



Impact des bactéries féminisantes du genre *Wolbachia* sur l'évolution des chromosomes sexuels d'isopodes terrestres

Thomas Becking

► To cite this version:

Thomas Becking. Impact des bactéries féminisantes du genre *Wolbachia* sur l'évolution des chromosomes sexuels d'isopodes terrestres. Sciences agricoles. Université de Poitiers, 2017. Français. NNT : 2017POIT2299 . tel-03188748

HAL Id: tel-03188748

<https://theses.hal.science/tel-03188748>

Submitted on 2 Apr 2021

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.

THESE

Pour l'obtention du Grade de

DOCTEUR DE L'UNIVERSITE DE POITIERS

(Faculté des Sciences Fondamentales et Appliquées)
(Diplôme National - Arrêté du 25 mai 2016)

Ecole Doctorale : Sciences pour l'environnement - Gay Lussac (La Rochelle)
Secteur de Recherche : Biologie de l'environnement, des populations, écologie

Présentée par :

Thomas BECKING

Impact des bactéries féminisantes du genre *Wolbachia* sur l'évolution des chromosomes sexuels d'isopodes terrestres

Directeurs de Thèse :

Richard CORDAUX, Clément GILBERT

Soutenue le 11 Décembre 2017 devant la Commission d'Examen

JURY

Président	Gabriel MARAIS	Directeur de recherche CNRS, Université de Lyon 1
Rapporteuse	Susana COELHO	Directrice de recherche CNRS, Station Biologique de Roscoff
Rapporteur	Nicolas PERRIN	Professeur, Université de Lausanne, Suisse
Membre	Sophie BELTRAN-BECH	Maître de conférences, Université de Poitiers
Membre	Richard CORDAUX	Directeur de recherche CNRS, Université de Poitiers
Membre	Clément GILBERT	Chargé de recherche CNRS, Université Paris-Sud

Je dédie ce manuscrit à ta mémoire.

Remerciements

Ces lignes sont peut-être les plus difficiles à rédiger, parce qu'elles marquent la fin d'une aventure, parce qu'elles sont toujours écrites en dernier (mais toujours lues en premier... Si si toi aussi je te vois lecteur !), parce que le poids de chaque mot compte. Malgré leur caractère anecdotique, ces lignes sont pourtant peut-être parmi les plus importantes, car elles témoignent de l'apport de toutes celles et ceux qui ont permis l'accomplissement de ces travaux, de près comme de loin et parfois même sans s'en rendre compte ! Une thèse n'est pas une histoire que l'on forge seul, et sans eux, rien n'aurait été possible.

Je tiens donc tout d'abord à remercier l'ensemble des membres du jury pour avoir accepté d'évaluer ces travaux. Un grand merci à mes rapporteurs Susana Coelho et Nicolas Perrin, que j'espère convaincre (si besoin en est) du formidable modèle que représentent les isopodes terrestres. Un grand merci bien entendu à Gabriel Marais pour avoir suivi mes travaux lors d'une étape charnière, pour ses conseils avisés et pour être de retour à Poitiers afin de voir le chemin parcouru depuis le comité. Un grand merci enfin à toi Sophie, qui continue de suivre mes travaux (mais pas d'écrevisse cette fois-ci) malgré une fin d'année chargée pour toi aussi.

Je tiens à remercier très vivement et chaleureusement mes directeurs de thèse, Richard Cordaux et Clément Gilbert, qui ont signé avec moi pour cette drôle d'histoire qu'est "la thèse". Clément, je crois que je me souviendrai encore longtemps de ce vendredi après-midi, lorsque tu m'avais annoncé le début de l'aventure (il m'a fallu au bas mot plus d'une heure pour finir cette solution tampon que je préparais au labo lorsque tu es venu !)... Tu es une des personnes les plus impressionnantes que je connaisse, un monstre de travail (je risquerai presque le terme de "machine" !), d'efficacité et de réactivité. Merci pour tout ce que tu m'as apporté techniquement, scientifiquement, humainement, et ce y compris à distance lors les derniers mois de thèse. Richard, c'est suite à notre voyage dans la "petite

bourgade" d'Oswego (ma première outre-Atlantique) que j'ai appris à mieux te connaître, non pas scientifiquement mais humainement. Ta culture scientifique m'impressionnera sans cesse, tu es toujours capable de tirer telle ou telle référence de ton chapeau, et ce sur de très nombreux domaines. Merci encore pour toutes ces discussions passionnantes, ces réunions "de 10 minutes, 15 grand maximum" qui au final pouvaient prendre des demies-journées entières (et c'était presque trop court !). Merci enfin pour ta disponibilité malgré un emploi du temps de plus en plus chargé, pour tes nombreux conseils, ton enthousiasme communicatif, et quoi qu'il en soit, rendez-vous quand tu veux pour taper du volant !

Bref, merci encore pour vos encouragements, vos conseils, votre bienveillance, votre enthousiasme et votre confiance (cette liste ne saurait être exhaustive) durant ces trois ans. Vous avez toujours su être stimulants, motivés et motivants. Je vous dois beaucoup, et même beaucoup plus que ce que j'ai parvenu à résumer ici.

Je souhaite bien évidemment remercier l'ensemble des membres de l'équipe, à commencer par Didier pour abattre ce travail de malade pour le labo, mais aussi pour être un des (si ce n'est le) premiers à m'avoir fait découvrir le monde de la biologie évolutive. Un grand merci à toi aussi Pierre pour permettre la bonne marche de l'équipe, mais aussi (surtout) pour les mots que tu as su me dire par écrit ou à l'oral lorsque je traversais de durs moments.

Un grand merci à toute l'équipe technique pour ces 3 (et même plus !) dernières années : Isa (de Charente Supérieure) pour ta pêche, ton dynamisme, tes conseils, bref pour être telle que tu es (et pour les billes) ! Maryline, ta patience et ta technique m'impressionneront toujours (les injections de glande androgène... Mais quelle galère !!), ainsi que ton enthousiasme communicatif (notamment pour la création d'un cloporte ailé de palladium). Carine, je sais que je me répète, mais voilà près de 5 ans que tu parviens encore et toujours à trouver 45 secondes de ton temps pour moi (t'aurais dû dire "non" dès le début !). Catherine, ta bonne humeur communicative fait toujours plaisir à voir, et à entendre ! Alex', change pas t'es une personne en or, et quand tu veux pour le combo fromage fondu + *Knacki Ball*, une tuerie ! Tiff', pour ton *smile*, ton énergie et ta disponibilité (je crois que c'est pas demain qu'on gagnera au loto, pas vrai ?). Hélène, pour ta bonne humeur p-e-

r-m-a-n-e-n-t-e, et pour tout ce que tu fais pour simplifier la vie de tous les jours des enseignants. Alina, la petite dernière de la bande, avec ta pêche communicative, je te souhaite plein de belles choses dans ce labo ! Merci aussi à toi Christelle, qui nous simplifie tellement la vie de tous les jours sans même qu'on s'en rende réellement compte. Merci enfin à toi Bouziane, pour tes conseils, ta disponibilité et ta bienveillance, et pour toujours trouver une solution à tous les problèmes bioinfo du monde...

Je souhaite remercier chaleureusement tous les enseignants-chercheurs de l'équipe, ceux que l'on appelle "les grands" ou encore "les adultes", pour ces 5 années passées dans l'équipe, et même plus encore pour ces 8 dernières années, déjà (n'est-ce pas ?). Merci donc à Joanne, Roland, Yves, Nico, Isa, Freddie-Jeanne, Jean et Julia. C'est tout de même aussi un peu grâce à vous que j'ai erré au cours de ces dernières années au sein des couloirs du labo...

Je tiens à remercier plus particulièrement Christine et Fred', sans qui ces lignes de remerciements n'existeraient même pas. Merci à vous pour votre confiance il y a quelques années déjà, et pour m'avoir permis de franchir ce pas aujourd'hui (Fred', je crois que je me souviendrai toujours de ton appel en Mai, tu m'as poussé à y aller, et tu as bien fait).

I would also like to kindly and sincerely thank Christopher Chandler (and his young team), for our truly constructive cooperation, for our exciting discussions and also for welcoming us (Isabelle, Richard and I) in the beautiful city of Oswego. I know now that you are an expert in androgenic gland injections, even on *Trachelipus rathkeii* !

Un grand merci à tous les "anciens", qui m'ont connu stagiaire eux aussi : Lise, Vico, Jessica, Mimi, Cloclo. Merci pour votre accueil, pour ces moments passés ensemble. Merci bien entendu à tous les "récents" ! Sylvine, compagne (si si ce mot existe) d'infortune depuis le M2, comme le temps passe... Mais on y est arrivé ! Plein de bonnes choses pour la suite, ne change pas (même si t'en avais pas l'intention), et merci à toi d'être telle que tu es (et pour en rajouter, encore merci pour le coup de main en stat' alors que tu avais bien d'autres choses à faire à ce moment là, paraît-il...). Merci à toi Benjamin (prononcer "BenneRamine", ça t'agace en plus) pour nos discussions scientifiques (et pas que), toutes plus

passionnées les unes que les autres. Un conseil : en cas de stress, pense *Mongozo/Kasteel rouge* : c'est la carotte qui fait avancer le bourriquot ! Marius, mon *padawan* de la ligne de commande (si on m'avait dit que j'écirai ça un jour...). Continue sur ta lancée, et répète après moi : I-S-T-I-B (et merci pour les repas lorsque ma carte ne passait plus !). Clémentine, la seule personne à travailler sur du vertébré au bureau (mais quelle idée ??). Toujours la banane et jamais la dernière pour un *mojito* (et t'as bien raison !). Et sûr la "chair fraîche" du bureau : Charlotte et Jérémy. Bienvenue à vous dans l'aventure "thèse" !

Je vous souhaite de vivre l'aventure de la même façon que j'ai pu la vivre, profitez. A tous, je vous souhaite le meilleur pour la suite (un peu le pire aussi, ça permet d'apprécier encore plus le meilleur ! Mouhahah !), ainsi que pour l'après, qui vient bien plus vite qu'on ne le croit, même si on le sait.

Un immense merci à toi Amine, mon maître *jedi* qui m'a tout appris ou presque des sombres arcanes de la bioinformatique, toujours disponible pour répondre à mes questions bêtes, et qui m'a dépanné tant de fois... Bon courage pour la suite ! Je suis optimiste ça va le faire !

Merci à toi Flo, pour nos discussions passionnées, et ce même si t'es un "paleu" ! Plein de bonnes choses pour la suite, et forcément : "Force et honneur" !

Merci aussi à toi Alexandre, t'es une chouette personne que j'apprends à connaître un peu mieux tous les jours...

Un merci aux autres membres de la Team 4 (Amber, Cyrilou, Ayalus), ça va en faire des docteurs mine de rien...

Merci aussi à vous, les *Bro'* : Sébichou (bon la prochaine fois tu te maries un peu plus près, hein ?), Thomassss (l'ingénieur à Grenoble...), Quentin (la route est encore longue mais accroche toi, muscle ton jeu !), Chachar (ENFIN on s'y est remis bon sang...). Mention spéciale à toi Niko, "programmeur" au grand cœur (ça rime donc c'est forcément vrai !) qui en connais un rayon sur les cloportes désormais, même si t'es un peu manichéen parfois (j'ai honte... Désolé je l'assume pas vraiment celui-là...).

Bien sûr, je remercie toute ma petite famille pour son soutien indéfectible : Maman, Claire, Léanne. Merci d'avoir encore d'avoir été là, et je sais que désormais vous ne regarderez plus jamais les cloportes de la même façon... (et pour être honnête, moi non plus !)

Enfin, merci à toi ma 'tite loutre, pour être là tous les jours à mes côtés. Merci pour ton soutien de chaque instant (alors que tu es une des seules personnes à savoir que je suis un faux-calme...). Merci d'avoir "tenu la baraque" lors de mes séances de "camping" au labo (pour faire écho à tes propres mots issus de ton manuscrit), merci pour cette pêche i-n-é-p-u-i-s-a-b-l-e (moyennant quelques siestes le samedi après-midi !), merci pour être telle que tu es. Enfin, parce que je ne saurai dire mieux, et comme ça tu auras le dernier mot : *Mais surtout... merci pour ton amour* (Fortin, 2016).

Sommaire

Liste des abréviations.....	7
Introduction générale.....	3
I. Contexte général	5
1. Le paradoxe évolutif du sexe.....	6
2. Les mécanismes moléculaires impliqués dans la reproduction sexuée	8
3. La question de l'identité sexuelle.....	10
4. La diversité des facteurs déclencheurs du déterminisme du sexe : de l'environnement à la génétique	12
5. L'évolution des chromosomes sexuels et ses implications	16
6. Les effecteurs moléculaires impliqués dans le déterminisme du sexe	20
7. Transitions entre systèmes de déterminisme du sexe	23
II. Le modèle isopode terrestre.....	26
1. Evolution des isopodes terrestres	26
2. Anatomie	28
3. Mécanismes de déterminisme du sexe.....	30
4. Différenciation sexuelle.....	33
III. Association isopodes terrestres/<i>Wolbachia</i>.....	35
1. Présentation et historique de <i>Wolbachia</i>	35
2. Conséquences de <i>Wolbachia</i> sur le phénotype de ses hôtes.....	38
a) <i>Wolbachia, un parasite de la reproduction</i>	38
b) <i>Wolbachia, un symbiote mutualiste</i>	43
3. Conflits génétiques et conséquences sur le déterminisme du sexe.....	44
IV. Projet de thèse	48

Chapitre 1

Séquençage et assemblage du génome d'*A. nasatum*51

Introduction53

Matériels et Méthodes58

I. Séquençage *Illumina* et assemblage du génome mâle d'*A. nasatum*.....58

1. Extraction d'ADN (séquençage *Illumina*).....58
2. Séquençage *Illumina*, contrôle qualité des lectures de séquençage et filtrage des adaptateurs de séquençage59
3. Assemblage des données *Illumina*.....60

II. Séquençage *Pacific Biosciences* et assemblages du génome mâle d'*A. nasatum*63

1. Extraction, concentration et contrôles qualité de l'ADN (séquençage *PacBio*)63
2. Séquençage *Pacific Biosciences*, correction des *reads* et premier assemblage avec *CeleraAssembler*64
3. Assemblage du génome de référence d'*A. nasatum* avec *DBG2OLC*65
4. Elimination des contaminants et vérification de l'intégrité fonctionnelle du génome de référence68
5. Caractérisation des éléments répétés et annotation structurale et fonctionnelle du génome70
6. Caractérisation des insertions de *Wolbachia* dans le génome d'*A. nasatum* 73

Résultats.....75

I. Assemblage du génome mâle d'*A. nasatum* à partir des données *Illumina*75

II. Assemblages du génome mâle d'*A. nasatum* à partir des données *Illumina* et des données *Pacific Biosciences*.....78

1. Assemblage du génome d'*A. nasatum* avec *CeleraAssembler*78
2. Assemblage du génome d'*A. nasatum* avec *DBG2OLC*79

III.	Annotation structurale et fonctionnelle du génome d' <i>A. nasatum</i> et caractérisation des éléments transposables.....	84
IV.	Caractérisation des insertions de <i>Wolbachia</i> dans le génome d' <i>A. nasatum</i>	88
V.	Comparaison de la structure des génomes d' <i>A. nasatum</i> et d' <i>A. vulgare</i>	90

Discussion97

I.	Validité de l'assemblage combinant <i>reads</i> courts et longs ..	97
II.	Comparaison avec des génomes de crustacés déjà publiés..	99
III.	Comparaison des génomes d' <i>A. nasatum</i> et d' <i>A. vulgare</i> ...	101
IV.	Perspectives de recherche	107

Chapitre 2

Identification de séquences Y-spécifiques et caractérisation du chromosome Y d'*A. nasatum*.....111

Introduction113

Matériels et Méthodes118

I.	Recherche de séquences spécifiques au chromosome Y dans le génome d' <i>A. nasatum</i>	118
1.	Recherche <i>in silico</i> par étude de ratio de couverture sur le génome assemblé avec les données <i>Illumina</i>	118
2.	Recherche <i>in silico</i> par comparaison de ratio de couverture sur les génomes <i>PacBio</i> - Méthode du <i>Chromosome Quotient (CQ)</i>	120
3.	Recherche <i>in silico</i> par comparaison de <i>k-mer</i> sur les génomes <i>PacBio</i> - Méthode du <i>Y chromosome Genome Scanning (YGS)</i>	122
II.	Tests de validation <i>in vivo</i> des séquences candidates Y-spécifiques	123

Résultats.....127

I. Recherche de séquences spécifiques au chromosome Y dans le génome d'*A. nasatum* 127

1. Recherche *in silico* par étude du ratio de couverture sur le génome assemblé avec les données *Illumina* 127
2. Recherche *in silico* par la combinaison des approches CQ et YGS sur le génome assemblé avec *CeleraAssembler* 129
 - a) Approche par ratio de couverture - méthode du Chromosome Quotient (CQ) 129
 - b) Approche par comparaison de k-mer - méthode du Y chromosome Genome Scanning (YGS) et influence de la taille de k-mer lors de l'analyse 132
 - c) Combinaison des approches CQ et YGS - sélection et test des candidats Y-spécifiques..... 134
3. Recherche *in silico* par la combinaison des approches CQ et YGS sur le génome de référence assemblé avec *DBG2OLC* 137
 - a) Réutilisation des approches CQ et YGS pour la détermination *in silico* de séquences Y-spécifiques 137
 - b) Comparaison entre les analyses *in silico* réalisées sur le génome de référence avec celles effectuées sur le génome assemblé avec *CeleraAssembler* 140
4. Tests des marqueurs Y-spécifiques sur des lignées d'individus infectés par des *Wolbachia* féminisantes 144

Discussion148

I. Recherche de séquences liées au chromosome Y d'*A. nasatum*..... 148

1. Analyse des assemblages du génome d'*A. nasatum* 148
2. Perspectives concernant la caractérisation de nouvelles séquences Y-spécifiques 152

II. Relations entre *A. nasatum* et sa souche de *Wolbachia* féminisante : *wNas* 157

III. Caractérisation de l'homomorphie des chromosomes sexuels d'*A. nasatum* 160

Chapitre 3

Caractérisation de gènes candidats du déterminisme du sexe et mesure de leur expression au cours du développement167

Introduction169

Matériels et Méthodes172

I. Caractérisation *in silico* de gènes potentiellement impliqués dans le déterminisme du sexe..... 172

II. Mesure de l'expression des gènes candidats au cours du développement..... 174

1. Echantillonnage de juvéniles pour étudier l'expression des gènes candidats au cours du développement 174
2. Sexage des juvéniles *A. nasatum* 174
3. Quantification de l'expression des gènes candidats au cours du développement des mâles et femelles *A. nasatum* et *A. vulgare* 175
4. Analyses statistiques 179

Résultats.....180

I. Caractérisation *in silico* de gènes candidats impliqués dans le déterminisme du sexe..... 180

II. Description structurale des séquences candidates identifiées chez *A. nasatum* et *A. vulgare* 187

III. Comparaison de l'expression des gènes candidats entre mâles et femelles au cours du développement 195

1. Comparaison de l'expression des gènes candidats entre mâles et femelles chez *A. nasatum* 195
2. Comparaison de l'expression des gènes candidats entre *A. nasatum* et *A. vulgare* 198

Discussion	200
I. Caractérisation des gènes à domaine DM et suivi de l'expression au cours du développement	200
II. Caractérisation des gènes <i>Transformer-2</i> et suivi de l'expression au cours du développement	204
III. Perspectives de recherche	206

Chapitre 4

Diversité et évolution des systèmes de déterminisme du sexe chez les isopodes terrestres	209
---	------------

Avant-propos.....	211
--------------------------	------------

Perspectives	230
---------------------------	------------

Synthèse générale.....	251
-------------------------------	------------

Références.....	261
------------------------	------------

Annexes	293
----------------------	------------

Liste des abréviations

ADN : Acide Désoxyribonucléique.

ADNc : ADN complémentaire.

ARN : Acide Ribonucléique.

ARNm : ARN messenger.

BUSCO : Benchmarking Universal Single-Copy Orthologs.

CCS : Circular Consensus Sequence.

CpG : Cytosine-phosphate-Guanine.

CQ : Chromosome Quotient.

Ct : Cycle threshold.

DMRT : Doublesex and Mab-3 Related Transcripts.

ESD : Environmental Sex Determination.

ET : Élément Transposable.

EVE : Endogenous Viral Element.

GLM : Generalized Linear Model.

GO : Gene Ontology.

GSD : Genetic Sex Determination.

HMM : Hidden Markov Model.

IC : Incompatibilité Cytoplasmique.

LINE : Long Interspersed Nuclear Elements.

LTR : Long Terminal Repeat sequences.

Mb : Million de paires de bases.

MSCI : Meiotic Sex Chromosome Inactivation.

NGS : Next Generation Sequencing.

PAR : Pseudo-Autosomal Region.

pb : paire de bases.

PCR : Polymerase Chain Reaction.

qPCR : quantitative Polymerase Chain Reaction.

rpm : Rotations par minute.

RRM : RNA Recognition Motif.

RT : Réverse Transcription.

RT-PCR : Reverse Transcriptase Polymerase Chain Reaction.

RT-qPCR : Reverse Transcriptase quantitative Polymerase Chain Reaction.

SINE : Short Interspersed Nuclear Elements.

SMRT : Single-Molecule Real-Time.

SNP : Single Nucleotide Polymorphism.

SRY : Sex-determining Region of Y chromosome.

UTR : Untranslated Transcribed Region.

YGS : Y chromosome Genome Scanning.

Introduction générale

I. Contexte général

"Sex is the queen of problems in evolutionary biology". C'est par cette phrase que Graham Bell débute son ouvrage de référence intitulé *The masterpiece of Nature : The Evolution and Genetics of Sexuality* (Bell, 1982). L'origine du sexe, et donc de la reproduction sexuée, son maintien au cours de l'évolution, la diversité des mécanismes impliqués, sont autant de points qui font que la sexualité est considérée comme l'une des plus grandes et anciennes énigmes du Vivant. L'origine du sexe fut d'ailleurs discutée dès l'Antiquité par Aristote qui, suite à ses travaux sur l'embryologie, concluait que le sexe d'un individu était déterminé par la chaleur dégagée par le partenaire masculin lors d'un acte de reproduction (Gordon, 1979). Jusqu'au début du XX^{ème} siècle, il était essentiellement admis que le sexe était déterminé par des facteurs environnementaux, tels que la température ou la nutrition, favorisant ainsi le stockage de l'énergie et des nutriments pour l'embryon (se développant en femelle) ou au contraire leur utilisation (l'embryon se développant en mâle) (Geddes & Thompson, 1890). Le dogme du déterminisme environnemental du sexe demeura d'ailleurs la seule théorie majeure, jusqu'à la redécouverte des travaux de Mendel au début du XX^{ème} siècle, et notamment celle des chromosomes sexuels (McClung, 1902), en relation avec le développement d'un phénotype sexuel chez les insectes (Wilson, 1905).

La reproduction sexuée (appelée aussi sexualité) peut se définir à travers la notion de transferts de matériel génétique entre entités vivantes. Ces transferts sont caractérisés par le mouvement du matériel génétique, et dans le cas de la reproduction sexuée, ce transfert est vertical, puisque le matériel génétique est transféré des parents à la descendance. Cette notion de verticalité s'oppose à la notion de transfert horizontal (ou latéral) de matériel génétique entre deux entités, dans lequel un organisme donneur va transférer une part de son matériel génétique vers un organisme receveur, sans relation de descendance. Ces échanges horizontaux constituent d'ailleurs un mécanisme clé de l'évolution bactérienne, puisque les procaryotes peuvent échanger jusqu'à près de 15% de leur génome par des processus de conjugaison, de transformation ou de transduction (Garcia-Vallvé et al., 2000). Au delà des bactéries, de plus en plus de transferts horizontaux entre eucaryotes de différentes espèces sont décrits dans la littérature.

Le matériel génétique franchit alors non seulement la barrière physique que constitue l'enveloppe nucléaire, mais aussi la barrière reproductive entre espèces (Keeling & Palmer, 2008). Enfin, des phénomènes d'insertion de séquences procaryotes au sein de génomes eucaryotes sont eux aussi de plus en plus décrits, facilités notamment par la coévolution endosymbiotique ou parasitique (Cordaux & Gilbert, 2017).

1. Le paradoxe évolutif du sexe

Les différents cas d'échanges latéraux de matériel génétique évoqués précédemment sont générateurs d'innovation et de diversité génétique. La sexualité, elle, a aussi un rôle reproductif. Cependant, la reproduction sexuée a toujours été considérée comme extrêmement coûteuse pour les espèces la pratiquant. John Maynard Smith illustre ces coûts avec ce qu'il a appelé le double-coût du sexe : il faut tout d'abord deux individus pour générer une descendance, et de plus seul le sexe femelle donne naissance à la progéniture (à l'exception des espèces à hermaphrodisme simultané) (Maynard Smith, 1978). Il est aussi possible d'entrer en compétition avec des congénères du même sexe pour un accès à la reproduction. Avec l'accouplement vient le risque de contracter des agents infectieux en plus d'être potentiellement vulnérable à la prédation lors de l'accouplement. À ces risques s'ajoute la nécessité de passer d'un mécanisme de mitose à un mécanisme de méiose pour la production des gamètes. L'exemple de la levure *Saccharomyces cerevisiae*, chez qui la réplication mitotique s'effectue en moins de deux heures tandis que le processus de méiose peut durer plusieurs jours (Otto, 2009), illustre le coût de ce changement de mécanisme. Enfin, la reproduction sexuée ne permet à un individu de transmettre que la moitié de son patrimoine génétique à sa descendance, alors que la reproduction asexuée permet la transmission de la totalité du matériel génétique. Il s'agit du coût évolutif majeur de la sexualité selon Graham Bell (Bell, 1982). Pourtant, en plus d'être l'une des caractéristiques ancestrales rencontrée chez les eucaryotes (le plus ancien eucaryote à reproduction sexuée découvert est l'algue rouge *Bangiomorpha pubescens*, datée de 1.2 milliards d'années, Butterfield, 2000), près de 99.9% des eucaryotes se reproduisent, même occasionnellement, de façon sexuée

(Vrijenhoek, 1998). Cette prédominance se retrouve aussi au sein du règne végétal (même si la reproduction asexuée y est plus répandue) : moins de 1% des Angiospermes sont des organismes à reproduction exclusivement asexuée (Asker & Jerling, 1992; Whitton et al., 2008).

