (Amino Acid/Choline Transporter)	5	-
LAT (L-type Amino Acid Transporter)	1	-
YAT (Yeast Amino Acid Transporter)	11	-
Amino Acid/Auxin Permease (AAAP) Family	5	4
Acetate Uptake Transporter (AceTr) Family		4
Amino Acid-Polyamine-Organocation (APC) Family	21	13
Ammonium Transporter Channel (Amt) Family	4	4
Anion Exchanger (AE) Family	1	1
Ca2+:Cation Antiporter (CaCA) Family	3	4
Ca2+:H+ Antiporter-2 (CaCA2) Family	-	4
Calcium Transporter A (CaTA) (formerly Testis-Enhanced Gene Transfer (TEGT) Family		2
Calcium-Dependent Chloride Channel (Ca-ClC) Family		1
Cation Diffusion Facilitator (CDF) Family	5	6
Cation-Chloride Cotransporter (CCC) Family	1	1
Chloride Carrier/Channel (ClC) Family	3	3
Concentrative Nucleoside Transporter (CNT) Family	1	1
CorA Metal Ion Transporter (MIT) Family	5	1
Drug/Metabolite Transporter (DMT) Superfamily	9	17
General Secretory Pathway (Sec) Family	7	8
Glycerol Uptake (GUP) Family	1	1
Glycoside-Pentoside-Hexuronide (GPH):Cation Symporter Family	2	1
H+- or Na+-translocating F-type, V-type and A-type ATPase (F-ATPase) Superfamily	20	17
K+ Transporter (Trk) Family	1	1
K+ Uptake Permease (KUP) Family	1	1
Major Facilitator Superfamily (MFS)	88	44
Major Intrinsic Protein (MIP) Family	15	13
Metal Ion (Mn2+-iron) Transporter (Nramp) Family	4	3
Mitochondrial and Plastid Porin (MPP) Family	1	2
Mitochondrial Carrier (MC) Family	27	26
Mitochondrial Protein Translocase (MPT) Family	18	12
Mitochondrial Tricarboxylate Carrier (MTC) Family	1	1
Multidrug/Oligosaccharidyl-lipid/Polysaccharide (MOP) Flippase Superfamily	2	2
Non-selective Cation Channel-2 (NSCC2) Family	1	1
Nucleobase:Cation Symporter-1 (NCS1) Family	3	3
Oligopeptide Transporter (OPT) Family	22	10
Proton-dependent Oligopeptide Transporter (POT/PTR) Family	1	1
P-type ATPase (P-ATPase) Superfamily	21	11
Resistance-Nodulation-Cell Division (RND) Superfamily	1	1
Sulfate Permease (SulP) Family	3	4

Voltage-gated Ion Channel (VIC) Superfamily	1	1
Zinc (Zn2+)-Iron (Fe2+) Permease (ZIP) Family	1	2
Annexin	1	-
ArAE (Aromatic Acid Exporter)	1	-
ArsAB (Arsenite-Antimonite Efflux Family)	1	-
Arsenical Resistance-3 (ACR3)	1	1
ATP-E (ATP Exporter)	1	-
Carnitine O-Acyl Transferase (CrAT) Family	-	2
Cation Channel-forming Heat Shock Protein-70 (Hsp70) Family	-	6
Chloroplast Envelope Protein Translocase (CEPT or Tic-Toc) Family	-	2
Copper Transporter (Ctr) Family	1	2
CPA1 (Monovalent Cation:Proton Antiporter-1)	3	2
CPA2 (Monovalent Cation:Proton Antiporter-2)	1	1
DASS (Divalent Anion:Na^+ Symporter)	1	
Ductin (Putative Ductin Channel)	3	-
Endoplasmic Reticular Retrotranslocon (ER-RT) Family	-	3
Eukaryotic Nuclear Pore Complex (E-NPC) Family	-	24
Fatty Acid Transporter (FAT) Family	-	2
FP (Ferroportin)	1	-
H+ or Na+-translocating NADH Dehydrogenase (NDH) Family	-	22
Ion-translocating Microbial Rhodopsin (MR) Family	-	2
Iron/Lead Transporter (ILT) Family	1	1
Lysosomal Cystine Transporter (LCT) Family	-	1
Magnesium Transporter1 (MagT1) Family	-	1
Mg2+ Transporter-E (MgtE) Family	1	1
MHP (Metal Homeostasis Protein)	1	-
Mid1 (Yeast Stretch-Activated, Cation-Selective, Ca^{2+} Channel)	1	6
Mitochondrial Inner Membrane K+/H+ and Ca2+/H+ Exchanger (LetM1) Family	-	1
Na+-transporting Carboxylic Acid Decarboxylase (NaT-DC) Family	-	1
Nuclear mRNA Exporter (mRNA-E) Family	-	10
Oxa1 (Cytochrome Oxidase Biogenesis)	1	-
Peroxisomal Protein Importer (PPI) Family	-	5
PF27	1	-
Phagocyte (gp91phox) NADPH Oxidase Family	-	5
PLI (Phospholipid Importer)	1	-
Pore-forming Amphipathic Helical Peptide HP(2-20) (HP2-20) Family	-	1
Prokaryotic Succinate Dehydrogenase (SDH) Family	-	4
Proton-translocating Cytochrome Oxidase (COX) Superfamily	-	8
Proton-translocating Quinol:Cytochrome c Reductase (QCR) Superfamily	-	2
SHP (Stress-Induced Hydrophobic Peptide)	4	-
Synaptosomal Vesicle Fusion Pore (SVF-Pore) Family	-	2
TDT (Telurite-resistance/Dicarboxylate Transporter)	1	-
TMS Recognition/Insertion Complex (TRC) Family	-	1
Vacuolar Iron Transporter (VIT) Family	-	2
YaaH (ATO)	4	-

Partie 4: Contribution à l'annotation du génome de *Melampsora allii-populina*

Le génome de *M. allii-populina* a été séquencé par une combinaison des deux technologies PacBio et Illumina permettant un assemblage de 335,7Mb sans gaps sur un ensemble de 3637 scaffolds (dont la taille et le nombre N50 sont respectivement de 0,19Mb et 440 scaffolds). Cet assemblage a été réalisé par le JGI et mis à disposition sur le site Mycocosm. L'annotation *de novo* a identifié un total de 23 089 gènes (Table 4).

Tableau 4: Statistiques de réassemblage du génome de *Melampsora allii-populina* 12AY07 en comparaison avec *M. larici-populina* 2.0

Métrique	*M. allii-populina*	*M. larici-populina* 2.0
Taille du génome, Mb	335.73	109.88
Scaffolds/Groupes de liaison (GL)	3637	18 GL and 488 small scafolds
N50† par taille de scaffold, Mb	0.19	5.82
N50† par nombre de scaffolds	440	8
Taux de lacunes, %	0	5.8
Gènes codant des protéines	23089	19550
Taille moyenne de séquence codante, nt	1495	1468
Nombre moyen d'exon par séquence	4.51	4.37
Taille moyenne des exons, nt	248	254
Taille moyenne des introns, nt	109	108

†La métrique N50 correspond au nombre N des plus larges scaffolds couvrant la moitié du génome total. La taille N50 montre la taille du plus petit scaffold parmi la métrique N50.

Dans cette partie sont présentés les résultats obtenus pour l'annotation du génome de *M. allii-populina*.

1. Annotation des TEs

La proportion de TE retrouvée dans le génome de *M. allii-populina* est estimée à 77%, indiquant une invasion très importante du génome et expliquant sa taille (336Mb) vis à vis du génome de l'espèce voisine *M. larici-populina* (110Mb) (Table 5). L'annotation de novo des TE dans le génome de *M. allii-populina* montre 45,9% de transposons à ADN, 47,7% de rétrotransposons et seulement 5,17% de TE non assignés (NoCat), (Table 5). Ce faible nombre de transposons sans catégories indique une bonne qualité de l'annotation. Les

transposons de classe I et II sont représentés à part égale dans le génome de *M. allii-populina*, contrairement à ce qu'on a pu observer dans le génome de *M. larici-populina*. La part importante de rétrotransposons de type LTR (41%) par rapport à *M. larici-populina* indique que cette catégorie a vraisemblablement un rôle important dans l'inflation du génome.

Tableau 5: Classification REPET des séquences consensus « full-length fragments » identifiées chez *M. alii-populina*.

TE consensus	Map FLF	Percentage on 2,322 FLF for Map	
LTR	956	41.17%	Class I 1109 (47.7%)
DIRS	62	2.67%	
LARD	7	0.30%	
LINE	68	2.93%	
SINE	5	0.22%	
NoCat	1	0.04%	
TRIM	10	0.43%	
TIR	784	33.76 %	Class II 1068 (45.9%)
MITE	129	5.56%	
Helitron	12	0.52%	
NoCat	125	5.38%	
Maverick	18	0.78%	
NoCat	120	5.17%	-

*NoCat représente les éléments transposables non identifiés par le pipeline REPET

2. Analyse comparative MCL

Comme pour l'analyse MCL centrée sur le génome de *M. larici-populina*, les proétines présentant un domaine PFAM attribué généralement aux TE ont été identifié. Ainsi, 32 possible TE au sein des familles de gènes de *M. allii-populina* ont été exclus de l'analyse MCL. Sur les 23057 gènes restant, 5465 sont des singletons (copie unique chez *M. allii-populina*) et 17592 appartiennent à des familles de 2 gènes ou plus. La famille la plus large contient 852 gènes, ce qui est particulièrement remarquable pour un champignon, et cette famille présente des homologues avec les autres génomes de Pucciniales (dont *M. larici-populina* ; 278 membres) et 1 gène homologue dans le génome de *M. violaceum*. Un ensemble de 26 familles multigéniques de plus de 50 membres est observé, mais représentant un nombre conséquent de gènes à l'échelle du génome, puisque ces 26 familles totalisent 3911 gènes du génome de *M. allii-populina*. Comme pour *M. larici-populina*, à l'exception d'une famille de plus de 50

gènes retrouvée en large nombre dans tous les génomes explorés (la même que celle identifiée dans l'analyse MCL centrée sur *M. larici-populina*), une très large majorité de ces familles est spécifique des Pucciniales ou de quelques génomes au sein des Pucciniales. Seules quelques unes présentent quelques homologues en dehors des Pucciniales, le plus souvent chez *M. violaceum*. Parmi ces grandes familles multigéniques, huit (représentant 2687 gènes) possèdent plus de 100 membres dont trois familles (représentant 1913 gènes) avec plus de 500 membres.

La comparaison des gènes codant des protéines homologues entre 22 génomes fongiques dont les trois espèces de Melampsoraceae séquencées, révèle que 89,5% (20 637 gènes) des gènes de *M. allii-populina* ont au moins un orthologue parmi les 22 espèces fongiques analysés (Figure 15). *M. allii-populina* présente 28,5% (6572) de gènes homologues spécifiques des Pucciniales dont 26,2% (6032) sont spécifiques des Melampsoraceae. Parmi ces derniers, on retrouve 1666 gènes spécifiques des deux rouilles du peuplier *M. larici-populina* et *M. allii-populina* et 2420 (10,4% des gènes totaux) gènes spécifiques de *M. allii-populina* (Figure 15). La divergence moléculaire observée à partir des pourcentages d'homologie de séquences au sein des orthogroupes au sein des 22 génomes analysés, montre bien le profil singulier des génomes des Melampsoraceae (*M. allii-populina, M. larici-populina et M. lini*) vis à vis des autres champignons (Figure 16).

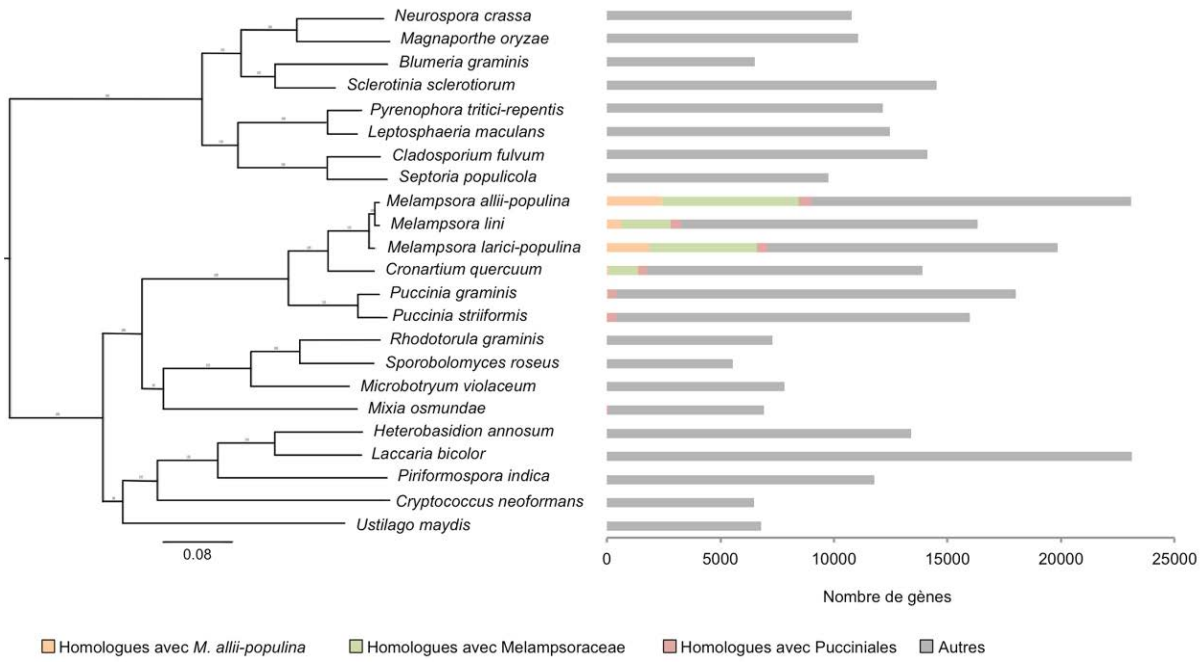


Figure 15 : Homologues partagés entre *M. allii-populina* et 22 espèces fongiques.
L'axe des abscisses montre le nombre de gènes parmi les génomes des 22 espèces fongiques. L'axe des ordonnées montre les 22 espèces fongiques sélectionnées pour l'analyse OrthoMCL. Les barres en orange représentent les homologues partagés avec *M. allii-populina* ; en vert les homologues partagés parmi deux Melampsoraceae (*M. allii-populina* ; *M. larici-populina* 2.0 et *M. lini*) ; en rouge les

homologues partagés parmi six Pucciniales (*M. allii-populina* ; *M. larici-populina* 2.0 ; *M. lini ; Cronartium quercuum ; Puccinia graminis* f. sp. *tritici ; Puccinia striifomis* f. sp *tritici*) et en gris les autres gènes de chaque génome fongique.

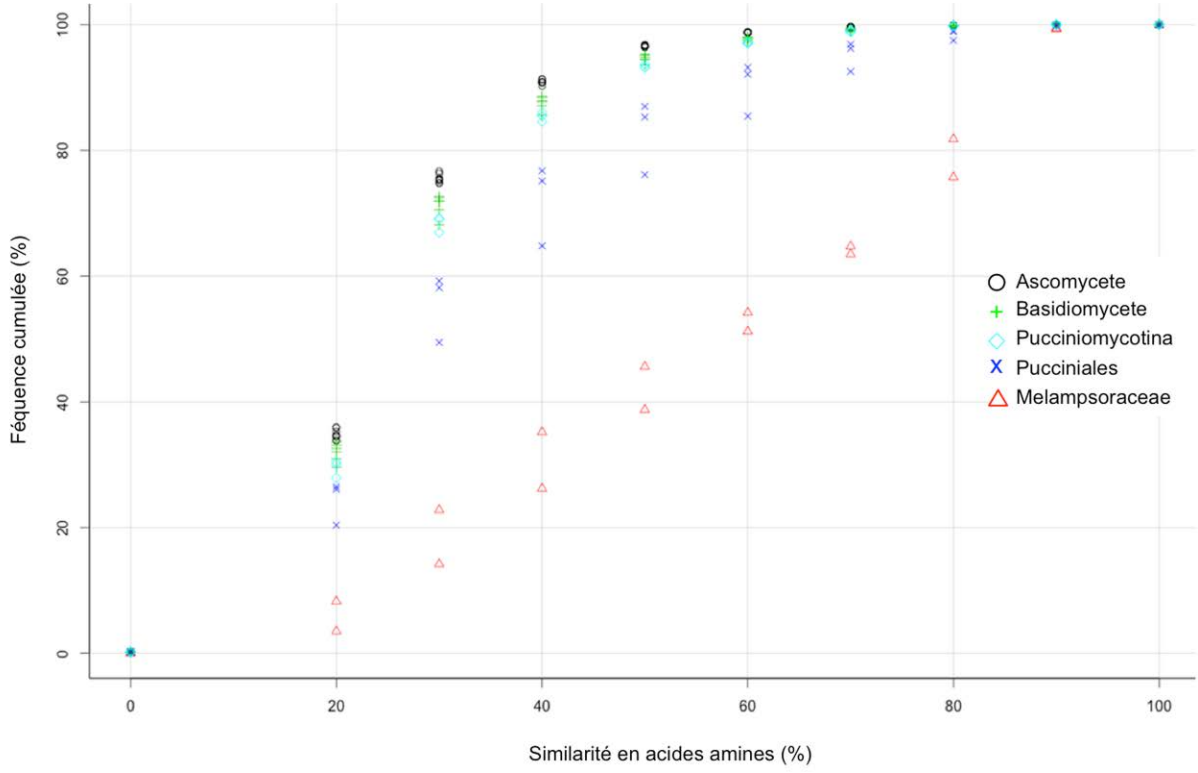


Figure 16 : Divergence moléculaire entre les 22 génomes fongiques analysés par OrthoMCL.
Les axes des abscisses et des ordonnées indiquent respectivement le pourcentage d'identité en acides aminés entre les protéomes des 22 génomes fongiques comparés et le pourcentage de la fréquence cumulative de la divergence moléculaire entre les 22 génomes fongiques groupés comme suit : Ascomycètes (*Blumeria graminis* f. sp. *tritici* 96224*; Septoria populicola ; Cladosporium fulvum; Leptosphaeria maculans ; Pyrenophora tritici-repentis; Sclerotinia sclerotiorum ; Magnaporthe oryzae ; Neurospora crassa* OR74A) ; Basidiomycètes (*Laccaria bicolor*, *Cryptococcus neoformans* var *neoformans*, *Piriformospora indica*, *Heterobasidion annosum*, *Ustilago maydis*) ; Pucciniomycotina (*Rhodotorula graminis* ; *Microbotryum violaceum*, *Mixia osmunde; Sporobolomyces roseus*) ; Pucciniales (*Puccinia graminis* f. sp. *tritici* v2.0; *Puccinia striiformis* f. sp. *tritici* PST130; *Cronartium quercuum* f. sp. *fusiforme* G11) ; Melampsoraceae (*M. allii-populina* 12AY07; *M. larici-populina* 98AG31 v2; *Melampsora lini* CH5).

3. Prédiction du sécrétome

Les outils de prédiction utilisés ont mis en évidence 2402 protéines possédant un peptide signal dont 2118 ne contiennent ni domaine transmembranaire, ni motif de rétention dans le réticulum endoplasmique (Figure 17). Grâce à l'annotation des familles multigéniques de *M. allii-populina* décrite précédemment, nous avons pu mettre en évidence 448 familles de SP de plus de deux membres. Ainsi le sécrétome final de *M. allii-populina* comprend 2177 SP dont 57% font partie de familles de SP de plus de deux membres, 20% font partie de familles multigéniques de trois membres au moins et 23% sont des singletons (Figure 17). Parmi ces

2177 SP, 61,1% (1337) sont des gènes codant des SSP de moins de 300 acides aminés. L'outil de prédiction EffectorP a permis de mettre en évidence 826 effecteurs candidats rassemblant des caractéristiques typiques des effecteurs fongiques. Le sécrétome de *M. allii-populina* contient également 1500 SP prédites comme non-apoplastiques et 674 sont prédites comme résidente de l'apoplaste. Nous avons également utilisé l'outil LOCALIZER sur le sécrétome de *M. allii-populina* et prédit 313 SP ciblant les chloroplastes, 177 les mitochondries et 411 les noyaux des cellules végétales.

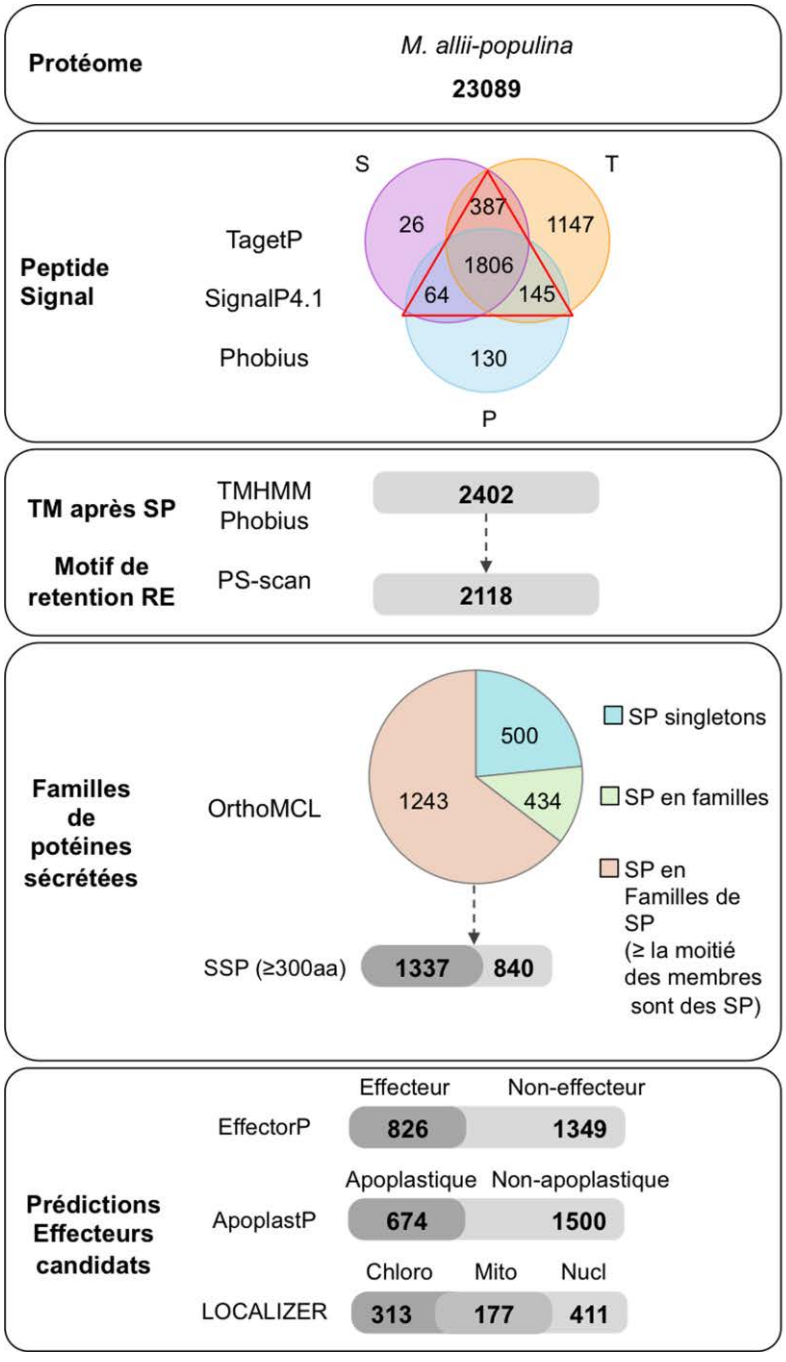


Figure 17 : Prédiction et analyse du sécrètome de *M. allii-populina*

Le sécrètome de *M. allii-populina* a été prédit en utilisant les trois outils de prédiction de peptide signal suivants : TargetP1.1 (T) ; SignalP4.1 (S) et Phobius (P). Une protéine donnée a été considérée comme prédite sécrétée « Secreted Protein » (SP) si au moins deux sur trois prédictions indiquent la présence d'un peptide signal. La présence dans la séquence protéique de chaque protéine de *M. allii-populina* de domaines transmembranaires et de motifs de rétention dans le réticulum endoplasmique est prédite par les outils de prédiction TMHMM et PS-scan, respectivement. Les SP protéines avec plus d'un domaine transmembranaire prédit ou avec un domaine transmembranaire prédit en dehors des 60 premiers acides aminés ne sont pas considérées comme sécrétées. De même les protéines contenant un motif de rétention dans le réticulum endoplasmique prédit sont éliminées du sécrètome. Les familles multigéniques déterminées par analyse MCL qui contiennent au moins la moitié des membres prédits comme sécrétés sont considérées comme des familles multigéniques de protéines sécrétées (Familles de SP). Parmi ces familles, les protéines non prédites comme sécrétées par au moins deux des trois outils de prédiction des protéines sécrétées mais montrant une structure introns/exons similaire aux autres

membres sécrétés de la

Figure 17 : Prédiction et analyse du sécrètome de M. allii-populina
famille, sont ajoutées au sécrètome de *M. allii-populina*. Les petites protéines sécrétées de moins de 300 acides aminés sont identifiées comme SSP (« small secreted proteins »). Enfin les effecteurs candidats de *M. allii-populina* sont prédits parmi le sécrètome avec l'outil de prédiction EffectorP et leur localisation intracellulaire *in planta* est prédite par les deux outils ApoplastP et LOCALIZER.

Dans le répertoire de 2177 protéines sécrétées de *M. allii-populina*, 232 SP (10,7%) sont spécifiques de l'espèce, 391 SP (18,0%) sont spécifiques des deux rouilles du peuplier, 480 SP (22,0%) sont spécifiques des Melampsoraceae et 383 SP (17,6%) sont spécifiques des Pucciniales. Au total, 1486 SP de *M. allii-populina* sont partagées parmi les Pucciniales. Parmi les protéines sécrétées de *M. allii-populina*, nous avons recherché les homologues des effecteurs caractérisés chez les autres Pucciniales. On retrouve des homologues des effecteurs connus de *M. lini* : 8 SP homologues de l'effecteur AvrL567, 5 SP (et 3 protéines non sécrétées) homologues de AvrM, 4 de AvrP4 et 1 de AvrP123 et 3 SP homologues de RTP1. Le sécrétome de *M. allii-populina* est composé à 66,5% (1448 SP) de protéines sans fonctions et sans domaine connus. Parmi les 33,5% (729 SP) avec une fonction prédite par les bases de données GO, KOG, IPR ou annotées spécifiquement, on trouve spécifiquement 9,2% (201 SP) de CAZymes % (53 SP) de prtoéases et 3 SP transporteurs. Les fonctions prédites sur la base de données KOG présentent 1738 SP sans fonction KOG et 546 SP avec des fonctions prédites (Figure 18). On retrouve notamment, 3 catégories avec plus de 50 gènes avec des 82 protéines impliquées dans le métabolisme et le transport carboné (catégorie KOG *Carbohydrate transport and metabolism*, en agnlais) ; 74 attribuées aux modifications post-traductionnelles, renouvellement des protéines et chaperonnes (catégorie KOG *Posttranslational modification, protein turnover, chaperones* en anglais) et 57 SP sont catégorisées dans les « Fonctions générales » (catégorie KOG *General function prediction* only en anglais). Les catégories *Lipid transport and metabolism* (22 SP), Transcription (24 SP), Inorganic ion transport and metabolism (27 SP), Extracellular structures (28 SP), Nuclear structure 29 SP), et Signal transduction mechanisms (37 SP) représentent chacune plus de 1% du sécrétome de *M. allii-populina*. Enfin, comme pour *M. larici-populina* on retrouve 16 protéines portant un domaine CFEM dont 11 sont sécrétées. Ces résultats montrent les caractéristiques du sécrétome de *M. allii-populina*.

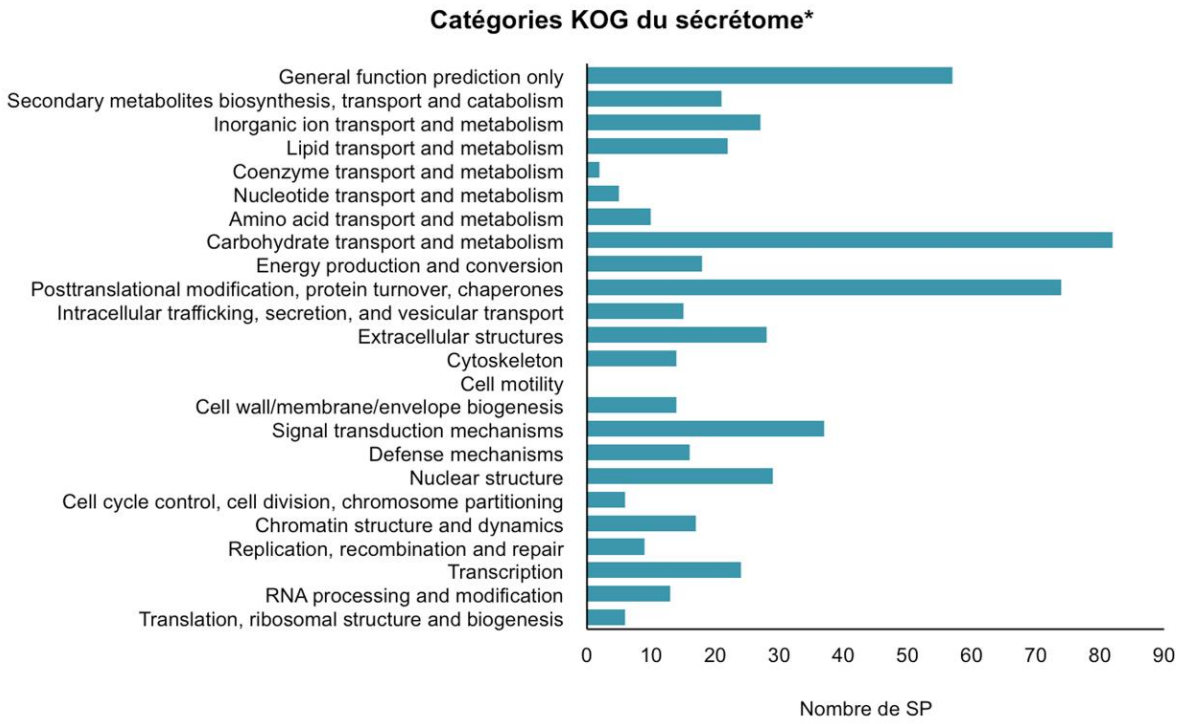


Figure 18 : Catégories KOG annotées dans le sécrétome de *M. allii-populina*
Les axes des abscisses et des ordonnées indiquent respectivement le nombre de protéines sécrétées (SP) et chaque catégorie KOG. *Les SP sans fonction KOG ne sont pas affichées sur ce graphique et représentent 1609 SP.

Avec un génome trois fois plus grand que celui de *M. larici-populina*, l'hypothèse d'une association entre régions riches en TE et en gènes SP/SSP pourrait être envisagée. Nous avons ainsi cartographié les régions intergéniques du génome et du sécrétome afin de la tester. Contrairement à ce qui a été décrit chez l'oomycète *P. infestans*, et de manière similaire à *M. larici-populina*, le sécrétome de *M. allii-populina* présente un profil similaire à celui du génome entier et se superpose principalement aux régions denses en gènes (Figure 19). Toutefois, la représentation des gènes de *M. allii-populina* en fonction de leur distance intergénique sur la Figure 13 présente un profil très particulier avec un groupe principal en bas à gauche du graphique (faible distance en 5' et 3' entre gènes adjacents) et trois autres groupes avec des distances distinguables (et similaires entre l'ensemble des gènes et ceux du sécrétome). Il serait particulièrement intéressant de regarder la distribution des catégories de gènes, et de TE, ainsi que de familles multigéniques appartenant à ces groupes et déterminer ainsi si la grande plasticité génomique générée par l'invasion en TE a pu avoir une influence sur des familles de gènes au cours de l'évolution. Il est possible que les groupes en haut à gauche et en bas à droite soient des artefacts créés par l'assemblage encore très incomplet du génome.

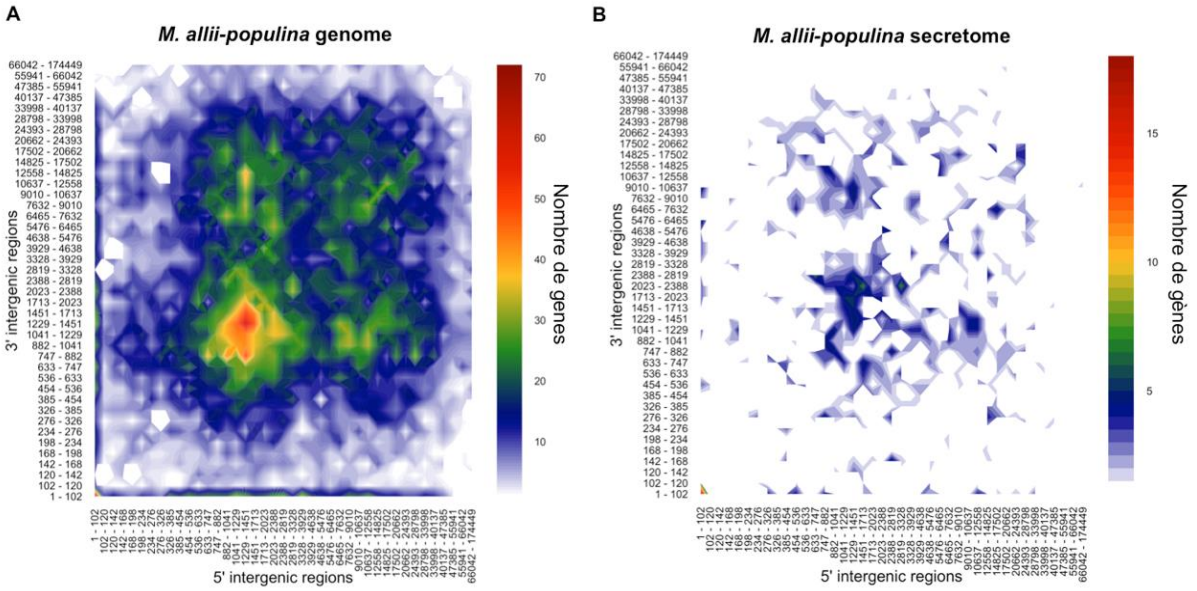


Figure 19 : Environnement génique des génome et sécrètome de *M. allii-populina*.
Les axes des abscisses et des ordonnées indiquent respectivement régions intergéniques 5' et 3' du génome de *M. allii-populina*. Les gènes de *M. allii-populina* ont été triés selon des « fenêtres » à deux dimensions sur la base de la distance avec les régions intergéniques environnantes en 5' et 3'. A. « Heatmap » représentant la distribution des distances intergéniques en 5' et 3' pour tous les gènes de *M. allii-populina*. B. pour le sécrètome de *M. allii-populina*. Les échelles représentent le nombre de gènes par « fenêtre ».

Cette première analyse du sécrétome de *M. allii-populina* permet de mettre à jour le sécrétome de *M. allii-populina*, qui montre les caractéristiques communes des sécrétomes des Pucciniales comme par exemple une majorité de SP de fonction inconnue. Il est à noté que 10,7% des SP sont spécifiques à l'espèce. Ce répertoire contient potentiellement des effecteurs impliqués dans les infections des hôtes écidiens (*Allium* spp.) de *M. allii-populina*.

4. Annotation des familles de gènes codant des CAZymes, des protéases et des transporteurs

Au total, 432 CAZymes ont été annotées dans le génome de *M. allii-populina* (Table S6 ; 11 CMB ; 49 CE ; 207 GH ; 16 PL ; 89 GT and 21 Expansines). Un ensemble de 207 GH est mis à jour, ce qui est comparable au nombre rapporté chez *M. larici-populina*. Une expansion plus ou moins modérée est observée pour certaines familles de GH impliquées dans la dégradation de la paroi végétale par clivage de la cellulose et hémicellulose dans ce génome (GH5 ; GH7, GH10, GH12, GH26, GH27 et GH47; Table S6). *M. allii-populina* contient un répertoire restreint de PL (16) et CBM (11) également impliquées dans la dégradation de la paroi végétale. Enfin, 21 protéines apparentées aux expansines végétales (Table S6 ; Cosgrove, 2000). Les profils en CAZymes des deux rouilles du peuplier ne présentent que quelques différences en termes de nombre de gènes, qui pourraient refléter une spécificité vis à vis des

hôtes alternants infectés par ces deux champignons. Sept CAZymes identifiées chez *M. allii-populina* ne sont pas retrouvées chez *M. larici-populina* (4 *Carbohydrate-Binding Module Family 48*; 1 *Carbohydrate-Binding Module Family 13*; 1 *Glycoside Hydrolase Family 1*; 1 *Glycoside Hydrolase Family 13 /Glycoside Hydrolase Family 133*). Trois GH sont présentes en nombre plus important chez *M. allii-populina* (7 GH2 ; 7 GH15 ; 6 GH92 contre 4 GH2 ; 3 GH15 et 2 GH92 chez *M. larici-populina*) et 1 famille de PL semble en expansion chez *M. allii-populina* (8 contre 1 chez *M. larici-populina*).

Au total, 417 protéases ont été annotées dans le génome de *M. allii-populina*, soit 72 de plus en comparaison aux 345 identifiées chez *M. larici-populina*. Il est intéressant de noter qu'aucune famille de protéases n'est présente que chez M. allii-populina et non chez M. larici-populina. Cependant quelques différences sont observables dans le nombre de membre de certaines familles. La différence la plus marquée se fait chez les inhibiteurs de protéases dont ont en retrouve 125 chez *M. allii-populina* contre 98 chez *M. larici-populina* (par exemple : 9 *Alpha-1-peptidase inhibitor* chez *M. allii-populina* contre 4 chez *M. larici-populina* ; 27 inhibiteurs protéases I39 contre 22, respectivement). On observe le même schéma dans les autres familles, très peu semblent en expansion chez M. allii-populina et non chez *M. larici-*populina comme par exemple : les peptidases N10 (respectivement 7 contre 1), les peptidases S09C (respectivement 6 contre 2), M23B (respectivement 6 contre 3) ou encore C50 (5 contre 1).

Au total, 341 transporteurs ont été annotés dans le génome de *M. allii-populina* (Table S5). Parmi les familles de transporteurs les plus abondantes dans le génome de *M. allii-populina* sont retrouvés les transporteurs ABC (26), les transporteurs MSF (45) et les transporteurs OPT (13). On retrouve également, comme pour *M. larici-populina*, 18 transporteurs de la famille E-NPC, mais aucun NDH ou mRNA-E (Table S5). Par ailleurs, 17 transporteurs sont annotés chez M. allii-populina uniquement, et non chez M. lairici-populina parmi lesquels on trouve notamment 4 *Mechanosensitive Calcium Channels* ; 4 Choline/Ethanolamine Phosphotransferase 1 ; 3 Zinc (Zn2+)-Iron (Fe2+) Permease. Les autres familles de transporteurs spécifiques de *M. allii-populina* n'ont qu'un ou deux membre (Table S5). Ces résultats montrent que les deux génomes des rouilles du peuplier partagent des catalogues de gènes codant des CAZymes, protéases et transporteurs similaires avec dans le détail des différences qui représentent autant d'indices de spécificité de chaque espèce.

5. Comparaison de *M. allii-populina* et *M. larici-populina* : orthologues sous pression de sélection

Des résultats préliminaires ont été obtenus pour évaluer si une pression évolutive s'exerçait sur les gènes des deux rouilles du peuplier (Figure 20). Le ratio *dN/dS* a été utilisé pour

déterminer le mode de sélection s'appliquant aux gènes. Ce paramètre correspond au ratio des substitutions non-synonymes sur les substitutions synonymes par paire de gènes orthologues. Un ratio *dN/dS* supérieur à 1 est interprété comme une signature de sélection positive c'est-à-dire que la sélection promeut les substitutions non-synonymes dans les séquences. Un *dN/dS* inférieur à 1 indique le maintien des séquences dans les génomes, les substitutions non-synonymes n'étant pas sélectionnées. L'estimation du ratio *dN/dS* entre *M. larici-populina* et *M. allii-populina* a été réalisée pour 8045 paires de gènes orthologues sélectionnées sur la base des homologies réciproques retrouvées à l'aide de l'outil BLASTP (*best-reciprocal hits* ; BRH). Le ratio moyen du nombre de substitutions non-synonymes par site sur le nombre de substitution synonymes par site entre les deux Melampsoraceae a été estimé à 0,28 (Figure 20). Au total, 98,8% des BRH ont un ratio *dN/dS* inférieur à 1 et 1,2% un ratio supérieur à 1 (Figure 20). Ces derniers correspondent à 91 BRH de *M. larici-populina* et *M. allii-populina* avec un ratio *dN/dS* supérieur à 1 parmi lesquels respectivement 98% (89 gènes) et 93% (85 gènes) sont de fonction inconnue. Les deux champignons présentent respectivement 23 (25%) et 27 (29%) gènes codant des petites protéines sécrétées dont le ratio *dN/dS* est supérieur à 1.

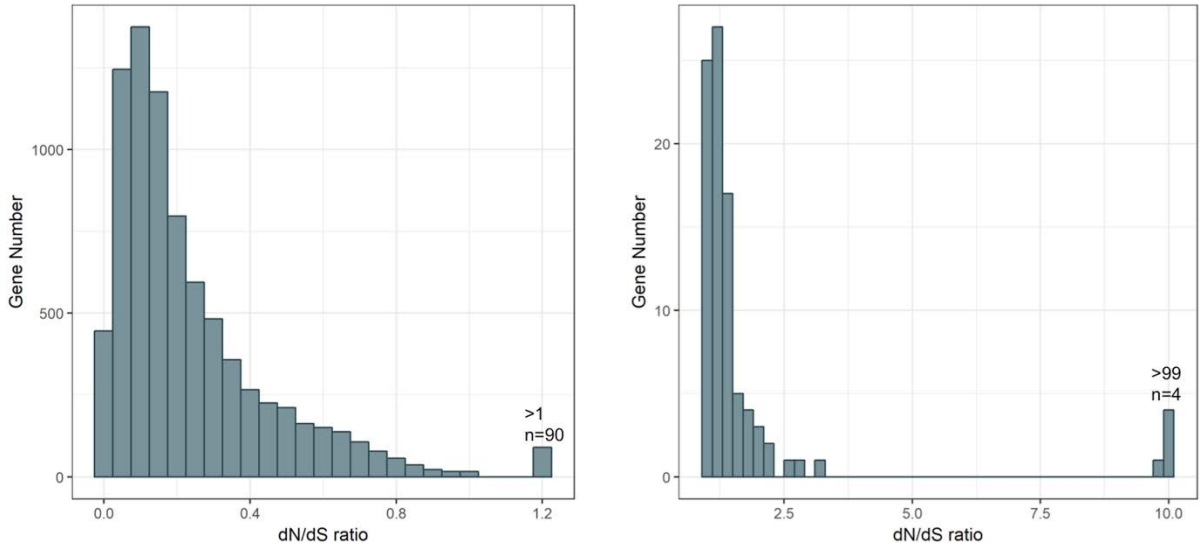


Figure 20 : Distribution des ratios *dN/dS* pour les orthologues identifiés entre *M. allii-populina* et *M. larici-populina*.

Les axes des abscisses et des ordonnées indiquent respectivement le ratio *dN/dS* et le nombre de gènes pour toutes les paires de gènes orthologues (gauche) et pour ceux dont le ratio est supérieur à 1 (à droite) identifiées entre *M. larici-populina* et *M. allii-populina*.

Partie 5 : Conclusions

M. larici-populina et *M. allii-populina* sont deux espèces de Pucciniales de la famille des Melampsoraceae, infectant des peupliers, partageant ainsi un hôte télien, mais alternant respectivement sur conifères (*Larix* spp.) ou monocots (*Allium* spp.). Le séquençage de ces deux génomes a permis de mettre en évidence, l'expansion remarquable d'un facteur trois de la taille du génome de *M. allii-populina*. Cette expansion est fortement liée à l'invasion du génome par de nombreux TE qui constituent 77% du génome (contre 42% chez *M. larici-populina*). De plus, la part de rétrotransposons dont le mécanisme de transposition se fait par « copier-coller » est nettement plus importante chez *M. allii-populina* par rapport à *M. larici-populina* (48% et 29%, respectivement). On trouve également chez *M. allii-populina* un nombre de gènes plus important (23089 gènes contre 19550 chez *M. larici-populina*). Cependant, l'annotation du sécrétome, des transporteurs, des CAZymes et des protéases montrent des répertoires très similaires en proportion et en composition. En effet, les familles de CAZymes, protéases et transporteurs qui étaient en expansion chez *M. larici-populina* vis à vis d'autres génomes fongiques, le sont également dans le génome de *M. allii-populina*.

De même les deux sécrétomes de *M. larici-populina* et *M. allii-populina* partagent de nombreux points communs à commencer par 22% des SP qui sont spécifiques aux deux rouilles du peuplier suggérant ainsi un core-sécrétome permettant l'infection d'un même hôte, le peuplier. Les deux Melampsoraceae possèdent des sécrétomes avec une majorité de protéines de fonctions inconnues, une caractéristique commune des sécrétomes des Pucciniales (Lorrain *et al.,* 2015 ; Duplessis *et al.,* 2014). Dans la fraction des sécrétomes de *M. larici-populina* et *M. allii-populina* montrant des fonctions prédites par les bases de données, quelques catégories communes sont retrouvées en abondance dans le sécrétome comme les CAZymes et les protéines de fonction associées au transport et métabolisme carboné, des fonctions associées aux modifications post-tranductionnelles, renouvellement des protéines et chaperones. Il est également à noter que les deux rouilles du peuplier présentent des homologues aux effecteurs Avr de *M. lini* et à RTP1 de *U. fabae* et *U. striatus* (voir Petre *et al.,* 2014 pour synthèse). Dans le détail, il existe tout de même un certain nombre de petites différences qui représentent autant de spécificité à mettre en regard de la biologie propre de chaque espèce et son adaptation à des hôtes alternants différents.

Chez *M. larici-populina*, les gènes codant des protéines sécrétées (SP/SSP), des CAZYmes, des protéases et des transporteurs sont répartis uniformément le long des 18 groupes de liaisons du génome et nous n'observons pas de point chaud particulier (Figure 13 ; Figure 19). Une telle distribution mise en parallèle de celle des éléments transposables semble indiquer

que les génomes des Pucciniales diffèrent d'autres pathogènes avec une compartimentation permettant une évolution différente et accélérée pour certaines régions liées à la pathogénèse (Dong *et al.,* 2015).

L'annotation des deux génomes de *M. larici-populina* et *M. allii-populina* met en lumière que plus de la majorité des protéines n'ont pas de fonction connue. Ce chiffre doit tout de fois être modéré puisque certaines annotations dans ces bases restent floues et peu informatives. On peut donc considérer que plus de 60% des gènes sont de fonction inconnue. Donc depuis la première analyse de génomes de Pucciniales (Duplessis *et al.,* 2011a), et malgré sept années écoulées, la proportion de gènes avec une annotation reste similaire. Les génomes des Pucciniales ont donc la particularité de présenter une large majorité de gène sans fonction associée.

Un modèle de « génome à deux vitesses » a pu être mis en évidence chez certains pathogènes filamenteux, au sein duquel les gènes codant des effecteurs présentent des vitesses d'évolution plus rapides au sein de régions à faible densités de gènes et riches en TE (Raffaele & Kamoun, 2012). Nous avons cartographié l'environnement génique des sécrétomes dans ces deux génomes de Pucciniales. Chez *M. larici-populina* et *M. allii-populina* la part de TE se situe entre 40 et 70% et ils se répartissent relativement uniformément sur l'ensemble du génome. De ce fait, les SP ne ressortent pas plus proches de TE que le reste des autres gènes. Cette observation semble une caractéristique commune des Pucciniales. En effet les génomes récemment séquencés de *P. striiformis* f. sp. *tritici* 104E et de *Puccinia coronata* f. sp. *aveanae* ne montrent pas d'association de proximité entre TE et gènes codant des effecteurs candidats (Miller *et al.,* 2017; Schwessinger *et al.,* 2017). Il semblerait donc que les TE aient envahi de manière homogène l'ensemble des génomes des Pucciniales, incluant ceux des deux Melampsoraceae étudiées ici. Ce mécanisme d'invasion des génomes de Pucciniales par les TE pourrait donc être relativement ancien et il serait intéressant d'explorer l'évolution des familles de TE annotées dans ces deux génomes. Bien que nous n'ayons pas pu montrer de localisation particulière de gènes codant des protéines sécrétées dans les régions riches en TE, nous avons pu mettre à jour un profil particulier vis à vis de la densité en gènes dans le génome de *M. allii-populina*. Il serait tout à fait pertinent d'étudier les caractéristiques sous-jacentes à ces groupes et de voir notamment si des gènes, et/ou des TE sont plus particulièrement représentés, ce qui révèlerait une compartimentation spécifique dans ce génome. Il faut garder en mémoire que l'assemblage de ce génome est encore imparfait avec plus de 3000 scaffolds, et qu'il faudra aussi vérifier si ces observations ne sont pas simplement artéfactuelles.

L'annotation de plusieurs familles d'intérêt a été réalisée dans les génomes de *M. larici-populina* et de *M. allii-populina*. Les profils de ces familles de gènes représentent un support important pour mieux comprendre la croissance biotrophe du champignon chez ses hôtes. Il sera particulièrement intéressant de voir si les profils d'expression des gènes que nous mettons à jour sont associés à la spécificité vis à vis des hôtes de ces Melampsoraceae.

Enfin, nous avons pu identifier des gènes orthologues de *M. allii-populina* et *M. larici-populina* montrant des signatures de sélection. L'estimation moyenne du ratio dN/dS entre *M. larici-populina* et *M. allii-populina* (0,28) est comparable au ratio *dN/dS* estimé entre plusieurs espèces de *Phytophthora* (Raffaele *et al.,* 2013) et entre *Blumeria graminis* f. sp. *tritici* et *Blumeria graminis* f. sp. *hordei* (Wicker *et al.,* 2013). Notamment, 91 paires de BRH avec un ratio dN/dS supérieur à 1 dont près de 25% sont des petites protéines sécrétées. Ainsi, ces gènes identifiés sous sélection diversifiante sur la base de la comparaison entre les deux espèces sont principalement des gènes codant soit des protéines de fonction inconnue dont 26% pour *M. larici-populina* et 32% pour *M. allii-populina* sont des petites protéines sécrétées. En considérant que certains de ces gènes sous sélection pourraient être associés à la pathogénicité sur les plantes hôtes, les petites protéines sécrétées pourraient représenter des effecteurs candidats potentiels.

Une discussion plus détaillée des résultats obtenus dans ce Chapitre est présentée dans le chapitre V de ce manuscrit de thèse.

CHAPITRE IV

Analyse fonctionnelle d'effecteurs candidats de *M. larici-populina*

Partie 1 : Introduction

Les champignons de l'ordre des Pucciniales sont des biotrophes obligatoires qui ne peuvent être cultivés en dehors de leurs hôtes (Aime *et al.,* 2017). De ce fait, leur manipulation en conditions contrôlées est compliquée et elle nécessite de disposer de cultures de l'hôte. Jusqu'à présent, les exemples de succès de transformation génétique de Pucciniales sont très limités (Djulic *et al.,* 2011; Lawrence *et al.,* 2010; Schillberg *et al.*, 2000). Par ailleurs, la plupart des Pucciniales sont pathogènes sur des plantes cultivées non modèles difficilement manipulables génétiquement en laboratoire. Enfin, le cycle de vie complexe des Pucciniales est rarement réalisable de manière contrôlée en laboratoire (Aime *et al.,* 2017 ; Pernaci *et al.,* 2014). Les pathosystèmes non modèles établis entre les Pucciniales et leurs plantes hôtes sont particulièrement complexes pour leur manipulation en laboratoire, rendant difficile l'étude expérimentale de la biologie de ces champignons.

L'étude de la biologie des effecteurs est un thème central et un front de science actuellement en phytopathologie moléculaire afin de comprendre comment se réalise l'infection d'une plante hôte par un agent pathogène et afin d'aider la sélection de résistances efficaces et durables contre les pathogènes (Dangl *et al.,* 2013; Hogenhout *et al.,* 2009). De plus en plus d'effecteurs ont été décrits au cours de la dernière décennie chez les bactéries et les microorganismes filamenteux phytopathogènes (Deslandes & Rivas, 2012; Lo Presti *et al*., 2015; Tyler & Rouxel, 2012). Ainsi, des caractéristiques communes (mais non exclusives) ont été mises en évidence chez les effecteurs protéiques et sont utilisées comme critères de sélection dans des pipelines d'identification d'effecteurs candidats. Ainsi la prédiction d'un signal de sécrétion, la taille de la protéine et sa composition, certaines caractéristiques génomiques, ou la spécificité de l'expression lors de l'interaction sont autant d'éléments qui peuvent conduire à la sélection d'effecteurs candidats.Ainsi, chez les champignons, plusieurs centaines de candidats ont été rapportés (Sperschneider *et al.,* 2015).

Des approches dites « effectoromiques » peuvent être développées pour affiner la sélection parmi ces effecteurs candidats. Des pipelines effectoromiques de type cribles à moyen ou haut débit peuvent être appliqués afin de classer par ordre de priorité des effecteurs candidats d'intérêt et de formuler des hypothèses quant à leur fonction. De telles approches ont été proposées pour déterminer les cibles cellulaires et moléculaires potentielles de sets de candidats (Alfano, 2009). L'expression des effecteurs étiquetés d'un marqueur par le pathogène chez l'hôte serait une approche idéale pour suivre leur translocation dans la cellule hôte au cours de l'infection et identifier ses cibles. Cependant, une telle approche reste inaccessible pour de nombreux pathosystèmes dont les Pucciniales (Petre & Kamoun, 2014).

Les approches effectoromiques peuvent être appliquées dans des systèmes hétérologues et des cribles ont ainsi été réalisés pour étudier les effecteurs candidats de plusieurs oomycètes tels que *Hyaloperonospora parasitica*, *Phytophthora capsici* ou *P. infestans* permettant des avancées sur la localisation de certains effecteurs candidats et l'identification d'interactants potentiels (Caillaud *et al.,* 2012; Schornack *et al.,* 2010; Stam *et al.,* 2013). D'autres approches effectoromiques vont plutôt viser à tester si le rôle des effecteurs peut être relié à la réponse immunitaire de type PTI ou ETI (par exemple à travers la suppression ou augmentation de la mort cellulaire) (ex. chez la rouille du soja *P. pachyrhizi*, Qi *et al.,* 2017; Sharpee *et al.,* 2017).

Depuis 2011, les annotations et prédictions *in sillico* du sécrétome de *M. larici-populina* ont contribué à mettre en évidence un large répertoire de gènes codant des protéines sécrétées dont plusieurs centaines de gènes exprimés lors de l'interaction représentant autant d'effecteurs candidats (Joly *et al.,* 2010; Hacquard *et al.,* 2010; Duplessis *et al.,* 2011a; Duplessis *et al.,* 2011b; Hacquard *et al.,* 2012; Petre *et al.,* 2012; Hacquard *et al.,* 2013; Pernaci *et al.,* 2014; Persoons *et al.,* 2014). *M. larici-populina* présente plus d'un millier de protéines sécrétées de fonction inconnue qui pourraient correspondre à des effecteurs dont 70% sont exprimés lors du processus d'infection (Duplessis *et al.,* 2011a ; Hacquard *et al.,* 2012).

Dans le cadre d'une collaboration avec The Sainsbury Laboratory (TSL, Sophien Kamoun & Benjamin Petre, Norwich, UK) et de mon stage de Master 2 BIOMANE (Université de Lorraine, Nancy), j'ai participé à des travaux qui visaient à étudier des effecteurs candidats de *M. larici-populina* et du champignon *P. striiformis* f. sp. *tritici* (l'agent de la rouille jaune du blé) par une approche effectoromique. Il s'agissait de suivre l'expression transitoire d'effecteurs candidats sélectionnés fusionnés à un marqueur fluorescent et dépourvus de leur peptide signal par Agroinfiltration de feuilles de *Nicotiana benthamiana*. Cette approche permet ainsi d'étudier où les effecteurs candidats sont localisés dans la cellule végétale les exprimant et aussi de rechercher de potentiels interactants protéiques (Petre et al. 2015; Petre et al. 2016a; Petre et al. 2015b). J'ai donc participé à l'élaboration des constructions génétiques et aux premières observations lors de mon stage de Master 2 au TSL puis j'ai participé à l'analyse des résultats et la rédaction des publications correspondantes lors de ma première année de thèse. Ma contribution est marquée par mon positionnement en tant que co-auteur sur deux articles scientifiques tirés des travaux présentés en annexe de cette thèse (cf. **Annexe 1 et 2; articles n°7 et 8**). Chez *M. larici-populina*, à partir d'un panel initial de 20 candidats sélectionnés, cette approche a permis de montrer une localisation particulière pour six candidats par imagerie confocale *in vivo*. Par ailleurs, des protéines végétales interagissant avec cinq effecteurs candidats et représentant de potentielles cibles ont été identifiées par co-immunoprécipitation

couplée à de la spectrométrie de masse. A l'issue de ces travaux, au début de ma thèse, nous disposions ainsi d'une liste réduite d'effecteurs protéiques candidats prometteurs qu'il restait à analyser plus avant sur le plan fonctionnel et structural.

La biologie structurale contribue significativement à la compréhension des fonctions, de la localisation et des interactions des effecteurs avec leurs interactants protéiques végétaux (Franceschetti *et al.,* 2017). Afin de poursuivre la caractérisation fonctionnelle et structurale des effecteurs candidats de *M. larici-populina* mis en évidence par le crible mentionné précédemment, nous avons mis en place une approche de production de protéines recombinantes chez *Escherichia coli* pour ensuite tenter de résoudre la structure tridimensionnelle des effecteurs candidats produits. Ce crible est présenté sous forme d'une ébauche d'article en préparation (**Article n°4**). Trois effecteurs candidats ont pu être produits, dont deux en quantité suffisante pour pouvoir résoudre leur structure par résonnance magnétique nucléaire (RMN).

Au début de ma thèse, sur la base des cribles effectoromiques effectués chez *N. benthamiana*, je me suis intéressée plus particulièrement à deux effecteurs candidats: MLP124017 et MLP107772. Afin de réaliser la caractérisation fonctionnelle et structurale de ces deux effecteurs candidats j'ai utilisé des approches de biologie cellulaire et moléculaire, et de biochimie des protéines (Figure 21). La démarche expérimentale qui a été mise en place consistait en la combinaison (i) d'expression transitoire des effecteurs candidats fusionnés à un marqueur fluorescent, en système hétérologue *N. benthamiana* pour comprendre la localisation et rechercher des interactants protéiques, (ii) d'expression stable dans le peuplier 717-1B4 et chez *A. thaliana* et (iii) de production de protéines recombinantes afin de résoudre leurs structures tridimensionnelles par cristallographie ou RMN (Figure 21).

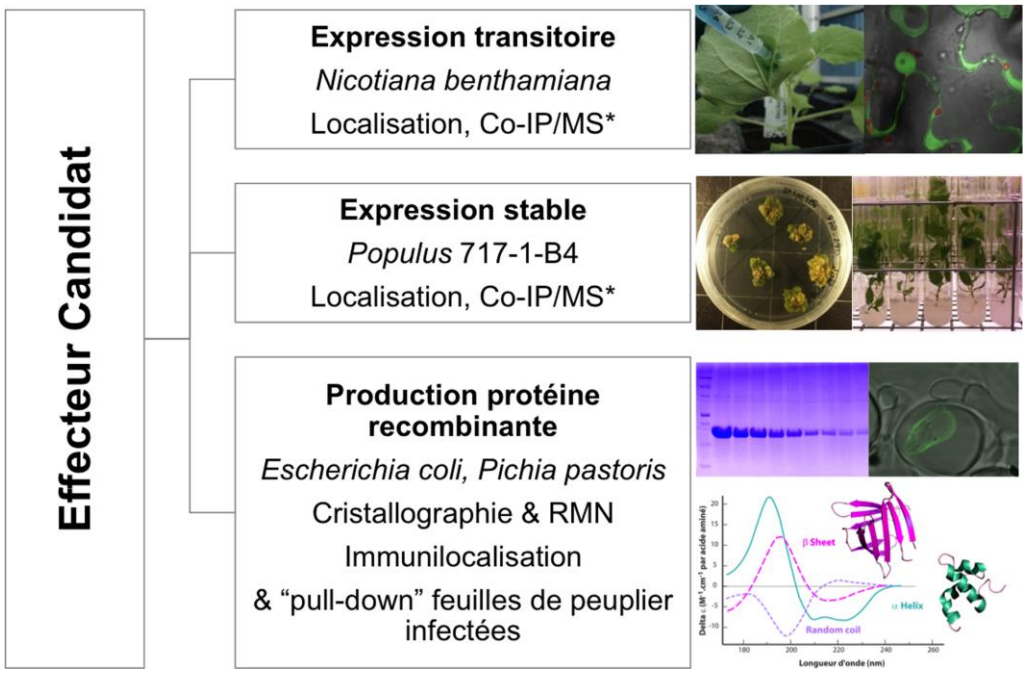


Figure 21 : Démarche expérimentale ayant conduit à la caractérisation fonctionnelle et structurale de deux effecteurs candidats de *M. larici-populina* MLPCTP1 et MLP124017.
Le plan expérimental a été établis comme suit : (i) expression des effecteurs candidats de façon transitoire en système hétérologue *Nicotioana benthamiana* (localisation subcellulaire, recherche d'interactants par co-immunoprécipitation couplée à de la spectrométrie de masse (Co-IP/MS), (ii) générer des transformants stables de peupliers *Populus tremula* x *Populus alba* 717-1B4 et d'*Arabidopsis thaliana* exprimant les effecteurs candidats (localisation subcellulaire, co-immunoprécipitation, recherche de phénotype particulier, tests d'infections Melampsoraceae), et (iii) produire en protéine recombinante et purifier les effecteurs candidats en système bactérien (*Escherichia coli*) ou levure (*Pichia pastoris* ; cristallographie, RMN, « *pull-down* », immunolocalisation).

Le premier effecteur candidat sélectionné nommé MLP124017 interagit avec les protéines de *P. trichocarpa* PopTOPLESS et PopTPR4 (**Annexe 1; article n°7**). MLP124017 présente également des structures secondaires prédites *in sillico* à partir de sa séquence primaire et la protéine recombinante a été produite en système bactérien. Les résultats de production de MLP124017 et de résolution de sa structure seront présentés la sous forme d'un article en préparation, rédigé en anglais et intitulé « **Structural characterization of two *Melampsora larici-populina* candidate effectors** »

Le second effecteur candidat auquel je me suis intéressée, MLP107772 montre une localisation chloroplastique lorsqu'il est exprimé transitoirement dans les feuilles de *N. benthamiana* (**Annexe 1; article n°7**). Une séquence d'adressage aux chloroplastes (peptide de transit) est prédite *in sillico* dans la partie N-terminale de la séquence protéique. De ce fait, MLP107772 a été renommée MLPCTP1 pour « chloroplast-targeted protein 1 ». En étroite collaboration avec Benjamin Petre, nous avons montré que le peptide de transit de MLPCTP1 est nécessaire et suffisant à l'accumulation de la protéine dans le stroma des chloroplastes et qu'il est clivé après translocation. Nous avons également identifié la famille

multigénique à laquelle appartient MLPCTP1. Cette famille de protéines sécrétées polymorphiques contient des membres également transloqués et clivés dans les chloroplastes. Cette famille CTP est spécifique des Melampsoraceae. Ces données seront présentées sous la forme d'un article intitulé « **Rust fungal effectors mimic host transit peptides to translocate into chloroplasts** » publié en 2016 dans la revue *Cellular Microbiology*.

Suite aux deux articles présentés ici, je présente des résultats complémentaires obtenus pour les deux effecteurs candidats sélectionnés, non intégrés aux articles et non publiés à ce jour.

Partie 2 : Article n°4 NMR structural characterization of MLP124266 and MLP124017, two candidate effector proteins produced by the leaf rust pathogen Melampsora larici-populina

Karine de Guillen, Pascale Tsan, Cécile Lorrain, Benjamin Petre, Natalya Saveleja, Nicolas Rouhier, Sébastien Duplessis, André Padilla et Arnaud Hecker

Article de recherche en préparation

NMR structural characterization of MLP124266 and MLP124017, two candidate effector proteins produced by the leaf rust pathogen *Melampsora larici-populina*

AUTHORS:

Karine de Guillen[1], Pascale Tsan[2], Benjamin Pêtre[3,4], Cécile Lorrain[3], Natalya Saveleva[3], Nicolas Rouhier[3], Sébastien Duplessis[3], André Padilla[1] and Arnaud Hecker[3,*]

AFFILIATIONS

[1] CNRS UMR 5048, INSERM U1054, Centre de Biochimie Structurale, Université Montpellier, 34090 Montpellier, France.

[2] CRM2, Equipe BioMod, Faculté des Sciences et Technologies, UMR 7036, Université de LorraineVandoeuvre-lès-Nancy, France; CNRS, CRM2, Equipe BioMod, Faculté des Sciences et Technologies, UMR 7036Vandoeuvre-lès-Nancy, France.

[3] Institut National de la Recherche Agronomique, Centre INRA Nancy Lorraine, UMR 1136 Interactions Arbres/MicroorganismesChampenoux, France; Faculté des Sciences et Technologies, UMR 1136 Interactions Arbres/Microorganismes, Université de Lorraine, Vandoeuvre-lès-Nancy, France.