Le paradoxe du sexe apparaît alors évident : pourquoi le sexe s'est-il maintenu au cours de l'évolution, et comment expliquer un tel succès évolutif à l'échelle du Vivant ? Une des plus anciennes explications possibles remonte à la fin du XIX^{ème} siècle, attribuée à August Weismann, qui proposait que le sexe avait évolué car il permet de générer la variation nécessaire à la sélection (Weismann, 1889). La pression de sélection est susceptible de varier dans l'espace comme dans le temps, ainsi de nouvelles associations générées par la reproduction sexuée peuvent augmenter en fréquence dans les populations lorsque les anciennes ne sont plus favorables, améliorant la valeur sélective de la descendance dans son nouvel environnement. Une étude modélisant l'évolution d'une population haploïde à reproduction asexuée a conclu que si les individus ne présentant aucune mutation délétère étaient peu nombreux et ne parvenaient pas à se reproduire, alors ces individus disparaîtraient et la nouvelle génération aurait ainsi au moins autant de mutations que leurs parents (Müller, 1964). Ce phénomène est surnommé le "cliquet de Müller". Par analogie avec le fonctionnement d'une roue à cliquet, il peut définir tout type de mécanisme maintenant un système en l'état, sans possibilité de retour à un état précédent. Une telle situation provoquerait donc inévitablement l'accumulation de mutations délétères. Or la reproduction sexuée, de par la recombinaison méiotique, permet l'élimination de ces mutations à chaque génération, suivant le principe de la méiose réparatrice (Kondrashov, 1988). Au-delà de ces modèles, les observations tirées de la Nature tendent à confirmer l'intérêt évolutif de la reproduction sexuée. En effet, les populations à reproduction asexuée s'avèrent incapables de s'adapter aussi rapidement et efficacement aux variations du milieu que les populations à reproduction sexuée (Malmberg, 1977; Colegrave et al., 2002; Goddard et al., 2005; Cooper, 2007). Cette capacité d'adaptation conférée par la reproduction sexuée peut même se réaliser à une échelle temporelle très courte, comme dans le cas de *Daphnia magna*, qui se reproduit par parthénogenèse dans un environnement stable et favorable, et de façon sexuée dans un environnement stressant (Roulin et al.,

2013). Une explication supplémentaire au succès évolutif de la reproduction sexuée propose que, bien que la reproduction asexuée soit favorisée à court terme, celle-ci souffre d'importants coûts évolutif sur le long terme et ce à l'échelle de l'espèce. Ainsi à long terme, la reproduction sexuée serait maintenue par la sélection de clade (Maynard Smith, 1978; Nunney, 1989). Cette théorie prédit que, en l'absence de recombinaison due à la reproduction sexuée, l'adaptation de l'espèce peut être entravée (Barton, 1994; Peck, 1994) et les mutations délétères peuvent s'accumuler de part le phénomène du cliquet de Müller (Müller, 1964). Selon cette théorie, la reproduction sexuée n'a donc pas besoin de compenser immédiatement son "double coût" à l'échelle de l'individu, et sera maintenue à long terme par la sélection de clade. Ainsi au-delà de l'aspect reproductif, c'est bien la notion de variabilité, et par extension de potentiel adaptatif, qui semble permettre non seulement le maintien mais aussi le succès évolutif de la reproduction sexuée, comme l'avait déjà suggéré August Weismann (Weismann, 1889).

2. Les mécanismes moléculaires impliqués dans la reproduction sexuée

La variabilité permise par la reproduction sexuée résulte de deux processus conservés dans la majorité des taxa : la méiose et la fécondation. La méiose est un mécanisme de double division cellulaire (Figure 1), prenant place au niveau des cellules diploïdes de la lignée germinale et résultant en la production des cellules reproductrices haploïdes : les gamètes. La première division méiotique de la gamétogenèse fait suite à la réplication de l'ADN qui permet le passage de chromosomes diploïdes monochromatidiens à bichromatidiens (Figure 1). Ensuite, les chromosomes homologues s'apparient et peuvent échanger des fragments de chromatide par un phénomène de recombinaison, constituant un brassage intra-chromosomique (Figure 1). Classiquement appelés événements de "crossing-over" ou d'enjambement (Janssens, 1909), il s'agit du premier niveau de brassage génétique. Lors de l'anaphase qui en découle, les chromosomes se séparent ensuite aléatoirement : il s'agit de la ségrégation indépendante des chromosomes homologues (Figure 1).

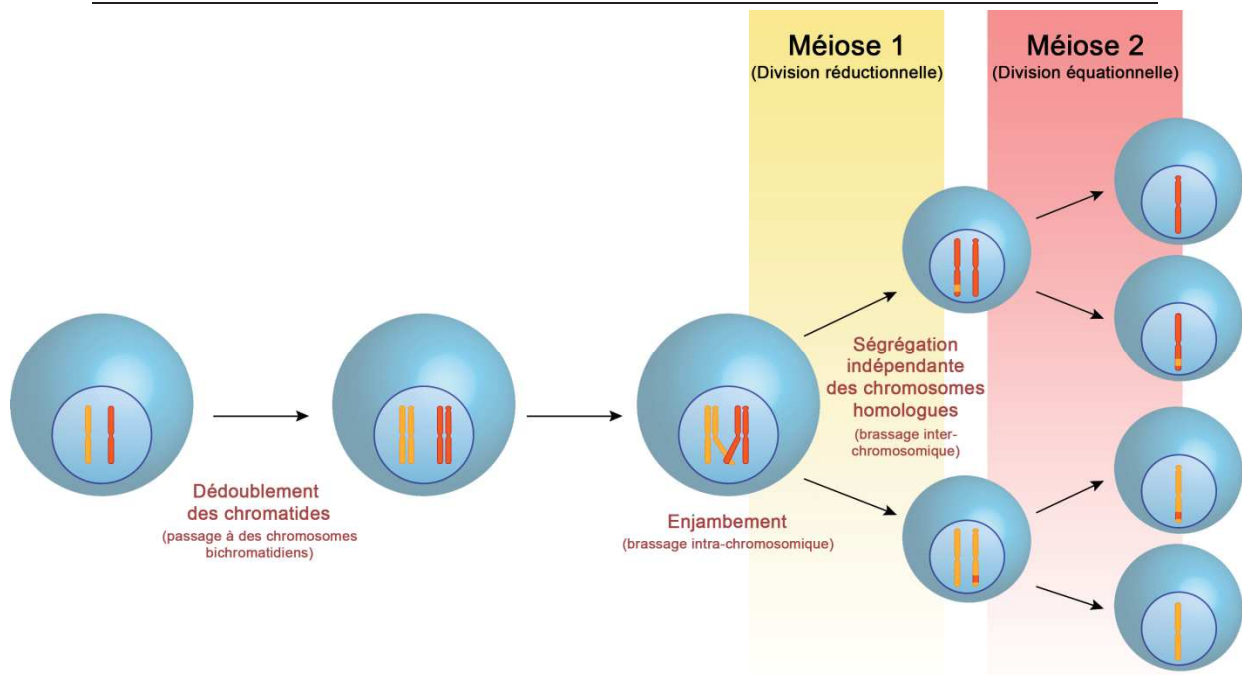


Figure 1 : Représentation de la méiose et des innovations génétiques qu'elle induit. La cellule originelle (à gauche) contient une paire de chromosomes homologues à un chromatide. Suite à ce dédoublement, un brassage intra-chromosomique a lieu lors de l'appariement des chromosomes homologues (phénomène d'enjambement). L'anaphase de la première méiose, qui consiste à séparer indépendamment les chromosomes homologues, constitue une phase de brassage inter-chromosomique. La méiose 2 (ou division équationnelle) correspond à une mitose classique. Elle correspond au passage d'une cellule à n chromosomes bichromatidiens à deux cellules à n chromosomes monochromatidiens.

Cette étape correspond à un second niveau de brassage génétique (brassage inter-chromosomique). A l'issue de la première division méiotique, les cellules filles disposent d'un matériel génétique haploïde original, grâce aux brassages intra et inter-chromosomiques. Chaque chromosome disposant de deux chromatides, la seconde division méiotique consiste en une division cellulaire sans étape de synthèse d'ADN. Les cellules ainsi produites, après maturation le plus souvent au sein des gonades, vont devenir les gamètes. Selon l'espèce considérée, les gamètes produits par le sexe mâle ou le sexe femelle peuvent être morphologiquement différents (anisogamie) ou non (isogamie).

L'anisogamie est la situation la plus fréquente rencontrée dans la Nature (Bell, 1978), l'isogamie s'avérant quant à elle bien plus rare (Matsuda & Abrams, 1999). Dans le cas d'espèces isogames, tous les gamètes d'une espèce donnée ne sont pas nécessairement compatibles entre eux. Les individus disposent alors d'un "type de compatibilité" (ou *mating type*) et ne peuvent se reproduire qu'avec des individus d'un type différent (Billiard et al., 2011). Il est suspecté que le passage de l'isogamie vers l'anisogamie serait dû à l'évolution de gènes contrôlant la taille

des gamètes au niveau de régions spécifiques du génome, appelées *mating type regions* (Ferris et al., 2010). D'ailleurs, pour les espèces anisogames, c'est justement l'anisogamie qui définit le sexe (mâle ou femelle) de l'individu. Ainsi, les gamètes mâles (spermatozoïdes) sont généralement mobiles, bien plus petits et nombreux que les gamètes femelles (ovules). La plus grande taille des gamètes femelles s'explique par un contenu cytoplasmique riche, à la base du métabolisme de l'œuf, et donc essentiel pour la survie du zygote. Ces différences morphologiques bénéfiques à l'embryon constituent l'explication de l'évolution d'un état ancestral d'isogamie vers l'anisogamie (Parker & Lehtonen, 2014). En effet, la formation d'un gamète volumineux est métaboliquement coûteuse, mais les réserves stockées constituent un avantage pour le développement de l'œuf fécondé, tandis que la production massive de gamètes de petite taille est moins coûteuse et maximise les chances de rencontre des deux types de gamètes. Des gamètes de même taille ne présentant pas d'avantage, l'évolution vers deux formes divergentes est donc favorisée, constituant ainsi un phénomène de sélection non-stabilisante (Parker et al., 1972). La rencontre aléatoire des gamètes mâles et femelles, et donc des deux lots parentaux de matériel génétique, est appelée fécondation. Il s'agit du troisième et dernier niveau de brassage génétique de la reproduction sexuée, qui formera ainsi un individu génétiquement original. La fusion des *pronuclei* des gamètes mâle et femelle dans l'ovocyte pour former le zygote diploïde est appelé caryogamie ou amphimixie. Deux types de fécondations peuvent être définies selon la provenance des gamètes : l'auto-fécondation si les deux gamètes proviennent d'un même individu, et l'allo-fécondation si les gamètes proviennent de deux individus différents.

3. La question de l'identité sexuelle

La reproduction sexuée présente deux grandes modalités : l'hermaphrodisme et le gonochorisme. Un même individu capable de produire les deux types de gamètes est appelé hermaphrodite. Les individus de telles espèces disposent de gonades mixtes appelées ovotestis (comme chez le gastéropode *Helix aspersa*; Antkowiak & Chase, 2003) ou des gonades mâles (testicule) et femelles (ovaire) comme c'est le cas chez de nombreuses espèces de vers de terre (Edwards

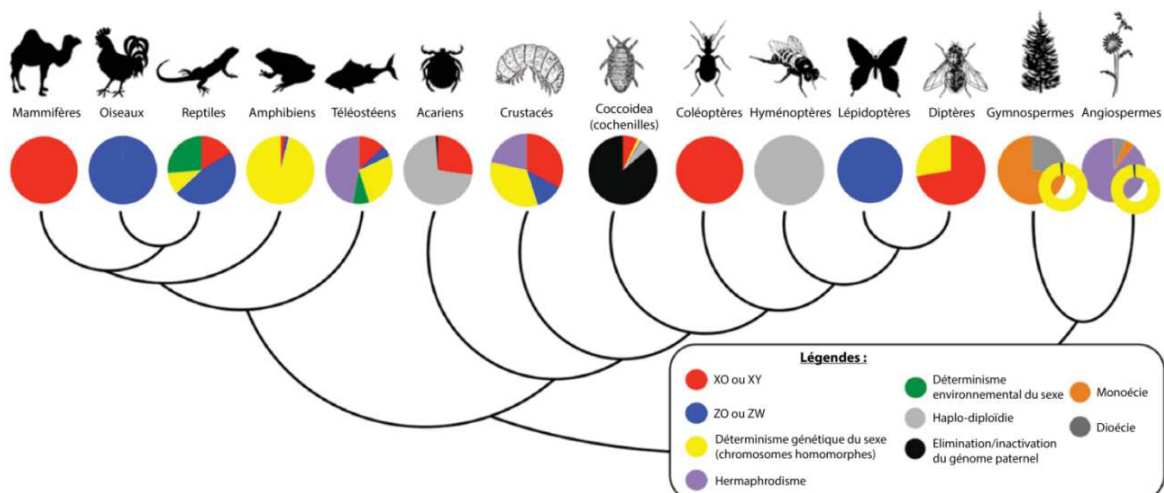
& Bohlen, 1996). Il s'agit d'une stratégie de reproduction particulièrement répandue puisqu'elle concerne près d'un tiers des métazoaires (hors insectes, Jarne & Auld, 2006) et plus de 90% des genres de plantes (Renner & Ricklefs, 1995). Selon les espèces, les gonades mâles et femelles peuvent être portées soit simultanément par le même individu (*e.g.* chez le poisson de récif *Hypoplectrus nigricans*, voir la revue de Wootton & Smith, 2014), soit successivement. Généralement, l'hermaphrodisme simultané se retrouve chez des espèces pour lesquelles les possibilités de rencontre de partenaires sexuels sont faibles, notamment dans le cas de faibles densités de populations (Ghiselin, 1969; Eppley & Jesson, 2008). Cependant, ce point est toujours discuté, puisque si la majorité des cirripèdes (Crustacea) sont hermaphrodites, certaines espèces ont évolué vers une séparation des sexes (gonochorisme, qui correspond au fait qu'un individu n'a qu'un seul sexe) (Yusa et al., 2012). En effet, les modèles prédisent que l'absence d'autofécondation (Barnes, 1989) associée à une faible probabilité de rencontrer des partenaires conduit à une séparation des sexes (Yusa et al., 2012). Dans le cas d'espèces à hermaphrodisme successif, une espèce peut être soit protandre (les individus sont tout d'abord mâles, puis femelles) tel que dans le cas du poisson-clown à trois bandes *Amphirion ocellaris* (Fautin & Allen, 1992), soit protogyne (les individus sont tout d'abord femelles, puis mâles), ce qui est retrouvé chez près de 75% des espèces de poissons à hermaphrodisme successif (Pandian, 2012).

Le gonochorisme concerne quant à lui la majorité du règne animal. Les espèces concernées ont ainsi deux sexes séparés : les mâles et les femelles. Par similarité avec le passage de l'isogamie à l'anisogamie, l'hermaphrodisme aurait évolué vers le gonochorisme via un compromis évolutif : un antagonisme entre le sexe mâle et le sexe femelle, caractérisé par des évolutions augmentant la valeur sélective d'un sexe aux dépens de la valeur sélective de l'autre sexe. La séparation des sexes se réaliserait par l'évolution d'allèles qui augmenteraient ou diminueraient la fertilité d'un des sexes (Charnov et al., 1976). Une seconde hypothèse majeure (compatible avec la précédente) suggère que la séparation des sexes aurait évolué comme moyen d'éviter l'auto-fécondation. Celle-ci peut en effet provoquer une dépression de consanguinité (par un phénomène de goulet d'étranglement, diminuant la variabilité allélique) responsable d'une faible valeur

sélective de la descendance par accumulation d'allèles délétères (Charlesworth & Charlesworth, 1978).

4. La diversité des facteurs déclencheurs du déterminisme du sexe : de l'environnement à la génétique

La mise en place de l'anisogamie et du gonochorisme a pour conséquence majeure de faire apparaître un rôle distinct pour chaque sexe. Leur apparition s'est accompagnée de celle de systèmes de déterminisme du sexe, afin de déclencher la différenciation en un individu sexué. Les mécanismes de déterminisme du sexe les plus connus sont ceux des chromosomes sexuels de *Drosophila melanogaster* ou encore des mammifères, où les mâles ont des chromosomes XY et les femelles XX (Bachtrog et al., 2014). Ce mécanisme de déterminisme du sexe, qui implique des chromosomes sexuels étant considérés comme des éléments stables à l'échelle génomique, est parfois considéré comme normal et généralisable à l'échelle du Vivant. Cependant, ces *a priori* ne résistent pas à un examen de la répartition des mécanismes de déterminisme du sexe au sein de multiples taxa, plantes comme animaux (Figure 2), puisque le plus souvent, une grande diversité de systèmes de déterminisme du sexe est observée (Figure 2). Cette diversité peut être décomposée en deux catégories distinctes selon la nature du facteur déclencheur du devenir sexuel des individus : le déterminisme environnemental du sexe, et le déterminisme génétique du sexe.



Le déterminisme environnemental du sexe (ESD, pour *Environmental Sex Determination*) est initié par des signaux environnementaux. Ces signaux déclenchent une cascade de mécanismes génétiques et régulent les gènes déterminants pour le sexe male ou le sexe femelle (Crews & Bull, 2009). Il peut dépendre de deux types de facteurs : (i) les caractéristiques physico-chimiques du milieu et (ii) les caractéristiques structurelles d'une population (caractéristiques dites "sociales"). Le premier cas constitue l'exemple le plus connu d'ESD, et se distingue par l'existence d'une période sensible au cours du développement embryonnaire, durant laquelle le devenir sexuel de l'individu sera influencé par un facteur physico-chimique du milieu. Par exemple, chez de nombreuses espèces d'ectothermes, le sexe d'un individu est déterminé par la température au cours du développement embryonnaire (Crain & Guillette, 1998; Valenzuela & Lance, 2004) (Figure 3).

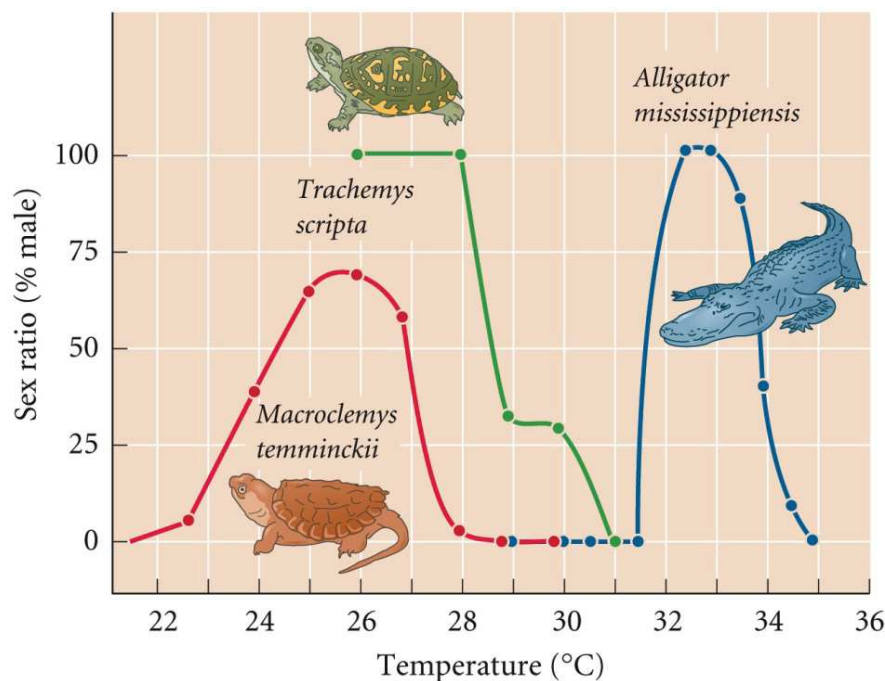


Figure 3 : Exemples d'espèces au déterminisme environnemental du sexe. Un déterminisme du sexe température-dépendant est constaté chez l'alligator américain (*Alligator mississippiensis*), la tortue de Floride (*Trachemys scripta elegans*) et la tortue alligator (*Macrolemys temminckii*) (adapté de Crain & Guillette, 1998).

La température va moduler la transcription de l'aromatase, une enzyme responsable de la conversion d'androgènes en œstrogènes chez les espèces à ESD température-dépendant (Wibbels & Crews, 1994; Navarro-Martín et al., 2011). La proportion relative de ces hormones influence le développement de gonades

indifférenciées en testicules ou en ovaires. Le déterminisme température-dépendant est particulièrement bien décrit chez les reptiles (Figure 3) tels que les tortues (Pieau, 1971), les alligators (Ferguson & Joanen, 1982), ou encore les lézards (Charnier, 1965), ainsi que chez les poissons (Ospina-Álvarez & Piferrer, 2008). La luminosité/photopériode représente un autre mécanisme de déterminisme du sexe, comme c'est le cas par exemple chez certaines espèces d'Orchidées appartenant aux genres *Catasetum* ou encore *Cycnoches*. Dans ces deux exemples les individus se différencient en mâles en présence d'un ensoleillement faible, et à l'inverse en femelles si l'exposition au soleil est plus intense (Charnov & Bull, 1977). Ce mécanisme existe aussi chez certains invertébrés tels que le crustacé *Gammarus duebeni* (Amphipoda), où les variations de photopériode résultent en des pics de production de mâles ou de femelles au cours de l'année (Bulnheim, 1978). Enfin, une étude récente menée sur la lamproie (*Petromyzon marinus*) suggère que le sexe de cette espèce serait directement influencé par la vitesse de croissance larvaire des individus (Johnson et al., 2017), cette croissance étant soumise directement à la qualité du milieu.

Dans le cas du déterminisme social du sexe, les facteurs essentiels sont la densité de population ou encore le sex-ratio des individus déjà présents (Haag & Doty, 2005). Chez le ver marin *Bonellia viridis* (Echiura, Annelida) par exemple, les larves se développent en femelles si la niche qu'elles occupent est vacante, ou en mâles si la larve indifférenciée se dépose sur une femelle déjà présente (Leutert, 1974).

Même si l'ESD est phylogénétiquement très répandu au sein du règne animal (Korpelainen, 1990), la catégorie de mécanismes déterminant le sexe la plus répandue reste le déterminisme génétique du sexe (GSD, pour *Genetic Sex Determination*). Dans ce cas, c'est un ou des éléments génétiques qui déterminent si l'individu se développe en femelle ou en mâle. Tout comme dans le cas de l'ESD, il existe de multiples modalités de GSD (Figure 2, Bachtrog et al., 2014).

Le plus décrit et le plus connu des GSD est le déterminisme chromosomique du sexe. Chez de nombreux métazoaire, les deux sexes ont une composition chromosomique différente. Le sexe est déterminé par un *locus* qui est hétérozygote chez un sexe et homozygote chez l'autre (Bull, 1983). Il existe deux types principaux d'hétérogamétie : l'hétérogamétie mâle dans le cas où les mâles

portent deux chromosomes différents (σ^{XY} - ϕ^{XX}), caractérisant par exemple *D. melanogaster* et l'ensemble des mammifères, comme évoqué plus haut (Figure 2), et l'hétérogamétie femelle dans le cas où les femelles portent deux chromosomes différents (σ^{ZZ} - ϕ^{ZW}) (Bull, 1983), comme chez l'ensemble des oiseaux ou encore des lépidoptères (Figure 2). D'autres combinaisons existent, allant de l'hétérogamétie femelle de type 00/W0 ou encore ZZ/Z0 observée chez certains amphibiens (Green, 1988) ou certains poissons (Devlin & Nagahama, 2002) à l'hétérogamétie mâle de type XX/X0 (Devlin & Nagahama, 2002). De plus, des systèmes complexes XY et ZW sont aussi possibles, dans le cas de fusions/scissions/translocations entre chromosomes sexuels et autosomes, pouvant mener à des systèmes de type X_1X_2Y , XY_1Y_2 ou encore Z_1Z_2W (voir la revue de Bachtrog et al., 2014). Les deux systèmes d'hétérogamétie peuvent d'ailleurs coexister au sein de la même espèce. C'est le cas par exemple chez la grenouille *Rana rugosa*, chez qui les chromosomes sexuels impliqués dans les deux systèmes se sont différenciés indépendamment au sein de différentes populations (Nishioka et al., 1993). Enfin, un dernier type d'hétérogamétie existe chez certaines espèces ayant une phase sexuée haploïde, qualifié de système U/V, où les femelles sont porteuses du chromosome U et les mâles porteurs du chromosome V (Ahmed et al., 2014). Dans l'exemple de l'algue brune *Ectocarpus sp.*, le sporophyte diploïde issu de la fécondation est tout d'abord de génotype UV, et le sexe ne sera déterminé que lors de la méiose, qui formera des individus haploïdes femelles U ou mâles V.

D'autres exemples de GSD, moins répandus, existent aussi en plus du déterminisme chromosomique du sexe. C'est le cas par exemple de l'haplo-diploïdie, notamment rencontrée chez l'ensemble des hyménoptères (Grimaldi & Engel, 2005), qui implique la totalité du génome, puisque les mâles se distinguent par un génome haploïde (développement à partir d'une cellule-œuf non fécondée), tandis que les femelles sont diploïdes (développement à partir d'une cellule-œuf fécondée). Il existe aussi des phénomènes d'inactivation (sous forme d'hétérochromatine), voire d'élimination du génome paternel post-fécondation, qui sont particulièrement répandus au sein des cochenilles (Coccoidea, Hemiptera) (Figure 2 ; Nur, 1980). Le système de déterminisme polygénique du sexe consiste quant à lui en de multiples *loci*/allèles potentiellement localisés sur différents

chromosomes (Moore & Roberts, 2013) étant impliqués dans le déterminisme du sexe, comme par exemple chez le poisson zèbre (Liew et al., 2012).

5. L'évolution des chromosomes sexuels et ses implications

S'il existe un dimorphisme phénotypique entre les individus du sexe mâle et ceux du sexe femelle, ce dimorphisme peut être retrouvé également à l'échelle chromosomique dans le cas des espèces à système de déterminisme chromosomique du sexe. Les chromosomes sexuels, appelés aussi gonosomes, sont susceptibles de présenter des différences structurales majeures entre le chromosome hétérogamétique (Y ou W) et le chromosome homogamétique (X ou Z). Le modèle consensuel d'évolution des hétérochromosomes (Charlesworth et al., 2005) prédit que les proto-chromosomes sexuels (dans l'exemple de l'hétérogamétie mâle : le proto-X et le proto-Y) dérivent d'une paire d'autosomes disposant du même set de gènes (I, Figure 4), sur lesquels seraient apparus des gènes déterminants du sexe responsables d'une stérilité mâles sur le proto-X ou d'une stérilité femelles pour le proto-Y (II, Figure 4). Afin de maintenir cette spécialisation, la zone du génome proche de ce *locus* déterminant le sexe cesse alors de recombiner avec la région située sur le chromosome homologue suite à des phénomènes d'hétérochromatisation et d'inversion chromosomique (III, Figure 4). La spécialisation pourrait d'ailleurs être accentuée par l'apparition de gènes bénéfiques au sexe mâle sur le chromosome Y (ou de gènes bénéfiques aux femelles sur le chromosome X) proches du *locus* déterminant le sexe hétérogamétique. Une des conséquences majeures de l'arrêt de la recombinaison est l'accumulation d'ADN répété sur l'hétérochromosome ainsi que la perte de fonction de certains gènes présents (pseudogénisation) (IV, Figure 4). Enfin, l'accumulation d'ADN répété associée, ainsi que la présence de pseudogènes non-fonctionnels, peut conduire à la délétion de larges segments chromosomiques (car non sélectionnés pour être conservés), et donc à la dégénérescence physique de l'hétérochromosome (V, Figure 4).