[4] The Sainsbury Laboratory, Norwich, United Kingdom of Great Britain and Northern Ireland.

* Corresponding author: arnaud.hecker@univ-lorraine.fr

Keywords: Fungal effector, *Melampsora larici-populina*, *Populus trichocarpa*, rust fungus, cysteine-knot, knottin Nuclear Transport Factor 2 like protein, NMR

ABSTRACT

Rust fungi are plant pathogens that secrete an arsenal of effector proteins to interfere with plant functions to promote parasitic infection. Effectors are often species-specific and are thought to evolve rapidly; they often display low sequence similarities with known proteins or functional domains. How rust fungal effectors function in host cells remain elusive and so far, biochemical and structural analysis of effectors have been scarcely used to screen effectors. In this study, we used a strategy based on recombinant protein production in *Escherichia coli* for eleven candidate effectors of the rust fungus *Melampsora larici-populina.* We purified three of them and solved the three-dimensional structure of two effectors using NMR spectroscopy to gain insights about their function. Although no sequence similarity could be determined, both effectors exhibit structural similarities with Knottins on one hand and Nuclear Transport Factor 2 like proteins on the other hand. Altogether these data demonstrate that sequence-unrelated effectors display similar structure to known proteins and support the use of protein production coupled to structural biology as an approach of utmost interest to screen effector candidate and make progress in the analysis of their biochemical function.

INTRODUCTION

To infect their host, filamentous pathogens secrete effectors (mostly proteins) to interfere with plant physiology and immunity to promote parasite growth (Win *et al.,* 2012). Although progress has been made in the past decade, how effectors act in host cells remains a central question in the field of effector biology. Effectors of filamentous pathogens can be secreted in the apoplast or inside the cell cavity through specialized infection structures such as haustoria (Petre and Kamoun 2014). They can be detected by the host plant by two layers of immune receptors at the cell surface or inside the cell, which triggers plant defence response (Jones *et al.,* 2016).

To evade plant immunity, effectors of filamentous plant pathogens are thought to evolve rapidly and they are quite diversified (Sperschneider *et al.,* 2015). A major difficulty in effector-mining is the identification of motifs that could indicate their function, activity or their secretion inside the host cell. Effectors are mainly small proteins whose rapid evolution leads to significant sequence divergence (Persoons *et al.,* 2014; Guyon *et al.,* 2014). Apart from few exceptions such as the well-established RxLR oomycete effectors, conserved motifs, known domains or functions are rarely observed (Franceschetti *et al.,* 2017). However, due to the link between protein structure and function, protein structures tend to be more conserved than primary sequences (Illergard *et al.*, 2009). Thus, structure resolution of candidate effectors could be a powerful approach to determine and understand effector functions. Efforts have been set in this direction and applied in order to determine the structure of effector proteins (Wirthmueller *et al.,* 2013; de Guillen *et al.*, 2015; Ve *et al.,* 2013; Wang *et al.,* 2007; Zhang *et al.,* 2017b).

Rust fungi are the largest group of obligate biotrophs and they infect almost all plant families causing serious damages to cultures (Dean *et al.,* 2012). In the past decade, progress made in rust fungi effector biology were mostly based on genomics and transcriptomics and common but not exclusive features have been used to identify candidate effectors in secretomes (Duplessis *et al.*, 2014). Catalogues of hundreds to thousands of predicted candidate effectors have been unraveled. Generally, criteria such as size, secretion signal, absence of functional information, richness in cysteines, expression during infection, or presence of signatures of rapid evolution have been used as a basis for their selection (e.g. Saunders *et al.,* 2012; Hacquard *et al.,* 2012; Nemri *et al.,* 2014; Cantu *et al.,* 2013). Due to the difficulty to genetically manipulate rust fungi and their host plants, only a handful of rust effectors have been characterized so far (Petre *et al.*, 2014). Apart from their avirulence properties (i.e. recognition by plant immune receptors inside the cell), the functions of these rust effectors remains unknown (Petre *et al.,* 2014). Effectoromic pipelines have been recently established to prioritize rust candidate effectors based on heterologous systems to get insights about their

plant cellular and molecular targets (Lorrain *et al.,* 2018; Petre *et al.,* 2016b; Petre *et al.,* 2015; Qi *et al.,* 2017; Germain *et al.,* 2016). But so far, only one study have set up a small-scale effort using production of candidate effectors in bacterial system to unravel their structure and function (Zhang *et al.*, 2017a). The identification of effector plant interactors associated with structure/function analyses of recombinant effectors can reveal how they interact with plant partners and how co-evolution with the plant promote the diversification of surface-exposed amino acids (Wang *et al.,* 2007; Yaeno *et al.,* 2011; Chou *et al.,* 2011; Leonelli *et al.,* 2011; Boutemy *et al.,* 2011; Win *et al.,* 2012). The avirulence proteins AvrL567, AvrM and AvrP of the flax rust fungus *Melampsora lini* are the three effector structures described in rust fungi so far (Wang *et al.,* 2007; Ve *et al.,* 2013; Zhang *et al.,* 2017b).

The poplar rust fungus, *Melampsora larici-populina* is the causal agent of the poplar leaf rust disease. It causes important damages in poplar plantations across Europe (Pinon and Frey 2005). It is also a tree-microbe interaction model pathosystem, and efforts have helped the identification of candidate effectors in the poplar rust secretome based on transcriptomics and assays developed in heterologous systems such as *Nicotiana benthamiana* and *Arabidopsis thaliana* (Petre *et al.*, 2015; Petre, *et al.,* 2016b; Germain *et al.,* 2017; Gaouar *et al.*, 2016). These studies highlighted that *M. larici-populina* candidate effectors target multiple cell compartments and plant proteins. The same conclusions have been drawn from similar effectoromic screens in other rust fungi (Lorrain *et al.,* 2018). It remains to make progress in the determination of the functions of unravelled priority effectors.

In this study, we used biochemical and structural approaches to explore further selected *M. larici-populina* candidate effectors. To this end, we used *Escherichia coli* as an heterologous system to express, produce and purify eleven candidate effectors that were previously described to target particular cell compartment and/or to interact with specific plant proteins and/or that are homologues of known avirulence effectors (Petre *et al.*, 2015; Petre, *et al.,* 2016b; Germain *et al.,* 2017; Gaouar *et al.*, 2016). Among the eleven selected effectors, only three were successfully expressed and purified from *E. coli* as recombinant proteins. We further scrutinized their structures and searched for potential structure similarity with known proteins. We highlight the structures of two *M. larici-populina* candidate effectors with one showing structural similarities with two bacterial proteins identified as putative Nuclear Transport Factor 2 like proteins and one candidate similar to Knottin proteins.

RESULTS

Selection of candidates secreted effector proteins

Candidates secreted effectors proteins selected for this study were part of a previous effectoromic pipeline performed on twenty *M. larici-populina* candidate effectors (Petre *et al*., 2015). Briefly, effectors were selected in *M. larici-populina* secretome (Hacquard e*t al.,* 2012) if they were expressed *in planta* or in haustoria, specific to rust fungi, with unknown function or with homologies to known rust effectors. Subcellular localization and identification of plant protein targets of these selected candidate effectors were realised by co-immunoprecitation/mass spectrometry in *Nicotiana benthamiana* (Petre *et al.,* 2015). Among the candidate effectors highlighted in this previous study, we took interest on candidate effectors showing a particular subcellular localization such as nucleus, nucleolus and nuclear bodies (MLP109567; MLP124448; MLP124530 respectively); chloroplasts and mitochondria or chloroplasts and aggregates (MLP 107772 and MLP124111 respectively); endomembranes and plamodesmata (MLP124202 and MLP37347 respectively; Table 1); effectors interacting with specific plant interactors (i.e. MLP124017, MLP37347, MLP124448, MLP124111; Table 1); effectors sharing similarities with *M. lini* Avr effectors (i.e. MLP124530, MLP37347, MLP124202, MLP124266; Table 1) and effectors belonging to small-secreted protein families (i.e. MLP124499, MLP124561; Table 1). In total, we selected eleven effectors for structural analyses (Figure 1).

Table 1: *M. larici-populina* candidate effectors screened for recombinant protein production in this study

Protein	MW (kDa)	Protein length (without the signal peptide)[a]	Cysteine residues[b]	Localization *in planta*[c]	Plant interactors[c]	Avr homologues[c]
MLP124478-(HIS)$_6$	8,1	70	6	Nucleolus	Ribosomal proteins	-
MLP124530-(HIS)6	12	96	10	Nuclear and cytosolic bodies	-	Avr123
MLP124111-(HIS)6	13,2	113	10	Chloroplasts and aggregates	Coproporphyrinogen-III oxdase	-
MLP124561-(HIS)6	13,7	110	3	Nucleus and cytosol	-	-
MLP37347-(HIS)6	15,9	128	2	Periphery of haustoria[d], and Plasmodesmata	Glutamate decarboxylase	AvrL567
MLP109567-(HIS)6	16,7	137	2	Nucleus	-	-
MLP124017-(HIS)6	18,5	150	1	Nucleus and cytosol	Topless/Topless-related proteins (TRPs)	-
MLP107772-(HIS)6	16,2	149	8	Chloroplasts and mitocondria	-	-
MLP124202-(HIS)6	45,4	398	1	Endomembranes	-	AvrM
MLP124266-(HIS)6	6,1	69	8	Nucleus and cytosol	-	AvrP4
MLP124499-(HIS)6	6	54	4	Nucleus and cytosol	-	-

[a] number of amino acids of the mature form of the protein are indicated (i.e. without the signal peptide)
[b] number of cysteine residues are indicated
[c] as describe in Petre *et al*. 2015
[d] immunolocalisation performed on infected poplar leaves by Hacquard et al. 2012

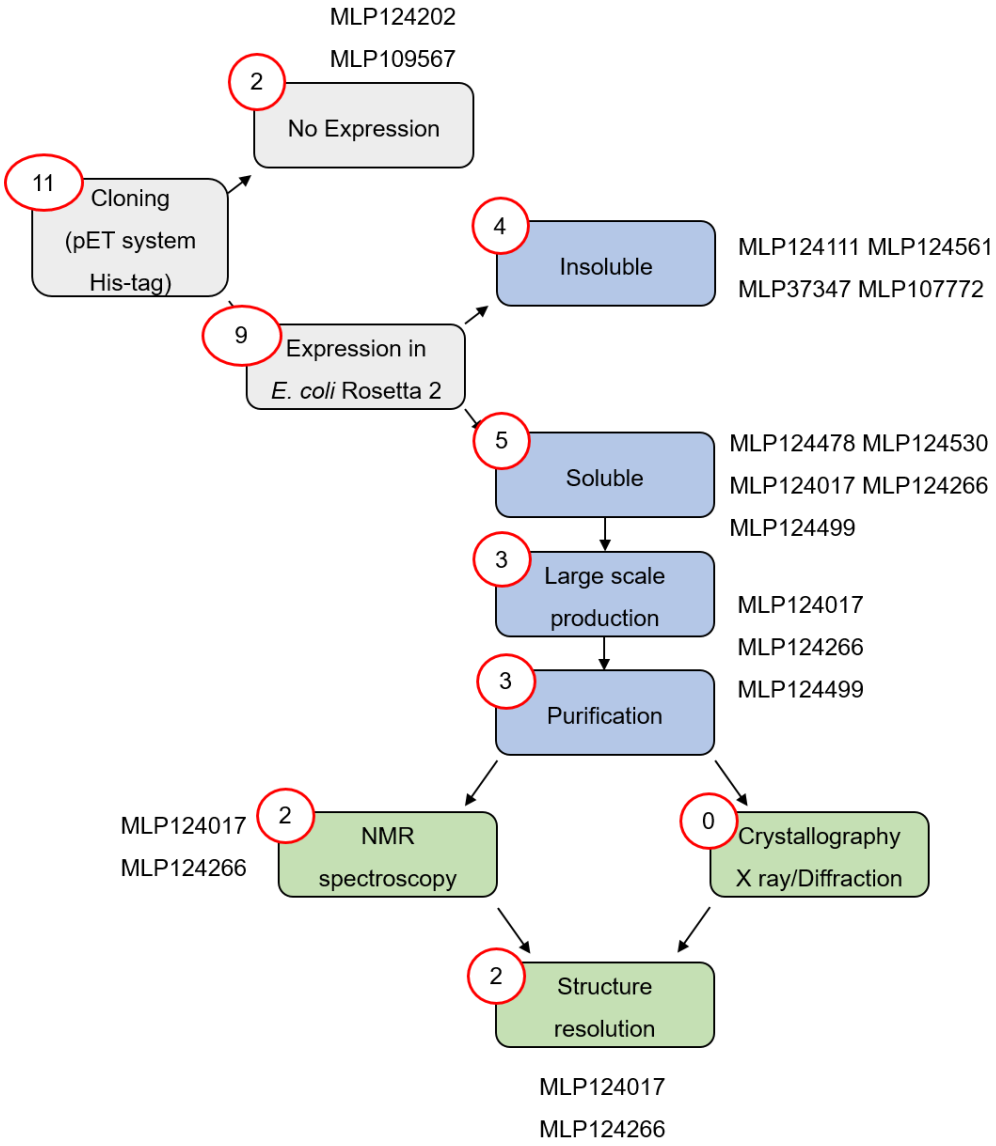


Figure 1: Overview of the effectoromic pipeline.
A total of 11 *M. larici-populina* candidate effectors selected from the previous study of Petre *et al.* (2015) (i.e. particular localization and/or specific plant interactors and/or homologies to *M. lini* Avr effectors). Effector candidates were expressed in *E. coli* SolBL21 (DE3) pRARE2, Rosetta2 (DE3) pLysS or RosettaGami2 (DE3) pLysS strains. Soluble recombinant proteins were produced, purified and their structure solved by NMR spectroscopy.

Three candidate effector recombinant proteins were produced and purified

Recombinant protein production is a *sine qua non* prerequisite to investigate structural properties of the selected effectors. The sequence encoding the eleven effectors was cloned as mature protein (i.e. without signal peptide) into pET26b (for Mlp124111; Mlp124478; Mlp124530; Mlp124561; Mlp37347; Mlp109567; Mlp107772 and Mlp124202) or pET28a (for Mlp124266; Mlp124499 and Mlp124017) expression vectors incorporating a C-terminal 6-histidine tag that helps in protein purification by immobilized metal ion affinity chromatography

(IMAC) (Table S1). His-tagged recombinant proteins corresponding to the mature forms (i.e. without signal peptide) of the selected effectors were over-expressed in *E. coli*. Small-scale expression assays performed on the 11 effectors initially selected indicated that only nine were able to be expressed in *E. coli* Rosetta2 (DE3) pLysS using a standard induction protocol (i.e. addition of 100 µM IPTG in mid-exponential growth phase and further growing for 3 to 4 hours at 37°C): four (MLP124111, MLP124561, MLP37347, MLP107772) were expressed as inclusion bodies (insoluble protein fraction) and five (MLP124478, MLP124530, MLP124017, MLP124266 and MLPH124499 albeit the first two were poorly expressed) were expressed as soluble proteins (Figure 2). Despite the use of other *E. coli* expression strains, such as SoluBL21 (DE3), Origami2 (DE3) pLysS as well as Rosetta-Gami2 (DE3) pLysS known to increase the solubility of recombinant proteins or to improve the maturation of disulfide bridges and despite the modification of induction conditions (induction time, temperature and osmolarity), we were not able to enhance the expression of MLP109567 and MLP124202 nor the solubility of MLP124111, MLP124561, MLP37347 and MLP107772. As MLP124017, MLP124266 and MLP124499 were highly expressed in the soluble protein fraction of *E. coli* as His-tagged recombinant proteins, they were purified in native conditions using a two-step protocol including affinity chromatography and exclusion chromatography (Figure 3). The purified proteins, yielding respectively 50 mg/L (cell culture), 0.5 mg/L and 0.5 mg/L for MLP124017, MLP124266 and MLPH1.1_124499, eluted in size exclusion chromatography as a single peak corresponding to an estimated apparent molecular mass compatible with a monomeric organization.

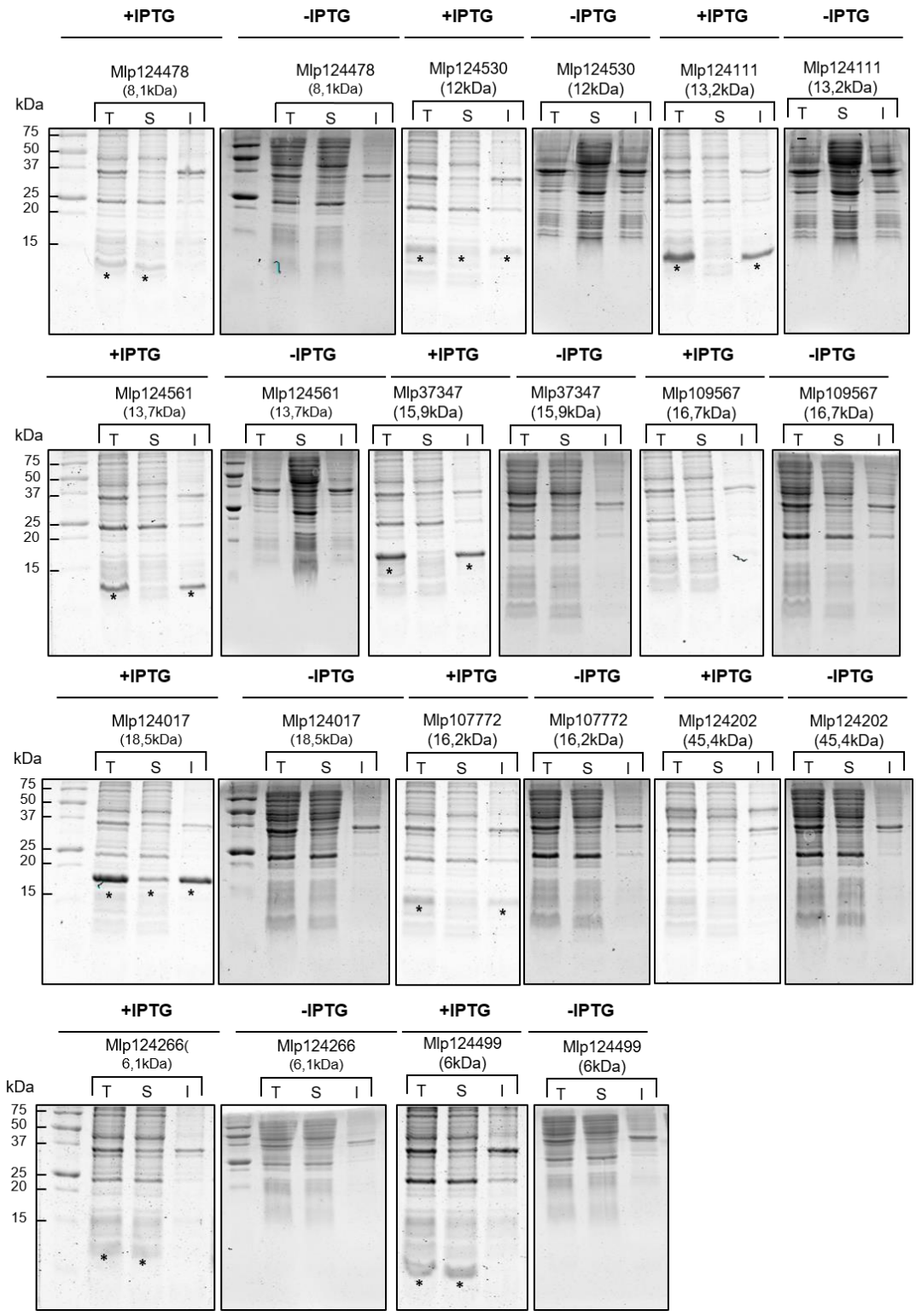


Figure 2: Small-scale expression test of selected effector proteins carried out in *Escherichia coli* Rosetta2 (DE3) pLysS expression strain.

Coomassie blue -stained sodium dodecylsulfate-polyacrylamide gel electrophoresis (SDS-PAGE) analysis of total (T), soluble (S) and (I) insoluble protein fractions of *E. coli* Rosetta2 grown in presence (+) or in absence (-) of 1mM isopropyl β-D21-thiogalactopyranoside (IPTG) to compare protein expression in *E. coli* Rosetta2 (DE3) pLysS expression strain. Asterisk indicate the expected migration of overexpressed proteins.

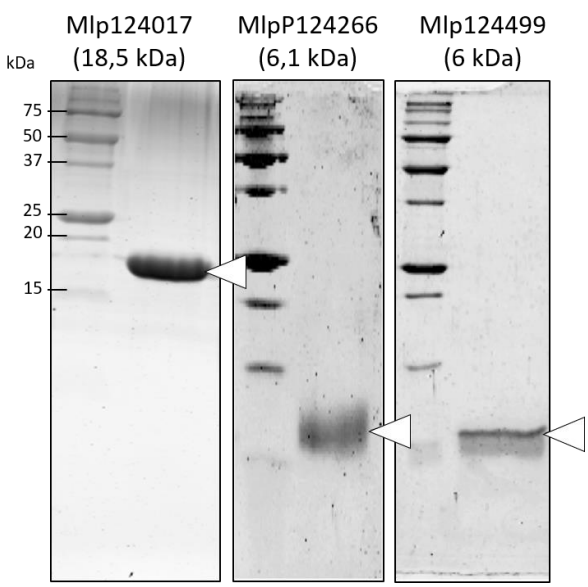


Figure 3: Candidate effectors soluble purified proteins

Ten micrograms of recombinant MLP124017-$(His)_6$; MLP124266-$(His)_6$, MLP124499-$(His)_6$ have been separated by SDS-PAGE (17%). The arrows indicate the band corresponding to the purified proteins.

MLP124266 is a thermostable protein that exhibits a cystine-knot

From a previous study, we reported that genes MLP124266 and MlpH1.1 are strongly expressed during poplar leaf colonization by *M. larici-populina* (Hacquard *et al.,* 2012). MLP124266 and MLP124499 encode a mature protein of 69 and 50 amino acids, respectively (Figure S1A). MLP124266 has an N-terminal part enriched in charged residues and a C-terminal region that possesses six conserved cysteines predicted to form a cystine-knot structure. This typical protein organization is shared by all members of the MlpP4 family as well as by alleles of *M. lini* AvrP4. In MLP124499, several acidic residues are found in the N-terminal part, whereas the C-terminal part contains 3 and 4 cysteines that are partially conserved in the family (Figure S1B). Prediction programs indicate that all members of both protein families exhibit highly conserved N-terminal signal peptides for protein secretion.

Following the production and the purification of MLP124266 and MLP124499, we undertook a structural characterization of each recombinant protein by using a NMR approach. According to their respective molecular mass (less than 30 kDa), mature MLP124266, MLP124499 are well suited for nuclear magnetic resonance (NMR) studies. Standard homonuclear 2D experiments and ^{15}N-edited TOCSY-HSQC and NOESY-HSQC experiments have been carried out on MlpH1.1 and allowed the assignment of ^{1}H and ^{15}N resonances except for the 4 N-terminal residues (Figure S2B). However, very few NOE correlations were observed. Moreover, several minor peaks were observed, especially for Ala14-Glu16, Gly25-Gln26, Glu30, Asp49 residues, suggesting the presence of multiple forms or conformations (Figure S2A). Changing the temperature, the ionic strength or adding dithiothreitol failed to improve the quality of the NMR spectra so MLP124499 structure could not be solved.

In contrast, the assignment of ^{1}H, ^{15}N and $^{13}C_{\alpha}$ resonances has been obtained for all residues except the five N-terminal amino acids (Figure S2A) for MLP124266 recombinant protein and its 3D structure could further be modelled by NMR (Figure 4; Figure S3). The Cys_{36}-Leu_{69} C-terminal region appears to exhibit a typical cysteine-knot structure involving Cys_{39}-Cys_{55}, Cys_{44}-Cys_{58} and Cys_{50}-Cys_{64} disulfide bonds, a β-sheet composed of anti-parallel strands between Thr_{42}-Cys_{44}, Gly_{57}-Ser_{59} and Val_{63}-Val_{65} segments and a short α-helix formed by Gln_{49}-Ala_{52} segment. In contrast, the Met_{1}-Asp_{35} N-terminal region displays large structural disorder, as shown by the superposition of the 20 NMR models (Figure S3). Very few NOE correlations are indeed observed for residues 1 to 35. A few sequential and medium-range NOE correlations characteristic of helical conformations can however be noticed (Figure S3) and explain the presence of short helices in some of these 20 models: 3_{10} helices in the 11-14 region are found in about half of the models while 3_{10} helices in the 7-9 and 21-23 regions are observed in nearly a quarter of the models. Backbone dynamic properties of MLP124266 have been investigated by 15N relaxation measurements. Heteronuclear 1H-15N NOE values showed a contrasted profile with low values for N-terminal residues (indicative of a flexible structure) and high values for C-terminal residues (indicative of a rigid structure). Indeed, amino acids D_6 to G_{38} and C_{39} to L_{69} presented heteronuclear NOE averaged values of 0.26 ± 0.05 and 0.66 ± 0.10, respectively. Interestingly 6 out of the 8 cysteines are gathered in the C-terminal region between Cys39 and Cys69, following a spacing (Cys-X_{2-7}-Cys-X_{3-10}-Cys-X_{0-7}-Cys-X_{1-17}-Cys-X_{4-19}-Cys) typical of cystine-knot structures, (i.e., 3 intricate disulfide bridges that confer very high stability to proteins; Figure 4; Figure S4; Daly *et al.*, 2011). Hence, it is very likely that rigidity originates from the structure formed by these cysteines that are highly conserved in the MLPCPG54 family as indicated by a Consurf analysis (Figure S5) that estimates the evolutionary conservation of amino/nucleic acid positions in a protein/DNA/RNA molecule based on the phylogenetic relations between homologous sequences. Thus, we sought to determine whether these disulfides are formed and whether they influence the stability and/or the oligomerization state of the protein by covalent bonds. A single peak corresponding to the theoretical mass of MLP124266 monomer was obtained by mass spectrometry (data not shown). The titration of free thiol groups in an untreated recombinant MLP124266 gave an averaged value of 1 mole SH per mol protein. Considering the presence of 8 cysteines in the protein, these results are consistent with the existence of 3 intramolecular disulfide bridges (Figure S4). The thermostability of MLP124266 was estimated by heating the protein for 10 min at 95°C. The fact that the protein remained in solution (*i.e.* no precipitation was observed) led us to conclude that it is thermosoluble. In order to investigate the role of the disulfides for such property, we should compare the results obtained with an oxidized and a reduced protein. However, as assessed by thiol titration experiments, we failed to obtain a complete reduction of these disulfides despite extensive incubation of the protein at high

temperatures, in denaturing and reducing environments. Altogether, these results indicate that recombinant MLP124266 is properly folded by *E. coli*, and that the disulfide bridges, which are somehow resistant to reduction, confer a high rigidity and stability to the protein.

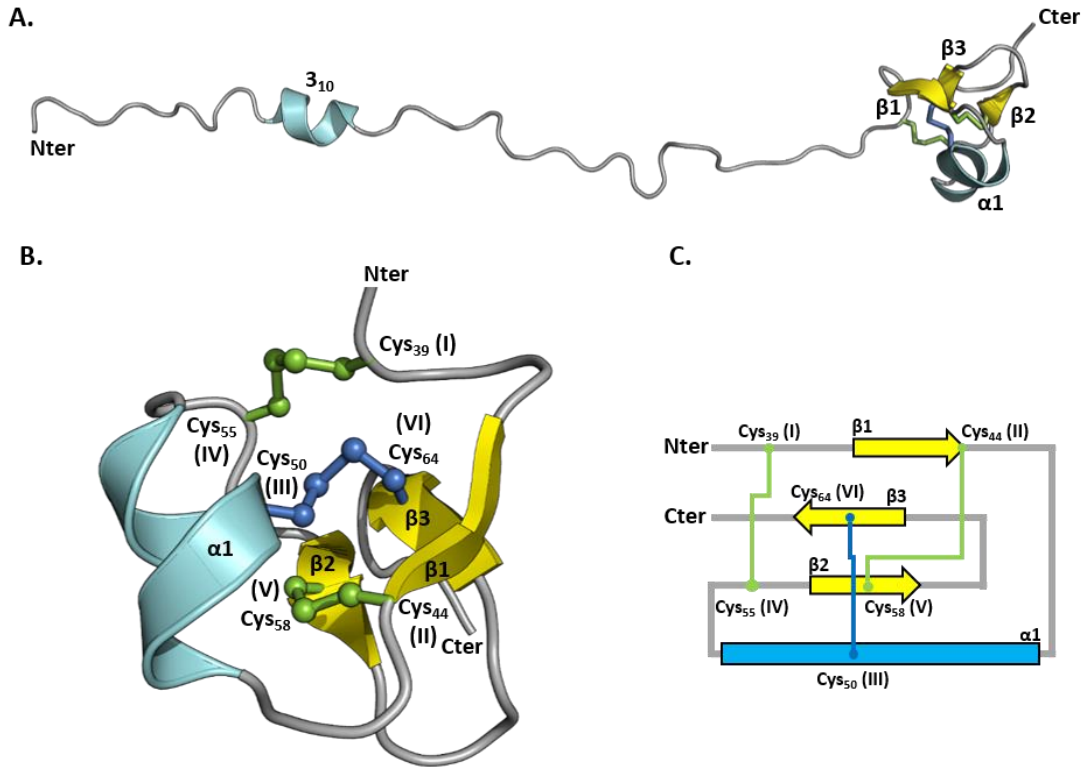


Figure 4: NMR solution structure of MLP124266.

(A). Cartoon representation of MLP124266 structural model deduced from a family of 20 lowest-energy NMR structures. (B). Structural and (C). schematic organization of cystine-knot. The two termini are labelled with Nter and Cter and the strands (β) and helix (3_{10} and α) are colored respectively in yellow and in cyan. Disulfide bridges are shown in a stick or in a ball and stick model colored in green (Cys39-Cys55, Cys44-Cys58) or in blue (Cys50-Cys64).

MLP124017 is part of the Nuclear-Transport Factor 2 like proteins superfamily

MLP124017 is a small-secreted protein (167 amino acids with its signal peptide; 150 amino acids its mature form, with a molecular weight of 18kDa) of unknown function, highly expressed during infection of poplar leaves by *M. larici-populina*. MLP124017 do not share sequence similarity with other *M. larici-populina* nor other rust fungal proteins. In a previous study, we demonstrated the nucleocytoplasmic localization of MLP124107 in *N. benthamiana* and its interaction with poplar TOPLESS-related 4 protein (Petre *et al.,* 2015). To further investigate MLP124017 structure and to get insights about its function, we first attempted to solve its 3D-structure by crystallization coupled to X-ray diffraction. After several assays and nearly 1500 conditions tested, we were unable to obtain exploitable diffracting crystal despite the use of

different version (untagged or N- or C-terminal His-tagged) of MLP124017 protein. We therefore switched to NMR methods. The recombinant 15N and 13C-labelled MLP124017 protein was used for structure determination by two- and three-dimensional NMR experiments. 3D HNCO, HNCA, HN(CO)CACB, HN(CA)CO, HNCACB and TOCSY(D2O) experiments were used for the backbone and aliphatic side chain resonance assignments. 3D 15N-edited NOESY-HSQC and 2D-NOESY(D2O) spectra were collected to confirm the chemical shift assignments and generate distance restraints for structure calculations (Table S3). The assigned 1H, 15N-HSQC spectra were well dispersed, however residues from the N-terminal 1-14 and from 86-95 are missing (Figure S6). The solution structure of MLP124017 was determined based on 1727 NOE-derived distance restraints, 214 dihedral angle restraints and 102 hydrogen bond restraints. All proline residues have been determined to be in a trans-conformation according to the 13Cβ chemical shift at 32.21; 32.46; 31.02 and 32.40 ppm for Pro36; Pro51; Pro54 and Pro146 respectively. The best conformers with the lowest energies, which exhibited no obvious NOE violations and no dihedral violations > 5° were selected for final analysis. The Ramachandran plot produced shows that 99.6% of the residues are in favoured regions (Table S4). MLP124017 is composed of a α+β barrel with seven β-strands forming one mixed β-sheet, four β-hairpins, four β-bulges and four α-helixes. Residues 1-14 and 150-151 are missing and not modelled. Residues 89-95 have no NOE correlation, and were represented as a β-hairpin (Figure 5A). This arrangement of secondary structure produces a cone-shaped fold for the protein, which generates a distinctive hydrophobic cavity (Figure 5B). To identify potential structural homologs of MLP124017, structural similarity searches were performed using the Dali server (Holm and Rosenström 2010). Queries identified SBAL_0622 (PDB code 3BLZ) and SPO1084 (PDB code 3FKA) as the closest structural homologs with the highest Z-score 7.3 and 7.2 respectively and a RMSD of 3.5Å (Figure S7). These two bacterial proteins of *Shewanella baltica* and *Ruegeria pomeroyi* share a common Nuclear Transport Factor 2 (NFT-2) like fold but with unknown functions. The Nuclear Transport factor 2 superfamily represents a large group of proteins sharing common fold and widespread in both prokaryotes and eukaryotes (Eberhardt *et al.,* 2013). Taken together these results showed that although MLP124017 do not shared sequence similarities nor domain with other proteins in sequence databases, its structure is similar to proteins of the NTF-2 folding superfamily.

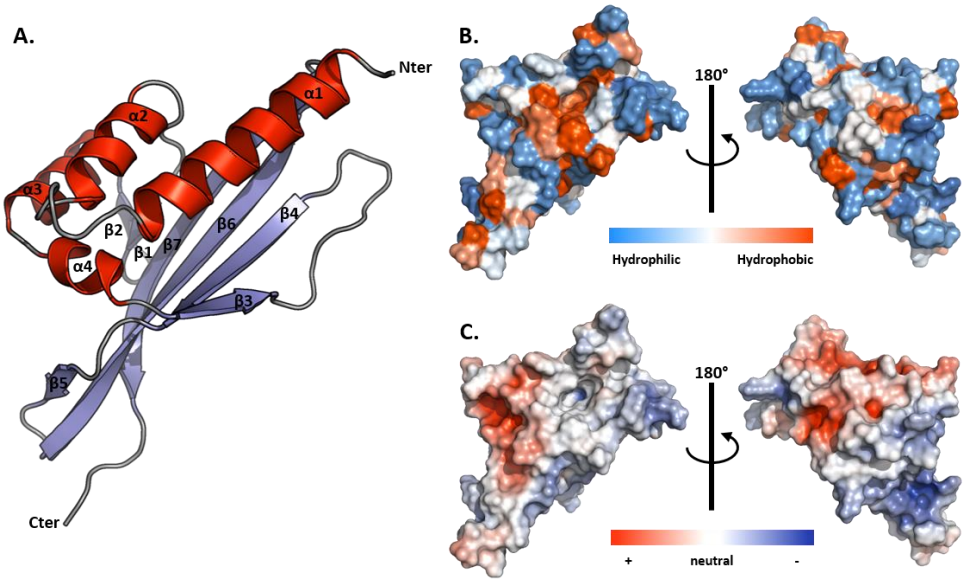


Figure 5: RMN structure of MLP124017
(A). Structure resolution of MLP124017. Cartoon représentation of MLP124017 with four α-helixes in red (α1:16-32; α2:41-48; α3:65-73; α4:80-82) and the mixed β-sheet composed by seven β-strands color in bleue (β1:55-58; β2:61-63; β3:84-86; β4:98-104; β5:108-109; β6:115-129; β7:132-144). The Figure was generated using PyMOL (http://www.pymol.org). (B-C) Surface representation of MLP124017. Front and rear views on the surface of MLP124017 illustrating the conservation of residues generated by Consurf server (B), the surface hydrophobic potential (C), and the surface electrostatic Coulomb potential at pH 7.0 using APBS plugin from Pymol 2.0 software with a contour of -10 kT/e to 10 kT/e.

DISCUSSION

In this study, we have set up a small-throughput effectoromic pipeline based on recombinant protein production and structural characterization to get insights on eleven candidate effectors from the poplar rust fungus *M. larici-populina*. Out of the 11 selected effectors, three were successfully purified as recombinant proteins. Among the latter, MLP124017 and MLP124266 were structurally characterized by NMR. MLP124266 is a homolog of the *M. lini* AvrP4 protein (Catanzariti *et al.,* 2006) and it exhibits a cysteine-knot structure that is found in small disulfide-rich proteins with a cystine-knot. MLP124017 is a singleton protein in *M. larici-populina* with no ortholog in Pucciniales. MLP124017 demonstrated interaction with the poplar TOPLESS-related protein 4 (TRP4) and the resolution of its structure revealed structural homologies with two bacterial proteins that are part of the Nuclear-Transport factor 2 like superfamily proteins.

The functional characterization of rust fungal effectors remains challenging due to the difficulty to experimentally manipulate these obligate biotrophs. In addition, rust fungi possess large catalogues of candidate effector that needs to be prioritized before characterization (Lorrain *et al.,* 2018b; **Article 6**). Effectoromic screens were recently set in heterologous systems for the rust fungi *M. larici-populina* (Petre *et al.,* 2015; Germain *et al.,* 2017), *Puccinia striiformis* f. sp.

tritici (Petre *et al.,* 2016b) and *Phakopsora pachyrizi* (Qi *et al.,* 2017) to pinpoint the most promising candidates for further functional studies. Recombinant protein production in *E. coli* is a valuable approach to perform biophysical and biochemical analysis of candidate effectors (Zhang *et al.,* 2017a). The main advantage to use *E. coli* for large-scale effectors production is that it is a fast, cheap and very well established model with a variety of strains and expression vectors optimized for protein production. However, it can also show some limits. Indeed, we have faced issues for the production of soluble small-cysteine rich proteins in this prokaryote system and among the eleven candidate effectors screened for expression only three were found in the soluble protein fraction. Although we tested different *E. coli* strains and induction protocols, the eight other candidate effectors were either not expressed at all or expressed as inclusion bodies. It is possible to purify recombinant proteins from inclusion bodies by using denaturating extraction conditions (urea quanidium chloride) and further refolding proteins (Palmer and Wingfield, 2012). However, this approach is not recommended for structural analysis as the refolding of the proteins may lead to quite different structure from the original one. Another limit of prokaryote system to produce eukaryote proteins is the lack of post-translational modification such as methylations. An alternative to this is to use the yeast *Pichia pastoris*. It has proven useful for several fungal effectors like *Leptospheria maculans* AvrLm4-7 or *Cladosporium fulvum* Avr2 and Avr4 (Blondeau *et al.,* 2015; Rooney *et al.,* 2005). *P. pastoris* combines many of the advantages of higher eukaryotic expression systems (protein processing, protein folding, and posttranslational modification) and the ease of use such as *E. coli*. It is also faster, easier, and less expensive to use than other eukaryotic expression systems (insect or mammalian cells). *P. pastoris* has also the advantage of excreting the recombinant protein in the culture medium making the purification faster and easier, with only one step (i.e. purification in *E. coli* usually take at least two purification steps) and yields to grams of proteins per liter (Ahmad *et al.,* 2014). We tried this system to test the production of one candidate effector (MLP107772) without noticeable expression in *P. pastoris* despite many tentatives. Nevertheless, this system may be useful and deserves to be considered as an alternative to assay other rust effectors for which we were not able to obtain production in *E. coli*.

NMR structure modelling revealed that MLP124266 possesses two regions with a contrasted profile. The C-terminal part is rigidified through the existence of a cysteine-knot motif whereas the N-terminal part displays a highly flexible structure. The three disulfide bridges at the C-terminal part of MLP124266 are likely responsible of its rigidity, which is confirmed by our NMR dynamic study and which probably contributes to its high stability (Daly *et al.,* 2011). As expected, a β-sheet structure typical of knottins was found but interestingly an additional helix has also been observed between β2 and β3 strands. As far as we know, the presence of this

helix has not been reported in other knottins and this is the first fungal knottin-like protein structure to be solved so far (Postic *et al.,* 2018). It would be interesting to collect structural data from other fungal knottins to investigate whether this helix is specific to rust fungi. Several properties of MLP124266 N-terminal region also deserve to be pointed out. This region, approximately extending up to residue 35 and thus representing half of its primary sequence, globally presents a higher flexibility, as demonstrated by the NMR dynamic results. Nevertheless, it is not totally unstructured, as the heteronuclear NOE values remain positive and as a few residues exhibit helical NMR features. This explains the presence of short helices in this region, although these secondary structures are not systematically found in all NMR models. Interestingly, several secondary structure prediction softwares predict a helix between residues 11 and 18 from its primary sequence. Besides many conserved amino acids have been found in this N-terminal region. These results altogether suggest that this N-terminal region may have some propensity to be structured and especially to form helical structures, which supports a biological role that remains to be elucidated. Interestingly, some effectors structurally characterized have a similar property, (i.e., an N-terminal region with a predicted disordered structure; Boutemy *et al.,* 2011). For instance, *M. lini* effectors AvrL567 and AvrM have predicted disordered N-terminal regions that are susceptible to protease degradation (Wang *et al.,* 2007; Catanzariti *et al.,* 2006). Flexible folds are known to be adaptable linkers that favour the ability to bind partners (Chouard, 2011). The N-terminus of many cytoplasmic effectors being the important part of the protein that mediate cell entry, it is tempting to speculate that this flexible part maybe necessary to bind a target important for cell entry. The knottins folded proteins display a variety of functions such as venoms and spider toxins (Lee and MacKinnon 2004; Garcia, 2004) but also antimicrobial properties such as the cyclotides (Tam *et al.,* 1999). Some are also found to interact with a receptor like protease inhibitors found in plants, insects and plant-parasites (Kim *et al.,* 2009).

The second structure that has been solved by NMR in this study is the one of MLP124017. This structure showed a fold similar of the NTF-2 superfamily. The NTF2-like superfamily is a versatile group of protein domains sharing a common fold but having nothing in common at the sequence level. These domains perform a range of different functions within the proteins in which they are found. MLP124017 is structurally close to two bacterial proteins despite the lack of sequence similarity between the three proteins. The structures of these proteins consist of a β-sheet surrounding a binding pocket and α-helixes acting as a lid (Marcos *et al.*, 2017). The NTF-2 family regroups catalytic and non-catalytic proteins that contains cone-like structured protein with a cavity that often act as a molecular container involved in diversified wide range of cellular functions (Eberhardt *et al.,* 2013). Interestingly, the cone-shaped structure of MLP124017 is widespread across both Prokaryotes and Eukaryotes. The first

proteins of the NTF-2 family were reported to play a role in the transport of molecules from the cytoplasm to the nucleus. It has been shown in Arabidopsis that a NTF-2 protein is required to import nuclear proteins by recognition of a nuclear localization signal (NLS). The Arabidopsis NTF-2 protein was also found to play a role in the nuclear importation of the small-GTPase Ran-GDP that is a central protein in signal transduction in response to various physiological processes (e.g. mitotic spindle formation, nuclear envelope assembly or responses to biotic stresses; Avis and Clarke 1996; Carazo-Salas *et al.,* 2001; Zhang and Clarke 2000; Hetzer *et al.,* 2000). In bacteria, some NTF-2 like proteins were found to play a role in bacterial conjugation as part of the type IV secretion system involved in the transfer of DNA from cell to cell (Goessweiner-Mohr *et al.,* 2013) whereas non-catalytic NTF2-like domains were shown to act as immunity proteins in bacterial toxin systems (Zhang *et al.,* 2012). As for fungi, it has been shown that a *Saccharomyces cerevisae* NTF2 knock-out mutant was defective to nuclear import (Corbett and Silver, 1996). Although it is known that the NFT-2 folded protein are widespread across kingdoms, very few is known about their role. A recent study presented that the silencing of NTF-2 in wheat decreased the resistance against avirulent isolates of the wheat stripe rust fungus *P. striiformis* f. sp. *tritici* (Zhang *et al.,* 2018). Since MLP124017 has been shown to interact with TOPLESS and TOPLESS-related proteins (Petre *et al.,* 2015), we may speculate that the cavity formed by the β-sheet could interact to these plant partners.

To conclude, this study demonstrates the usefulness of including biochemical and structural approaches in systematic screens for functional analysis of candidate effectors. Although it is too early to speculate about the function of these two candidate effectors, structural biology has proved to be a powerful tool to investigate effector functions through structural homology as exemplified by the recently unravelled fungal MAX effector family (de Guillen *et al.,* 2015). The structures for two poplar rust fungus effectors not only add new information about fungal effectors or rust effectors they also provide a basis for future comparative structural studies and bring complementary information on fungal knottins in a dedicated database.

EXPRERIMENTAL PROCEDURES

Sequence analyses and names

Alignment and phylogenetic analyses were performed on the phylogeny website (www.phylogeny.fr) with default parameters. Alignments were corrected and edited manually, and phylogenetic trees generated with FigTree v1.4.3. Physical and chemical parameters of proteins were estimated using protparam tool (http://web.expasy.org/protparam/). Common names and JGI ID of genes described in this study are as follow: MLP124266, Mlp124499, Mlp124111, Mlp124478, Mlp124530, Mlp124561, Mlp37347, Mlp109567, Mlp124017, Mlp107772 and Mlp124202.

Cloning of selected effector enconding sequences

Open reading frames (ORF) coding the mature forms (i.e. sequence devoid of the sequence encoding N-terminal secretion peptide) of Mlp124266 and Mlp124499 (gene ID in *M. larici-populina* strain 98AG31 reference genome) were ordered as synthetic genes cloned in pBSK(+) vectors (Genecust). Coding sequence of the mature form (i.e. without the signal peptide) of the nine candidate effectors Mlp124111, Mlp124478, Mlp124530, Mlp124561, Mlp37347, Mlp109567, Mlp124017, Mlp107772 and Mlp124202 were amplified by polymerase chain reaction (PCR) using cDNAs from Beaupré leaves infected by *M. larici-populina* (isolate 98AG31) and further cloned into pICSL01005 vector as described in Petre *et al.* (2015). The sequences encoding the mature form of each effector were subsequently cloned by PCR in either pET26b vector or pET-26b between *Nde*I and *Not*I or *Nco*I and *Not*I restriction sites respectively using primers detailed in Table S1.

Expression and purification of recombinant proteins in *Escherichia coli*

Expression of recombinants effectors was performed at 37°C using the transformed *E. coli* SolBL21 (DE3) pRARE2 (Amsbio Abington, UK), Rosetta2 (DE3) pLysS or RosettaGami2 (DE3) pLysS strains (Novagen) with adequate pET expression vector coding for selected effector (Table S1) in LB medium supplemented with kanamycin (50 µg/ml) and chloramphenicol (34 µg/ml). When the cell culture reached a OD_{600nm} of 0.7, protein expression was induced with 0.1 mM IPTG and cells were grown for a further 4 h.To improve the solubility of some recombinant effectors, other protocols were used as follows. First, we added 0.5% (v/v) of ethanol in the medium when culture reached an OD_{600nm} of 0.7. The cells were cooled to 4°C for 3 h, recombinant protein expression induced with 0.1 mM isopropyl-β-D-1-thiogalactopyranoside (IPTG) and cells further grown for 18 h at 20°C. We also tested tested a combination of an osmotic and thermal shocks (Oganesyan *et al.,* 2007). When the culture reached an OD_{600nm} of 0.5-0.6, 29.2 g/l of NaCl and 2 mM of betaine were added to the culture medium and the culture incubated at 47°C for 1 hour under stirring. Cells were then cooled to 20°C and the expression of recombinant proteins induced with 0.1 mM IPTG. After induction, cells were harvested by centrifugation, resuspended in a 30 mM Tris-HCl pH 8.0 and 200 mM NaCl lysis buffer and stored at −20°C. Cell lysis was completed by sonication (three times for 1 min with intervals of 1 min). The cell extract was then centrifuged at 35000 g for 25 min. at 4 °C to remove cellular debris and aggregated proteins. C-terminal His-tagged recombinant proteins were then purified by by gravity-flow chromatography on a nickel nitrilotriacetate (Ni-NTA) agarose resin (Qiagen) according to the manufacturer's recommendations followed by an exclusion chromatography on a Superdex75 column connected to an ÄKTA Purifier™ (GE Healthcare). The fractions containing recombinant MLP124017 were pooled, concentrated and stored at -20°C as such whereas MLP124266 and MLPH1.1, fractions were pooled, dialyzed

against 30 mM Tris-HCl, 1 mM EDTA (TE) pH 8.0 buffer and stored at 4°C until use. For NMR analyses, MLP124266-(His6) and MLPH1.1-(His6) recombinant proteins, were ^{15}N-labeled in M9 minimal synthetic medium containing ^{15}NH4Cl (1 g/L)). MLP124017 was single ^{15}N or double ^{15}N and ^{13}C labelled in M9 minimal medium containing 1 g/l NH4Cl (^{15}N) and 2 g/l glucose (^{13}C) supplemented with 2,5% thiamine (m/v), 1mg/ml biotine, 50mM $FeCl_3$, 10mM $MnCl_2$, 10mM $ZnSO_4$, 2mM $CoCl_2$, 2mM $NiCl_2$, 2mM $NaSeO_3$ and 2mM H_3BO_3. After purification as described above, labelled MLP124266/MlpH1.1 and Mlp124017 were dialyzed respectively into a 50 mM phosphate pH 6.0 buffer and 20 mM phosphate pH 6,8, 200mM NaCl buffer and concentrated for immediate analysis.

NMR samples

Uniformly 15N or 13C/15N labelled proteins (1 mM 50 mM phosphate pH 6.0 buffer dor MLP124266 and MlpH1.1 or 20 mM phosphate pH 6.8, 200mM NaCl for Mlp124017) was supplemented with 10% D_2O, 5 mM 4,4-dimethyl-4-silapentane-1-sulfonic acid (DSS) as a reference and 0.02% sodium azide.. For the D_2O experiments, samples were lyophilized and dissolved in 200 µL D_2O. For the ^{13}C assignments experiments, the sample was prepared with 200 µL of 600 µM MLP124017 uniformly labeled with ^{13}C and ^{15}N in the same phosphate buffer previously used supplied with 20 µL of D_2O and 5 mM DSS as a reference.

Nuclear magnetic resonance spectroscopy

For MLP124017, spectra were acquired on 800 and 700 MHz Avance Bruker spectrometers equipped with triple-resonance (^{1}H, ^{15}N, ^{13}C) z-gradient cryo-probe at 298 K. Experiments were recorded using the TOPSPIN pulse sequence library (v. 2.1) (Table S2). All spectra are referenced to the internal reference DSS (4,4-dimethyl-4-silapentane-1-sulfonic acid) for the ^{1}H dimension and indirectly referenced for the ^{15}N and ^{13}C dimensions [Whishart et al., 1995]. Sequential assignment was performed using 3D ^{15}N-NOESY-HSQC, ^{15}N-TOCSY-HSQC, HNCO, HNCACO, HNCA, HNCOCACB and HNCACB. Side chain ^{1}H assignments were carried out using combined analysis with 3D ^{15}N-NOESY-HSQC, ^{15}N-TOCSY-HSQC and 2D NOESY and TOCSY with D_2O samples. A series of three HSQC spectra were performed after lyophilization and dilution of the first sample in D_2O to determine amides protons in slow exchange (Table S2). For MLP124266 and MLPH1.1, NMR spectra were acquired on a Bruker DRX 600 MHz spectrometer equipped with a TCI cryoprobe. For MLP124266 and MLPH1.1, COSY, TOCSY (mixing time of 60 ms) and NOESY (mixing time of 150 ms) experiments were run at 298 K on 0.85 and 0.65 mM unlabeled samples of MLP124266 and MlpH1.1, respectively. HSQC, TOCSY-HSQC (mixing time of 60 ms) and NOESY-HSQC (mixing time of 150 ms) experiments were recorded with 1.8 and 0.135 mM uniformly 15N-labeled samples of MLP124266 and MlpH1.1, respectively. For MLP124266, HNHA, HNHB, R1 and R2 15N

relaxation rates, 1H-15N heteronuclear NOE, HNCA (with 24 (15N) x 28 (13C) complex points and 192 transients per increment) standard experiments were recorded. Spectra were processed using Topspin3.0 (Bruker) and analyzed with NmrViewJ (Johnson and Blevins, 1994), CcpNmr (Vranken *et al.,* 2005) and ARIA2 (Rieping *et al.,* 2007).

Structure calculation

NOE peaks identified in 3D ^{15}N-NOESY-HSQC and 2D NOESY experiments were automatically assigned during structure calculations performed by the program CYANA 2.1 (Güntert, 2004). The ^{15}N, H_N, ^{13}C', ^{13}Cα, Hα and ^{13}Cβ chemical shifts were converted into β/β dihedral angle constraints using TALOS+ (v. 1.2) (Shen *et al.,* 2009). Hydrogen bonds constraints were determined "par" by exchange experiments ^{1}H / ^{2}H of "skeleton" amides protons (H_N). After identification each hydrogen bonds is forced using following constraints: 1.8–2.0 Å for H_N,O distance and 2.7–3.0 Å for N_H,O distance. Final structure calculations were performed with CYANA (v. 2.1) using all distance and angle restraints (Table S3). 600 structures were calculated with CYANA 2.1, of which the 20 conformers with the lowest target function were refined by CNS (v. 1.2) using the refinement in water of RECOORD (Nederveen *et al.,* 2005) and validated using PROCHECK (Laskowski *et al.,* 1993).

ACKNOWLEDGMENTS

GK and AP are supported by a FRISBI project. CL and BP are supported by INRA, in the framework of a Contrat Jeune Scientifique and by the Labex ARBRE. SD and AH acknowledge the support of the French ANR for the young scientist grant POPRUST (ANR-2010-JCJC-1709-01) and of the Région Lorraine. All authors benefit from the support of the ANR in the frame of a grant part of the "Investissements d'Avenir" program (ANR-11-LABX-0002-01,LabofExcellenceARBRE).

LITTERATURE CITED

Ahmad, M., Hirz, M., Pichler, H., & Schwab, H. (2014). Protein expression in Pichia pastoris: recent achievements and perspectives for heterologous protein production. *Applied Microbiology and Biotechnology*, 98(12), 5301‑5317.

Avis, J. M., & Clarke, P. R. (1996). Ran, a GTPase involved in nuclear processes: its regulators and effectors. *Journal of Cell Science*, 109(10), 2423‑2427.

Blondeau, K., Blaise, F., Graille, M., Kale, S. D., Linglin, J., Ollivier, B., Labarde, A., Lazar, N., Daverdin, G., Balesdent, M.-H., Choi, D. H. Y., Tyler, B. M., Rouxel, T., van Tilbeurgh, H., & Fudal, I. (2015). Crystal structure of the effector AvrLm4-7 of Leptosphaeria maculans reveals insights into its translocation into plant cells and recognition by resistance proteins. The Plant Journal : For Cell and Molecular Biology, 83(4), 610–624.

Boutemy, L. S., King, S. R. F., Win, J., Hughes, R. K., Clarke, T. A., Blumenschein, T. M. A., Kamoun, S., & Banfield, M. J. (2011). Structures of Phytophthora RXLR Effector Proteins:

a conserved but adaptable fold underpins functional diversity. *The Journal of Biological Chemistry*, 286(41), 35834‑35842.
Cantu, D., Segovia, V., MacLean, D., Bayles, R., Chen, X., Kamoun, S., Dubcovsky, J., Saunders, D.G.O., & Uauy, C. (2013). Genome analyses of the wheat yellow (stripe) rust pathogen *Puccinia striiformis* f. sp. *tritici* reveal polymorphic and haustorial expressed secreted proteins as candidate effectors. *BMC Genomics*, 14: 270.
Carazo-Salas, R. E., Gruss, O. J., Mattaj, I. W., & Karsenti, E. (2001). Ran-GTP coordinates regulation of microtubule nucleation and dynamics during mitotic-spindle assembly. *Nature Cell Biology*, 3(3), 228‑234.
Catanzariti, A.-M., Dodds, P. N., Lawrence, G. J., Ayliffe, M. A., & Ellis, J. G. (2006). Haustorially Expressed Secreted Proteins from Flax Rust Are Highly Enriched for Avirulence Elicitors. *The Plant Cell*, 18(1), 243‑256.
Chou, S., Krasileva, K. V., Holton, J. M., Steinbrenner, A. D., Alber, T., & Staskawicz, B. J. (2011). *Hyaloperonospora arabidopsidis* ATR1 effector is a repeat protein with distributed recognition surfaces. *Proceedings of the National Academy of Sciences of the United States of America*, 108(32).
Chouard, T. (2011). Structural biology: Breaking the protein rules. *Nature*, 471(7337), 151‑153.
Daly, N. L., Gruber, C. W., Göransson, U., & Craik, D. J. (2011). Cystine Knot Folding in Cyclotides. Chang & S. Ventura (Éd.), Folding of Disulfide Proteins (p. 43‑61).
Dean R., van Kan J. A. L., Pretorius Z. A., Hammond-Kosack K. E., di Pietro A., Spanu P. D., Rudd J. J., Dickman M., Kahmann R., Ellis J. & Foster G. D. (2012) The Top 10 fungal pathogens in molecular plant pathology. *Molecular Plant Pathology*, 13: 414–430.
de Guillen, K., Ortiz-Vallejo, D., Gracy, J., Fournier, E., Kroj, T., & Padilla, A. (2015). Structure Analysis Uncovers a Highly Diverse but Structurally Conserved Effector Family in Phytopathogenic Fungi. *PLoS Pathogens*, 11(10), e1005228.
Duplessis, S., Bakkeren, G., & Hamelin, R. (2014). Advancing knowledge on biology of rust fungi through genomics. *Advances in Botanic Research*, Vol. 70, p. 173‑209.
Eberhardt, R. Y., Chang, Y., Bateman, A., Murzin, A. G., Axelrod, H. L., Hwang, W. C., & Aravind, L. (2013). Filling out the structural map of the NTF2-like superfamily. *BMC Bioinformatics*, 14, 327‑327.
Franceschetti, M., Maqbool, A., Jimenez-Dalmaroni, M. J., Pennington, H. G., Kamoun, S., & Banfield, M. J. (2017). Effectors of Filamentous Plant Pathogens: Commonalities amid Diversity. *Microbiology and Molecular Biology Reviews : MMBR*, 81(2).
Gaouar, O., Morency, M.-J., Letanneur, C., Séguin, A., & Germain, H. (2016). The 124202 candidate effector of *Melampsora larici-populina* interacts with membranes in *Nicotiana* and *Arabidopsis*. *Canadian Journal of Plant Pathology,* 38(2), 197‑208.
Garcia, M. L. (2004). Ion channels: gate expectations. *Nature*, 430(6996), 153‑155.
Germain H., Joly D.L., Mireault C., Plourde M.B., Letanneur C., Stewart D., Morency M. J., Petre B., Duplessis S., & Seguin A. (2017) Infection assays in Arabidopsis reveal candidate effectors from the poplar rust fungus that promote susceptibility to bacteria and oomycete pathogens. *Molecular Plant Pathology*, 19: 191-200.
Goessweiner-Mohr, N., Arends, K., Keller, W., & Grohmann, E. (2013). Conjugative type IV secretion systems in Gram-positive bacteria. *Plasmid*, 70(3): 289‑302.
Hacquard S., Joly D.L., Lin Y.-C., Tisserant E., Feau N., Delaruelle C., Legue V., Kohler A., Tanguay P., Petre B., Frey P., Van de Peer Y., Rouzé P., Martin F., Hamelin R. C., & Duplessis S. (2012). A comprehensive analysis of genes encoding small secreted proteins

identifies candidate effectors in *Melampsora larici-populina* (poplar leaf rust). *Mol. Plant-Microbe Interact. MPMI*, 25(3):279-293.
Hetzer, M., Bilbao-Cortes, D., Walther, T. C., Gruss, O. J., & Mattaj, I. W. (2000). GTP hydrolysis by Ran is required for nuclear envelope assembly. *Molecular Cell*, 5(6): 1013‑1024.
Illergard, K., Ardell, D. H., & Elofsson, A. (2009). Structure is three to ten times more conserved than sequence--a study of structural response in protein cores. *Proteins*, 77(3):499‑508.
Jones, J. D. G., & Dangl, J. L. (2006). The plant immune system. *Nature*, 444(7117), 323‑329.
Kim, J.-Y., Park, S.-C., Hwang, I., Cheong, H., Nah, J.-W., Hahm, K.-S., & Park, Y. (2009). Protease Inhibitors from Plants with Antimicrobial Activity. *International Journal of Molecular Sciences*, 10(6)
Johnoson, B. A., & Belvin, B.A. (1994) NMR View: A computer program for the visualization and analysis of NMR data. *J Biomol NMR,* 4(5):603-14.
Lee, S.-Y., & MacKinnon, R. (2004). A membrane-access mechanism of ion channel inhibition by voltage sensor toxins from spider venom. *Nature*, 430(6996), 232‑235.
Leonelli, L., Pelton, J., Schoeffler, A., Dahlbeck, D., Berger, J., Wemmer, D. E., & Staskawicz, B. (2011). Structural Elucidation and Functional Characterization of the *Hyaloperonospora arabidopsidis* Effector Protein ATR13. PLoS Pathogens, 7(12), e1002428.
Lorrain, C., Petre, B., & Duplessis, S. (2018b) Show me the way: rust effector targets in heterologous plant systems. *Current Opinion in Microbiology*, *accepted in press*.
Marcos, E., Basanta, B., Chidyausiku, T. M., Tang, Y., Oberdorfer, G., Liu, G., Swapna, G.V.T., Guan, R., Silva, D.-A., Dou, J., Pereira, J. H., Xiao, R., Sankaran, B., Zwart, P. H., Montelione, G. T., & Baker, D. (2017). Principles for designing proteins with cavities formed by curved β sheets. *Science*, 355(6321), 201.
Nemri, A., Saunders, D. G. O., Anderson, C., Upadhyaya, N. M., Win, J., Lawrence, G. J., Jones, D. A., Kamoun, S., Ellis, J. G., & Dodds, P. N. (2014). The genome sequence and effector complement of the flax rust pathogen *Melampsora lini. Frontiers in Plant Science*, 5, 98.
Oganesyan, N., Ankoudinova, I., Kim, S.-H., & Kim, R. (2007). Effect of osmotic stress and heat shock in recombinant protein overexpression and crystallization. Protein Expression and Purification, 52(2): 280‑285.
Palmer, I., & Wingfield, P. T. (2012). Preparation and Extraction of Insoluble (Inclusion-Body) In Proteins from *Escherichia coli.* Current Protocols in Protein Science, eds John E. Coligan.
Persoons, A., Hayden, K. J., Fabre, B., Frey, P., De Mita, S., Tellier, A., & Halkett, F. (2017). The escalatory Red Queen: Population extinction and replacement following arms race dynamics in poplar rust. *Molecular Ecology*, 26(7), 1902‑1918.
Petre B., Saunders D.G.O., Sklenar J., Lorrain C., Krasileva K.V., Win J., Duplessis S., & Kamoun S. (2016b) Heterologous Expression Screens in *Nicotiana benthamiana* Identify a Candidate Effector of the Wheat Yellow Rust Pathogen that Associates with Processing Bodies. *PloS One*, 11, e0149035.
Petre B., Saunders D.G.O., Sklenar J., Lorrain C., Win J., Duplessis S., & Kamoun S. (2015) Candidate Effector Proteins of the Rust Pathogen *Melampsora larici-populina* Target Diverse Plant Cell Compartments. *Mol. Plant-Microbe Interact*. 28:689–700
Petre B., & Kamoun S. (2014) How do filamentous pathogens deliver effector proteins into plant cells? *PLoS Biololgy*. 12, e1001801.

Postic, G., Garcy, J., Perin, C., Chiche, L., & Gelly, J.-C. (2018) KNOTTIN: the database of inhibitor cystine knot scaffold after 10 years, toward a systematic structure modelling. *Nucleic Acids Res*, 46: D454–D458.

Qi M., Grayczyk J.P., Seitz J.M., Lee Y., Link T.I., Choi D., Pedley K., Voegele R., Baum T., & Whitham S.A. (2017) Suppression or activation of immune responses by predicted secreted proteins of the soybean rust pathogen *Phakopsora pachyrhizi*. *Mol. Plant. Microbe Interact*. 31(1):163-174.

Rieping, W., Habeck, M., Bardiaux, B., Bernard, A., Malliavin, T.E., & Nilges, M. (2005) ARIA2: automated NOE assignment and data integration in NMR structure calculation**.** *Bioinformatics*, 23(3):381-2.

Rooney, H. C. E., Van't Klooster, J. W., van der Hoorn, R. A. L., Joosten, M. H. A. J., Jones, J. D. G. & de Wit, P. J. G. M. (2005) *Cladosporium* Avr2 inhibits tomato rcr3 protease required for cf-2-dependent disease resistance. *Science*, 308 (5729): 1783–86.

Saunders D.G.O., Win J., Cano L.M., Szabo L.J., Kamoun S., & Raffaele S. (2012) Using hierarchical clustering of secreted protein families to classify and rank candidate effectors of rust fungi. *PloS One*, 7, e29847.

Tam, J. P., Lu, Y.-A., Yang, J.-L., & Chiu, K.-W. (1999). An unusual structural motif of antimicrobial peptides containing end-to-end macrocycle and cystine-knot disulfides. Proceedings of the National *Academy of Sciences of the United States of America,* 96(16), 8913‑8918.

Terpe, K. (2006). Overview of bacterial expression systems for heterologous protein production: from molecular and biochemical fundamentals to commercial systems. *Applied Microbiology and Biotechnology*, 72(2), 211‑222.

Vranken, W.F., Boucher, W., Stevens, T.J., Fogh, R.H., Pajon, A., Llinas, M., Ulrich, E.L., Markley, J.L., Ionides, J., & Laue, E.D. (2005) The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins*. 59(4):687-96.