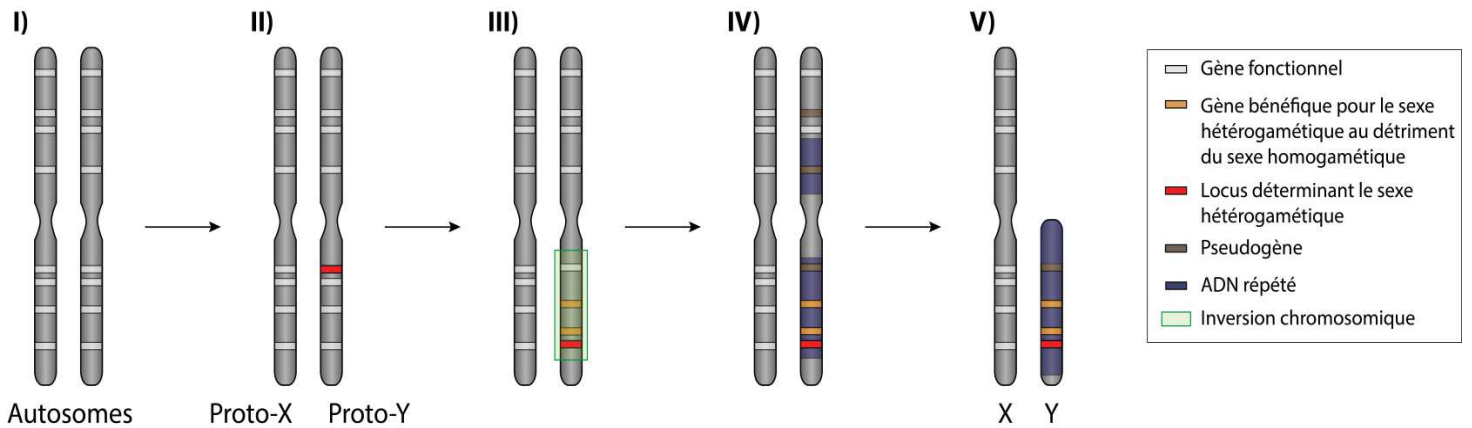


Figure 4 : Modèle de différenciation des chromosomes sexuels. I) Les chromosomes sexuels sont formés à partir d'une paire d'autosomes disposant du même set de gènes. II) Acquisition d'un *locus* déterminant le sexe sur un proto-chromosome sexuel (ici avec l'exemple d'une hétérogamétie mâle, et donc une différenciation entre proto-X et proto-Y). III) Accumulation de mutations sexuellement antagonistes avec possibilité d'inversions chromosomiques, et extension progressive de la zone non-recombinante. IV) L'arrêt de la recombinaison contribue à l'accumulation de séquences répétées et à une pseudogénéisation de certains gènes liés au Y. V) Délétion de larges segments d'ADN non-fonctionnels, réduisant la taille physique de l'hétérochromosome.

De ce scénario résulte un hétérochromosome Y (ou W) morphologiquement et génétiquement différent du X (ou du Z). La partie recombinante demeurant éventuellement entre les deux chromosomes sexuels est appelée région pseudo-autosomique. La région spécifique à l'hétérochromosome (Y-spécifique ou W-spécifique) est le plus souvent riche en séquences répétées, pauvre en contenu génique, et hétérochromatique (Bachtrog, 2013). Selon l'espèce considérée, différents degrés de dégénérescence de l'hétérochromosome sont observés, allant de l'homomorphie (les deux hétérochromosomes sont sensiblement de même taille) comme chez le poisson globe *Takifugu rubripes* (Kikuchi et al., 2007), à la dégénérescence totale de l'hétérochromosome (suivie parfois de sa disparition) comme chez *D. melanogaster* (Carvalho et al., 2003) (Figure 5). Une conséquence de ce scénario évolutif est que, jusqu'à encore assez récemment, les chromosomes homomorphes étaient considérés comme évolutivement "jeunes" et récents, leur dégénérescence vers un état d'hétéromorphie semblant inévitable (Steinemann & Steinemann, 2005). Grâce à des études effectuées sur des espèces non modèles, ce paradigme ne semble plus être systématiquement vrai, puisque l'âge de certains gonosomes homomorphes a été estimé à plus de 120 millions d'années (exemple des serpents de la famille des Boidae, Vicoso et al., 2013a ; ou encore de l'émeu *Dromaius novaehollandiae*,

Vicoso et al., 2013b) et à l'inverse, certains chromosomes Y rencontrés chez les drosophiles, fortement hétéromorphes, ne sont âgés que de quelques dizaines de millions d'années, pouvant même redevenir des autosomes (Vicoso & Bachtrog, 2013).

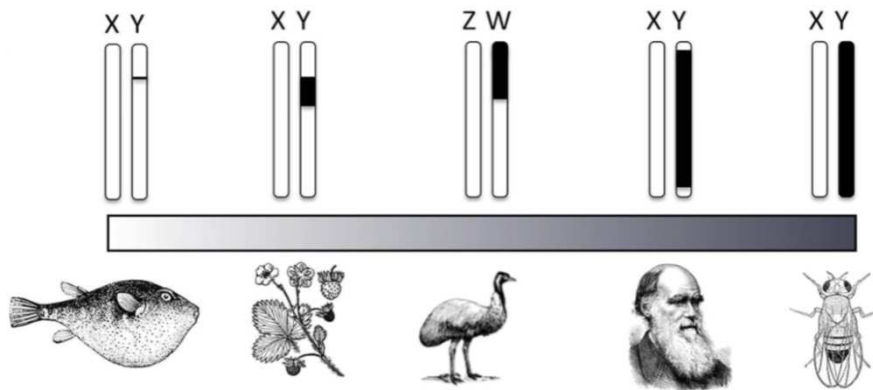


Figure 5 : Différents niveaux de dégénérescence de l'hétérochromosome (Y ou W). Selon l'espèce considérée, des gonosomes homomorphes sont observés chez le poisson globe *Takifugu rubripes*, une hétéromorphie intermédiaire est observée chez le fraisier de Virginie *Fragaria virginiana* ou encore chez l'émou *Dromaius novaehollandiae*, une très forte hétéromorphie est constatée chez l'être humain, jusqu'à atteindre une hétéromorphie totale dans le cas de *Drosophila melanogaster* (adapté de Bachtrog et al., 2014)

Une conséquence directe de l'hétéromorphie apparaît à la méiose. En effet, les chromosomes Y et X sont de compositions géniques différentes et ne peuvent s'apparier que grâce à leurs *loci* communs, situés dans les régions pseudo-autosomiques (PAR : *Pseudo-autosomal regions*). Leur reconnaissance et leur appariement sont facilités par une condensation chromatinienne des gonosomes du sexe hétérogamétique, et donc par un arrêt de leur transcription. Ce phénomène est qualifié d'inactivation méiotique des chromosomes sexuels (MSCI : *Meiotic Sex Chromosome Inactivation*) (Turner, 2007). A l'inverse dans le cas femelles homogamétiques XX, le chromosome X inactivé est quant à lui réactivé afin de faciliter l'appariement avec son homologue déjà actif (Jablonka & Lamb, 1988). Une autre conséquence directe de l'hétéromorphie des chromosomes sexuels réside dans le déséquilibre de dosage génique, à la fois entre les sexes mais aussi entre les chromosomes sexuels et les autosomes (Charlesworth, 1996; Mank, 2009). Concernant les gènes situés sur les chromosomes X et Z, ils ont potentiellement des fonctions généralistes importantes dans le développement de l'individu. Un déséquilibre dans la dose de ces gènes présents en une seule copie

chez les individus hétérogamétiques est susceptible de diminuer la valeur sélective de ces individus, de la même façon que dans le cas de l'aneuploïdie (Veitia, 2005). Pour compenser cette charge génique différentielle, deux axes stratégiques ont évolué indépendamment : (i) surexprimer les gènes situés sur l'unique X (mâle) ou l'unique Z (femelle) comme dans le cas de *D. melanogaster* (Figure 6, Lucchesi & Kuroda, 2015) ou (ii) inactiver l'expression des gènes situés sur l'un des deux chromosomes homologues du sexe homogamétique (X chez la femelle ou Z chez le mâle) tout en surexprimant les gènes portés par le X ou le Z au sein du sexe hétérogamétique, comme c'est le cas chez les mammifères (Figure 6, Brockdorff & Turner, 2015). Un mécanisme intermédiaire existe aussi chez *Caenorhabditis elegans*, où une inactivation partielle de l'expression des gènes situés sur les deux X des femelles est observée (Figure 6, Meyer, 2000). Lorsque la compensation de dosage est complète, le ratio moyen entre l'expression des gènes gonosomaux et l'expression des gènes autosomaux est égal à 1. En revanche, une compensation incomplète (le ratio d'expression moyen entre les gènes portés par les chromosomes sexuels et ceux portés par les autosomes est différent de 1 a été observée chez les oiseaux (Ellegren et al., 2007; Itoh et al., 2007), les serpents (Vicoso et al., 2013a) ou encore chez les lépidoptères (Zha et al., 2009). Plus généralement, les espèces à hétérogamétie mâle semblent posséder des mécanismes complets de compensation de dosage, tandis que les espèces à hétérogamétie femelle disposent apparemment d'un mécanisme local et incomplet (Mank, 2009; Vicoso & Bachtrog, 2009).

	Male	Female
<i>Drosophila melanogaster</i>	 Y AA	XX AA
<i>Homo sapiens</i>	 Y AA	  AA
<i>Caenorhabditis elegans</i>	 AA	  AA

Figure 6 : Principe de la compensation de dosage génique. La présence d'un unique chromosome X chez *D. melanogaster*, l'Homme et *C. elegans* est compensée par une surexpression de leurs gènes. Dans le cas des Mammifères (ici l'Homme) et de *C. elegans*, soit les femelles/hermaphrodites inactivent un des deux chromosomes X (Mammifères) soit elles inactivent partiellement les deux chromosomes X (*C. elegans*). Les lettres de grande taille encadrées en vert symbolisent une surexpression, tandis que les encadrés oranges indiquent une inactivation (Straub & Becker 2007).

6. Les effecteurs moléculaires impliqués dans le déterminisme du sexe

Le plus souvent, le GSD s'organise selon une cascade de gènes : un signal primaire déclenche l'expression d'une série de gènes déterminant le sexe menant à la différenciation sexuelle de l'individu. En général, on observe que les gènes situés en aval de cette cascade sont davantage conservés que les gènes situés plus en amont, ce qui s'explique par une évolution du type *bottom-to-top* (Wilkins, 1995). Cette cascade du déterminisme du sexe repose le plus souvent sur un gène déterminant clé dont l'effet peut suivre deux types de modalités : présence/absence (exemple du gène SRY - *Sex-determining Region of Y chromosome* - chez les mammifères, Koopman et al., 1991) ou dose-dépendance (exemple du gène *DMRT1* - *Doublesex and Mab-3-Related Transcription factor 1* - chez le poulet *Gallus gallus*, Smith et al., 2009). Si une grande variété de mécanismes de déterminisme du sexe existe dans le règne du Vivant (pour rappel, Figure 2), il faut noter que les acteurs moléculaires de ces mécanismes sont souvent conservés (Bachtrog et al., 2014).

La famille des gènes à domaine DM, codant pour des facteurs de transcription, semble jouer un rôle central dans le déterminisme du sexe tant chez les vertébrés que chez les invertébrés (Raymond et al., 1998; Matson & Zarkower, 2012). Le domaine DM a été défini suite à la comparaison entre le gène *Doublesex* (*Dsx*) identifié chez *D. melanogaster* et le gène *Male Abnormal-3* (*MAB-3*) de *C. elegans* (Raymond et al., 1998). Ce domaine protéique forme une structure en double doigts-de-zinc (motif de liaison à l'ADN) entrecroisés composée de six cystéines et de deux histidines (Erdman & Burtis, 1993; Zhu et al., 2000). La famille des gènes à domaine DM semble être ubiquiste à l'échelle du règne animal, mais n'est pas présente au sein des autres règnes (Figure 7). Cette large distribution y compris au sein des non-bilatériens sous-entend une origine pré-métazoaire, suivie d'une diversification indépendante au cours de l'évolution (Wexler et al., 2014). La famille des gènes à domaine DM (gènes *DMRT*, pour *Doublesex and Mab-3 Related Transcripts*) comprend neuf sous-familles, en plus de *Doublesex* pour les insectes et de *MAB-3* chez *C. elegans* (Volff et al., 2003; Wexler et al., 2014).

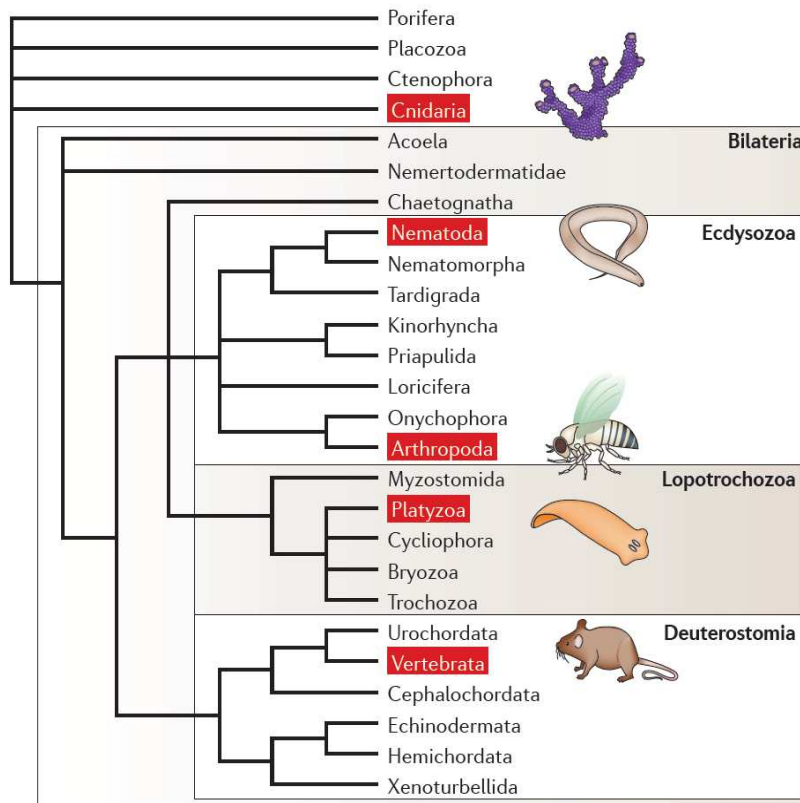


Figure 7 : Distribution de la présence de gènes à domaine DM impliqués dans le développement sexuel des Métazoaires. L'implication des gènes à domaine DM a été déterminée soit par des analyses fonctionnelles, soit par l'observation d'une expression sexe-biaisée (phyla surlignés en rouge) (Matson & Zarkower, 2012).

Pour la majorité des espèces chez qui les gènes à domaine DM ont été identifiés, le rôle de ces gènes dans le déterminisme ou la différenciation sexuelle demeure inconnu. Il existe néanmoins des exemples pour lesquels l'implication de ces gènes dans les mécanismes de déterminisme ou de différenciation du sexe est clairement établie. Par exemple, c'est le cas du gène *DMRT1* qui est présent chez l'ensemble des vertébrés, et qui peut être situé au niveau de régions autosomiques (cas des espèces à hétérogamétie mâle) ou au niveau des chromosomes sexuels (cas des espèces à hétérogamétie femelle) (Veyrunes et al., 2008). Ainsi, chez les mammifères placentaires, il est impliqué dans le maintien de la fonction mâle (Zarkower, 2013) tandis qu'il intervient directement dans le déterminisme sexuel du poulet (Smith et al., 2009). Ce gène est aussi responsable à la fois de la différenciation testiculaire et du maintien de la fonction mâle chez la grenouille *Rana rugosa* (Shibata et al., 2002).

Un autre exemple de gène à domaine DM impliqué dans le déterminisme ou la différenciation sexuel est le gène *Doublesex*. Ce gène a été trouvé chez tous les insectes chez lesquels il a été recherché, couvrant ainsi les ordres des diptères, des hyménoptères, des coléoptères ainsi que des lépidoptères (Geuverink & Beukeboom, 2014; Picard et al., 2015). Classiquement, ce gène subit un épissage alternatif, menant à la production de variants protéiques sexe-spécifiques (voir la revue de Geuverink & Beukeboom, 2014). Parmi les crustacés, deux copies du gène *Doublesex* ont été trouvées chez le branchiopode *D. magna* (*DapmaDsx1* et *DapmaDsx2*), espèce présentant un ESD. Dans cet exemple, ces homologues ne sont pas épissés alternativement, mais l'une des copies constitue l'effecteur majeur de la différenciation des individus en mâles (Kato et al., 2011). En effet, non seulement le gène *DapmaDsx1* est surexprimé au sein des embryons mâles, mais son extinction provoque le développement des embryons mâles en femelles phénotypiques (Kato et al., 2011). La famille des gènes à domaine DM présente la particularité de contenir des acteurs moléculaires impliqués dans le déterminisme ou la différenciation du sexe. Ces gènes sont ainsi retrouvés chez un vaste panel d'espèces animales, indépendamment de leurs stratégies reproductives et de leurs mécanismes de déterminisme du sexe (Picard et al., 2015).

D'autres acteurs moléculaires du déterminisme du sexe sont aussi fortement conservés à l'échelle du règne animal, mais disposent de fonctions plus variables, et d'une distribution phylogénétique plus éparse (Geuverink & Beukeboom, 2014). C'est le cas du gène *Transformer* (*Tra*), identifié pour la première fois chez *D. melanogaster*, et qui constitue un facteur d'épissage alternatif intervenant dans la cascade de différenciation sexuelle, en amont du gène *Dsx* (Walthour & Schaeffer, 1994). L'association de *Tra* avec le gène *Transformer-2* (dont l'expression est constitutive) va permettre le contrôle de l'épissage alternatif de l'ARNm produit par le gène *Dsx* (Hedley & Maniatis, 1991; Ryner & Baker, 1991; Hertel et al., 1996). En revanche chez d'autres diptères, le gène *Transformer* peut directement avoir un rôle de déterminant sexuel, comme c'est le cas chez *Ceratitis capitata* (Pane et al., 2002) ou encore chez *Musca domestica* (Hediger et al., 2010). Au-delà des hexapodes, ce gène a aussi été identifié chez *D. magna* sans qu'un rôle dans le développement sexuel n'ait pu lui être attribué (Kato et al., 2010). Néanmoins, des patrons d'expression

sexe-spécifiques ont été relevés pour les gènes *Transformer-2* des crustacés *Fenneropenaeus chinensis* et *Eriocheir sinensis*, suggérant une possible implication dans le déterminisme du sexe (Li et al., 2012 et Luo et al., 2017, respectivement). Il faut noter que la découverte des gènes *Transformer/Transformer-2* à la fois chez les hexapodes et les crustacés suggère que son origine est antérieure à la divergence entre ces deux clades (Geuverink & Beukeboom, 2014), suggérant un âge d'au moins 400 millions d'années (Ronshaugen et al., 2002; Glenner et al., 2006). D'autre part, la forte conservation de ces gènes suggère que d'autres pressions de sélection ont agi au cours de l'évolution pour maintenir leur présence, notamment car ils pourraient être impliqués dans d'autres fonctions essentielles (Geuverink & Beukeboom, 2014).

7. Transitions entre systèmes de déterminisme du sexe

Qu'ils appartiennent à la catégorie des ESD ou des GSD, les mécanismes de déterminisme du sexe sont répartis de façon hétérogène le long de l'arbre du Vivant (Figure 2), et ce y compris entre espèces apparentées dans certains cas (Graves, 2008; Sarre et al., 2011; Vicoso & Bachtrog, 2015). Ceci suggère que des transitions récurrentes entre ces mécanismes ont eu lieu au cours de l'évolution. Dans le cadre des espèces à GSD suivant un déterminisme chromosomique du sexe, Ohno (1967) a proposé que les chromosomes sexuels dériveraient d'une même paire ancestral. Cette hypothèse tend à être soutenue par le fait qu'une partie des gènes impliqués dans le déterminisme du sexe sont fortement conservés. Pourtant, ce paradigme est remis en question. En effet, des études de cartographie génique comparant les chromosomes sexuels des mammifères (XY), des oiseaux et des serpents (ZW), ont démontré que ces gonosomes ne sont pas homologues, et donc qu'ils dérivent de paires d'autosomes différentes (Graves & Shetty, 2001; Matsubara et al., 2006; Veyrunes et al., 2008). Certaines homologies décelées entre les chromosomes XY et ZW observées chez des poissons (Ezaz et al., 2006) ne seraient pas non plus expliquées par une origine commune, mais par des phénomènes récents de transitions entre ces systèmes d'hétérogamétie. En effet, les transitions seraient

suffisamment récentes pour que les chromosomes sexuels soient toujours homologues.

Les mécanismes évolutifs sous-jacents à ces transitions sont actuellement encore peu connus, même si plusieurs forces évolutives ont été proposées pour expliquer des phénomènes de turnover entre systèmes d'hétérogamétie. Une étude théorique suggère ainsi qu'un gène autosomique sous sélection antagoniste (profitable à un sexe au détriment de l'autre sexe) peut causer l'augmentation en fréquence d'un nouveau déterminant génétique du sexe lié à ce gène autosomique (van Doorn & Kirkpatrick, 2007). Une autre piste serait que l'accumulation de mutations délétères sur le gonosome non-recombinant, en l'absence de mécanisme efficace de compensation de dosage génique, diminuerait la valeur sélective du sexe hétérogamétique à tel point que la sélection de gènes antagonistes ne parviendrait pas à compenser l'effet de ces mutations délétères, et provoquerait donc un turnover de chromosomes sexuels (Blaser et al., 2012). Plus récemment, une combinaison de ces deux hypothèses, nommée le "*hot potato model*", propose un cycle potentiellement infini de turnovers de chromosomes sexuels (Blaser et al., 2014). Ce modèle suggère qu'après l'arrêt de la recombinaison au niveau de la zone déterminant le sexe (suite à l'accumulation de mutations antagonistes), l'hétérochromosome accumule des mutations délétères à tels point que la transition devient inévitable. Ce modèle implique que le type d'hétérogamétie soit conservé lors de ces transitions, mais aussi que le recrutement des autosomes afin de générer de nouveaux chromosomes sexuels (par phénomène de fusion autosome/gonosome) n'est pas aléatoire (Blaser et al., 2014). Dans les deux derniers cas évoqués, une nouvelle paire de gonosomes est formée (Yoshida & Kitano, 2012). Un modèle de génétique des populations, basé sur l'évolution des termites, propose que des fusions/translocations entre autosomes et gonosomes auraient un avantage sélectif lorsque la sélection est en faveur des individus hétérozygotes, et donc que des turnovers de chromosomes sexuels seraient sélectionnés (Charlesworth & Wall, 1999). Une dernière hypothèse met en avant le rôle des conflits génétiques comme force évolutive générant de nouveaux chromosomes sexuels, dans le cadre de la sélection liée au sex-ratio (Werren & Beukeboom, 1998; Mank et al., 2014). Cette force évolutive va tendre à restaurer un sex-ratio équilibré chez des espèces produisant une descendance biaisée en

faveur d'un sexe (Beukeboom & Perrin, 2014), rétablissant ainsi un sex-ratio équilibré selon le principe de Fisher qui stipule que la majorité des espèces à reproduction sexuée produit une descendance au sex-ratio équilibré (Fisher, 1930).

Si plusieurs modèles pouvant expliquer ces transitions ont été proposés au cours des vingt dernières années, l'importance relative de ces différents modèles est encore peu éprouvée. De plus, les mécanismes moléculaires sous-jacents à ces transitions sont quant à eux toujours méconnus. Ces travaux théoriques n'ont eu encore que peu d'occasions d'être confrontés à des données empiriques. Pour pleinement comprendre la diversité des mécanismes de déterminisme du sexe, mais aussi les forces évolutives responsables de cette multiplicité de systèmes, il apparaît nécessaire d'étudier une grande diversité d'organismes, allant bien au-delà des modèles classiques. Le nombre croissant d'études portant sur les poissons téléostéens (Mank et al., 2006), les amphibiens (Sarre et al., 2011), les reptiles (Sarre et al., 2011) ou encore de nombreuses plantes (Charlesworth, 2002), laisse présager de prometteuses perspectives pour expliquer l'évolution du déterminisme du sexe. Parmi ces modèles émergents, les isopodes terrestres (Oniscidea, Crustacea) présentent une combinaison de caractéristiques unique, en faisant un sujet d'étude de choix.

II. Le modèle isopode terrestre

1. Evolution des isopodes terrestres

Les isopodes terrestres (ou cloportes) ont été étudiés depuis déjà de nombreux siècles, puisqu'Aristote (IV^{ème} siècle avant J.-C.) les classait parmi le groupe des insectes (Michel-Salzat, 2000). Ils ne seront considérés comme appartenant aux crustacés (à juste titre) qu'à partir du XVIII^{ème} siècle par Cuvier, qui dénombre alors 7 espèces de cloportes dans son essai *Mémoire sur les cloportes terrestres* (Cuvier, 1792). Les cloportes sont classés au sein de l'ordre des isopodes. Il s'agit d'un ordre particulièrement important en nombre et en diversité d'espèces, puisque plus de 10 000 espèces ont été décrites à l'heure actuelle (Poore & Bruce, 2012). Si leur position phylogénétique au sein des Pecarida (crustacés dits "supérieurs") est toujours sujette à débat, l'ordre des Isopoda est considéré majoritairement comme un ordre monophylétique (Tabacaru & Danielopol, 1999; Poore & Bruce, 2012). Les isopodes sont généralement divisés en 8 sous-ordres, dont les relations de parenté sont elles aussi méconnues (Figure 8; Schmidt, 2008). Ainsi, la monophylie de ces sous-ordres n'est pas résolument établie, comme pour l'exemple des Flabellifera ou encore les Gnathiidea (Wägele, 1989; Brusca & Wilson, 1991). De plus, les relations phylogénétiques au sein des sous-ordres sont elles aussi non résolues, c'est le cas en particulier pour le sous-ordre des isopodes terrestres (Oniscidea).

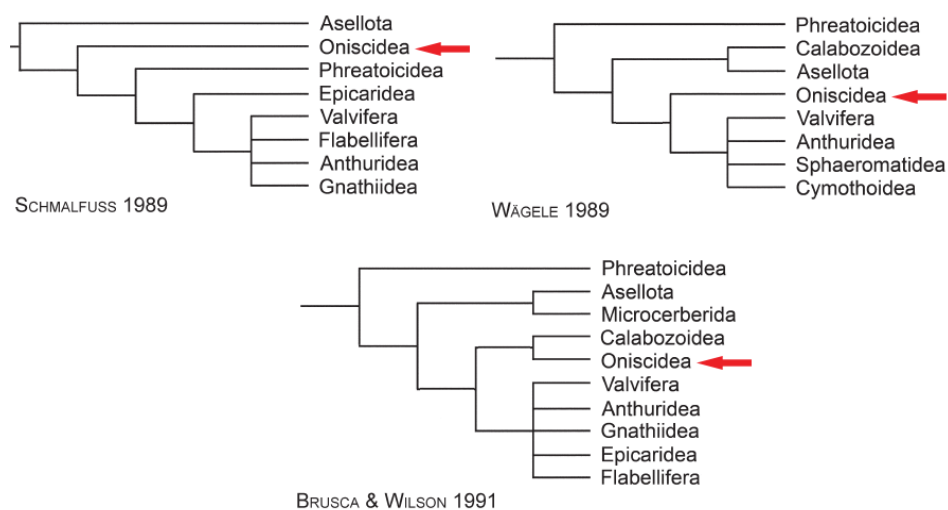


Figure 8 : Hypothèses phylogénétiques des relations basales entre les sous-ordres d'isopodes. Les Oniscidea sont indiqués par une flèche rouge. Les trois phylogénies, établies à partir de critères morphologiques, ne sont pas congruentes (adapté de Schmidt, 2008)

Les Oniscidea constituent un groupe remarquable de crustacés puisqu'il s'agit du seul comptant exclusivement des espèces terrestres. Ce sous-ordre présente donc un intérêt évolutif particulier puisqu'il illustre l'adaptation des crustacés à la vie terrestre. Actuellement, si près de 3 700 espèces de cloportes ont été décrites, la richesse spécifique des Oniscidea est estimée entre 5 000 et 7 000 espèces (Sfenthourakis & Taiti, 2015). L'ensemble des espèces décrites se répartissent dans 527 genres et 37 familles. Il s'agit d'un sous-ordre quasi ubiquiste, puisque les isopodes terrestres sont présents dans presque la totalité des biomes à l'exception des pôles et des régions de haute altitude (aucune espèce n'est recensée au-delà de 4 800 mètres d'altitude) (Schmalfuss, 2003; Sfenthourakis & Taiti, 2015).

Les relations phylogénétiques au sein des cloportes sont encore méconnues. En effet, les premiers travaux visant à décrire l'évolution des isopodes terrestres par des approches phylogénétiques étaient uniquement basés sur des critères morphologiques, et menaient à des conclusions différentes selon les études (Figure 8; Kussakin, 1979; Bruce, 1981; Schmalfuss, 1989; Wägele, 1989; Brusca & Wilson, 1991; Tabacaru & Danielopol, 1999). Ces divergences sont dues à la difficulté de déterminer des caractères morphologiques pertinents dans le cas d'études phylogénétiques (Barriel et al., 1993; Milinkovitch, 1995). Ces difficultés sont particulièrement prononcées dans le cas des isopodes terrestres puisqu'ils présentent de nombreuses convergences évolutives (Tabacaru & Danielopol, 1996; Wägele, 1996). A titre d'exemple, la capacité de volvation (se rouler en boule), présente chez les familles Armadillidae, Armadillidiidae ou encore Cylisticidae (Tuf et al., 2015), est clairement reconnue comme une convergence, comme le sont aussi certaines caractéristiques liées aux systèmes reproducteur et respiratoire (Vandel, 1960; Tabacaru & Danielopol, 1996).

Face à l'absence de congruence entre les phylogénies basées sur des critères morphologiques, et avec l'essor de la biologie moléculaire et des techniques de séquençage, deux études ont tenté de résoudre les relations phylogénétiques à l'aide de marqueurs moléculaires (Michel-Salzat & Bouchon, 2000; Mattern & Schlegel, 2001). Ces deux travaux, basés sur des gènes ribosomaux mitochondriaux, ne sont pas pour autant parvenus à résoudre les relations phylogénétiques des cloportes, de par la faible quantité de marqueurs

utilisés ainsi que leur signal phylogénétique insuffisant. Des travaux apparaissent encore nécessaires afin de pleinement comprendre l'histoire évolutive de ces crustacés terrestres, puisqu'aucune phylogénie robuste des Oniscidea n'est encore établie. L'augmentation du nombre de marqueurs phylogénétiques moléculaires semble être la meilleure solution pour résoudre ces relations encore mal comprises (Sfenthourakis & Taiti, 2015).

2. Anatomie

Sur le plan anatomique, les isopodes terrestres présentent un système digestif composé principalement du tube digestif en position dorsale ainsi que d'une paire de *caeca* (A, Figure 9). Le système nerveux se compose d'une chaîne nerveuse ventrale et du cerveau (B, Figure 9), tandis que le système reproducteur est formé d'une paire de gonades (A, Figure 9). Ces organes sont présents dans la cavité générale de l'individu et baignent dans l'hémolymphe (qui abrite potentiellement de nombreux organismes tels que des bactéries, des nématodes ou des acantocéphales, Cordaux et al., 2001; Dittmer et al., 2014; Sicard et al., 2014). La grande majorité des isopodes terrestres sont des espèces gonochoriques comportant des individus mâles et femelles. Les mâles disposent de deux paires de pléopodes différenciés en organes copulateurs, sous la forme de deux paires d'endopodites (Figure 10). La première paire est différenciée en stylets copulateurs tandis que la deuxième paire forme une gouttière accolée à la première paire d'endopodites, permettant l'écoulement du sperme (Vandel, 1960). Quant aux femelles, elles disposent de deux ouvertures génitales, situées à la base de la 5^{ème} paire de patte, qui permettent l'acheminement du sperme vers l'oviducte (Vandel, 1960).

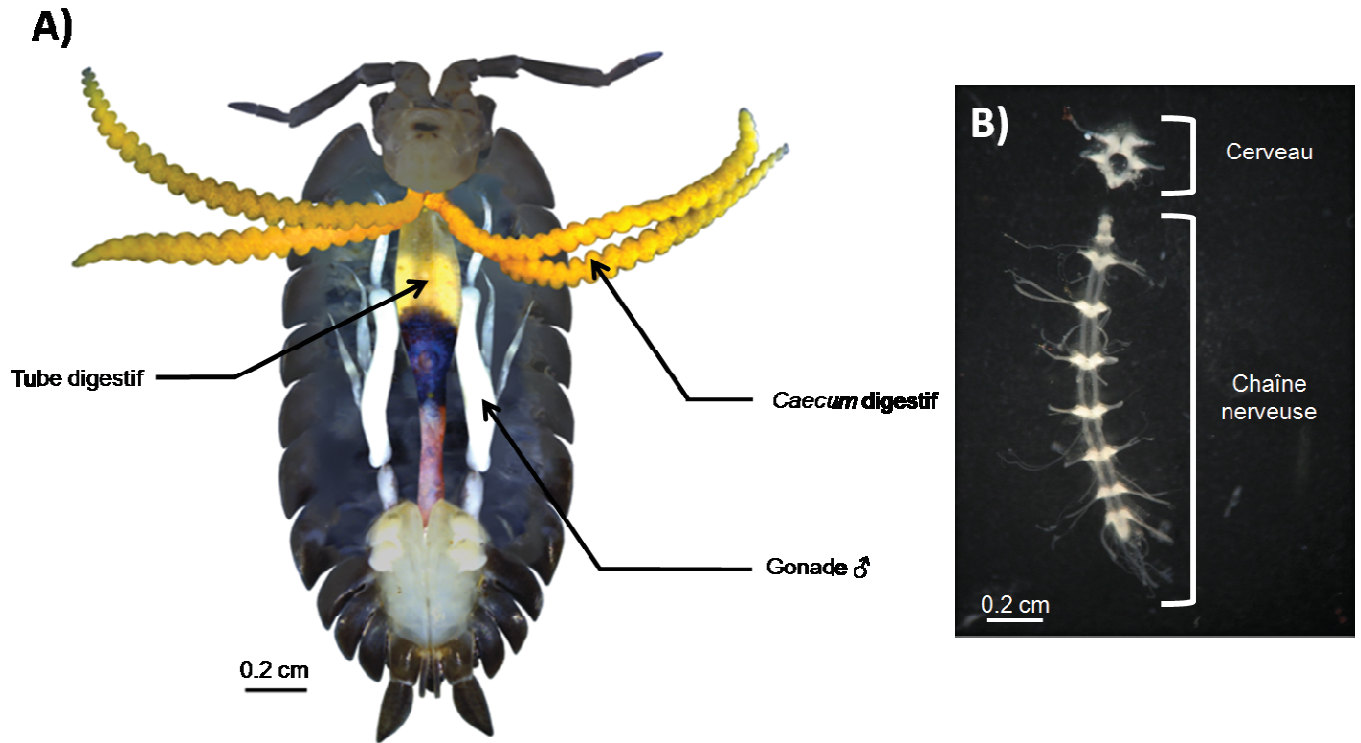


Figure 9 : Anatomie des différents organes de cloportes. A) Photographie recomposée des organes internes d'un isopode terrestre (exemple d'un mâle appartenant à l'espèce *Oniscus asellus*) observé en vue ventrale. B) Photographie du système nerveux d'un isopode terrestre (exemple d'un individu appartenant à l'espèce *Armadillidium vulgare*) (crédit : Laboratoire EBI, Poitiers).

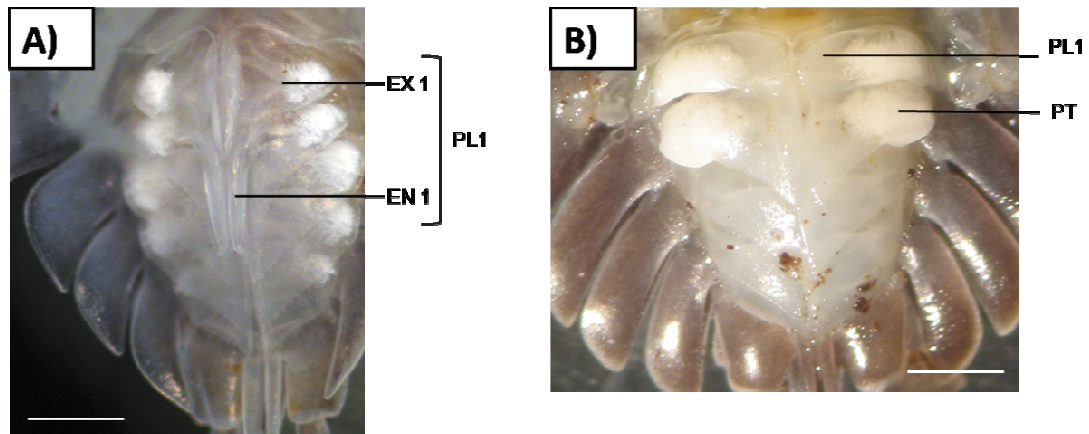


Figure 10 : Photographie de la face ventrale d'un mâle (A) et d'une femelle (B) *Cylisticus convexus*. EX 1 : exopodite 1 ; EN1 : endopodite 1 ; PL1 : pléopode 1 ; PT : pseudotrachée. La barre d'échelle correspond à 1.5 mm (crédit : laboratoire EBI, Poitiers)

3. Mécanismes de déterminisme du sexe

Chez les isopodes terrestres, comme chez de nombreux organismes, la base du déterminisme sexuel est génétique. Les premiers travaux effectués sur le déterminisme du sexe des isopodes terrestres ont consisté à établir les systèmes d'hétérogamétie suivant deux méthodes. Tout d'abord, la première méthode basée sur des études cytogénétiques a permis, dans certains cas, de révéler directement quel est le sexe hétérogamétique (grâce à l'hétéromorphie des chromosomes sexuels chez certaines espèces). Cependant, l'analyse caryotypique est le plus souvent complexe et peu informative, compte tenu du grand nombre de petits chromosomes généralement présents chez de nombreuses espèces de crustacés, et notamment chez les Oniscidea (Artault, 1977). Les chromosomes sexuels ont malgré tout pu être identifiés par cette approche chez quelques espèces d'isopodes terrestres (Mittal & Pahwa, 1980, 1981). La seconde méthode, originale, permet de déterminer le sexe hétérogamétique chez les isopodes terrestres en se basant sur l'analyse du sex-ratio de la descendance obtenue par croisement entre des néo-mâles phénotypiques (qui sont en fait des femelles génétiques transformées expérimentalement en mâles phénotypiques) avec des femelles génétiques (Figure 11, Legrand et al., 1987).

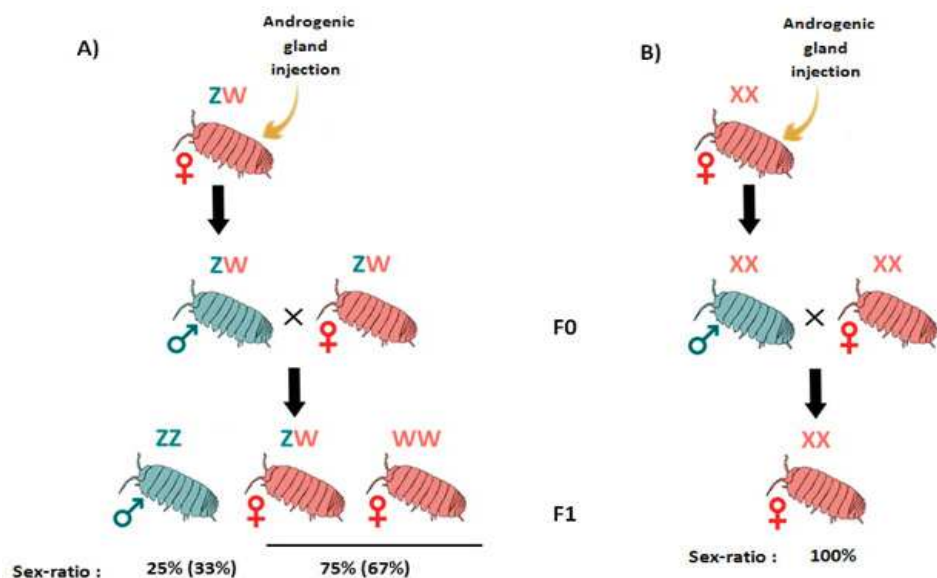


Figure 11 : Principe des croisements expérimentaux entre néo-mâles (i. e. femelles masculinisées) et femelles génétiques. A) Cas d'un système hétérogamétique femelle (ZZ/ZW). Le sex-ratio attendu de la descendance est de 75% de femelles (ZW et WW) et de 25% de mâles (ZZ) si le génotype WW est viable. Si le génotype WW n'est pas viable, alors le sex-ratio attendu est de 67% de femelles (ZW) et de 33% de mâles (ZZ). B) Cas d'un système hétérogamétique mâles (XY/XX). Le sex-ratio attendu de la descendance est de 100% de femelles (XX) (Becking et al., 2017).

Les différentes études menées au cours du XX^{ème} siècle ont montré que la plupart des espèces d'isopodes terrestres suivait un déterminisme hétérochromosomique du sexe (Rigaud et al., 1997). Ainsi les deux systèmes d'hétérogamétie (XY/XX et ZZ/ZW) ont été rapportés chez les Oniscidea, parfois au sein d'un même genre (exemple d'*Armadillidium vulgare* [ZZ/ZW] et d'*Armadillidium nasatum* [XY/XX] ; respectivement Juchault & Legrand, 1972, 1979), voir au sein d'une même espèce (exemple de *Porcellio dilatatus dilatatus* [XY/XX] et *P. d. petiti* [ZZ/ZW] ; respectivement Juchault & Legrand, 1964; Legrand et al., 1974). Cependant, les informations concernant la diversité des systèmes d'hétérogamétie rencontrés chez les cloportes demeurent encore largement parcellaires, puisque les systèmes d'hétérogamétie ne sont connues que chez neuf espèces d'isopodes terrestres (Figure 12; Juchault & Rigaud, 1995; Rigaud et al., 1997). Néanmoins, ces informations (bien que très fragmentaires à l'échelle des Oniscidea) laissent présager une grande variabilité des systèmes de déterminisme du sexe. Ainsi, le clade des isopodes terrestres constitue une excellente opportunité d'étudier l'évolution des systèmes d'hétérogamétie, et donc de mieux comprendre non seulement la diversité des mécanismes de déterminisme du sexe rencontrée chez ces espèces mais aussi les forces évolutives responsables de cette diversité.

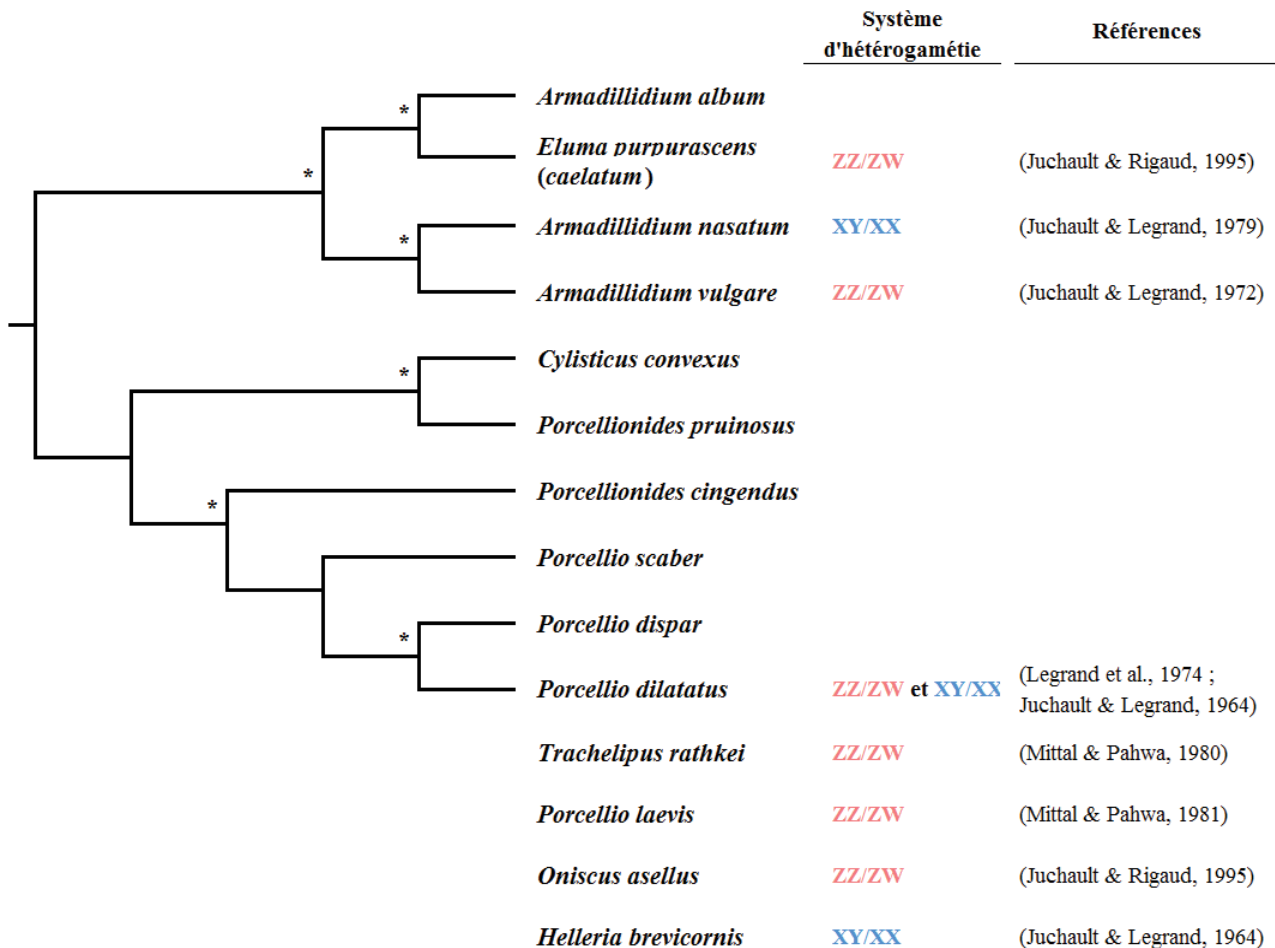


Figure 12 : Relations phylogénétiques entre espèces d'isopodes terrestres associées aux systèmes d'hétérogamétie connus. Les relations affichées sont tirées de Michel-Salzat & Bouchon (2000). Les astérisques correspondent aux nœuds soutenus par une valeur de bootstrap supérieure à 70%. Les espèces *T. rathkei*, *P. laevis*, *O. asellus* et *H. brevicornis* n'étant pas incluses dans la publication de Michel-Salzat & Bouchon (2000), leur position phylogénétique demeure indéterminée.

Comme évoqué précédemment, les études cytogénétiques réalisées ont le plus souvent décrit une absence d'hétéromorphie entre les deux hétérochromosomes (quel que soit le système considéré). La quasi-absence d'hétéromorphie ainsi que la présence de différents systèmes d'hétérogamétie parfois au sein d'une même espèce sont généralement considérés comme des indices suggérant que les gonosomes sont à une étape particulièrement précoce de leur processus de dégénérescence (Bull, 1983; Charlesworth, 1991). Basée sur un ensemble de données expérimentales, une étude a proposé que dans le cas du cloporte commun *A. vulgare*, mâles et femelles partageraient pour l'essentiel un même génotype (Legrand et al., 1987). Les gonosomes de mâles et de femelles différeraient seulement par la présence ou l'absence d'un unique facteur génétique déterminant le sexe, simplement sous la forme d'un gène déclencheur induisant ultimement la différenciation sexuelle.

Même si les relations phylogénétiques des isopodes terrestres sont encore méconnues, la distribution discontinue des systèmes d'hétérogamétie connus dans l'arbre des isopodes (Figure 12) suggère que des changements de mécanismes se sont produits de façon récurrente au cours de l'évolution de ces taxons. Cependant, les relations phylogénétiques incertaines empêchent de connaître le nombre et la direction de ces transitions de XY/XX à ZZ/ZW (ou inversement). La question est également de savoir quels facteurs pourraient expliquer une telle variabilité et une telle labilité des mécanismes de déterminisme du sexe chez les Oniscidea.

4. Différenciation sexuelle

Le seul effecteur connu de la cascade de différenciation sexuelle des isopodes terrestres est l'hormone androgène, qui est responsable du développement en mâle. Cet effecteur est d'ailleurs le dernier de la cascade de différenciation sexuelle (Suzuki, 1999). Il s'agit d'une hormone circulante, principalement sécrétée par la glande androgène (A, Figure 13) qui est située à l'extrémité de chaque utricule des gonades mâles (Martin & Kowallik, 1999; Okuno et al., 1999; Suzuki, 1999). Les isopodes terrestres mâles comptent six glandes androgènes (trois par gonade), qui sont constituées de nombreuses cellules enveloppées dans une gaine conjonctive, et sont localisées dans les filaments suspenseurs des utricules testiculaires (Juchault, 1966).

Les individus naissent sexuellement indifférenciés et se différencient au cours d'une fenêtre temporelle restreinte. En l'absence de l'hormone androgène, un phénotype femelle se développe par défaut (Charniaux-Cotton, 1959). Chez le cloporte commun *A. vulgare*, la différenciation sexuelle débute à partir de la 4^{ème} mue post-natale (C, Figure 13) (Katakura, 1984; Suzuki & Yamasaki, 1995a; Badawi et al., 2015) tandis qu'elle débute à partir de la 3^{ème} mue chez *Cylisticus convexus* (C, Figure 13) (Badawi et al., 2015) et dès la 2^{nde} mue chez *Helleria brevicornis* (Juchault, 1966). Ces quelques exemples illustrent la diversité observée dans la chronologie de la différenciation sexuelle des isopodes terrestres. Cette différenciation sexuelle particulière (différenciation hormone-dépendante des mâles ; existence d'une fenêtre temporelle de différenciation) permet de

manipuler expérimentalement le phénotype sexuel des individus. En effet, il est possible d'induire le développement de femelles génétiques en mâles phénotypiques fonctionnels (ces individus sont alors qualifiés de "néo-mâles") par l'injection de glandes androgènes au cours des premières étapes de la différenciation sexuelle.

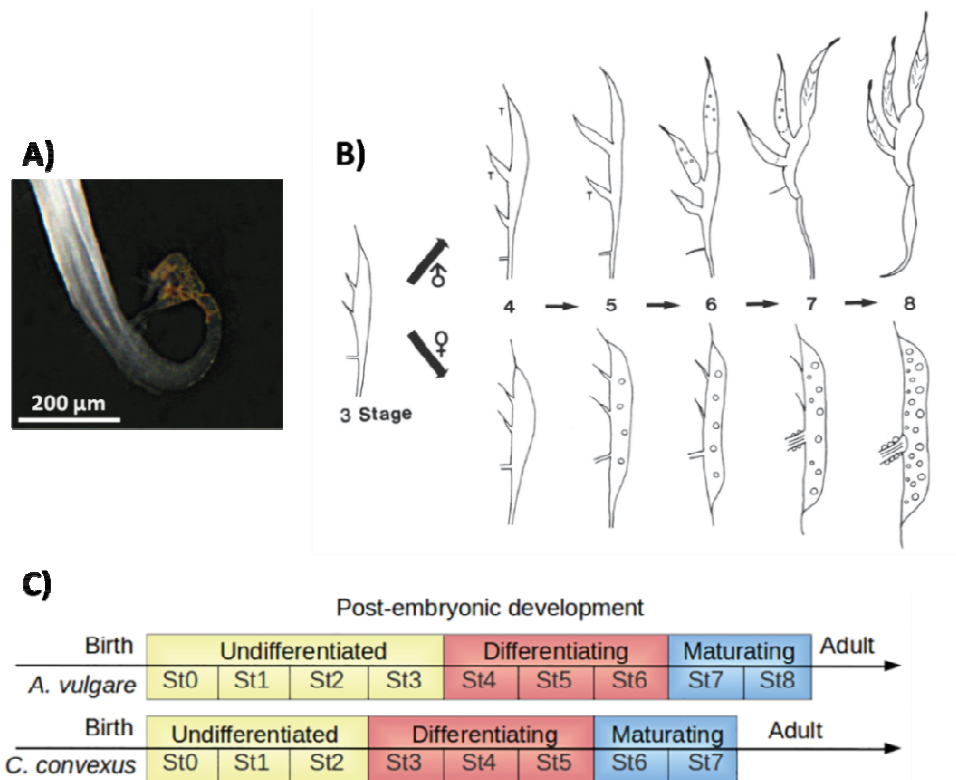


Figure 13 : Différenciation des gonades d'isopodes terrestres. A) Glande androgène d'*Armadillidium versicolor* observée à la loupe binoculaire (crédit : T. Becking). Cette glande est composée d'un ensemble de cellules sécrétant l'hormone androgène. Elle est située à l'extrémité de chaque testicule, incluse dans le filament suspenseur. B) Schéma du développement morphologique des gonades mâles et femelles d'*A. vulgare*, du stade 3 (indifférenciée) au stade 8 (différenciation totale). La partie supérieure du schéma correspond au développement des gonades mâles et la partie inférieure correspond au développement des gonades femelles. Chaque stade est séparé par une mue (d'après Suzuki et al., 1995). C) Chronologie de la différenciation sexuelle d'*A. vulgare* et de *C. convexus*. La différenciation est un stade plus précoce chez *C. convexus* et dure le même nombre de stades de mues (d'après Badawi et al., 2015).

III. Association isopodes terrestres/*Wolbachia*

1. Présentation et historique de *Wolbachia*

Une caractéristique importante de nombreuses espèces d'isopodes terrestres est leur association avec la bactérie *Wolbachia pipientis*, un micro-organisme endosymbiotique strict (c'est-à-dire que le symbiote ne peut vivre que dans les cellules de son hôte). Elle doit son nom au fait qu'elle a été découverte pour la première fois chez le moustique *Culex pipiens* par Simeon B. Hertig et Marshall Wolbach en 1924 (Hertig & Wolbach, 1924). Il s'agit d'une Alphaprotéobactérie gram-négative appartenant à la famille des Rickettsiales (Hertig, 1936). Elle est connue pour infecter un large spectre d'hôtes (Tableau 1), à tel point qu'elle est considérée comme un des endosymbiotes les plus répandus sur Terre (LePage & Bordenstein, 2013; Makepeace & Gill, 2016), infectant à la fois des nématodes (Taylor et al., 2005) et plus de 40% des arthropodes (Bouchon et al., 2008; Zug & Hammerstein, 2012). Chez ces derniers, *Wolbachia* est présente chez près de 66% des espèces d'insectes (Hilgenboecker et al., 2008), chez les chélicérates (Duron et al., 2008), et chez les crustacés (Cordaux et al., 2012). En particulier, la présence de cet endosymbiote a été rapportée dans près de 61% des espèces d'isopodes terrestres (Bouchon et al., 2008).