Ve, T., Williams, S. J., Catanzariti, A.-M., Rafiqi, M., Rahman, M., Ellis, J. G., Hardham, A. R., Jones, D. A., Anderson, P. A., Dodds, P. N., & Kobe, B. (2013). Structures of the flax-rust effector AvrM reveal insights into the molecular basis of plant-cell entry and effector-triggered immunity. *Proceedings of the National Academy of Sciences of the United States of America*, 110(43), 17594‑17599.

Wang C.I.A., Guncar G., Forwood J.K., Teh T., Catanzariti A.M., Lawrence G.J., Loughlin F.E., Mackay J.P., Schirra H.J., Anderson P.A., Ellis J.G, Dodds P.N., & Kobe B. (2007) Crystal structures of flax rust avirulence proteins AvrL567-A and -D reveal details of the structural basis for flax disease resistance specificity. *Plant Cell*, 19:2898–2912.

Win J., Chaparro-Garcia A., Belhaj K., Saunders D.G., Yoshida K., Dong S., Schornack S., Zipfel C., Robatzek S., Hogenhout S.A. & Kamoun S. (2012) Effector biology of plant-associated organisms: concepts and perspectives. *Cold Spring Harb Symp Quant Biol*, 77.

Wirthmueller, L., Maqbool, A., & Banfield, M. J. (2013). On the front line: structural insights into plant-pathogen interactions. *Nature Reviews*. *Microbiology*, 11(11):761‑776.

Yaeno, T., Li, H., Chaparro-Garcia, A., Schornack, S., Koshiba, S., Watanabe, S., Kigawa, T., Kamoun, S., & Shirasu, K. (2011). Phosphatidylinositol monophosphate-binding interface in the oomycete RXLR effector AVR3a is required for its stability in host cells to modulate plant immunity. *Proceedings of the National Academy of Sciences of the United States of America*, 108(35), 14682‑14687.

Zhang, Q., Wang, B., Wei, J., Wang, X., Han, Q., & Kang, Z. (2018). TaNTF2, a contributor for wheat resistance to the stripe rust pathogen. *Plant Physiology and Biochemistry*, 123, 260–267.

Zhang, X., Nguyen, N., Breen, S., Outram, M. A., Dodds, P. N., Kobe, B., Solomon, P. S., & Williams, S. J. (2017a). Production of small cysteine-rich effector proteins in *Escherichia coli* for structural and functional studies. *Molecular Plant Pathology*, 18(1), 141‐151.

Zhang X., Farah N., Rolston L., Ericsson D.J., Catanzariti A.M., Bernoux M., Ve T., Bendak K., Chen C., Mackay J.P., Lawrence G.J., Hardham A., Ellis J.G., Williams S.J., Dodds P.N., Jones D.A.,& Kobe B. (2017b) Crystal structure of the *Melampsora lini* effector AvrP reveals insights into a possible nuclear function and recognition by the flax disease resistance protein. *P. Mol. Plant Pathol*.,12597.

Zhang, H., Gao, S., Lercher, M. J., Hu, S., & Chen, W.-H. (2012). EvolView, an online tool for visualizing, annotating and managing phylogenetic trees. *Nucleic Acids Research*, 40:W569‐W572.

Zhang, C., & Clarke, P. R. (2000). Chromatin-Independent Nuclear Envelope Assembly Induced by Ran GTPase in XenopusEgg Extracts. *Science*, 288(5470), 1429.

SUPPLEMENTAL MATERIAL

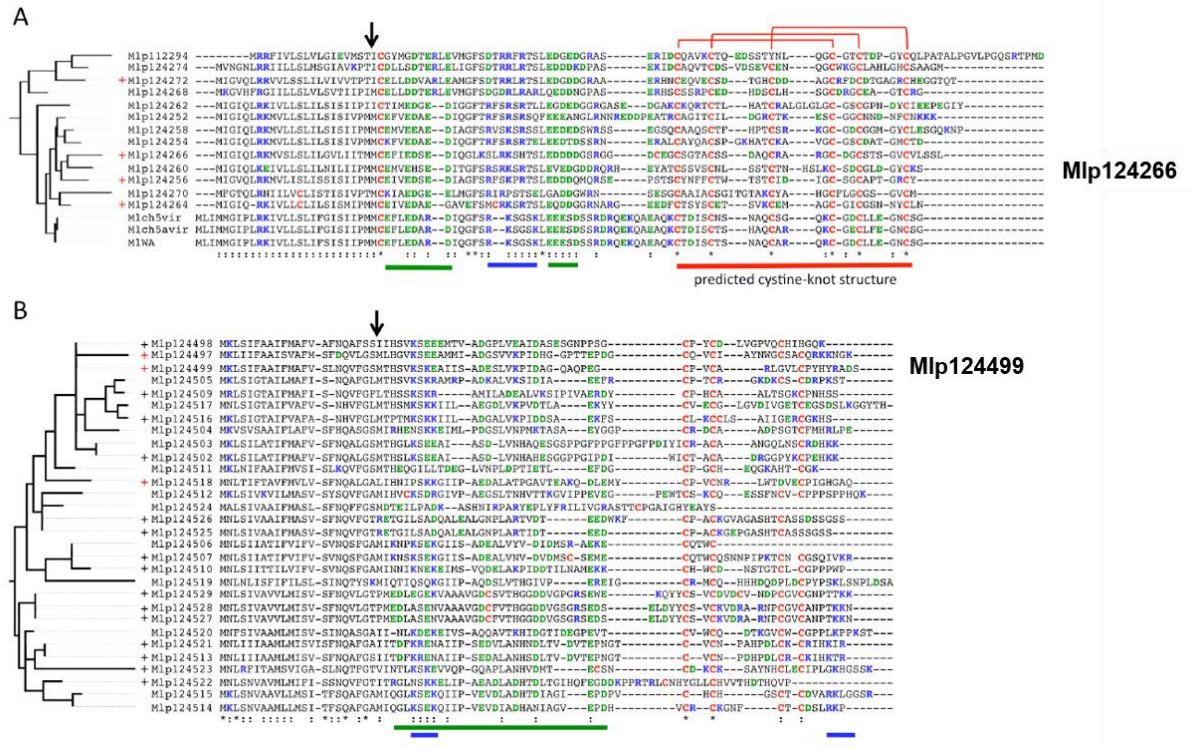


Figure S1: Sequence characteristics of proteins from MlpCPG54 and MlpCPGH1 families
Phylogenetic trees were built from corresponding aligned protein sequences from MlpCPG54 (A) and MlpCPGH1 (B) families. Branch length is proportional to phylogenetic distance. Black '+' indicate that expression of the gene has been detected in planta (Joly *et al.,* 2010; Duplessis *et al.,* 2011), red '+' indicate that the RT-qPCR expression profile of the gene has been established (Hacquard *et al.,* 2012). The black arrows indicate the predicted cleavage site of the signal peptide. Cysteine residues are in red, basic residues (lysine/K and arginine/R) are in blue, acid residues (aspartic acid/D and glutamic acid/E) are in green. Blue and green lines show basic and acid stretches of amino acids, whereas the red line delimited the predicted cystine-knot motif. An amino acid conservation of 50% or more is marked by a colon. An amino acid conservation of 90% or more is marked by a star.

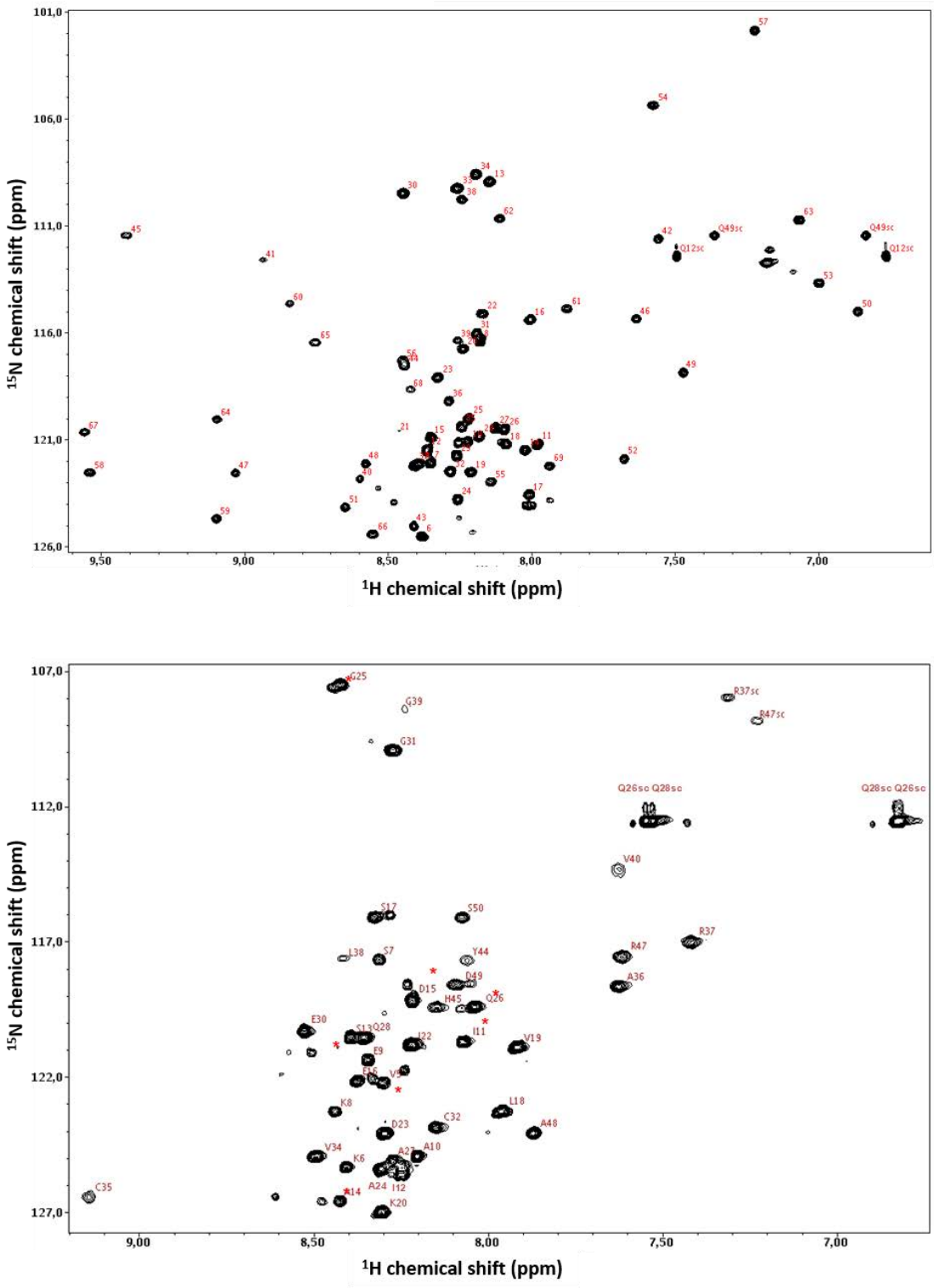


Figure S2: 1H-15N HSQC spectra of recombinant MLP124266.1 and MLP124499.
The NMR experiment was recorded on a 600 MHz spectrometer, at 298 K, on a 1.8 mM uniformly 15N-labeled sample in phosphate buffer pH 6.0. Backbone amide signals of MLP124266 (upper panel) or MLP124499 (lower panel) are labelled with the residue number and « sc » refers to side-chain amide signals. Minor peaks observed in MLP124499 1 are indicated by asteriks (*).

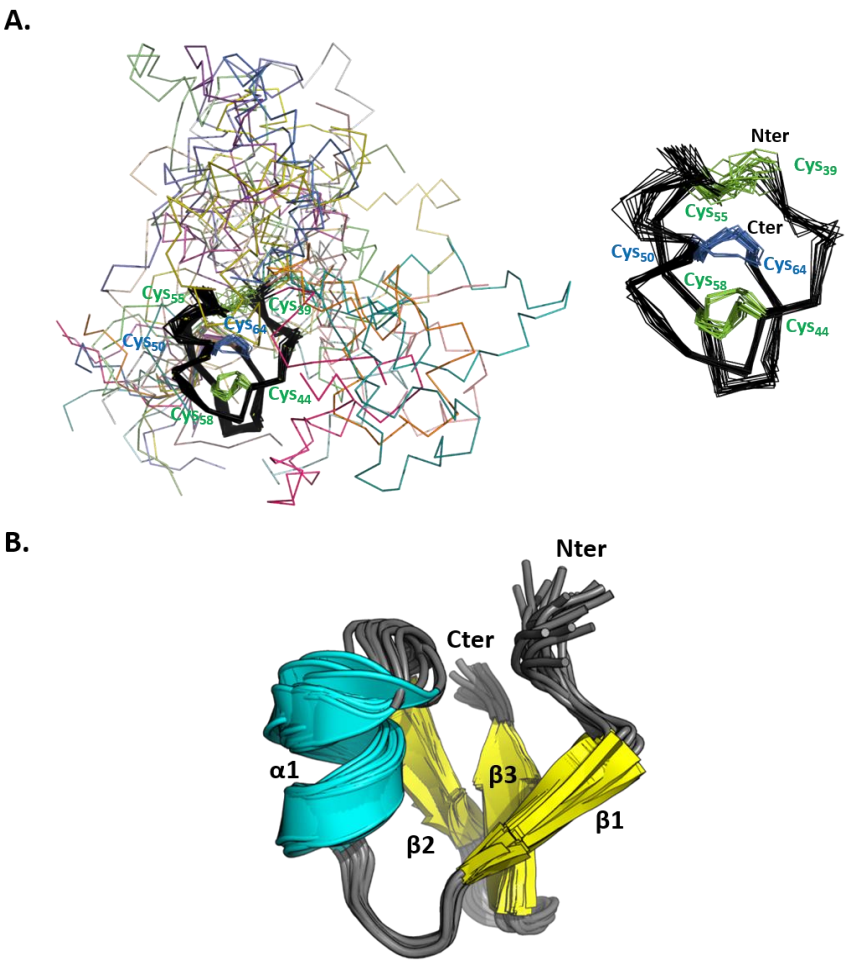


Figure S3: Superposition of the 20 lowest-energy NMR structures of MLP124266

A. Superimposition of backbone heavy atoms (N, Cα and C) for a family of 20 lowest energy NMR structures (left) and zoom on cystine-knot (right). Each structure is rendered as ribbon. Cystine-knot (region 39-65) is colored in black and the three disulfide brides are labeled and colored in green (Cys39-Cys55, Cys44-Cys58) or in blue (Cys50-Cys64). Nter and Cter extremities of the cystine-knot are indicated. B. NMR models were aligned by their cysteine-knot region (region 39-65). Secondary structures are labelled and represented as cartoon. The three disulfide brides are not represented to clarify the figure. Nter and Cter extremities of the cystine-knot region are indicated.

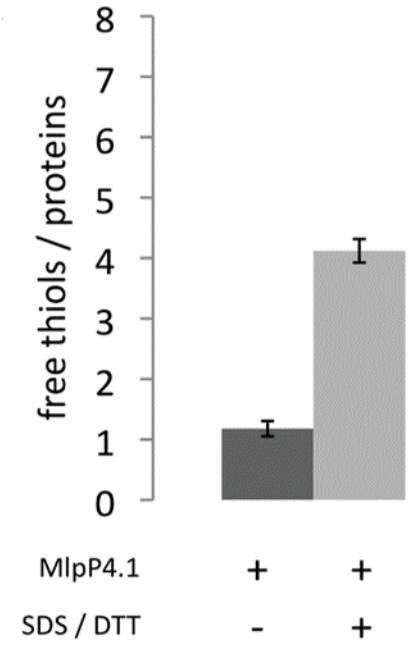


Figure S4. Thiol titration of recombinant MLP124266
Thiol titration of either native or denatured and reduced recombinant MLP124266

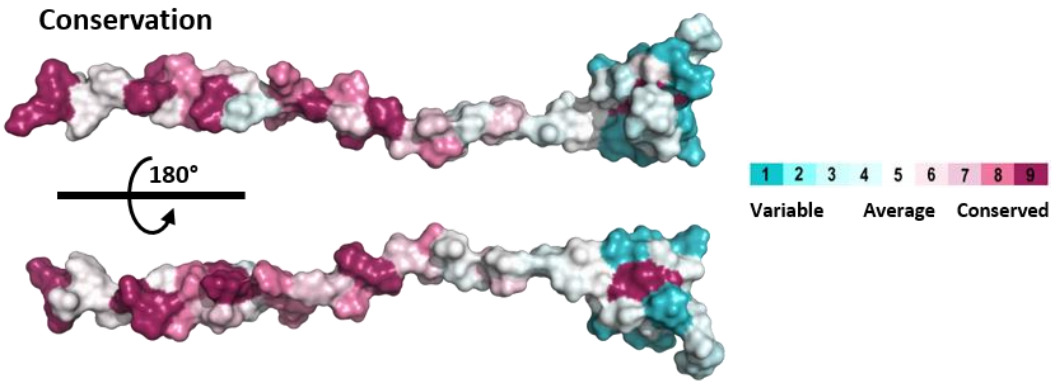


Figure S5: ConSurf distribution of conserved amino acids surfaces of MLP124266.

The conservation scale ranged from high (purple) to poor conservation (light blue), hydrophobic and hydrophilic patches are shown in orange and in blue respectively, positive-charge and negative-charge density are colored in blue and red respectively.

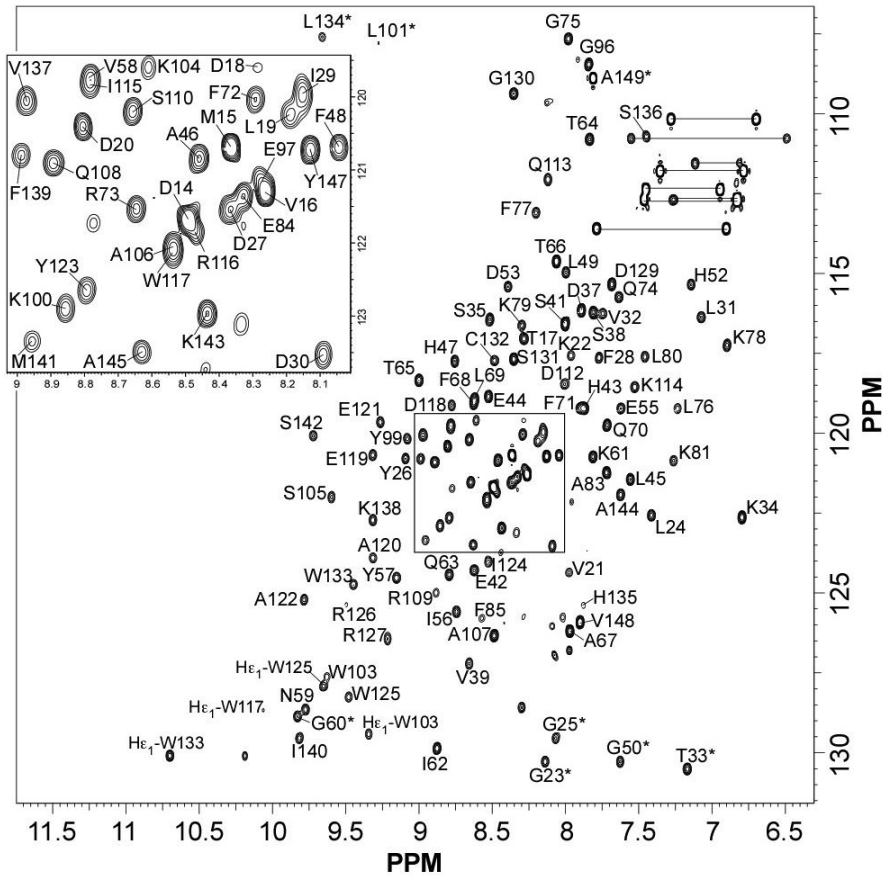


Figure S5: 15N-HSQC spectra of MLP124017.

Cross peak assignments are indicated using the one-letter amino acid and number (the asterisk indicates a folded peak). The central part of the spectrum is expended in the insert. Missing or not assigned residues: 1-15, 40, 82, 86-95, 98, 102, 111, 128, 150,151.

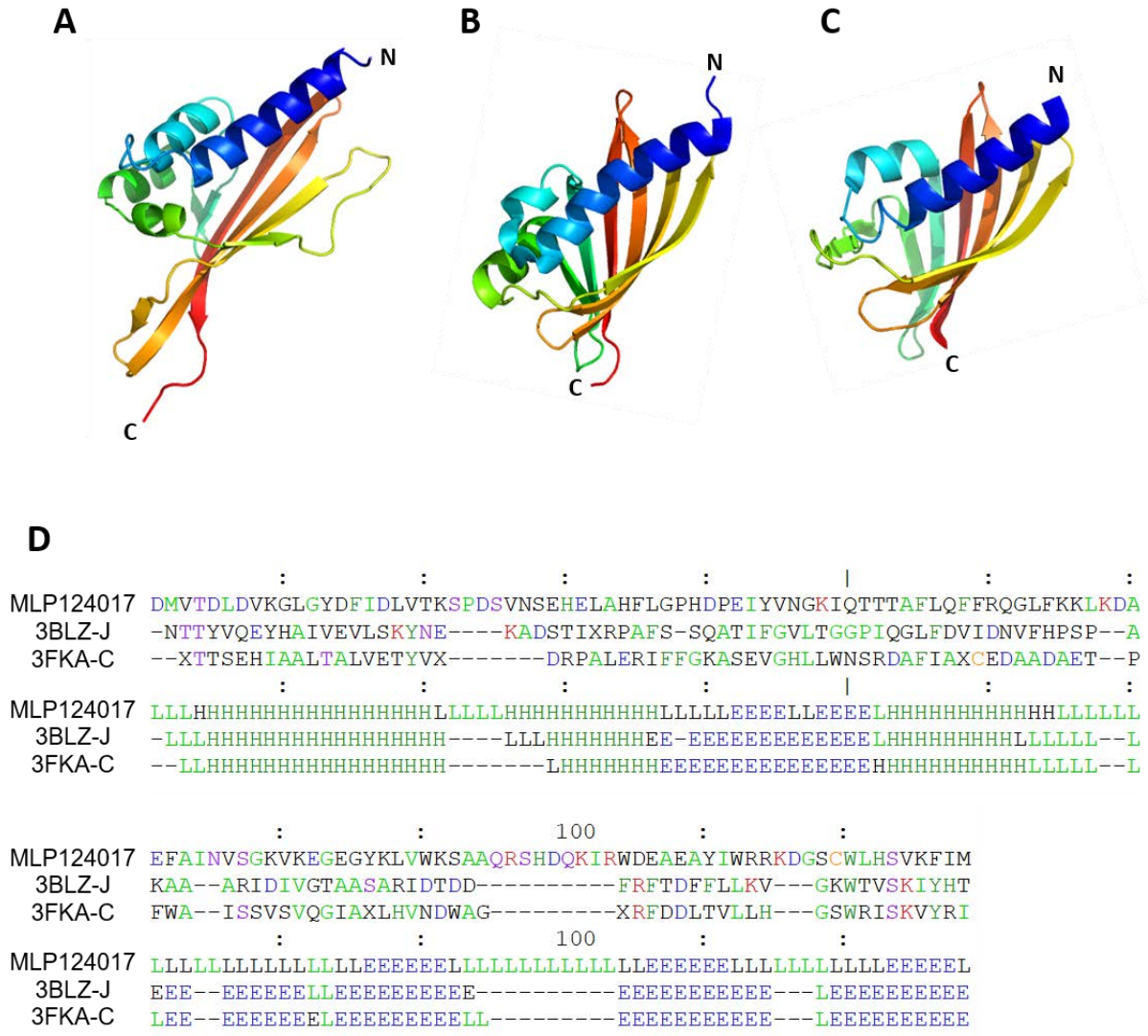


Figure S7: MLP124017 and two bacterial proteins SBAL_0622 and SPO1084 have similar Nuclear-transport factor 2 like fold

Structures of (A) MLP120117, (B) SBAL_0622 (PDBcode: 3BLZ) and (C) SPO1084 (PDBcode: 3FKA) were represented in cartoon with a rainbow coloration and generated using PYMOL (http://www.pymol.org). (D) Multiple structural alignment for the three proteins was done with Dali with the MLP124017 as the reference for numbering. The first part shows the amino acid sequences of the selected neighbors. The second part shows the secondary structure assignments by DSSP (H/h: helix, E/e: strand, L/l: coil). The most frequent amino acid type is colored in each column.

Table S1: Primers used to clone and express the 11 candidates in *Escherichia coli* Rosetta2

Protein ID	Cloning Enzymes	Primer Forward	Primer Reverse	Expression vector
MLP124478-$(His)_6$	NdeI/XhoI	gggggCATATGAAAGGTCGACACAAAAATG	gggggCTCGAGACATGTAACTTTCACGTTCC	pET-26b
MLP124530-(His)6	NdeI/XhoI	gggggCATATGAAAGACATCAATTTAGTTG	gggggCTCGAGTTCTGCAGACTCATTTTTGAAG	pET-26b
MLP124111-(His)6	NdeI/XhoI	gggggCATATGCAAACTCCCACCGGAGG	gggggCTCGAGGGGTACAACACATTTGTAACAAG	pET-26b
MLP124561-(His)6	NdeI/XhoI	gggggCATATGGCAGACTGGCACACAGCAG	gggggCTCGAGATAATCCCTCGTTATGACCAAAG	pET-26b
MLP37347-(His)6	NdeI/XhoI	gggggCATATGGTAGAACCACTCCCACTCG	gggggCTCGAGCCACTTCTTGGGTTTTGGTAC	pET-26b
MLP109567-(His)6	NdeI/XhoI	gggggCATATGGGACTGGTTTCTAGCTTAGC	gggggCTCGAGGTGAACCTCTTGCTCGGGAC	pET-26b
MLP124017-(His)6	NcoI/XhoI	gggggCCATGGAGCTACCAGAAAGTTTTG	gggggCTCGAGGTCTATGGCAACATATGGTGC	pET-28a
MLP107772-(His)6	NdeI/XhoI	gggggCATATGCAGGgtcgaGCAATCTGTC	gggggCTCGAGAGGACAATTTGGTCCGTTTCC	pET-26b
MLP124202-(His)6	NdeI/XhoI	gggggCATATGAATTTCTGTGTTAAGGGAC	gggggCTCGAGTTTGTGAGATGTCTTTCTTCC	pET-26b
MLP124266-(His)6	NdeI/NotI	gggggCATATGTGTGAATTTATAGAGGATTCAG	gggggGCGGCCGCTTAGTGATGGTGATGGTGATGTAGAGAGCTCAATACACATACACC'	pET-28a
MLPH1.1-(His)6	NdeI/NotI	gggggCATATGACTCACAGTGTCAAGAGCAAAG	gggggGCGGCCGCTCAGTGATGGTGATGGTGATGAGAATCCGCGCGGTAGTGATAG	pET-28a

Table S2: Statistics for 20 NMR structures of MLP124266
Structures were calculated using ARIA2 and analyzed using PROCHECK.

	MLP124266
NOE restraints	653
Short range (|i-j|≤1)	510
Medium range (1<|i-j|<5)	102
Long range (|i-j|≥5)	41
H-bond restraints	7
Dihedral restraints [(a)]	75
Number of NOE violations	
> 0.0 Å	82
> 0.3 Å	45
> 0.5 Å	18
Dihedral violations	
> 0°	63
> 5°	23
Ramachandran plot statistics	
most favorable regions (%)	81,9
allowed regions (%)	15,3
disallowed regions (%)	2,8
RMSD (Å) [(b)]	
Backbone	1.44 ± 0.52
Heavy atoms	2.08 ± 0.53

[(a)] Dihedral restraints were generated using DANGLE.

[(b)] RMSD were calculated over residues 39 to 69.

Table S3: NMR experiments acquired for structure calculations and chemical shift assignments for MLP124017

		Size			Sweep width (ppm)				
Experiments	nuclei	F3	F2	F1	F3	F2	F1	Mix(ms)	NS
^{15}N-HSQC	^{1}H, ^{15}N	1800	180	-	13.94	36	-	-	4
HNCO(*)	^{1}H,^{15}N,^{13}C	1366	60	100	13.94	36	15	-	8
HNCA(*)	^{1}H,^{15}N,^{13}C	1362	50	60	13.94	36	30	-	16
HNCOCACB(*)	^{1}H,^{15}N,^{13}C	1366	60	100	13.94	36	70	-	32
HNCACO(**)	^{1}H,^{15}N,^{13}C	1366	40	100	13.94	36	15	-	16
HNCACB(*)	^{1}H,^{15}N,^{13}C	1362	50	60	13.94	36	60	-	32
^{15}N-NOESY-HSQC	^{1}H,^{15}N, ^{1}H	1800	64	360	13.94	25	13.94	150	8
^{15}N-TOCSY-HSQC	^{1}H,^{15}N, ^{1}H	1800	64	360	13.94	25	13.94	54	8
NOESY (D_2O)	^{1}H, ^{1}H	2048	512	-	12.03	12.03	-	150	128
TOCSY (D_2O)	^{1}H, ^{1}H	2048	512	-	12.03	12.03	-	57.6	128
^{15}N-HSQC (***)	^{1}H, ^{15}N	1112	256	-	15.95	36	-	-	2

Experiments were recorded using the TOPSPIN Library (v. 2.1) at 298 K.
(*) BEST pulse sequences
(**) Watergate version
(***)SOFAST version(from Institut de Biologie Structurale, Grenoble).
(http://www.ibs.fr/research/scientific-output/software/pulse-sequence-tools/article/ibs-pulse-sequence-tools-for-bruker-spectrometers)

Table S4: Statistics for 20 NMR structures of MLP124017

	MLP124017
NOE restraints	1727
Short range (|i-j|≤1)	999
Medium range (1<|i-j|<5)	345
Long range (|i-j|≥5)	383
H-bond restraints	102
Dihedral restraints (a)	214
Number of NOE violations	
> 0.0 Å	186.35
> 0.3 Å	0.05
> 0.5 Å	0
Dihedral violations	
> 0°	22.25
> 5°	0
Ramachandran plot statistics	
Most favorable regions (%)	89.8
Additionally allowed regions (%)	9.8
Generously allowed regions (%)	0.4
Disallowed regions (%)	0.1
RMSD (Å) (b)	
Backbone	1.24 ± 0.39
Heavy atoms	1.81 ± 0.40

(a) Residues in regular secondary structures were derived from the chemical shifts using TALOS+ software.

(b) Main chain atoms (N, Cα, C) over the residues 20-81 and 100-141.

Partie 3: Article n°5 Rust fungal effectors mimic host transit peptides to translocate into chloroplasts

Benjamin Petre, Cécile Lorrain, Diane G.O. Saunders, Joe Win, Jan Sklenar, Sébastien Duplessis et Sophien Kamoun

Article de recherche publié dans la revue *Cellular Microbiology*, le 21 Mars 2016

Information supplémentaires disponibles : http://onlinelibrary.wiley.com/wol1/doi/10.1111/cmi.12530/suppinfo

Rust fungal effectors mimic host transit peptides to translocate into chloroplasts

Benjamin Petre[1,2,3], Cécile Lorrain[1,2,3], Diane G.O. Saunders[1,4,5], Joe Win[1], Jan Sklenar[1], Sébastien Duplessis[2,3*], Sophien Kamoun[1*]

[1] The Sainsbury Laboratory, Norwich Research Park, NR4 7UH Norwich, United Kingdom
[2] INRA, UMR 1136 Interactions Arbres/Microorganismes, Centre INRA Nancy Lorrain, 54280 Champenoux, France
[3] Université de Lorraine, UMR 1136 Interactions Arbres/Microorganismes, Faculté des Sciences et Technologies, 54506 Vandoeuvre-lès-Nancy, France
[4] The Genome Analysis Centre, Norwich Research Park, NR4 7UH Norwich, United Kingdom
[5] The John Innes Centre, Norwich Research Park, NR4 7UH Norwich, United Kingdom

* Sophien Kamoun, The Sainsbury Laboratory, Norwich Research Park, NR4 7UH Norwich, United Kingdom
sophien.kamoun@tsl.ac.uk

* Sébastien Duplessis, UMR 1136 Interactions Arbres/Microorganismes, Centre INRA Nancy Lorrain, 54280 Champenoux, France
duplessi@nancy.inra.fr

RUNNING TITLE

Effectors use molecular mimicry to target chloroplasts

SUMMARY

Parasite effector proteins target various host cell compartments to alter host processes and promote infection. How effectors cross membrane-rich interfaces to reach these compartments is a major question in effector biology. Growing evidence suggests that effectors use molecular mimicry to subvert host cell machinery for protein sorting. We recently identified CTP1 (chloroplast-targeted protein 1), a candidate effector from the poplar leaf rust fungus *Melampsora larici-populina* that carries a predicted transit peptide and accumulates in chloroplasts and mitochondria. Here, we show that the CTP1 transit peptide is necessary and sufficient for accumulation in the stroma of chloroplasts. CTP1 is part of a *Melampsora*-specific family of polymorphic secreted proteins. Two members of that family, CTP2 and CTP3, also translocate in chloroplasts in a N-terminal signal-dependent manner. CTP1, CTP2 and CTP3 are cleaved when they accumulate in chloroplasts, while they remain intact when they do not translocate into chloroplasts. Our findings reveal that fungi have evolved effector proteins that mimic plant-specific sorting signals to traffic within plant cells.

INTRODUCTION

Parasitic microbes deliver effector proteins into host cell compartments to manipulate a variety of processes and promote parasitic colonization (Dodds & Rathjen, 2010; Win *et al.*, 2012). One driving theme in plant pathology is to uncover how effector proteins traffic to these compartments; most notably how they cross a multitude of membrane-bound biological interfaces. Bacterial effectors cross the host plasma membrane through a complex molecular syringe (or secretion system), whereas effectors of filamentous parasites, namely fungi and oomycetes, enter host cells via mechanisms that remain to be discovered (Petre & Kamoun, 2014). Once inside host cells, some effectors subsequently traffic to distinct compartments including nuclei, plasma membrane, endoplasmic reticulum, tonoplast, vesicles, chloroplasts and mitochondria (Lindeberg *et al.*, 2012; Deslandes & Rivas, 2012; Hicks & Galan, 2013). To do so, several effectors possess domains that mimic host-targeting (or addressing) sequences. For example, effectors targeted to host nuclei through nuclear pores carry nuclear-localization signals (NLS) (Schornack *et al.*, 2010; Wirthmueller *et al.*, 2014). However, the mechanisms by which filamentous pathogens target effectors to membrane-bound organelles, such as chloroplasts, are poorly understood.

Chloroplasts are plant cell organelles composed of a double membrane envelope that surrounds a fluid called stroma, in which important structures such as thylakoids and starch granules reside. The envelope can form elongated structures called stromules (*stroma*-containing tub*ules*), which are thought to improve exchange of molecules with other compartments (Krenz *et al.*, 2012). Chloroplasts selectively internalize proteins from the cytosol via the TOC/TIC (Translocon at Outer membrane of Chloroplasts/Translocon at Inner membrane of Chloroplasts) complexes, by recognizing transit peptides that are present at the N-termini of proteins (Shi and Theg, 2013). Upon translocation, transit peptides are typically cleaved off by processing peptidases and mature proteins are released in the stroma (Texeira & Glaser, 2013). Transit peptides are variable in term of amino acid sequence and length, and no clear consensus has been identified (Demarsy *et al.*, 2014).

To date, only a few plant pathogen effectors have been reported to target chloroplasts (Jelenska *et al.*, 2007; Rogriguez-Herva *et al.*, 2012; Li *et al.*, 2014; Petre *et al.*, 2015; Torres Zabala *et al.*, 2015). HopI1, HopN1, HopK1 and AvrRps4 are type-III effector proteins of the bacterial plant pathogen *Pseudomonas syringae* pv *tomato* that function as virulence factors by modulating chloroplast structures and functions. How HopI1 and HopN1 translocate into chloroplasts is unknown, as they lack a N-terminal transit peptide. In contrast, HopK1 and AvrRps4 carry N-terminal cleavable regions that are sufficient for import into chloroplasts,

suggesting that these proteins carry transit peptides (Li *et al.*, 2014). However, whereas both proteins accumulate in chloroplasts in stable *Arabidopsis thaliana* transgenics, they diffuse in the cytosol and the nucleus when transiently expressed in *Nicotiana benthamiana* leaf cells (Li *et al.*, 2014). Besides, several studies have reported AvrRps4 in the nucleus and cytosol rather than, or in addition to chloroplasts (Heidrich *et al.*, 2011; Bhattacharjee *et al.*, 2011; Sohn *et al.*, 2012). Recently, Torres Zabala and colleagues also reported that the bacterial effectors HopR1 and HopO1-2 translocate into chloroplasts *in vitro* (Torres Zabala *et al.*, 2015).

Rust fungi (Basidiomycetes, Pucciniales) are obligate biotrophic parasites of plants (Duplessis *et al.*, 2015). Several rust species are devastating pathogens of crops, and a constant threat to agrosystems worldwide (Pennisi, 2010). For instance, the leaf rust fungus *Melampsora larici-populina* triggers annual epidemics on cultivated poplars and dramatically reduces plantations yield (Gérard *et al.* 2006). We recently took advantage of the predicted effector complement of *M. larici-populina* to perform an expression screen of the most promising candidates in leaf cells of the model plant *N. benthamiana*. These analyses revealed that the effector MLP107772, renamed chloroplast-targeted protein 1 (CTP1) in the present study, accumulates in chloroplasts and mitochondria (Petre *et al.*, 2015). Interestingly, CTP1 carries a predicted transit peptide of 83 amino acids at its N-terminus, which may explain its ability to translocate into chloroplasts (Petre *et al.*, 2015).

In this study, we further exploited the *N. benthamiana* experimental system to study effector trafficking across the chloroplast envelope. We combined genetic transformation with live cell imaging, cell fractionation and biochemical analyses to demonstrate that CTP1 resides in the stroma of chloroplasts. We then showed that the predicted CTP1 transit peptide is necessary and sufficient for accumulation in chloroplasts. CTP1 is a member of a *Melampsora*-specific effector family. We also provide evidence that other members of the family (named CTP2 and CTP3) also enter chloroplasts in a N-terminal signal-dependent manner. CTP1, CTP2 and CTP3 are cleaved *in planta*, and fusion proteins that do not accumulate in chloroplasts are not cleaved. Our data indicate that rust fungi have evolved a class of effector proteins that translocate into chloroplasts by functionally mimicking plant-targeting sequences.

RESULTS

CTP1-mCherry accumulates in the stroma of chloroplasts and in mitochondria

We recently reported that the mature form of CTP1 fused to a GFP ($CTP1_{28\text{-}171}$-GFP) accumulates inside chloroplasts and mitochondria when it is heterologously expressed in *N. benthamiana* leaf cells (Petre *et al.*, 2015). To test the robustness of this observation, we

constructed a fusion protein with a different fluorescent tag (CTP1$_{28-171}$-mCherry) and assayed its localization in *N. benthamiana* leaf cells by confocal microscopy. CTP1$_{28-171}$-mCherry fluorescent signal overlapped with chlorophyll autofluorescence, but also labeled stromules and mitochondria-like bodies (Figure 1A). Co-expression of CTP1$_{28-171}$-mCherry and ScCOX4-GFP, a marker of mitochondria (Nelson *et al.*, 2007), showed overlapping signals, confirming the presence of CTP1$_{28-171}$-mCherry in mitochondria (Figure 1B). The co-expression of CTP1$_{28-171}$-GFP and CTP1$_{28-171}$-mCherry revealed perfectly overlapping fluorescent signals in cells accumulating the two fusion proteins, indicating that the nature of the fluorescent tag does not affect the localization. Saturation of the fluorescent signals revealed no accumulation in the cytosol and the nucleus, suggesting that CTP1$_{28-171}$-fluorescent protein fusions are efficiently internalized into these organelles (Figure S1).

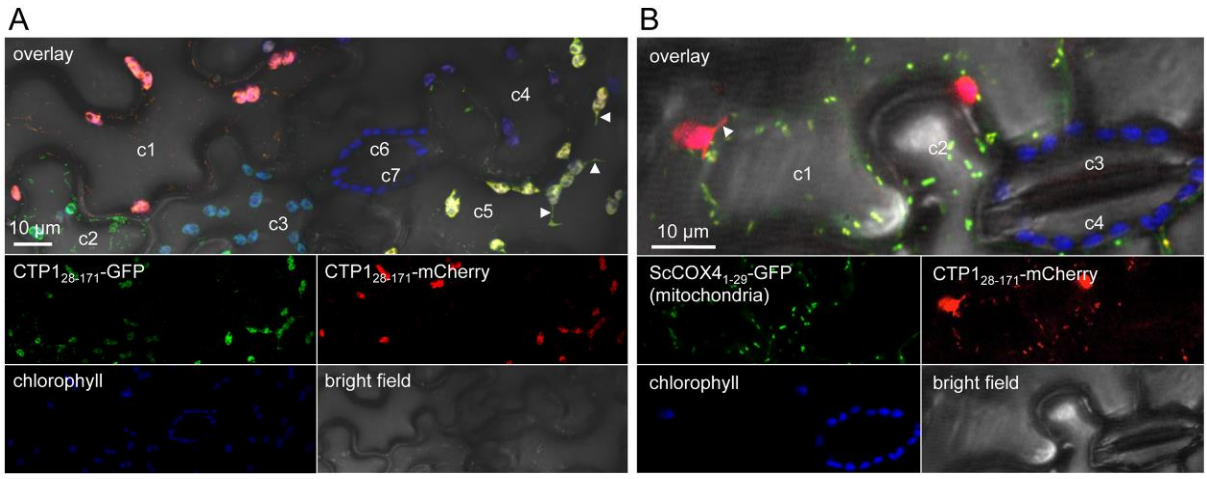


Figure 1: CTP1$_{28-171}$-fluorescent protein fusions accumulate in chloroplasts and mitochondria
(A) Live-cell imaging of CTP1$_{28-171}$-GFP and CTP1$_{28-171}$-mCherry in *N. benthamiana* leaf pavement cells. Images present a maximal projection of 23 optical sections (z-stack: 18.4 µm). Cells marked c1 to c5 are pavement cells differentially accumulating the two fusion proteins (c1 and c5 accumulate the mCherry fusion; c1, c2, c3 and c5 accumulate the GFP fusion). Cells marked c6 and c7 are guard cells. (B) Live-cell imaging of CTP1$_{28-171}$-mCherry and ScCOX4$_{1-29}$-GFP (mitochondria marker) in *N. benthamiana* leaf pavement cells. Images present a single optical section. Cells marked c1 and c2 are pavement cells. Cells marked c3 and c4 are guard cells. Proteins were transiently expressed in *N. benthamiana* leaf cells by agroinfiltration. Live-cell imaging was performed with a laser-scanning confocal microscope with a sequential scanning mode two days after infiltration. The GFP and the chlorophyll were excited at 488 nm; the mCherry was excited at 561 nm. GFP (green), mCherry (red) and chlorophyll (blue) fluorescent signals were collected at 505-525 nm, 580-620 nm and 680-700 nm, respectively. White arrowheads indicate stromules.

To unambiguously determine whether CTP1 translocates inside chloroplasts, we performed a series of co-localizations of CTP1$_{28-171}$-fluorescent protein fusions with various markers of sub-chloroplastic compartments. First, we co-expressed CTP1$_{28-171}$-GFP with NbpMSRA-mCherry, a marker of the stroma of chloroplasts (Lorrain *et al.*, 2014a). In this assay, GFP and mCherry fluorescent signals overlapped perfectly, indicating that CTP1 was present in the stroma

(Figure 2A). Second, co-expression of CTP1$_{28-171}$-GFP with a GBSSI-YFP fusion, a starch granule marker (Wang *et al.*, 2013), allowed a better visualization of chloroplasts at high magnification. The uniform distribution of the GFP fluorescent signal around GBSSI-YFP labeled starch granules further confirmed the accumulation of CTP1$_{28-171}$-GFP in the stroma (Figure 2B and 2C). Finally, we used cell fractionation to isolate chloroplasts from *N. benthamiana* leaves expressing CTP1$_{28-171}$-GFP. Confocal microscopy performed on purified organelles revealed a strong fluorescent signal in some chloroplasts, which was uniformly distributed around the starch granules (Figure 2D). Note that most of the chloroplasts we isolated come from mesophyll, guard and abaxial epidermal cells, which were not efficiently transformed in transient assays under our conditions. Therefore, as expected, only a small fraction of chloroplasts showed a fluorescent signal (Figure 2D). Altogether, these experiments indicate that CTP1 enters the chloroplasts to accumulate in the stroma.

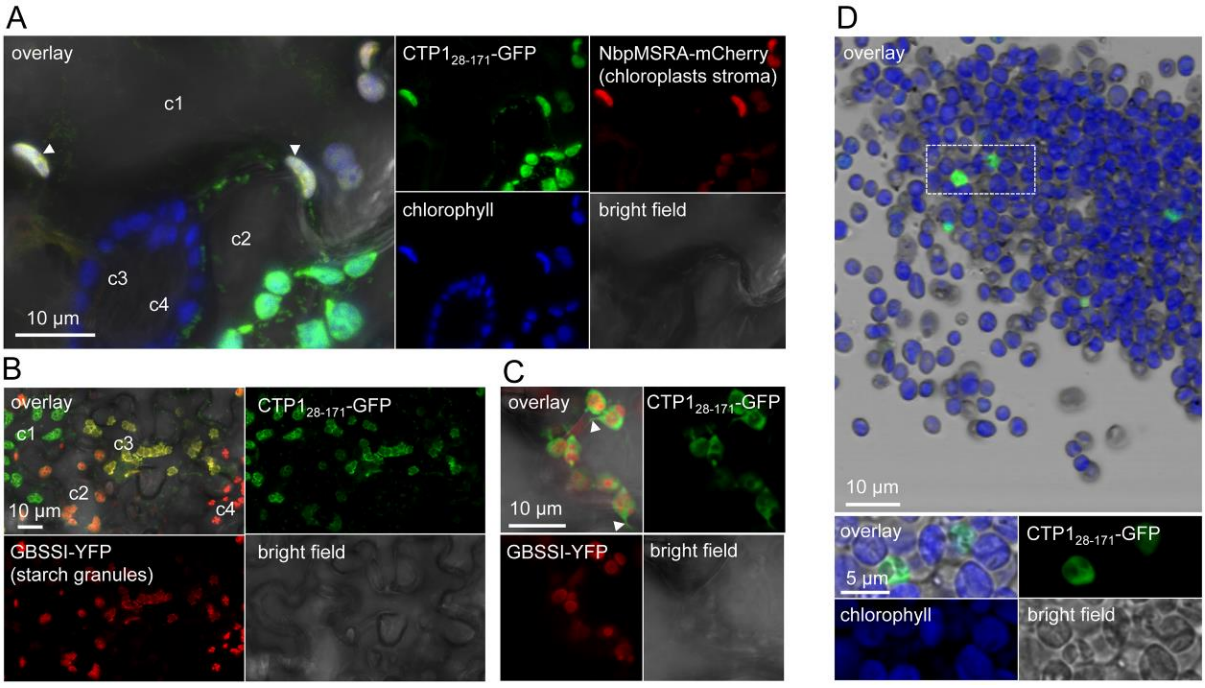


Figure 2. CTP1$_{28-171}$-fluorescent protein fusions accumulate in the stroma of chloroplasts
(A) Live-cell imaging of CTP1$_{28-171}$-GFP and NbpMSRA-mCherry (chloroplast stroma marker) in *N. benthamiana* leaf pavement cells. Images present a maximal projection of 21 optical sections (Z-stack: 16.8 µm). Cells marked c1 and c2 are pavement cells differentially accumulating the two fusion proteins. Cells marked c3 and c4 are guard cells. White arrowheads show chloroplasts from c1 where

---

Figure 2. CTP1$_{28-171}$-fluorescent proteins accumulate in the stroma of chloroplasts
GFP, mCherry and chlorophyll fluorescent signals overlap. (B) Live-cell imaging of CTP1$_{28-171}$-GFP and GBSSI-YFP (starch granule marker) in *N. benthamiana* leaf pavement cells. Images present a maximal projection of 15 optical sections (z-stack: 12.0 µm). Cells marked c1 to c4 are pavement cells differentially accumulating the two fusion proteins. (C) Live-cell imaging of CTP1$_{28-171}$-GFP and GBSSI-YFP in *N. benthamiana* leaf pavement cells. Images present a single optical section. White arrowheads show stromules with GFP fluorescent signal. (D) Live-cell imaging of CTP1$_{28-171}$-GFP in chloroplasts isolated from agrotransformed *N. benthamiana* leaf cells. Images show a single optical section. The top

panel shows a low magnification overlay image (GFP fluorescence, chlorophyll fluorescence and bright field). The white dotted-line rectangle indicates the close-up shown in the bottom panel. Proteins were expressed in *N. benthamiana* leaf cells by agroinfiltration. Live-cell imaging was performed with a laser-scanning confocal microscope with a sequential scanning mode two days after infiltration. For (D), leaves were homogenized and the lysate was filtered and centrifuged to retrieve organelles. The chloroplasts were separated on a Percoll gradient and collected with a pipet for immediate use. The GFP and the chlorophyll were excited at 488 nm; the YFP was excited at 514 nm; the mCherry was excited at 561 nm. YFP (red), mCherry (red) and chlorophyll (blue) fluorescent signals were collected at 530-550 nm, 580-620 nm and 680-700 nm, respectively. GFP (green) fluorescent signal was collected at 505-525 nm (A and D) or 505-510 nm (B and C).

CTP1 carries a functional chloroplast transit peptide

We hypothesized that the predicted 83-amino-acid N-terminal transit peptide of CTP1 mediates trafficking to chloroplasts and mitochondria (Petre *et al.*, 2015; Figure 3A). To test the functionality of this predicted transit peptide, we generated two truncations: $CTP1_{28\text{-}114}$ (the predicted transit peptide plus four amino acids), and $CTP1_{107\text{-}171}$ (CTP1 without its predicted transit peptide). We fused the truncations with GFP ($CTP1_{28\text{-}114}$-GFP and $CTP1_{107\text{-}171}$-GFP) and determined their subcellular localization in *N. benthamiana* leaf cells. $CTP1_{28\text{-}114}$-GFP signal overlapped with chlorophyll, but also showed a weak signal in the cytosol and the nucleus (Figure 3B). We used $CTP1_{28\text{-}171}$-mCherry as a marker of both the stroma and mitochondria, and co-expressed it with $CTP1_{28\text{-}114}$-GFP. Fluorescent signals overlapped in the stroma but not in mitochondria, suggesting that $CTP1_{28\text{-}114}$-GFP accumulates only in chloroplasts. Further observations performed on chloroplasts isolated from leaves expressing $CTP1_{28\text{-}114}$-GFP confirmed the presence of the fusion protein in the stroma (Figure 3C). On the other hand, $CTP1_{107\text{-}171}$-GFP signal was not detected in chloroplasts or mitochondria, and showed a nucleocytosolic distribution (Figure 3D). This indicates that CTP1 lacking its transit peptide can no longer enter these organelles. In stable *N. benthamiana* transgenics, $CTP1_{28\text{-}114}$-GFP fluorescent signal specifically labeled chloroplasts and stromules, validating the observations made in transient assays (Figure 3E). We conclude that CTP1 carries a functional transit peptide, which is necessary and sufficient to mediate trafficking to chloroplasts. However, this transit peptide is necessary but not sufficient for accumulation in mitochondria.

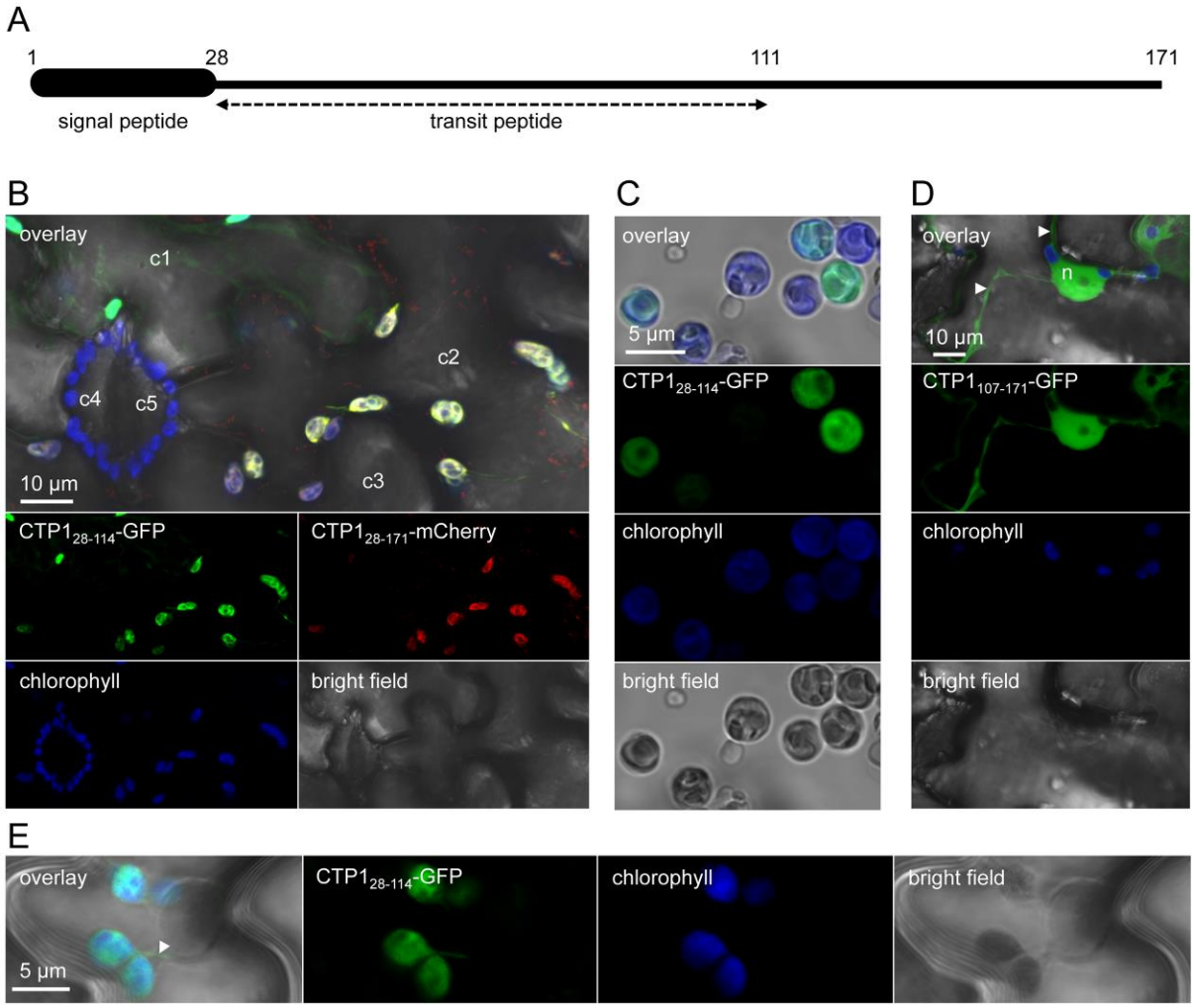


Figure 3. CTP1 carries a functional chloroplast transit peptide

(A) Protein architecture of CTP1, with its two predicted targeting sequences. (B) Live-cell imaging of $CTP1_{28\text{-}114}$-GFP and $CTP1_{28\text{-}171}$-mCherry (here used as a marker of the stroma of chloroplasts and mitochondria) in *N. benthamiana* leaf pavement cells. Images present a maximal projection of 11 optical sections (z-stack: 8.8 µm). Cells marked c1 to c3 are pavement cells differentially accumulating the two fusion proteins. Cells marked c4 and c5 are guard cells. (C) Live-cell imaging of $CTP1_{28\text{-}114}$-GFP in chloroplasts isolated from agrotransformed *N. benthamiana* leaf cells. Leaves were homogenized and the lysate was filtered and centrifuged to retrieve organelles. The chloroplasts were separated on a Percoll gradient and collected with a pipet for immediate use. Images present a single optical section. (D) Live-cell imaging of $CTP1_{107\text{-}171}$-GFP in *N. benthamiana* leaf pavement cells. Images present a maximal projection of three optical sections (z-stack: 2.4 µm). White arrowheads indicate cytosol. n: nucleus. (E) Live-cell imaging of $CTP1_{28\text{-}114}$-GFP in leaf pavement cells of a stable transgenic *N. benthamiana* plant. Images show a single optical section. The white arrowhead indicates a stromule. For (B) to (D), proteins were expressed in *N. benthamiana* leaf cells by agroinfiltration.
Live-cell imaging was performed with a laser-scanning confocal microscope with a sequential scanning mode two days after infiltration. The GFP and the chlorophyll were excited at 488 nm. The mCherry was excited at 561 nm. GFP (green), mCherry (red) and chlorophyll (blue) fluorescent signals were collected at 505-525 nm, 580-620 nm, and 680-700 nm, respectively.

CTP1 is part of a *Melampsora*-specific family of modular and polymorphic proteins

CTP1 was first described as an orphan gene in *M. larici-populina* (Hacquard *et al.*, 2012; Saunders *et al.*, 2012). However, a more recent analysis reported that CTP1 displays sequence similarity to other secreted proteins of *M. larici-populina* (Cantu *et al.*, 2013). To establish a list of proteins related to CTP1, we performed sequence similarity searches as well as protein and gene structure similarity analyses. We identified four proteins with sequence similarity to CTP1: Mlp52246 (hereafter CTP2), Mlp55690 and Mlp116083 in *M. larici-populina*, and Melli_sc2834 (hereafter CTP3) in the flax rust fungus *Melampsora lini*. The four proteins showed amino acid similarity to CTP1 that ranged from 16 to 28%. Although the sequence similarity was low, these five proteins display a similar modular protein structure. They all have a predicted signal peptide followed by a predicted transit peptide of variable length, as well as two to three repeated domains or modules, with the first one overlapping with the predicted transit peptide (Figure 4A). Also, gene structures matched protein organization, with the first exon encoding the signal peptide, and each of the other exons encoding a module. These modules are highly divergent (average paired amino acid identity below 30%, ranging from 15 to 100%) but present stretches of amino acids with similar properties and four conserved cysteine residues (Figure 4B). Some stretches of amino acids are predicted to have a secondary structure conserved across the different repeats (Figure 4C). Interestingly, the phylogeny of the modules does not fully match the gene organization, suggesting that the CTP family evolved potentially by exon shuffling. We conclude that CTP1 is part of a family of secreted, modular proteins that have probably emerged and diversified recently in *Melampsora* species.

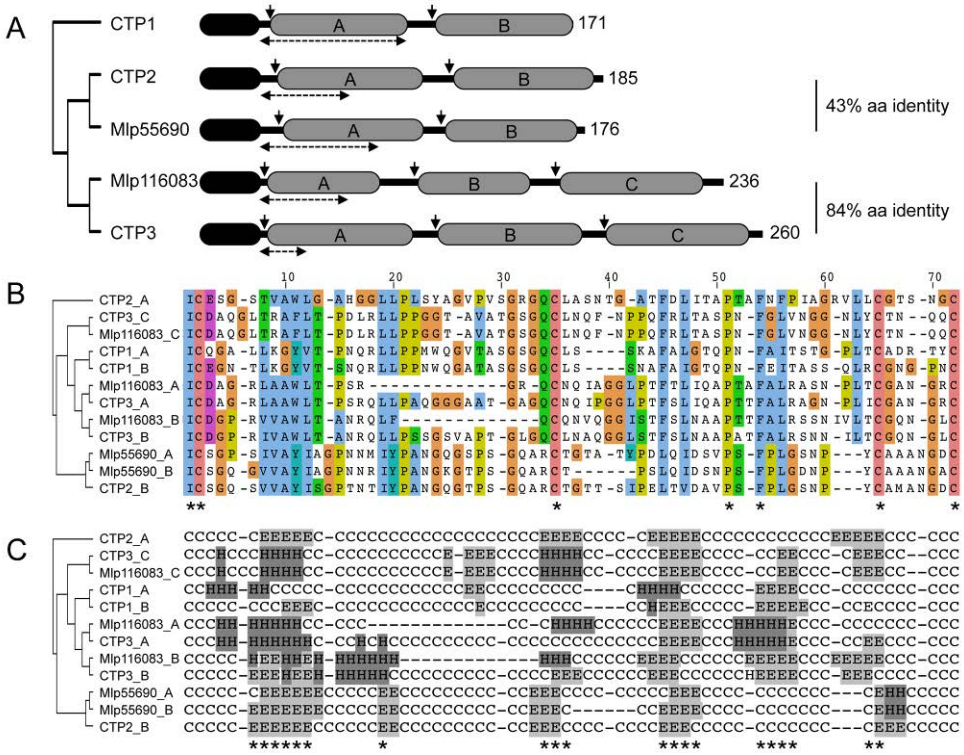


Figure 4. CTP1 is part of a family of modular and polymorphic proteins

(A) Gene and protein architecture of CTP1 (alias Mlp107772), CTP2 (alias Mlp52246), Mlp55690, Mlp116083, and CTP3 (alias Melli_sc2834.1). Black: predicted signal peptide; horizontal double-arrowed dotted lines: predicted transit peptide; vertical arrows: intron positions; grey: repeated domains or modules. Numbers on the right-hand side indicate the number of amino acids for each protein. Chloroplast transit peptides were predicted with ChloroP 1.1 and the following Score/CS-Score/cTP-length values were obtained: CTP1 (0.533/1.040/83); CTP2 (0.503/4.184/38); Mlp55690 (0.484/5.625/67); Mlp116083 (0.510/2.432/40); CTP3 (0.493/1.934/23). (B) Phylogeny and amino acid alignment of the 12 repeated modules displayed in (A). Modules are named according to their protein of origin and their position in the protein sequence (A, B or C) as indicated in (A). The phylogenetic tree was built with Jalview (neighbor joining using % identity) and the amino acid alignment was performed with ClustalX. Amino acid residues are colored according to the ClustalX scheme. Black asterisks mark the seven fully conserved residues. (C) Predicted secondary structure of the repeated modules. The phylogenetic tree and the amino acid residue positions are the same as in (B). Secondary structures were predicted using Scratch Protein Predictor with the SSPro: secondary structure (3 class) mode. C: no particular prediction; H: alpha-helix; E: extended strand. H and E are highlighted in dark and light grey, respectively, for better visualization. Asterisks indicate amino acid positions where at least 50% of the residues are predicted to be part of a secondary structure.

CTP2 and CTP3, two proteins similar to CTP1, accumulate in chloroplasts

To determine whether CTP1-like proteins also target chloroplasts and mitochondria, we studied CTP2 and CTP3, two proteins representative of the diversity of the family, and which carry two and three modules, respectively. We generated fusions between the mature forms of CTP2 and CTP3 (i.e. without the signal peptide) with GFP or mCherry and expressed the fusion proteins in *N. benthamiana*. Confocal microscopy revealed that the two fusion proteins display a signal that overlapped with chlorophyll. CTP2 fluorescent signal accumulated in the stroma of the chloroplasts but also in the nucleus and cytosol (Figure 5A and 5B). CTP3

fluorescent signal distributed similarly to CTP2, but also labeled discrete areas of approximately 0.1 to 0.3 µm of diameter within the chloroplasts (Figure 5C to E). A *Populus trichocarpa* Coproporphyrinogen III Oxidase-GFP fusion (PtCPO-GFP) was previously observed to accumulate in similar sub-chloroplastic structures in *N. benthamiana* (Lorrain *et al.*, 2014b). Co-expression of $CTP3_{28-260}$-mCherry and PtCPO-GFP revealed overlapping signals, indicating that $CTP3_{28-260}$-fluorescent protein fusions reside in PtCPO-containing compartments whose nature remains to be determined (Figure 5F). Co-expression of $CTP2_{29-185}$- and $CTP3_{28-260}$-fluorescent protein fusions with mitochondria markers (ScCOX4-GFP or ScCOX4-mCherry) revealed no overlapping signals, suggesting that CTP2 and CTP3 do not accumulate in mitochondria (Figure 5B and D). We conclude that CTP2 and CTP3 translocate into chloroplasts when expressed in *N. benthamiana* cells.

CTP1, CTP2 and CTP3 are processed *in planta*

Upon translocation in chloroplasts and mitochondria, processing peptidases cleave off transit peptides (Teixeira and Glaser, 2013). To determine whether CTPs undergo processing *in planta*, we performed anti-GFP western blotting with total proteins isolated from *N. benthamiana* leaves expressing the CTP-GFP fusions. The band signals for all fusion proteins were smaller than their theoretical size, suggesting that they undergo cleavage *in planta*. Indeed, $CTP1_{28-171}$-GFP showed a unique band signal at 34 kDa, approximately 8 kDa below its expected size of 41.8 kDa (Figure 6A), confirming previous observations (Petre *et al.*, 2015). Both $CTP2_{29-185}$-GFP and $CTP3_{28-260}$-GFP displayed a band signal that coincided with their predicted molecular weight at 42.7 and 51.0 kDa, respectively, but also a second band signal that was 8 and 16 kDa smaller, respectively. To identify the region of CTP1 that is cleaved off, we performed a liquid chromatography - tandem mass spectrometry (LC-MS/MS) analysis of $CTP1_{28-171}$-GFP fusions purified by anti-GFP immunoprecipitation (Figure 6B). Tryptic peptides matching the N-terminal region of the mature form of CTP1 (amino acid residues 28 to 91) were not detected in two of the three replicates, and weekly detected in the third replicate, whereas tryptic peptides matching the C-terminal region (amino acid residues 92 to 178) were abundantly detected (Figure 6C; Table S2). This result suggests that the N-terminal region of CTP1 is missing from the purified fusion protein, and that the processing event occurs between the amino acid residues 69 and 91 (i.e. the last tryptic peptide with low coverage).

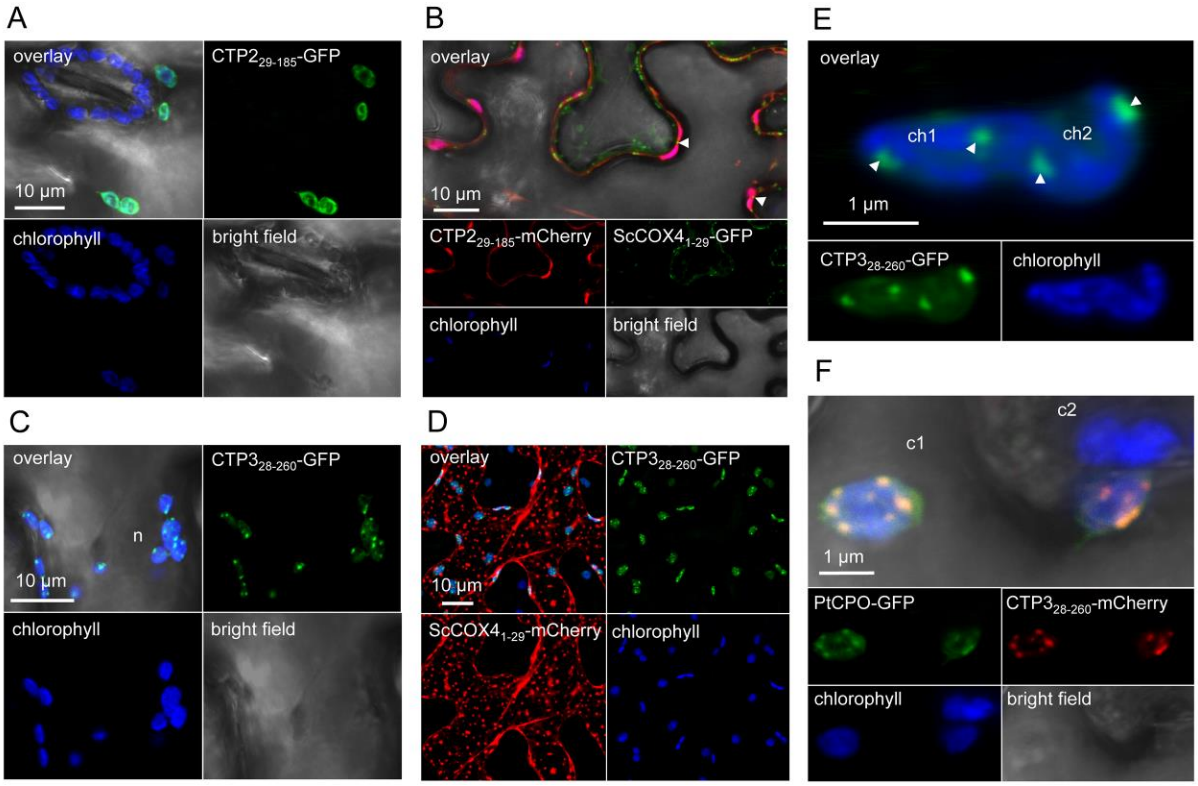


Figure 5. CTP2$_{29-185}$- and CTP3$_{28-260}$-fluorescent protein fusions accumulate in chloroplasts but not in mitochondria

(A) Live-cell imaging of CTP2$_{29-185}$-GFP in *N. benthamiana* leaf pavement cells. Images show a maximal projection of seven optical sections (z-stack: 5.6 µm). (B) Live-cell imaging of CTP2$_{29-185}$-mCherry and ScCOX4$_{1-29}$-GFP (mitochondria marker) in *N. benthamiana* leaf pavement cells. Images show a single optical section. White arrowheads indicate mitochondria. (C) Live-cell imaging of CTP3$_{28-260}$-GFP in *N. benthamiana* leaf pavement cells. Images show a maximal projection of 15 optical sections (z-stack: 12.0 µm). (D) Live-cell imaging of CTP3$_{28-260}$-GFP and ScCOX4$_{1-29}$-mCherry in *N. benthamiana* leaf pavement cells. Images show a maximal projection of 41 optical sections (z-stack: 32.8 µm). Due to the thickness of the Z-stack, the bright field is not interpretable and thus is not displayed. (E) Live-cell imaging of CTP3$_{28-260}$-GFP in *N. benthamiana* leaf pavement cells. Images show a single optical section. ch1 and ch2 indicate two closely associated chloroplasts. White arrowheads indicate discrete areas inside chloroplasts that are labeled by CTP3$_{28-260}$-GFP. (F) Live-cell imaging of CTP3$_{28-260}$-mCherry and PtCPO-GFP (marker of undetermined discrete chloroplast bodies) in *N. benthamiana* leaf pavement cells. Images present a maximal projection of 11 optical sections (z-stack: 8.8 µm). The cell marked c1 is a pavement cell that accumulates both protein fusions, the cell marked c2 is a guard cell. Proteins were expressed in *N. benthamiana* leaf cells by agroinfiltration. Live-cell imaging was performed with a laser-scanning confocal microscope with a sequential scanning mode two days after infiltration. The GFP and the chlorophyll were excited at 488 nm; the mCherry was excited at 561 nm. GFP (green), mCherry (red) and chlorophyll (blue) fluorescent signals were collected at 505-525 nm, 580-620 nm, and 680-700 nm, respectively.

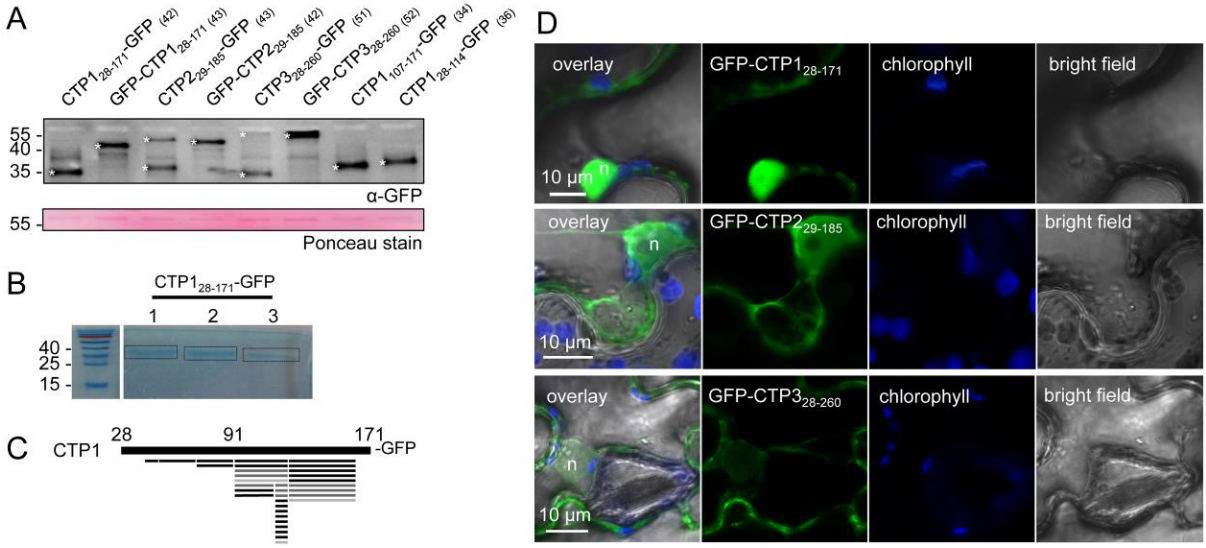


Figure 6. CTP1, CTP2 and CTP3 undergo processing *in planta*

(A) Western blotting analyses of GFP fusions. Proteins were transiently expressed in *N. benthamiana* leaf cells by agroinfiltration. Total protein extraction was performed two days after infiltration by grinding leaves in liquid nitrogen and immediately reducing and denaturing proteins from the leaf powder in a Laemmli buffer 15 min at 95°C. Proteins were separated by 15% SDS-PAGE and transferred onto a nitrocellulose membrane. Immune detection was performed with anti-GFP antibodies. For each fusion protein the theoretical size is indicated in parentheses in kiloDaltons (kDa). Protein sizes are indicated on the left-hand side of the image in kDa. Asterisks indicate protein bands discussed in the text. Ponceau S staining of the RubisCO on the membrane was used as a loading control. (B) Separation by SDS-PAGE of the protein mixtures obtained by anti-GFP immunoprecipitation of CTP1$_{28-171}$-GFP. Proteins were visualized by Coomassie blue staining. Numbers (1, 2, and 3) indicate biological replicates. Rectangles indicate the gel band that were excised and used for further trypsin digestion and LC-MS/MS analysis. Replicate 1 was previously reported in Petre *et al*., 2015. The page ruler on the left-hand side and the protein bands on the right-hand side originate from the same gel. Numbers on the left-hand side of the page ruler indicate protein sizes in kDa. (C) Tryptic peptides matching the CTP1 sequence (large black bar). Small bars indicate tryptic peptides that were identified by LC-MS/MS analysis of the trypsin-digested protein bands shown in (B). Black, dark grey, and light grey colors indicate peptides identified in replicates 1, 2, and 3, respectively. Numbers indicate amino acid positions. (D) Live-cell imaging of GFP-CTP1$_{28-171}$ (top panel), GFP-CTP2$_{29-185}$ (middle panel), and GFP-CTP3$_{28-260}$ in *N. benthamiana* leaf pavement cells. Images show a single optical section (top and bottom panels) or a maximal projection of three optical sections (z-stack: 2.4 µm; middle panel). Proteins were expressed in *N. benthamiana* leaf cells by agroinfiltration. Live-cell imaging was performed with a laser-scanning confocal microscope with a sequential scanning mode two days after infiltration. The GFP and the chlorophyll were excited at 488 nm. GFP (green) and chlorophyll (blue) fluorescent signals were collected at 505-525 nm and 680-700 nm, respectively. n: nucleus.

DISCUSSION

The N-terminus of the rust fungal protein CTP1 is necessary and sufficient to enter the chloroplast organelle of plants indicating that it mimics host transit peptides to subvert host cell machinery. Host mimicking domains are frequent in pathogen effectors (Hicks and Galan, 2013). However, how did a chloroplast transit peptide evolve in an organism that lacks chloroplasts? It should be noted that mitochondria and chloroplast transit peptides have similar

properties (Carrie and Small 2013). The CTP1 transit peptide may have evolved from a mitochondrial transit peptide endogenous to the fungus. The observation that CTP1 can accumulate in mitochondria somewhat supports this hypothesis although the transit peptide does not fully account for this (Figure 4). On the other hand, the transit peptide may have evolved *de novo*. Transit peptides typically do not display conserved amino acid sequences, and their functionality is based on conserved physical properties (Bruce 2001). Therefore, they are more likely to evolve *de novo* compared with other domains whose function is based on strict amino acid sequence conservation (Tonkin *et al.*, 2008). This possibility is perhaps more likely given that CTP1 does not share sequence similarity to any known protein sequence, and belongs to a polymorphic and lineage-specific family. Hence, the CTP1 transit peptide may have evolved *de novo* possibly by neofunctionalization after a duplication or recombination event.

Host-translocated effectors of oomycete and fungal plant pathogens usually carry host-targeting sequences just downstream of their signal peptide (Petre and Kamoun 2014). These targeting sequences have no known function in host cells. We have shown that the N-terminal region immediately downstream of CTP1 signal peptide functions as transit peptide in plant cells. This finding implies that, if CTP1 encodes a N-terminal host-targeting sequence, it overlaps with its transit peptide. It is also possible that CTP1 does not require a host-targeting sequence.

CTP1 accumulates in both chloroplasts and mitochondria. Although accumulation in mitochondria is weak, CTP1 could very well target the two organelles. Many plant proteins have evolved to target and function in two cell compartments, and are called 'dually-targeted' proteins (Carrie and Whelan 2013). For instance, over a hundred plant proteins target both chloroplasts and mitochondria (Carrie and Small 2013). Some pathogen effectors have also been reported in multiple host cell compartments (Heidrich *et al.*, 2011; Hicks and Galan, 2013; Li *et al.*, 2014) and may also be dually targeted proteins. However, the disparate localizations observed by different authors, as for instance with the RTP1 family of rust fungal proteins, could be due to different tags or methods of localization (Kemen *et al.*, 2005; Hacquard *et al.*, 2012; Kemen *et al.*, 2013).

We found that CTP1 transit peptide (amino acid positions 28 to 114) is necessary but not sufficient for targeting mitochondria (Figure 4). This implies that additional amino acids from the C-terminal part are necessary for targeting mitochondria. In proteins targeted to both chloroplasts and mitochondria, the determinants for dual targeting remain unclear, and there is no general rule about how targeting signals are distributed within the proteins (Berglund *et al.*, 2009; Carrie and Small 2013). For some proteins, short C-terminal domains are sufficient

to target mitochondria (Li *et al.*, 2010). However, the C-terminus of CTP1 alone remains nucleocytosolic; and is therefore unlikely to carry a mitochondrial targeting sequence.

CTP1, CTP2 and CTP3 display various patterns of accumulation in chloroplasts and mitochondria. Plant proteins evolve to form large families, whose members diversify to target and function in different cell compartments (Rouhier *et al.*, 2006). The same holds true for effectors of plant parasites. For instance, several members of the Avrblb2 family of *Phythophtora infestans* accumulate in the nuclei, the cytosol or at the plasma membrane (Bozkurt *et al.*, 2011; Zheng *et al.*, 2014). Also, effector alleles can target different compartments (Leonelli *et al.*, 2011; Caillaud *et al.*, 2012). Whether effector variants carry similar or different functions in distinct cellular compartments remains to be determined. In plants, members of large protein families often carry similar biochemical activities, although they are performed with substrates or protein partners that are specific to a given compartment (Rouhier *et al.*, 2006).

We have shown that three candidate effectors from *Melampsora* species accumulate in chloroplasts (and in mitochondria in the case of CTP1) in *N. benthamiana* leaf cells. Growing evidence indicates that effectors from unrelated parasites converge towards chloroplasts and mitochondria (Win *et al.*, 2012; Hicks and Galan 2013; Li *et al.*, 2014; Petre *et al.*, 2015). This implies that manipulating chloroplasts and mitochondria is crucial for parasites to colonize host plants. Surprisingly, to date no oomycete effectors have been reported in these organelles, despite several large-scale subcellular localization screens (Schornack *et al.*, 2010; Caillaud *et al.*, 2012; Stam *et al.*, 2013; Win *et al.*, unpub). However, all these studies used N-terminally tagged effectors, which may have impaired the proper translocation of the proteins into the organelles, resulting in false negatives.

The ability of CTPs to enter chloroplasts suggests that they have a function inside this organelle. What could be that function? CTPs display no sequence similarity to known proteins or protein domains, which prevents the prediction of a putative function based on their primary structure. The only similar feature among the CTPs is the four fully conserved cysteine residues within repeated modules. Such patterns of conserved cysteine residues have been observed previously for several families of secreted proteins in *M. larici-populina* (Duplessis *et al.*, 2011; Hacquard *et al.*, 2012). Repeated domains with conserved residues are common within effectors (Kloppholz *et al.*, 2011; Mak *et al.*, 2012; Chou *et al.*, 2011) and can provide conserved folds that promote rapid evolution (Win *et al.*, 2012). The cysteines could have different roles, for instance by being involved in intra- or extra-molecular disulfide bridges or by having a catalytic activity. Importantly, cysteine residues are also conserved in the repeats that overlap with transit peptides. It is tempting to speculate that repeats could be independently functional, and that the ones at the N-terminus may function as targeting

sequences in addition to other activities. Importantly, CTPs are highly polymorphic and accumulate in various locations, which suggests that they may have different activities. One way to proceed further in elucidating CTP1's function will be to evaluate the relevance of plant protein interactors identified previously (Petre *et al.*, 2015).

EXPERIMENTAL PROCEDURES

Biological material and growth conditions

Escherichia coli (Subcloning Efficiency DH5 α Competent Cells; Invitrogen, Carlsbad, California, USA) and *Agrobacterium tumefaciens* (electrocompetent strain GV3101) were conserved at -80°C and grown in L medium at 37°C and 28°C, respectively. *N. benthamiana* plants were grown in greenhouses at 22°C under 16/8h day/night conditions. The poplar hybrid Beaupré (*P. trichocarpa* x *Populus deltoides*) was propagated in greenhouse as previously described (Rinaldi *et al.*, 2007). Urediniospores of *Melampsora larici-populina* (isolate 98AG31) were conserved at -80°C and propagated as previously described (Rinaldi *et al.*, 2007). Beaupré leaf infections by *M. larici-populina* (isolate 98AG31, virulent on Beaupré) as well as *M. larici-populina* urediniospores germination *in vitro* were performed as previously described (Rinaldi *et al.*, 2007).

Sequence analysis

Protein sequences were retrieved as follow: *M. larici-populina* (http://genomeportal.jgi-psf.org.html ; Duplessis *et al.*, 2011)*, M. lini* (Nemri *et al*, 2014)*, Puccinia striiformis* f sp *tritici* and *Puccinia graminis* f sp *tritici* (Cantu *et al*, 2013), *N. benthamiana* (http://solgenomics.net/ ; Bombarely *et al.*, 2012) *and P. trichocarpa* (http://www.phytozome.net/poplar ; Tuskan *et al.*, 2006). Protein sequence analyses were performed with ClustalX and Jalview softwares. Protein targeting sequences were predicted with SignalP 3.0 (http://www.cbs.dtu.dk/services/SignalP-3.0/), TargetP 1.1 (http://www.cbs.dtu.dk/services/TargetP/), ChloroP 1.1 (http://www.cbs.dtu.dk/services/ChloroP/) and Wolf P Sort (http://psort.hgc.jp/). Protein parameters were calculated with Protparam (http://web.expasy.org/protparam/). Protein secondary structures were predicted with Scratch Protein Predictor (http://www.ics.uci.edu/~baldig/scratch/index.html).

Cloning procedures and plasmids

The open reading frame (ORF) encoding $CTP1_{28\text{-}171}$ and $CTP2_{29\text{-}185}$ were amplified by polymerase chain reaction (PCR) using cDNA from Beaupré leaves infected by *M. larici-populina* (isolate 98AG31; Petre *et al.*, 2012), with primers designed to generate BbsI

AATG/ggTTCG compatible overhangs. The ORF encoding $Mlp70965_{20\text{-}211}$, $Mlp123218_{20\text{-}209}$, and $Mlp123438_{21\text{-}179}$ were obtained following a similar procedure in the course of a different project. The ORF of $CTP3_{28\text{-}260}$ was obtained through gene synthesis (Genewiz, London, UK), with codon optimization for plant expression and removal of internal BbsI and BsaI restriction sites. Sequences suitable for N-terminal tagging were generated by PCR amplifications with primers designed to insert BbsI AATG/GGCT compatible overhangs. Truncated sequences of CTP1 suitable for C-terminal tagging were generated by PCR amplifications with primers designed to insert BbsI AATG/ggTTCG compatible overhangs. DNA fragments were inserted into golden gate level 0 vectors pICSL01005 (BbsI compatible overhangs AATG/ggTTCG) or pICH41308 (BbsI compatible overhangs AATG/GGCT) following a digestion/ligation procedure (Weber *et al.*, 2011; Engler *et al.*, 2013; http://synbio.tsl.ac.uk/). Sequences of DNA fragments obtained by PCR were verified by sequencing. *To generate C-terminally tagged proteins, the ORF of the protein of interest* (BsaI compatible overhangs AATG/ggTTCG) *was assembled* with the ORF of a mCherry or a GFP (BsaI compatible overhangs ggTTCG/GCTT) into the golden gate level 1 binary vector pICH86988 (BsaI compatible overhangs AATG/GCTT; 35S promoter/OCS terminator). To generate N-terminally tagged proteins, *the ORF of the protein of interest* (BsaI compatible overhangs AATG/GCTT) *was* assembled with a 35S promoter (BsaI compatible overhangs GGAG/CCAT), the ORF of a mCherry or a GFP (BsaI compatible overhangs CCAT/AATG) and an octopine synthase (OCS) terminator (BsaI compatible overhangs GCTT/CGCT) into the golden gate level 1 vector pICH47742 (BsaI compatible overhangs GGAG/CGCT). Fusion protein sequences are shown in Table S1. Plasmids were multiplied using *E. coli*. Level 1 vectors were inserted in *A. tumefaciens* strain GV3101, and transformed bacteria were conserved at -80°C in 20% glycerol.

Expression of proteins in *N. benthamiana*

A. tumefaciens strain GV3101 was used to deliver T-DNA constructs into leaf cells of three-week-old *N. benthamiana* plants, following the agroinfiltration method previously described (Win *et al.*, 2011). Briefly, overnight-grown bacterial cultures were adjusted at an OD_{600} of 0.2 into an infiltration buffer (10 mM $MgCl_2$, 150 µM acetosyringone) one hour before leaf infiltration. For all co-transformations, *A. tumefaciens* strains were mixed in a 1:1 ratio in infiltration buffer to a final OD_{600} of 0.2. The leaves were collected two days after infiltration for further cell fractionation, protein isolation or microscopy. Stable *N. benthamiana* transgenics were obtained by *A. tumefaciens* mediated transformation (same strain as used for agroinfiltration), selected on Kanamycin and genotyped by PCR.

Chloroplasts isolation

Three agrotransformed *N. benthamiana* leaves were collected two days after infiltration, cut into 1 cm^2 pieces and incubated in 50 ml cold Isolation Buffer (IB; 400 mM Sorbitol, 50 mM HEPES KOH 8M, 2 mM EDTA) for 15 min. Leaf pieces were homogenized with a PT 1200 C Polytron homogenizer (3 x 5 sec full speed), and the lysate was immediately filtrated onto doubled Miracloth to remove cell debris. The filtrate was collected in 50 mL conical tubes and centrifuged at 3,000 rpm for 5 min at 4°C. The pellet containing the organelles was resuspended in 1 mL of cold IB and chloroplasts were isolated by centrifugation on a Percoll gradient (2 mL 80 % v/v Percoll/IB; 6 mL 40 % v/v Percoll/IB in a 15 mL conical tube) at 4,500 rpm for 10 min at 4°C. The lower green band corresponding to intact chloroplasts was collected by pipetting and immediately used for further experiments.

Live-cell imaging by laser-scanning confocal microscopy

Small pieces of leaves (abaxial surface towards the objective) or solutions of isolated chloroplasts were mounted in water between a slide and a coverslip and immediately observed. Live-cell imaging was performed with a Leica DM6000B/TCS SP5 laser-scanning confocal microscope (Leica microsystems, Bucks, UK), using 10x (air) and 63x (water immersion) objectives. The GFP and the chlorophyll were excited at 488nm, the YFP was excited at 514 nm, and the mCherry was excited at 561 nm. Specific emission signals corresponding to the GFP, the YFP, the mCherry and the chlorophyll were collected between 505-525 nm, 530-550 nm, 580-620 nm and 680-710 nm, respectively, except where otherwise stated. Scanning was performed in sequential mode when needed. We selected leaf areas with cells showing low to moderate levels of fluorescence intensity. For co-localization assays, we imaged clusters of leaf cells showing different levels of fusion protein accumulation to demonstrate the robustness of the observations. Image analysis was performed with Fiji (http://fiji.sc/Fiji).

Protein isolation and western blotting experiments

N. Benthamiana leaves were harvested two days after infiltration, frozen in liquid nitrogen and ground into powder with mortar and pestle. Total protein extraction was performed by reducing and denaturing proteins form the leaf powder 10 minutes at 95°C in Laemmli buffer (1 M pH 6.8 Tris-HCL, 10 mM dithiothreitol [DTT], 2% SDS, 20% glycerol) in order to avoid *in vitro* non-specific degradation of the fusion proteins. Then, five to twenty microliters of isolated proteins were separated by 15% SDS-PAGE, and protein content was estimated by Coomassie blue staining. Western blotting experiments were performed as previously described (Bozkurt *et al.*, 2014). The following antibodies were used: rabbit anti-GFP (Invitrogen), goat anti-rabbit IRW800 (LI-COR Biosciences) and GFP (B2): sc-9996

(Santa-Cruz Biotechnology). Membrane revelation was carried out on an Odyssey infrared imager (LI-COR Biosciences) or on an ImageQuant LAS 4000 luminescent imager (GE Healthcare Life Sciences).

Coimmunoprecipitation and liquid chromatography - tandem mass spectrometry

Coimmunoprecipitation and liquid chromatography / tandem mass spectrometry (LC-MS/MS) analyses were performed as in Petre *et al*., 2015, using GFP_Trap_A beads (Chromotek, Munich, Germany), a hybrid mass spectrometer LTQ-Orbitrap XL (ThermoFisher Scientific) and a nanoflow-UHPLC system (nanoAcquity, Waters Corp.).

ACKNOWLEDGMENTS

We thank Y. Dagdas (TSL, Norwich, UK) and T. Bozkurt (Imperial College, London, UK) for critical reading of an early version of that manuscript. We thank P. Dodds (CSIRO, Canberra, Australia) for sharing protein sequences and discussions. B. Petre was supported by an INRA Contrat Jeune Scientifique (CJS), by the European Union, in the framework of the Marie-Curie FP7 COFUND People Programme, through the award of an AgreenSkills' fellowship (under grant agreement n° 267196) and by the laboratory of excellence (Labex) ARBRE, through the award of a mobility grant (12RW53). D. Saunders was supported by Leverhulme early career fellowship and a fellowship in computational biology at TGAC, in partnership with the John Innes Centre, and strategically supported by BBSRC. C. Lorrain is supported by an INRA CJS and by the Labex ARBRE. S. Duplessis is supported by the French National Research Agency through the Labex ARBRE (ANR-12-LABXARBRE-01) and the Young Scientist Grant POPRUST (ANR-2010-JCJC-1709-01). Research at TSL is supported by the Gatsby Charitable Foundation and the BBSRC.

REFERENCES

Bhattacharjee, S., Halane, M.K., Kim, S.H., and Gassmann, W. (2011) Pathogen effectors target Arabidopsis EDS1 and alter its interactions with immune regulators. *Science* 334:1405-1408

Bombarely, A., Rosli, H.G., Vrebalov, J., Moffett, P., Mueller, L.A., and Martin, G.B. (2012) A draft genome sequence of *Nicotiana benthamiana* to enhance molecular plant-microbe biology research. *Mol Plant Microbe Interact* 25:1523-1530

Bozkurt, T.O., Richardson, A., Dagdas, Y.F., Mongrand, S., Kamoun, S., and Raffaele, S. (2014) The Plant Membrane-Associated REMORIN1.3 Accumulates in Discrete Perihaustorial Domains and Enhances Susceptibility to *Phytophthora infestans*. *Plant Physiol* 165:1005-1018

Bozkurt, T.O., Schornack, S., Win, J., Shindo, T., Ilyas, M., Oliva, R., *et al* (2011) *Phytophthora infestans* effector AVRblb2 prevents secretion of a plant immune protease at the haustorial interface. *Proc Natl Acad Sci U S A* 108:20832-20837

Bruce, B.D. (2001) The paradox of plastid transit peptides: conservation of function despite divergence in primary structure. *Biochim Biophys Acta* 1541:2-21

Berglund, A-K., Pujol, C., Duchene, A-M., and Glaser, E. (2009) Defining the determinants for dual targeting of amino acyl-tRNA synthetases to mitochondria and chloroplasts. *J Mol Biol* 393:803-814

Caillaud, M.C., Piquerez, S.J., Fabro, G., Steinbrenner, J., Ishaque, N., Beynon, J., and Jones, J.D. (2012) Subcellular localization of the Hpa RxLR effector repertoire identifies a tonoplast-associated protein HaRxL17 that confers enhanced plant susceptibility. *Plant J* 69:252-265

Cantu, D., Segovia, V., MacLean, D., Bayles, R., Chen, X., Kamoun, S., *et al* (2013) Genome analyses of the wheat yellow (stripe) rust pathogen *Puccinia striiformis* f. sp. *tritici* reveal polymorphic and haustorial expressed secreted proteins as candidate effectors. *BMC Genomics* 14:270

Carrie, C., and Whelan, J. (2013) Widespread dual targeting of proteins in land plants: when, where, how and why. *Plant Signal Behav* 8:8

Carrie, C., and Small, I. (2013) A reevaluation of dual-targeting of proteins to mitochondria and chloroplasts. *Biochim Biophys Acta* 1833:253-259

Carrie, C., Kühn, K., Murcha, M.W., Duncan, O., Small, I.D., O'Toole, N., and Whelan, J. (2009) Approaches to defining dual-targeted proteins in Arabidopsis. *Plant J* 57:1128-1139

Chou, S., Krasileva, K.V., Holton, J.M., Steinbrenner, A.D., Alber, T., and Staskawicz, B.J. (2011) *Hyaloperonospora arabidopsidis* ATR1 effector is a repeat protein with distributed recognition surfaces. *Proc Natl Acad Sci U S A* 108:13323-13328

Demarsy, E., Lakshmanan, A.M., and Kessler, F. (2014) Border control: selectivity of chloroplast protein import and regulation at the TOC-complex. *Front Plant Sci* 5:483

Deslandes, L., and Rivas, S. (2011) Catch me if you can: bacterial effectors and plant targets. *Trends Plant Sci* 17:644-655

Dodds, P.N., and Rathjen, J.P. (2010) Plant immunity: towards an integrated view of plant-pathogen interactions. *Nat Rev Genet* 11:539-548

Duplessis, S., Bakkeren, G., and Hamelin, R. (2014) Advancing Knowledge on Biology of Rust Fungi Through Genomics. *Adv Bot Res* 70:173-209

Duplessis, S., Cuomo, C.A., Lin, Y.C., Aerts, A., Tisserant, E., Veneault-Fourrey, C., *et al* (2011) Obligate biotrophy features unraveled by the genomic analysis of rust fungi. *Proc Natl Acad Sci U S A* 108:9166-9171

Engler, C., and Marillonnet, S. (2013) Combinatorial DNA assembly using Golden Gate cloning. *Methods Mol Biol* 1073:141-156

Gérard, P.R., Husson, C., Pinon, J., and Frey, P. (2006) Comparison of Genetic and Virulence Diversity of *Melampsora larici-populina* Populations on Wild and Cultivated Poplar and Influence of the Alternate Host. *Phytopathology* 96:1027-1036

Hacquard, S., Joly, D.L., Lin, Y.C., Tisserant, E., Feau, N., Delaruelle, C., *et al* (2012) A comprehensive analysis of genes encoding small secreted proteins identifies candidate effectors in *Melampsora larici-populina* (poplar leaf rust). *Mol Plant Microbe Interact* 25:279-293

Heidrich, K., Wirthmueller, L., Tasset, C., Pouzet, C., Deslandes, L., and Parker, J.E. (2011) Arabidopsis EDS1 connects pathogen effector recognition to cell compartment-specific immune responses. *Science* 334:1401-1404

Hicks, S.W., and Galán J.E. (2013) Exploitation of eukaryotic subcellular targeting mechanisms by bacterial effectors. *Nat Rev Microbiol* 11:316-326

Jelenska, J., Yao, N., Vinatzer, B.A., Wright, C.M., Brodsky, J.L., and Greenberg, J.T. (2007) A J domain virulence effector of *Pseudomonas syringae* remodels host chloroplasts and suppresses defenses. *Curr Biol* 17:499-508

Kemen, E., Kemen, A., Ehlers, A., Voegele, R., and Mendgen, K. (2013) A novel structural effector from rust fungi is capable of fibril formation. *Plant J* 75:767-780

Kemen, E., Kemen, A.C., Rafiqi, M., Hempel, U., Mendgen, K., Hahn, M., and Voegele, R.T. (2005) Identification of a protein from rust fungi transferred from haustoria into infected plant cells. *Mol Plant Microbe Interact* 18:1130-1139

Kloppholz, S., Kuhn, H., and Requena, N. (2011) A secreted fungal effector of *Glomus intraradices* promotes symbiotic biotrophy. *Curr Biol* 21:1204-1209

Krenz, B., Jeske, H., and Kleinow, T. (2012) The induction of stromule formation by a plant DNA-virus in epidermal leaf tissues suggests a novel intra- and intercellular macromolecular trafficking route. *Front Plant Sci* 3:291

Leonelli, L., Pelton, J., Schoeffler, A., Dahlbeck, D., Berger, J., Wemmer, D.E., and Staskawicz, B. (2011) Structural elucidation and functional characterization of the *Hyaloperonospora arabidopsidis* effector protein ATR13. *PLoS Pathog* 7:e1002428

Li, G., Froehlich, J.E., Elowsky, C., Msanne, J., Ostosh, A.C., Zhang, C., *et al* (2014) Distinct Pseudomonas type-III effectors use a cleavable transit peptide to target chloroplasts. *Plant J* 77:310-321

Lindeberg, M., Cunnac, S., and Collmer, A. (2012) *Pseudomonas syringae* type III effector repertoires: last words in endless arguments. *Trends Microbiol* 20:199-208

Lorrain, C., Kamoun, S., and Petre, B. (2014a) *Nicotiana benthamiana* plastidial Methionine Sulfoxide Reductase A (NbpMSRA) localizes in the stroma of chloroplasts. figshare. http://dx.doi.org/10.6084/m9.figshare.1035057

Lorrain, C., Kamoun, S., and Petre, B. (2014b) *Populus trichocarpa* Coproporphyrinogen III oxidase (PtCPO) localizes in sub-chloroplastic structures. figshare. http://dx.doi.org/10.6084/m9.figshare.1035063

Mak, A.N., Bradley, P., Cernadas, R.A., Bogdanove, A.J., and Stoddard, B.L. (2012) The crystal structure of TAL effector PthXo1 bound to its DNA target. *Science* 335:716-719

Nelson, B.K., Cai, X., and Nebenführ, A. (2007) A multicolored set of in vivo organelle markers for co-localization studies in Arabidopsis and other plants. *Plant J* 51:1126-1136

Nemri, A., Saunders, D.G., Anderson, C., Upadhyaya, N.M., Win, J., Lawrence, G.J., *et al* (2014) The genome sequence and effector complement of the flax rust pathogen *Melampsora lini*. *Front Plant Sci* 5:98

Pennisi, E. (2010) Armed and dangerous. *Science* 327:804-805

Petre, B., Saunders, D.G., Sklenar, J., Lorrain, C., Win, J., Duplessis, S., and Kamoun, S. (2015) Candidate Effector Proteins of the Rust Pathogen *Melampsora Larici-Populina* Target Diverse Plant Cell Compartments. *Mol Plant Microbe Interact* 28:689-700

Petre, B., and Kamoun, S. (2014) How do filamentous pathogens deliver effector proteins into plant cells? *PLoS Biol* 12:e1001801

Petre, B., Morin, E., Tisserant, E., Hacquard, S., Da Silva, C., Poulain, J., *et al* (2012) RNA-Seq of early-infected poplar leaves by the rust pathogen *Melampsora larici-populina* uncovers PtSultr3;5, a fungal-induced host sulfate transporter. *PLoS One* 7:e44408

Rinaldi, C., Kohler, A., Frey, P., Duchaussoy, F., Ningre, N., Couloux, A., *et al* (2007) Transcript profiling of poplar leaves upon infection with compatible and incompatible strains of the foliar rust *Melampsora larici-populina*. *Plant Physiol* 144:347-366

Rodríguez-Herva, J.J., González-Melendi, P., Cuartas-Lanza, R., Antúnez-Lamas, M., Río-Alvarez, I., Li, Z., *et al* (2012) A bacterial cysteine protease effector protein interferes with photosynthesis to suppress plant innate immune responses. *Cell Microbiol* 14:669-681

Rouhier, N., Vieira Dos Santos, C., Tarrago, L., and Rey, P. (2006) Plant methionine sulfoxide reductase A and B multigenic families. *Photosynth Res* 89:247-262

Saunders, D.G., Win, J., Cano, L.M., Szabo, L.J., Kamoun, S., and Raffaele, S. (2012) Using hierarchical clustering of secreted protein families to classify and rank candidate effectors of rust fungi. *PLoS One* 7:e29847

Schornack, S., van Damme, M., Bozkurt, T.O., Cano, L.M., Smoker, M., Thines, M., *et al* (2010) Ancient class of translocated oomycete effectors targets the host nucleus. *Proc Natl Acad Sci U S A* 107:17421-17426

Shi, L.X., and Theg, S.M. (2013) The chloroplast protein import system: from algae to trees. *Biochim Biophys Acta* 1833:314-331

Sohn, K.H., Hughes, R.K., Piquerez, S.J., Jones, J.D., and Banfield, M.J. (2012) Distinct regions of the *Pseudomonas syringae* coiled-coil effector AvrRps4 are required for activation of immunity. *Proc Natl Acad Sci U S A* 109:16371-16376

Stam, R., Jupe., J, Howden, A.J., Morris, J.A., Boevink, P.C., Hedley, P.E., Huitema, E. (2013) Identification and Characterisation CRN Effectors in *Phytophthora capsici* Shows Modularity and Functional Diversity. *PLoS One* 8:e59517

Teixeira, P.F., and Glaser, E. (2013) Processing peptidases in mitochondria and chloroplasts. *Biochim Biophys Acta* 1833:360-370

Tonkin, C.J., Foth, B.J., Ralph, S.A., Struck, N., Cowman, A.F., and McFadden, G.I. (2008) Evolution of malaria parasite plastid targeting sequences. *Proc Natl Acad Sci U S A* 105:4781-4785

Torres Zabala, M., Littlejohn, G., Jayamaran, S., Studholme, D., Bailey, T., Lawson, T., *et al.* (2015) Chloroplasts play a central role in plant defence and are targeted by pathogen effectors. *Nat Plants* 1:15074

Tuskan, G.A., Difazio, S., Jansson, S., Bohlmann, J., Grigoriev, I., Hellsten, U., *et al* (2006) The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). *Science* 313:1596-1604

Wang, Y., Yu, B., Zhao, J., Guo, J., Li, Y., Han, S., *et al* (2013) Autophagy contributes to leaf starch degradation. *Plant Cell* 25:1383-1399

Weber, E., Gruetzner, R., Werner, S., Engler, C., and Marillonnet, S. (2011) Assembly of designer TAL effectors by Golden Gate cloning. *PLoS One* 6:e19722

Win, J., Chaparro-Garcia, A., Belhaj, K., Saunders, D.G., Yoshida, K., Dong., S, *et al* (2012) Effector biology of plant-associated organisms: concepts and perspectives. *Cold Spring Harb Symp Quant Biol* 77:235-247

Win, J., Kamoun, S., and Jones, A.M. (2011) Purification of effector-target protein complexes via transient expression in *Nicotiana benthamiana*. *Methods Mol Biol* 712:181-194

Wirthmueller, L., Roth, C., Fabro, G., Caillaud, M.C., Rallapalli, G., Asai, S., *et al* (2015) Probing formation of cargo/importin-α transport complexes in plant cells using a pathogen effector. *Plant J* 81:40-52

Zheng, X., McLellan, H., Fraiture, M., Liu, X., Boevink, P.C., Gilroy, E.M., *et al* (2014) Functionally redundant RXLR effectors from *Phytophthora infestans* act at different steps to suppress early flg22-triggered immunity. *PLoS Pathog* 10:e1004057

SUPPLEMENTARY FIGURES :

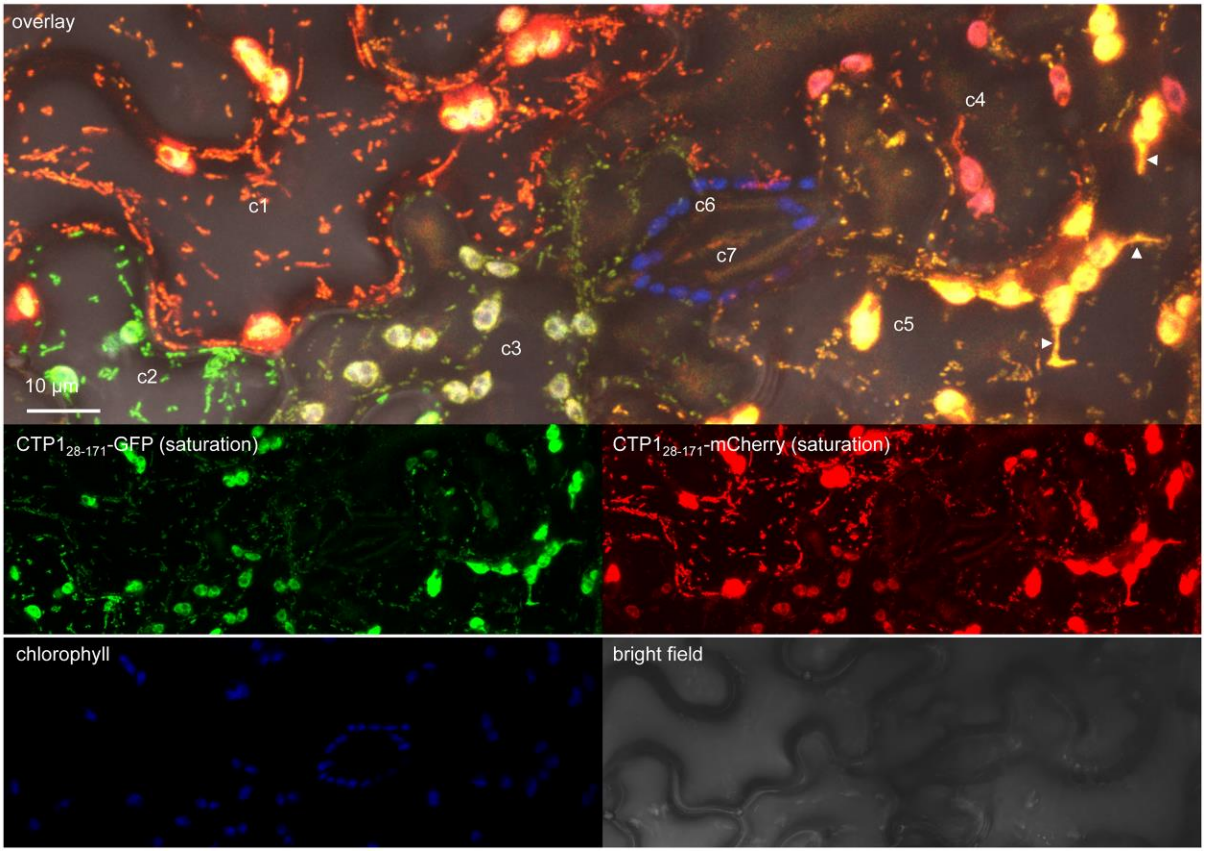


Figure S1: CTP1$_{28-171}$-fluorescent protein fusions show no background in the nucleocytosol Same images as Figure 1A, but with saturated GFP and mCherry fluorescent signals. See Figure 1A for details.

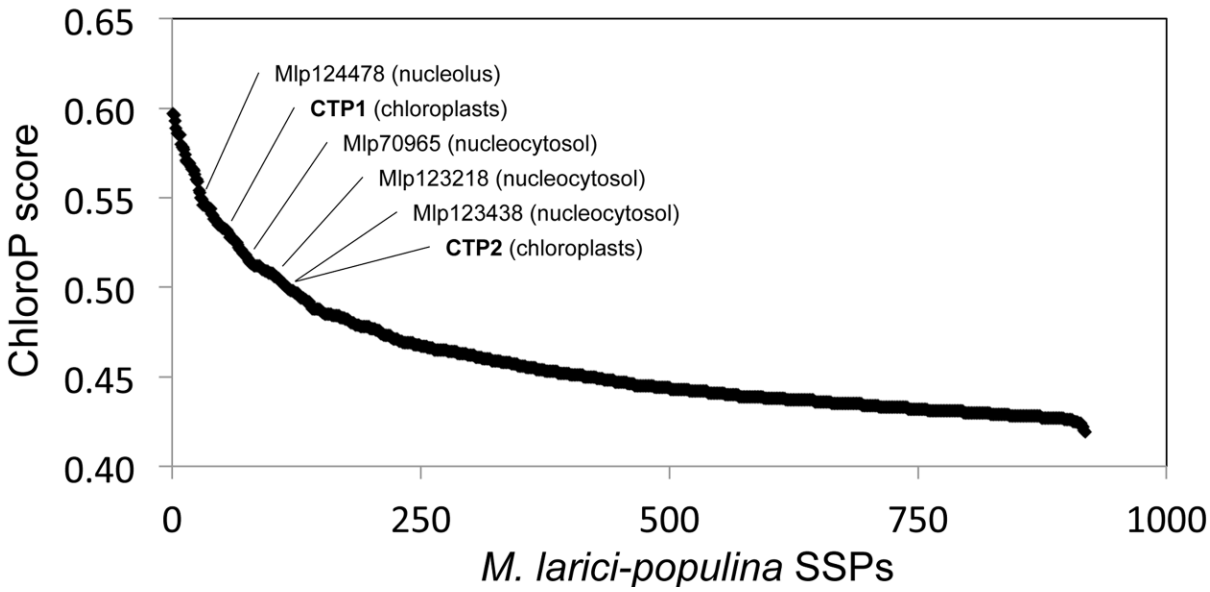


Figure S2. Several *M. larici-populina* candidate effectors carry a predicted chloroplast transit peptide downstream of their signal peptide

The ChloroP program was used to predict chloroplast transit peptides over a set of 918 predicted small-secreted proteins (SSPs) of *M. larici-populina* lacking their signal peptide. The resulting ChloroP scores are displayed, sorted in descending order from left to right. Selected proteins with a score ≥ 0.5 and for which the subcellular localization in leaf cells was experimentally determined are shown on the graph.

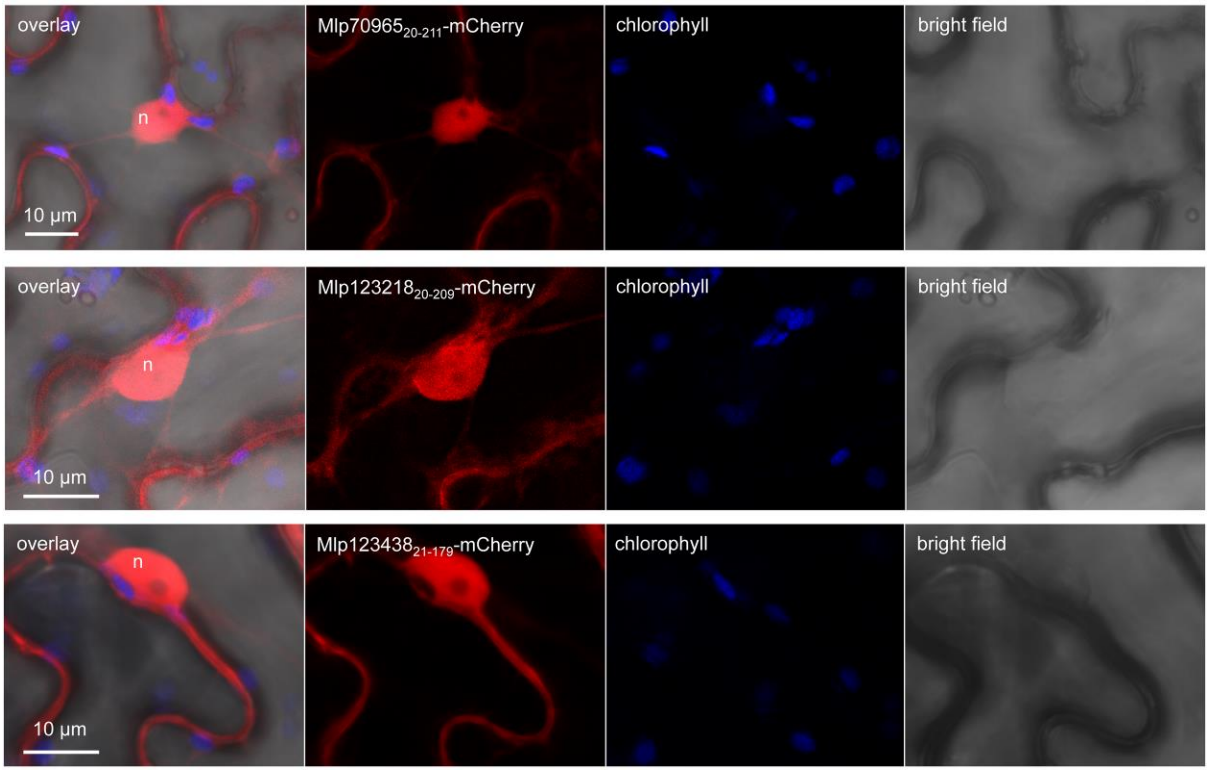


Figure S3. Mlp70965, Mlp123218 and Mlp123438 do not accumulate in chloroplasts
Live-cell imaging of Mlp70965$_{20-211}$-mCherry, Mlp123218$_{20-209}$-mCherry, and Mlp123438$_{21-179}$-mCherry in *N. benthamiana* leaf pavement cells. Images show a single optical section. Proteins were expressed in *N. benthamiana* leaf cells by agroinfiltration. Live-cell imaging was performed with a laser-scanning confocal microscope two days after infiltration. The mCherry and the chlorophyll were excited at 561 nm. mCherry (red) and chlorophyll (blue) fluorescent signals were collected at 580-620 nm and 680-700 nm, respectively. n: nucleus.

Partie 4 : Informations complémentaires

1. Matériel et Méthodes

Cultures des plantes et microorganismes

Les plants d'*Arabidopsis thaliana* Col-0 utilisés ont été cultivés en terre dans des chambres de cultures à 23°C, 60% d'humidité et avec une photopériode de 14h d'éclairage et 10h d'obscurité (135 $\mu mol.m^{-2}.s^{-1}$). Les cultures *in vitro* d'Arabidopsis et de *Populus tremula x Populus alba* 717-1B4 ont été réalisées sur milieu Murashige-Skoog dilué de moitié (½ MS) contenant 0,7% d'agar. Pour la sélection des plants transgéniques, de la kanamycine (100 $mg.L^{-1}$) a été ajoutée au milieu ½ MS. Les souches d'*Escherichia coli* (souche DH5α) et d'*Agrobacterium tumefaciens* (souche électrocompétente GV3101 et C58) ont été conservée à -80°C et cultivée sur milieu LB avec kanamycine (50 $mg.L^{-1}$) à 28°C. Les autres cultures réalisées pour la production de plantes transgéniques sont décrites plus loin.

Clonage et Plasmides

La séquence codante de $MlpCTP1_{28-171}$-GFP a été générée par Petre *et al*. (2016), insérée dans le vecteur Golden Gate pICH86988. Le plasmide a été multiplié dans la souche DH5α d'*E.coli* et inséré dans la souche d'*A. tumefaciens* C58. Les bactéries transformées ont été conservées à -80°C dans 20% de glycérol. Pour la transformation d'Arabidopsis, la séquence codante de $MlpCTP1_{28-171}$-GFP a été insérée dans le vecteur GATEWAY pK7WGF2 (Karimi *et al.*, 2002) pour permettre l'expression de la protéine fusion $MLPCTP1_{28-171}$-GFP. La séquence codante $MlpCTP1_{28-171}$-GFP a été placée sous contrôle du promoteur 35S du virus de la mosaïque de chou-fleur (CaMV35S) et d'un terminateur. La séquence codante de gène de résistance à la kanamycine contrôlée par le promoteur de la nopaline synthase (nos) a été placé en amont de $MlpCTP1_{28-171}$-GFP pour la sélection des lignées transgéniques.

Production d'Arabidopsis transgéniques

Le vecteur pK7WGF2::$MLPCTP1_{28-171}$-GFP a été transféré dans la souche électrocompétente *A. tumefaciens* C58. Les plants d'Arabidopsis Col-0 ont été transformés par la méthode de trempage des fleurs (Mireault *et al.,* 2014). Brièvement, les fleurs d'Arabidopsis Col-0 de plants de 6 semaines ont été trempées dans une suspension bactérienne d'*A. tumefaciens* C58 contenant le vecteur pK7WGF2 :: $MLPCTP1_{28-171}$-GFP. La suspension bactérienne d'une densité optique de 0,6 à 1, est réalisée dans le milieu suivant: 5% (m/v) de saccarose ; 50mM d'acétosyringone ; 0,05% (v/v) Sylgard 309. Les graines transformées ont été sélectionnées sur du milieu ½ MS contenant de la kanamycine ($100mg.L^{-1}$) et les plants régénérés ont été criblés pour l'expression de la fusion $MLPCTP1_{28-171}$-GFP par microscopie confocale. Les lignées T2 montrant une ségrégation 3:1 resistants/sensibles à la kanamycine ont été

sélectionnées comme lignées avec une seule insertion. Les lignées T3 avec 100% de résistants ont été sélectionnées pour l'insertion homozygote de MLPCTP1$_{28-171}$-GFP.

Production de peupliers *Populus tremula x Populus alba* 717-1B4 transgéniques

La souche d'*A. tumefaciens* C58 a été utilisée pour transférer la construction pICH86988::MLPCTP1$_{28-171}$-GFP dans les entre-nœuds de peupliers *P.tremula x P. alba* 717-1B4. La transformation des peupliers a été réalisée selon le protocole de Leplé et al. (Leple *et al.*, 1992). Des entre-nœuds de 5mm à 10mm ont été prélevés sur des peupliers de 6 semaines cultivés *in vitro* sur ½ MS, fendus dans la longueur et placés sur milieu M1 consistant en un milieu MS additionné de vitamines 1X (L-cysteine 1mg.L^{-1} ; acide nicotinique 1mg.L^{-1} ; pyridoxine 1mg.L^{-1} ; thiamine 1mg.L^{-1} ; pantothenate de calcium 1mg.L^{-1} ; biotine 1mg.L^{-1}); L-glumatine 200mg.L^{-1} ; 2iP 5µM ; ANA 10 µM 48h à 24°C. La souche d'*A. tumefaciens* C58 est étalée en stries denses sur milieu MYA (0,5% (m/v) extrait de levure, 0,05% (m/v) hydrolysat de caséine ; 0,5% (m/v) NaCl ; 0,8% (m/v) mannitol ; 0,2% $(NH4)_2SO_4$; 1,5% (m/v) agar) avec de la rifampicine (50mg. L^{-1}), de la kanamycine (50mg. L^{-1}) et de la gentamicine (50mg. L^{-1}). Les bactéries sont incubées 48h à 24°C puis resuspendues dans du milieu MS liquide à une densité optique DO_{600nm} de 0,2. Les entre-nœuds de peupliers sont placés dans la suspension bactérienne et incubés sous agitation pendant 16h à l'obscurité, à 24°C. Les entre-nœuds sont ensuite transférés sur milieu M1 pour 48h à l'obscurité, à 24°C. Après rinçage, les entre-nœuds sont transférés sur milieu M2 (milieu MS additionné en vitamines 1X (L-cysteine 1mg.L^{-1} ; acide nicotinique 1mg.L^{-1} ; pyridoxine 1mg.L^{-1} ; thiamine 1mg.L^{-1} ; pantothenate de calcium 1mg.L^{-1} ; biotine 1mg.L^{-1}); L-glumatine 200mg.L^{-1} ; 2iP 5µM ; ANA 10 µM ; 50mg.L^{-1} carbéniciline ; 100mg.L^{-1} céfotaxime) à 24°C pendant 3 semaines à l'obscurité. Les explants sont transférés sur milieu de sélection M2 additionné de kanamycine (50mg.L^{-1}), placés à la lumière à 24°C jusqu'à apparition de cals verts. Les cals verts sont excisés et placés sur milieu M3 (milieu MS additionné en vitamines 1X (L-cysteine 1mg.L^{-1} ; acide nicotinique 1mg.L^{-1} ; pyridoxine 1mg.L^{-1} ; thiamine 1mg.L^{-1} ; pantothenate de calcium 1mg.L^{-1} ; biotine 1mg.L^{-1}); L-glumatine 200mg.L^{-1} ; 2iP 5µM ; ANA 10 µM ; 50mg.L^{-1} carbéniciline ; 100mg.L^{-1} céfotaxime ; thidiazuron 0,25 µM) additionné en kanamycine (50mg.L^{-1}) pendant deux semaines à 24°C. Quand les cals ont doublé de volume, les cals ont été placés sur milieu ½ MS pour permettre l'apparition de bourgeons, leur enracinement et leur élongation. Les peupliers transgéniques exprimant MLPCTP1$_{28-171}$-GFP sont criblés au microscope confocal puis maintenus en culture *in vitro* sur milieu ½ MS ou transférés en terre.

Imagerie confocale *in vivo*

Des fragments de feuilles d'Arabidopsis et de peupliers ont été montés entre lame de lamelle de verre et observés directement. L'imagerie confocale *in vivo* des feuilles d'Arabidopsis a été réalisée avec un microscope confocal laser Leica TC SP8 (Leica, Microsystems, Germany). Les images ont été réalisées avec les objectifs 40x (immersion dans l'eau) et 63x (immersion dans l'huile). L'imagerie confocale *in vivo* des feuilles de peupliers a été réalisée avec un microscope confocal laser Zeiss LSM780 (Carl Zeiss AG, Germany). Les images ont été réalisées avec les objectifs 10x (air), 40x (immersion dans l'eau) et 63x (immersion dans l'huile). La GFP et la chlorophylle ont été excitées à 488nm. Les signaux d'émission de la GFP et de la chlorophylle ont été collectés entre 505-525nm et 680-700nm, respectivement. Les images ont été analysées en utilisant Fiji (http:// fiji.sc/Fiji).

Dichroïsme circulaire

La presence de structures secondaires chez MLP124017 a été anaysée par dichroïsme circulaire (CD)

Circular Dichroism spectropolarimetry

Each purified recombinant protein was analysed by circular dichroism (CD) to estimate protein secondary structure. CD experiments were carried out with 75 µM of recombinant protein in 50 mM sodium phosphate buffer pH 7.0 supplemented with 200 mM NaF using a Chirascan^{TM}-plus Circular Dichroism spectrophotometer (Applied Photophysics).

2. Résultats

2.1 Expression des effecteurs CTP1 et 124017 de *M. larici-populina* dans le peuplier et *A. thaliana*

Nous avons pu montrer que les fusions MLPCTP1-GFP exprimées dans *N. benthamiana* sont transloquées dans les chloroplastes (Petre et al. 2016a ; **Article 5**). Pour tester la spécificité de cette localisation chloroplastique pour l'effecteur candidat MLPCTP1, nous avons généré des transformants stables d'*A. thaliana* exprimant la même fusion MLPCTP1-GFP (cf. Annexe 1 ; **Article 5**). Le signal GFP de MLPCTP1-GFP se superpose au signal d'autofluorescence de la chlorophylle dans les chloroplastes des plants d'*A. thaliana* régénérés (Figure 22). MLPCTP1 cible donc les chloroplastes lorsqu'il est exprimé de façon stable dans le système hétérologue *A. thaliana*.

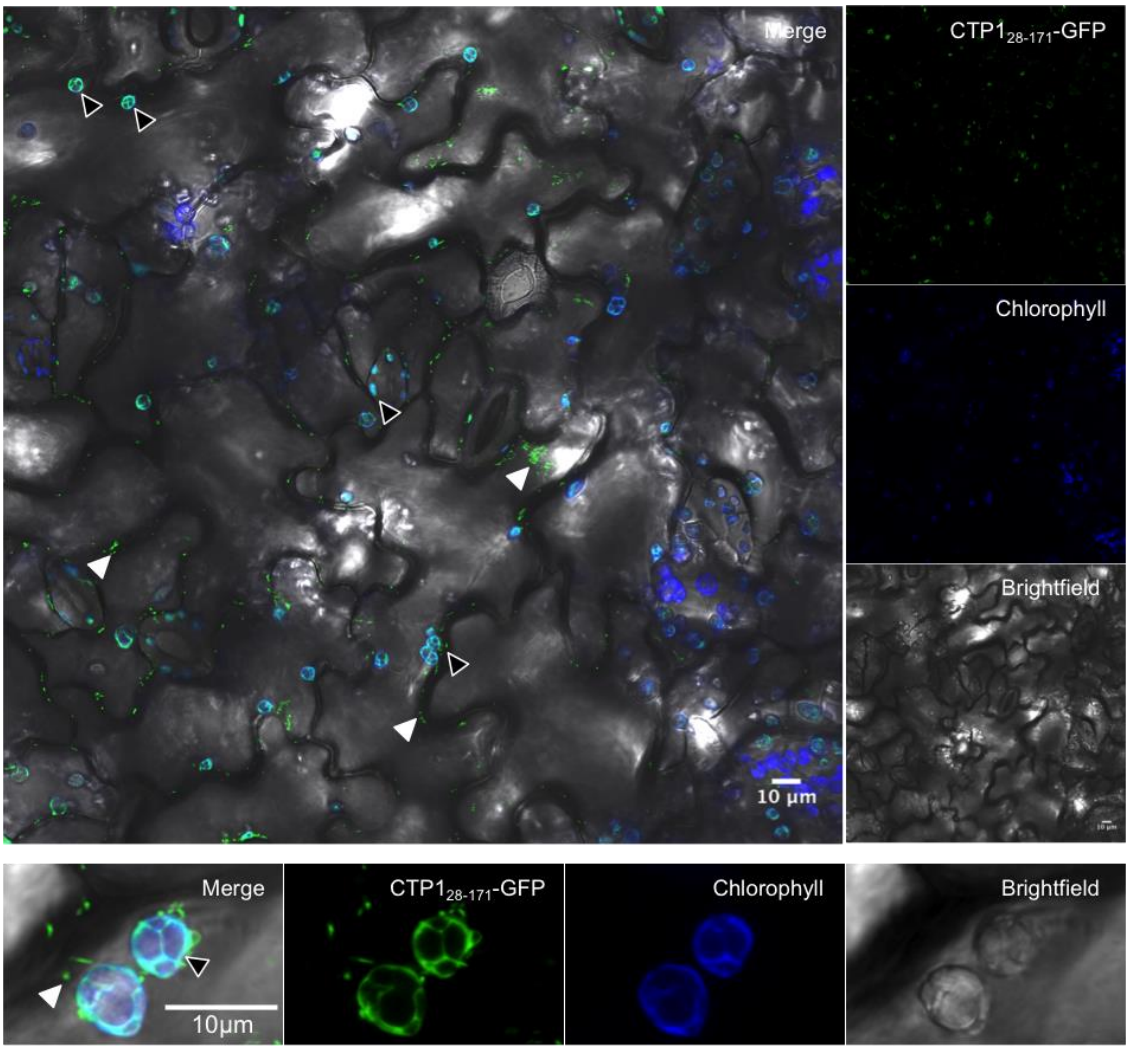


Figure 22 : MLPCTP1-GFP cible les chloroplastes de transformants stables d'Arabidopsis. Imagerie confocale *in vivo* de MLPCTP1-GFP dans les cellules foliaires épidermiques d'*A. thaliana* (génotype Col-0). Le panel du haut montre une section optique de 0,8µm d'une feuille de 717-1B4, grossissement x40 ; le panel du bas montre section optique de 0,8µm, grossissement x60. Les signaux relatifs à la GFP (vert) et chlorophylle (bleu) sont collectés à 508-526 nm et 680-700 nm, respectivement. Les flèches noires indiquent la superposition des signaux GFP et d'autofluorescence de la chlorophylle dans les chloroplastes. Les flèches blanches indiquent le signal GFP dans des corps cytoplasmiques.

Pour confirmer la localisation chloroplastique de l'effecteur candidat MLPCTP1 dans le peuplier, nous avons généré des transformants stables du clone de peuplier *Populus tremula x Populus alba* 717-1B4 exprimant la fusion MLPCTP1-GFP (cf. Annexe 1 ; **Article 5**). Le signal GFP de la fusion MLPCTP1-GFP co-localise avec l'autofluorescence de la chlorophylle dans les feuilles de peupliers régénérés (Figure 23). Le signal de la fusion MLPCTP1-GFP est également détecté dans des corps cytoplasmiques et en bruit de fond dans le cytoplasme. MLPCTP1-GFP est exprimé dans les cellules de l'épiderme et dans les cellules de garde des stomates. La forme mature de MLPCTP1-GFP localise dans le stroma des chloroplastes lorsqu'il est exprimé de façon stable dans les cellules de peuplier.