Tableau 1 : Classification taxonomique des différents hôtes de *Wolbachia* (d'après Makepeace & Gill, 2016).

Phylum	Sous-phylum	Classe	Ordre	Nom commun
Arthropoda	Hexapoda	Insecta	Archaeognatha	archaeognathes
			Blattodea	blattes et termites
			Coleoptera	coléoptères
			Diptera	mouches
			Hemiptera	punaïses
			Hymenoptera	guêpes, abeilles et fourmis
			Lepidoptera	mites et papillons
			Mantodea	mantes
			Neuroptera	névroptères
			Odonata	libellules et demoiselles
			Orthoptera	sauterelles et criquets
			Phthiraptera	poux
			Plecoptera	plécoptères
			Psocoptera	psocoptères
			Siphonaptera	puces
			Strepsiptera	strepsiptères
			Thysanoptera	thrips
		Entognatha		
		Collembola		collemboles
	Crustacea	Maxillopoda		
		Pedunculata		pousse-pieds
		Ostracoda		
		Podocopida		ostracodes
		Malacostraca		
	Chelicerata	Isopoda		cloportes, poux de mer
		Amphipoda		puces de mer
		Arachnida		
		Trombidiformes		tétranyques
		Sarcoptiformes		oribates et sarcoptoïdés
		Parasitiformes		tiques
		Araneae		araignées
		Pseudoscorpionida		pseudoscorpions
		Scorpiones		scorpions
Nematoda	-	Chromadoria		
		Rhabditida		vers ronds

La transmission de *Wolbachia* s'effectue principalement par voie ovocytaire, ce qui signifie qu'elle doit coloniser les cellules de la lignée germinale pour pouvoir être transmise. Cependant, les phylogénies de *Wolbachia* et de ses hôtes ne sont pas congruentes, suggérant que des transferts horizontaux de *Wolbachia* se sont produits au cours de l'évolution (Werren et al., 1995; Shoemaker et al., 2002). Ces transferts horizontaux ont pu s'effectuer naturellement par ingestion, par parasitisme ou même encore par contact direct d'hémolymphe (Juchault & Rigaud, 1995; Heath et al., 1999; Cordaux et al., 2001; Huigens et al., 2004; Le Clec'h et al., 2013a, 2013b; Ahmed et al., 2015). En plus de sa présence dans les tissus germinaux, selon l'hôte considéré, cet endosymbiote intracellulaire est susceptible d'occuper d'autre tissus de la lignée somatique. Plusieurs études *in vivo* ont décrit la présence de *Wolbachia* dans le système nerveux, le système digestif ou encore dans l'hémolymphe d'insectes et d'isopodes terrestres notamment (Dobson et al., 1999; Dittmer et al., 2014). Cette bactérie dispose d'une double membrane sans paroi ainsi que d'une troisième membrane vraisemblablement dérivée du réticulum endoplasmique ou d'un autre organe de son hôte (Figure 14, Martin et al., 1973; Voronin et al., 2004).

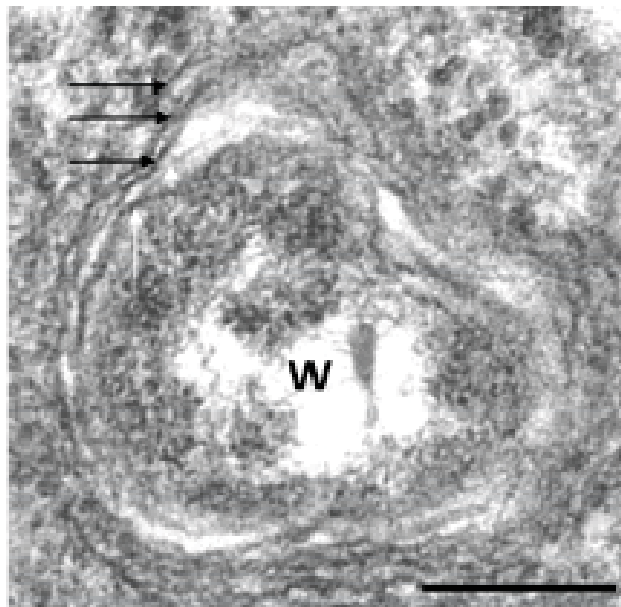


Figure 14 : Observation de *Wolbachia* en microscopie électronique à transmission. La bactérie (notée W) est ici présente dans le cytoplasme d'une cellule nerveuse d'*A. vulgare*. L'échelle représente 500 nm. Les flèches noires indiquent les membranes entourant la bactérie (image tirée de Sicard et al., 2014)

Une telle organisation pourrait être un moyen de faciliter la communication entre *Wolbachia* et la cellule hôte car les facteurs sécrétés à la surface de ces membranes (tant par la bactérie que par son hôte) se situeraient directement à l'interface physique entre les deux acteurs de cette symbiose. La bactérie serait ainsi en mesure de transporter ses facteurs vers le cytoplasme de l'hôte tout en étant en contact avec les nutriments de l'hôte (Serbus et al., 2008). L'ubiquité cellulaire et tissulaire de *Wolbachia* associée à sa proximité avec le cytoplasme de son hôte pourraient expliquer le fort impact de cette bactérie sur les traits d'histoire de vie de ses hôtes, ainsi que les conséquences phénotypiques qui en découlent.

2. Conséquences de *Wolbachia* sur le phénotype de ses hôtes

a) Wolbachia, un parasite de la reproduction

Pour la majorité de ses hôtes arthropodes, *Wolbachia* est considérée comme un parasite de la reproduction (Werren et al., 2008; Cordaux et al., 2011). La bactérie se transmet verticalement et a su développer plusieurs stratégies lui permettant d'accroître son taux de transmission dans les populations hôtes (Bordenstein & Reznikoff, 2005; Jaenike, 2012). Si d'autres bactéries endosymbiotiques sont connues pour perturber la reproduction de leur hôte (Duron et al., 2008; Cordaux et al., 2011), *Wolbachia* est à ce jour la seule capable d'induire les quatre types de manipulations de la reproduction connus : l'incompatibilité cytoplasmique, le *male-killing*, la parthénogenèse thélytoque et la féminisation de mâles génétiques.

L'incompatibilité cytoplasmique (IC), premier effet connu de *Wolbachia* (Yen & Barr, 1971), observée à la fois chez les insectes, les chélicérates et les crustacés, constitue la manipulation de la reproduction la plus répandue et supposée la plus ancestrale induite par *Wolbachia* (Rousset et al., 1992; Werren et al., 2008). Parmi les quatre effets connus de *Wolbachia*, il s'agit par ailleurs du seul à ne pas induire de biais de sex-ratio. Les femelles infectées par la bactérie disposent d'une meilleure valeur sélective que les femelles non infectées, favorisant ainsi la propagation du symbiote dans les populations infectées (Xi et

al., 2005; Engelstädter & Telschow, 2009). Les femelles porteuses de *Wolbachia* peuvent ainsi avoir une descendance viable en s'accouplant avec des mâles infectés ou non, tandis que les femelles asymbiotiques avortent lorsqu'elles se reproduisent avec des mâles infectés (A, Figure 15) (Yen & Barr, 1971, 1973). Ceci serait causé par l'existence de deux facteurs générés par *Wolbachia* au sein des gamètes : le facteur "mod" (modificateur) présent dans les ovocytes, et le facteur "resc" (restaurateur) retrouvé au niveau des spermatozoïdes (Zabalou et al., 2008). La non-concordance de ces deux facteurs aboutit à une mauvaise séparation des chromosomes lors de la première mitose post-fécondation, conduisant à la mort de l'embryon (Tram et al., 2006). Enfin, deux types d'IC peuvent être différenciés : l'IC unidirectionnelle lorsqu'une seule souche de *Wolbachia* est impliquée, ou l'IC bidirectionnelle qui concerne plusieurs souches de *Wolbachia*, avec potentiellement plusieurs facteurs "mod" ou "resc" différents (A, Figure 15) (Bordenstein et al., 2001; Telschow et al., 2005; Jaenike et al., 2006; Miller et al., 2010). A l'échelle cellulaire, il apparaît que lors de l'anaphase, les chromosomes paternels sont partiellement condensés, et ainsi soit séquestrés au niveau des centrosomes, soit exclus (Lassy & Karr, 1996; Tram et al., 2006). Si les mécanismes responsables de l'IC sont encore méconnus, deux études fonctionnelles récentes, menées sur les souches de *Wolbachia* infectant les hôtes *D. melanogaster* et *Culex pipiens*, ont permis de mettre en évidence l'importance de plusieurs *loci* bactériens responsables de l'IC (Beckmann et al., 2017; LePage et al., 2017).

Le *male killing* provoqué par la présence de *Wolbachia* a pour effet la mort ciblée des embryons mâles (B, Figure 15). Cet effet est observé chez cinq ordres d'insectes mais aussi chez des acariens (Hurst et al., 2003; Cordaux et al., 2011). Cette manipulation est considérée comme avantageuse pour les femelles de la descendance car elles vont obtenir les ressources allouées théoriquement à leurs frères, avec une possibilité de consommer ces frères morts dans le cas d'hôtes cannibales (Hurst & Majerus, 1993; Elnagdy et al., 2011). Il s'agit d'un *male killing* qualifié de précoce car ce processus est induit au cours de l'embryogenèse, si ce n'est lors des premiers stades du développement larvaire (Kageyama et al., 2012). Si les mécanismes moléculaires responsables du male-killing demeurent encore méconnus, *Wolbachia* semble induire des biais de ségrégation chromatique

durant les premières phases de développement de *Drosophila bifasciata*, conduisant à la mort des embryons mâles avant leur éclosion (Riparbelli et al., 2012). Plus récemment, une étude effectuée sur les lépidoptères a révélé que *Wolbachia*, en réprimant l'expression du gène masculinisant *Masc* (situé sur le chromosome Z), empêche une compensation de dosage génique efficace et conduit à la mort des embryons mâles (Fukui et al., 2015).

Dans le cas d'insectes et de quelques acariens ayant un déterminisme du sexe haplo-diploïde, *Wolbachia* est capable d'induire une parthénogenèse thélytoque chez les individus infectés (Stouthamer et al., 1990; Huigens & Stouthamer, 2003). Pour cela, la bactérie peut induire la diploïdisation des œufs non-fécondés qui se développent ainsi en femelles (C, Figure 15). Cette

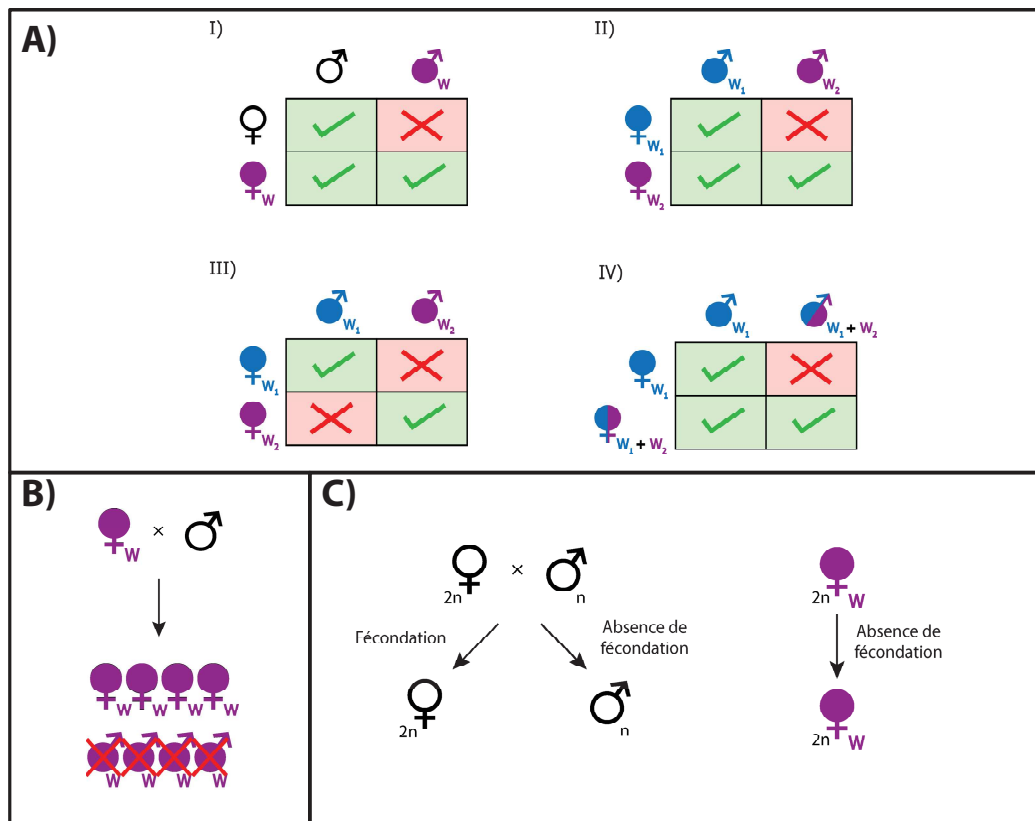


Figure 15 : Mécanismes de manipulation de la reproduction induite par *Wolbachia*. A) Description des différents types d'incompatibilités cytoplasmiques (IC). Les individus peuvent être non-infectés ou infectés (notés W) par une seule souche ou par plusieurs souches différentes de *Wolbachia*. I) IC unidirectionnelle entre individus infectés et non infectés. II) IC unidirectionnelle entre individus infectés par les souches 1 et 2. III) IC bidirectionnelle entre individus infectés par les souches 1 et 2. IV) IC unidirectionnelle entre individus bi-infectés par 1 et 2 et individus mono-infectés par 1. Les croisements compatibles sont en vert (+) et les croisements incompatibles en rouge (-) B) Représentation du *male killing* induit par *Wolbachia*. La transmission de la bactérie est responsable de la mort des embryons mâles. C) Induction de la parthénogenèse thélytoque, caractérisée par la conversion de mâles génétiques haploïdes en femelles génétiques diploïdes. n correspond à des individus haploïdes, et 2n correspond à des individus diploïdes. La couleur violette (et bleue dans le cas de l'IC) indique les individus porteurs de *Wolbachia*.

diploïdisation peut s'effectuer selon trois modalités distinctes : (i) par absence de ségrégation des chromatides lors de l'anaphase, comme chez les guêpes appartenant au genre *Trichogramma* (Huigens & Stouthamer, 2003), (ii) par la fusion des deux cellules filles issues de la première division mitotique, comme chez la guêpe *Muscidifurax uniraptor* (Gottlieb et al., 2002), (iii) par des perturbations au niveau méiotique aboutissant à la formation d'œufs diploïdes, comme c'est le cas chez l'acarien *Brevipalpus phoenicis* (Weeks et al., 2001).

Enfin, un autre effet particulièrement spectaculaire de *Wolbachia* est la féminisation des mâles génétiques en femelles fonctionnelles (Figure 16). Il s'agit du second effet de la bactérie découvert chez les arthropodes (Martin et al., 1973). La féminisation a été découverte tout d'abord chez le cloporte commun *A. vulgare* (Martin et al., 1973; Cordaux et al., 2004), avant d'être établie dans d'autres espèces d'isopodes terrestres (chez *A. nasatum*, *P. pruinus* et *O. asellus*; Juchault & Legrand, 1989; Marcadé et al., 1999; Rigaud et al., 1999, respectivement). Elle est de plus suspectée chez de nombreuses autres espèces d'isopodes (Bouchon et al., 1998, 2008). L'hypothèse d'une bactérie symbiotique féminisante a été proposée dès les années 1940 avec l'observation d'individus intersexués (partageant des caractères à la fois mâles et femelles) dans des populations naturelles d'*A. vulgare* (Legrand, 1941).

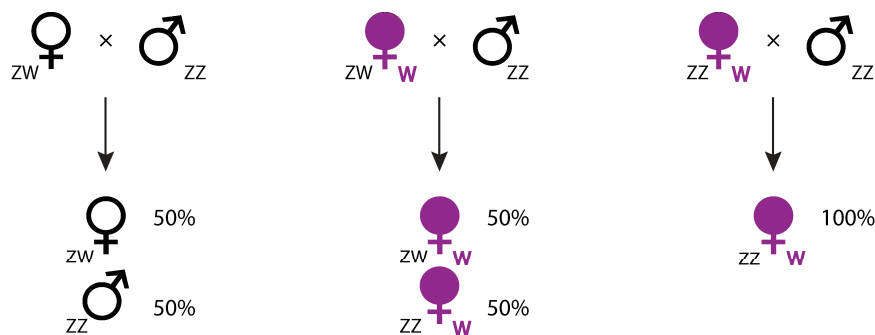


Figure 16 : Principe de la féminisation induite par *Wolbachia* chez des individus présentant une hétérogamétie femelle. Les mâles génétiques ZZ infectés par *Wolbachia* se développent en femelles phénotypiques. Au fur et à mesure des générations, le génotype ZZ tend à envahir la population. Les individus infectés sont colorés en violet (adapté de Cordaux et al., 2011). (Par souci de simplicité, le taux de transmission de *Wolbachia* est considéré comme étant de 100%).

La combinaison entre des observations de forts biais de sex-ratio en faveur des femelles et la mise en place de lignées matrilineaires (suivies de mères en filles) ont permis une investigation plus précise, jusqu'à permettre une observation par microscopie électronique d'une bactérie intracytoplasmique en 1973 (Martin et

al., 1973). Par ailleurs, la féminisation induite par *Wolbachia* n'est pas restreinte aux seuls crustacés, puisque ce phénotype induit a aussi été démontré chez le papillon *Eurema hecabe* (Hiroki et al., 2002; Narita et al., 2007) ou encore chez la cicadelle *Zyginidia pullula* (Negri et al., 2006, 2009). Il est cependant causé par des souches de *Wolbachia* génétiquement distantes de celles présentes chez les cloportes.

Les processus moléculaires responsables de la féminisation sont encore méconnus. Ils ne sont probablement pas les mêmes entre crustacés et insectes, puisque la base de la différenciation sexuelle est différente. Chez la plupart des insectes, la différenciation sexuelle s'effectue par un mécanisme cellulaire, qui se déclenche très tôt et de façon autonome au cours du développement embryonnaire (Sánchez, 2008; Gempe & Beye, 2011). L'action de *Wolbachia* s'effectue donc probablement de manière continue au cours de l'ensemble du développement embryonnaire, et ce dans toutes les cellules de l'organisme (Cordaux et al., 2011; Kageyama et al., 2012). Si la bactérie n'est présente que dans une partie des cellules ou si elle est éliminée par traitement antibiotique au cours du développement, alors les individus développent des phénotypes intersexués, présentant des tissus à la fois mâles et femelles (Narita et al., 2007). En revanche chez l'ensemble des malacostracés (crustacés dits "supérieurs"), comme indiqué précédemment, la différenciation sexuelle est sous contrôle hormonal avec la présence de l'hormone androgène, synthétisée par la glande androgène (Charniaux-Cotton, 1962; Martin et al., 1999; Grève et al., 2004). Par un mécanisme encore inconnu aujourd'hui, *Wolbachia* parviendrait à empêcher la production de cette hormone en prévenant la formation de la glande androgène, ce qui convertit les mâles génétiques en femelles phénotypiques fonctionnelles (Rigaud, 1997; Rigaud et al., 1997; Bouchon et al., 2008; Cordaux et al., 2011). Son effet pourrait être analogue à celui du déterminant génétique primaire porté par le chromosome W, à savoir qu'elle pourrait cibler le ou les gènes du déterminisme ou de la différenciation mâle (Juchault & Legrand, 1985).

b) *Wolbachia*, un symbiote mutualiste

L'endosymbiote *Wolbachia* peut aussi devenir non plus un parasite de la reproduction pour maximiser sa transmission, mais un acteur mutualiste essentiel au développement et à la reproduction de ses hôtes. C'est notamment le cas chez ses hôtes nématodes, puisque l'élimination de la bactérie par traitements antibiotiques a permis de démontrer son utilité pour la fertilité et le développement larvaire des nématodes (Taylor et al., 2005). Le séquençage des génomes de *Wolbachia* a permis de démontrer que des souches infectant les nématodes pouvaient générer des molécules vitales pour l'hôte, telles que l'hème nécessaire à l'embryogenèse (Warbrick et al., 1993; Taylor et al., 2005), la riboflavine et l'adénine flavine dinucléotide (Foster et al., 2005). Des analyses basées sur des données protéomiques et transcriptomiques ont aussi révélé que *Wolbachia* aurait non seulement un rôle antibactérien mais aussi un rôle de générateur d'ATP pour l'hôte *Onchocerca ochengi* (Darby et al., 2012).

La symbiose mutualiste de *Wolbachia* dépasse le cadre des nématodes et concerne aussi les arthropodes. Cette bactérie est connue pour entretenir une relation mutualiste avec une espèce de punaise de lit, *Cimex lectularius* (Hemiptera), en lui fournissant la biotine (vitamine B8) (Hosokawa et al., 2010). La souche de *Wolbachia* concernée est en fait la seule décrite disposant de la voie de biosynthèse complète de cette vitamine (Nikoh et al., 2014). Une étude récente souligne aussi l'influence de *Wolbachia* dans le cadre de la résistance antibactérienne de la drosophile (Gupta et al., 2017), en plus d'avoir un effet antiviral (Hedges et al., 2008; Teixeira et al., 2008). Enfin, l'infection par *Wolbachia* est aussi nécessaire à l'ovogenèse de la guêpe parasitoïde *Asobara tabida* (Dedeine et al., 2001). Des études ultérieures ont démontré que *Wolbachia* modifiait la régulation de l'apoptose des ovogonies (cellules souches de la lignée germinale femelle), rendant ainsi sa présence indispensable pour l'hôte (Pannebakker et al., 2007; Kremer et al., 2009). Ceci constitue une véritable transition d'une symbiose parasitaire vers une symbiose mutualiste par dépendance de l'hôte vis-à-vis de son symbiote.

3. Conflits génétiques et conséquences sur le déterminisme du sexe

L'élaboration de modèles mathématiques a permis de poser pour hypothèse que la féminisation induite par *Wolbachia* pourrait entraîner, à l'échelle évolutive, des changements récurrents de systèmes d'hétérogamétie via l'existence de "conflits génétiques" (Caubet et al., 2000; Werren & Beukeboom, 1998). L'association isopodes terrestres/*Wolbachia* représente le modèle historique ayant permis la découverte de la féminisation (Martin et al., 1973). Cette dernière est susceptible de causer de nombreux conflits génétiques, ce qui rend cette association parfaitement adéquate pour l'étude des transitions de systèmes d'hétérogamétie.

Des conflits entre deux éléments génétiques apparaissent lorsque ces éléments subissent des pressions de sélections opposées. Ainsi l'augmentation de la valeur sélective (ou fitness) d'un élément déclenche la diminution de la valeur sélective de l'autre élément. Les *Wolbachia* féminisantes, en temps qu'endosymbiotes perturbateurs du sex-ratio, sont potentiellement responsables de tels conflits. La bactérie favorise les femelles dans les populations infectées et augmente ainsi leur nombre. Les mâles constituent alors le sexe rare. Ainsi les gènes codant pour la fonction mâle sont favorisés par la sélection, contrebalançant le biais de sex-ratio en faveur des femelles (Bull, 1983) et ainsi de permettre le retour à un sex-ratio équilibré (dit "équilibre de Fisher"; Fisher, 1930). De tels conflits sont observables dans des lignées maternelles d'*A. vulgare*, où des mâles génétiques ZZ sont convertis en femelles phénotypiques fonctionnelles par l'action de *Wolbachia*, induisant des sex-ratios biaisés en faveur des femelles (Rigaud et al., 1992). En effet, si les femelles génétiques (ZW) produisent 50% de femelles (elles aussi ZW), les portées issues d'individus ZZ féminisées seront typiquement composées de 80% à 90% de femelles ZZ (car féminisées). Si les deux types de femelles (ZW et ZZ infectées par *Wolbachia*) ont une fécondité similaire, alors les modèles théoriques prédisent la disparition des femelles génétiques ZW (Taylor, 1990; Ferdy et al., 2016), ce qui est confirmé par l'analyse de données empiriques (Juchault et al., 1980, 1993). Dans un tel cas, puisque tous les individus ont un génotype ZZ, le sexe n'est alors déterminé que

par la présence ou l'absence de la bactérie. Le déterminant du sexe n'étant plus nucléaire (le chromosome W ayant disparu) mais cytoplasmique (*Wolbachia*), il est alors possible de parler de "déterminisme cytoplasmique du sexe" (Rigaud, 1997; Rigaud et al., 1997; Bouchon et al., 2008; Cordaux et al., 2011). Cette situation, qui provoque d'importants biais de sex-ratio en faveur des femelles, est donc propice à l'évolution de facteurs rétablissant un sex-ratio équilibré. Ainsi, un système polygénique limitant la transmission de la bactérie a été mis en évidence dans certaines lignées maternelles, permettant le retour à un équilibre fisherien (Rigaud & Juchault, 1992).

Un autre perturbateur du sex-ratio, d'origine nucléaire, a aussi été mis en évidence chez certaines lignées d'*A. vulgare*. Cet élément génétique, appelé "élément f", constitue tout comme *Wolbachia* un facteur féminisant (Legrand & Juchault, 1984). A l'époque de sa découverte, il a été proposé que ce facteur pourrait résulter de l'intégration d'une partie du génome de *Wolbachia* dans le génome d'*A. vulgare* (Legrand & Juchault, 1984). Ce n'est que plus de 30 ans plus tard que cette hypothèse a été validée, grâce à l'utilisation de technologies de séquençage à haut débit (Leclercq et al., 2016). En effet, il a été démontré récemment que plus de 80% du génome de *Wolbachia* s'était intégré au sein du génome d'*A. vulgare* avec un impact fonctionnel direct puisque le chromosome portant cet insert bactérien a la même fonction qu'un nouveau chromosome W (Leclercq et al., 2016). Il est suspecté que face à ce nouveau facteur féminisant, un gène masculinisant *M*, épistatique sur l'élément f mais insuffisant pour contrer l'effet féminisant de *Wolbachia*, a été sélectionné dans certaines lignées générant un équilibre entre facteurs féminisants et masculinisant (Juchault et al., 1992; Rigaud & Juchault, 1992, 1993).

A partir des différents facteurs identifiés chez *A. vulgare*, un scénario a été élaboré pour expliquer l'évolution dynamique des mécanismes du sexe chez cette espèce (Figure 17). Selon ce scénario, des *Wolbachia* féminisantes se disperseraient dans des lignées maternelles à hétérogamétie femelle. Les males féminisés ZZ envahiraient progressivement la population et remplaceraient les femelles ZW (qui à terme disparaîtraient). La population ne serait donc plus composée que de mâles et de femelles ZZ (le sexe n'étant déterminé que par la présence ou par l'absence de la bactérie). Ce scénario prévoit que *Wolbachia*

transfère une part de son génome à l'hôte, générant ainsi l'élément f. Le biais de sex-ratio induit par *Wolbachia* ou par l'élément f favoriserait ensuite la sélection d'un gène autosomique masculinisant, afin de rétablir un équilibre fisherien.