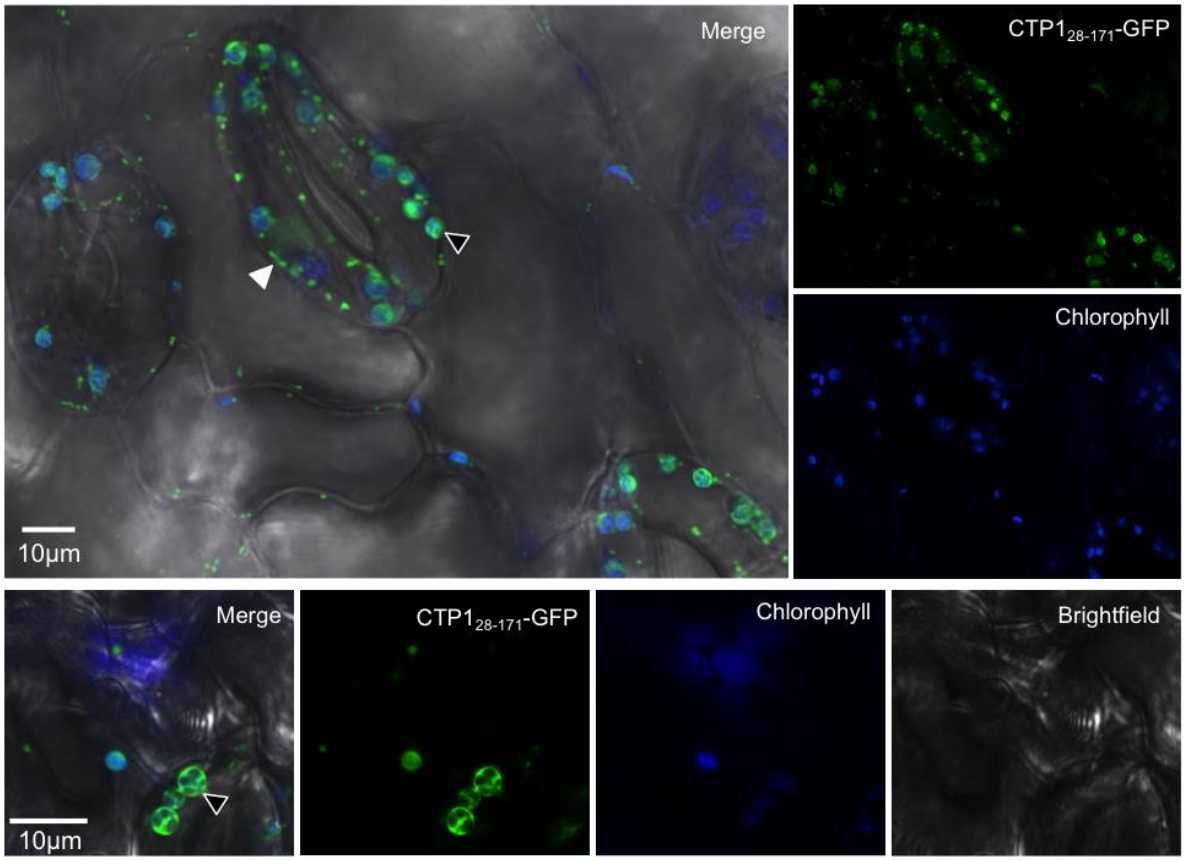


Figure 23 : MLPCTP1-GFP cible les chloroplastes de transformants stables de peupliers 717-1B4. Imagerie confocale *in vivo* de MLPCTP1-GFP dans les cellules foliaires épidermiques de *P. tremula x P. alba* (génotype 717-1B4). Le panel du haut montre une section optique de 0,8µm d'une feuille de 717-1B4, grossissement x40 ; le panel du bas montre section optique de 0,8µm, grossissement x60. Les signaux relatifs à GFP (vert) et chlorophylle (bleu) sont collectés à 508-526 nm et 680-700 nm, respectivement. Les flèches noires indiquent la superposition des signaux GFP et d'autofluorescence de la chlorophylle dans les chloroplastes. Les flèches blanches indiquent le signal GFP dans des corps cytoplasmiques.

J'ai également tenté de générer des plants d'*A. thaliana* et de *P. tremula x P. alba* 717-1B4 transformés avec la construction MLP124017-GFP (sans le peptide signal) utilisée pour l'expression transitoire chez le tabac. Aucun des 20 explants régénérés de peupliers, ni des 32 plants T1 d'Arabidopsis récupérés n'ont montré de signal GFP dans les feuilles.

2.2. Expression de protéines recombinantes pour les effecteurs candidats MLPCTP1 et MLP124017 en vue de leur caractérisation structurale.

La production et la purification de protéines recombinantes des deux effecteurs candidats sont nécessaires pour réaliser leur caractérisation structurale. L'expression de MLP124017-(His)$_6$ dans la fraction soluble d'*E. coli* a permis sa production et sa purification en grande quantité en conditions natives (Figure 24 ; **article n°4**). La production et purification de la protéine recombinante montre que MLP124017 est une protéine monomérique de 18kDa. Une analyse

in sillico a permis de prédire la présence de structures secondaires dans la séquence de MLP124017. Cette prédiction a ensuite été confirmée par dichroïsme circulaire. Le spectre de difraction de MLP124017 décrit la présence d'hélices α et de feuillets β (Figure 24; article 4). Pour poursuivre l'analyse structurale de MLP124017, nous avons dans un premier temps, cherché à obtenir la structure tridimensionnelle de MLP124017 par une approche de cristallographie (en collaboration avec Thomas Roret, Post-doctorant Equipe *Réponse au Stress et Régulation Redox* de l'UMR IAM). En effet, la cristallographie représente une méthode de choix pour l'analyse des structures des protéines inconnues, sous condition d'obtenir des cristaux.

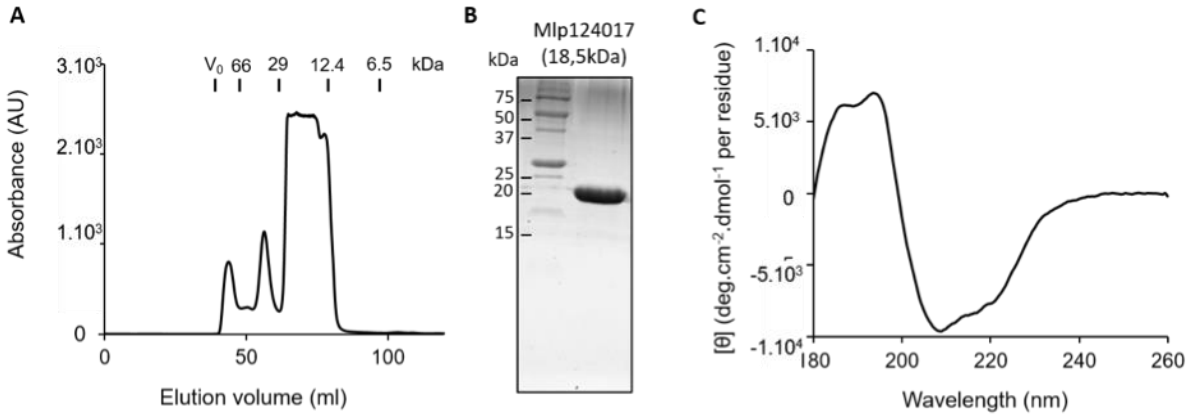


Figure 24 : Production, purification de MLP124017 et analyse des structures secondaires
(A) Profil de MLP124017 en chromatographie d'exclusion (Superdex 75). (B) Gel SDS-PAGE 15% coloré au Bleu de Coomassie de 10µg de MLP124017-$(His)_6$ purifiée par chromatographie d'affinité (résine agarose saturée en ions Ni^{2+}) puis par chromatographie d'exclusion (Superdex 75). (C) Dichroïsme circulaire de MLP124107. Les axes des abscisses et des ordonnées indiquent respectivement les longueurs d'ondes (nm) et l'angle de diffraction par résidu ($deg.cm^{-2}.dmol^{-1}$).

Après plusieurs essais au laboratoire et près de 450 conditions testées, aucun cristal n'a pu être obtenu. Dans le cadre d'une collaboration établie avec Ahmed Haouz (Responsable de la plateforme de Cristallographie de l'Institut Pasteur à Paris), plus de 1500 conditions de cristallisation ont été testées de manière automatisée pour la protéine MLP124017 étiquetée ou non de six résidus histidine. De nouveau, aucun cristal n'a pu être détecté. Finalement, dans le cadre du réseau Effectome, puis d'un projet FRISBI avec l'infrastructure française pour la biologie structurale intégrée (http://frisbi.eu/), j'ai pu établir une collaboration avec Karine de Guillen et André Padilla (UMR 5048 CNRS/INSERM Structure, dynamique et fonction des biomolécules par RMN de Montpellier) pour tenter résoudre la structure de l'effecteur candidat MLP124017 par une approche de RMN. Cette approche a donné d'excellents résultats avec MLP124017 qui ont été présentés dans l'**article n°4**.

Comme cela est présenté dans l'article n°4, l'effecteur candidat MLPCTP1 (MLP107772) est exprimé majoritairement dans la fraction protéique insoluble d'*E. coli* Rosetta2. Nous avons aussi testé plusieurs souches d'*E. coli* (par exemple, les souches BL21 et Origami) sans toutefois réussir à produire MLPCTP1 en conditions natives (ce qui permet de préserver la structuration des protéines recombinantes) en quantité suffisante dans la fraction protéique soluble pour la purifier (Figure 25 ; données non présentées). Nous avons également tenter de produire la protéine recombinante chez la levure *Pichia pastoris* en collaboration avec Fabrice Neiers (Université de Bourgogne) et Loïc briand (DR INRA, Centre des sciences du Goût et de l'alimentation, UMR CNRS 6265, UMR INRA 1334) en testant plusieurs construction avec ou sans étiquette histidine, avec une étiquette CPD (*Self-Cleaving Protease Tag* en anglais, Shen, 2014) favorisant la solubilisation de protéines insolubles ou avec seulement la partie C-terminale de MLPCTP1 sans le peptide d'adressage aux chloroplastes correspondant à la protéine mature. Cependant, aucune bande n'a pu être observée sur gel pour toutes les constructions testées (Figure 25).

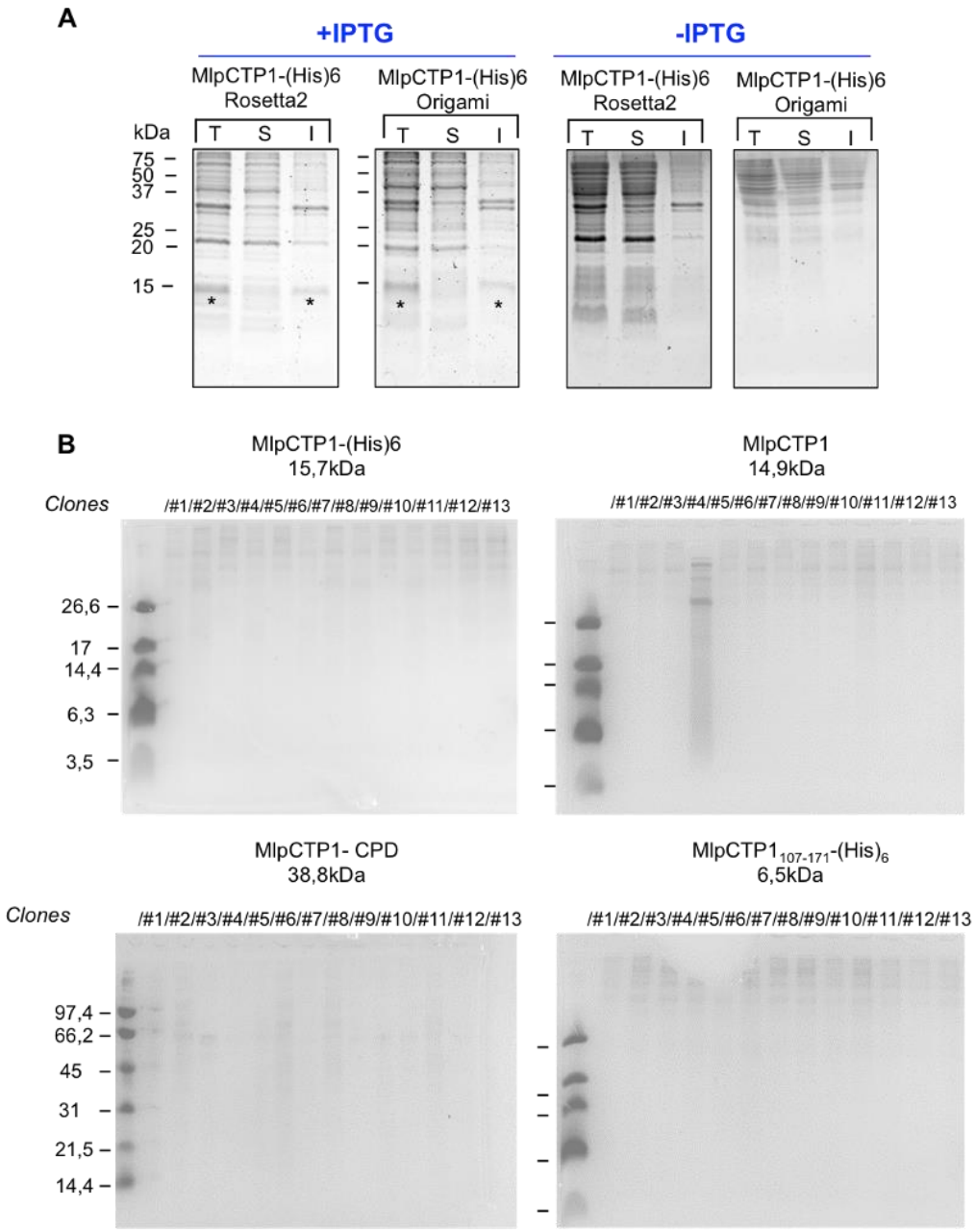


Figure 25 : Expression de MLPCTP1 chez *E. coli* et *Pichia pastoris*.

(A). Gels SDS-PAGE 17% colorés au Bleu de Coomassie après séparation des protéines totales (T), solubles (S) et insolubles (I) de MLPCTP1-$(His)_6$ exprimé chez la bactérie Rosetta2 et Origami, avec (+) ou sans (-) induction à l'isopropyl β-D21-thiogalactopyranoside (1mM ; IPTG). Les astérisques représentent la migration attendue de MLPCTP1-$(His)_6$ surexprimée. (B) Gels SDP-PAGE 16% colorés au Bleu de Coomassie après séparation des protéines surexprimées dans 50µl de surnageant de cultures de *P. pastoris* GS115 exprimant respectivement MLPCTP1-$(His)_6$; MLPCTP1 ; MLPCTP1-CDP et $MLPCTP1_{107\text{-}171}$-$(His)_6$. Les dépôts correspondent à 13 cultures de 13 clones pour chaque construction.

3. Discussion

Pendant des années, l'étude des effecteurs chez les Pucciniales a été ralentie par le manque d'outils moléculaires applicables à ces pathosystèmes (Petre *et al.,* 2014). Les approches effectoromiques utilisant des systèmes hétérologues ont été popularisées sur d'autres microorganismes phytopathogènes et ont récemment été appliquées aux Pucciniales pour étudier la localisation subcellulaire et rechercher les interactants protéiques d'effecteurs candidats (**Annexe 1 et 2 ; articles n°7 et 8** ; De Carvalho *et al.,* 2016; Germain *et al.,* 2016; Maia *et al.,* 2017; Qi *et al.,* 2017). Si ces cribles ne permettent pas de déterminer les fonctions biologiques des effecteurs candidats, ils représentent une opportunité pour connaître les compartiments cellulaires et les interactants potentiellement ciblés dans la cellule végétale. Ces systèmes hétérologues sont peu onéreux et rapides en termes d'obtention de résultats. Ils permettent aussi de cribler un grand nombre d'effecteurs candidats. La simplicité de leur mise en place en laboratoire les rend facilement accessibles pour la communauté et ils peuvent être appliqués à différentes espèces. Ces aspects sont développés dans une *Mini-Review* invitée dans le journal *Current Opinion in Microbiology* intitulée « **Show me the way: rust effector targets in heterologous plant systems** » dont je suis première auteure. Cet article qui sera publié dans le courant de l'année 2018 est adossé à la discussion générale. Il discute en détail les cribles effectoromiques en systèmes hétérologues pour l'étude des effecteurs chez les Pucciniales (cf. **Article n°6**).

Les résultats complémentaires obtenus pour MLPCTP1 montrent que le mode d'expression stable ou transitoire n'influence pas la localisation chloroplastique de l'effecteur candidat MLPCTP1 (**Annexe 1; article n°7**). Notamment l'expression de MLPCTP1 dans le tabac, Arabidopsis et le peuplier montre dans tous les cas une accumulation dans le stroma des chloroplastes. Li et al. (2014) ont montré que les deux effecteurs bactériens de *Pseudomonas syringae* pv. *tomato*, HopK1 et AvrRps4 présentent une localisation différente en fonction du système et du mode d'expression. En effet, ils sont localisés dans les chloroplastes dans les plants transgéniques stables d'*A. thaliana* alors qu'ils sont nucléo-cytoplasmiques en cas d'expression transitoire chez *N. benthamiana* (Li *et al.,* 2014). La localisation chloroplastique de MLPCTP1 dans les trois plantes utilisées montre la robustesse de l'adressage de cet effecteur candidat dans les chloroplastes. Il est à noter que je n'ai pas pu régénérer de peupliers exprimant la GFP seule, ni de plants exprimant le peptide de transit seul de MLPCTP1. L'obtention de peupliers transgéniques exprimant ces constructions sont des contrôles qui seront nécessaires à la poursuite des analyses des transformants exprimant MLPCTP1. Aucune différence de phénotype notable n'a pu être observée directement entre les peupliers transgéniques exprimant MLPCTP1 et des témoins non transformés, qu'ils soient

maintenus *in vitro* en tube ou en boîte Magenta, mais aussi en pot en terre après acclimatation et croissance dans une serre S2 OGM (données non présentées). Des suivis de paramètres physiologiques chez les plants transformés, notamment pour des paramètres associés aux fonctions chloroplastiques (assimilation du carbone, conductance stomatique, activités enzymatiques chloroplastiques) pourront être réalisés pour rechercher le rôle potentiel de MLPCTP1. Les peupliers exprimant MLPCTP1-GFP pourront aussi permettre de rechercher des interactants protéiques potentiels par co-immunoprécipitation couplée à de la spectrométrie de masse (Co-IP/MS) comme cela a été réalisé chez *N.* benthamiana (**Annexe 1; article 7**). L'approche de Co-IP/MS sur feuilles de *N. benthamiana* exprimant MLPCTP1-GFP, n'avait pas permis d'identifier d'interactants potentiels en nombre restreint et robuste. Cependant, il est envisageable que la cible de l'effecteur candidat MLPCTP1 ne soit pas présente ou présente des divergences suffisantes pour ne pas permettre d'interaction chez le tabac. Une telle approche de CoIP/MS pourrait être envisagée sur des chloroplastes de peuplier isolés. Le clone INRA *P. tremula x P. alba* 717-1B4 est largement utilisé par la communauté scientifique s'intéressant à cet arbre modèle pour sa facilité de transformation due à sa sensibilité à *A. tumefasciens*. C'est un clone de choix pour réaliser des approches fonctionnelles. Toutefois, il n'est pas compatible avec *M. larici-populina* et il est sensible à d'autres espèces de la famille des Melampsoraceae. Dans le cadre de ma thèse, la génération des mutants de peuplier a demandé beaucoup de temps entre la transformation effective et la régénération et l'acclimatation de plants en serre. Je n'ai pas pu établir de pathosystème avec une espèce compatible de *Melampsora* spp. afin de tester si l'expression de l'effecteur MLPCTP1 avait un effet notable sur l'infection. Il sera intéressant de tester ce point à la suite de ma thèse. Les clones de 717-1-B4 étant régulièrement utilisés par l'équipe *Ecogénomique des interactions* pour la mycorhization avec le champignon *Laccaria bicolor* S238N dans des systèmes *in vitro* ou en serres, il sera intéressant de tester si l'expression d'un effecteur de rouille affecte la colonisation par un biotrophe mutualiste. Les plants d'arabettes transgéniques exprimant MLPCTP1 n'ont pas présenté de phénotype particulier en comparaison avec des témoins non transformés. L'obtention de mutants chez cette plante modèle ouvre des perspectives intéressantes pour étudier le rôle des effecteurs d'une manière générale, du fait de l'existence d'un grand nombre de mutants et de phénotypes répertoriés chez cette plante notamment vis à vis des réactions immunitaires lors d'interactions avec des micro-organismes pathogènes. J'ai obtenu ces transformants lors d'un séjour au Canada dans le cadre d'une collaboration avec le Pr. Hugo Germain (UQTR, Québec, Canada). Au sein de ce laboratoire, la plante modèle *A. thaliana* est utilisée pour exprimer les effecteurs de *M. larici-populina*, et tester si un effet peut être observé lors d'interactions avec la bactérie *P. syringae* ou l'oomycete *Hyaloperonospora arabidopsidis* qui forme des haustoria comme les rouilles

(Germain *et al.,* 2017). J'ai obtenu les mutants d'*A. thaliana* exprimant les fusions MLPCTP1-GFP tardivement au cours de ma thèse et je n'ai pas pu exploiter ce système pour avancer sur l'étude de la fonction de cet effecteur. Il sera intéressant de tester par la suite l'impact de l'expression de cet effecteur candidat de rouille sur différentes fonctions chez *A. thaliana*.

Très peu de données ont été publiées quant à l'étude de la structure d'effecteurs chez les Pucciniales (**Cf. article n°1**). Les seules informations disponibles concernent le système modèle de la rouille du lin *M. lini* (Wang *et al.,* 2007 ; Ve *et al.,* 2013 ; Zhang *et al.,* 2017). Dans le pipeline d'analyse établi pour étudier les effecteurs candidats de *M. larici-populina*, nous avons mis à profit l'expertise de l'Equipe *Réponse au stress et Régulation Redox* de l'UMR 1136 IAM pour la production de protéines recombinantes en vue d'obtenir des informations quant à la structure de ces effecteurs, permettant d'envisager des approches de structure/fonction dans la compréhension de leur rôle lors de l'infection. Toutefois, la production de petites protéines sécrétées dans les systèmes bactériens habituellement utilisés s'est avérée difficile. Après diverses collaborations, nous avons finalement pu obtenir une première ébauche de la structure de l'effecteur candidat MLP124017 par RMN. Ayant déjà à disposition la structure d'un autre effecteur de *M. larici-populina* MLP124266 homologue du facteur d'avirulence AvrP4 de *M. lini* (thèse de Benjamin Pêtre, 2012), nous proposons de publier ces structures pour les présenter à la communauté, notamment parce que leur accès présente certains intérêts. Dans un cas, cela permettra de proposer la première structure d'une protéine fongique dans la base de données Knottin qui compile les protéines à domaines Cys-Knot et dans l'autre cas, la structure permet d'identifier une homologie structurale là où aucune homologie de séquence n'était disponible. Au delà, les informations structurales à disposition permettent d'envisager des analyses complémentaires pour la compréhension des mécanismes dans lesquels les effecteurs candidats sont engagés. En effet, nous avions pu montrer l'interaction entre MLP124017 et des protéines TOPLESS et TOPLESS-RELATED PROTEIN. A titre d'exemple, la connaissance de la structure de MLP124017 peut permettre des approches mutationnelles pour identifier les acides aminés engagés dans l'interaction avec la protéine cible.

En conclusion, au cours de ma thèse je me suis intéressée à deux effecteurs candidats du champignon *M. larici-populina*. Dans le cas de MLPCTP1, j'ai pu générer des mutants chez différentes plantes qui permettront de poursuivre son étude dans les années à venir. Dans le cas de MLP124017, nous avons pu obtenir des informations quant à sa structure, ce qui permettra là aussi de poursuivre l'étude de son rôle dans les mécanismes d'interaction. Dans ce projet, nous nous sommes essentiellement intéressés au rôle de l'effecteur sans

présupposé quant aux processus visés chez l'hôte. Il serait envisageable de tester si ces effecteurs auraient un quelconque effet sur l'immunité en mettant en place des cribles vis à vis de la suppression ou de l'induction de la réponse immunitaire dans des systèmes hétérologues, tel que cela a été réalisé chez la rouille du Soja (Qi *et al.,* 2016 ; Qi *et al,.* 2017). Un enseignement à tirer des travaux effectués au cours de cette thèse est le temps nécessaire à la réalisation de certaines expérimentations, notamment pour l'obtention et l'utilisation de mutants chez le peuplier.

CHAPITRE V : Discussion générale

Partie 1 : Questions et stratégie de recherche pour l'étude de l'interaction peuplier-rouille du peuplier

Depuis plus de cinq ans, un projet transversal d'étude de l'interaction peuplier-rouille du peuplier est mené au sein de l'UMR 1136 IAM entre les différentes équipes composant l'UMR. Le but est de comprendre les mécanismes et l'évolution de l'interaction à travers différents axes de recherche centrés sur l'étude du dialogue moléculaire établi entre l'hôte et son pathogène lors d'interactions compatibles (maladie) et incompatibles (résistance) ; l'étude de la typologie de l'infection au-delà des traits d'infection classiquement étudiés et l'étude de l'évolution des populations du pathogène fongique. Mon travail de thèse s'est inséré dans ce projet, et plus particulièrement dans le premier axe sur l'étude du rôle des effecteurs dans le processus infectieux chez le champignon.

Mon travail a permis de compléter nos connaissances de la biologie de *M. larici-populina* en décrivant dans un premier temps le programme génétique exprimé au cours de son cycle de vie à des stades non étudiés jusqu'à présent et notamment lors de l'alternance sur hôte écidien, le mélèze. La comparaison des profils d'expression génique sur les deux hôtes a permis de mettre en évidence l'expression préférentielle de gènes codant des petites protéines sécrétées lors de l'infection du peuplier ou du mélèze. La possibilité de discriminer des effecteurs candidats exprimés spécifiquement chez un hôte représente une avancée conceptuelle majeure pour la compréhension des interactions moléculaires entre les Pucciniales et leurs hôtes. De plus, l'apport de données d'expression à de nouveaux stades du cycle de vie contribue à améliorer l'annotation du génome de *M. larici-populina*. Dans un deuxième volet de ma thèse, j'ai contribué à l'effort d'analyse de la nouvelle version du génome de *M. larici-populina*, ainsi que de celui d'un autre agent de la rouille foliaire du peuplier, *M. allii-populina*. Sur la base des données mises à disposition par le JGI, je me suis attachée à réaliser l'annotation des éléments transposables et de certaines catégories de gènes jouant potentiellement un rôle important lors de l'infection (sécrétome, CAZYmes, protéases et transporteurs) à la fois chez *M. larici-populina* et chez *M. allii-populina*. Ces travaux permettent ainsi d'amorcer l'analyse comparative de deux génomes de Melampsoraceae partageant un même hôte télien et un même environnement. Le troisième volet de mon projet de thèse a permis d'amorcer l'analyse fonctionnelle et structurale d'effecteurs candidats dont la première famille d'effecteurs fongiques décrite ciblant les chloroplastes (Oliveira-garcia & Valent, 2015). Les résultats obtenus au cours de ma thèse permettent d'ouvrir de nouvelles perspectives dans l'étude des mécanismes des interactions entre plantes et agents des rouilles.

Ce dernier chapitre replace mes travaux de thèse dans le contexte général des interactions moléculaires entre plantes et microorganismes et celui de questions auxquelles le domaine est confronté. Ici, j'aborderai et mettrai en perspective les points de discussion liés aux articles et résultats présentés dans les chapitres précédents et je montrerai dans quelle mesure ceux-ci nous permettent d'avancer dans la compréhension des mécanismes d'interaction plantes-microorganismes. Je vais tout d'abord me concentrer sur les aspects de génomique et de transcriptomique abordés dans les chapitres II et III, notamment en discutant de la spécificité d'hôte et de la coévolution plante-champignon chez les Pucciniales. Je ferai ensuite le bilan de l'étude de la biologie des effecteurs et plus particulièrement leur analyse fonctionnelle chez *M. larici-populina* et plus largement chez les pathogènes filamenteux. Je discuterai des avancées réalisées et des limites inhérentes imposées aux pathosystèmes non-modèles, particulièrement pour les microorganismes biotrophes obligatoires que sont les Pucciniales. Pour finir, une partie de la discussion sera dédiée plus spécifiquement à la caractérisation fonctionnelle d'effecteurs candidats chez les Pucciniales par des approches effectoromiques en systèmes hétérologues pour l'identification de leur localisation et de leurs cibles. Cette partie est rédigée sous la forme d'un article de synthèse invité qui a été accepté pour publication en janvier 2018 dans la revue *Current Opinion in Microbiology* (cf. **Article 6**).

Partie 2 : Les apports de la génomique à l'étude de la biologie des Pucciniales

Le séquençage de génomes de Pucciniales a révélé des caractéristiques communes comme leur taille importante (de 80Mb à plus de 2Gb) et leur richesse intrinsèque en éléments transposables (Aime *et al.,* 2017). Les Pucciniales possèdent aussi des répertoires de gènes étendus au sein desquels la part codant des protéines sécrétées est importante (plus de 10%) comparée aux nécrotrophes et hémibiotrophes (Lo presti *et al.,* 2015). La combinaison de ces caractéristiques communes aux Pucciniales est unique au sein du règne fongique (Duplessis *et al.,* 2013). Ces caractéristiques représentent un challenge notable pour le séquençage et surtout l'assemblage des génomes des Pucciniales (Aime *et al.*, 2017). Toutefois, les récentes technologies de séquençage générant de longs fragments, combinées à de nouvelles méthodes pour leur assemblage laissent entrevoir la production de séquences de génomes complexes de bonne qualité pour cet ordre fongique (Schwessinger *et al.,* 2017, Miller *et al.,* 2017, Aime *et al.,* 2017).

L'analyse détaillée du contenu du génome, l'analyse comparative entre génomes, ainsi que les approches de transcriptomique permettent d'aborder certaines questions clés de la biologie des Pucciniales comme i) quelles sont les bases moléculaires des mécanismes d'infection et notamment, au sein des répertoires d'effecteurs candidats quels sont ceux communs aux Pucciniales et ceux spécifiques à une famille taxonomique ou à une espèce donnée, ii) quelles sont les bases de la spécificité d'hôte et les mécanismes de sélection sous-jacents à l'adaptation à un hôte donné et iii) comment les éléments transposables ont envahis les génomes de Pucciniales et quel est leur impact sur la biotrophie et les mécanismes de pathogénèse, par exemple sur la diversification des familles multigéniques dont celles codant des effecteurs ?

1. Apports de la transcriptomique à l'étude de l'hétéroécie chez les Pucciniales

La transcriptomique est une méthode de choix pour comprendre les programmes génétiques mis en œuvre dans les interactions entre un champignon pathogène et son hôte (Tan *et al.,* 2009 ; Schenk *et al.,* 2012). Divers outils sont disponibles pour quantifier l'abondance des transcrits d'un organisme donné, depuis les puces à ADN jusqu'au séquençage d'ARN. De telles approches ont notamment été utilisées pour mieux comprendre la biologie des Pucciniales au cours de l'infection (de Carvalho *et al.,* 2017; Cuomo *et al.,* 2017; Dobon *et al*., 2016; Jing *et al.,* 2017; Huang *et al.,* 2011; Liu *et al.,* 2015; Rochi *et al.,* 2016; Rutter *et al.,* 2017; Talhinhas *et al.,* 2014; Tao *et al.,* 2017 ; Duplessis *et al.,* 2012, Duplessis *et al.,* 2014). L'étude ciblée des programmes génétiques exprimés lors de la formation de structures d'infection de Pucciniales comme des tubes germinatifs issus d'urédiniospores, des appressoria ou encore des haustoria purifiés a conduit à une meilleure compréhension des processus biologiques mis en place lors de l'établissement d'une interaction compatible avec une plante hôte. Les profils d'expression révélés chez plusieurs espèces de Pucciniales ont montré que les catégories de gènes impliquées dans les mécanismes de transcription et de traduction, dans le métabolisme primaire ou la production d'énergie, sont particulièrement représentées parmi les gènes fortement exprimés, soulignant l'activité biologique intense du pathogène lors de son développement. D'autres catégories de gènes associées à des mécanismes actifs de dégradation ou de transports de nutriments comme les protéases, les CAZymes ou encore les transporteurs font partie des gènes fortement régulés lors de l'interaction avec l'hôte. Les gènes codant des protéines sécrétées montrent quant à eux des profils d'expression très dynamiques, sous la forme de vagues d'expression successives au cours des phases de colonisation et de croissance biotrophe. De tels profils ont notamment été décrits chez plusieurs espèces de Pucciniales comme *M. larici-populina* ou *P. striiformis* f.

sp. *tritici*, respectivement au cours de l'infection de feuilles de peuplier et de blé (Duplessis *et al.,* 2011; Cantu *et al.,* 2013; Dobon *et al.,* 2017; Huang *et al.,* 2011; Petre *et al.,* 2014).

Les analyses de transcriptomes conduites chez les Pucciniales ont largement visé l'infection d'un seul hôte, l'hôte tellien, et un seul type de spores produites lors de la multiplication clonale sur l'hôte tellien, les urédiniospores. En effet, les dommages les plus conséquents observés au cours de la phase épidémique sont le plus souvent occasionnés sur cet hôte (Duplessis *et al.,* 2014). Il existe un nombre limité d'information quant à l'expression des gènes de Pucciniales tout au long du cycle de vie, et notamment très peu d'études se sont intéressées au programme génétique mis en place los de l'infection de l'hôte alternant. Pourtant, ce stade est critique puisque c'est celui où la reproduction sexuée prend place (Duplessis *et al.,* 2014). Chez l'agent de la rouille du blé *Puccinia triticina*, le séquençage d'étiquettes de séquences exprimées (« expressed sequence tags », EST) et le séquençage d'ARN (RNAseq) ont permis d'identifier, à une échelle modeste, des groupes de gènes exprimés sur chacun des deux hôtes, *Thalictrum* et blé (Xu *et al.,* 2011; Cuomo *et al.,* 2017). Du RNAseq a pu être réalisé à une profondeur de séquençage plus conséquente sur des spores de *Cronartium ribicola* collectées sur les deux hôtes *Pinus monticola* (écidiospores) et *Ribes nigrum* (urédiniospores) puis comparé à deux stades d'infection sur le pin argenté (Liu *et al.,* 2015). Bien que le génome ne soit pas disponible pour cette espèce, l'assemblage du transcriptome a permis de montrer que la majorité des gènes sont exprimés dans tous les stades collectés alors que seuls 4-5% des gènes sont spécifiquement exprimés à chaque stade (Liu *et al.,* 2015). Malgré l'annotation partielle du sécrétome de *C. ribicola*, les auteurs ont pu mettre en évidence quelques centaines de petites protéines sécrétées exprimées *in planta* et qui représentent de potentiels effecteurs candidats.

L'analyse des profils d'expression de *M. larici-populina* à trois stades associés à l'infection de l'hôte écidien, le mélèze, présentée dans le **chapitre II** de cette Thèse, a mis en évidence un transcriptome partagé (*core transcriptome*) important avec 62% des gènes exprimés aux stades basides, pycnies et écidies. Cependant, les profils d'expression génique lors de l'infection des aiguilles de mélèze (pycnies et écidies) sont plus proches entre eux, en comparaison avec le stade basides suggérant une reprogrammation importante du programme génétique de *M. larici-populina* lors du passage à l'hôte écidien. Il est intéressant de noter que les stades pycnien et écidien correspondent respectivement aux stades avant et après la fertilisation qui conduit à la production d'un mycélium dicaryotique. Toutefois, l'ensemble des stades marquant l'infection de cet hôte est complexe. En effet, des structures telles que les proto-écidies sont formées dans les tissus végétaux avant la fertilisation entre

pycniospores et hyphes flexueux compatibles pour permettre la fertilisation, ce qui pourrait expliquer ce profil d'expression global similaire entre les stades pycnies et écidies (Figure 19 ; Harder, 1984 ; Voegele *et al.,* 2009). Les gènes codant des protéines sécrétées de fonctions inconnues représentent la seule catégorie significativement surexprimée entre les stades pycnies et écidies et le stade basides, suggérant un rôle important de ces protéines sécrétées dans l'infection des aiguilles de mélèze. La représentation significative de cette catégorie parmi les gènes les plus fortement régulés a également été observée dans des études transcriptomiques portant sur l'infection de feuilles de peuplier (Duplessis *et al.,* 2011b ; Hacquard *et al.,* 2010 ; 2012). Ces résultats démontrent l'importance des protéines sécrétées pour l'infection des deux hôtes et indiquent de potentiels effecteurs candidats. Toutefois, devant la complexité du processus d'infection réalisé chez l'hôte écidien, il est tout à fait envisageable que des vagues d'expression puissent être observées pour ces catégories de gènes, comme cela a pu être rapporté chez l'hôte peuplier. Il serait intéressant d'envisager une dissection plus fine de ce processus d'infection afin de distinguer des gènes éventuellement spécifiques de la formation des protoécidies, puis de la production des gouttelettes de nectar, de la fertilisation, puis de la constitution des écidies et des écidiospores avant leur libération dans les pustules visibles à la surface des aiguilles de mélèze (Figure 26).

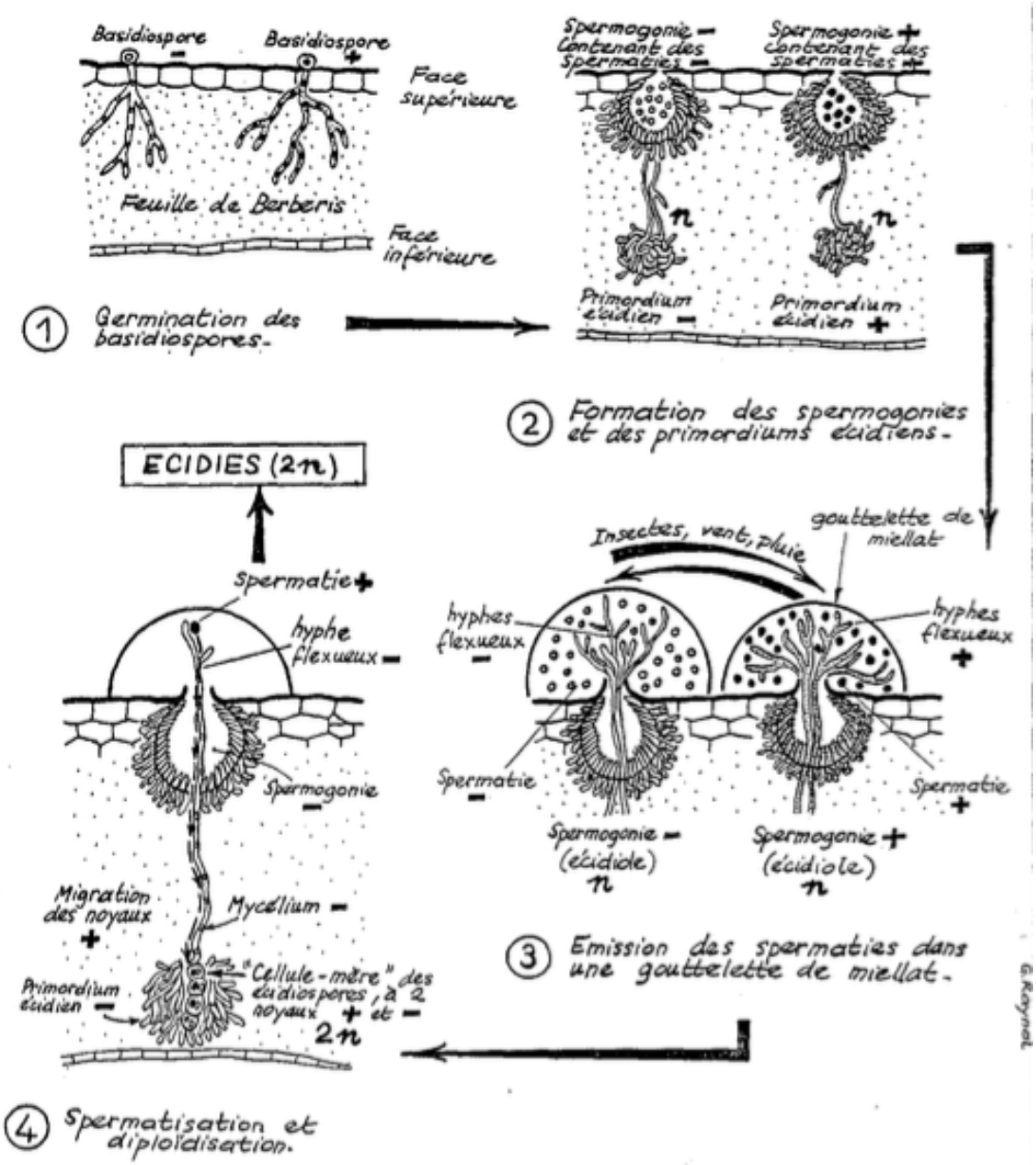


Figure 26 : Le processus d'infection complexe de l'hôte écidien de *P. striiformis* f. sp. *tritici*.

1) Au début du printemps, les basidiospores (n) des deux types sexuels + et - infectent l'hôte écidien par la face supérieure des feuilles de *Berberis* pour produire 2) des spermogonies (ou pycnies ; n) contenant des spermaties (ou pycniospores ; n) côté face supérieure de la feuille, ainsi que des primordia écidiens (n) côté face inférieure. 3) Les spermogonies se développent et produisent des gouttes de nectar contenant des hyphes flexueux et des spermaties. La fertilisation se produit par la rencontre de spermaties de chaque type sexuel, par l'intermédiaire d'insectes par exemple. Lorsqu'une spermatie atteint un hyphe flexueux de type opposé, les noyaux + (spermatie) et – (spermogonie) migrent le long du mycélium pour atteindre la cellule-mère des primordia écidien et produire des écidiospores à deux noyaux (n+n). Une fois les écidiospores produites, celles-ci sont libérées côté inférieure des feuilles.*Raynal (INRA), Cours sur les Pucciniales*

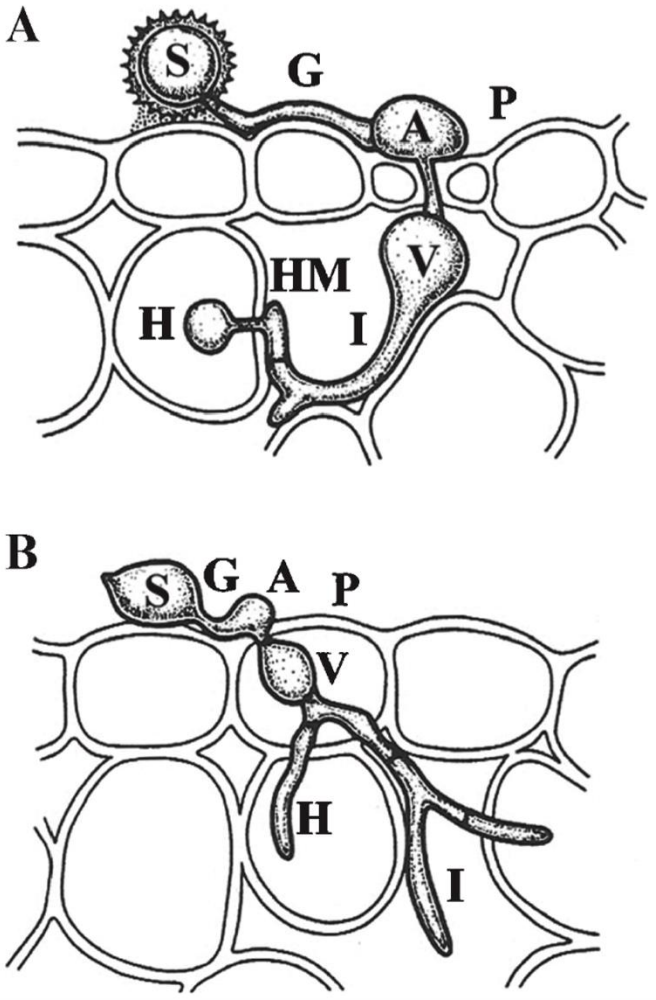


Figure 27 : Structures d'infection des urédiniospores et des basidiospores de Pucciniales : deux mécanismes de pénétration différents.

(A) les urédiniospores pénètrent dans les tissus foliaires par les stomates et (B) les basidiospores par la cuticule. A appressorium, G tube germinatif, H haustorium, HM cellule mère haustoriale, I hyphe d'infection, P hyphe de pénétration, S spore, V vésicule. *Voegele et al., 2009*

La question de l'existence de répertoires d'effecteurs différents propre à l'infection des différents hôtes chez les Pucciniales est centrale et est posée de manière récurrente au sein de la communauté des mycologues (Shulze-Lefert & Panstruga, 2011). Si les quelques études mentionnées auparavant apportaient déjà quelques éléments de réponse, l'analyse plus globale et systématique sur la base de génomes à disposition restait à réaliser (Duplessis *et al.,* 2014). L'étude menée lors de l'infection du mélèze par *M. larici-populina* permet de distinguer un répertoire de gènes communs exprimés lors de l'infection chez les deux hôtes et des répertoires de gènes plus spécifiques à un hôte ou à l'autre (Lorrain *et al.*, 2018). Ainsi, seule une fraction des gènes codant des protéines sécrétées est exprimée spécifiquement sur chacun des hôtes confirmant les résultats rapportés chez *P. triticina* et *C. ribicola* (Xu *et al.,*

2011, Liu *et al.,* 2015, Cuomo *et al.,* 2017). Ceci suggère que la majorité des effecteurs de *M. larici-populina* présents parmi ce groupe de gènes pourraient avoir des fonctions identiques lors de la colonisation des deux hôtes. Il est intéressant de noter que parmi les familles de petites protéines sécrétées identifiées dans les études antérieures (Hacquard *et al.,* 2012), certaines d'entre elles montrent un profil tout à fait singulier avec une partie des membres exprimés lors de l'infection du peuplier et l'autre partie chez le mélèze. Ces familles particulières pourraient soit présenter des fonctions différentes chez les deux hôtes, ou des fonctions conservées mais des cibles différentes, ou encore avoir une même cible qui aurait évoluée différemment dans les deux hôtes à partir d'un ancêtre commun (Lorrain *et al.,* 2018).

Comme le présente la figure 27, les étapes de pénétration dans les organes infectés chez le peuplier et le mélèze sont radicalement différents et nécessitent vraisemblablement dans le cas des basidiospores un arsenal spécifique pour assurer la pénétration à travers la cuticule des aiguilles du mélèze puis pour traverser l'épiderme avant de poursuivre avec un mode de croissance intercellulaire (Ragazzi *et al.,* 2005). Parmi les catégories de gènes distinguables à cette étape de colonisation, certaines enzymes comme les CAZymes sécrétées, sont exprimées spécifiquement lors de l'infection du mélèze ou du peuplier (Lorrain *et al.,* 2018). La même observation a été faite chez *C. ribicola* lors des infections de *Ribes* et du pin (Liu *et al.,* 2015). En effet, les CAZymes sont des enzymes essentielles pour l'altération des polysaccharides composant la paroi cellulaire des plantes (Henrissat *et al.,* 2017). Malgré le répertoire restreint en CAZymes observé chez *M. larici-populina*, et chez les biotrophes obligatoires en général, elles restent essentielles pour le succès de l'infection et notamment pour la pénétration de l'hyphe dans les étapes précoces d'infection (Duplessis *et al.,* 2011a). Aussi, la formation de l'haustorium à des étapes plus tardives de colonisation chez le peuplier comme chez le mélèze altère ponctuellement la paroi cellulaire (Ragazzi *et al.,* 2005). Il est communément admis que la structure de la paroi végétale varie en composition et en structure selon les espèces végétales, le stade de développement, l'organe et même selon le type cellulaire (Kubicek *et al.,* 2014). Des répertoires de CAZymes différents exprimés au cours de l'infection du peuplier et du mélèze peuvent indiquer une adaptation à la composition des parois cellulaires chez un angiosperme ou un gymnosperme. Par exemple, certains gènes codant des glycosides hydrolases des familles GH5, GH17 et GT32 remodelant la paroi cellulaire sont spécifiquement surexprimés pendant l'infection des aiguilles de pin par *C. ribicola* (Liu *et al.,* 2015). Chez *M. larici-populina*, la famille GH5 est en expansion avec 29 membres dont certains ont été identifiés comme surexprimés lors de l'infection des feuilles de peuplier (Duplessis *et al.,* 2011a) et certains membres montrent une expression spécifique lors de l'infection du mélèze (Lorrain *et al.,* 2018).

De même, lors de la croissance biotrophe, on pourrait s'attendre à une spécialisation dans la régulation de l'expression de certaines voies métaboliques lors des infections des différents hôtes. Cependant, l'analyse des voies métaboliques annotées avec la base de données KEGG ne montre pas de différence majeure en termes d'expression, ce qui corrobore également l'expression des catégories de gènes KOG similaire dans les deux hôtes. Ces résultats suggèrent que *M. larici-populina* utilise un programme génétique similaire lors des phases biotrophes de la colonisation du mélèze et du peuplier aux stades qui ont été étudiés. *M. larici-populina* utilise apparemment des stratégies similaires d'acquisition et d'utilisation des nutriments dérivés des deux hôtes. Encore une fois, il serait intéressant de vérifier si cela s'applique à tous les stades d'infection, comme par exemple lors de la formation des gouttelettes de nectar qui disposent peut-être d'une composition particulière pour l'attraction des mouches qui participent à la fertilisation. Il est aussi à noter que comme chez l'ensemble des Pucciniales dont le génome a été séquencé jusqu'à présent, plus de la moitié du génome de *M. larici-populina* correspond à des gènes de fonctions inconnues ne présentant aucune homologie en dehors des Pucciniales (Aime *et al.,* 2017). Ces gènes représentent 58% des gènes communs exprimés chez les deux hôtes et plus de 10% sont exprimés spécifiquement sur peuplier et plus de 10% sont exprimés spécifiquement sur mélèze. Une analyse plus approfondie de l'infection des deux hôtes, chez différentes espèces de Pucciniales, ou à travers des cinétiques d'infection plus détaillées, pourrait aider à l'élaboration de nouvelles hypothèses quant aux mécanismes dans lesquels ces gènes sont impliqués. Au delà, de nombreuses étapes du cycle de vie de *M. larici-populina* (et cela est vrai pour toutes les autres espèces de Pucciniales) ne sont pas étudiées, comme l'entrée ou la sortie de dormance observée pour les téliospores. Etendre les approches de transcriptomique à l'étude des cycles de vie des Pucciniales d'une manière plus systématique permettrait d'apporter de nouvelles perspectives et décrire de nouvelles fonctions associées à la biotrophie.

2. Revisiter l'annotation des génomes afin d'affiner les répertoires de gènes clés associées à la biotrophie

Au cours des années 2000, des avancées majeures dans la compréhension des interactions moléculaires entre les plantes et les Pucciniales ont été réalisées à travers des approches de génomique et de transcriptomique (Duplessis *et al.,* 2014). L'obtention de génomes de référence de qualité est primordiale pour l'analyse de leur contenu. Plusieurs facteurs entrent en jeu pour atteindre une qualité de séquence suffisante. Par exemple, l'amélioration des outils de séquençage et d'assemblage, l'acquisition de données d'expression à différents stades de développement ou encore le re-séquençage d'isolats et l'obtention d'une carte génétique

peuvent permettre d'améliorer l'annotation de génomes de référence. Dans le cadre d'une collaboration établie entre l'UMR 1136 IAM et le JGI, un nouvel assemblage du génome de référence de l'isolat 98AG31 de *M. larici-populina* a été ancré sur la carte génétique générée par Michaël Pernaci au cours de sa thèse (Pernaci, 2015). Par ailleurs, les données d'expression de RNAseq obtenues sur l'hôte écidien décrites dans le **chapitre II** ont permis l'annotation de nouveau modèles de gènes par le JGI faisant évoluer le catalogue de gènes de 16399 à 19550 gènes dans la version 2 du génome. L'annotation *de novo* des TE dans cette nouvelle version du génome indique qu'ils représentent 42% de la séquence totale. Aussi, les groupes de gènes codant des CAZymes, des protéases, des transporteurs et des protéines sécrétées ont pu être réévalués. Les différences observées en termes de nombre de gènes entre les deux versions pour les catégories ciblées sont à la fois dues à l'amélioration des séquences nucléotidiques dans le nouvel assemblage mais aussi aux approches utilisées pour identifier et retenir les séquences. Ainsi l'analyse expertisée des CAZymes révèle surtout des différences en termes de nombre, mais pas pour les proportions au sein des différentes catégories. En revanche, l'analyse du transportome paraît plus compliquée et une analyse sur la base d'outils de bioinformatique ne semble pas suffisante et nécessite une validation experte humaine systématique. C'est ce que montre l'analyse de la famille des gènes OPT. L'effort d'annotation demandé sera donc plus important que pour d'autres catégories. Il est toutefois essentiel car les transporteurs jouent un rôle essentiel chez un biotrophe obligatoire pour l'importation de nutriments lors de l'infection.

Une perspective directe à la nouvelle annotation des gènes pour la version 2 du génome de *M. larici-populina* sera de replacer les données d'expression acquises ces dernières années à l'aide de différents outils de transcriptomique. L'assignation des séquences issues du RNAseq (Lorrain *et al.,* 2018) pourra être réalisée simplement sur la base des outils déjà utilisés sur la version 1 du génome. Il sera aussi possible d'exploiter les données issues des microréseaux à oligonucléotides NimbleGen lors de l'infection chez le peuplier (Duplessis *et al.*, 2011b). Du fait de la modification de l'annotation d'un nombre important de gènes, il ne sera pas possible de simplement transférer les données d'expression sur la base d'homologies entre les deux catalogues de gènes. Il sera nécessaire de ré-assigner les oligonucléotides sur les nouveaux modèles de gènes et ainsi redéfinir les niveaux d'expression aux stades étudiés. Avec ces jeux de données en main, et avec l'ancrage sur la carte génétique, il sera intéressant de rechercher s'il existe un niveau d'organisation spécifique dans le génome pour l'infection des deux hôtes avec des clusters d'expression propres à l'infection du peuplier et du mélèze. En effet, la version 1 du génome correspondait à 462 *scaffolds* pour lesquels l'ordre relatif d'était pas connu (Duplessis *et al.,* 2011a).

Le sécrétome de *M. larici-populina* déterminé dans la nouvelle version du génome par des outils de prédiction contient 2238 SP dont 1439 SSP. L'approche communément retenue pour prédire le sécrétome des micro-organismes phytopathogènes à partir de leur génome comprend plusieurs étapes dont i) la prédiction de la présence d'un peptide signal dans la région N-terminale de la protéine, ii) l'absence de domaines transmembranaires dans la protéine mature, iii) l'absence de signaux d'adressage spécifiques. Un des défis pour les outils de prédiction est de réussir à distinguer les peptides signaux des domaines transmembranaires situés dans les régions N-terminales, qui représentent tous deux des régions hydrophobes. Il est intéressant de noter que la dernière version en date disponible de l'outil de prédiction utilisé actuellement pour déterminer les peptides signaux est SignalP4.1, accessible depuis 2011 (Petersen *et al.,* 2011). Cet outil a été évalué comme plus performant dans la distinction entre peptide signal et domaine transmembranaire, mais montre moins de sensibilité pour prédire les sites de clivage des peptides de sécrétion (Sperschneider *et al.,* 2017a). La version précédente de SignalP montrait une meilleure sensibilité pour la prédiction des sécrétomes chez les champignons (Sperschneider *et al.,* 2015). Il en est de même pour le programme TargetP largement utilisé pour déterminer certains signaux d'adressage. De plus, il est admis que certaines protéines sécrétées ne présentent pas de peptide signal et sont sécrétées de manière non conventionnelle, par exemple deux isochorismatases impliqués dans la pathogénicité de deux oomycètes *Phytophthora sojae* et *Verticilium dahliae* sont sécrétées de manière non-conventionnelle (Liu, Song *et al.,* 2014). SecretomeP est un outil de prédiction qui a été développé pour rechercher de telles protéines au sein des protéomes, cependant il induit un fort taux de faux positifs dans les prédictions réalisées chez les eucaryotes (Lonsdale *et al.,* 2016). La combinaison de plusieurs outils de prédiction pour les peptides signaux paraît l'approche la plus appropriée pour tenter de corriger les particularités de chaque algorithme. De plus, il est important de garder à l'esprit que la prédiction de peptide signal est dépendante de la qualité de l'annotation des modèles de gènes réalisée en amont. En effet, les régions N-terminales peuvent parfois être mal annotées et par conséquent ces gènes ne seront pas reconnus comme sécrétés par les outils de prédiction. Ainsi, toute amélioration de l'annotation des génomes, par des données d'expression notamment, permet aussi d'améliorer la qualité de la prédiction des sécrétomes. Aussi, sur la base d'une nouvelle annotation et d'une analyse comparative des familles multigéniques au sein du génome étudié ou avec des espèces voisines, il est possible d'identifier des erreurs d'annotation de la région N-terminale par comparaison de la composition des différents gènes en exons pour une famille donnée.

Le but d'une telle identification du sécrétome chez les micro-organismes phytopathogènes est de pouvoir identifier des effecteurs candidats. Contrairement aux oomycètes, aucun motif particulier ne permet de discriminer des effecteurs candidats au sein des génomes des champignons et l'identification d'effecteurs candidats est basée sur la recherche de plusieurs critères souvent retrouvés chez les effecteurs déjà caractérisés (ex. taille, richesse en cystéines, expression pendant l'infection, homologies aves des effecteurs connus, etc ; voir Lorrain *et al.,* 2015). Dans le but de pouvoir classer les candidats remplissant ces critères, des pipelines basés sur du regroupement hiérarchisé de familles de protéines sécrétées ont été proposés, notamment pour *P. striiformis* f. sp*. tritici* et *M. larici-populina* (Saunders *et al.,* 2012). Toutefois, il n'y a pas de règle absolue pour déterminer l'importance et donc le poids à attribuer à chaque critère, car il existe de nombreux cas d'effecteurs dérogeant à la présence de tous ces critères (Lorrain *et al.,* 2015).

Malgré l'établissement de tels pipelines, l'identification des effecteurs candidats repose encore largement sur des *a priori* sans justification. Une illustration est la taille des protéines avec des seuils arbitraires à 300, 250 ou 200 acides aminés classiquement utilisés pour définir les groupes de petites protéines sécrétées comme effecteurs candidats alors que plusieurs effecteurs de plus de 300 acides aminés ont été rapportés. Récemment, une approche de prédiction par apprentissage automatique (*machine learning*) appelée EffectorP a été développée pour la détection d'effecteurs candidats au sein des sécrétomes prédits (Sperschneider *et al.,* 2016). EffectorP a été spécifiquement entrainé sur des effecteurs fongiques caractérisés et est donc dédié à la prédiction d'effecteurs au sein des sécrétomes fongiques (Sperschneider *et al.,* 2016). Les effecteurs prédits sont assignés selon leur probabilité à être des effecteurs. Cette approche a été utilisée pour définir les catalogues d'effecteurs candidats dans la version 2 du génome de *M. larici-populina* et dans le génome de *M. allii-populina* (**Chapitre III**). Cette approche reste cependant une prédiction basée sur des critères observés et aura donc tendance à exclure des effecteurs dont les caractéristiques de composition sont différentes. Ainsi, les effecteurs PGTAUSPE-10-1 de *P. graminis* f. sp. *tritici* et AvrM de *M. lini* ont été prédit comme n'étant pas des effecteurs par EffectorP, montrant les limites de cet outil (Sperschneider *et al.,* 2016). Avec la caractérisation fonctionnelle d'un plus large nombre d'effecteurs, cet outil devrait toutefois pouvoir améliorer ses performances. Ainsi une version 2 du programme de prédiction EffectorP vient juste d'être mise à disposition de la communauté (Sperschneider *et al.,* 2018). Dans la continuité, deux autres outils basés sur de l'apprentissage automatique ont été développés pour aider la sélection d'effecteurs candidats présentant une localisation particulière. D'une part, ApoplastP permet de distinguer des effecteurs candidats apoplastiques et cytoplasmiques (Sperschneider *et al.,* 2017c).

D’autre part, LOCALIZER permet de prédire une localisation subcellulaire potentielle avec une attention particulière pour la prédiction dans les chloroplastes, dans les mitochondries et dans le noyau (Sperschneider *et al.,* 2017b). Ces outils récents sont autant d'atouts supplémentaires pour affiner les listes d’effecteurs candidats au sein des larges sécrétomes des Pucciniales. LOCALIZER met en évidence la présence d'une proportion non négligeable d’effecteurs candidats ciblant potentiellement les chloroplastes (ex. 282 et 313 protéines sécrétées prédites comme adressées aux chloroplastes respectivement chez *M. larici-populina* et *M. allii-populina*). Un tel résultat suggèrerait que les chloroplastes représentent une cible importante lors de l’infection des plantes hôtes par les Pucciniales.

3. L’architecture des génomes de *M. larici-populina* et *M. allii-populina* : invasion par les éléments transposables (TE)

M. allii-populina et *M. larici-populina* partagent le même hôte télien et sont deux espèces phylogénétiquement proches (Vialle *et al.*, 2013). Elles sont aussi très proches de *M. lini* l'agent de la rouille du lin qui n'infecte qu'un seul hôte en tant qu'espèce autoïque (Lawrence *et al.*, 2007). Malgré un nombre de gènes prédits relativement proche pour ces différentes espèces (entre 20000 et 24000 gènes), on note des différences de taille de génome notables avec 110 Mb pour *M. larici-populina*, 220 Mb estimés chez *M. lini* et 336Mb chez *M. allii-populina*. Cette différence est due à une invasion de TE dans le génome de *M. allii-populina* qui représentent 77% du génome (contre 42% chez *M. larici-populina*). Une telle proportion de TE a été relevée pour la plupart des espèces de Pucciniales dont le génome a été séquencé (Aime *et al.*, 2017). Cette observation soulève des questions quant à l'impact des éléments transposables sur l’architecture des génomes chez les Pucciniales et leur évolution. De telles expansions de la taille des génomes est aussi observée chez d'autres organismes phytopathogènes filamenteux comme chez certains oomycètes du clade des *Phytophthora* avec des génomes de 220 à 280 Mb ou encore chez les champignons ascomycètes responsables de l’oïdium comme *Golovinomyces orontii* avec un génome de 160Mb ou encore l’ascomycète *Tuber melanosporum* avec 125 Mb (Haas *et al.,* 2009, Spanu *et al.,* 2010, Martin *et al.,* 2010, Duplessis *et al.*, 2013). Il apparait que les pathogènes filamenteux biotrophes obligatoires ou présentant une phase biotrophe tendent à avoir des génomes de taille importante (Raffaele & Kamoun, 2012).

Les TE jouent un rôle crucial dans l’organisation des génomes fongiques. Grâce au séquençage de plus en plus d’espèces de champignons, il est apparu que l’invasion et l’expansion des TE ont affecté l’évolution des génomes de différentes façons. Des variations importantes du nombre de TE sont observées dans les génomes des champignons ainsi que

divers mécanismes de défense contre l'invasion de ces TE (ex. RIP *Repeat-Induced Point Mutation* ; MIP ; et *Quelling* qui utilise des ARN interférants; Amselem *et al.,* 2015). Cependant, les TE peuvent également favoriser la diversité des gènes par rearrangement chromosomiques, délétions ou duplications (Raffaele & Kamoun, 2012). Il a été montré chez plusieurs espèces d'organismes phytopathogènes filamenteux, dont *P. infestans*, que des niveaux de diversité génétique associées aux TE représentaient un avantage adaptatif (Dong *et al.,* 2015). La combinaison d'approches génomiques et fonctionnelles ont aussi confirmé l'importance des TE dans la création de nouveaux gènes et la diversification des effecteurs, ainsi que dans l'expansion de familles multigéniques (Stukenbrock & Croll, 2014). Chez de nombreuses espèces d'oomycètes et de champignons phytopathogènes filamenteux dont les génomes sont riches en TE, la distribution des TE n'est pas uniforme. Ainsi, plusieurs études ont montré la localisation de gènes codant des effecteurs dans des régions riches en TE. Ces régions spécifiques paraissent déterminantes dans la spécificité d'hôte suggérant que ces îlots riches en éléments répétés participeraient à l'adaptation des pathogènes à leurs hôtes (de Jonge *et al.,* 2012 ; 2013 ; Dong *et al.,* 2015). Dans les génomes de *M. larici-populina* et de *M. allii-populina*, la répartition des TE est uniforme au sein de leurs génomes (Figure 28). Il en est de même pour les autres espèces de Pucciniales dont le génome a été séquencé (Aime *et al.*, 2017). De plus aucune corrélation entre la présence de TE à proximité de gènes codant des protéines sécrétées n'a pu être mise en évidence (Duplessis *et al.,* 2011a). Ces résultats sont en accord avec ce qui a récemment été montré dans les génomes de *P. striiformis* f. sp. *tritici* et de *Puccinia coronata* f. sp. *aveane* (Shwessinger *et al.,* 2017, Miller *et al.,* 2017). Il reste donc de nombreuses questions quant aux mécanismes qui ont permis aux TE d'envahir des génomes des Pucciniales, leur rôle potentiel dans les expansions de familles multigéniques largement observées dans ces génomes fongiques et leur éventuel avantage dans l'évolution et l'adaptation des Pucciniales à leurs hôtes.