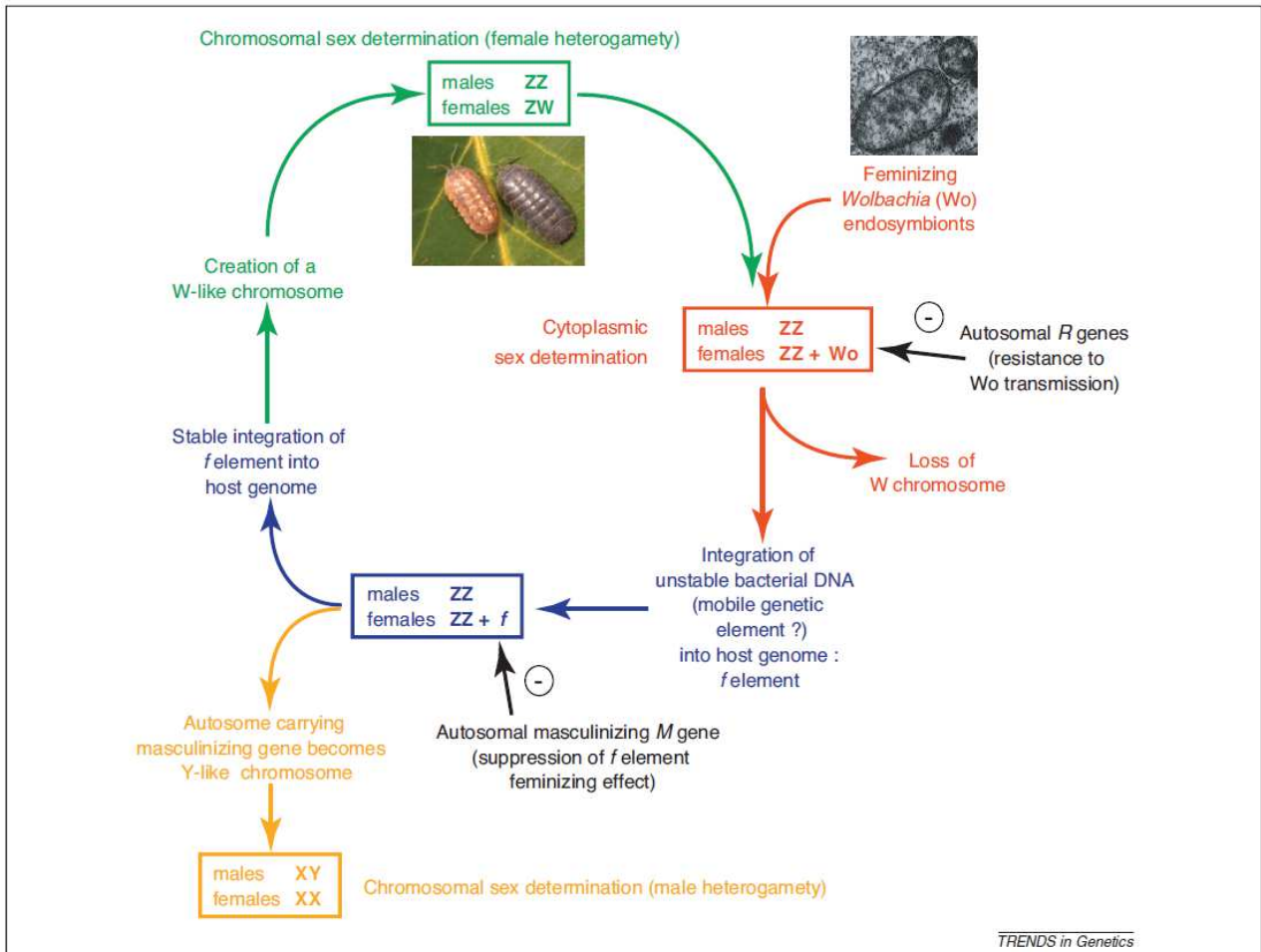


Figure 17 : Scénario évolutif du déterminisme du sexe dans l'hypothèse d'un conflit génétique entre *Wolbachia* féminisante et l'hôte *A. vulgare*. La féminisation des mâles ZZ en femelles fonctionnelles conduit à la perte du chromosome W. Les mâles féminisés ZZ envahissent progressivement la population et remplacent les femelles génétiques ZW, qui à terme disparaissent et laissent place des femelles ZZ symbiotiques. Une fois perdu, seule la présence ou l'absence de *Wolbachia* détermine le sexe de l'individu (symbiotique : femelle ; asymbiotique : mâle). Par transfert horizontal, une partie du génome de la bactérie s'insère dans celui de son hôte (élément f) en conservant son effet féminisant. Des gènes autosomiques de résistance à la transmission de la bactérie sont sélectionnés et conduisent à la disparition de la bactérie. L'élément f est possiblement réprimé par l'action d'un gène masculinisant (nommé M). A terme, l'un des deux facteurs génétiques va induire la formation d'un nouveau chromosome sexuel : Y pour le gène M masculinisant, et W pour l'élément f féminisant (d'après Cordaux et al., 2011).

La fixation de cet allèle masculinisant dans les populations n'est possible que si le coût de cet allèle est très faible/nul pour son hôte (Caubet et al., 2000). Ce gène autosomique pourrait ainsi constituer un nouveau chromosome sexuel Y, et donc permettre l'établissement d'un système hétérogamétique mâle. A l'inverse, la fixation de l'élément f donnerait naissance à un nouveau chromosome W (Juchault & Mocquard, 1993; Cordaux et al., 2011).

Ce scénario est d'ailleurs généralisable à l'ensemble des isopodes terrestres. Il faut aussi noter que l'ensemble de ce scénario contraint à un faible degré d'hétéromorphie des gonosomes, tant morphologiquement que fonctionnellement, car les changements récurrents de positions des allèles déterminant le sexe limiteraient l'accumulation de différences entre les chromosomes sexuels, ces derniers demeurant recombinants (Charlesworth, 1991, 1996). D'ailleurs, de nombreux chromosomes sexuels observés chez les espèces d'isopodes terrestres semblent être homomorphes (Juchault & Rigaud, 1995). A l'échelle de l'évolution des Oniscidea, les conflits génétiques entre le compartiment nucléaire hôte et le compartiment cytoplasmique bactérien pourraient donc avoir induit des changements récurrents des déterminants génétiques du sexe chez l'hôte. Ces déterminants seraient régulièrement relocalisés sur différentes paires d'autosomes, et généreraient constamment de nouveaux chromosomes sexuels (Juchault & Rigaud, 1995; Rigaud, 1997; Rigaud et al., 1997; Cordaux et al., 2011).

IV. Projet de thèse

Dans ce contexte où l'essentiel de nos connaissances repose sur l'étude approfondie d'une seule espèce (*A. vulgare*), mon projet de thèse vise à évaluer l'impact des bactéries féminisantes *Wolbachia* sur l'évolution des mécanismes de déterminisme du sexe chez les isopodes terrestres. Ces derniers sont connus pour présenter des systèmes d'hétérogamétie mâle (XX/XY) ou femelle (ZZ/ZW) selon l'espèce considérée (Rigaud et al., 1997). De plus, *Wolbachia* est connue pour être très largement répandue chez de nombreuses espèces d'isopodes terrestres (Bouchon et al., 1998; Cordaux et al., 2012). Ces bactéries sont susceptibles d'occasionner des conflits génétiques avec leurs hôtes, potentiellement générateurs de nouveaux mécanismes de déterminisme du sexe. Les isopodes terrestres constituent ainsi un modèle unique pour étudier les transitions entre systèmes de déterminisme du sexe. Plus globalement, ces travaux visent à mieux comprendre l'influence de l'endosymbiose sur l'évolution de la structuration et de la complexité des génomes hôtes, mais aussi à caractériser les mécanismes de transitions entre systèmes de déterminisme du sexe.

Une première stratégie a consisté à se focaliser sur une transition d'hétérogamétie entre deux espèces apparentées appartenant au même genre : *A. vulgare* (ZZ/ZW, Juchault & Legrand, 1972) et *A. nasatum* (XX/XY, Juchault & Legrand, 1964). Une seconde stratégie a consisté à caractériser les différents systèmes d'hétérogamétie rencontrés au sein d'un large panel d'isopodes terrestres afin d'évaluer plus globalement l'impact de *Wolbachia* sur l'évolution des mécanismes de déterminisme du sexe chez les Oniscidea. Ainsi, plusieurs approches combinant l'étude des génomes d'*A. vulgare* et d'*A. nasatum* (comparaisons génomiques, prédictions structurales et fonctionnelles, recherche de séquences sexe-spécifiques), l'étude de la transcription de gènes candidats du déterminisme du sexe (au cours du développement), et l'étude des systèmes de déterminisme du sexe rencontrés au sein des isopodes terrestres (approche phylo-transcriptomique, expériences de réversion de sexe) ont été réalisées.

Le premier chapitre de cette thèse traite de l'assemblage et de l'analyse structurale et fonctionnelle du génome d'*A. nasatum* (hétérogamétie mâle : XY/XX). Le but est de disposer d'un génome de référence pour permettre la

caractérisation des séquences et gènes potentiellement impliqués dans le déterminisme du sexe chez cette espèce, et aussi plus globalement de mieux évaluer le paysage génomique caractérisant *A. nasatum*. Des travaux similaires sont menés en parallèle sur le modèle *A. vulgare* (thèse menée par Amine Chebbi), permettant une comparaison fine entre ces deux génomes.

Le second chapitre de cette thèse fait le lien entre étude génomique et phénotypique dans le cadre du déterminisme du sexe. Les génomes annotés et assemblés d'*A. nasatum* et d'*A. vulgare* constituent des outils de choix pour caractériser les mutations et réarrangements génomiques impliqués dans une transition de système d'hétérogamétie. L'objectif ici est donc de caractériser des séquences sexe-spécifiques au sein du génome d'*A. nasatum*. Ces travaux ont été effectués à partir d'une approche de génomique comparative à partir de données issues du séquençage indépendant de mâles et de femelles, menant à la caractérisation de séquences Y-spécifiques chez *A. nasatum*.

Le troisième chapitre est focalisé sur la caractérisation de gènes candidats connus pour être impliqués dans les cascades de déterminisme ou de différenciation sexuelle de nombreux métazoaires. Cette caractérisation a été possible grâce à l'assemblage et à l'annotation du génome d'*A. nasatum* réalisés au cours du premier chapitre. Suite à leur identification, le suivi du profil d'expression de ces gènes candidats a été réalisé cours du développement post-embryonnaire d'*A. nasatum*.

Le quatrième et dernier chapitre de cette thèse est centré sur la diversité ainsi que sur l'évolution du déterminisme du sexe à l'échelle des Oniscidea, dans le but d'évaluer le nombre et la direction des transitions entre ces systèmes. L'objectif est d'évaluer l'impact global des *Wolbachia* féminisantes sur l'ensemble des mécanismes de déterminisme du sexe rencontrés chez les isopodes terrestres. Ces travaux se basent d'une part sur une approche phylotranscriptomique afin d'établir une phylogénie robuste et datée des isopodes terrestres, et d'autre part sur une approche originale de croisements entre des femelles génétiques et des néo-mâles phénotypiques (obtenus à partir de femelles génétiques masculinisées par l'injection de glandes androgènes au cours du développement), permettant d'établir le système d'hétérogamétie des espèces étudiées.

Chapitre 1

Séquençage et assemblage du génomme d'*A. nasatum*

Introduction

Si le premier génome d'eucaryote séquencé fut celui de la levure *Saccharomyces cerevisiae* en 1996 (Goffeau et al., 1996), il existe désormais plus de 4 600 génomes d'eucaryotes disponibles dans les banques de données publiques de NCBI (consulté le 01/08/2017). Cet essor a été permis par le développement et l'utilisation croissante du séquençage à haut débit (appelé couramment NGS pour *Next-Generation Sequencing*) depuis la dernière décennie. Ces NGS, majoritairement popularisées par la technologie *Illumina*, permettent la production d'un très grand nombre de lectures de séquençage (*reads*), le plus souvent de très petite taille (entre 50 pb et 150 pb). Cette taille limitée est en partie compensée par d'importantes profondeurs de séquençage (Salzberg et al., 2012), ce qui signifie qu'un même nucléotide est séquencé un grand nombre de fois.

Cependant, les grands génomes (de plusieurs milliards de nucléotides) sont généralement difficiles à assembler *de novo* (sans référence) à partir de telles technologies, notamment à cause de leur complexité causée par la présence parfois massive d'éléments répétés. Il s'agit d'ailleurs là d'un des enjeux majeurs des projets de séquençage de grands génomes (Wessler, 2006). En effet, ces répétitions peuvent empêcher de clore l'assemblage et donc d'assurer sa continuité : deux séquences occupant des positions différentes du génome peuvent être parfaitement identiques, l'assembleur ne peut cependant pas les distinguer et les assembler correctement (Treangen & Salzberg, 2011). Ainsi, les régions génomiques flanquant ces répétitions peuvent être assemblées de façon erronée en constituant une séquence chimérique entre deux régions chromosomiques différentes (Treangen & Salzberg, 2011). Un assemblage chimérique peut être minimisé par le séquençage de *reads* appariés (générés à partir d'un même fragment d'ADN de taille fixe, dont les deux extrémités ont été séquencées), puisque les assembleurs utilisent aussi cette information d'appariement (distance entre les deux *reads*, orientation). Cependant, si ces *reads* appariés contiennent

des séquences répétées présentes en de nombreuses copies dans le génome, leur assemblage restera compliqué (Treangen & Salzberg, 2011).

Les éléments répétés émergent d'une variété de mécanismes biologiques, qui résultent en la production d'exemplaires supplémentaires d'une même séquence insérée dans le génome (Treangen & Salzberg, 2011). Ces répétitions varient en formes et en tailles : il peut s'agir de séquences répétées dispersées, ou encore de séquences répétées en tandem. Elles peuvent être présentes en seulement deux copies ou en des millions d'exemplaires, et il peut s'agir de répétitions de quelques nucléotides comme de répétitions de plusieurs milliers de paires de bases (Treangen & Salzberg, 2011). Comme évoqué précédemment, ces séquences répétées sont le plus souvent classées selon deux catégories. Tout d'abord, les séquences courtes répétées en tandem (qui regroupes l'ADN satellite, les minisatellites et les microsatellites) sont constituées d'un même motif de nucléotides répétés adjacents les uns aux autres. Les séquences répétées dispersées, appelées aussi éléments transposables (ETs), sont quant à elles présentes dans plusieurs régions chromosomiques différentes. Ces ETs sont capables de se répliquer de façon autonome ou semi-autonome selon différents mécanismes, et ce à partir d'intermédiaires à ARN (ET de classe I, ou rétrotransposons) ou à ADN (ET de classe II, ou transposons) (Wicker et al., 2007). Les ET de classe I se répliquent selon un mécanisme communément appelé "copier-coller" (l'ET produit sa réplique qui est ensuite intégrée à un autre emplacement du génome), et se distinguent selon qu'ils possèdent de longues séquences terminales répétées (LTR pour *Long Terminal Repeats*) ou non. Les éléments non-LTR sont classiquement classés selon s'ils sont de petite taille (de 100 à 300 pb, appelés SINE pour *Short Interspersed Nuclear Element*) ou de grande taille (plusieurs kb, appelés LINE pour *Long Interspersed Nuclear Element*) (Prak & Kazazian, 2000). Les éléments de classe II transposent quant à eux selon un mécanisme de "couper-coller" (l'élément est typiquement excisé puis intégré à un autre emplacement du génome) et peuvent être de taille variable (de 100 pb à plusieurs kb) (Prak & Kazazian, 2000).

Si la composition des chromosomes, principalement en ETs, constitue un frein à leur assemblage, la nature même des chromosomes peut avoir une influence sur la fragmentation de leur assemblage. En effet, cette fragmentation

est d'autant plus accentuée dans le cas de séquences situées sur les chromosomes sexuels hétérogamétiques (Y ou W) que ces chromosomes tendent à accumuler de telles répétitions au cours de leur évolution. Pour cette raison, la majorité des projets de séquençage de génomes complets est focalisée sur l'assemblage de génomes homogamétiques (d'individus XX ou ZZ). A titre d'exemple, à peine 20% des génomes de mammifères disponibles publiquement sont ceux du sexe hétérogamétique (Tomaszkiewicz et al., 2017). Le séquençage et l'assemblage de des hétérochromosomes constituent toujours un véritable défi. En effet, la nature hémizygote de ces chromosomes complique leur assemblage simplement à cause d'une profondeur de séquençage moindre par rapport aux autosomes (Carvalho et al., 2003). La conséquence directe est que pour assembler des séquences appartenant à un hétérochromosome, il faut une profondeur de séquençage deux fois plus élevée que celle prévue pour l'assemblage de séquences homogamétiques (Tomaszkiewicz et al., 2017). De plus, des régions dites ampliconiques (issues des autosomes et fortement répétées) peuvent être présentes sur l'hétérochromosome, comme c'est le cas pour le chromosome Y des mammifères (Wei et al., 2015; Ghenu et al., 2016; Oetjens et al., 2016; Tomasziewicz et al., 2016). Ces régions aussi contribuent aux difficultés d'assemblage des hétérochromosomes (Tomaszkiewicz et al., 2017). Enfin, les chromosomes Y et W présentent souvent une structure inhabituellement riche en hétérochromatine (Carvalho et al., 2003; Skaletsky et al., 2003; Hoskins et al., 2007), souvent composée de nombreuses séquences répétées, ce qui constitue un élément compliquant fortement les assemblages génomiques *de novo* comme évoqué plus haut. L'une des principales difficultés concernant l'analyse de l'évolution des chromosomes sexuels à l'échelle du génome entier réside donc dans la qualité de l'assemblage.

Une solution pour améliorer les assemblages, et notamment ceux des séquences hétérochromosomiques, est d'augmenter considérablement la taille des *reads* produits (Tomaszkiewicz et al., 2016). C'est ce que proposent des technologies récentes, telles que celle basée sur le séquençage d'une seule molécule d'ADN en temps réel (SMRT pour *Single Molecule Real Time*). Démocratisé par la technologie *Pacific Biosciences (PacBio)*, cette méthode de séquençage repose sur la lecture en temps réel de nucléotides marqués par fluorescence issus de la réplication d'une unique molécule d'ADN, permettant la

lecture de plusieurs milliers de nucléotides en une seule fois (et donc la production de longs *reads*). Le taux d'erreur élevé de cette méthode (proche de 10%, Rhoads & Au, 2015) peut être compensé par une profondeur de séquençage élevée et/ou par l'utilisation de *reads Illumina* permettant de corriger les longs *reads* (Roberts et al., 2013).

L'objectif de cette première partie de thèse était donc de générer un génome de référence de l'espèce *A. nasatum* (XX/XY) pour permettre à terme la caractérisation des séquences et gènes potentiellement impliqués dans le déterminisme du sexe chez cette espèce (Chapitre 2, page 113), mais aussi plus globalement de mieux caractériser le paysage génomique d'*A. nasatum*. Trois phases ont été nécessaires afin de générer un assemblage de ce génome de référence.

La première phase a consisté en une première exploration du génome d'*A. nasatum* dont rien n'était connu jusqu'alors. Ainsi des données de séquençage (*reads*) *Illumina* ont été générées à partir de l'ADN de 2 mâles et de 5 femelles *A. nasatum*. A partir des données mâles, une première version de l'assemblage a été réalisée *de novo*, tandis que les données femelles ont été générées pour la recherche de séquences mâles-spécifiques (voir Chapitre 2, page 113). Une première phase de tests a été effectuée dans le but de déterminer les meilleurs paramètres d'assemblage. L'objectif était de maximiser la quantité de nucléotides assemblés afin de disposer du plus grand jeu de données possible et donc de maximiser l'information génétique présente. Les fragments assemblés de ce génome ont ensuite été associés pour générer de plus grands fragments continus (*scaffolding*). Cette association a généré l'introduction de nucléotides indéterminés, qui ont ensuite été en partie résolus dans la dernière phase de cet assemblage. Enfin, la qualité de cet assemblage *de novo* a été évaluée par l'alignement des *reads* mâles sur le génome. La forte fragmentation de l'assemblage final nous a amené à affiner notre stratégie.

La seconde phase a consisté à diminuer l'importante fragmentation du génome assemblé issu de la première étape. Des données de séquençage *PacBio* ont été générées à partir de l'ADN de 36 mâles, et ont ensuite été utilisées en association avec les données *Illumina* pour générer une nouvelle version de

l'assemblage. Puisque les données *PacBio* sont susceptibles de contenir un fort taux d'erreurs (indels ou substitutions), il a tout d'abord été nécessaire de les corriger. Une autocorrection des *reads PacBio* a été réalisée à partir de *reads* de haute fiabilité nommés *reads CCS* (*Circular Consensus Sequences*), suivi d'une correction par les *reads Illumina*. Un premier assemblage a été réalisé à partir du pipeline basé sur le logiciel *CeleraAssembler* (Berlin et al., 2015), puis un second assemblage a été réalisé à partir du logiciel *DBG2OLC* (Ye et al., 2016). Les séquences provenant potentiellement d'organismes autres qu'*A. nasatum* ont ensuite été éliminées de cette dernière version de l'assemblage. Cet assemblage a été évalué par un alignement des *reads Illumina*, suivi d'une estimation de la proportion de gènes considérés comme universels au sein des arthropodes afin d'évaluer l'exhaustivité fonctionnelle de cet assemblage.

Enfin, la troisième et dernière phase était de décrire le paysage génomique d'*A. nasatum*. Les répétitions présentes ont tout d'abord été caractérisées par homologie avec une banque de données de répétitions connues, et ont aussi été complétées par une prédiction *ab initio* des éléments répétés. Une fois identifiés, ces répétitions ont été masquées dans le génome afin de faciliter l'étape d'annotation génique. Celle-ci consistait tout d'abord à caractériser structurellement l'ensemble des gènes présents au sein de ce génome de référence, en comparant les séquences génomiques à des séquences transcriptomiques tout d'abord, puis par une approche probabiliste. Cette caractérisation structurale a été suivie d'une annotation fonctionnelle en combinant une recherche des domaines protéiques présents, complétée par une recherche par similarité avec une base de données de séquences protéiques connues. La structure du génome d'*A. nasatum* a ensuite été comparée avec celle du génome d'une espèce apparentée : *A. vulgare* (ZW,ZZ), dont le séquençage et l'assemblage ont été réalisés en parallèle au sein du laboratoire (Chebbi et al., en préparation).

Matériels et Méthodes

I. Séquençage *Illumina* et assemblage du génome mâle d'*A. nasatum*

1. Extraction d'ADN (séquençage *Illumina*)

Afin de séquencer le génome d'*A. nasatum*, l'ADN de cinq femelles et de deux mâles issus d'une même portée asymbiotique a été extrait (origine des individus : Thuré, 79, France). Pour chaque extraction, les individus ont été rincés dans trois bains consécutifs (javel 3%, et deux bains d'eau déminéralisée) afin de limiter les contaminations dues aux organismes présents sur la cuticule des animaux. Les individus ont ensuite été disséqués afin de prélever trois catégories de tissus par animal : (i) tête, antennes et pattes ; (ii) organes reproducteurs (ovaires ou testicules et utricules) ; et (iii) chaîne nerveuse. L'ADN de ces différents tissus a ensuite été extrait indépendamment. Cette méthode permet une maximisation du rendement de l'extraction ADN par individu. Dans le but d'accroître le rendement de l'extraction, ces tissus ont été broyés mécaniquement avec des pistons en plastique. L'extraction d'ADN a ensuite été effectuée avec le kit d'extraction en colonne *DNeasy Blood and Tissue* (Qiagen) en suivant le protocole fourni. Une fois les extractions d'ADN achevées, les trois échantillons d'ADN issus d'un même individu ont été réunis.

L'ADN extrait a ensuite été concentré dans le but d'atteindre les concentrations requises pour un séquençage à haut débit. En effet, un individu d'*A. nasatum* adulte (âgé d'environ un an) ne pèse qu'entre 15 et 25 mg. En suivant la méthode d'extraction décrite précédemment, la quantité d'ADN extraite par individu est d'environ 1.5 μg avec une concentration moyenne de 2.5 $\text{ng} \cdot \mu\text{L}^{-1}$ (concentration mesurée au *Qubit 2.0*, Invitrogen). Or la concentration de l'ADN extrait par individu est trop faible pour le protocole de séquençage *Illumina* que nous avons utilisé (pré-requis pour générer une banque *Illumina* : 0.5 μg d'ADN avec une concentration conseillée d'au moins 50 $\text{ng} \cdot \mu\text{L}^{-1}$) et nécessite une étape de précipitation. Ainsi, ont été ajoutés aux ADN extraits 1/10ème du volume

d'extraction d'acétate de sodium (3 M, pH 7) et 1 volume d'isopropanol froid. La précipitation de l'ADN a été facilitée en plaçant les tubes 10 minutes dans l'azote liquide. Afin de rincer ces échantillons, ces derniers ont été centrifugés 30 minutes à 15 000 g (4°C). Le surnageant a ensuite été éliminé, et 500 µL d'éthanol 70% froid ont été ajoutés. Les échantillons ont ensuite été centrifugés pendant 5 minutes à 15 000 g (4°C). Une fois le surnageant retiré, le rinçage a été répété une seconde fois. Enfin, après avoir retiré l'éthanol, les culots d'ADN ont été solubilisés dans 100 µL d'eau stérile, puis agités avec un *MixMate* (Eppendorf) durant 30 minutes à 300 g (assurant une dissolution totale de l'ADN dans l'eau).

Suite à la quantification de l'ADN extrait, sa qualité a été évaluée par migration sur gel afin de s'assurer que l'ADN n'était pas dégradé. Ainsi, 5 µL d'ADN concentré mélangé avec 10 µL de Bleu de bromophénol ont été déposés dans un gel d'agarose à 0.8%. Les échantillons ont ensuite migré durant 45 minutes à 100 V. La présence d'une bande de migration nette et de haut poids moléculaire après révélation aux UV est synonyme de très faible dégradation (et donc de bonne qualité) de l'ADN. A l'inverse, la présence d'une trainée d'ADN ("smear") est synonyme de dégradation. De plus, l'amplification des gènes mitochondriaux *coI* (Folmer et al., 1994) et *ARNr 16S* (Kocher et al., 1989) suivie de leur séquençage Sanger a permis de vérifier l'identité de l'espèce séquencée. L'absence de *Wolbachia* a aussi été vérifiée par PCR avec les amorces *wsp* et *ftsZ* (Werren et al., 1995; Braig et al., 1998).

2. Séquençage *Illumina*, contrôle qualité des lectures de séquençage et filtrage des adaptateurs de séquençage

Suite à ces contrôles qualité, deux banques d'ADN, une mâle et une femelle, ont été construites avec les kits de préparation *Nextera* à partir de 7.2 µg (les deux mâles poolés de façon équimolaire) et 9,7 µg (les cinq femelles poolées de façon équimolaire) d'ADN, respectivement. Les banques ont ensuite été séquencées en mode *paired-end* sur un séquenceur *Illumina HiSeq2000*, ce qui correspond au séquençage des deux extrémités de fragments courts (*inserts*) générant des lectures de séquençage (*reads*) de 100 pb (Beckman Coulter

Genomics). Le détail des données brutes séquencées est indiqué dans le Tableau 2. Le contrôle qualité des *reads* a été réalisé à l'aide de *FastQC* (version v0.11.5, www.bioinformatics.babraham.ac.uk/projects/fastqc/). Les extrémités des *reads* présentant un mauvais score de qualité *phred* (< 20) ont été tronquées avec *FastX Toolkit* (version 0.0.13). Les adaptateurs de séquençage *Illumina*, potentiellement séquencés en même temps que les *inserts*, ont été supprimés avec *Cutadapt* (version 1.2.1). Enfin, seuls les *reads* appariés ont été conservés pour l'assemblage génomique.