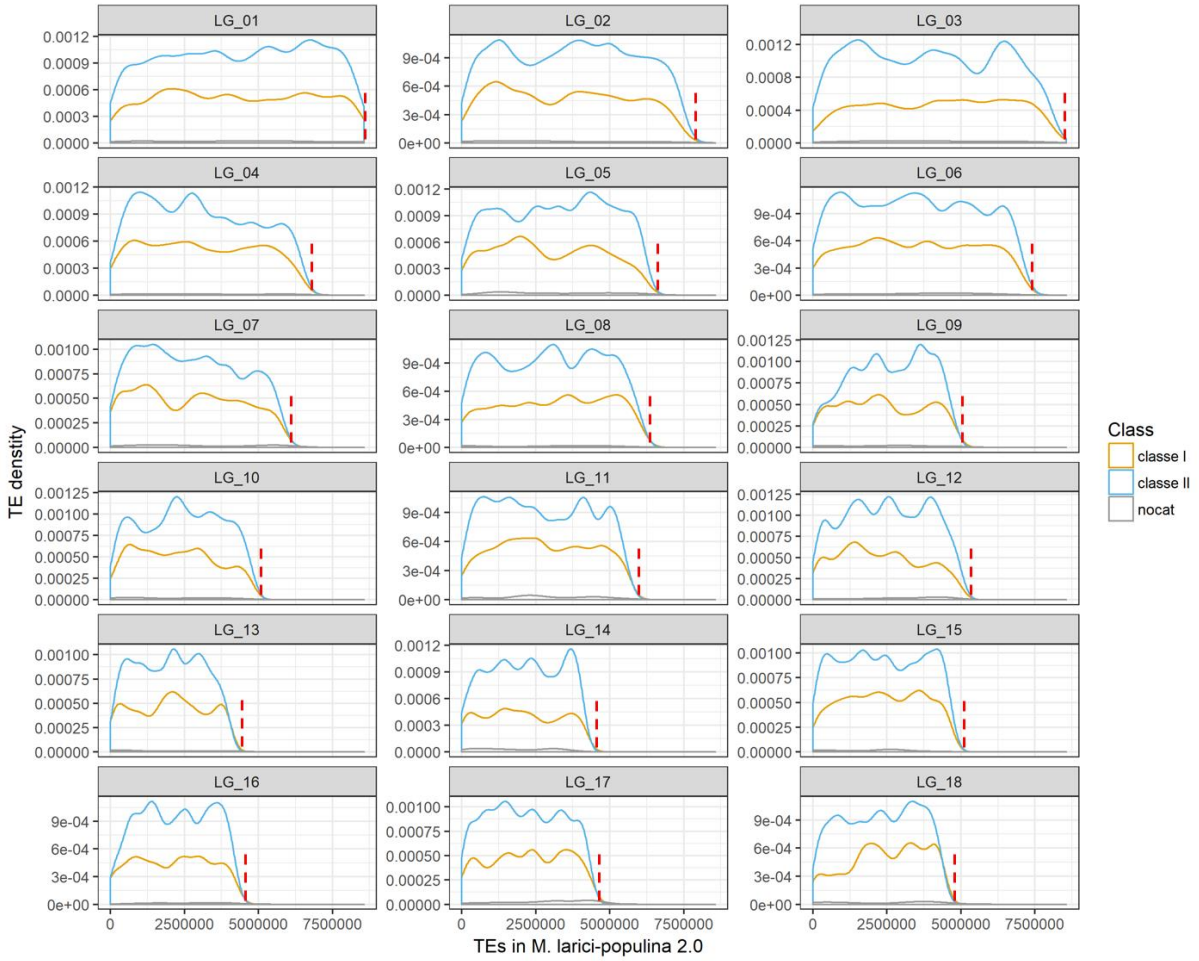


Figure 28 : Densité en TE sur les 18 groupes de liaison du génome de *M. larici-populina*.
Les axes des abscisses et des ordonnées représentent respectivement les groupes de liaison (taille en pb) et la densité en nombre de TE le long les groupes de liaison. En jaune sont représentés les TE de classe I ; en bleu ceux de classe II et en gris les TE non catégorisés. Les barres en pointillés rouges représentent la fin des groupes de liaison.

4. Vers la génomique comparative des Pucciniales

M. larici-populina et *M. allii-populina* sont donc des espèces proches qui partagent le même environnement et réalisent leurs cycles de multiplication végétative sur un même hôte. En France, *M. larici-populina* est plus agressif que *M. allii-populina* en plantations et certaines résistances développées contre *M. larici-populina* sont également actives sur *M. allii-populina*. La comparaison des génomes de ces deux espèces montre que malgré la présence de 3539 gènes supplémentaires chez *M. allii-populina*, les répertoires en CAZymes, protéases, transporteurs et protéines sécrétées sont relativement proches en proportion (432 CAZymes, 417 protéases, 341 transporteurs et 2177 protéines sécrétées pour *M. allii-populina* et 377 CAZymes, 345 protéases, 347 transporteurs et 2238 protéines sécrétées pour *M. larici-populina*). On retrouve également les mêmes familles en expansion pour les catégories des

CAZymes, des protéases et des transporteurs chez ces deux Melampsoraceae. Le sécrétome de *M. allii-populina* représente 9,4% du génome contre 11,4% pour celui de *M. larici-populina*. Les différences se trouvent dans la composition des familles avec des spécificités propres qui sont très certainement à mettre sur le compte de la spécificité d'hôte et qu'il reste encore à analyser dans le détail. Des différences sont plus notables dans les gènes de fonctions inconnues qu'ils codent des protéines sécrétées ou non.

L'estimation des ratios *dN/dS* entre les BRH des deux génomes génomes de *M. allii-populina* et *M. larici-populina* a permis de mettre en évidence des gènes montrant des signatures de sélection dont un part importante sont des petites protéines sécrétées. Il serait intéressant d'étendre cette analyse à d'autres Melampsoraceae comme *M. lini* qui est aussi très proche sur le plan phylogénétique. Cela permettrait aussi de voir si un effet est observé entre des génomes de Pucciniales hétéroïques et celui d'une espèce autoïque. En effet, *M. lini*, qui présente un cycle de vie dit macrocyclique avec cinq formes de spores comme les agents des rouilles foliaires du peuplier, réalise tout son cycle sur un hôte unique, le lin. Il est cependant important de noter que les génomes de *M. allii-populina* et de *M. lini* sont relativement imparfaits sur le plan de leurs assemblages avec respectivement plus de 3000 et de 20 000 *scaffolds*. Les profils d'expression des gènes mis à jour pourront être analysés chez *M. larici-populina*, afin de voir s'ils présentent une expression préférentielle lors de l'infection des hôtes télien et écidien. La même démarche pourra être réalisée chez *M. allii-populina*. Une telle démarche nécessite de réussir à réaliser en laboratoire les deux cycles de vie complet pour les deux espèces. Des expériences ont démarrées en ce sens à l'automne 2017 et devrait permettre d'être exploitée courant 2018. Ces quelques gènes candidats pourraient être testés prioritairement par RT-qPCR. L'analyse transcriptomique comparative sur les deux cycles de vie est envisagée à moyen terme.

Au-delà, un projet de recherche intitulé *Rust pangenomics* est en cours dans le cadre d'un consortium scientifique international en collaboration avec le JGI et a pour objet le séquençage de 50 génomes de Pucciniales appartenant à toutes les familles taxonomiques représentées au sein de cet ordre. Il est prévu d'obtenir les séquences des génomes mais aussi des transcriptomes chez les différents hôtes dans le cas des Pucciniales hétéroïques. Les pipelines uniformisés d'annotation de gènes, d'annotation des éléments transposables et de prédiction des sécrétomes qui sont à disposition seront systématiquement appliqués à ces espèces permettant de réaliser des analyses comparatives de meilleure qualité que lors de comparaison entre des génomes de source hétérogènes (Aime *et al.,* 2017).

Partie 3 : La biologie des effecteurs dans un pathosystème non-modèle : avancées et limites

1. Vers la caractérisation fonctionnelle des effecteurs des Pucciniales

L'analyse « post-génomique » de la biologie des effecteurs peut se découper en trois étapes principales : i) l'identification d'effecteurs candidats par des approches globales de génomique et transcriptomique (*effector-mining*), ii) mise en place de cribles effectoromiques *in planta* pour préciser les effecteurs candidats et iii) caractérisation fonctionnelle détaillée afin d'élucider les fonctions biochimique et biologique des effecteurs candidats (Alfano, 2009). La prédiction de larges catalogues de protéines sécrétées et l'identification d'effecteurs candidats exprimés au cours de l'infection par des approches de génomique ont permis de faire progresser nos connaissances sur la biologie des Pucciniales. Ainsi, les espèces représentant les menaces les plus sérieuses pour l'agriculture (les rouilles du blé *P. striiformis* f. sp. *tritici*, *P. graminis* f. sp. *tritici*, *P. triticina* ; la rouille du soja *P. pachyrizi* ; la rouille du caféier *H. vastatrix*) ont toutes des listes d'effecteurs candidats établies (**Article n°1 ; Article n°6**). Ces répertoires contiennent des centaines de gènes candidats représentant des pistes potentielles d'étude pour lutter contre les maladies des rouilles. En effet, certains candidats peuvent se montrer essentiels au processus infectieux et/ou représenter des cibles pour les récepteurs de l'immunité chez l'hôte (protéines d'avirulence). L'identification des cibles des effecteurs pourrait révéler les points de sensibilité chez l'hôte et ainsi permettre d'envisager de nouvelles méthodes de lutte par une sélection basée sur les cibles des effecteurs et non pas les moyens de les détecter (Moscou & van Esse, 2017). Toutefois, le nombre de gènes candidats à analyser en détail reste très conséquent chez les Pucciniales avec plusieurs centaines de gènes codant des petites protéines sécrétées exprimés lors de l'interaction pour chaque espèce (Aime *et al.,* 2017). Par ailleurs, ces champignons biotrophes obligatoires sont notables pour la difficulté de leur manipulation en laboratoire. En conséquence, des approches de criblage à moyen- et haut-débit ont été développées pour amorcer la caractérisation fonctionnelle de larges groupes de candidats et aider à l'émergence de nouvelles hypothèses quant à leurs possibles fonctions dans la cellule hôte. Ces approches dites effectoromiques ont pour but d'identifier les cibles cellulaires et moléculaires des effecteurs et elles reposent sur l'expression d'effecteurs candidats dans des systèmes hétérologues, comme *N. benthamiana* ou encore *A. thaliana*. Ces approches présentent l'avantage d'être rapides, peu couteuses et facilement reproductibles, avec toutefois quelques limites avec la génération de données non-informatives. Nous abordons les avantages et les limites de tels systèmes hétérologues dans le cadre d'un article de synthèse invité dans la revue *Curent Opinion in Microbiology* qui est présenté à la suite de cette discussion générale (cf. **Article n°6**).

Brièvement, ces systèmes sont indispensables pour l'avancée des connaissances de la biologie des effecteurs dans les pathosystèmes non-modèles comme les Pucciniales. Cependant, il faut les utiliser en ayant connaissance de leurs limitations en termes d'interprétation.

2. La biologie structurale appliquée aux effecteurs

Au-delà de la connaissance des cibles cellulaires et moléculaires des effecteurs candidats, il est possible d'étudier leur biochimie à travers leur production *in vivo* ou *in vitro*. Ainsi, l'identification d'une activité enzymatique ou d'interactions moléculaires peut permettre de contribuer à une meilleure compréhension de leur rôle dans l'interaction. La biologie structurale peut guider vers des approches de biochimie et de biophysique afin d'élucider leurs fonctions (Figure 29 ; Varden *et al.,* 2017), et la petite taille de la plupart des effecteurs candidats en font des cibles de choix pour la résolution de leur structure par RMN.

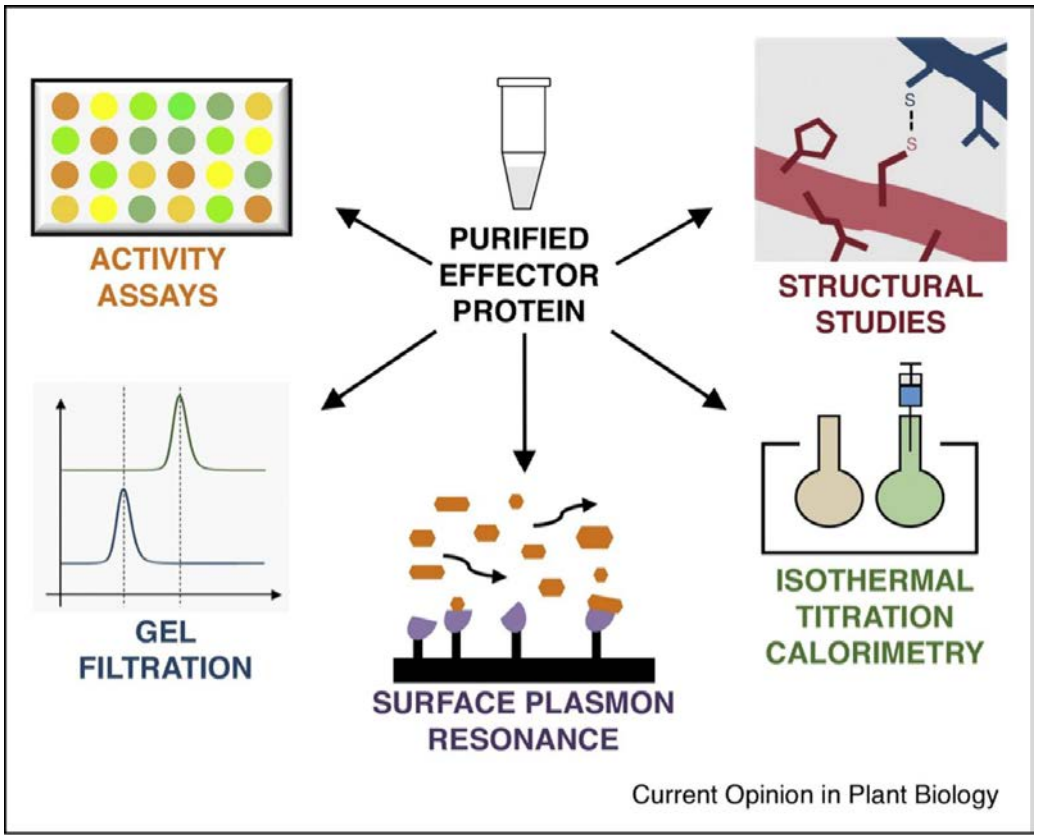


Figure 29 : Exemples d'approches *in vitro* applicables à la caractérisation des effecteurs.
De gauche à droite, à partir d'un effecteur protéique purifié : recherche d'activité enzymatique pour les effecteurs présentant des domaines enzymatiques prédits ; utilisation de méthodes de gel filtration pour déterminer le statut d'oligomérisation et tester la formation de complexes protéiques quantification des affinités d'interactions moléculaires par résonnance plasmonique de surface (SPR, *Surface plasmon resonance*) et titration calorimétrique isotherme (ITC, *Isothermal titration calorimetry*) ; analyse de la

structure des effecteurs par des approches de cristallographie, de RMN et de microscopie à cryo-microscopie électronique (cryo-EM) permettant la modélisation des structures 3D des effecteurs seuls ou en complexe. D'après Varden *et al.* (2017)

La condition *sine qua non* pour caractériser les effecteurs candidats par des approches biochimiques et biophysiques est de réussir à les produire et à les purifier en quantité suffisante. Pour cela, le système bactérien *E. coli* représente une approche de choix peu couteuse, rapide et avec de nombreuses souches à disposition pour optimiser la production. Une des limites de ce système est que les bactéries peuvent être récalcitrantes à la production de petites protéines riches en cystéines (Rosano & Ceccarelli, 2014; cf. **Article 4**). Ainsi, certains des effecteurs candidats que nous avons tenté de produire chez *E. coli* n'ont pas montré d'expression (cf. **Article 4**). Le système en levure *P. pastoris* représente un autre modèle biologique couramment utilisé pour la production de protéines recombinantes. Ce système eucaryote présente l'avantage de procéder à des modifications post-traductionnelles intéressantes pour des protéines d'origine fongique. De plus, *P. pastoris* excrète les protéines dans le milieu de culture ce qui réduit les étapes de purification. Cependant, nos tests de production pour l'effecteur MLPCTP1 (quatre constructions et 13 clones par construction testés) n'ont pas été plus concluants chez *P. pastoris* que chez *E. coli* (cf. **chapitre V**). Un système de production d'effecteurs candidats a récemment été rapporté chez *M. lini* (Zhang *et al.,* 2017a). Les auteurs ont testé la production de quatre effecteurs Avr (AvrP, AvrP123, AvrP4 et AvrM) dans plusieurs souches d'*E. coli*. Cette étude s'intéressait particulièrement au rôle des résidus cystéines dans la structuration des effecteurs de la rouille du lin. Chez les Pucciniales, seules trois structures d'effecteurs ont été résolues (AvrL567, AvrM et AvrP de *M. lini*). Ces trois effecteurs présentent des structures complètement différentes (Wang *et al.,* 2007 ; Ve *et al.,* 2013 ; Zhang *et al.,* 2017b). Cependant, toutes montrent un polymorphisme de surface impliqué dans la reconnaissance par les protéines de résistances de l'hôte, respectivement L5/L6/L7, M et P (Wang *et al.,* 2007 ; Ve *et al.,* 2013 ; Zhang *et al.,* 2017b). Outre leurs propriétés d'avirulence qui ont été décrites et analysées en détail (Ellis *et al.,* 2007), les fonctions de ces effecteurs dans le processus d'infection restent inconnues. Toutefois, la résolution de la structure d'AvrL567 a permis de montrer une activité de liaison aux acides nucléiques *in vitro* (Wang *et al.,* 2007).

Ces trois effecteurs de *M. lini* ne présentent pas à ce jour de similarité de structure avec des protéines connues ou avec d'autres effecteurs. Au contraire, nous avons pu mettre en évidence pour MLP124017 une homologie structurale avec deux protéines bactériennes identifiées comme *Nuclear-transport fator 2 like* alors que nous ne disposions d'aucune information d'homologie sur la base des séquences nucléotidique ou protéique. Aussi,

MLP124266, qui est un homologue de l'effecteur AvrP4 de *M. lini*, appartient à la large classe des *Knottins*, un groupe de protéines présentant un motif de type cystine-knot (cf. **Article n°4**). Ces résultats d'analyse des premières structures d'effecteurs chez l'agent de la rouille du peuplier sont prometteurs et montrent que la connaissance de leurs structures peut apporter des éléments d'information supplémentaires pouvant contribuer à poser des hypothèses pour étudier plus avant leur rôle dans l'interaction moléculaire avec la plante hôte. L'identification de l'organisation structurale des effecteurs MAX de *Magnaporthe oryzae* (« *Magnaporthe Avrs and ToxB like* ») qui ne partagent aucune homologie de séquence mais qui montrent des structures similaires, a démontré l'importance des apports de la biologie structurale à l'étude des effecteurs (de Guillen *et al.,* 2016). Les données structurales sur les effecteurs sont encore relativement limitées. Si l'on considère que tous les effecteurs de micro-organismes phytopathogènes ciblent des protéines plus ou moins conservées dans les cellules végétales, la résolution de nouvelles structures d'effecteurs permet d'amender les bases de données et contribue ainsi à améliorer l'état général des connaissances du domaine.

3. Les effecteurs chloroplastiques

L'analyse fonctionnelle des effecteurs de *M. larici-populina* a permis de mettre en évidence la première famille d'effecteurs fongiques ciblant les chloroplastes en mimant les peptides d'adressage des plantes (Oliveira-garcia & Valent, 2015 ; Petre *et al.,* 2016). Peu d'effecteurs candidats ciblant les chloroplastes ont été décrit chez les bactéries (Jelenska *et al.,* 2007, Rodríguez-Herva *et al.,* 2012, Li *et al.,* 2014, Petre *et al.,* 2015, de Torres Zabala *et al.,* 2015). La toxine protéique ToxA du champignon nécrotrophe *Pyrenophora tritici-repentis* cible aussi les chloroplastes (Manning *et al.,* 2007). Certaines localisations sont discutées, par exemple Jelenska et al. (2007) décrivent la localisation de HopI1 dans les chloroplastes des cellules de garde des stomates de *N. benthamiana*, alors que les cellules de gardes ne sont pas transformées lors d'essais d'agroinfiltration. Plusieurs études montrent des localisations différentes pour HopK1 et AvrRps4 (Liu *et al.,* 2014 ; (Heidrich *et al.,* 2011, Bhattacharjee *et al.,* 2011, Sohn *et al.,* 2012). La plupart des cribles de localisation d'effecteurs candidats conduits chez des oomycètes ou certaines espèces de Pucciniales ne montrent pas de localisation chloroplastique (Caillaud *et al.,* 2012 ; Schornach *et al.,* 2010 ; Qi *et al.,* 2017). Cette observation peut s'expliquer par le fait que la plupart de ces études de localisation ont été réalisées avec des protéines fluorescentes fusionnées dans la région N-terminale des effecteurs candidats. Ces fusions peuvent ainsi empêcher la reconnaissance des peptides de transit. Enfin, jusqu'à très récemment, les outils de prédiction des peptides de transits n'étaient pas forcément adaptés à la prédiction dans les génomes de pathogènes filamenteux, ce que corrige le prédicteur LOCALIZER (Sperschneider *et al.,* 2017c). Il est à noter que plus d'une

centaine d'effecteurs candidats de *M. larici-populina* et *M. allii-populina* ont des peptides de transits prédits. Les rôles des effecteurs ciblant les chloroplastes restent encore à élucider. Aucun interactant protéique robuste n'a pu être identifié pour MLPCTP1, rendant hasardeuse la formulation d'hypothèses quant à la fonction de cet effecteur candidat. De plus, les mutants de peupliers exprimant de manière stable MLPCTP1 en fusion avec la GFP avec une localisation effective au niveau des chloroplastes, ne présentent pas de phénotype visible particulier. Outre leur rôle central dans la production de photosynthétats, les chloroplastes jouent également un rôle dans l'immunité des plantes, notamment au niveau de la signalisation hormonale et ils représentent donc des cibles de choix pour les effecteurs des pathogènes (Torres-Zabala *et al.,* 2015). Il est à noter que dans le cadre de l'interaction avec le peuplier, après une colonisation rapide du parenchyme lacuneux des feuilles infectées par *M. larici-populina* (moins de 24h après inoculation), les hyphes d'infection gagnent très rapidement le parenchyme palissadique. Ce dernier consiste chez le cultivar de peuplier Beaupré en deux assises de cellules jointives très denses et très riches en chloroplastes. Le champignon *M. larici-populina* forme de nombreux haustoria dans cette zone afin de détourner les nutriments des cellules de l'hôte. L'hypothèse d'une prise de contrôle de la machinerie photosynthétique par l'arsenal d'effecteurs du champignon est une piste de travail séduisante qu'il reste à explorer. Au vu du large nombre d'effecteurs prédits pour cibler cet organite cellulaire chez différents champignons (Sperschneider *et al.,* 2017c), il sera intéressant de déterminer s'il existe une convergence évolutive sur des mêmes fonctions cibles et avec une même finalité chez ces différents organismes phytopathogènes et s'il existe une stratégie d'infection partagée chez les Pucciniales.

4. La caractérisation fonctionnelle des effecteurs: la partie émergée de l'iceberg

Ces dix dernières années, les effecteurs ont été le centre d'attention des interactions moléculaires entre plantes et micro-organismes. De nombreuses avancées ont permis de faire évoluer le concept d'effecteur depuis leur définition comme protéine ayant pour rôle de perturber le système immunitaire, à une vision plus large de molécules qui promeuvent la colonisation du parasite à travers la modulation de la structure et des fonctions des cellules végétales infectées. Tout un pan de recherche s'est appliqué à élucider les fonctions des effecteurs chez les bactéries et les pathogènes filamenteux. Un large nombre d'effecteurs des Pucciniales ont été caractérisés pour leurs propriétés d'avirulence et encore peu rapportent leurs fonctions premières dans l'infection (cf. **Article 1**).

Un large nombre de travaux chez les Pucciniales sont basés sur l'utilisation de systèmes hétérologues, limitant ainsi les conclusions quant aux fonctions des effecteurs candidats dans

leurs pathosystèmes natifs. De nombreuses études portant sur la caractérisation fonctionnelle d'effecteurs candidats accumulent des résultats partiels qui ne conduisent pas nécessairement à une publication, à l'image des résultats que j'ai obtenus au cours de ma thèse, mais aussi de ceux de la thèse de mes prédécesseur Benjamin Petre (Petre, 2012) et Stéphane Hacquard (Hacquard, 2010) sur le pathosystème peuplier-rouille du peuplier. J'ai pu échanger sur la difficulté d'appréhender le rôle des effecteurs avec d'autres doctorants et des chercheurs lors de congrès ou de rencontres du réseau de chercheurs *Effectome,* s'intéressant à la biologie des effecteurs dans différents pathosystèmes, et ce ressenti m'a semblé assez général. Cela traduit bien la difficulté que représente l'étude des effecteurs dans les interactions plantes-microorganismes et souligne les efforts expérimentaux encore à produire pour réussir à aborder leur rôle chez des organismes complexes tels que les pathogènes biotrophes obligatoires que sont les rouilles.

Article n°6: Show me the way: rust effector targets in heterologous plant systems

Cécile Lorrain, Benjamin Petre, Sébastien Duplessis

Article de synthèse invité, accepté dans la revue internationale *Current Opinion in Microbiology*

Show me the way: rust effector targets in heterologous plant systems

Cécile Lorrain[1], Benjamin Petre[1,2], Sébastien Duplessis[1]

Address

[1] INRA/Université de Lorraine, UMR 1136 Interactions Arbres/Microorganismes, INRA Centre Grand Est-Nancy, Champenoux, France

[2] The Sainsbury Laboratory, Norwich Research Park, Norwich, United Kingdom

Corresponding author: Duplessis, Sébastien (sebastien.duplessis@inra.fr)

Abstract

For years, the study of rust fungal effectors has been impeded by the lack of molecular genetic tools in rust pathosystems. The recent use of heterologous plants to perform effector screens (effectoromics) -including effector localisation (cellular targets) and protein interactors (molecular targets) in plant cells- has changed the game. These screens revealed that many candidate effectors from various rust fungi target specific plant cell compartments, including chloroplasts, and associate with specific plant protein complexes. Such information represents unparalleled opportunities to understand how effectors sustain extreme parasitic interactions and obligate biotrophy. Despite their limitations, we here portray how the use of heterologous expression systems has been essential for gaining new insight into rust effectors.

Introduction

Rust fungi (Pucciniales) are the largest group of obligate biotrophic plant pathogens across the fungal kingdom, which threatens agriculture and human food production [1, 2, 3]. They feed exclusively on their host plants and they cannot be grown in laboratory conditions. Rust fungi also have complex life cycles; they often alternate on host plants from different taxa and produce different spore stages [1, 4]. Most agronomic hosts of rust fungi are not amenable to genetic transformation. All these features impair research progress in rust fungal biology.

Plant pathogens deliver effector proteins into host tissues to promote infection by interfering with host structures and functions [5]. Effector biology is a central theme in molecular plant pathology; it aims to understand how pathogens establish successful infection and to support the development of resistant plants [6, 7]. To these ends, many effectors have been studied in model pathogenic bacteria and filamentous pathogens (i.e. fungi and oomycetes) [8, 9, 10]. Computational tools have been designed to identify candidate secreted effector proteins (CSEPs) in filamentous pathogens and to predict their localization in host tissues (i.e. ApoplatP, EffectorP, LOCALIZER) [11, 12•, 13•, 14]. In rust fungi, transcriptomics of host infection has helped prioritize lists of CSEPs, which were further analysed with dedicated analytical pipelines (e.g. [15•, 16•], for details see [17, 18]).

Here, we summarize recent effector screens (so called 'effectoromics' approaches) performed in the non-host plants *Arabidopsis thaliana* and *Nicotiana benthamiana* to determine localization of rust fungal CSEPs. We also discuss the pros and cons of using non-host systems.

Heterologous systems enable fast-forward effector screens (effectoromics)

Effectoromic screens are performed to gain knowledge about CSEPs, in order to further prioritize their analysis and formulate hypotheses about their function. Several time- and cost-effective assays exist to perform effectoromics [19, 20, 21, 22]. These assays often rely on the heterologous expression of tagged effectors (without their predicted secretion signals) directly inside leaf cells of model plant species such as *A. thaliana* or *N. benthamiana*. For oomycete CSEPs, screens in heterologous plant systems provided relevant information about specific localisation in plant cells and putative plant interactors [24, 25, 26]. Such information was key to draw hypotheses regarding the function of specific effectors and move forward with mechanistic studies. Collectively, these studies showed how effectors manipulate host structures and functions, and significantly moved forward our understanding of the molecular interactions between pathogens and plants [8, 10].

Despite its recognizable benefits, the use of non-host systems for effector screens is sometimes debated or criticized during informal discussions at scientific meetings. Notably,

'artificial' screens can lead to spurious effector localisation and interactors (see pros and cons in Box 1). Nevertheless, in the case of pathosystems with limited methodological options to study CSEPs, we argue that the use of heterologous plants is a reasonable way to circumvent bottlenecks and to move forward. Noteworthy, by 'moving forward' we mean to get the information needed before moving back to the native system to tackle biologically-relevant questions.

Rust candidate effectors target multiple plant cell compartments

Over the past two years, four effector screens identified cellular targets of CSEPs for three different rust species: the poplar leaf rust fungus *Melampsora larici-populina* [27••, 28••], the wheat yellow rust fungus *Puccinia striiformis* f. sp. *tritici* [29••], and the Asian soybean rust fungus *Phakopsora pachyrhizi* [30••]. All screens were based on the expression of effector-GFP fusions (effectors were cloned without their signal peptide) in heterologous plant systems (*A. thaliana* or *N. benthamiana*). All studies reported that rust CSEPs target multiple plant cell compartments (Figure 1; Table 1). For the *M. larici-populina* pathosystem, 29 CSEPs were assayed either in stably transformed *A. thaliana* plants [28••] or in transiently transformed *N. benthamiana* leaves [27••]. Among the 29 CSEPs tested, nine showed an informative subcellular localisation (i.e. different from free GFP) in the nuclei, nucleolus, nuclear bodies, plasma membrane, cytosolic bodies, plasmodesmata, chloroplasts, and mitochondria. Five *M. larici-populina* CSEPs showed a similar localization in the two different heterologous systems (Table 1) [27••, 28••]. Sixteen CSEPs from *P. striiformis* f. sp. *tritici* were assayed in transiently transformed *N. benthamiana* leaves; among which seven showed informative localization in cellular bodies, nucleus, chloroplasts, and endomembranes [29••]. Finally, 82 CSEPs from *Phakopsora pachyrhizi* were assayed in transiently transformed *N. benthamiana* leaves [30••]. A total of 30 CSEPs were assigned to the nucleus (17), the nucleolus (1) or the cytosol (12); in addition, 16 CSEPs accumulated in aggregates of unknown nature, which could be organelles or artefacts (Table 1) [30••]. In this study, a particular focus was given to 17 CSEPs that showed complementary immune-related functions (i.e. immune suppression related phenotypes). Among these, 10 showed informative localisation in the cytosol or the nucleus (pictured in Figure 1), which helps prioritization for future functional studies. Beside these recent screens, separate studies were conducted with a single or a small set of CSEPs (from *M. larici-populina* [31, 32••], the flax rust fungus *M. lini* [32••, 33], the wheat stem rust fungus *P. graminis* f. sp. *tritici* [13•], *P. striiformis* f. sp. *tritici* [34-38] and *P. pachyrhizi* [39•, 40]). These studies also identified plant cell compartments (such as the plasma membrane, the nucleus, the nucleolus, and chloroplasts) targeted by CSEPs. Together, these studies indicate that rust fungi likely address suites of effectors to different subcellular locations in host cells.

Box 1. Pros and Cons of heterologous protein expression systems

Heterologous expression means that a gene is expressed outside of its natural context, in a 'foreign' host system (often a model system). For instance, a fungal gene (e.g. coding a candidate effector) is expressed in plant cells (e.g. the leaf cells of *Nicotiana benthamiana*). Such approaches are often performed to take advantage of the analytical tools available in the heterologous host system, and which are not available in the organism from which the gene originates. Here, we present important pros and cons to be considered when using heterologous protein expression systems and associated assays.

Pros

- **Assays in model systems are reliable**. A large community of researchers developed and optimized assays in model systems used for heterologous protein expression. The drawbacks and pitfalls for data interpretation are known. This is particularly true for protein-protein interaction assays.
- **Using model systems is fast, cheap, and easy.** The information needed can be generated easily and quickly in a cost-effective manner, which is ideal for performing screens.
- **Everyone can revisit the findings.** Heterologous systems promote reproducibility, as they are more accessible to a wide community than more emerging systems. Other researchers can reproduce experiments, and confirm, build on, or challenge them.

Cons

- **Limited conclusion can be drawn regarding biological significance**. When using a heterologous system, the biological significance of the observed phenomenon cannot be tested. Thus conclusions are restricted to molecular or cellular mechanisms.
- **Data may be non-informative**. Heterologous systems may not be suitable for assaying molecules or process that occur only in specific biological settings. For instance, heterologous systems may lack critical components found in native systems, many proteins may not fold properly, will not be in the right cellular context, will not find their molecular targets, …etc
- **You get spoiled.** Getting used to work with heterologous systems may lead to less investment of time and energy to develop methods and tools for the original systems studied. Going back and forth to the studied systems is necessary.

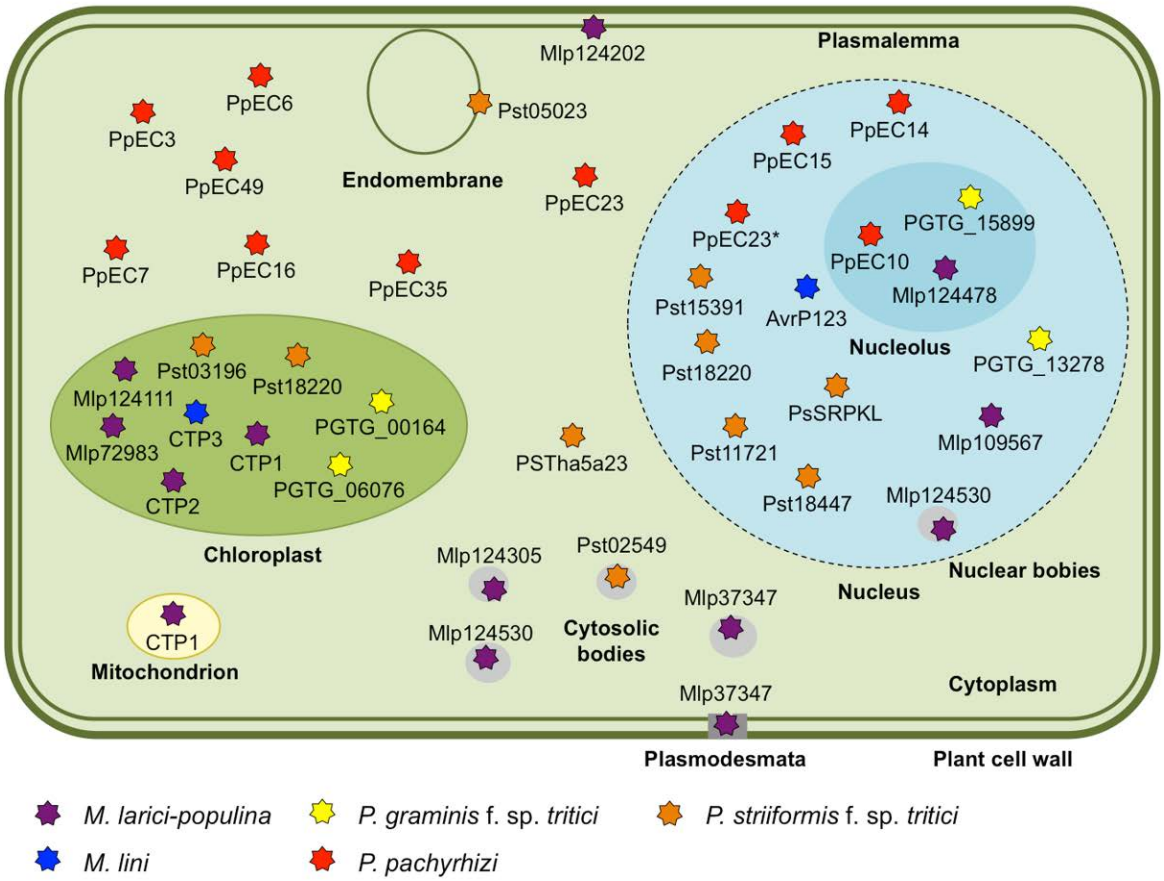


Figure 1: Localization of rust effectors in cell compartments of heterologous plants.

Subcellular localization of CSEPs determined by expression in heterologous plant systems: *Arabidopsis thaliana* and *Nicotiana benthamiana* for *Melampsora larici-populina*; and *N. benthamiana* for *Melampsora lini*, *Phakopsora pachyrhizi*, *Puccinia graminis* f. sp. *tritici*, and *Puccinia striiformis* f. sp. *tritici*. Only 10 *P. pachyrhizi* CSEPs that showed complementary immune-related functions in [30••] are pictured. *P. pachyrhizi* PpEC23* localizes in the nucleus when co-expressed with its interactor (transcription repressor GmSPL21). CSEPs presenting localization in the nucleus and the cytosol similar to GFP expressed in heterologous systems are not represented in this figure.

Multiple rust candidate effectors target chloroplasts

Chloroplast Targeting Protein (CTP1) is an effector from *M. larici-populina* that targets chloroplasts [27••]. Detailed analysis of CTP1 in *N. benthamiana* has revealed that it possesses an N-terminal region mimicking plant transit peptides, which is sufficient and necessary for the accumulation of CTP1 in the stroma of chloroplasts [32••]. CTP1 belong to a small gene family specific to *Melampsora* species; two other members of the family, CTP2, and CTP3, also possess a transit peptide for translocation in chloroplasts [32••]. The presence of a transit peptide suggests that this effector family subverts the plant cell machinery to enter chloroplasts. How this transit peptide evolved, and how CTP1 functions in chloroplasts are still unknown.

In addition to the CTP1 family, other effectors targeting chloroplasts have been identified in three other rust species [13•, 28••, 29••, 32••]. LOCALIZER - a recent computational method that predicts effector localisation in chloroplasts, mitochondria and nuclei - has predicted other chloroplast-targeting CSEPs in rust species [13•]. Two such CSEPs from *P. graminis* f. sp. *tritici* were shown to translocate in chloroplasts when expressed in the cytosol of the leaf cells of *N. benthamiana* [13•]. From these experiments, it appears that the chloroplast may be a central target organelle for rust fungi. How rust effectors modulate chloroplast structure and function is a fascinating question that remains to be answered.

As a note, many subcellular localization screens use N-terminally-tagged effectors [24-26]. A N-terminal GFP may prevent the proper function of a transit peptide. As a result, many effector screens may have failed at revealing effectors targeted to chloroplasts. For instance, LOCALIZER predicted a chloroplastic location for several CSEPs of P. *pachyrhizi* [30••]. However, those CSEPs were assayed as N-terminally-tagged fusions. In the future, (re-)performing subcellular localization screens with CSEPs carrying a GFP tag at their C-Terminus may reveal additional effectors targeting chloroplasts.

Rust candidate effectors target multiple plant protein complexes

Immunoprecipitation of CSEP-GFP fusions from plant cell extracts coupled with liquid chromatography-tandem mass spectrometry (in short coIP/MS) is commonly used to identify plant protein complexes with which CSEPs associate [22]. CoIP/MS identified putative interactors in *N. benthamiana* for five and six CSEPs from *M. larici-populina* and *P. striiformis* f. sp. *tritici*, respectively (Table 1) [27••, 29••]. In both studies, one CSEP was shown to associate with the interactor homolog from the native host. This established proof-of-concept that heterologous plant systems can be used for CSEP protein interactor screens. The CoIP/MS method applied in these studies identified larger lists of plant proteins interacting with CSEPs. A scoring system based on specificity and peptide count was established to distinguish specific interactors that are relevant for further study from 'usual suspects' commonly found with many CSEPs [22, 27••, 29••]. Other studies have used yeast-two-hybrid screens to identify putative CSEP interactors (Table 1). For instance, in *P. pachyrhizi*, PpEC23 interacts with the soybean transcription repressor GmSPL12 [39•]. In other work, nucleocytosolic effectors from the wheat stripe rust fungus were shown to associate with specific plant proteins (Table 1) [36, 37]. To conclude, coIP/MS screens in heterologous systems permit the identification of plant interactors of CSEPs, and should be more often considered for CSEP-plant protein interaction screens.

Conclusion: the half full glass of rust biologists

In just two years, the field of rust effector biology has moved from a couple of well established effectors (mostly studied for their avirulence activities) [18] to a growing list of candidate effectors with valuable information regarding their putative function in plant cells. The use of heterologous plant systems has been key to this progress and paves the way for future rust effector screens. Future investigations will also be directed towards testing the biological significance of observations made in non-host plants, notably by assessing the ability of the CSEPs to suppress cell death and promote pathogen growth in the context of infection [28••, 30••]. Even if heterologous approaches present many pros and cons (Box 1), enough progress has been achieved to see the half full glass. Although it is reasonable to consider that some spurious data may result from assays in heterologous plant systems, it is a wager well worth taking for advancing fundamental knowledge of the biology of rust pathogens and potentially developing new opportunities for disease control.

Acknowledgements

The present review takes its roots at the Rust Workshop premeeting of the 29th Fungal Genetics Conference in Asilomar (USA), March 2017. We are grateful to all colleagues in the molecular phytopathology community for sharing ideas and concepts through the years, notably in the context of the French network Effectome supported by INRA (France). The fruitful interactions with the groups of Sophien Kamoun at The Sainsbury Laboratory (Norwich, UK) and Hugo Germain at the Université du Québec à Trois-Rivières (Québec, Canada) are particularly warmly acknowledged. CL and SD are supported by the French 'Investissements d'Avenir' program (ANR-11-LABX-0002-01, Lab of Excellence ARBRE). CL and BP were supported by a young scientist grant (CJS) from INRA.

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:
• of special interest
•• of outstanding interest

[1] Aime M.C., McTaggart A.R., Mondo S.J., Duplessis S. **Phylogenetics and phylogenomics of rust fungi**. *Advances in Genetics* vol 99 In Press (2017)

[2] Pennisi E. **Armed and dansgerous**. *Science*, 327:5967, (2010), pp. 804-805.

[3] Dean R., van Kan J. A. L., Pretorius Z. A., Hammond-Kosack K. E., di Pietro A., Spanu P. D., Rudd J. J., Dickman M., Kahmann R., Ellis J. and Foster G. D. **The Top 10 fungal pathogens in molecular plant pathology**. *Molecular Plant Pathology,* 13, (2012), pp. 414–430.

[4] Pernaci M., De Mita S., Andrieux A., Petrowski J., Halkett F., Duplessis S., and Frey P. **Genome-wide patterns of segregation and linkage disequilibrium: the construction of a linkage genetic map of the poplar rust fungus *Melampsora larici-populina***. *Frontiers in Plant Science*, 5 (2014), 454.

[5] Win J., Chaparro-Garcia A., Belhaj K., Saunders D.G., Yoshida K., Dong S., Schornack S., Zipfel C., Robatzek S., Hogenhout S.A. and Kamoun S. **Effector biology of plant-associated organisms: concepts and perspectives**. *Cold Spring Harb Symp Quant Biol* 77, (2012)

[6] Hogenhout S.A., Van der Hoorn R.A.L., Terauchi R., and Kamoun S. **Emerging concepts in effector biology of plant-associated organisms**. *Mol. Plant-Microbe Interact. MPMI* 22, (2009), pp. 115–122.

[7] Dangl J.L., Horvath D.M., and Staskawicz B.J. **Pivoting the plant immune system from dissection to deployment**. *Science* 341, (2013), pp. 746–751.

[8] Tyler B.M., and Rouxel T. **Effectors of Fungi and Oomycetes: Their Virulence and Avirulence Functions and Translocation From Pathogen to Host Cells**. *Molecular Plant Immunity*, Wiley-Blackwell, (2012), pp. 123–167.

[9] Deslandes L., and Rivas S. **Catch me if you can: bacterial effectors and plant targets.** *Trends Plant Sci.* 17, (2012), pp. 644–655

[10] Lo Presti L., Lanver D., Schweizer G., Tanaka S., Liang L., Tollot M., Zuccaro A., Reissmann S., and Kahmann R. **Fungal effectors and plant susceptibility**. *Annu. Rev. Plant Biol*. 66, (2015) pp. 513–545.

[11] Sperschneider J., Taylor J., Dodds P.N., and Duplessis S. **Computational methods for predicting effectors in rust pathogens**. *Wheat Rust Disease - Methods and Protocols*, Springer in press, (2017a).

[12•] Sperschneider J., Gardiner D.M., Dodds P.N., Tini F., Covarelli L., Singh K.B., Manners, J.M., and Taylor J.M. **EffectorP: predicting fungal effector proteins from secretomes using machine learning**. *New Phytologist*. 210, (2016), pp. 743–761.
This study presents EffectorP a machine learning approach designed to predict fungal effectors based on compositional features that discriminate effectors from non-effectors among secreted proteins.

[13•] Sperschneider J., Catanzariti A.M., DeBoer K., Petre B., Gardiner D.M., Singh K.B., Dodds P.N., and Taylor J.M. **LOCALIZER: subcellular localization prediction of both plant and effector proteins in the plant cell**. *Scientifique Report* 7, (2017b) 44598.
LOCALIZER is a machine-learning tool designed to predict subcellular localization of plant proteins and plant pathogen effectors. LOCALIZER is a valuable tool to help prioritization of fungal candidate effectors.

[14] Sperschneider J., Dodds P.N., Singh K.B., and Taylor J.M. **ApoplastP: prediction of effectors and plant proteins in the apoplast using machine learning**. *bioRxiv*, (2017c)

[15•] Hacquard S., Joly D.L., Lin Y.-C., Tisserant E., Feau N., Delaruelle C., Legue V., Kohler A., Tanguay P., Petre B., Frey P., Van de Peer Y., Rouzé P., Martin F., Hamelin R. C., and Duplessis S. **A comprehensive analysis of genes encoding small secreted proteins identifies candidate effectors in *Melampsora larici-populina* (poplar leaf rust)**. *Mol. Plant-Microbe Interact. MPMI* 25, (2012), pp. 279–293.
A detailed survey of *M. larici-populina* CSEPs based on genomic and transcriptomic information, including classification of manually curated CSEP gene families, identification of conserved motifs such as the [FYW]xC reported in cereal obligate biotrophs, identification of clusters of paralogous genes under selection. Immunolocalization of four CSEPs including orthologs of effectors previously reported in *Melampsora lini* and *Uromyces* spp. indicates a diversity of localization in the infected host leaves.

[16•] Saunders D.G.O., Win J., Cano L.M., Szabo L.J., Kamoun S., and Raffaele S. **Using hierarchical clustering of secreted protein families to classify and rank candidate effectors of rust fungi**. *PloS One,* 7, (2012), e29847.
This paper presents an elegant way of prioritizing candidate effectors in two rust fungi, *Puccinia striifomis* f. sp. *tritici* and *Melampsora larici-populina*, using a hierarchical classification of known effector features.

[17] Duplessis S., Bakkeren G., and Hamelin R. **Advancing knowledge on biology of rust fungi through genomics**. *Adv. Bot. Res.* 70, (2014)

[18] Petre B., Joly D.L., and Duplessis S. **Effector proteins of rust fungi**. *Front Plant Sci* 5, (2014)

[19] Alfano J.R. **Roadmap for future research on plant pathogen effectors**. *Mol. Plant Pathol*. 10, (2009), pp. 805–813

[20] Ma L., Lukasik E., Gawehns F., and Takken F.L.W. **The use of agroinfiltration for transient expression of plant resistance and fungal effector proteins in *Nicotiana benthamiana* leaves**. *Methods Mol. Biol. Clifton NJ* 835, (2012), pp. 61–74.

[21] Win J., Kamoun S., and Jones A.M.E. **Purification of effector-target protein complexes via transient expression in *Nicotiana benthamiana***. *Methods Mol. Biol. Clifton NJ* 712, (2011), pp.181–194

[22] Petre B., Win, J., Menke F.L.H., and Kamoun S. **Protein-Protein Interaction Assays with Effector-GFP Fusions in *Nicotiana benthamiana***. *Methods Mol. Biol. Clifton NJ* 1659, (2017), pp. 85–98.

[23] Petre B., and Kamoun S. **How do filamentous pathogens deliver effector proteins into plant cells?** *PLoS Biol*. 12, (2014), e1001801

[24] Caillaud M. C., Piquerez S.J.M., Fabro G., Steinbrenner J., Ishaque N., Beynon J., and Jones J.D.G. **Subcellular localization of the Hpa RxLR effector repertoire identifies a tonoplast-**

associated protein HaRxL17 that confers enhanced plant susceptibility. *Plant J. Cell Mol. Biol*. 69, (2012), pp. 252–265

[25] Schornack S., Van Damme M., Bozkurt T.O., Cano L.M., Smoker M., Thines M., Gaulin E., Kamoun S., and Huitema E. **Ancient class of translocated oomycete effectors targets the host nucleus**. *Proc. Natl. Acad. Sci. U.S.A.* 107, (2010), pp. 17421–17426

[26] Stam R., Jupe J., Howden A.J.M., Morris J.A., Boevink P.C., Hedley P.E., and Huitema E. **Identification and Characterisation CRN Effectors in Phytophthora capsici Shows Modularity and Functional Diversity**. *PloS ONE* 8, (2013), e59517.

[27••] Petre B., Saunders D.G.O., Sklenar J., Lorrain C., Win J., Duplessis S., and Kamoun S. **Candidate Effector Proteins of the Rust Pathogen Melampsora larici-populina Target Diverse Plant Cell Compartments**. *Mol. Plant-Microbe Interact. MPMI* 28, (2015), pp. 689–700
This paper presents an effectoromic pipeline based on transient expression of *M. larici-populina* CSEPs in *Nicotiana benthamiana* designed to determine subcellular localization in plant cell and potential plant interactors. Among 20 assayed CSEPs, nine accumulate in diverse cell compartments and/or present specific plant interactors. The interaction between MLP124017 and a poplar TOPLESS-related protein 4 was confirmed by co-immunoprecipitation in *N. benthamiana*. This paper reports a scoring system to distinguish specific effectors from conspicuous interactors, and establishes a proof-of-concept that heterologous plant systems can be used for rust effectoromics.

[28••] Germain H., Joly D.L., Mireault C., Plourde M.B., Letanneur C., Stewart D., Morency M. J., Petre B., Duplessis S., and Seguin A. **Infection assays in Arabidopsis reveal candidate effectors from the poplar rust fungus that promote susceptibility to bacteria and oomycete pathogens.** *Mol. Plant Pathol.,* (2016), 12514
This study reports an effectoromic pipeline based on the stable expression of 16 poplar rust CSEPs in *Arabidopsis thaliana* to assay subcellular localization and impact on bacterial and oomycete infection. Three CSEPs accumulate in diverse cell compartments and two promote leaf colonization of both bacteria and oomycete infection.

[29••] Petre B., Saunders D.G.O., Sklenar J., Lorrain C., Krasileva K.V., Win J., Duplessis S., and Kamoun S. **Heterologous Expression Screens in Nicotiana benthamiana Identify a Candidate Effector of the Wheat Yellow Rust Pathogen that Associates with Processing Bodies**. *PloS One* 11, (2016a), e0149035
This study applies the same effectoromic pipeline described in [27••] to determine the subcellular localization and the plant interactors of 16 CSEPs from the wheat yellow rust fungus *Puccinia striiformis* f. sp. *tritici*. Nine CSEPs accumulate in diverse cell compartments and/or interact with plant proteins in *Nicotiana benthamianai* leaf cells. One CSEP accumulates in processing bodies (P-bodies) and interacts with the P-body protein ENHANCER OF mRNA DECAPPING PROTEIN 4 from wheat (TaEDC4). This paper also illustrates how *N. benthamiana* (a dicot plant) can be used as a heterologous system for studying rust fungi infecting cereals (monocot plants).

[30••] Qi M., Grayczyk J.P., Seitz J.M., Lee Y., Link T.I., Choi D., Pedley K., Voegele R., Baum T., and Whitham S.A. **Suppression or activation of immune responses by predicted secreted proteins of the soybean rust pathogen *Phakopsora pachyrhizi***. *Mol. Plant. Microbe Interact.* (2017)
This paper reports a multi-screen effectoromic pipeline designed to pinpoint CSEPs of the Asian soybean rust *Phakopsora pachyrhizi*. Out of 156 predicted CSEPs, the authors assayed 82 CSEPs in various heterologous plant systems regarding their capacity to activate or suppress diverse immune responses and determined their subcellular localization in *Nicotiana benthamiana*.

[31] Gaouar O., Morency M.J., Letanneur C., Séguin A., and Germain H. **The 124202 candidate effector of *Melampsora larici-populina* interacts with membranes in *Nicotiana* and *Arabidopsis***. *Can. J. Plant Pathol*. 38, (2016), pp. 197–208.

[32••] Petre B., Lorrain C., Saunders D.G.O., Win J., Sklenar J., Duplessis S., and Kamoun S. **Rust fungal effectors mimic host transit peptides to translocate into chloroplasts**. *Cell. Microbiol.* 18, (2016b), pp. 453–465.

This study is a follow up of one poplar rust effector identified in [27••] targeting chloroplasts. It shows that the chloroplast-targeted protein 1 (CTP1) carries a transit peptide necessary and sufficient for accumulation in chloroplasts when expressed transiently in *Nicotiana benthamiana*, mimicking plant sorting signals. CTP1 is part of a Melampsoraceae-specific family and other members also accumulate in chloroplasts.

[33] Zhang X., Farah N., Rolston L., Ericsson D.J., Catanzariti A.M., Bernoux M., Ve T., Bendak K., Chen C., Mackay J.P., Lawrence G.J., Hardham A., Ellis J.G., Williams S.J., Dodds P.N., Jones D.A., Kobe B. **Crystal structure of the *Melampsora lini* effector AvrP reveals insights into a possible nuclear function and recognition by the flax disease resistance protein P**. *Mol. Plant Pathol.,* (2017), 12597

[34] Cheng Y., Wang X., Yao J., Voegele R.T., Zhang Y., Wang W., Huang L., and Kang Z. **Characterization of protein kinase PsSRPKL, a novel pathogenicity factor in the wheat stripe rust fungus.** *Environ. Microbiol.* 17, (2015), pp. 2601–2617

[35] Cheng Y., Wu K., Yao J., Li S., Wang X., Huang L., and Kang Z. **PSTha5a23, a candidate effector from the obligate biotrophic pathogen *Puccinia striiformis* f. sp. *tritici*, is involved in plant defense suppression and rust pathogenicity**. *Environ. Microbiol*., 19, (2016), pp. 1717-1729

[36] Liu C., Pedersen C., Schultz-Larsen T., Aguilar G.B., Madriz-Ordenana K., Hovmoller M.S., and Thordal-Christensen H. **The stripe rust fungal effector PEC6 suppresses pattern-triggered immunity in a host species-independent manner and interacts with adenosine kinases**. *New Phytol.,* (2016), 14034

[37] Wang X., Yang B., Li K., Kang Z., Cantu D., and Dubcovsky J. **A Conserved *Puccinia striiformis* Protein Interacts with Wheat NPR1 and Reduces Induction of Pathogenesis-Related Genes in Response to Pathogens**. *Mol. Plant-Microbe Interact. MPMI* 29, (2016), pp. 977–989.

[38] Dagvadorj B., Ozketen A.C., Andac A., Duggan C., Bozkurt T.O., and Akkaya M.S. **A *Puccinia striiformis* f. sp. *tritici* secreted protein activates plant immunity at the cell surface.** *Sci. Rep.* 7, (2014), 1141

[39•] Qi M., Link T.I., Muller M., Hirschburger D., Pudake R.N., Pedley K.F., Braun E., Voegele R.T., Baum T.J., and Whitham S.A. **A Small Cysteine-Rich Protein from the Asian Soybean Rust Fungus, *Phakopsora pachyrhizi*, Suppresses Plant Immunity**. *PLoS Pathog*. 12, (2016), e1005827.

This work describes the functional analysis of the CSEPs PpEC23 from the Asian soybean rust *Phakopsora pachyrhizi* that suppress plant immunity. PpEC23 with GmSPL12l, a transcription repressor of soybean immunity, associate in yeast and in the nuclei of *N. benthamiana* leaf cells.

[40] De Carvalho M.C. da C.G., Nascimento L.C., Darben L.M., Polizel-Podanosqui A.M., Lopes-Caitar V.S., Qi M., Rocha C.S., Carazzolle M.F., Kuwahara M.K., Pereira G.A.G., Abdelnoor R.V.,

Whitham S.A., Marcelino-Guimaraes F.C. (2016). **Prediction of the *in planta Phakopsora pachyrhizi* secretome and potential effector families**. *Mol. Plant Pathol.,* 18, (2017), 12405

[41] Wang C.I.A., Guncar G., Forwood J.K., Teh T., Catanzariti A.M., Lawrence G.J., Loughlin F.E., Mackay J.P., Schirra H.J., Anderson P.A., Ellis J.G, Dodds P.N., Kobe B. **Crystal structures of flax rust avirulence proteins AvrL567-A and -D reveal details of the structural basis for flax disease resistance specificity**. *Plant Cell* 19, (2007), pp. 2898–2912.

CONCLUSIONS & PERSPECTIVES

En 2011, *M. larici-populina* et *P. graminis* f. sp. *tritici* étaient les premières espèces de Pucciniales dont le génome était séquencé. On compte désormais 11 espèces avec un génome de référence, ce qui représente moins de 0,1% de l'ordre des Pucciniales qui comprend plus de 8000 espèces. Ces travaux de thèse montrent que la comparaison entre deux espèces proches peut apporter des éléments d'information pertinents contribuant à une meilleure connaissance des spécificités des Pucciniales. L'analyse détaillée et comparative de l'expansion des TE à travers leur annotation experte dans des génomes de Pucciniales devrait permettre d'identifier les trajectoires évolutives de certaines familles, et l'existence d'évènements de type *burst* de transposition et leur possible impact sur les familles de gènes elles aussi en expansion. Par ailleurs, il serait intéressant de pouvoir identifier les mécanismes qui ont permis aux TE d'envahir les génomes des Pucciniales au sens large par des analyses comparatives au sein des Pucciniales mais aussi avec des espèces proches dans la classe des Pucciniomycètes (perte ou prise de contrôle, accommodation et régulation de leur activité, maintien). Parmi les nombreux gènes mis à jour dans les génomes des deux Melampsoraceae étudiés, près de 10% codent des petites protéines sécrétées dont une part non négligeable (22%) est partagée par les deux espèces. Aussi, de nombreux gènes de fonctions inconnues sont présents dans les génomes des Pucciniales dont une partie est exprimée spécifiquement chez les différentes plantes hôtes. Le séquençage d'autres génomes de Pucciniales pourra aider à définir la part des gènes réellement spécifiques pour chacune des espèces, notamment pour les gènes qui codent des effecteurs candidats qui pourraient être impliqués dans la spécificité d'hôte. Dans ce cadre, l'étude de génomes de Pucciniales infectant une même plante hôte et des hôtes alternants différents comme c'est le cas pour *M. larici-populina* et *M. allii-populina* serait très pertinent, mais il serait aussi intéressant de pouvoir croiser des analyses avec des espèces de Pucciniales appartenant à des familles différentes et infectant des mêmes plantes hôtes (par exemple *Puccinia allii* qui infecte des *Allium* spp. et *M. allii-populina*). De telles canalyses comparatives devraient voir le jour dans le cadre du projet *Rust pangenomics* débuté par le JGI et un consortium de chercheurs travaillant sur les Pucciniales. Des analyses comparatives plus étendues seront aussi très pertinentes pour comprendre l'évolution de fonctions clés (expansion, contraction, diversification, néo-fonctionalisation) depuis des lignées basales des Pucciniales jusqu'aux espèces les plus récentes. Mon expérience d'annotation de génomes des Pucciniales m'a montré qu'il existe une grande variabilité au sein des bases de données génomiques à disposition avec des qualités de séquence et d'annotation très variables. Or toutes les hypothèses qui peuvent être posées sur la base d'analyses comparatives nécessitent d'avoir une assise solide avec des annotations

homogènes en terme de qualité. Il apparait important, au sein de la communauté scientifique qui s'intéresse aux Pucciniales, et au delà aux génomes fongiques, de définir des routines d'annotation et des pipelines d'analyse communs afin de standardiser ces démarches et gagner en robustesse. Pour mieux comprendre les bases moléculaires de la spécificité d'hôte chez les Pucciniales, il serait nécessaire de réaliser plus systématiquement des approches transcriptomiques par RNAseq sur l'ensemble des cycles de vie. Des analyses globales de l'expression par RNAseq au cours du cycle de vie de *M. allii-populina* pourront permettre de déterminer la part des gènes exprimée chez les deux rouilles sur leur hôte commun et quels sont les groupes de gènes exprimés communément ou spécifiquement chez les deux hôtes alternants (*Allium* spp. et *Larix* spp.). Aussi la comparison avec *M. lini*, qui est une espèce autoïque pourrait permettre de séparer un peu plus précisément les gènes qui sont partagés par toutes les modalités d'infection et qui seraient donc propres aux mécanismes infectieux, de ceux spécifique d'un hôte donné. De plus, le lin et le peuplier sont des espèces taxonomiquement proches vis à vis des espèces alternantes, ce qui rend la comparaison pertinente.

La caractérisation fonctionnelle d'effecteurs de Pucciniales était jusqu'il y a quelques années un domaine peu exploré du fait des difficultés associées à l'étude de ces organismes biotrophes obligatoires. Ainsi en 2014, seuls six effecteurs avaient été caractérisés chez les Pucciniales, dont cinq étaient identifiés par leurs propriétés d'avirulence mais pour lesquels les fonctions n'étaient pas définies (Petre *et al.,* 2014). En seulement quelques années, l'application d'approches effectoromiques à travers l'utilisation de systèmes hétérologues a permis de nombreuses avancées pour préciser des effecteurs à étudier en priorité à partir de listes de candidats sélectionnés.

Dans le cadre de ce document de thèse, nous montrons que ces cribles, bien qu'imparfaits, s'avèrent être essentiels pour l'amorce de la caractérisation fonctionnelle des effecteurs. Au total, plus de 300 effecteurs candidats ont été testés ces trois dernières années chez différentes espèces de Pucciniales, et plus de 70 ont été sélectionnés en vue d'analyses fonctionnelles plus poussées (cf. **Article 1**). Cela illustre l'évolution très rapide de l'étude des effecteurs des Pucciniales, qui est vraisemblablement associée à l'intérêt que représente l'obtention de nouveaux moyens de lutte contre ces maladies. Parmi les approches utilisées pour tester les effecteurs candidats, celles basées sur l'extinction de gènes par HIGS ou VIGS ont été développées chez différentes Pucciniaceae (*P. triticina* et *P. striiformis* f. sp. *tritici*). Toutefois, cette approche est parfois controversée du fait que les mécanismes sous-jacents à cette extinction d'expression génique (mise en place, contrôle, régulation) sont peu étudiés et

mal compris. Cependant cette approche permet de se placer directement dans la plante hôte et son développement à un plus large nombre de pathosystèmes permettrait des avancées remarquables chez les Pucciniales. Il faut souligner que cette approche ne permet de réaliser que de l'extinction de gène partielle (*knock-down*) qui peut être complexe à interpréter en termes de phénotype vis à vis d'une désactivation complète de gène (*gene knock-out*). Une autre approche encore peu développée chez les Pucciniales est la production systématique de protéines recombinantes pour analyser en détail la structure tridimensionnelle des effecteurs candidats. La caractérisation structurale d'effecteurs de *M. larici-populina* a permis de mettre en évidence des similarités structurales avec des protéines aux motifs ou au rôle défini comme les *Knottins* et les *Nuclear transport factor 2 like* ouvrant des hypothèses quant aux rôles possibles de ces effecteurs et ouvrant des perspectives pour progresser dans leur analyse fonctionnelle. En considérant les spécificités des Pucciniales, une telle approche de production systématique pourrait éventuellement révéler des familles d'effecteurs partageant des structures partagées propres aux Pucciniales comme cela a été montré pour la famille des effecteurs MAX chez différents ascomycètes phytopathogènes dont *M. oryzae* (De guilen et al. 2015). De plus, les approches biophysiques disponibles pour étudier les interactions protéine-protéine peuvent également apporter des éléments de réponse quant aux interactions entre les effecturs candidats et les protéines cibles chez l'hôte. Ces approches sont encore sous-exploitées à l'heure actuelle. Elles nécessitent des compétences bien spécifiques en biochimie et elles peuvent s'avérer extrêmement pertinentes et complémentaires d'autres approches comme les cribles réalisés de façon plus systématique en système hétérologues.

Parmi les effecteurs étudiés au cours de cette thèse, la famille MLPCTP ciblant les chloroplastes est particulièrement intéressante car elle amène à de nombreuses questions quant aux rôles que pourraient jouer les effecteurs dans ces organites cellulaires. Les chloroplastes jouent un rôle important dans l'immunité via la génération de molécules impliquées dans la signalisation et les réponses de défense mais aussi en tant que source énergétique primaire et centrale dans le fonctionnement cellulaire. Il n'est donc pas surprenant que les champignons ciblent ces compartiments lors de l'infection. Toutefois, leur rôle précis n'est que spéculatif à l'heure actuelle. Par ailleurs, l'outil de prédiction LOCALIZER permet de prédire la présence de signaux d'adressage à ces organites dans les effecteurs candidats des champignons. Les quelques analyses réalisées chez les Pucciniales et au delà chez les champignons phytopathogènes indiquent que les chloroplastes seraient particulièrement ciblés par les champignons biotrophes et hémibiotrophes (Sperschneider *et al.,* 2017). Jusqu'à présent, au delà des screens effectoromiques, les approches de caractérisation fonctionnelle

chez des champignons modèles comme *U. maydis* se sont plutôt concentrées sur un seul effecteur (cf. Lo Presti *et al,.* 2015). Toutefois, et notamment en considérant les vagues d'expression coordonnées mises à jour par des approches transcriptomiques, il semble évident qu'une action coordonnée de multiples effecteurs prend place de manière simultanée. A plus ou moins long terme, il sera intéressant de considérer l'analyse de groupes d'effecteurs et non plus d'effecteurs isolés. Les approches prédictives par apprentissage de type *machine Learning* couplée aux méthodes d'annotation et aux approches transcriptomiques peuvent s'avérer puissantes pour orienter de telles recherches.

L'étude ciblée des effecteurs a marqué un tournant dans le domaine des interactions moléculaires entre plantes et micro-organismes. Le foisonnement de publications récent du domaine montre bien que l'effort des scientifiques s'est concentré sur ces déterminants moléculaires. Ainsi, d'autres activités associées à la signalisation ou au transport de nutriments qui étaient plus en vogue il y a quelques années ne sont quasiment plus abordées sauf en quelques occasions (par exemple chez *P. striiformis* f. sp. *tritici*). Il est pourtant très important de comprendre comment les Pucciniales sont capables de détourner les nutriments de leur hôte pour assurer leur croissance et leur multiplication et quels sont les conséquences physiologiques sur leurs hôtes. La sélection des fonctions d'intérêt à étudier peut être priorisée par les analyses globales d'expression génique déjà disponibles, et leur étude peut plus spécifiquement être envisagée à travers des approches telles que le HIGS. Même si jusqu'à présent la transformation génétique des Pucciniales n'a pas été possible, certains groupes continuent de tenter de réaliser une telle transformation. Un succès représenterait une avancée considérable pour l'étude de la biologie des pucciniales et nous permettrait d'aborder l'analyse moléculaire des interactions plantes-Pucciniales sous un autre angle.

REFERENCES

Abramovitch, R. B., Janjusevic, R., Stebbins, C. E., & Martin, G. B. (2006). Type III effector AvrPtoB requires intrinsic E3 ubiquitin ligase activity to suppress plant cell death and immunity. *Proceedings of the National Academy of Sciences of the United States of America*, 103(8), 2851‑2856.

Ahmad, M., Hirz, M., Pichler, H., & Schwab, H. (2014). Protein expression in Pichia pastoris: recent achievements and perspectives for heterologous protein production. *Applied Microbiology and Biotechnology*, 98(12), 5301‑5317.

Aime, M.C., MacTaggart A.R, Mondo, S.J., & Duplessis, S. (2017). Phylogenetics and phylogenomics of rust fungi. *Advances in fungal Genetics*.100: 267-307.

Aime, M., Toome, M., & McLaughlin, D. (2014). The Pucciniomycotina. *In* eds *McLaughlin D., Spatafora J.W.*, The Mycota VII Part A. Springer-Verlag, 271–294.

Aime, M. C., (2006). Toward resolving family-level relationships in rust fungi (Uredinales). *Mycoscience*, 47(3), 112–122.

Aime, M. C., Matheny, P. B., Henk, D. A., Frieders, E. M., Nilsson, R. H., Piepenbring, M., et al. (2006). An overview of the higher level classification of Pucciniomycotina based on combined analyses of nuclear large and small subunit rDNA sequences. *Mycologia*, 98(6), 896–905.

Alfano, J. R. (2009). Roadmap for future research on plant pathogen effectors. Molecular Plant Pathology, 10(6), 805–813.

Amselem, J., Lebrun, M.-H., & Quesneville, H. (2015) Whole genome comparative analysis of transposable elements provides new insight into mechanisms of their inactivation in fungal genomes. *BMC genomics*, 16(1): 141.

Amselem, J., Cuomo, C. A., van Kan, J. A. L., Viaud, M., Benito, E. P., Couloux, A., Coutinho, P. M., de Vries, R. P., Dyer, P. S., Fillinger, S., Fournier, E., Gout, L., Hahn, M., Kohn, L., Lapalu, N., Plummer, K. M., Pradier, J.-M., Quévillon, E., Sharon, A., Simon, A., ten Have, Tudzynski, A., B., Tudzynski, P., Wincker, P., Andrew, M., Anthouard, V., Beever, R. E., Beffa, Benoit, R., I., Bouzid, O., Brault, B., Chen, Choquer, Z., M., Collémare, J., Cotton, P., Danchin, E. G., Da Silva, C., Gautier, A., Giraud, C., Giraud, T., Gonzalez, C., Grossetete, S., Güldener, U., Henrissat, B., Howlett, B. J., Kodira, Kretschmer, C., M., Lappartient, A., Leroch, M., Levis, C., Mauceli, E., Neuvéglise, Oeser, C., Pearson, B., M., Poulain, J., Poussereau, N., Quesneville, H., Rascle, C., Schumacher, J., Ségurens, B., Sexton, A., Silva, E., Sirven, C., Soanes, D. M., Talbot, N. J., Templeton, M., Yandava, C., Yarden, O., Zeng, Q., Rollins, J. A., Lebrun, M.-H., & Dickman, M. (2011). Genomic Analysis of the Necrotrophic Fungal Pathogens Sclerotinia sclerotiorum and Botrytis cinerea. *PLOS Genetics*, 7(8), e1002230.

Anders, S., & Huber, W. (2010). Differential expression analysis for sequence count data. *Genome Biology*, 11(10), R106‑R106.

Anderson, C., Khan, M. A., Catanzariti, A.-M., Jack, C. A., Nemri, A., Lawrence, G. J., Jones, D. A. (2016). Genome analysis and avirulence gene cloning using a high-density RADseq linkage map of the flax rust fungus, *Melampsora lini. BMC Genomics*, 17, 667.

Anikster, Y., Szabo, L. J., Eilam, T., Manisterski, J., Koike, S. T., & Busnell, W. R., (2004) Morphology, life cycle biology and DNA sequence analysis of rust fungi on garlic and chives from California. Phytopathology, **9:** 569-577.

Arias, S. L., Theumer, M. G., Mary, V. S., & Rubinstein, H. R. (2012). Fumonisins: probable role as effectors in the complex interaction of susceptible and resistant maize hybrids and *Fusarium verticillioides*. *Journal of Agricultural and Food Chemistry*, 60(22), 5667‐5675.

Avis, J. M., & Clarke, P. R. (1996). Ran, a GTPase involved in nuclear processes: its regulators and effectors. *Journal of Cell Science*, 109(10), 2423‐2427.

Bailey, T L., Boden, M., Buske, F A., Frith, M., Grant, C E., Clementi, Ren,J., Li, W.W., & Noble, W.S. (2009). MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Research* 37: W202–8.

Bhattacharjee, S., Halane, M.K., Kim, S.H., & Gassmann, W. (2011) Pathogen effectors target Arabidopsis EDS1 and alter its interactions with immune regulators. *Science* 334:1405-1408

Bakkeren, G., Joly, D. L., & Duplessis, S. (2016). Editorial: Genomics Research on Non-model Plant Pathogens: Delivering Novel Insights into Rust Fungus Biology. *Frontiers in Plant Science*, 7, 216.

Bao, Z., & Eddy, S. R. (2002). Automated De Novo Identification of Repeat Sequence Families in Sequenced Genomes. *Genome Research*, 12(8), 1269‐1276.

Barrett, L. G., Thrall, P. H., Dodds, P. N., van der Merwe, M., Linde, C. C., Lawrence, G. J., & Burdon, J. J. (2009). Diversity and Evolution of Effector Loci in Natural Populations of the Plant Pathogen *Melampsora lini*. *Molecular Biology and Evolution*, 26(11), 2499–2513.

Bartoli, C., & Roux, F. (2017). Genome-Wide Association Studies In Plant Pathosystems: Toward an Ecological Genomics Approach. *Frontiers in Plant Science*, 8, 763.

Beckers, G. J. M., & Spoel, S. H. (2006). Fine-Tuning Plant Defence Signalling: Salicylate versus Jasmonate. *Plant Biology*, 8(1), 1‐10.

Beddow, J. M., Pardey, P. G., Chai, Y., Hurley, T. M., Kriticos, D. J., Braun, H.-J., & Yonow, T. (2015). Research investment implications of shifts in the global geography of wheat stripe rust. *Nature Plants*, 1, 15132.

Berglund, A-K., Pujol, C., Duchene, A-M., & Glaser, E. (2009) Defining the determinants for dual targeting of amino acyl-tRNA synthetases to mitochondria and chloroplasts. J *Mol Biol* 393:803-814

Block, A., & Alfano, J. R. (2011). Plant targets for Pseudomonas syringae type III effectors: Virulence targets or guarded decoys? *Current opinion in microbiology*, 14(1), 39‐46.

Blondeau, K., Blaise, F., Graille, M., Kale, S. D., Linglin, J., Ollivier, B., Labarde, A., Lazar, N., Daverdin, G., Balesdent, M.-H., Choi, D. H. Y., Tyler, B. M., Rouxel, T., van Tilbeurgh, H., & Fudal, I. (2015). Crystal structure of the effector AvrLm4-7 of Leptosphaeria maculans reveals insights into its translocation into plant cells and recognition by resistance proteins. The Plant Journal : For Cell and Molecular Biology, 83(4), 610–624.

Boch, J., Bonas, U., & Lahaye, T. 2014. TAL effectors--pathogen strategies and plant resistance engineering. *New Phytologist*, **204**(4):823-32.

Boch, J., Scholze, H., Schornack, S., Landgraf, A., Hahn, S., Kay, S., Lahaye, T., Nickstadt, A., & Bonas, U. (2009). Breaking the code of DNA binding specificity of TAL-type III effectors. *Science*, 326(5959), 1509‐1512.

Boehm, E. W. A., & Bushnell, W. R. (1992). An Ultrastructural Pachytene Karyotype for *Melampsora lini*. *Phytopathology*, 82, 1212‐1218.

Bombarely, A., Rosli, H.G., Vrebalov, J., Moffett, P., Mueller, L.A., & Martin, G.B. (2012) A draft genome sequence of Nicotiana benthamiana to enhance molecular plant-microbe biology research. *Mol Plant Microbe Interact* 25:1523-1530

Bolton, M. D., Kolmer, J. A., & Garvin, D. F. (2008). Wheat leaf rust caused by *Puccinia triticina*. *Molecular Plant Pathology*, 9(5), 563–575.

Boutemy, L. S., King, S. R. F., Win, J., Hughes, R. K., Clarke, T. A., Blumenschein, T. M. A., Kamoun, S., & Banfield, M. J. (2011). Structures of Phytophthora RXLR Effector Proteins: a conserved but adaptable fold underpins functional diversity. *The Journal of Biological Chemistry*, 286(41), 35834‐35842.

Boutrot, F., & Zipfel, C. (2017). Function, Discovery, and Exploitation of Plant Pattern Recognition Receptors for Broad-Spectrum Disease Resistance. *Annual Review of Phytopathology*, 55, 257‐286.

Bozkurt, T.O., Richardson, A., Dagdas, Y.F., Mongrand, S., Kamoun, S., & Raffaele, S. (2014) The Plant Membrane-Associated REMORIN1.3 Accumulates in Discrete Perihaustorial Domains and Enhances Susceptibility to *Phytophthora infestans. Plant Physiol* 165:1005-1018

Bozkurt, T. O., Schornack, S., Banfield, M. J., & Kamoun, S. (2012). Oomycetes, effectors, and all that jazz. *Biotic interactions*, 15(4), 483‐492.

Bozkurt, T.O., Schornack, S., Win, J., Shindo, T., Ilyas, M., Oliva, R., et al (2011) Phytophthora infestans effector AVRblb2 prevents secretion of a plant immune protease at the haustorial interface. *Proc Natl Acad Sci U S A* 108:20832-20837

Bruce, M., Neugebauer, K., Joly, D., Migeon, P., Cuomo, C., Wang, S., & Fellers, J. (2014). Using transcription of six *Puccinia triticina* races to identify the effective secretome during infection of wheat. *Frontiers in Plant Science*, 4, 520.

Bruce, B.D. (2001) The paradox of plastid transit peptides: conservation of function despite divergence in primary structure. *Biochim Biophys Acta* 1541:2-21

Bueno-Sancho, V., Persoons, A., Hubbard, A., Cabrera-Quio, L. E., Lewis, C. M., Corredor-Moreno, P., Bunting, D.C.E., Ali, S., Cheng, S., Hodson, D. P., Madariaga-Burrow, R., Bryson, R., Thomas, J., Holdgate, S., & Saunders, D. G. O. (2017). Pathogenomic Analysis of Wheat Yellow Rust Lineages Detects Seasonal Variation and Host Specificity. *Genome Biology and Evolution*, 9(12), 3282–3296.

Caillaud M. C. Piquerez S. J. Fabro, G. Steinbrenner J., Ishaque N., Beynon J., & Jones J. D. (2012) Subcellular localization of the Hpa RxLR effector repertoire identifies a tonoplast-associated protein HaRxL17 that confers enhanced plant susceptibility, *The Plant Journal*, 69(2), 252-265.

Caillaud M. C., Asai S., Rallapalli G., Piquerez S., Fabro G., & Jones J. D. (2013) A downy mildew effector attenuates salicylic Acid-triggered immunity in Arabidopsis by interacting with the host mediator complex, *PLoS Biology*, 11(12).

Cantarel, B. L., Coutinho, P. M., Rancurel, C., Bernard, T., Lombard, V., & Henrissat, B. (2009). The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. *Nucleic Acids Research*, 37:D233‐D238.

Cantu, D., Govindarajulu, M., Kozik, A., Wang, M., Chen, X., Kojima, K.K., Jurka, J., Michelmore, R.W., Dubcovsky, J., & Harris, S. (2011). Next generation sequencing provides rapid access to the genome of *Puccinia striiformis* f. sp. *tritici*, the causal agent of wheat stripe rust. *PLoS ONE*, 6(8): e24230.

Cantu, D., Segovia, V., MacLean, D., Bayles, R., Chen, X., Kamoun, S., Dubcovsky, J., Saunders, D.G.O., & Uauy, C. (2013). Genome analyses of the wheat yellow (stripe) rust pathogen *Puccinia striiformis* f. sp. *tritici* reveal polymorphic and haustorial expressed secreted proteins as candidate effectors. *BMC Genomics*, 14: 270.

Carazo-Salas, R. E., Gruss, O. J., Mattaj, I. W., & Karsenti, E. (2001). Ran-GTP coordinates regulation of microtubule nucleation and dynamics during mitotic-spindle assembly. *Nature Cell Biology*, 3(3), 228‐234.

Carrie, C., & Whelan, J. (2013) Widespread dual targeting of proteins in land plants: when, where, how and why. *Plant Signal Behav* 8:8

Carrie, C., & Small, I. (2013) A reevaluation of dual-targeting of proteins to mitochondria and chloroplasts. *Biochim Biophys Acta* 1833:253-259

Carrie, C., Kühn, K., Murcha, M.W., Duncan, O., Small, I.D., O'Toole, N., & Whelan, J. (2009) Approaches to defining dual-targeted proteins in Arabidopsis. *Plant Journal*, 57:1128-1139

Castanera, R., López-Varas, L., Borgognone, A., LaButti, K., Lapidus, A., Schmutz, J., Grimwood, J., Pérez, G., Pisabarro, A.G., Grigoriev, I.V., Stajich, J.E., & Ramírez, L. (2016). Transposable Elements versus the Fungal Genome: Impact on Whole-Genome Architecture and Transcriptional Profiles. *PLoS Genetics*, 12(6), e1006108.

Catanzariti, A.-M., Dodds, P. N., Lawrence, G. J., Ayliffe, M. A., & Ellis, J. G. (2006). Haustorially Expressed Secreted Proteins from Flax Rust Are Highly Enriched for Avirulence Elicitors. *The Plant Cell*, 18(1), 243‐256.

Chang, Q., Liu, J., Lin, X., Hu, S., Yang, Y., Li, D., Chen, L., Huai, B., Huang, L., Voegele, R.T., & Kang, Z. (2017). A unique invertase is important for sugar absorption of an obligate biotrophic pathogen during infection. *New Phytologist*, 215(4), 1548–1561.

Chang, S., Puryear, J., & Cairney, J. (1993). A Simple and Efficient Method for Isolating RNA from Pine Tree. *Plant Molecular Biology Reporter*, 11(2), 113‐116.

Chen, X. & Kang, Z. 2017. Stripe Rust, 719.

Chen J., Upadhyaya N.M., Ortiz D., Sperschneider J., Li F., Bouton C., Breen S., Dong C., Xu B., Zhang X., Mago R., Newell K., Xia X., Bernoux M., Taylor J.M., Steffenson B., Jin Y., Zhang P., Kanyuka K., Figueroa M., Ellis J.G., Park R.F., & Dodds P.N. (2017) Loss of AvrSr50 by somatic exchange in stem rust leads to virulence for Sr50 resistance in wheat. *Science,* 358:1607-1610.

Chen, W., Wellings, C., Chen, X., Kang, Z., & Liu, T. (2014). Wheat stripe (yellow) rust caused by *Puccinia striiformis* f. sp. *tritici*. *Molecular Plant Pathology*, 15(5), 433–446.

Chen, Z., Nunes, M. A., Silva, M. C., & Rodrigues, C. J. (2004). Appressorium turgor pressure of *Colletotrichum kahawae* might have a role in coffee cuticle penetration. *Mycologia*, 96(6), 1199–1208.

Cheng Y., Wang X., Yao J., Voegele R.T., Zhang Y., Wang W., Huang L., & Kang Z. (2015) Characterization of protein kinase PsSRPKL, a novel pathogenicity factor in the wheat stripe rust fungus. *Environmental Microbiology*, 17(8), 2601–2617.

Cheng Y., Wu K., Yao J., Li S., Wang X., Huang L., & Kang Z. (2016) PSTha5a23, a candidate effector from the obligate biotrophic pathogen *Puccinia striiformis* f. sp. *tritici*, is involved in plant defense suppression and rust pathogenicity. *Environmental Microbiology*, 19(5), 1717-1729. .

Cheng, Y., Wang, W., Yao, J., Huang, L., Voegele, R. T., Wang, X., & Kang, Z. (2016). Two distinct Ras genes from *Puccinia striiformis* exhibit differential roles in rust pathogenicity and cell death. *Environmental Microbiology*, 18(11), 3910–3922.

Chou, S., Krasileva, K. V., Holton, J. M., Steinbrenner, A. D., Alber, T., & Staskawicz, B. J. (2011). *Hyaloperonospora arabidopsidis* ATR1 effector is a repeat protein with distributed recognition surfaces. *Proceedings of the National Academy of Sciences of the United States of America*, 108(32).

Chouard, T. (2011). Structural biology: Breaking the protein rules. *Nature*, 471(7337), 151‐153.

Collemare, J., & Lebrun, M. (2011). Fungal Secondary Metabolites: Ancient Toxins and Novel Effectors in Plant–Microbe Interactions. *Effectors in Plant-Microbe Interactions*.(Wiley-Blackwell), pp:377-400.

Cook, D. E., Mesarich, C. H., & Thomma, B. P. H. J. (2015). Understanding plant immunity as a surveillance system to detect invasion. *Annual Review of Phytopathology*, 53, 541‐563.

Cooper, B., & Campbell, K. B. (2017). Protection Against Common Bean Rust Conferred by a Gene-Silencing Method. *Phytopathology*, 107(8), 920–927.

Cooper, B., Campbell, K. B., Beard, H. S., Garrett, W. M., & Islam, N. (2016). Putative Rust Fungal Effector Proteins in Infected Bean and Soybean Leaves. *Phytopathology*, 106(5), 491–499.

Corbett, A. H., & Silver, P. A. (1996). The NTF2 gene encodes an essential, highly conserved protein that functions in nuclear transport in vivo. *The Journal of Biological Chemistry*, 271(31), 18477‐18484.

Cosgrove, D. J. (2000). Loosening of plant cell walls by expansins. *Nature*, 407(321).

Cristancho, M. A., Botero-Rozo, D. O., Giraldo, W., Tabima, J., Riaño-Pachón, D. M., Escobar, C., Rozo, Y., Rivera, L.F., Duran, A., Restrepo, S., Eilam, T., Anikster, Y., & Gaitan, A.L. (2014). Annotation of a hybrid partial genome of the coffee rust (*Hemileia vastatrix*) contributes to the gene repertoire catalog of the Pucciniales. *Frontiers in Plant Science*, 5, 594.

Cui, H., Tsuda, K., & Parker, J. E. (2015). Effector-Triggered Immunity: From Pathogen Perception to Robust Defense. *Annual Review of Plant Biology*, 66(1), 487‐511.

Cummins, G., & Hiratsuka, Y. (2003). Illustrated genera of rust fungi (APS, St Paul). pp.225.

Cunningham, G. H. (1931). The rust fungi of New Zealand (Dunnedin).pp.261.

Cuomo, C.A., Bakkeren, G., Khalil, H.B., Panwar, V., Joly, D., Linning, R., Sakthikumar, S., Song, X., Adiconis, X., Fan, L., Goldberg, J.M., Levin, J.Z., Young, S., Zeng, Q., Anikster, Y., Bruce, M., Wang, M., Yin, C., McCallum, B., Szabo, L.J., Hulbert, S., Chen, X., & Fellers, J.P. (2017). Comparative analysis highlights variable genome content of wheat rusts and divergence of the mating loci. *G3*, 7: 361-376.

Dagvadorj, B., Ozketen, A. C., Andac, A., Duggan, C., Bozkurt, T. O., & Akkaya, M. S. (2017). A *Puccinia striiformis* f. sp. *tritici* secreted protein activates plant immunity at the cell surface. *Scientific Reports*, 7(1), 1141.

Daly, N. L., Gruber, C. W., Göransson, U., & Craik, D. J. (2011). Cystine Knot Folding in Cyclotides. Chang & S. Ventura (Éd.), Folding of Disulfide Proteins (p. 43‑61).

Dangl, J. L., Horvath, D. M., & Staskawicz, B. J. (2013). Pivoting the plant immune system from dissection to deployment. *Science*, 341(6147), 746‑751.

Dawkins, R. (1982). The Extended Phenotype: The Gene as the Unit of Selection. Oxford University Press. pp320.

de Carvalho M.C. da C.G., Nascimento L.C., Darben L.M., Polizel-Podanosqui A.M., Lopes-Caitar V.S., Qi M., Rocha C.S., Carazzolle M.F., Kuwahara M.K., Pereira G.A.G., Abdelnoor R.V., Whitham S.A., & Marcelino-Guimaraes F.C. (2016). Prediction of the in planta Phakopsora pachyrhizi secretome and potential effector families. *Molecular Plant Pathology*, 18(2017),12405.

de Castro, E., Sigrist, C. J. A., Gattiker, A., Bulliard, V., Langendijk-Genevaux, P. S., Gasteiger, E., Bairoch, A., & Hulo, N. (2006). ScanProsite: detection of PROSITE signature matches and ProRule-associated functional and structural residues in proteins. *Nucleic Acids Research*, 34, W362‑W365.

de Guillen, K., Ortiz-Vallejo, D., Gracy, J., Fournier, E., Kroj, T., & Padilla, A. (2015). Structure Analysis Uncovers a Highly Diverse but Structurally Conserved Effector Family in Phytopathogenic Fungi. *PLoS Pathogens*, 11(10), e1005228.

de Torres Zabala, M., Littlejohn, G., Jayaraman, S., Studholme, D., Bailey, T., Lawson, T., Tillich, M., Licht, D., Bölter, B., Delfino, L, Truman, W., Mansfield, J., Smirnoff, N., & Grant, M. (2015). Chloroplasts play a central role in plant defence and are targeted by pathogen effectors. *Nature Plants*, 1, 15074.

De Wit, P. J. G. M. (2016). Apoplastic fungal effectors in historic perspective; a personal view. *New Phytologist*, 212(4), 805‑813.

Dean R., van Kan J. A. L., Pretorius Z. A., Hammond-Kosack K. E., di Pietro A., Spanu P. D., Rudd J. J., Dickman M., Kahmann R., Ellis J. & Foster G. D. (2012) The Top 10 fungal pathogens in molecular plant pathology. *Molecular Plant Pathology*, 13: 414–430.

Dean, R. A., Talbot, N. J., Ebbole, D. J., Farman, M. L., Mitchell, T. K., Orbach, M. J., et al. (2005). The genome sequence of the rice blast fungus *Magnaporthe grisea*. *Nature*, 434, 980.

Demarsy, E., Lakshmanan, A.M., & Kessler, F. (2014) Border control: selectivity of chloroplast protein import and regulation at the TOC-complex. *Front Plant Sci* 5:483

Deslandes, L., & Rivas, S. (2012). Catch me if you can: bacterial effectors and plant targets. *Trends in Plant Science*, 17(11), 644‑655.

Djamei, A., Schipper, K., Rabe, F., Ghosh, A., Vincon, V., Kahnt, J., Osorio, S., Tohge, T., Fernie, A., Feussner, I., Feussner, K., Meinicke, P., Stierhof, Y.-D., Schwarz, H., Macek, B., Mann, M., & Kahmann, R. (2011). Metabolic priming by a secreted fungal effector. *Nature*, 478(7369), 395‑398.

Djulic, A., Schmid, A., Lenz, H., Sharma, P., Koch, C., Wirsel, S. G. R., & Voegele, R. T. (2011). Transient transformation of the obligate biotrophic rust fungus Uromyces fabae using biolistics. *Fungal Biology*, 115(7), 633‑642.

Dobon, A., Bunting, D. C. E., Cabrera-Quio, L. E., Uauy, C., & Saunders, D. G. O. (2016). The host-pathogen interaction between wheat and yellow rust induces temporally coordinated waves of gene expression. *BMC Genomics*, 17(1), 380.

Dodds, P. N., & Rathjen, J. P. (2010). Plant immunity: towards an integrated view of plant-pathogen interactions. *Nature Reviews Genetics*, 11(8)

Dodds, P.N., Rafiqi, M., Gan, P. H. P., Hardham, A. R., Jones, D. A., & Ellis, J. G. (2009). Effectors of biotrophic fungi and oomycetes: pathogenicity factors and triggers of host resistance. *New Phytologist* 183 (4): 993–1000.

Dodds, P. N., Lawrence, G. J., Catanzariti, A. M., Ayliffe, M. A., & Ellis, J. G. (2004). The *Melampsora* lini Avr567 avirulence genes are expressed in haustoria and their products are recognized inside plant cell. *Plant Cell*, 16.

Dong, S., Raffaele, S., & Kamoun, S. (2015). The two-speed genomes of filamentous pathogens: waltz with plants. *Genomes and evolution*, 35, 57‐65.

Dong, S., Remco, S., Cano, L.M., Song, S., Sklenar, J., Yoshida, K., Bozkurt, T.O., Oliva, R., Liu, Z., Tian, M., Win, J., Banfield, M.J., Jones, A.M.E., van der Hoorn, R.A.L., & Kamoun, S. (2014). Effector specialization in a lineage of the Irish Potato famine pathogen. *Science,* 343: 552-555.

Duplessis, S., Bakkeren, G., & Hamelin, R. (2014). Advancing knowledge on biology of rust fungi through genomics. *Advances in Botanic Research*, Vol. 70, p. 173‐209.

Duplessis, S., Cuomo, C. A., Lin, Y, C., Aerts, A., Tisserant, E., Veneault- Fourrey, C., Joly, D. L., Hacquard, S., Amselem, J., Cantarel, B., Chiu, R., Coutinho, P., Feau, N., Field, M., Frey, P., Gelhaye, E., Goldberg, J., Grabherr, M., Kodira, C., Kohler, A., Hües, U., Lindquist, E. A., Lucas, S., Mago, R., Mauceli, E., Morin, E., Murat, C., Pangilinan, J. L., Park, R., Pearson, M., Quesneville, H., Rouhier, N., Sakthikumar, S., Schmutz, J., Selles, B., Shapiro, H., Tanguay, P., Tuskan, G.A., Henrissat, B., Van de Peer, Y., Rouzé, P., Ellis, J. G., Dodds, P. N., Schein, J. E., Zhong, S., Hamelin, R. C., Grigoriev, I. V., Szabo, L. J., & Martin, F. (2011a). Obligate biotrophy features unravelled by the genomic analysis of rust fungi. *Proceedings of the National Academy of Sciences of the United States of America*. 108: 9166-9171.

Duplessis, S., Hacquard, S., Delaruelle, C., Tisserant, E., Frey, P., Martin, F., Kohler, A. (2011b). *Melampsora larici-populina* transcript profiling during germination and timecourse infection of poplar leaves reveals dynamic expression patterns associated with virulence and biotrophy. *Molecular Plant-Microbe Interactions : MPMI* 24 (7): 808–18.

Duplessis, S., Joly, D. J., & Dodds, P. N. (2012). Rust effectors. *Effectors in Plant-Microbes Interactions*, eds F. Martin and S. Kamoun (Oxford: Wiley-Blackwell), 155–193.

Duplessis, S., Spanu, P. D., & Schirawski, J. (2013). Biotrophic fungi (powdery mildews, Rusts and Smuts). *In Ecological genomics of the fungi, Plant-interacting fungi section,* ed R. Martin, (Wiley-Blackwell, pp. 149-168).

Eberhardt, R. Y., Chang, Y., Bateman, A., Murzin, A. G., Axelrod, H. L., Hwang, W. C., & Aravind, L. (2013). Filling out the structural map of the NTF2-like superfamily. *BMC Bioinformatics*, 14, 327‐327.

Edgar, R. C., & Myers, E. W. (2005). PILER: identification and classification of genomic repeats. *Bioinformatics*, 21, i152‐i158.

Ellis, J. G., Dodds, P. N., & Lawrence, G. J. (2007). Flax rust resistance gene specificity is based on direct resistance-avirulence protein interactions. *Annual Review of Phytopathology*, 45, 289–306.

Emanuelsson, O., Brunak, S., von Heijne, G., & Nielsen, H. (2007). Locating proteins in the cell using TargetP, SignalP and related tools. *Nature Protocols*, 2, 953.

Emanuelsson, O., H. Nielsen, S. Brunak, & G. von Heijne. (2000). Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *Journal of Molecular Biology,* 300(4): 1005–16.

Engler, C., & Marillonnet, S. (2013) Combinatorial DNA assembly using Golden Gate cloning. *Methods Mol Biol,* 1073:141-156

Enright, A. J., Van Dongen, S., & Ouzounis, C. A. (2002). An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Research,* 30(7): 1575–84.

Faino, L., Seidl, M. F., Shi-Kunne, X., Pauper, M., van den Berg, G. C., Wittenberg, A. H., & Thomma, B. P. (2016). Transposons passively and actively contribute to evolution of the two-speed genome of a fungal pathogen. *Genome Research*, 26(8), 1091‑1100.

Fernandez, D., Tisserant, E., Talhinhas, P., Azinheira, H., Vieira, A., Petitot, A.-S., Loureiro, A., Poulain, J., Da Silva, C., Silva, M. D.C., & Duplessis, S. (2012). 454-pyrosequencing of Coffea arabica leaves infected by the rust fungus Hemileia vastatrix reveals in planta-expressed pathogen-secreted proteins and plant functions in a late compatible plant-rust interaction. *Molecular Plant Pathology*, 13(1), 17–37.

Figueroa, M., Upadhyaya, N. M., Sperschneider, J., Park, R. F., Szabo, L. J., Steffenson, B., Ellis, G.J., & Dodds, P. N. (2016). Changing the Game: Using Integrative Genomics to Probe Virulence Mechanisms of the Stem Rust Pathogen *Puccinia graminis* f. sp. *tritici*. *Frontiers in Plant Science*, 7, 205.

Flor, H. (1971). Current status of the gene-for-gene concept. *Annual Review of Phytopathology*, 9, 275‑296.

Flutre, T., Duprat, E., Feuillet, C., & Quesneville, H. (2011). Considering Transposable Element Diversification in De Novo Annotation Approaches. *PLOS ONE*, 6(1), e16526.

Franceschetti, M., Maqbool, A., Jimenez-Dalmaroni, M. J., Pennington, H. G., Kamoun, S., & Banfield, M. J. (2017). Effectors of Filamentous Plant Pathogens: Commonalities amid Diversity. *Microbiology and Molecular Biology Reviews : MMBR*, 81(2).

Frey, P., & Pinon, J. (1997). Variability in pathogenicity of *Melampsora allii-populina* expressed on poplar cultivars. *European Journal of Forest Pathology*, 27(6), 397‑407.

Frey, P., Gérard, P. R., Feau, N., Husson, C., & Pinon, J. (2005). Variability and population biology of *Melampsora* rusts on poplars. In *Rust Disease in Willow and Poplar* (McCracken A.R and Wallingford), p. 63‑73.

Galagan, J. E., Calvo, S. E., Borkovich, K. A., Selker, E. U., Read, N. D., Jaffe, D., FitzHugh, W., Ma, L.-J., Smirnov, S., Purcell, S., Rehman, B., Elkins, T., Engels, R., Wang, S.,g Nielsen, C. B., Butler, J., Endrizzi, M., Qui, D., Ianakiev, P., Bell-Pedersen, D., Nelson, M. A., Werner-Washburne, M., Selitrennikoff, C. P., Kinsey, J. A., Braun, E. L., Zelter, A., Schulte, U., Kothe, G. O., Jedd, G., Mewes, W., Staben, C., Marcotte, E., Greenberg, D., Roy, A., Foley, K., Naylor,

J., Stange-Thomann, N., Barrett, R., Gnerre, S., Kamal, M., Kamvysselis, M., Mauceli, E., Bielke, C., Rudd, S., Frishman, D., Krystofova, S., Rasmussen, C., Metzenberg, R. L., Perkins, D. D., Kroken, S., Cogoni, C., Macino, G., Catcheside, D., Li, W., Pratt, R. J., Osmani, S. A., DeSouza, C. P., Glass, L., Orbach, M. J. Berglund, J. A., Voelker, R., Yarden, O., Plamann, M., Seiler, S., Dunlap, J., Radford, A., Aramayo, R., Natvig, D. O., Alex, L. A., Mannhaupt, G., Ebbole, D. J., Freitag, M., Paulsen, I., Sachs, M. S., Lander, E. S., Nusbaum, C., & Birren, B. (2003). The genome sequence of the filamentous fungus *Neurospora crassa*. *Nature*, 422(6934): 859-868.

Gaouar, O., Morency, M.-J., Letanneur, C., Séguin, A., & Germain, H. (2016). The 124202 candidate effector of *Melampsora larici-populina* interacts with membranes in *Nicotiana* and *Arabidopsis*. *Canadian Journal of Plant Pathology,* 38(2), 197‑208.

Garcia, M. L. (2004). Ion channels: gate expectations. *Nature*, 430(6996), 153‑155.

Garnica, D. P., Nemri, A., Upadhyaya, N. M., Rathjen, J. P., & Dodds, P. N. (2014). The Ins and Outs of Rust Haustoria. *PLOS Pathogens*, 10(9), e1004329.

Garnica, D. P., Upadhyaya, N. M., Dodds, P. N., & Rathjen, J. P. (2013). Strategies for Wheat Stripe Rust Pathogenicity Identified by Transcriptome Sequencing. *PloS One*, 8(6), e67150.

Gaulin, E. (2017). Effector-mediated communication of filamentous plant pathogens with their hosts. *Advances in botanical research*, Vol. 862: 161‑185.

Germain H., Joly D.L., Mireault C., Plourde M.B., Letanneur C., Stewart D., Morency M. J., Petre B., Duplessis S., & Seguin A. (2017) Infection assays in Arabidopsis reveal candidate effectors from the poplar rust fungus that promote susceptibility to bacteria and oomycete pathogens. *Molecular Plant Pathology*, 19: 191-200.

Godfrey, D., Böhlenius, H., Pedersen, C., Zhang, Z., Emmersen, J. & Thordal-Christensen, H. (2010). Powdery mildew fungal effector candidates share N-terminal Y/F/WxC-motif. BMC Genomics 11: 317–317.

Godoy, C.V., Seixas, C.D.S., Soares, R.M., Marcelino-Guimaraes, F.C., Meyer, M.C., Costamilan, L.M. 2016. Asian soybean rust in Brazil: past, present and future. *Pesq*. *Agropec*. Bras. 5:407-421.

Goellner, K., Loehrer, M., Langenbach, C., Conrath, U., Koch, E., & Schaffrath, U. (2010). *Phakopsora pachyrhizi*, the causal agent of Asian soybean rust. *Molecular Plant Pathology*, 11(2):169-177.

Goessweiner-Mohr, N., Arends, K., Keller, W., & Grohmann, E. (2013). Conjugative type IV secretion systems in Gram-positive bacteria. *Plasmid*, 70(3): 289‑302.

Goswami, R. S., & Kistler, H. C. (2004). Heading for disaster: Fusarium graminearum on cereal crops. *Molecular Plant Pathology*, 5(6): 515‑525.

Guo, J., Dai, X., Xu, J.-R., Wang, Y., Bai, P., Liu, F., Duan, Y., Zhang, H., Huang, L., & Kang, Z. (2011). Molecular Characterization of a Fus3/Kss1 Type MAPK from *Puccinia striiformis* f. sp. *tritici*, PsMAPK1. *PLOS ONE*, 6(7), e21895.

Guttman, D. S., McHardy, A. C., & Schulze-Lefert, P. (2014). Microbial genome-enabled insights into plant–microorganism interactions. *Nature Reviews Genetics*, 15, 797.

Guyon, K., Balagué, C., Roby, D., & Raffaele, S. (2014) Secretome analysis reveals effector candidates associated with broad host range necrotrophy in the fungal plant pathogen *Sclerotinia sclerotiorum*. BMC Genomics, 15:336.

Haas, B.J., Kamoun, S., Zody, M.C., Jiang, R. H. Y., Handsaker, R. E., Cano, L. M., Grabherr, M., Kodira, C. D., Raffaele, S., Torto-Alalibo, T., Bozkurt, T. O., Ah-Fong, A. M. V., Alvarado, L., Anderson, V. L., Armstrong, M. R., Avrova, A., Baxter, L., Beynon, J., Boevink, P. C., Bollmann, S. R., Bos, J. I. B., Bulone, V., Cai, G., Cakir, C., Carrington, J. C., Chawner, M., Conti, L., Costanzo, S., Ewan, R., Fahlgren, N., Fischbach, M. A., Fugelstad, J., Gilroy, E. M., Gnerre, S., Green, P. J., Grenville-Briggs, L. J., Griffith, J., Grünwald, N. J., Horn, K., Horner, N. R., Hu, C.-H., Huitema, E., Jeong, D.-H., Jones, A. M. E., Jones, J. D. G., Jones, R. W., Karlsson, E. K., Kunjeti, S. G., Lamour, K., Liu, Z., Ma, L., MacLean, D., Chibucos, M. C., McDonald, H., McWalters, J., Meijer, H. J. G., Morgan, W., Morris, P. F., Munro, C. A., O'Neill, K., Ospina-Giraldo, M., Pinzón, A., Pritchard, L., Ramsahoye, B., Ren, Q., Restrepo, S., Roy, S., Sadanandom, A., Savidor, A., Schornack, S., Schwartz, D. C., Schumann, U. D., Schwessinger, B., Seyer, L., Sharpe, T., Silvar, C., Song, J., Studholme, D. J., Sykes, S., Thines, M., van de Vondervoort, P., Phuntumart, V., Wawra, S., Weide, R., Win, J., Young, C., Zhou, S., Fry, W., Meyers, B. C., van West, P., Ristaino, J., Govers, F. Birch, P. R. J., Whisson, S. C., Judelson, H. S., & Nusbaum, C. (2009). Genome sequence and analysis of the Irish potato famine pathogen Phytophthora infestans. *Nature* 461 (7262): 393–98.

Hacquard, S., Delaruelle, C., Frey, P., Tisserant, E., Kohler, A., & Duplessis, S. (2013). Transcriptome analysis of poplar rust telia reveals overwintering adaptation and tightly coordinated karyogamy and meiosis processes. *Frontiers in Plant Science*, 4, 456.

Hacquard S., Joly D.L., Lin Y.-C., Tisserant E., Feau N., Delaruelle C., Legue V., Kohler A., Tanguay P., Petre B., Frey P., Van de Peer Y., Rouzé P., Martin F., Hamelin R. C., & Duplessis S. (2012). A comprehensive analysis of genes encoding small secreted proteins identifies candidate effectors in *Melampsora larici-populina* (poplar leaf rust). *Mol. Plant-Microbe Interact. MPMI*, 25(3):279-293.

Hacquard, S., Petre, B., Frey, P., Hecker, A., Rouhier, N., & Duplessis, S. (2011a). The poplar-poplar rust interaction: insights from genomics and transcriptomics. *Journal of Pathogens*, 2011, 716041.

Hacquard, S., Veneault-Fourrey, C., Delaruelle, C., Frey, P., Martin, F., & Duplessis, S. (2011b). Validation of *Melampsora larici-populina* reference genes for in planta RT-quantitative PCR expression profiling during time-course infection of poplar leaves. *Physiol. Mol. Plant Pathol*. 75: 106-112.

Hacquard, S., Delaruelle, C., Legue, V., Tisserant, E., Kohler, A., Frey, P., Martin, F., & Duplessis, S (2010). Laser Capture microdissection of uredinia formed by *Melampsora larici-populina* revealed a transcriptional switch between biotrophy and sporulation. *Mol. Plant-Microbe Interact.MPMI*, 23 (10): 1275–86.

Hahn, M., (2000) The rust fungi: cytology, physiology and molecular biology of infection. *Fungal Pathology* (Kronstad J, ed), pp. 267–306. Kluwer Academic Publishers, Dordrecht.

Hahn, M., & Mendgen, K. (1997). Characterization of in plant-induced rust genes isolated from a haustorium-specific cDNA library. *Mol Plant Microbe Interact*, 10(4): 427-437.

Hahn, M., Neef, U., Struck, C., Göttfert, M., & Mendgen, K. (1997). A putative amino acid transporter is specifically expressed in haustoria of the rust fungus *Uromyces fabae*. *Molecular Plant-Microbe Interactions*, 10(4) : 438–445.

Hamel, L.-P., Nicole, M.-C., Duplessis, S., & Ellis, B. E. (2012). Mitogen-Activated Protein Kinase Signaling in Plant-Interacting Fungi: Distinct Messages from Conserved Messengers. *The Plant Cell*, 24(4): 1327–1351

Harder, D. (1984). Developmental ultrastructure of hyphae and spores. *In The Cereal Rust*, (Bushnell W. R., Roelfs A. P), pp322-361.

Henrissat, B., Surolia, A., & Stanley, P. (2015). A Genomic View of Glycobiology. *In Essentials of Glycobiology*, A. Varki, R.D. Cummings, J.D. Esko, et al., eds. (Cold Spring Harbor Laboratory Press) 2015-2017.

Hetzer, M., Bilbao-Cortes, D., Walther, T. C., Gruss, O. J., & Mattaj, I. W. (2000). GTP hydrolysis by Ran is required for nuclear envelope assembly. *Molecular Cell*, 5(6): 1013‑1024.

Heidrich, K., Wirthmueller, L., Tasset, C., Pouzet, C., Deslandes, L., & Parker, J.E. (2011) Arabidopsis EDS1 connects pathogen effector recognition to cell compartment-specific immune responses. *Science* 334:1401-1404

Hicks, S.W., & Galán J.E. (2013) Exploitation of eukaryotic subcellular targeting mechanisms by bacterial effectors. *Nat Rev Microbiol*, 11:316-326

Hoch, H. C., Staples, R. C., Whitehed, B., Comeau, J., & Wolf, E. D. (1987). Signaling for growth orientation and cell differentiation by surface topography in *Uromyces*. *Science*, 235, 1659–1663.

Hofmann, N. R. (2017). An Emerging Paradigm? RxLR Cleavage before Effector Secretion. The Plant Cell, 29(6): 1177‑1178.

Hogenhout, S. A., Van der Hoorn, R. A. L., Terauchi, R., & Kamoun, S. (2009). Emerging concepts in effector biology of plant-associated organisms. **Molecular Plant-Microbe Interactions** : MPMI, 22(2): 115‑122.

Holm, L., & Rosenström, P. (2010). Dali server: conservation mapping in 3D. *Nucleic Acids Research*, 38,, W545‑W549.

Horbach, R., Navarro-Quesada, A. R., Knogge, W., & Deising, H. B. (2011). When and how to kill a plant cell: Infection strategies of plant pathogenic fungi. Quantitative disease resistance and fungal pathogenicity in Triticeae, 168(1): 51‑62.

Huang, X. (1994). On global sequence alignment. *Bioinformatics*, 10(3): 227‑235.

Huang, X., Chen, X., Coram, T., Wang, M., & Kang, Z. (2011). Gene expression profiling of *Puccinia striiformis* f. sp. *tritici* during development reveals a highly dynamic transcriptome. *J Genet Genomics* 38: 357-71.

Hug, L. A., Baker, B. J., Anantharaman, K., Brown, C. T., Probst, A. J., Castelle, C. J., Butterfield, C.N., Hernsdorf, A.W., Amano, Y., Ise K., Suzuki, Y., Dudek, N., Relman, D.A., Finstad, K.M., Amundson, R., Thomas, B.C., & Banfield, J. F. (2016). A new view of the tree of life. *Nature Microbiology*, 1, 16048.

Illergard, K., Ardell, D. H., & Elofsson, A. (2009). Structure is three to ten times more conserved than sequence--a study of structural response in protein cores. *Proteins*, 77(3):499‑508.

Jacobs, K. A., Collins-Racie, L. A., Colbert, M., Duckett, M., Evans, C., Golden-Fleet, M., Kelleher, K., Kriz, R., La Vallie, E.R., Merberg, D., Spaulding, V., Stover, J., Williamson, M.J., & McCoy, J. M. (1999). A genetic selection for isolating cDNA clones that encode signal peptides. *Methods in Enzymology*, 303:468–479.

Jaouannet, M., Rodriguez, P. A., Thorpe, P., Lenoir, C. J. G., MacLeod, R., Escudero-Martinez, C., & Bos, J. I. (2014). Plant immunity in plant–aphid interactions. *Frontiers in Plant Science*, 5, 663.

Jelenska, J., Yao, N., Vinatzer, B. A., Wright, C. M., Brodsky, J. L., & Greenberg, J. T. (2007). A J-domain Virulence Effector of *Pseudomonas syringae* Remodels Host Chloroplasts and Suppresses Defenses. *Current Biology : CB*, *17*(6), 499–508.

Jiao, M., Yu, D., Tan, C., Guo, J., Lan, D., Han, E., Guo, J., Lan, D., Han, E., Qi, T., Voegele, R.T., Kang, Z., & Guo, J. (2017). Basidiomycete-specific PsCaMKL1 encoding a CaMK-like protein kinase is required for full virulence of *Puccinia striiformis* f. sp. *tritici*. *Environmental Microbiology*, 19(10):4177-4189.

Jin, Y., Szabo, L. J., & Carson, M. (2010). Century-old mystery of *Puccinia striiformis* life history solved with the identification of Berberis as an alternate host. *Phytopathology*, 100(5):432-435.

Jing, L., Guo, D., Hu, W., & Niu, X. (2017). The prediction of a pathogenesis-related secretome of Puccinia helianthi through high-throughput transcriptome analysis. *BMC Bioinformatics* 18(1):166.

Joly, D. L., Feau, N., Tanguay, P. and Hamelin, R. C. (2010). Comparative analysis of secreted protein evolution using expressed sequence tags from four poplar leaf rusts (*Melampsora* spp.). *BMC Genomics* 11:422–422.

Jones, J. D. G., Vance, R. E., & Dangl, J. L. (2016). Intracellular innate immune surveillance devices in plants and animals. *Science*. 354(6316).

Jones, J. D. G., & Dangl, J. L. (2006). The plant immune system. *Nature*, 444(7117), 323‑329.

Käll, L., Krogh, A., & Sonnhammer, E. L. (2004). A Combined Transmembrane Topology and Signal Peptide Prediction Method. *Journal of Molecular Biology*, 338(5), 1027‑1036.

Kamel, L., Tang, N., Malbreil, M., San Clemente, H., Le Marquer, M., Roux, C., & Frei dit Frey, N. (2017). The comparison of expressed candidate secreted proteins from two arbuscular mycorrhizal fungi unravels common and specific molecular tools to invade different host plants. *Front. Plant Sci*. 8:124.

Kamoun, S. (2007). Groovy times: filamentous pathogen effectors revealed. *Current Opinion in Plant Biology*, 10 (4): 358–65.

Kamoun, S. (2009). The secretome of plant-associated fungi and oomycetes. In Plant Relationships, edited by Holger B. Deising, The Mycota, 5:173–80.

Kamper, J., Kahmann, R., Bolker, M., Ma, L.-J., Brefort, T., Saville, B. J., Banuett, F., Kronstad, J.W., Gold, S.E., Muller, O., Perlin, M.H., Wosten, H.A.B., deVries, R., Ruiz-Herrera, J., Reynaga-Pena, C.G., Snetselaar, K., McCann, M., Perez-Martin, J., Feldbrugge, M., Basse, C.W., Steinberg, G., Ibeas, J., Holloman, W., Guzman, P., Farman, M., Stajich, J.E., Sentandreu, R., Gonzalez-Prieto, J.M., Kennell, J.C., Molina, L., Schirawski, J,. Mendoza-Mendoza, A., Greilinger, D., Munch, K., Rossel, N., Scherer, M., Vranes, M., Ladendorf, O., Vincon, V., Fuchs, U., Sandrock, B., Meng, S., Ho, E.C.H., Cahill, M.J., Boyce, K.J., Klose, J., Klosterman, S.J., Deelstra, H.J., Ortiz-

Castellanos, L., Li, W., Sanchez-Alonso, P., Schreier, P.H., Hauser-Hahn, I., Vaupel, M., Koopmann, E., Friedrich, G., Voss, H., Schluter, T., Margolis, J., Platt, D., Swimmer, C., Gnirke, A., Chen, F., Vysotskaia, V., Mannhaupt, G., Guldener, U., Munsterkotter, M., Haase, D., Oesterheld, M., Mewes, H.-W., Mauceli, E.W., DeCaprio, D., Wade, C.M., Butler, J., Young, S., Jaffe, D.B., Calvo, S., Nusbaum, C., Galagan, J., & Birren, B.W. (2006). Insights from the genome of the biotrophic fungal plant pathogen *Ustilago maydis*. *Nature*, 444(7115), 97‐101.

Kanneganti, T-D., Bai, X., Tsai, C-W., Win, J., Meulia, T., Goodin, M., Kamoun, S., & Hogenhout, S. A. (2007). A functional genetic assay for nuclear trafficking in plants. *The Plant Journal : For Cell and Molecular Biology*, 50 (1): 149–58.

Karimi, M., Inze, D., & Depicker, A. (2002). GATEWAY vectors for Agrobacterium-mediated plant transformation. *Trends in Plant Science,* 7(5), 193‐195.

Katoh, K., Misawa, K., Kuma, K., & Miyata, T. (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research*, 30(14), 3059‐3066.

Kemen A. C., Alger M. T., & Kemen E. (2015). Host-microbe and microbe-microbe interactions in the evolution of obligate plant parasitism. *New phytologist*. 206:1207-1228.

Kemen, E., Kemen, A., Ehlers, A., Voegele, R., & Mendgen, K. (2013). A novel structural effector from rust fungi is capable of fibril formation. *The Plant Journal*, 75(5), 767–780.

Kemen, E., and Jones, J. D. G. (2012). Obligate biotroph parasitism: can we link genomes to lifestyles? *Trends in Plant Science*, 17(8), 448‐457.

Kemen, E., Kemen, A. C., Rafiqi, M., Hempel, U., Mendgen, K., Hahn, M., & Voegele, R. T. (2005). Identification of a protein from rust fungi transferred from haustoria into infected plant cells. *Molecular Plant-Microbe Interactions : MPMI*, 18(11), 1130–1139.

Khang, C. H., Berruyer, R., Giraldo, M. C., Kankanala, P., Park, S.-Y., Czymmek, K., Kang, S., Valent, B. (2010). Translocation of Magnaporthe oryzae Effectors into Rice Cells and Their Subsequent Cell-to-Cell Movement. *The Plant Cell*, 22(4): 1388‐1403.

Khatib, M., Lafitte, C., Esquerré-Tugayé, M.-T., Bottin, A., & Rickauer, M. (2004). The CBEL elicitor of *Phytophthora parasitica* var. *nicotianae* activates defence in Arabidopsis thaliana via three different signalling pathways. *New Phytologist*, 162(2), 501‐510.

Kim, J.-Y., Park, S.-C., Hwang, I., Cheong, H., Nah, J.-W., Hahm, K.-S., & Park, Y. (2009). Protease Inhibitors from Plants with Antimicrobial Activity. *International Journal of Molecular Sciences*, 10(6)

Kloppholz, S., Kuhn, H., & Requena, N. (2011). A Secreted Fungal Effector of *Glomus intraradices* Promotes Symbiotic Biotrophy. *Current Biology*, 21(14): 1204‐1209.

Kohler, A., Kuo, A., Nagy, L.G., Morin, E., Barry, K.W., Buscot, F., Canbäck, B., Choi, C., Cichocki, N., Clum, A., Colpaert, J., Copeland, A., Costa, M.D., Doré, J., Floudas, D., Gay, G., Girlanda, M., Henrissat, B., Herrmann, S., Hess, J., Högberg, N., Johansson, T., Khouja, H-R., LaButti, K., Lahrmann, U., Levasseur, A., Lindquist, E.A., Lipzen, A., Marmeisse, R., Martino, E., Murat, C., Ngan, C.Y., Nehls, U., Plett, J.M., Pringle, A., Ohm, R.A., Perotto, S., Peter, M., Riley, R., Rineau, F., Ruytinx, J., Salamov, A., Shah, F., Sun, H., Tarkka, M., Tritt, A., Veneault-Fourrey, C., Zuccaro, A., Mycorrhizal Genomics Initiative Consortium, Tunlid, A., Grigoriev, I.V., Hibbett, D.S., &

Francis Martin (2015). Convergent losses of decay mechanisms and rapid turnover of symbiosis genes in mycorrhizal mutualists. *Nature Genetics*. 47: 410-415.

Kolmer, J. (2013). Leaf Rust of Wheat: Pathogen Biology, Variation and Host Resistance. *Forests*, 4(1).

Krenz, B., Jeske, H., & Kleinow, T. (2012) The induction of stromule formation by a plant DNA-virus in epidermal leaf tissues suggests a novel intra- and intercellular macromolecular trafficking route. *Front Plant Sci*, 3:291

Kryazhimskiy, S., & Plotkin, J. B. (2008). The Population Genetics of dN/dS. PLOS Genetics, 4(12), e1000304.

Kubicek, C. P., Starr, T. L., & Glass, N. L. (2014). Plant cell wall-degrading enzymes and their secretion in plant-pathogenic fungi. *Annual Review of Phytopathology*, 52:427‐451.

Kunjeti, S. G., Iyer, G., Johnson, E., Li, E., Broglie, K. E., Rauscher, G., & Rairdan, G. J. (2016). Identification of *Phakopsora pachyrhizi* Candidate Effectors with Virulence Activity in a Distantly Related Pathosystem. *Frontiers in Plant Science*, 7, 269.

Lahrmann, U., Ding, Y., Banhara, A., Rath, M., Hajirezaei, M. R., Döhlemann, S., von Wirén, N., Parniske, M., & Zuccaro, A. (2013). Host-related metabolic cues affect colonization strategies of a root endophyte. *Proc Natl Acad Sci USA* 110: 13965-13970.

Langfelder, P., & Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*, 9(1): 559.

Large, E. (1950). The advance of the fungi (Dover Publication).

Lawrence, G. J., Dodds, P. N., & Ellis, J. G. (2010). Transformation of the flax rust fungus, Melampsora lini: selection via silencing of an avirulence gene. *The Plant journal : for cell and molecular biology*, 61(2), 364‐369.

Lawrence, G. J., Dodds, P. N., & Ellis, J. G. (2007). Rust of flax and linseed caused by Melampsora lini. *Molecular Plant Pathology*, 8(4):349–364.

Lee, S.-J., & Rose, J. K. (2010). Mediation of the transition from biotrophy to necrotrophy in hemibiotrophic plant pathogens by secreted effector proteins. *Plant Signaling & Behavior*, 5(6), 769‐772.

Lee, S.-W., Han, S.-W., Sririyanum, M., Park, C.-J., Seo, Y.-S., & Ronald, P. C. (2009). A Type I–Secreted, Sulfated Peptide Triggers XA21-Mediated Innate Immunity. *Science*, 326(5954), 850.

Lee, S.-Y., & MacKinnon, R. (2004). A membrane-access mechanism of ion channel inhibition by voltage sensor toxins from spider venom. *Nature*, 430(6996), 232‐235.

Leonelli, L., Pelton, J., Schoeffler, A., Dahlbeck, D., Berger, J., Wemmer, D. E., & Staskawicz, B. (2011). Structural Elucidation and Functional Characterization of the *Hyaloperonospora arabidopsidis* Effector Protein ATR13. PLoS Pathogens, 7(12), e1002428.

Leple, J. C., Brasileiro, A. C., Michel, M. F., Delmotte, F., & Jouanin, L. (1992). Transgenic poplars: expression of chimeric genes using four different constructs. *Plant Cell Reports*, 11(3) :137‐141.

Lepoivre, P. (2003). Phytopathologie: bases moléculaires et biologiques des pathosysytèmes et fondements des stratégies de lutte (De Boeck Supérieur). Louvain-la-Neuve.

Leppik, E. E. (1965). Some Viewpoints on the Phylogeny of Rust Fungi. V. Evolution of Biological Specialization. *Mycologia*, 57(1):6–22.

Li, G., Froehlich, J. E., Elowsky, C., Msanne, J., Ostosh, A. C., Zhang, C., Awada, T., & Alfano, J. R. (2014). Distinct Pseudomonas type-III effectors use a cleavable transit peptide to target chloroplasts. *The Plant Journal : For Cell and Molecular Biology*, 77(2):310‑321.

Liao, H.-L., Chen, Y., & Vilgalys, R. (2016) Metatranscriptomic Study of Common and Host-Specific Patterns of Gene Expression between Pines and Their Symbiotic Ectomycorrhizal Fungi in the Genus *Suillus*. *PLOS Genetics*. 12: e1006348.

Lin K., Limpens E., Zhang Z., Ivanov S., Saunders D.G.O., Mu, D., Pang, F., Cao, H., Cha, H., Li T., Zhou, Q., Shang, Y., Li, Y., Sharma, T., van Velzen, R., de Ruijter, N., Aanen,D.K., Win, J., Kamoun, S., Bisseling, T., Geurts, R., & Huang, S. (2014) Single Nucleus Genome Sequencing Reveals High Similarity among Nuclei of an Endomycorrhizal Fungus. *PLoS Genetics*, 10(1): e1004078.

Lindeberg, M., Cunnac, S., & Collmer, A. (2012) Pseudomonas syringae type III effector repertoires: last words in endless arguments. *Trends Microbiol*, 20:199-208

Link, T. I., & Voegele, R. T. (2008). Secreted proteins of Uromyces fabae: similarities and stage specificity. *Molecular Plant Pathology*, 9(1): 59–66.

Link, T. I., Lang, P., Scheffler, B. E., Duke, M. V., Graham, M. A., Cooper, B., Tucker, M. L., van de Mortel, M., Voegele, R.T., Mendgen, K., Baum, T. J., & Whitham, S. A. (2014a). The haustorial transcriptomes of Uromyces appendiculatus and *Phakopsora pachyrhizi* and their candidate effector families. *Molecular Plant Pathology,* 15(4): 379–393.

Link, T., Seibel, C., & Voegele, R. T. (2014b). Early insights into the genome sequence of *Uromyces fabae*. *Frontiers in Plant Science*, 5, 587.

Link, T., Lohaus, G., Heiser, I., Mendgen, K., Hahn, M., & Voegele, R. T. (2005). Characterization of a novel NADP(+)-dependent D-arabitol dehydrogenase from the plant pathogen *Uromyces fabae*. *Biochemical Journal*, 389(2), 289–295.

Littlefield, L. J., & Heath, M. C. (1979). Ultrasrtucure of Rust Fungi. Academic Press.p.293

Liu, J., Guan, T., Zheng, P., Chen, L., Yang, Y., Huai, B., Li, D., Chang, Q., Huang, L., & Kang, Z. (2016). An extracellular Zn-only superoxide dismutase from Puccinia striiformis confers enhanced resistance to host-derived oxidative stress. *Environmental Microbiology*, 18(11), 4118–4135.

Liu C., Pedersen C., Schultz-Larsen T., Aguilar G.B., Madriz-Ordenana K., Hovmoller M.S., & Thordal-Christensen H. (2016) The stripe rust fungal effector PEC6 suppresses pattern-triggered immunity in a host species-independent manner and interacts with adenosine kinases. *New Phytologist*, 14034.

Liu, L., Sonbol, F.-M., Huot, B., Gu, Y., Withers, J., Mwimba, M., Yao, J., He, S. Y., & Dong, X. (2016). Salicylic acid receptors activate jasmonic acid signalling through a non-canonical pathway to promote effector-triggered immunity. *Nature Communications*, 7, 13099.

Liu, J.-J., Sturrock, R. N., Sniezko, R. A., Williams, H., Benton, R., & Zamany, A. (2015). Transcriptome analysis of the white pine blister rust pathogen *Cronartium ribicola*: de novo assembly, expression profiling, and identification of candidate effectors. *BMC Genomics*, 16, 678.

Liu, J., Wang, Q.-L., Chang, Q., Han, L.-N., Pei, G.-L., Xue, Y.-Q., Jia, L.-M., Zhang, K., Duan, Y.-Y., & Kang, Z. (2014). Isocitrate lyase is required for urediniospore germination of *Puccinia striiformis* f. sp. *tritici*. *Molecular Biology Reports*, 41(12), 7797–7806.

Liu, T., Song, T., Zhang, X., Yuan, H., Su, L., Li, W., Xu, L., Liu, S., Chen, L., Chen, T., Zhang, M., Gu, L., Zhang, B., & Dou, D. (2014). Unconventionally secreted effectors of two filamentous pathogens target plant salicylate biosynthesis. *Nature Communications*, 5, 4686.

Liu, T., Ye, W., Ru, Y., Yang, X., Gu, B., Tao, K., Lu, S., Dong, S., Zheng, X., Shan, W., Wang, Y., & Dou, D. (2011). Two host cytoplasmic effectors are required for pathogenesis of Phytophthora sojae by suppression of host defenses. *Plant Physiology* 155 (1).

Lo Presti, L., Lanver, D., Schweizer, G., Tanaka, S., Liang L., Tollot, M., Zuccaro, A., Reissmann, S., Kahmann, R. (2015). Fungal effectors and plant susceptibility. *Annual Review of Plant Biology*, 66(1): 513–545.

Loehrer, M., Botterweck, J., Jahnke, J., Mahlmann, D. M., Gaetgens, J., Oldiges, M., Horbach, R., Deising, H., & Schaffrath, U. (2014a). In vivo assessment by Mach-Zehnder double-beam interferometry of the invasive force exerted by the Asian soybean rust fungus (*Phakopsora pachyrhizi*). *The New Phytologist*, 203(2), 620–631.

Loehrer, M., Vogel, A., Huettel, B., Reinhardt, R., Benes, V., Duplessis, S., Usadel, B., Schaffrath, U. (2014b). On the current status of *Phakopsora pachyrhizi* genome sequencing. *Frontiers in Plant Science*, 5, 377.

Lorrain, C., Marchal, C., Hacquard, S., Delaruelle, C., Petrowski, J., Petre, B., Hecker, A., Frey, P., & Duplessis, S. (2018a). The rust fungus *Melampsora larici-populina* expresses a conserved genetic program and distinct sets of secreted protein genes during infection of its two host plants, larch and poplar. *Molecular Plant-Microbe Interactions : MPMI.*

Lorrain, C., Petre, B., & Duplessis, S. (2018b) Show me the way: rust effector targets in heterologous plant systems. *Current Opinion in Microbiology*, *accepted in press*.

Lorrain, C., Hecker, A., & Duplessis, S. (2015). Effector-Mining in the Poplar Rust Fungus Melampsora larici-populina Secretome. *Frontiers in Plant Science*, 6, 1051.

Lorrain, C., Kamoun, S., & Petre, B. (2014a) Nicotiana benthamiana plastidial Methionine Sulfoxide Reductase A (NbpMSRA) localizes in the stroma of chloroplasts. *figshare*.

Lorrain, C., Kamoun, S., & Petre, B. (2014b) Populus trichocarpa Coproporphyrinogen III oxidase (PtCPO) localizes in sub-chloroplastic structures. *figshare*.

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, 15(12), 550.

Lowe, R. G. T., & Howlett, B. J. (2012). Indifferent, Affectionate, or Deceitful: Lifestyles and Secretomes of Fungi. *PLOS Pathogens*, 8(3), e1002515.

Ma L., Lukasik E., Gawehns F., & Takken F.L.W. (2012) The use of agroinfiltration for transient expression of plant resistance and fungal effector proteins in Nicotiana benthamiana leaves. *Methods Mol. Biol*. 835:61–74.

MacLean, A.M., Orlovskis Z., Kowitwanich K., Zdziarska A.M., Angenent G.C., Immink, R. G. H., & Hogenhout, S. A. (2014) Phytoplasma Effector SAP54 Hijacks Plant Reproduction by Degrading

MADS-box Proteins and Promotes Insect Colonization in a RAD23-Dependent Manner. *PLoS Biol,* 12(4): e1001835.

Maia, T., Badel, J. L., Marin-Ramirez, G., Rocha, C. de M., Fernandes, M. B., da Silva, J. C. F., de Azevedo-Junior, G.-H., & Brommonschenkel, S. H. (2017). The *Hemileia vastatrix* effector HvEC-016 suppresses bacterial blight symptoms in coffee genotypes with the SH 1 rust resistance gene. *The New Phytologist*, 213(3), 1315–1329.

Mak, A.N., Bradley, P., Cernadas, R.A., Bogdanove, A.J., & Stoddard, B.L. (2012) The crystal structure of TAL effector PthXo1 bound to its DNA target. *Science*, 335:716-719

Maqbool, A., Saitoh, H., Franceshetti, M., Stevenson, C.E.M., Uremura, A., Kanzaki, H., Terauchi, R., & Banfield, M.J. (2015) Structural basis of pathogen recognition by an integrated HMA domain in a plant NLR immune receptor. *eLife*, 4:e08709.

Marcos, E., Basanta, B., Chidyausiku, T. M., Tang, Y., Oberdorfer, G., Liu, G., Swapna, G.V.T., Guan, R., Silva, D.-A., Dou, J., Pereira, J. H., Xiao, R., Sankaran, B., Zwart, P. H., Montelione, G. T., & Baker, D. (2017). Principles for designing proteins with cavities formed by curved β sheets. *Science*, 355(6321), 201.

Martin, F. M., Uroz, S., & Barker, D. G. (2017). Ancestral alliances: Plant mutualistic symbioses with fungi and bacteria. *Science*, 356(6340).