Tableau 2 : Statistiques sur les banques de données séquencées.

Technologie de séquençage	<i>Illumina Hiseq2000</i> - 2×100 pb - <i>paired end</i>		<i>PacBio RS II</i>
Sexe des individus séquencés	Mâles XY	Femelles XX	Mâles XY
Origine des individus séquencés	Thuré (79, France)	Thuré (79, France)	Thuré (79, France)
Nombre d'individus poolés	2	5	36
Centre de séquençage	Beckman Coulter Genomics	Beckman Coulter Genomics	Genome Québec
Nombre de bibliothèques construites	1	1	5
Nombre d'unités de séquençage	2 lignes	2 lignes	50 cellules
Nombre de <i>reads</i> bruts	657 263 842	651 769 122	5 962 613
Nombre de bases brutes	65 726 384 200	65 176 912 200	43 152 605 152

3. Assemblage des données *Illumina*

Les *reads* conservés après le contrôle qualité ont été assemblés avec *SOAPdenovo* (version 2.04, Luo et al., 2012). Cet assembleur *de novo* (capable d'assembler des données en l'absence de génome de référence) se base sur l'utilisation d'un *k-mer* (correspondant à toutes les sous-séquences de longueur *k* présentes au sein de *reads*; A, Figure 18) fixé pour réaliser les graphes de *de Bruijn* (qui correspondent à des graphes orientés représentant les chevauchements de longueur $k - 1$ entre les *k-mers* de longueur *k*) permettant l'assemblage des *reads* en *contigs* (B, Figure 18).

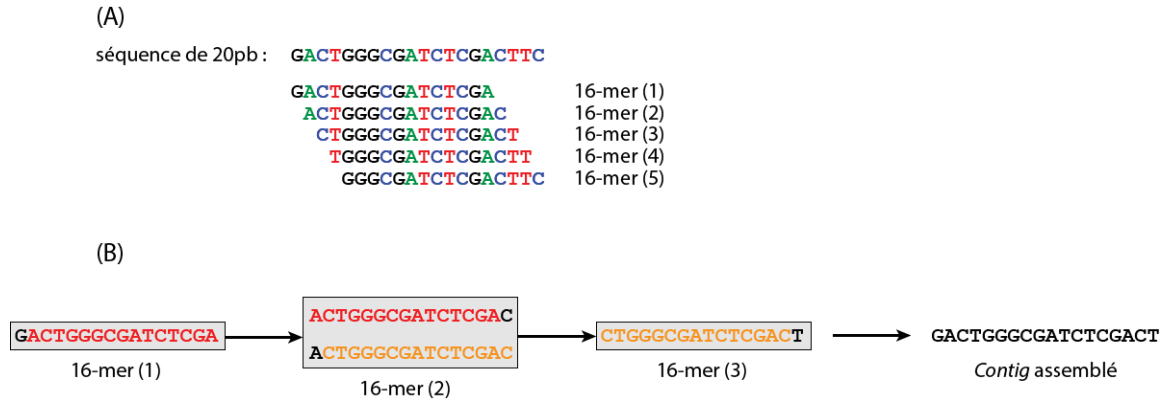


Figure 18 : Exemple de différents *k*-mers présents au sein d'une séquence. (A) Tous les *k*-mers de 16 pb de longueur (appelés 16-mers) présents au sein d'une séquence de 20 pb sont représentés (nombre de *k*-mers au sein d'une séquence = taille de la séquence (pb) - taille du *k*-mer (pb) + 1). (B) Graphe de *de Bruijn* réalisé à partir des 16-mers (1), (2) et (3) permettant la formation d'un *contig*. Les séquences d'une même couleur (de longueur : taille du *k*-mer - 1 = 15 pb) entre deux *k*-mers sont identiques (taille du *contig* assemblé = nombre de *k*-mers inclus dans le graphe de *de Bruijn* + taille des *kmers* - 1).

Un *contig* correspond à un assemblage de *reads* par similarité de séquences chevauchantes, et ne prend pas en compte l'information donnée par les *reads paired-end*. Le logiciel d'assemblage permet ensuite de relier les *contigs* formés en *scaffolds*, qui résultent de la connexion de *contigs* qui contiennent des *reads* appariés assemblés sur deux *contigs* différents. Le *scaffold* ainsi formé aura donc une longueur connue, mais contiendra potentiellement des zones contenant des nucléotides indéterminés (Figure 19).

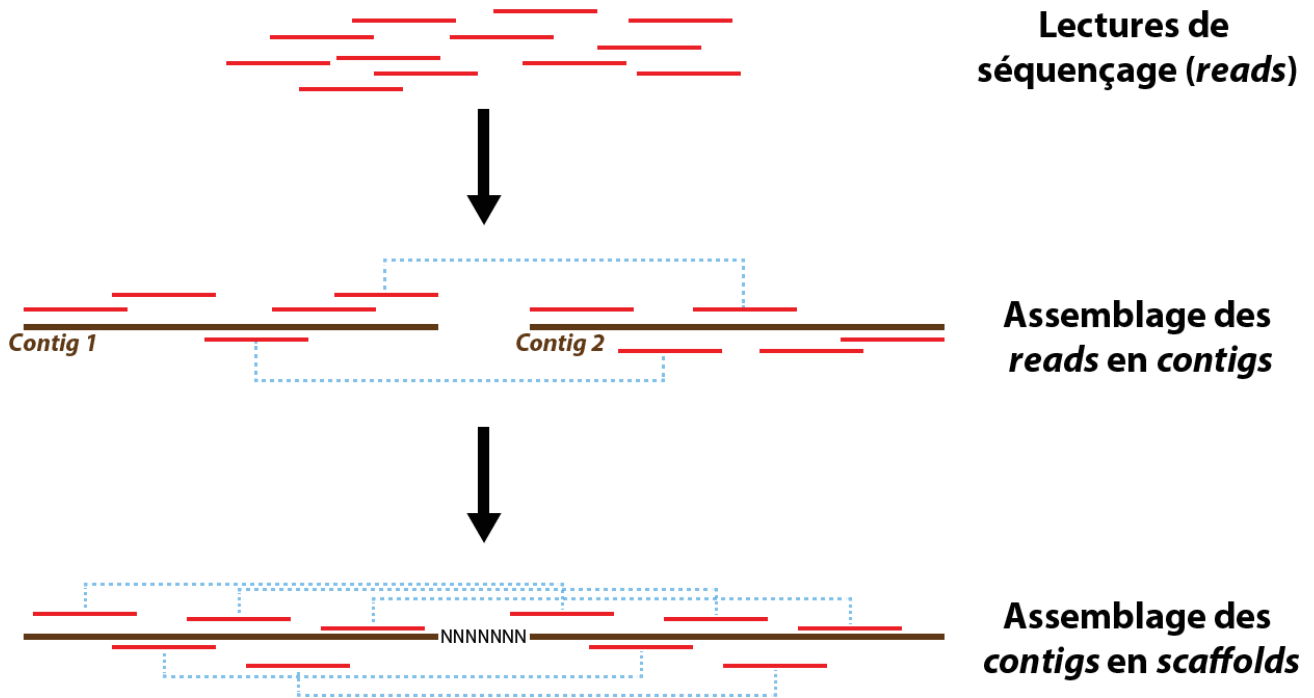


Figure 19 : Représentation de la méthode utilisée pour assembler les *reads* *Illumina* en *contigs*, puis ces *contigs* en *scaffolds*. Tous les *reads* sont initialement assemblés en *contigs* par l'intermédiaire de la méthode des graphes de *de Bruijn* sans tenir compte de l'information d'appariement (représentée par des pointillés bleus). Un *contig* constitue une séquence assemblée connue sur toute sa longueur. Si possible, ces *contigs* sont ensuite assemblés en séquences plus longues, appelées *scaffolds*, lorsque des *reads* appariés sont contenus dans deux *contigs* différents. L'assemblage de ces *scaffolds* permet de créer des séquences de plus longue taille, mais pouvant contenir des lacunes correspondant à des séquences inconnues (notées N). Les *reads* sont affichés en rouge, tandis que les *contigs* assemblés sont colorés en brun.

Ainsi, des assemblages utilisant des valeurs de *k-mer* fixées à 41, 51, 61, 71 et 81 ont été réalisés, dans le but d'identifier l'assemblage maximisant la quantité finale de données assemblées. Les *scaffolds* formés au cours de l'assemblage ont causé l'insertion de nucléotides indéterminés. Afin de combler ces séquences avec des nucléotides connus, une étape supplémentaire a consisté à combler ces lacunes (étape de *Gapfilling*) avec *GapFiller* (Boetzer & Pirovano, 2012). La méthode consiste ici à repérer les paires de *reads* pour lesquelles un des deux *reads* correspond à une région assemblée et l'autre *read* correspond partiellement à la région indéterminée. Ce *read* est alors utilisé pour combler la lacune par chevauchement de *k-mer* avec la région assemblée connue. Ce processus est répété de façon itérative jusqu'à ce qu'aucune lacune ne puisse plus être comblée.

II. Séquençage *Pacific Biosciences* et assemblages du génome mâle d'*A. nasatum*

1. Extraction, concentration et contrôles qualité de l'ADN (séquençage *PacBio*)

Le protocole d'extraction d'ADN a été le même que celui utilisé pour le séquençage *Illumina*. Cependant, les conditions de séquençage de la technologie de *Pacific Biosciences* nécessitaient d'extraire une quantité bien plus élevée d'ADN. Ainsi, des extractions ont été réalisées sur 36 mâles (3 pools de 12 individus extraits). Ils étaient issus de croisements entre des frères et des sœurs des individus utilisés pour le séquençage *Illumina*, limitant ainsi la variabilité génétique entre les individus séquencés précédemment et les individus utilisés pour le séquençage *PacBio* (Figure 20).

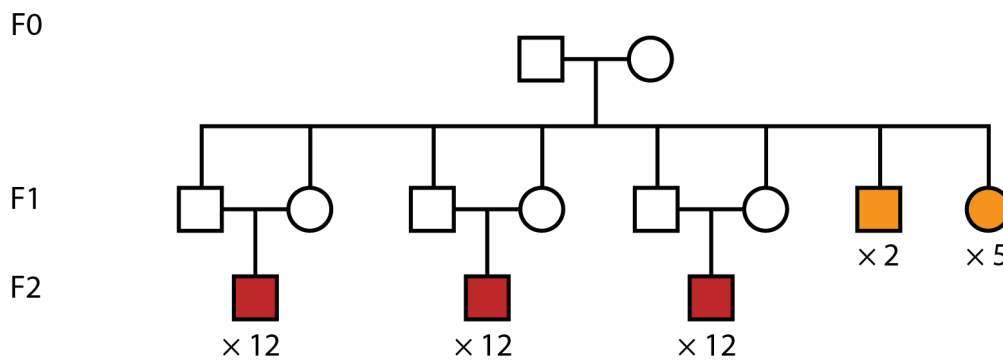


Figure 20 : Généalogie des individus *A. nasatum* séquencés. Les carrés représentent des mâles, et les cercles correspondent aux femelles. Les individus colorés en orange correspondent aux deux mâles et aux cinq femelles utilisés pour le séquençage *Illumina*. Les mâles colorés en rouge correspondent aux individus utilisés pour le séquençage *Pacific Biosciences*.

L'ADN a ensuite été concentré avec le kit de purification *Agencourt AMPure XP* (Beckman Coulter Genomics) afin de limiter au maximum la dégradation de l'ADN tout en augmentant sa concentration. Les pré-requis *Pacific Biosciences* étaient de 10 µg d'ADN par banque, avec une concentration conseillée d'au moins 100 ng.µL⁻¹. Le kit de purification contient des billes magnétiques pour retenir l'ADN, et permet donc de s'affranchir d'étapes de centrifugation pouvant détériorer l'ADN extrait. Afin d'atteindre les concentrations requises pour le séquençage, l'étape de concentration a été effectuée deux fois consécutivement. Les contrôles qualité ont été les mêmes que

ceux réalisés pour le séquençage *Illumina*, à la différence que la concentration d'ADN a été mesurée avec un *Qubit 2.0* (Invitrogen).

2. Séquençage *Pacific Biosciences*, correction des *reads* et premier assemblage avec *CeleraAssembler*

L'utilisation de l'ADN de 48 mâles a permis de générer 50 cellules SMRT séquencées grâce à la chimie P6-C4 de *Pacific Biosciences* (*PacBio* RSII, Génome Québec). Les statistiques des données brutes séquencées sont indiquées dans le Tableau 2. Les données *PacBio* présentant un fort taux d'erreurs de séquençage (autour de 10%, Rhoads & Au, 2015), il était nécessaire de procéder à une étape de correction des *reads* bruts. Cette correction a été effectuée avec *LoRDEC* (version 0.4, Salmela & Rivals, 2014), qui construit des graphes de *de Bruijn* à partir de *reads Illumina* et d'un sous-ensemble des *reads PacBio* de haute fiabilité appelées *reads CCS* (*Circular Consensus Sequence*, Figure 21) afin de corriger les *reads PacBio* non *CCS*. Ainsi ce logiciel a été utilisé avec les *reads* des individus mâles séquencés en *Illumina* avec une taille de *k-mer* de 23 pb combinés aux *reads CCS*. L'ensemble des *reads* corrigés a été utilisé pour l'assemblage du génome à l'aide de *CeleraAssembler* (version 8.2) en suivant le pipeline spécifique aux *reads PacBio* mis au point par Berlin et al. (2015).

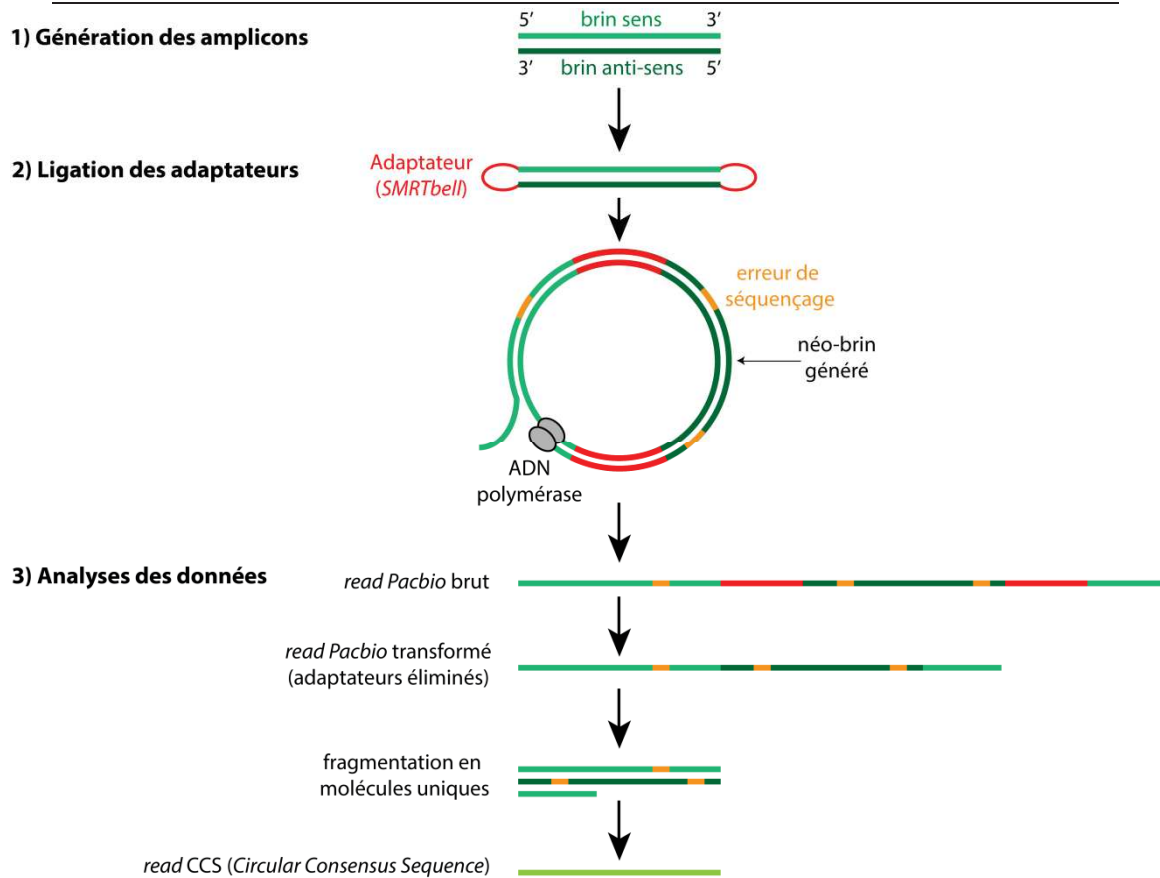


Figure 21 : Description des *reads* CCS (Circular Consensus Sequence) générés par le séquençage *PacBio*. L'ADN-polymérase réplique le brin circularisé, et produit un *read* répété de grande taille. Les adaptateurs sont ensuite éliminés par traitement bioinformatique et les fragments de molécules uniques sont alignés permettant d'éliminer les erreurs ponctuelles de séquençage présentes sur les différents fragments (indiquées en orange), afin de générer une séquence consensuelle de haute fidélité (appelée *read CCS*).

3. Assemblage du génome de référence d'*A. nasatum* avec *DBG2OLC*

Un nouveau pipeline d'assemblage *PacBio*, plus fiable concernant la prise en compte des erreurs de séquençage et plus performant en termes de temps de calcul, a été publié peu de temps après celui proposé par Berlin et al. (2015). Ce pipeline utilise principalement *DBG2OLC* (Ye et al., 2016) qui permet de générer un assemblage hybride à partir de *reads* de petite taille à haute fiabilité (*Illumina*) et de *reads* de grande taille présentant un taux d'erreur important (*PacBio*). Pour cet assemblage, seuls les *reads PacBio* corrigés (cf paragraphe précédent) d'au moins 5 000 pb ont été conservés, ce qui correspondait à 3 430 398 *reads* (57.5% de l'ensemble des *reads*) et à 36 468 068 637 pb (84.5% du nombre total de nucléotides séquencés).

La première étape de ce pipeline consiste à construire un pré-assemblage uniquement à partir des *reads* *Illumina* afin de produire des *contigs* de haute fiabilité, sans étape de construction de *scaffolds*. Cette première étape est réalisée par *SparseAssembler* (C. Ye et al., 2012), qui procède à un assemblage basé sur des graphes de *de Bruijn*. Les *contigs* de haute fiabilité ainsi générés, appelés *unitigs* (car assemblés à partir de *reads* chevauchants sans ambiguïté possible), sont ensuite alignés sur les *reads* *PacBio* (A, Figure 22). L'ensemble formé par les *reads* *PacBio* et les *unitigs* est appelé *reads* compressés (*compressed reads*). Habituellement, ces multiples alignements de séquences permettent d'éliminer les *reads* présentant des erreurs structurales (appelées "chimères"). Cependant, pour éviter l'élimination de *reads* causée par des séquences fortement répétées, cette étape n'a pas été effectuée (option - RemoveChimera 0).

A partir des *reads* compressés, *DBG2OLC* procède ensuite à la construction des *contigs* en maximisant le recouvrement des *reads* compressés (B, Figure 22). Afin de maximiser la précision de ces chevauchements, leur longueur minimale devait être de 150 pb (paramètre -MinOverlap 150). Ces *contigs* ont ensuite été associés pour former l'ossature (*backbone assembly*) à la base de l'assemblage final (C, Figure 22). Enfin, un nouvel alignement des *reads* sur l'ossature précédemment formée a été effectué afin de calculer la séquence consensuelle la plus probable (D, Figure 22). Cette étape a été réalisée à l'aide du module *Sparc* (Ye & Ma, 2016).

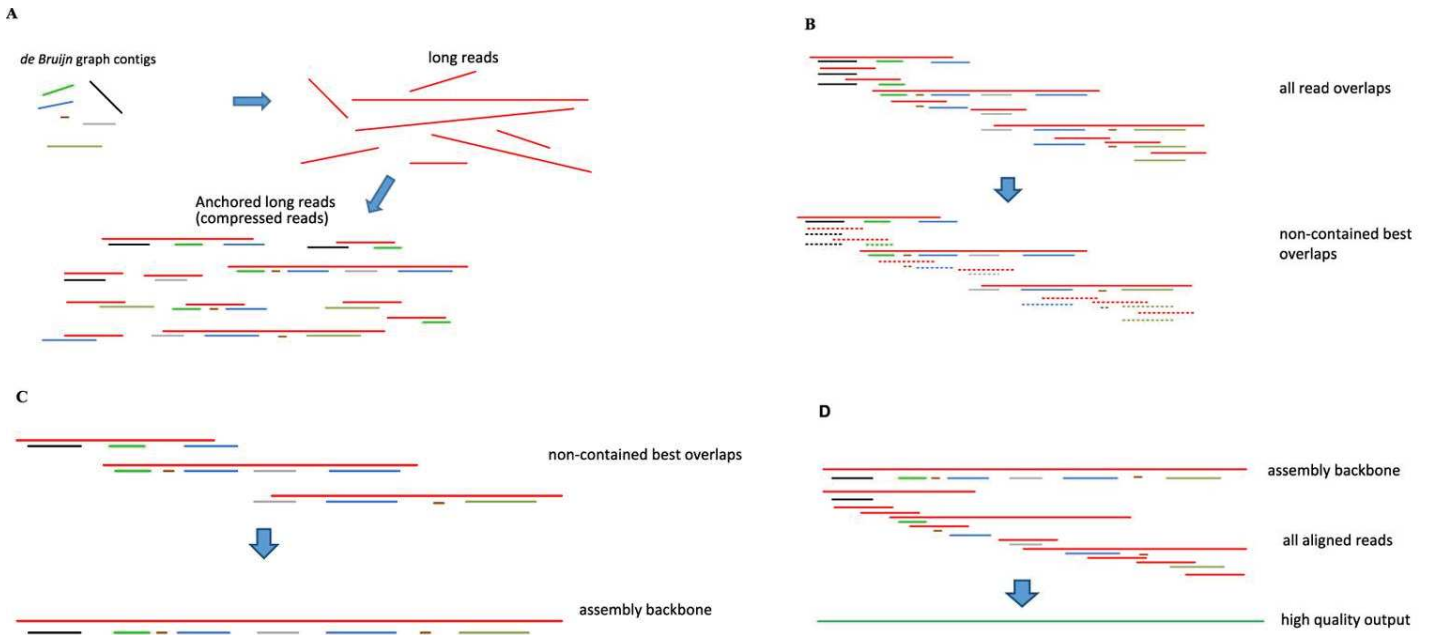


Figure 22 : Principe de l'assemblage par DBG2OLC. (A) Alignement des *unitigs* (de différentes couleurs sauf rouge) sur les *reads PacBio* (en rouge), formant les *reads compressés* (*compressed reads*). (B) Calcul du recouvrement entre les *reads compressés* et association de ceux-ci en maximisant leur recouvrement. (C) Construction de l'ossature de l'assemblage (*backbone assembly*). (D) Alignement des *reads* ayant formé l'ossature et calcul de la séquence consensuelle la plus probable (Ye et al., 2016).

Une étape de polissage (*polishing*) du génome a ensuite été réalisée à partir des *reads Illumina* avec le logiciel *Pilon* (version 1.18, Walker et al., 2014). Cette étape permet l'élimination des erreurs de séquençage (indels ou substitutions) inhérentes à la technologie *PacBio*. Enfin une étape supplémentaire d'amélioration de la qualité de l'assemblage a été réalisée grâce au transcriptome d'*A. nasatum*, afin de construire des *scaffolds*.

Le transcriptome d'*A. nasatum* a été assemblé à partir des données générées par Romiguier et al. (2014), en utilisant l'ARN de deux individus. Il contient 15 552 480 *reads* de 100 pb séquencés en *single-end* sur une plateforme Hiseq2000 (*Illumina*) (Romiguier et al., 2014). Les *reads* de faible qualité ont été éliminés avec *Trimmomatic* (version 0.33, Bolger et al., 2014), de sorte que seuls les *reads* avec un score de qualité *phred* supérieur à 15 le long d'une fenêtre de 4 nucléotides ont été conservés. Les queues poly-A ont aussi été éliminées sur les 5 dernières positions de chaque *reads* avec *PRINSEQ lite* (version 0.20.4, Schmieder & Edwards, 2011). Enfin, l'assemblage *de novo* de ces *reads* filtrés a été effectué avec le pipeline *Trinity*, en utilisant l'option adaptée aux données non-appariées en orientation sens, c'est-à-dire de 5' vers 3' (option "- SS_lib_type F").

Le transcriptome assemblé a été aligné contre le génome post-polissage à l'aide du logiciel *BLAT* (version 36x2; Kent, 2002) afin de construire des *scaffolds*. Seuls les alignements hautement fiables (% d'identité de l'alignement $\geq$ 98% de la longueur des séquences du transcriptome) ont été conservés et exploités par le logiciel *L_RNA_scaffolder* (Xue et al., 2013) pour procéder à la formation de *scaffolds* génomiques.

4. Elimination des contaminants et vérification de l'intégrité fonctionnelle du génome de référence

Les séquences potentiellement contaminantes ont été enlevées de l'assemblage, puisque des organismes pouvaient être encore présents au sein de l'hémolymphe ou sur la cuticule de l'animal lors de l'extraction d'ADN (malgré le rinçage à la javel). Le matériel génétique de ces organismes aurait donc pu être séquencé en même temps que l'ADN d'*A. nasatum*. Une approche par similarité de séquences a été employée pour éliminer ces contaminants. Les séquences assemblées ont été comparées à la base de données non-redondante d'acides nucléiques de NCBI (version de Juillet 2016) avec *BLASTn* (version 2.2.30+, Camacho et al., 2009), ainsi qu'à la base de données protéiques UniRef90 (version de Septembre 2016) avec *diamond* (version 0.7.12, Buchfink et al., 2015) en suivant le mode *BLASTx* (recherche de similarité entre des nucléotides traduits et une base de données protéiques). Pour les deux approches, un alignement a été considéré significatif (*hit*) si sa *e-value* était inférieure ou égale à 10^{-25} . De plus, l'assignation taxonomique de chaque séquence assemblée a été réalisée en effectuant la somme des scores des *hits* d'un même rang taxonomique pour une même séquence assemblée (en utilisant les bases de données nucléiques et protéiques).

La profondeur de séquençage (qui correspond au nombre de *reads* incluant un nucléotide donné au sein de séquences assemblées) a été estimée par mapping avec *Bowtie2* (version 2.2.9, Langmead & Salzberg, 2012). L'option "*very-sensitive*" a été choisie, qui correspond à un alignement strict faisant intervenir l'ensemble des nucléotides présents au sein d'une *read* (dit "*end-to-end*", par opposition à un alignement local, Figure 23).