Martin, F., Kohler, A., Murat, C., Veneault-Fourrey, C., & Hibbett, D.S. (2016) Unearthing the roots of ectomycorrhizal symbioses. *Nat. Rev. Microbiol*. 14: 760-773.

Martin, F., Kohler, A., Murat, C., Balestrini, R., Coutinho, P. M., Jaillon, Montanini, B., Morin, E., Noel, B., Percudani, R., Porcel, B., Rubini, A, Amicucci, A., Amselem, J., Anthouard, V., Arcioni, S., Artiguenave, F., Aury, J.-M., Ballario, P., Bolchi, A., Brenna, A., Brun, A., Buée, M., Cantarel, B., Chevalier, G., Couloux, A., Da Silva, C., Denoeud, F., Duplessis, S., Ghignone, S., Hilselberger, B. Iotti, M., Marçais, B., Mello, A., Miranda, M., Pacioni, G., Quesneville, H., Riccioni, C., Ruotolo, R., Splivallo, R., Stocchi, V., Tisserant, E., Viscomi, A.R., Zambonelli, A., Zampieri, E., Henrissat, B., Lebrun, M.-H., Paolocci, F., Bonfante, P., Ottonello, S., Wincker, P. (2010). Périgord black truffle genome uncovers evolutionary origins and mechanisms of symbiosis. *Nature*, 464, 1033.

Martin, F., Aerts, A., Ahrén, D., Brun, A., Danchin, E.G.J., Duchaussoy, F., Gibon, J., Kohler, A., Lindquist, E., Pereda, V., Salamov, A., Shapiro, H.J., Wuyts, J., Blaudez, D., Buée, M., Brokstein, P., Canbäck, B., Cohen, D., Courty, P.E., Coutinho, P.M., Delaruelle, C., Detter, J.C., Deveau, A., DiFazio, S., Duplessis, S., Fraissinet-Tachet, L., Lucic, E., Frey-Klett, P., Fourrey, C., Feussner, I., Gay, G., Grimwood, J., Hoegger, P.J., Jain, P., Kilaru, S., Labbé, J., Lin, Y.C., Legué, V., LeTacon, F., Marmeisse, R., Melayah, D., Montanini, B., Muratet, M., Nehls, U., Niculita-Hirzel, H., Secq, M.P.Oudot-Le, Peter, M., Quesneville, H., Rajashekar, B., Reich, M., Rouhier, N., Schmutz, J., Yin, T., Chalot, M., Henrissat, B., Kües, U., Lucas, S., VandePeer, Y., Podila, G.K., Polle, A., Pukkila, P.J., Richardson, P.M., Rouzé, P., Sanders, I.R., Stajich, J.E., Tunlid, A., Tuskan, G., Grigoriev, I.V. (2008). The genome of Laccaria bicolor provides insights into mycorrhizal symbiosis. *Nature*, 452, 88.

McCann, H. C., & Guttman, D. S. (2008). Evolution of the type III secretion system and its effectors in plant-microbe interactions. *The New Phytologist*, 177(1), 33‐47.

McDowell, J. M. (2011). Genomes of obligate plant pathogens reveal adaptations for obligate parasitism. *Proceedings of the National Academy of Sciences of the United States of America*, 108(22), 8921‐8922.

McTaggart, A. R., Shivas, R. G., van der Nest, M. A., Roux, J., Wingfield, B. D., & Wingfield, M. J. (2016). Host jumps shaped the diversity of extant rust fungi (Pucciniales). *The New Phytologist*, 209(3), 1149‐1158.

Mendgen, K. (1984) Development and physiology of teliospores. *In The Cereal Rust*, (Bushnell W. R., Roelfs A. P), pp. 362-384.

Miller, M. E., Zhang, Y., Omidvar, V., Sperschneider, J., Schwessinger, B., Raley, C., Raley, C., Palmer, J. M., Garnica, D., Upadhyaya, N., Rathjen, J., Taylor, J .M., Park, R. F., Dodds, P. N., Hirsch, C. D., Kianian, S. F., & Figueroa, M. (2017). De novo assembly and phasing of dikaryotic genomes from two isolates of *Puccinia coronata* f. sp. *avenae*, the causal agent of oat crown rust. *Preprint in BioRxiv.*

Mims, C. W. (1977). Ultrastructure of teliospore formation in the cedar-apple rust fungus *Gymnosporangium juniperi-virginianae*. *Canadian Journal of Botany*, 55(17): 2319-2329.

Mireault, C., Paris, L.-E., & Germain, H. (2014). Enhancement of the Arabidopsis floral dip method with XIAMETER OFX-0309 as alternative to Silwet L-77 surfactant. *Botany*, 92(7) 523‐525.

Mitchum, M. G., Hussey, R. S., Baum, T. J., Wang, X., Elling, A. A., Wubben, M., & Davis, E. L. (2013). Nematode effector proteins: an emerging paradigm of parasitism. *New Phytologist*, 199(4): 879‐894.

Moller, M., & Stukenbrock, E. H. (2017). Evolution and genome architecture in fungal plant pathogens. *Nature Reviews. Microbiology*, 15(12): 756–771.

Moller, S., Croning, M. D., & Apweiler, R. (2001). Evaluation of methods for the prediction of membrane spanning regions. *Bioinformatics*,17(7): 646‐653.

Moscou, M. J., & van Esse, H. P. (2017) The quest for durable resistance. Science 358:1541-1542.

Nazareno, E. S., Li, F., Smith, M., Park, R. F., Kianian, S. F., & Figueroa, M. (2017). *Puccinia coronata* f. sp. *avenae*: a threat to global oat production. *Molecular Plant Pathology*. In press.

Nelson, R., Wiesner-Hanks, T., Wisser, R., & Balint-Kurti, P. (2018). Navigating complexity to breed disease-resistant crops. *Nature genetics reviews*, 19: 21.

Nelson, B.K., Cai, X., and Nebenführ, A. (2007) A multicolored set of in vivo organelle markers for co-localization studies in Arabidopsis and other plants. *Plant Journal*, 51:1126-1136

Nemri, A., Saunders, D. G. O., Anderson, C., Upadhyaya, N. M., Win, J., Lawrence, G. J., Jones, D. A., Kamoun, S., Ellis, J. G., & Dodds, P. N. (2014). The genome sequence and effector complement of the flax rust pathogen *Melampsora lini. Frontiers in Plant Science*, 5, 98.

Oganesyan, N., Ankoudinova, I., Kim, S.-H., & Kim, R. (2007). Effect of osmotic stress and heat shock in recombinant protein overexpression and crystallization. Protein Expression and Purification, 52(2): 280‐285.

Okmen, B. and Doehlemann, G. (2014). Inside plant: biotrophic strategies to modulate host immunity and metabolism. Current Opinion in Plant Biology 20.

Oksanen, J., Blanchet, G.F., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., Minchin, P.R., O'Hara, R.B., Simpson, L., Solymos, P., Stevens, M.H.H.S., Szoecs, E., & Wagner, H. (2017) The vegan package. *Community Ecol.* Package.

Oliveira-Garcia, E., & Valent, B. (2015). How eukaryotic filamentous pathogens evade plant recognition. *Host–microbe Interactions: Fungi/Parasites/Viruses*, 26, 92–101.

Panwar, V., McCallum, B., & Bakkeren, G. (2013) Host-induced gene silencing of wheat leaf rust fungus *Puccinia triticina* pathogenicity genes mediated by the Barley stripe mosaic virus. *Plant Mol Biol* 81(6):595-608.

Palmer, I., & Wingfield, P. T. (2012). Preparation and Extraction of Insoluble (Inclusion-Body) In Proteins from *Escherichia coli.* Current Protocols in Protein Science, eds John E. Coligan.

Pearson, R. C., Aldwinckle, H. S., & Seem, R. C. (1977). Teliospore germination and basidiospore formation in *Gymnosporangium juniperi-virginianae*: a regression model of temperature and time effects. *Canadian Journal of Botany*, 55(22), 2832–2837.

Pendleton, A. L., Smith, K. E., Feau, N., Martin, F. M., Grigoriev, I. V., Hamelin, R., Nelson, C. D., Burleigh, J. G., & Davis, J. M. (2014). Duplications and losses in gene families of rust pathogens highlight putative effectors. *Frontiers in Plant Science*, 5, 299.

Pennisi, E. (2010). Armed and dangerous. *Science*, 327(5967):804–805.

Periyannan, S. 2017. Wheat Rust Disease, *Methods in Molecular Biology*, vol 1659.

Perlin, M. (2012). Essential letters in the fungal alphabet : ABC and MFS transporters and their roles in survival and pthogenicity. *Advances in Genetics,* 85 : 211-249.

Pernaci, M. (2015). Etude des traits d'histoire de vie de Melampsora laricipopulina, agent de la rouille du peuplier : de leur déterminisme génétique à leurs conséquences évolutives (Université de Lorraine). *Thèse de doctorat.*

Pernaci, M., De Mita, S., Andrieux, A., Pétrowski, J., Halkett, F., Duplessis, S. & Frey, P. (2014). Genome-wide patterns of segregation and linkage disequilibrium: the construction of a linkage genetic map of the poplar rust fungus Melampsora larici-populina. *Frontiers in Plant Science*, 5:454.

Persoons, A., Hayden, K. J., Fabre, B., Frey, P., De Mita, S., Tellier, A., & Halkett, F. (2017). The escalatory Red Queen: Population extinction and replacement following arms race dynamics in poplar rust. *Molecular Ecology*, 26(7), 1902‐1918.

Persoons, A. (2015). Les contournements de résistance par Melampsora larici-populina, l'agent de la rouille du peuplier : impact démographique et déterminisme génétique (Université de Lorraine). *Thèse de doctorat.*

Persoons, A., Morin, E., Delaruelle, C., Payen, T., Halkett, F., Frey, P., De Mita, S. & Duplessis, S. (2014). Patterns of genomic variation in the poplar rust fungus Melampsora larici-populina identify pathogenesis-related factors. *Frontiers in Plant Science,* 5: 450.

Petersen, T. N., Brunak, S., von Heijne, G., & Nielsen, H. (2011). SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nature Methods*, 8(10), 785‐786.

Peterson, P., D. (1992). James E. Teschemacher and the Cause and Management of Potato Blight in the United States. *Plant Disease*, 76(7), 754‑756.

Petre B., Win, J., Menke F.L.H., & Kamoun S. (2017) Protein-Protein Interaction Assays with Effector-GFP Fusions in Nicotiana benthamiana. *Methods Mol. Biol*. 1659: 85–98.

Petre B., Saunders D.G.O., Sklenar J., Lorrain C., Krasileva K.V., Win J., Duplessis S., & Kamoun S. (2016a) Heterologous Expression Screens in *Nicotiana benthamiana* Identify a Candidate Effector of the Wheat Yellow Rust Pathogen that Associates with Processing Bodies. *PloS One*, 11, e0149035.

Petre, B., Lorrain, C., Saunders, D. G. O., Win, J., Sklenar, J., Duplessis, S., & Kamoun, S. (2016b). Rust fungal effectors mimic host transit peptides to translocate into chloroplasts. *Cellular Microbiology*, 18(4), 453‑465.

Petre B., Saunders D.G.O., Sklenar J., Lorrain C., Win J., Duplessis S., & Kamoun S. (2015) Candidate Effector Proteins of the Rust Pathogen *Melampsora larici-populina* Target Diverse Plant Cell Compartments. *Mol. Plant-Microbe Interact*. 28:689–700

Petre B., Joly D.L., & Duplessis S. (2014) Effector proteins of rust fungi. *Front Plant Sci* 5, 416.

Petre B., & Kamoun S. (2014) How do filamentous pathogens deliver effector proteins into plant cells? *PLoS Biololgy*. 12, e1001801.

Petre, B., Morin, E., Tisserant, E., Hacquard, S., Da Silva, C., Poulain, J., Delaruelle, C., Martin, F., Rouhier, N., Kohler, A., and Duplessis, S. (2012) RNA-Seq of early-infected poplar leaves by the rust pathogen *Melampsora larici-populina* uncovers PtSultr3;5, a fungal-induced host sulfate transporter. *PloS One,* 7: e44408.

Pfaffl, M.W. (2001) A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res*. 29: e45.

Pinon, J., & Frey, P. (2005). Interactions between poplar clones and Melampsora populations and their implications for breeding for durable resistance. *In Rust diseases of Willow and Poplar* (p.139‑154). CAB International, Wallingford.

Plett, J.M., Tisserant, E., Brun, A., Morin, E., Grigoriev, I.V., Kuo, A., Martin, F., & Kohler, A. (2015) The mutualist *Laccaria bicolor* expresses a core gene regulon during the colonization of diverse host plants and a variable regulon to counteract host-specific defenses. *Mol. Plant-Microbe Interact*. 28: 261-273.

Plett, J. M., Daguerre, Y., Wittulsky, S., Vayssières, A., Deveau, A., Melton, S. J., Kohler, A., Morrell-Falvey, J. L., Brun, A., Veneault-Fourrey, C., & Martin, F. (2014). Effector MiSSP7 of the mutualistic fungus *Laccaria bicolor* stabilizes the *Populus* JAZ6 protein and represses jasmonic acid (JA) responsive genes. *Proceedings of the National Academy of Sciences of the United States of America*, 111(22), 8299‑8304.

Pretsch, K., Kemen, A., Kemen, E., Geiger, M., Mendgen, K., & Voegele, R. (2013). The rust transferred proteins-a new family of effector proteins exhibiting protease inhibitor function. *Molecular Plant Pathology,* 14(1), 96–107.

Pritchard, L., & Birch, P. R. J. (2014). The zigzag model of plant–microbe interactions: is it time to move on? *Molecular Plant Pathology*, 15(9): 865‑870.

Punta, M., Coggill, P. C., Eberhardt, R. Y., Mistry, J., Tate, J., Boursnell, C., Pang, F., Forslund, K., Ceric, G., Clements, J., Heger, A., Holm, A., Sonnhammer, E. L. L., Eddy, S. R., Bateman,A., & Finn, R. D. (2012). The Pfam protein families database. *Nucleic Acids Research*, 40, D290‑D301.

Pusztahelyi, T., Hold, I.J., & Pósci, I. (2015). Secondary metabolites in fungus-plant interactions. *Frontiers in Plant Science*, 6, 573.

Qi M., Grayczyk J.P., Seitz J.M., Lee Y., Link T.I., Choi D., Pedley K., Voegele R., Baum T., & Whitham S.A. (2017) Suppression or activation of immune responses by predicted secreted proteins of the soybean rust pathogen *Phakopsora pachyrhizi*. *Mol. Plant. Microbe Interact*. 31(1):163-174.

Qi M., Link T.I., Muller M., Hirschburger D., Pudake R.N., Pedley K.F., Braun E., Voegele R.T., Baum T.J., & Whitham S.A. (2016) A Small Cysteine-Rich Protein from the Asian Soybean Rust Fungus, Phakopsora pachyrhizi, Suppresses Plant Immunity. *PLoS Pathog*. 12, e1005827.

Qi, T., Zhu, X., Tan, C., Liu, P., Guo, J., Kang, Z., & Guo, J. (2017). Host-induced gene silencing of an important pathogenicity factor PsCPK1 in *Puccinia striiformis* f. sp. *tritici* enhances resistance of wheat to stripe rust. *Plant Biotechnology Journal*.pp1-11.

Quesneville, H., Bergman, C. M., Andrieu, O., Autard, D., Nouaud, D., Ashburner, M., & Anxolabehere, D. (2005). Combined Evidence Annotation of Transposable Elements in Genome Sequences. *PLOS Computational Biology*, 1(2), e22.

Quesneville, H., Nouaud, D., & Anxolabehere, D. (2003). Detection of new transposable element families in Drosophila melanogaster and Anopheles gambiae genomes. *Journal of Molecular Evolution*, 57 Suppl 1, S50-59.

Raffaele, S., & Kamoun, S. (2012). Genome evolution in filamentous plant pathogens: why bigger can be better. *Nature Reviews Microbiology*, 10, 417.

Rafiqi, M., Ellis, J. G., Ludowici, V.A., Hardham, A. R. & Dodds, P.N. (2012). Challenges and progress towards understanding the role of effectors in plant-fungal interactions. *Current Opinion in Plant Biology*, 15 (4): 477–82.

Rafiqi, M., Gan, P. H. P., Ravensdale, M., Lawrence, G. J., Ellis, J. G., Jones, D. A., Hardham, A. R., Dodds, P. N. (2010). Internalization of Flax Rust Avirulence Proteins into Flax and Tobacco Cells Can Occur in the Absence of the Pathogen. *The Plant Cell*, 22(6), 2017.

Ragazzi, A., Longo, N., Naldini, B., Moricca, S., & Dellavalle, I. (2005) Basidiospore-derived penetration by species of Cronartium and Melampsora an outline. *In Rust Diseases of Willow and Poplar*, (M.H Pei and A.R McCracken), pp. 161-174.

Ramachandran, S. R., Yin, C., Kud, J., Tanaka, K., Mahoney, A. K., Xiao, F., & Hulbert, S. H. (2017). Effectors from Wheat Rust Fungi Suppress Multiple Plant Defense Responses. *Phytopathology*, 107(1), 75–83.

Ramos, A. P., Tavares, S., Tavares, D., Silva, M. D. C., Loureiro, J., & Talhinhas, P. (2015). Flow cytometry reveals that the rust fungus, *Uromyces bidentis* (Pucciniales), possesses the largest fungal genome reported--2489 Mbp. *Molecular Plant Pathology*, 16(9), 1006–1010.

Ricci, P., Bonnet, P., Huet, J. C., Sallantin, M., Beauvais-Cante, F., Bruneteau, M., Billard, V., Michel, G., & Pernollet, J. C. (1989). Structure and activity of proteins from pathogenic fungi

Phytophthora eliciting necrosis and acquired resistance in tobacco. *European Journal of Biochemistry*, 183(3), 555‑563.

Rinaldi, C., Kohler, A., Frey, P., Duchaussoy, F., Ningre, N., Couloux, A., Wincker, P., Le Thiec, D., Fluch, S., Martin, F., & Duplessis, S. (2007). Transcript Profiling of Poplar Leaves upon Infection with Compatible and Incompatible Strains of the Foliar Rust *Melampsora larici-populina*. *Plant Physiology*, 144(1), 347–366.

Rinke, C., Schwientek, P., Sczyrba, A., Ivanova, N. N., Anderson, I. J., Cheng, J.-F., Darling, A., Malfatti, S., Swan, B. K., Gies, E. A., Dodsworth, J. A., Hedlund, B.P., Tsiamis, G., Sievert, S. M., Liu, W.-T., Eisen, J. A., Hallam, S. J., Kyrpides, N. C., Stepanauskas, R., Rubin, E. M., Hugenholtz, P., & Woyke, T. (2013). Insights into the phylogeny and coding potential of microbial dark matter. *Nature*, 499, 431.

Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., & Smyth, G. K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research*, 43(7), e47‑e47.

Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, 26(1), 139‑140.

Rodríguez-Herva, J.J., González-Melendi, P., Cuartas-Lanza, R., Antúnez-Lamas, M., Río-Alvarez, I., Li, Z., et al (2012) A bacterial cysteine protease effector protein interferes with photosynthesis to suppress plant innate immune responses. *Cell Microbiol* 14:669-681

Rochi, L., Dieguez, M.J., Burguener, G., Darino, M.A., Pergolesi, M.F., Ingala, L.R., Cuyeu, A.R., Turjanski, A., Kreff, E.D., & Sacco, F. (2016) Characterization and comparative analysis of the genome of *Puccinia sorghi* Schwein, the causal agent of maize common rust. *Fungal Genet. Biol*. In press.

Rooney, H. C. E., Van't Klooster, J. W., van der Hoorn, R. A. L., Joosten, M. H. A. J., Jones, J. D. G. & de Wit, P. J. G. M. (2005) *Cladosporium* Avr2 inhibits tomato rcr3 protease required for cf-2-dependent disease resistance. *Science*, 308 (5729): 1783–86.

Rosano, G. L., & Ceccarelli, E. A. (2014). Recombinant protein expression in *Escherichia coli*: advances and challenges. *Frontiers in Microbiology*, 5, 172.

Rouhier, N., Vieira Dos Santos, C., Tarrago, L., & Rey, P. (2006) Plant methionine sulfoxide reductase A and B multigenic families. *Photosynth Res*, 89:247-262

Rouxel, T., Grandaubert, J., Hane, J. K., Hoede, C., van de Wouw, A. P., Couloux, A., Dominguez, V., Anthouard, V., Bally, P., Bourras, S., Cozijnsen, A. J., Ciuffetti, L. M., Degrave, A., Dilmaghani, A., Duret, L., Fudal, I., Goodwin, S. B., Gout, L., Glaser, N., Linglin, J., Kema, G. H. J., Lapalu, N., Lawrence, C., May, K., Meyer, B., Ollivier, B., Poulain, J., Schoch, C. L., Simon, A., Spatafora, J. W., Stachowiak, A., Turgeon, B. G., Tyler, B. M., Vincent, D., Weissenbach, J., Amselem, J., Quesneville, H., Oliver, R. P., Wincker, P., Balesdent, M.-H, & Howlett, B. J. (2011). Effector diversification within compartments of the *Leptosphaeria maculans* genome affected by Repeat-Induced Point mutations. *Nature Communications*, 2, 202.

Rovenich, H., Boshoven, J. C., & Thomma, B. P. H. J. (2014). Filamentous pathogen effector functions: of pathogens, hosts and microbiomes. *Current Opinion in Plant Biology*, 20, 96‑103.

Rutter, W.B., Salcedo, A., Akhunova, A., He, F., Wang, S., Liang, H., Bowden, R.L., & Akhunov, E. (2017) Divergent and convergent modes of interaction between wheat and *Puccinia graminis* f. sp. *tritici* isolates revealed by the comparative gene co-expression network and genome analyses. *BMC Genomics*, 18: 291.

Salcedo, A., Rutter, W., Wang, S., Akhunova, A., Bolus, S., Chao, S., Anderson, N., De Soto, M.F., Rouse, M., Szabo, L., Bowden, R.L., Dubcovsky, J., & Akhunov, E. (2017) Variation in the AvrSr35 gene determines Sr35 resistance against wheat stem rust race Ug99. *Science*, 358:1604-1606.

Saunders D.G.O., Win J., Cano L.M., Szabo L.J., Kamoun S., & Raffaele S. (2012) Using hierarchical clustering of secreted protein families to classify and rank candidate effectors of rust fungi. *PloS One*, 7, e29847.

Saunders, D. G. O., Win, J., Kamoun, S., & Raffaele, S. (2014). Two-Dimensional Data Binning for the Analysis of Genome Architecture in Filamentous Plant Pathogens and Other Eukaryotes. *In Plant-Pathogen Interactions: Methods and Protocols (*Paul Birch et al., Vol. 1127, p. 29‑51). Springer Science

Savile, D. B. O. (1976). Evolution of the Rust Fungi (Uredinales) as Reflected by Their Ecological Problems. In M. K. Hecht, W. C. Steere, & B. Wallace (Eds.), *Evolutionary Biology*: Volume 9 (pp. 137–207).

Schenk, P.M., Carvalhais, L.C., & Kazan, K. (2012). Unraveling plant-microbe interactions: can multi-species transcriptomics help? *Trends Biotechnol.* 30: 177-184.

Schillberg, S., Tiburzy, R., & Fischer, R. (2000). Transient transformation of the rust fungus *Puccinia graminis* f. sp. *tritici*. *Molecular & General Genetics*, 262(6), 911‑915.

Schornack S., Van Damme M., Bozkurt T.O., Cano L.M., Smoker M., Thines M., Gaulin E., Kamoun S., & Huitema E. (2010) Ancient class of translocated oomycete effectors targets the host nucleus. *Proc. Natl. Acad. Sci. U.S.A.,* 107:17421–17426

Schulze-Lefert, P., & Panstruga, R. (2011). A molecular evolutionary concept connecting nonhost resistance, pathogen host range, and pathogen speciation. *Trends in Plant Science*, 16(3), 117‑125.

Schulze-Lefert, P., & Panstruga, R. (2011) A molecular evolutionary concept connecting nonhost resistance, pathogen host range, and pathogen speciation. *Trends Plant Sci*. 16: 117-125.

Schwessinger, B. (2017). Fundamental wheat stripe rust research in the 21st century. *The New Phytologist*, 213(4), 1625–1631.

Schwessinger, B., Sperschneider, J., Cuddy, W., Miller, M., Garnica, D., Taylor, J., Dodds, P. N., Figueroa, M., Park, R., & Rathjen, J. (2017). A near complete haplotype-phased genome of the dikaryotic wheat stripe rust fungus *Puccinia striiformis* f. sp. *tritici* reveals high inter-haplome diversity. *bioRxiv*.

Selosse, M.-A. (2000). La symbiose. Structure et fonctions, rôle écologique et évolutif (Broché). Vuibert.

Sharpee, W., Oh, Y., Franck, W., Eyre, A., Okagaki, L., Valent, B., & Dean, R. A. (2017). Identification and characterization of suppressors of plant cell death (SPD) effectors from *Magnaporthe oryzae*. *Molecular Plant Pathology*, 6(18), 850‑863.

Sharpee, W.C., & Dean, R.A. (2016) Form and function of fungal and oomycete effectors. *Fungal Biol. Rev*. 30: 62-73.

Shen, A. (2014). Simplified Protein Purification Using an Autoprocessing, Inducible Enzyme Tag. *Methods in molecular biology*, 1177, 59‐70.

Shen, Q-H., Saijo, Y., Mauch, S., Biskup, C., Bieri, S., Keller, B., Seki, H., Ulker, B., Somssich, I. E., & Schultze-Lefert, P. (2007). Nuclear Activity of MLA Immune receptors links isolate-specific and basal disease-resistance responses. *Science*, 315 (5815): 1098–1103.

Shi, L.X., & Theg, S.M. (2013) The chloroplast protein import system: from algae to trees. *Biochim Biophys Acta*, 1833:314-331

Sohn, K.H., Hughes, R.K., Piquerez, S.J., Jones, J.D., & Banfield, M.J. (2012) Distinct regions of the *Pseudomonas syringae* coiled-coil effector AvrRps4 are required for activation of immunity. *Proc Natl Acad Sci USA*, 109:16371-16376

Singh, R. P., Hodson, D. P., Jin, Y., Lagudah, E. S., Ayliffe, M. A., Bhavani, S., Rouse, B. M., Pretorius, Z. A., Szabo, L. J., Huerta-Espino, J., Basnet, B. R., Lan, C., & Hovmoller, M. S. (2015). Emergence and Spread of New Races of Wheat Stem Rust Fungus: Continued Threat to Food Security and Prospects of Genetic Control. *Phytopathology*, 105(7), 872–884.

Sjamsuridzal, W., Nishida, H., Ogawa, H., Kakishima, M., & Sugiyama, J. (1999). Phylogenetic positions of rust fungi parasitic on ferns: Evidence from 18S rDNA sequence analysis. *Mycoscience*, 40(1), 21–27.

Slaminko, T. L., Miles, M. R., Frederick, R. D., Bonde, M. R., & Hartman, G. L. (2008). New Legume Hosts of *Phakopsora pachyrhizi* Based on Greenhouse Evaluations. *Plant Disease*, 92(5), 767–771.

Sniezko, R.A., Yanchuk, A.D., Kliejunas, J.T., Palmieri, K.M., Alexander, J.M., Frankel, S.J. (2012) Proceedings of the Fourth International Workshop on the Genetics of Host-Parasite Interactions in Forestry: Disease and Insect Resistance in Forest Trees. General Technical Report PSW-GTR-240. Albany, CA: U.S. Department of Agriculture, Forest Service, Pacific Southwest Research Station. 372 p

Song, J., Win, J., Tian, M., Schornack, S., Kaschani, F., Ilyas, M., van der Hoorn, R., & Kamoun, S. (2009). Apoplastic effectors secreted by two unrelated eukaryotic plant pathogens target the tomato defense protease Rcr3. *Proceedings of the National Academy of Sciences*, 106(5), 1654.

Spanu, P.D. (2015) RNA-protein interactions in plant disease: hackers at the dinner table. New Phytol 207(4):991-995.

Spanu, P. D., & Panstruga, R. (2012). Powdery mildew genomes in the crosshairs. *New Phytologist*, 195(1), 20‐22.

Spanu, P. D., Abbott, J. C., Amselem, J., Burgis, T. A., Soanes, D. M., Stuber, K., et al. (2010). Genome expansion and gene loss in powdery mildew fungi reveal tradeoffs in extreme parasitism. *Science*, 330(6010), 1543‐1546.

Sperschneider, J., Dodds, P. N., Gardiner, D. M., Singh, K. B., & Taylor, J. M. (2018). Improved prediction of fungal effector proteins from secretomes with EffectorP 2.0. *BioRxiv*.

Sperschneider J., Taylor J., Dodds P.N., & Duplessis S. (2017a). Computational methods for predicting effectors in rust pathogens. *Wheat Rust Disease - Methods and Protocols*, Springer in press,

Sperschneider J., Catanzariti A.M., DeBoer K., Petre B., Gardiner D.M., Singh K.B., Dodds P.N., & Taylor J.M. (2017b) LOCALIZER: subcellular localization prediction of both plant and effector proteins in the plant cell. *Scientifique Report*, 7,44598.

Sperschneider J., Dodds P.N., Singh K.B., & Taylor J.M. (2017c) ApoplastP: prediction of effectors and plant proteins in the apoplast using machine learning. *bioRxiv*.

Sperschneider J., Gardiner D.M., Dodds P.N., Tini F., Covarelli L., Singh K.B., Manners, J.M., & Taylor J.M. (2016) EffectorP: predicting fungal effector proteins from secretomes using machine learning. *New Phytologist*. 210:743–761.

Sperschneider, J., Dodds, P. N., Gardiner, D. M., Manners, J. M., Singh, K. B. & Taylor, J. M. (2015). Advances and Challenges in Computational Prediction of Effectors from Plant Pathogenic Fungi. *PLoS Pathogens*, 11(5), e1004806.

Stam R., Jupe J., Howden A.J.M., Morris J.A., Boevink P.C., Hedley P.E., & Huitema E. (2013) Identification and Characterisation CRN Effectors in Phytophthora capsici Shows Modularity and Functional Diversity. *PloS ONE*, 8,e59517.

Staples, R. C. (2000). Research on the rust fungi during the twentieth century. *Annual Review of Phytopathology*, 38(1), 49‑69.

Staples, R. C. (2001). Nutrients for a rust fungus: the role of haustoria. *Trends in Plant Science*, 6.

Stergiopoulos, I., Cordovez, V., Ökmen, B., Beenen, H. G., Kema, G. H. J., & de Wit, P. J. G. M. (2014). Positive selection and intragenic recombination contribute to high allelic diversity in effector genes of *Mycosphaerella fijiensis*, causal agent of the black leaf streak disease of banana. *Molecular Plant Pathology*, 15(5), 447–460.

Stergiopoulos, I., Collemare, J., Mehrabi, R. & De Wit, P. J. G. M. (2013). Phytotoxic secondary metabolites and peptides produced by plant pathogenic dothideomycete fungi. *FEMS Microbiology Reviews*, 37 (1).

Stergiopoulos, I., & De Wit, P. J. G. M. (2009) Fungal effector proteins. *Annual Review of Phytopathology*, 47: 233–63.

Stewart, A., & Cromey, M. (2011). Identifying disease threats and management practices for bio-energy crops. *Current Opinion in Environmental Sustainability*, 3(1), 75–80.

Struck, C. (2015). Amino acid uptake in rust fungi. *Frontiers in Plant Science*, 6, 40.

Struck, C., Siebels, C., Rommel, O., Wernitz, M., & Hahn, M. (1998). The plasma membrane H(+)-ATPase from the biotrophic rust fungus *Uromyces fabae*: molecular characterization of the gene (PMA1) and functional expression of the enzyme in yeast. *Molecular Plant-Microbe Interactions*, 11(6), 458–465.

Struck, Hahn, & Mendgen. (1996). Plasma Membrane H+-ATPase Activity in Spores, Germ Tubes, and Haustoria of the Rust Fungus Uromyces viciae-fabae. *Fungal Genetics and Biology*, 20(1), 30–35.

Stukenbrock, E. H. (2013). Evolution, selection and isolation: a genomic view of speciation in fungal plant pathogens. *The New Phytologist*, 199(4), 895–907.

Talhinhas, P., Batista, D., Diniz, I., Vieira, A., Silva, D. N., Loureiro, A., Tavares, S., Pereira, A.-P., Azinheira, H.-G., Guerra-Guimaraes, L., Varzea, V., & Silva, M. do C. (2016). The coffee leaf rust pathogen Hemileia vastatrix: one and a half centuries around the tropics. *Molecular Plant Pathology*.

Talhinhas, P., Azinheira, H.G., Vieira, B., Loureiro, A., Tavares, S., Batista, D., Morin, E., Petitot, A.S., Paulo, O.S., Poulain, J., Da Silva, C., Duplessis, S., Silva, Mdo. C., & Fernandez, D. (2014) Overview of the functional virulent genome of the coffee leaf rust pathogen Hemileia vastatrix with an emphasis on early stages of infection. *Front Plant Sci,* 5: 88.

Tam, J. P., Lu, Y.-A., Yang, J.-L., & Chiu, K.-W. (1999). An unusual structural motif of antimicrobial peptides containing end-to-end macrocycle and cystine-knot disulfides. Proceedings of the National *Academy of Sciences of the United States of America,* 96(16), 8913‐8918.

Tan, K.-C., Ipcho, S. V. S., Trengove, R. D., Oliver, R. P., & Solomon, P. S. (2009). Assessing the impact of transcriptomics, proteomics and metabolomics on fungal phytopathology. *Molecular Plant Pathology*, 10(5), 703‐715.

Tanaka, S., Brefort, T., Neidig, N., Djamei, A., Kahnt, J., Vermerris, W., Koenig, S., Feussner, K., Feussner, I., & Kahmann, R. (2014). A secreted *Ustilago maydis* effector promotes virulence by targeting anthocyanin biosynthesis in maize. *eLife*, 3, e01355.

Tang C.L., Wang X.J., Cheng Y.L., Liu M.J., Zhao M.X., Wei J. P., & Kang Z. S. (2015) New insights in the battle between wheat and *Puccinia striiformis*. *Front Agr Sci Eng*, 2:101–14

Tang, C., Xu, Q., Zhao, M., Wang, X., & Kang, Z. (2017). Understanding the lifestyles and pathogenicity mechanisms of obligate biotrophic fungi in wheat: The emerging genomics era. *The Crop Journal*.

Tang, C., Wei, J., Han, Q., Liu, R., Duan, X., Fu, Y., Huang, X., Wang, X., & Kang, Z. (2015). PsANT, the adenine nucleotide translocase of *Puccinia striiformis*, promotes cell death and fungal growth. *Scientific Reports*, 5, 11241.

Tao, S.-Q., Cao, B., Tian, C.-M., & Liang, Y.-M. (2017). Comparative transcriptome analysis and identification of candidate effectors in two related rust species (*Gymnosporangium yamadae* and *Gymnosporangium asiaticum*). *BMC Genomics*, 18(1), 651.

Tavares, S., Ramos, A. P., Pires, A. S., Azinheira, H. G., Caldeirinha, P., Link, T., Abranches, R., Silva, M. D. C., Voegele, R. T., Loureiro,J., & Talhinhas, P. (2014). Genome size analyses of Pucciniales reveal the largest fungal genomes. *Frontiers in Plant Science*, 5, 422.

Teixeira, P. J. P. L., Thomazella, D. P. de T., Reis, O., do Prado, P. F. V., do Rio, M. C. S., Fiorin, G. L., José, J., Costa, C. J. L., Negri, V. A., Mondego, J. M. C., Mieczkowski, P., & Pereira, G. A. G. (2014). High-Resolution Transcript Profiling of the Atypical Biotrophic Interaction between *Theobroma cacao* and the Fungal Pathogen *Moniliophthora perniciosa*. *The Plant Cell*, 26(11), 4245‐4269.

Teixeira, P.F., & Glaser, E. (2013) Processing peptidases in mitochondria and chloroplasts. *Biochim Biophys Acta,* 1833:360-370

Terhune, B. T., Bojko, R. J., & Hoch, H. C. (1993). Deformation of Stomatal Guard Cell Lips and Microfabricated Artificial Topographies during Appressorium Formation by Uromyces. Experimental *Mycology*, 17(1), 70–78.

Terpe, K. (2006). Overview of bacterial expression systems for heterologous protein production: from molecular and biochemical fundamentals to commercial systems. *Applied Microbiology and Biotechnology*, 72(2), 211‐222.

Thomma, B. P. H. J., Nürnberger, T., & Joosten, M. H. A. J. (2011). Of PAMPs and Effectors: The Blurred PTI-ETI Dichotomy. *The Plant Cell*, 23(1), 4–15.

Tobias, P. A., Guest, D. I., Kulheim, C., Hsieh, J.-F., & Park, R. F. (2016). A curious case of resistance to a new encounter pathogen: myrtle rust in Australia. *Molecular Plant Pathology*, 17(5), 783–788.

Tonkin, C.J., Foth, B.J., Ralph, S.A., Struck, N., Cowman, A.F., & McFadden, G.I. (2008) Evolution of malaria parasite plastid targeting sequences. *Proc Natl Acad Sci USA*, 105:4781-4785

Toome-Heller, M. (2016). Latest developments in the research of rust fungi and their allies (Pucciniomycotina). In D.-W. Li (Ed.), Biology of microfungi (pp. 147–168). Cham: Springer International Publishing.

Toome, M., Ohm, R. A., Riley, R. W., James, T. Y., Lazarus, K. L., Henrissat, B., Albu, S., Boyd, A., Chow, J., Clum, A., Heller, G., Lipzen, A., Nolan, M., Sandor, L., Zvenigorodsky, N., Grigoriev, I. V., Spatafora, J. W., & Aime, M. C. (2014). Genome sequencing provides insight into the reproductive biology, nutritional mode and ploidy of the fern pathogen Mixia osmundae. *New Phytologist*, 202(2), 554–564.

Torres Zabala, M., Littlejohn, G., Jayamaran, S., Studholme, D., Bailey, T., Lawson, T., Tillich, M., Licht, D., Bölter, B., Delfino, L., Truman, W., Mansfield, J., Smirnoff, N., & Grant, M. (2015) Chloroplasts play a central role in plant defence and are targeted by pathogen effectors. *Nat Plants*, 1:15074

Tremblay, A., Hosseini, P., Li, S., Alkharouf, N. W., & Matthews, B. F. (2013). Analysis of Phakopsora pachyrhizi transcript abundance in critical pathways at four time-points during infection of a susceptible soybean cultivar using deep sequencing. *BMC Genomics*, 14(1), 614.

Tuskan, G. A., DiFazio, S., Jansson, S., Bohlmann, J., Grigoriev, I., Hellsten, U., *et al.* (2006). The Genome of Black Cottonwood, *Populus trichocarpa*. *Science*, 313(5793), 1596.

Tyler, B. M., & Rouxel, T. (2012). Effectors of Fungi and Oomycetes: Their Virulence and Avirulence Functions and Translocation From Pathogen to Host Cells. In *Molecular Plant Immunity* (p. 123–167). Wiley-Blackwell.

Upadhyaya, N. M., Garnica, D. P., Karaoglu, H., Sperschneider, J., Nemri, A., Xu, B., Mago, E., Cuomo, C. A., Rathjen, J. P., Park, R. F., Ellis, J. G., Dodds, P. N. (2014). Comparative genomics of Australian isolates of the wheat stem rust pathogen *Puccinia graminis* f. sp. *tritici* reveals extensive polymorphism in candidate effector genes. *Frontiers in Plant Science*, 5, 759.

Valent, B., & Khang, C. H. (2010). Recent advances in rice blast effector research. *Current Opinion in Plant Biology*, 13(4), 434–441.

van der Heijden, M.G.A., Martin, F.M., Selosse, M.-A., & Sanders, I.R. (2015) Mycorrhizal ecology and evolution: the past, the present, and the future. *New Phytol*. 205: 1406-1423.

Ve, T., Williams, S. J., Catanzariti, A.-M., Rafiqi, M., Rahman, M., Ellis, J. G., Hardham, A. R., Jones, D. A., Anderson, P. A., Dodds, P. N., & Kobe, B. (2013). Structures of the flax-rust effector AvrM reveal insights into the molecular basis of plant-cell entry and effector-triggered immunity. *Proceedings of the National Academy of Sciences of the United States of America*, 110(43), 17594–17599.

Vialle, A., Frey, P., Hambleton, S., Bernier, L., & Hamelin, R. C. (2011). Poplar rust systematics and refinement of *Melampsora* species delineation. Fungal Diversity, 50(1), 227.

Vialle A., Feau N., Frey P., Bernier L., & Hamelin R.C. (2013). Phylogenetic species recognition reveals host-specific lineages among poplar rust fungi. *Molecular Phylogenetics and Evolution* 66: 628–644.

Viennot-Bourgin, G. (1981). The downy mildews (Academic Press). Spencer DM.

Voegele, R. T., & Mendgen, K. (2011). Nutrient uptake in rust fungi: how sweet is parasitic life? *Euphytica*, 179, 41–55.

Voegele, R. T., Hahn, M., & Mendgen, K. (2009). The Uredinales: Cytology, Biochemistry, and Molecular Biology. *In The Mycota* (A Comprehensive Treatise on Fungi as Experimental Systems for Basic and Applied Research) (Deising H.B. (eds) Plant Relationships, Vol. 5, pp. 69–98)

Voegele, R. T., Wirsel, S., Moll, U., Lechner, M., & Mendgen, K. (2006). Cloning and characterization of a novel invertase from the obligate biotroph *Uromyces fabae* and analysis of expression patterns of host and pathogen invertases in the course of infection. *Molecular Plant-Microbe Interactions*, 19(6), 625–634.

Voegele, R. T., Hahn, M., Lohaus, G., Link, T., Heiser, I., & Mendgen, K. (2005). Possible Roles for Mannitol and Mannitol Dehydrogenase in the Biotrophic Plant Pathogen Uromyces fabae. Plant Physiology, 137(1), 190–198.

Voegele, R. T., & Mendgen, K. (2003). Rust haustorium: Nutrient uptake and beyond. *New Phytol*, 159

Voegele, R. T., Struck, C., Hahn, M., & Mendgen, K. (2001). The role of haustoria in sugar supply during infection of broad bean by the rust fungus *Uromyces fabae*. *Proc Natl Acad Sci USA*, 98.

Wang C.I.A., Guncar G., Forwood J.K., Teh T., Catanzariti A.M., Lawrence G.J., Loughlin F.E., Mackay J.P., Schirra H.J., Anderson P.A., Ellis J.G, Dodds P.N., & Kobe B. (2007) Crystal structures of flax rust avirulence proteins AvrL567-A and -D reveal details of the structural basis for flax disease resistance specificity. *Plant Cell*, 19:2898–2912.

Wang, Y., Yu, B., Zhao, J., Guo, J., Li, Y., Han, S., Huang, L., Du, Y., Hong, Y., Tang, D., & Liu, Y. (2013) Autophagy contributes to leaf starch degradation. *Plant Cell*, 25:1383-1399

Wang X., Yang B., Li K., Kang Z., Cantu D., & Dubcovsky J. (2016) A Conserved *Puccinia striiformis* Protein Interacts with Wheat NPR1 and Reduces Induction of Pathogenesis-Related Genes in Response to Pathogens. *Mol. Plant-Microbe Interact*., 29:977–989.

Wawra, S., Trusch, F., Matena, A., Apostolakis, K., Linne, U., Zhukov, I., Stanek, I., Koźmiński, W., Davidson, I., Secombes, C. J., Bayer, P., & van West, P. (2017). The RxLR Motif of the Host Targeting Effector AVR3a of *Phytophthora infestans* Is Cleaved Before Secretion. *The Plant Cell*. In press.

Weber, E., Gruetzner, R., Werner, S., Engler, C., & Marillonnet, S. (2011) Assembly of designer TAL effectors by Golden Gate cloning. *PLoS One*, 6:e19722

Weber, H., & Roitsch, T. (2000). Invertases and life beyond sucrose cleavage. *Trends in Plant Science*, 5(2), 47–48.

Weiberg, A., Wang, M., Lin, F.-M., Zhao, H., Zhang, Z., Kaloshian, I., Huang, D.-H., & Jin, H. (2013). Fungal Small RNAs Suppress Plant Immunity by Hijacking Host RNA Interference Pathways. *Science*, 342(6154), 118‑123.

Wicker, T., Oberhaensli, S., Parlange, F., Buchmann, J. P., Shatalina, M., Roffler, S., Roi, B.-D., Dolezel, J., Simkova, H., Schulze-Lefert, P., Spanu, P. D., Bruggmann, R., Amselem, J., Quesneville, H., Ver Loren van Themaat, E., Paape, K., Shimizu, K. K., & Keller, B. (2013). The wheat powdery mildew genome shows the unique evolution of an obligate biotroph. *Nature Genetics*, 45(9), 1092‑1096.

Wiese, J., Kleber, R., Hampp, R., & Nehls, U. (2000). Functional Characterization of the *Amanita muscaria* Monosaccharide Transporter, AmMst1. *Plant Biology*, 2(3), 278–282.

Williams, P. G. (1984). Obligate parasitism and axenic culture. In *The cereal rust volume* I origins, specificity, structure and pysiology (Academic Press, Vol. 1, pp. 399–430).

Wilson, R. A., & Talbot, N. J. (2009). Under pressure: investigating the biology of plant infection by *Magnaporthe oryzae*. *Nature Reviews Microbiology*, 7, 185.

Win J., Chaparro-Garcia A., Belhaj K., Saunders D.G., Yoshida K., Dong S., Schornack S., Zipfel C., Robatzek S., Hogenhout S.A. & Kamoun S. (2012) Effector biology of plant-associated organisms: concepts and perspectives. *Cold Spring Harb Symp Quant Biol*, 77.

Win J., Kamoun S., & Jones A.M.E. (2011) Purification of effector-target protein complexes via transient expression in *Nicotiana benthamiana*. *Methods Mol. Biol. Clifton*, NJ 712:181–194.

Win, J., Morgan, W., Bos, J., Krasileva, K. V., Cano, L. M., Chaparro-Garcia, A., Ammar, R., Staskawicz, B. J., & Kamoun, S. (2007). Adaptive Evolution Has Targeted the C-Terminal Domain of the RXLR Effectors of Plant Pathogenic Oomycetes. The Plant Cell, 19(8), 2349‑2369.

Wirthmueller, L., Roth, C., Fabro, G., Caillaud, M.C., Rallapalli, G., Asai, S., *et al* (2015) Probing formation of cargo/importin-α transport complexes in plant cells using a pathogen effector. *Plant J* 81:40-52

Wirthmueller, L., Maqbool, A., & Banfield, M. J. (2013). On the front line: structural insights into plant-pathogen interactions. *Nature Reviews*. *Microbiology*, 11(11):761‑776.

Wishart, D. S., Bigam, C. G., Holm, A., Hodges, R. S., & Sykes, B. D. (1995). 1H, 13C and 15N random coil NMR chemical shifts of the common amino acids. I. Investigations of nearest-neighbor effects. *Journal of Biomolecular NMR*, 5(1), 67‑81.

Wu, J., Wang, Q., Liu, S., Huang, S., Mu, J., Zeng, Q., Huang, L., Han, D., & Kang, Z. (2017). Saturation Mapping of a Major Effect QTL for Stripe Rust Resistance on Wheat Chromosome 2B in Cultivar Napo 63 Using SNP Genotyping Arrays. *Frontiers in Plant Science*, 8, 653.

Xu, J., Linning, R., Fellers, J., Dickinson, M., Zhu, W., Antonov, I., Joly, D.L., Donaldson, M.E., Eilam, T., Anikster, Y., Banks, T., Munro, S., Mayo, M., Wynhoven, B., Ali, J., Moore, R., McCallum, B., Borodovsky, M., Saville, B., & Bakkeren, G. (2011) Gene discovery in EST sequences from the wheat leaf rust fungus *Puccinia triticina* sexual spores, asexual spores and haustoria, compared to other rust and corn smut fungi. *BMC Genomics*, 12: 161.

Yaeno, T., Li, H., Chaparro-Garcia, A., Schornack, S., Koshiba, S., Watanabe, S., Kigawa, T., Kamoun, S., & Shirasu, K. (2011). Phosphatidylinositol monophosphate-binding interface in the

oomycete RXLR effector AVR3a is required for its stability in host cells to modulate plant immunity. *Proceedings of the National Academy of Sciences of the United States of America*, 108(35), 14682–14687.

Yin, C., Downey, S. l., Klages-Mundt, N. L., Ramachandran, S., Chen, X., Szabo, L. J., Kigawa, T., Kamoun, S., & Hulbert, S. H. (2015). Identification of promising host-induced silencing targets among genes preferentially transcribed in haustoria of *Puccinia*. *BMC Genomics*, 16(1), 579.

Yin, C., Park, J.-J., Gang, D. R., & Hulbert, S. H. (2014). Characterization of a tryptophan 2-monooxygenase gene from *Puccinia graminis* f. sp. *tritici* involved in auxin biosynthesis and rust pathogenicity. *Molecular Plant-Microbe Interactions*, 27(3), 227–235.

Yin, C., Jurgenson, J. E., & Hulbert, S. H. (2011). Development of a host-induced RNAi system in the wheat stripe rust fungus *Puccinia striiformis* f. sp. *tritici*. *Molecular Plant-Microbe Interactions*, 24(5), 554–561.

Yuan, C., Wang, M., Skinner, D. Z., See, D. R., Xia, C., Guo, X., & Chen, X. (2018). Inheritance of Virulence, Construction of a Linkage Map, and Mapping Dominant Virulence Genes in *Puccinia striiformis* f. sp. *tritici* Through Characterization of a Sexual Population with Genotyping-by-Sequencing. *Phytopathology*, 108(1), 133–141.

Zale, J., Freshour, L., Agarwal, S., Sorochan, J., Ownley, B. H., Gwinn, K. D., & Castlebury, L. A. (2008). First Report of Rust on Switchgrass (*Panicum virgatum*) Caused by *Puccinia emaculata* in Tennessee. *Plant Disease*, 92(12), 1710–1710.

Zhang, Q., Wang, B., Wei, J., Wang, X., Han, Q., & Kang, Z. (2018). TaNTF2, a contributor for wheat resistance to the stripe rust pathogen. *Plant Physiology and Biochemistry*, 123, 260–267.

Zhang, X., Nguyen, N., Breen, S., Outram, M. A., Dodds, P. N., Kobe, B., Solomon, P. S., & Williams, S. J. (2017a). Production of small cysteine-rich effector proteins in *Escherichia coli* for structural and functional studies. *Molecular Plant Pathology*, 18(1), 141–151.

Zhang X., Farah N., Rolston L., Ericsson D.J., Catanzariti A.M., Bernoux M., Ve T., Bendak K., Chen C., Mackay J.P., Lawrence G.J., Hardham A., Ellis J.G., Williams S.J., Dodds P.N., Jones D.A.,& Kobe B. (2017b) Crystal structure of the *Melampsora lini* effector AvrP reveals insights into a possible nuclear function and recognition by the flax disease resistance protein. *P. Mol. Plant Pathol.*,12597

Zhang, C., & Clarke, P. R. (2000). Chromatin-Independent Nuclear Envelope Assembly Induced by Ran GTPase in XenopusEgg Extracts. *Science*, 288(5470), 1429.

Zhang, H., Gao, S., Lercher, M. J., Hu, S., & Chen, W.-H. (2012). EvolView, an online tool for visualizing, annotating and managing phylogenetic trees. *Nucleic Acids Research*, 40:W569–W572.

Zheng, X., McLellan, H., Fraiture, M., Liu, X., Boevink, P.C., Gilroy, E.M., *et al* (2014) Functionally redundant RXLR effectors from *Phytophthora infestans* act at different steps to suppress early flg22-triggered immunity. *PLoS Pathog* 10:e1004057

Zheng, W., Huang, L., Huang, J., Wang, X., Chen, X., Zhao, J., Guo, J. et al. (2013). High genome heterozygosity and endemic genetic recombination in the wheat stripe rust fungus. *Nature Communications* 4: 2673.

Zhou, J.-M., & Chai, J. (2008). Plant pathogenic bacterial type III effectors subdue host responses. *Current Opinion in Microbiology*, 11(2).

Zhu, X., Liu, W., Chu, X., Sun, Q., Tan, C., Yang, Q., Jiao, M., Guo, J., & Kang, Z. (2017). The transcription factor PstSTE12 is required for virulence of *Puccinia striiformis f. sp. tritici. Molecular Plant Pathology*.

ANNEXES

Liste des annexes :

Annexe 1 : Article n°7: « *Candidate effector proteins of the rust pathogen Melampsora larici-populina target diverse plant cell compartments* »

Annexe 2 : Article n°8: « *Heterologous expression screens in Nicotiana benthamiana identify a candidate effector of the wheat yellow rust pathogen that associates with processing bodies* »

Annexe 1

Article n°7: « **Candidate effector proteins of the rust pathogen *Melampsora larici-populina* target diverse plant cell compartments** »

Benjamin Petre, Diane G. O. Saunders, Jan Sklenar, Cécile Lorrain, Joe Win, Sébastien Duplessis, Sophien Kamoun

Article de recherche publié dans la revue *Molecular plant-microbe interactions*, le 26 janvier 2015.
https://apsjournals.apsnet.org/doi/abs/10.1094/MPMI-01-15-0003-R

CANDIDATE EFFECTOR PROTEINS OF THE RUST PATHOGEN *MELAMPSORA LARICI-POPULINA* TARGET DIVERSE PLANT CELL COMPARTMENTS

Benjamin Petre[1,2,3], Diane G.O. Saunders[1,4,5], Jan Sklenar[1], Cécile Lorrain[1,2,3], Joe Win[1], Sébastien Duplessis[2,3*], Sophien Kamoun[1*]

[1] The Sainsbury Laboratory, Norwich Research Park, NR4 7UH Norwich, United Kingdom
[2] INRA, UMR 1136 Interactions Arbres/Microorganismes, Centre INRA Nancy Lorraine, 54280 Champenoux, France
[3] Université de Lorraine, UMR 1136 Interactions Arbres/Microorganismes, Faculté des Sciences et Technologies, 54506 Vandoeuvre-lès-Nancy, France
[4] The Genome Analysis Centre, Norwich Research Park, NR4 7UH Norwich, United Kingdom
[5] The John Innes Centre, Norwich Research Park, NR4 7UH Norwich, United Kingdom

* Sophien Kamoun, The Sainsbury Laboratory, Norwich Research Park, NR4 7UH Norwich, United Kingdom
Sophien.Kamoun@tsl.ac.uk

* Sébastien Duplessis, UMR 1136 Interactions Arbres/Microorganismes, Centre INRA Nancy Lorraine, 54280 Champenoux, France
duplessi@nancy.inra.fr

ABSTRACT

Rust fungi are devastating crop pathogens that deliver effector proteins into infected tissues to modulate plant functions and promote parasitic growth. The genome of the poplar leaf rust fungus *Melampsora larici-populina* revealed a large catalogue of secreted proteins, some of which have been considered candidate effectors. Unravelling how these proteins function in host cells is key to understanding pathogenicity mechanisms and developing resistant plants. In this study, we used an effectoromics pipeline to select, clone, and express 20 candidate effectors in *Nicotiana benthamiana* leaf cells to determine their subcellular localisation and identify the plant proteins they interact with. Confocal microscopy revealed that six candidate effectors target the nucleus, nucleoli, chloroplasts, mitochondria and discrete cellular bodies. We also used coimmunoprecipitation and mass spectrometry to identify 606 *N. benthamiana* proteins that associate with the candidate effectors. Five candidate effectors specifically associated with a small set of plant proteins that may represent biologically relevant interactors. We confirmed the interaction between the candidate effector MLP124017 and the TOPLESS-Related Protein 4 from poplar by *in planta* coimmunoprecipitation. Altogether, our data enable us to validate effector proteins from *M. larici-populina* and reveal that these proteins may target multiple compartments and processes in plant cells. It also shows that *N. benthamiana* can be a powerful heterologous system to study effectors of obligate biotrophic pathogens.

Key words: Pucciniales, cell biology, live-cell imaging, liquid-chromatography/tandem mass spectrometry, agroinfiltration, obligate biotroph, TOPLESS-Related Proteins

Annexe 2

Article n°8: « **Heterologous expression screens in *Nicotiana Benthamiana* identify a candidate effector of the wheat yellow rust pathogen that associates with processing bodies** »

Benjamin Petre, Diane G.O. Saunders, Jan Sklenar, Cécile Lorrain, Ksenia V. Krasileva, Joe Win, Sébastien Duplessis, Sophien Kamoun

Article de recherche publié dans la revue *PLOS one*, le 10 février 2016.
http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0149035

Heterologous expression screens in *Nicotiana benthamiana* identify a candidate effector of the wheat yellow rust pathogen that associates with processing bodies

Benjamin Petre[1,2,3], Diane G.O. Saunders[1,4,5], Jan Sklenar[1], Cécile Lorrain[1,2,3], Ksenia V. Krasileva[1,4], Joe Win[1], Sébastien Duplessis[2,3*], Sophien Kamoun[1*]

[1] The Sainsbury Laboratory, Norwich Research Park, Norwich, United Kingdom
[2] INRA, UMR 1136 Interactions Arbres/Microorganismes, Centre INRA Nancy Lorraine, Champenoux, France
[3] Université de Lorraine, UMR 1136 Interactions Arbres/Microorganismes, Faculté des Sciences et Technologies, Vandoeuvre-lès-Nancy, France
[4] The Genome Analysis Centre, Norwich Research Park, Norwich, United Kingdom
[5] The John Innes Centre, Norwich Research Park, Norwich, United Kingdom
* Corresponding author
E-mail: Sophien.Kamoun@tsl.ac.uk (SK)

Abstract

Rust fungal pathogens of wheat (*Triticum* spp.) affect crop yields worldwide. The molecular mechanisms underlying the virulence of these pathogens remain elusive, due to the limited availability of suitable molecular genetic research tools. Notably, the inability to perform high-throughput analyses of candidate virulence proteins (also known as effectors) impairs progress. We previously established a pipeline for the fast-forward screens of rust fungal candidate effectors in the model plant *Nicotiana benthamiana*. This pipeline involves selecting candidate effectors *in silico* and performing cell biology and protein-protein interaction assays *in planta* to gain insight into the putative functions of candidate effectors. In this study, we used this pipeline to identify and characterize sixteen candidate effectors from the wheat yellow rust fungal pathogen *Puccinia striiformis* f sp *tritici*. Nine candidate effectors targeted a specific plant subcellular compartment or protein complex, providing valuable information on their putative functions in plant cells. One candidate effector, PST02549, accumulated in processing bodies (P-bodies), protein complexes involved in mRNA decapping, degradation, and storage. PST02549 also associates with the P-body-resident ENHANCER OF mRNA DECAPPING PROTEIN 4 (EDC4) from *N. benthamiana* and wheat. We propose that P-bodies are a novel plant cell compartment targeted by pathogen effectors.

Key words

Agrobacterium tumefaciens, coimmunoprecipitation, confocal microscopy, live-cell imaging, mass spectrometry, Pucciniales, *Triticum aestivum*, varicose, wheat stripe rust

Résumé :
La maladie de la rouille foliaire du peuplier est causée par des espèces de champignons de la famille des Melampsoraceae et de l'ordre des Pucciniales. Trois espèces infectent principalement les peupleraies en France : *Melampsora larici-populina* qui est la plus agressive, *Melampsora allii-populina* et *Melampsora medusae* f. sp. *deltoidae*. Dans leurs cycles de vie, ces trois espèces partagent les peupliers comme hôtes communs pour assurer leur multiplication clonale au cours de l'été et elles infectent des hôtes différents pour réaliser leur cycle sexué au cours du printemps. Chez les champignons pathogènes biotrophes obligatoires comme les Pucciniales, la sécrétion de molécules appelées effecteurs conditionne le succès d'infection de la plante hôte. Ainsi, les effecteurs permettent de manipuler les fonctions cellulaires de l'hôte et de contourner son système immunitaire. Comprendre le rôle fondamental de ces molécules dans la cellule-hôte est une des questions clés de la biologie des effecteurs. Au cours de cette thèse, des approches de génomique, de transcriptomique et de caractérisation fonctionnelle et structurale d'effecteurs candidats ont été déployées afin de contribuer à l'analyse moléculaire de l'interaction entre peupliers et *Melampsora* spp. Premièrement, l'analyse du transcriptome de *M. larici-populina* au cours de son cycles sexué a mis en évidence que les protéines sécrétées sont surreprésentées parmi les gènes surexprimés dans les basides et au cours de l'infection de l'hôte mélèze. La comparaison des transcriptomes de *M. larici-populina* lors de l'infection des deux hôtes, le peuplier et le mélèze, révèle la présence d'une majorité de gènes exprimés communément chez les deux hôtes et une fraction exprimée spécifiquement chez chacun des deux hôtes. Ainsi, nous révélons des gènes codant de petites protéines sécrétées qui pourraient être associés à la spécificité d'hôte. Deuxièmement, le séquençage du génome de *M. allii-populina* et sa comparaison au génome de *M. larici-populina* a révélé une expansion de sa taille (336 Mb contre 110 Mb). Cette dernière est associée à l'invasion du génome par de nombreux éléments transposables qui représentent près de 75% du génome et par un enrichissement en gènes par rapport à *M. larici-populina* (23 089 gènes contre 19 550). Cependant, les répertoires de gènes potentiellement impliqués dans le processus infectieux sont proportionnellement sensiblement similaires entre les deux espèces. Les quelques différences observées pourraient être associées à la spécificité d'hôte (*Populus* spp./*Larix* spp. dans un cas et *Populus* spp./*Allium* spp. dans l'autre) et à l'adaptation à des hôtes alternants différents. Troisièmement, des cribles effectoromiques réalisés sur un répertoire d'effecteurs candidats de *M. larici-populina* ont révélé deux candidats d'intérêt. Nous avons mis en place des approches fonctionnelles pour tenter d'élucider leur rôle dans l'interaction avec le peuplier. Nous avons ainsi pu produire l'effecteur MLP124017 qui interagit avec des protéines de la famille TOPLESS-RELATED PROTEINS et résoudre sa structure tridimensionnelle qui présente une similarité à des protéines NUCLEAR TRANSPORT FACTOR 2 LIKE qui sont des protéines cargo transportant diverses molécules entre le noyau et le cytoplasme. L'effecteur CHLOROPLAST-TARGETED PROTEIN MLPCTP1 est adressé vers et est localisé dans les chloroplastes *via* un peptide de transit dans le système hétérologue *Nicotiana benthamiana* et nous montrons une localisation similaire chez le peuplier et l'arabette. Les fonctions de ces effecteurs restent encore à déterminer mais les travaux amorcés dans cette thèse ouvrent de nouvelles perspectives quant à la diversité et au rôle des effecteurs chez les Pucciniales.

Abstract:
The poplar rust disease is caused by fungi belonging to the order Pucciniales in the family Melamsporaceae. Three main species cause damages in poplar plantations in France: *Melampsora larici-populina*, which is the most aggressive, *Melampsora allii-populina* and *Melampsora medusae* f. sp. *deltoidae*. During their life cycles, these three fungi share poplars as main host in which they achieve their clonal multiplication during summer and they infect different alternate hosts to complete their sexual reproduction in spring. In fungal obligate biotrophs like rust fungi, the success of host infection is conditioned by the secretion of molecules called effectors. Effectors allow the pathogen to manipulate host cell processes and to evade the plant immune system. A key question in effector biology is to understand how these molecules function in the host cell. In the frame of this PhD thesis, genomics, transcriptomics, functional and structural approaches were used to contribute to the molecular analysis of the poplar-*Melamspora* spp. interaction. Firstly, the transcriptome analysis of *M. larici-populina* during its sexual cycle highlighted that small-secreted proteins were over-represented among genes up-regulated in basidia and during larch infection. Comparison of *M. larici-populina* transcriptomes during infection of the two hosts, poplar and larch, revealed that a majority of the genes are commonly expressed on both hosts and a fraction is specifically expressed on each host. Small-secreted protein genes specifically expressed on each host maybe related to host specificity. Secondly, the sequencing of the *M. allii-populina* genome and the comparison with *M. larici-populina* revealed a size expansion in *M. allii-populina* (336Mb *vs.* 110Mb). This expansion is mostly due to the invasion of the genome by numerous transposable elements that represent more than 75% of the genome and by a gene enrichment compared to *M. larici-populina* (23,089 and 19,550 genes, respectively). However, catalogues of gene potentially involved in the infection process are proportionally similar between the two species. The few differences observed could be linked to host specificity (*Populus* spp./*Larix* spp. on one hand and *Populus* spp./*Allium* spp. on the other hand) and adaptation to their different hosts. Thirdly, effectoromic screens developed on a panel of *M. larici-populina* candidate effectors revealed two candidates of interest. We set up a pipeline of functional approaches in order to elucidate their role in the interaction with poplar. We produced the candidate effector MLP124017 that interacts with proteins of the TOPLESS-RELATED PROTEINS family and solved its structure that presents similarities with NUCLEAR TRANSPORT FACTOR 2 LIKE proteins. These proteins are cargos that transport diverse molecules between the nucleus and cytosol in plant cells. The effector CHLOROPLAST-TARGETED PROTEIN MLPCTP1 is addressed and translocated inside chloroplasts through a transit peptide in the heterologous system *Nicotianana benthamiana* and we showed that the same localization occurs in poplar and in Arabidopsis. The functions of these two effectors remain to be determined but the functional characterization initiated in this thesis opens new perspectives in term of diversity and roles of effectors in rust fungi.