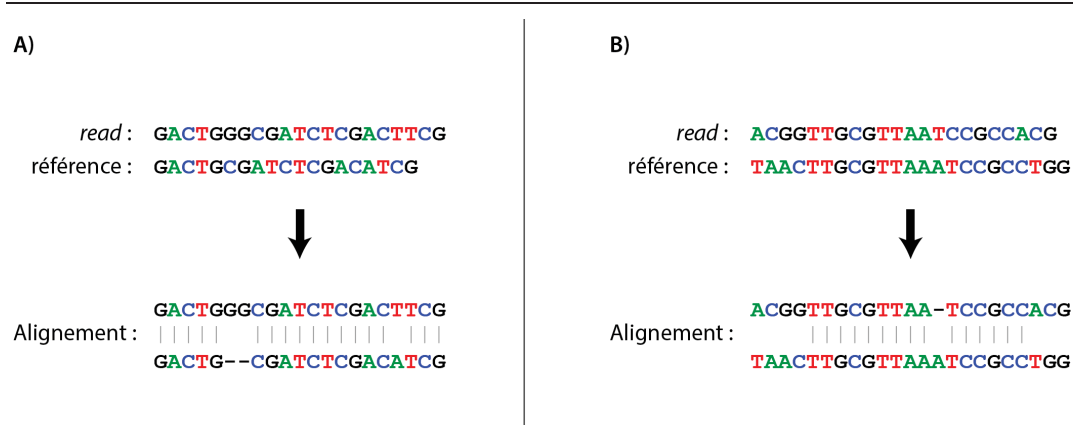


Figure 23 : Illustration des catégories d'alignement proposées par Bowtie2. A) Exemple d'un alignement *end-to-end*. Cet alignement prend en compte l'ensemble des nucléotides présents au sein du *read*. B) Exemple d'un alignement local. Cet alignement ne prend pas nécessairement en compte les nucléotides situés aux extrémités du *read*. Dans l'exemple affiché, 4 nucléotides au début du *read* et 3 nucléotides à la fin du *read* ne sont pas pris en compte pour l'alignement.

Les séquences assemblées ont ensuite été visualisées avec le logiciel *Blobtools* (version 0.9.17, Kumar et al., 2013), qui permet d'associer la couverture moyenne de chaque contig en fonction de son taux de GC tout en affichant l'assignation taxonomique de chaque *contig/scaffold* (graphiques *Taxa-annotated-GC-Coverage*). Toutes les séquences assignées aux métazoaires ont été conservées, en supposant qu'elles appartenaient probablement au génome d'*A. nasatum*. Toutes les séquences de procaryotes ont été éliminées (exceptées celles assignées à *Wolbachia* puisqu'elles peuvent correspondre à des séquences intégrées au génome d'*A. nasatum*). Enfin, les séquences assemblées n'étant assignées à aucun taxa connu ont aussi été conservées puisque l'absence de *hit* significatif peut s'expliquer par l'absence de séquences appartenant à des espèces proches dans les bases de données examinées.

L'intégrité fonctionnelle du génome a été évaluée par une recherche par similarité de séquences avec une base de données de gènes fortement conservés à l'échelle du Vivant. Cette recherche a été effectuée avec *BUSCO* (pour *Benchmarking Universal Single Copy Orthologs*) en utilisant l'option "genome" (version 3.0.1, Simão et al., 2015). Le génome a été comparé à une base de données de 1066 gènes d'arthropodes, généralement présents en copie unique et considérés comme à la fois quasi-universels et indispensables à la survie des arthropodes (appelés "*core genes*"). La sensibilité de recherche de ces gènes a été optimisée avec l'option "--long" du module *Augustus* (implémenté dans *BUSCO*), afin de maximiser le nombre de gènes détectés au sein du génome d'*A. nasatum*.

5. Caractérisation des éléments répétés et annotation structurale et fonctionnelle du génome

L'annotation des éléments répétés au sein du génome a tout d'abord été réalisée *de novo* avec *RepeatModeler* (version open-1.0.8, Tarailo-Graovac & Chen, 2009), à partir d'une base de données d'éléments répétés issue de la banque d'éléments répétés *Repbase* (version 21.08, Jurka et al., 2005). Afin d'identifier et de classer les éléments répétés, le pipeline *RepeatModeler* nécessitait l'utilisation des outils *RECON* (Bao & Eddy, 2002), *RepeatScout* (Price et al., 2005) et *Tandem Repeats Finder* (Benson, 1999). Les éléments répétés n'ayant pas été assignés à des éléments connus par *RepeatModeler* ont été classés selon leur mécanisme de transposition par le logiciel *TEclass* (version 2.1.3, Abrusán et al., 2009). Ils ont ainsi été répartis dans les catégories suivantes : transposons à ADN, rétrotransposons à LTR (séquences terminales longues inversées), longs éléments nucléaires intercalés (LINEs) et petits éléments nucléaires intercalés (SINEs).

Cependant, les gènes situés à proximité d'ETs pouvaient être considérés eux-mêmes comme des ETs, et donc générer des faux-positifs dans la détection d'éléments répétés. Ainsi, la base de données UniProt-Swiss-Prot (version de Septembre 2016) a été modifiée de sorte qu'elle ne contienne plus aucune séquence protéine d'ETs. Afin d'éliminer les transposases de la base de données UniProt-Swiss-Prot, celle-ci a été comparée aux bibliothèques *RepeatPeps* (implémentée dans *RepeatMasker*, version open-4.0.6, (Tarailo-Graovac & Chen, 2009) et *TELibrary* (issue du package *TESeeker*, Kennedy et al., 2011) par *BLASTx* (version 2.2.29+, Camacho et al., 2009). Les *hits* de *BLASTx* ont été considérés comme significatifs lorsque leur *e-value* était inférieure ou égale à 10^{-10} . La base de données modifiée a été comparée par *BLASTx* (version 2.2.29+, Camacho et al., 2009) aux ETs détectés afin d'éliminer les gènes non ETs présents et de réduire le nombre de faux-positifs (ces séquences ont été éliminées ainsi que les 50 pb en amont et 50 pb en aval du *hit* à l'aide du script *ProtExcluder1.2*, disponible à l'adresse : <http://www.hrt.msu.edu/uploads/535/78637/ProtExcluder1.2.tar.gz>). Les répétitions identifiées par *RepeatMasker* n'ont été conservées que si elles étaient plus longues que 50 pb. Pour chaque famille d'ETs, la distance génétique entre l'ET identifié et la séquence consensus représentative de la famille a été calculée

en corrigeant les forts taux de mutation causés par les dinucléotides CpG (Cytosine-Phosphate-Guanine) avec le modèle de substitution nucléotidique "Kimura 2-parameters" (Kimura, 1980).

L'annotation du génome de référence a été réalisée avec le pipeline *MAKER* (version 2.31.8, Holt & Yandell, 2011). La première étape consistait à caractériser la structure des gènes présents au sein du génome. Afin de faciliter cette étape, tous les éléments répétés (incluant les éléments ETs) ont été masqués avec les logiciels *RepeatMasker* et *RepeatRunner* (tous les deux inclus dans la suite *MAKER*). La présence de gènes a tout d'abord été prédite à partir d'une approche basée sur l'homologie entre le génome et les données transcriptomiques issues de l'assemblage du transcriptome d'*A. nasatum*. Cette première identification ensuite a été complétée par une comparaison entre le génome et la base de données Uniprot_Sprot (version de Septembre 2016). L'identification de sites d'épissage par le logiciel *Exonerate* (Slater & Birney, 2005) a ensuite permis de générer une première version du modèle décrivant la structure des gènes. Enfin, une prédiction de gènes *ab initio* avec le logiciel *SNAP* (Korf, 2004) a été réalisée. Cette méthode de détection probabiliste se base sur deux propriétés permettant de détecter un gène : (i) la détection de signaux caractéristiques de la présence de gènes (comme les régions promotrices ou les codons stop), (ii) l'analyse statistique de la composition en nucléotides des séquences (car les exons, introns ou séquences intergéniques notamment peuvent présenter des biais de composition nucléotidique dus à des pressions de sélection différentes selon la nature des séquences). Basé sur des modèles de Markov cachés (HMM, pour *Hidden Markov Model*), le logiciel *SNAP* génère des modèles de gènes qui sont ensuite évalués par *MAKER* par une métrique appelée *Annotation Edit Distance* (*AED*). Seuls les 1 000 meilleurs modèles de gènes ont ainsi été conservés. L'ensemble du pipeline *MAKER* est résumé dans la Figure 24.

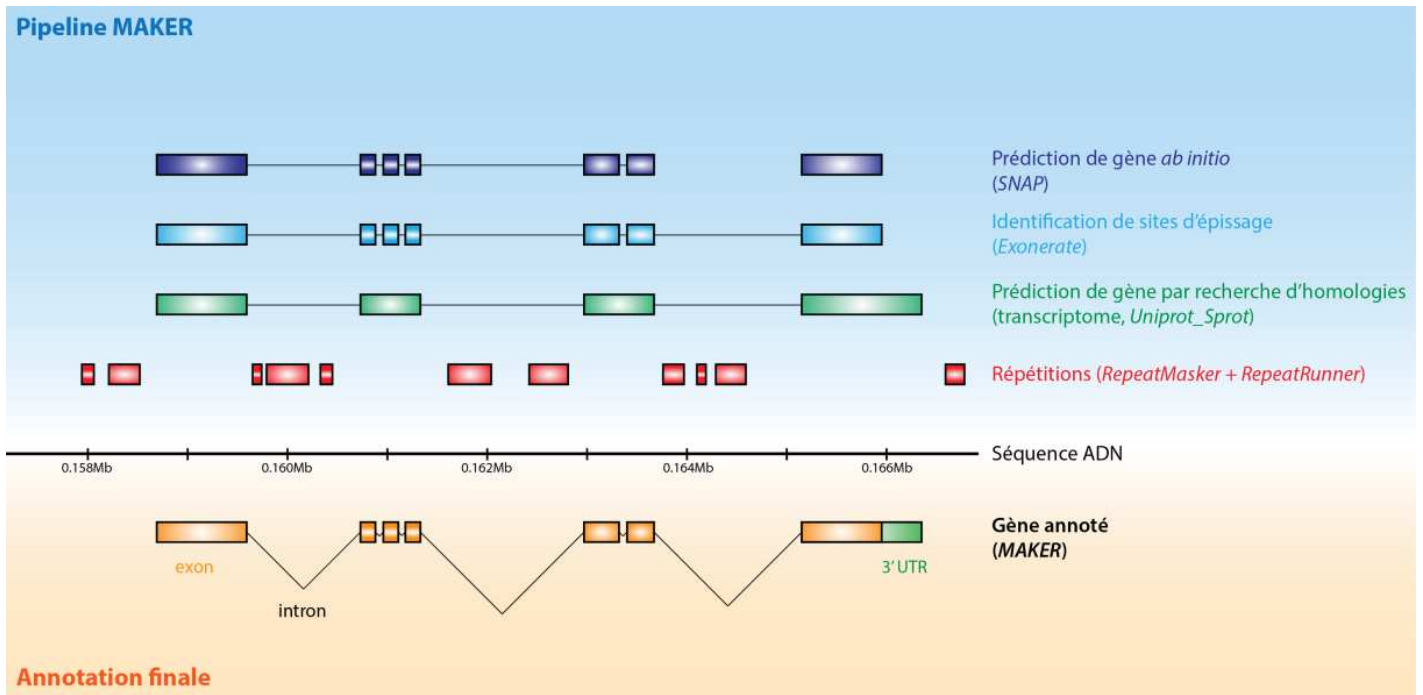


Figure 24 : Principe du pipeline d'annotation structurale *MAKER*. La séquence d'ADN analysée est représentée en noir. La première étape consiste à masquer les zones répétées (en rouge) afin de faciliter les analyses prédictives. La prédiction par homologie de la structure des gènes est effectuée à partir de données transcriptomiques et d'une base de données protéiques (étape indiquée en vert). Cette prédiction est affinée par la détection de sites d'épissage par le logiciel *Exonerate* (structuration déduite indiquée en bleu clair). Ces détections sont complétées par des analyses probabilistes *ab initio* par le logiciel *SNAP* (étape indiquée en violet). La combinaison de ces différentes analyses permet d'obtenir une structuration consensuelle, qui constitue l'annotation finale du gène, effectuée par *MAKER* (partie inférieure du schéma).

Les gènes ainsi prédits par l'annotation structurale ont été annotés fonctionnellement (une fonction leur a donc été attribuée) par la combinaison de deux approches. Tout d'abord, les domaines protéiques ont été caractérisés avec le logiciel *InterProScan* (version 5.21-60.0, Quevillon et al., 2005) à partir de la base de données PfamA (version 30.0 datée du 01/07/2016). L'option "-goterms" a été utilisée afin de décrire les fonctions moléculaires des gènes identifiés, leur compartiment cellulaire d'expression et les grands processus biologiques auxquels ils appartiennent. Ceci se traduit par une attribution de termes GO (*Gene Ontology*) pour chaque gène. Enfin, l'annotation a été complétée par une recherche par similarité de séquences protéiques grâce à *BLASTp* (version 2.2.30+, Camacho et al., 2009) en comparant les gènes prédits avec la base de données UniProt-Swiss-Prot (version de Septembre 2016).

Afin de procéder à une comparaison de l'annotation du génome d'*A. nasatum* avec celle d'*A. vulgare*, les relations d'orthologie entre ces gènes ont été établies avec le pipeline *OrthoMCL* (Li et al., 2003). Ce pipeline est basé des *BLASTp* réciproques entre les jeux de données (ici l'ensemble des protéines

prédites dans les génomes d'*A. nasatum* et d'*A. vulgare*). La *e-value* seuil a été fixée à 10^{-5} .

6. Caractérisation des insertions de *Wolbachia* dans le génome d'*A. nasatum*

Les insertions du génome de *Wolbachia* dans le génome d'*A. nasatum* ont été caractérisées par l'intermédiaire d'une approche par similarité de séquence. Tout d'abord, l'ensemble du génome a été comparé avec *BLASTn* (version 2.4.0+, Camacho et al., 2009) à une base de données composée de l'ensemble des génomes de *Wolbachia* disponibles dans les banques de données publiques (35 génomes, les numéros d'accès sont indiqués dans l'Annexe I) à laquelle a été ajouté le génome de la souche *wVulC* (connu pour infecter naturellement *A. vulgare*, Cordaux et al., 2004). Seuls les *hits* d'au moins 100 pb ont été retenus. Les séquences correspondant à ces *hits* ainsi que les 50 pb flanquant ces régions ont été extraites et comparées par *BLASTn* (version 2.4.0+, Camacho et al., 2009) à une base de données comprenant l'ensemble des génomes procaryotes disponibles sur NCBI considérés comme représentatifs ou de référence (soient 88 677 génomes, téléchargés à partir de l'adresse suivante : <ftp://ftp.ncbi.nlm.nih.gov/genomes/genbank/bacteria/>), les 36 génomes de *Wolbachia* précédemment utilisés, et une base de données de génomes mitochondriaux (disponible à l'adresse suivante : <https://ftp.ncbi.nih.gov/refseq/release/mitochondrion/>). Seuls les *hits* correspondant à des séquences de *Wolbachia* ont été conservés, et les séquences du génome d'*A. nasatum* correspondant à ces *hits* ont ensuite été conservées et considérées comme étant originaires du génome de *Wolbachia*.

Une recherche parallèle a été effectuée en comparant l'ensemble des protéines prédites par l'annotation du génome d'*A. nasatum* à la base de données protéique non redondante de NCBI, par l'intermédiaire du logiciel *diamond* (version 0.7.12, Buchfink et al., 2015). Toutes les protéines présentant un *hit* pour une séquence ayant pour origine *Wolbachia* ont été dans un premier temps retenues. Ces séquences ont ensuite été comparées par *BLASTp* (version 2.4.0+, Camacho et al., 2009) contre la base de données protéiques non redondante de NCBI avec une *e-value* seuil de 10^{-6} .

La redondance de l'ensemble des séquences sélectionnées a été éliminée avec le logiciel *R* (version 3.3.0, R Core Team, 2014), et a été confirmée après visualisation des *contigs* porteurs des insertions de *Wolbachia* avec le logiciel *Genomeview* (version 2250, Abeel et al., 2012). Les séquences nucléotidiques ont ensuite été alignées par *MAFFT* (version v6.953b, Katoh et al., 2002) avec les 36 génomes de *Wolbachia* utilisés précédemment. Les analyses phylogénétiques des inserts de *Wolbachia* ont ensuite été réalisées indépendamment pour chaque insertion caractérisée, et ont été effectuée avec le logiciel *MEGA* (version Tamura et al., 2013) en suivant la méthode du *Neighbor-joining* avec les paramètres par défaut. La robustesse des différentes topologies obtenues a été évaluée à partir de 1 000 tests de *bootstraps*.

Résultats

I. Assemblage du génome mâle d'*A. nasatum* à partir des données *Illumina*

Avant de procéder à l'assemblage *de novo* du génome mâle d'*A. nasatum*, les *reads* ont été filtrés en suivant un critère de qualité (les *reads* ont été éliminés si leur score de qualité *phred* < 20), et les adaptateurs de séquençage ont été retirés. Ce premier filtre a éliminé plus de 5% des *reads* (sens + anti-sens; Tableau 3). Afin de faciliter l'assemblage, seuls les *reads* encore appariés ont été conservés, soit près de 93.16% de l'ensemble des *reads* (Tableau 3).

Tableau 3 : Statistiques des *reads Illumina* conservés pour l'assemblage *de novo*.

	Nombre de <i>reads</i> (sens + anti-sens)	% du nombre de <i>reads</i> total
Nombre total de <i>reads</i> séquencés	657 263 842	100.00
Nombre de <i>reads</i> après élimination des adaptateurs + <i>reads</i> de faible qualité	620 732 629	94.44
Nombre de <i>reads</i> appariés conservés	612 319 530	93.16

Les *reads* conservés ont été assemblés *de novo* avec *SOAPdenovo* en utilisant différentes valeurs de *k-mer* : 41-mer, 51-mer, 61-mer, 71-mer et 81-mer. Les statistiques des différents assemblages obtenus (formation de *contigs* puis de *scaffolds*) sont indiquées dans le Tableau 4.

Tableau 4 : Statistiques des assemblages générés avec *SOAPdenovo* à partir des *reads Illumina*. Le N50 correspond à la taille de la séquence assemblée (*contig/scaffold*) telle que 50% des bases de l'assemblage sont comprises dans des *contig/scaffold* de taille supérieure au N50. Les N présents dans l'assemblage correspondent aux nucléotides de nature indéterminée issus de la formation des *scaffolds* à partir de l'information d'appariement.

	Assemblages <i>SOAP de novo</i>				
	41-mer	51-mer	61-mer	71-mer	81-mer
Nombre de séquences assemblées	4 252 071	5 274 048	5 056 750	4 191 167	2 947 060
Taille totale du génome assemblé (pb)	1 231 480 571	1 486 467 326	1 510 407 617	1 351 385 012	948 506 415
Séquence assemblée la plus longue (pb)	27 683	16 070	13 369	15 771	16 930
Séquence assemblée la plus courte (pb)	100	100	100	100	100
Nombre de séquences > 1kb	145 971	206 315	233 299	228 596	161 419
Taille moyenne des séquences assemblées (pb)	290	282	299	322	322
Taille médiane des séquences assemblées (pb)	173	159	154	153	163
N50 (pb)	395	424	463	511	512
% GC	28.26	29.93	30.87	31.95	33.8
% N	7.56	4.84	3.32	2.41	1.98

L'assemblage utilisant le paramètre 61-mer a été retenu. Il s'agit de l'assemblage maximisant la quantité de nucléotides assemblés (1 510 407 617 pb, contre 1 486 467 326 pb pour le second meilleur assemblage suivant ce critère, réalisé avec le paramètre 51-mer), mais aussi de l'assemblage disposant du plus grand nombre de séquences assemblées de plus de 1 000 pb (233 299 séquences, contre 228 596 pour le second meilleur assemblage suivant ce critère, réalisé avec le paramètre 71-mer). Il est notable que plus le *k-mer* utilisé est élevé, plus le taux de GC est élevé, à l'inverse du taux de N présent qui est de plus en plus faible avec l'augmentation du paramètre de *k-mer* (indiquant que la proportion de *scaffolds* est de plus en plus faible par rapport au nombre de *contigs* assemblés). De plus, plus le *k-mer* utilisé est élevé, plus le N50 est élevé (paramètre qui correspond à la taille de la séquence assemblée de sorte que 50% des bases de l'assemblage sont comprises dans des *contigs/scaffolds* de taille supérieure au N50).

Une étape de *Gapfilling* a ensuite été réalisée sur le génome assemblé sélectionné, afin de diminuer la proportion de N présents, et donc de combler les lacunes présentes au sein des *scaffolds*. Le génome de 61-mer présentait 1 370 268 lacunes, pour un total de 50 084 571 N (Tableau 5).

Tableau 5 : Statistiques générales du génome 61-mer avant et après l'étape de *Gapfilling*.

	Génome 61-mer original	Génome 61- mer <i>post- Gapfilling</i>	Diminution (%)
Nombre de séquences assemblées	5 056 750	5 056 750	0.00
Taille totale du génome assemblé (pb)	1 510 407 617	1 432 273 152	5.17
Taille totale du génome assemblé sans N (pb)	1 460 323 046	1 408 912 510	3.52
Nombre total de N	50 084 571	23 360 642	53.36
Séquence assemblée la plus longue (pb)	13 369	13 369	0.00
Séquence assemblée la plus courte (pb)	100	100	0.00
Taille moyenne des séquences assemblées (pb)	298	283	5.03
N50 (pb)	463	407	12.10
% N	3.32	1.87	43.67

L'étape de *Gapfilling* a légèrement diminué la taille totale du génome assemblé (5.17% de nucléotides en moins, soit 78 134 465 nucléotides), sachant que 26 723 929 de ces nucléotides étaient indéterminés (soit plus de la moitié des N présents, Tableau 5). L'élimination des autres nucléotides est due à la jointure de différentes lacunes, qui a supprimé des positions redondantes de part et d'autre de la lacune. Si un peu moins de la moitié des N a été retirée (43.67%, Tableau 5), ces N ne constituaient que 30 396 lacunes parmi les 370 268 identifiées avant le *Gapfilling*. Le comblement des lacunes se traduit aussi par une diminution du N50 et de la taille moyenne des séquences assemblées (de 12.1% et de 5.03%, respectivement). C'est cet assemblage *post-Gapfilling* qui a ensuite été utilisé pour la première phase de recherche de séquences candidates Y-spécifiques (cf : chapitre de thèse suivant).

II. Assemblages du génome mâle d'*A. nasatum* à partir des données *Illumina* et des données *Pacific Biosciences*

1. Assemblage du génome d'*A. nasatum* avec *CeleraAssembler*

Le séquençage *PacBio* a permis la génération de 5 962 613 *reads*, qui représentaient 43 152 605 152 nucléotides. Parmi ces lectures de séquençage, 40 539 d'entre-elles étaient des *reads CCS* (0.23% de l'ensemble des *reads*), représentant 100 893 323 nucléotides (0.68% du total des nucléotides séquencés). La première phase pré-assemblage était donc de corriger les erreurs des *reads*, et ce via l'utilisation des *reads CCS* en association avec les *reads Illumina*. Cette étape a permis la correction de 28 904 326 297 nucléotides (soit près de 67% de l'ensemble des nucléotides séquencés). Au final, cette correction a permis d'obtenir un ensemble de données corrigées (*reads* sans erreur à corriger + *reads* corrigés + *reads CCS*) de 41 890 327 966 nucléotides (soit 97.1% de tous les nucléotides issus du séquençage *PacBio*). Le N50 de ces lectures de séquençage corrigées était de 10 824 nucléotides.

Ce jeu de données filtré a été ensuite utilisé pour générer un assemblage avec le logiciel *CeleraAssembler*. Les statistiques détaillées de cette nouvelle version d'assemblage sont indiquées dans le Tableau 6.

Tableau 6 : Statistiques de l'assemblage génomique réalisé avec *CeleraAssembler* à partir des données *PacBio* corrigées. L'assemblage réalisé uniquement à partir des données *Illumina* (*post-Gapfilling*) est indiqué à titre comparatif.

	Génome <i>Illumina</i>	Génome assemblé avec <i>CeleraAssembler</i>
Nombre de séquences assemblées	5 056 750	343 357
Taille totale du génome assemblé (pb)	1 432 273 152	2 052 692 966
Séquence assemblée la plus longue (pb)	13 369	92 419
Séquence assemblée la plus courte (pb)	100	1007
Nombre de séquences > 1kb	233 299	343 357
Taille moyenne des séquences assemblées (pb)	283	5 978
Taille médiane des séquences assemblées (pb)	154	4 450
N50 (pb)	407	7 774
% GC	30.87	34.14
% N	1.87	0

L'assemblage réalisé avec *CeleraAssembler* est près de 14 fois moins fragmenté que l'assemblage réalisé uniquement avec les données *Illumina*. Il contient aussi 620 419 814 nucléotides supplémentaires, et des séquences assemblées en moyenne 28.9 fois plus longues (de plus, elles sont toutes supérieures à 1 000 pb). Aucun nucléotide indéterminé n'est présent puisque l'assemblage effectué avec *CeleraAssembler* a été réalisé sans étape de formation de *scaffolds*. Cet assemblage a été utilisé pour la 2nde phase de recherche de séquences candidates Y-spécifiques à partir des méthodes *CQ* et *YGS* (cf : Chapitre 2, page 113).

2. Assemblage du génome d'*A. nasatum* avec *DBG2OLC*

La première étape d'assemblage en suivant le pipeline *DBG2OLC* était de générer un pré-assemblage *de novo* à partir des *reads Illumina*. Celui-ci contenait 11 658 845 contigs, et rassemblait 1 435 423 395 nucléotides (pour un N50 de 176 pb). Ce pré-assemblage a ensuite été utilisé en association avec les *reads PacBio* (uniquement les *reads* corrigés de plus de 5 000 pb ont été utilisés) afin de générer l'ossature du génome qui a ensuite été affinée. Les statistiques de cette ossature (*backbone assembly*), ainsi que la proportion de *reads Illumina* alignés sur cet assemblage sont indiquées dans le Tableau 7. Ce *backbone assembly* contenait déjà plus de 12 fois moins de *contigs* que le génome assemblé via *CeleraAssembler* pour un N50 10 fois supérieur (respectivement 27 247 *contigs* vs. 343 357 *contigs*, et 76 343 pb vs. 7 774 pb, Tableau 7), témoignant d'une fragmentation bien plus limitée de l'assemblage. Ceci est appuyé par le fait que 94% des *reads Illumina* ont pu être alignés sur ce génome (dont 79.97% ont été alignés dans le bon sens de lecture avec une taille d'*insert* cohérente, dites "paires propres", Tableau 7).

Tableau 7 : Statistiques des assemblages obtenus avec le pipeline *DBG2OLC*. L'assemblage qualifié de "*backbone assembly*" correspond à la première ossature de l'assemblage obtenue à partir des données *PacBio*. La partie supérieure du tableau correspond aux statistiques directes des assemblages, tandis que la partie inférieure concerne l'alignement des *reads Illumina* effectué sur les différents génomes. Les "-" indiquent que les valeurs n'ont pas été mesurées. Les paires "propres" correspondent aux paires de *reads* alignés dans le bon sens de lecture avec une taille d'*insert* cohérente. La couverture par nucléotide correspond au nombre de *reads* incluant un nucléotide donné au sein de séquences assemblées.

	Pipeline <i>DBG2OLC</i>				
	<i>Backbone assembly</i>	Assemblage post-polissage	Assemblage post-décontamination	Assemblage final (post- <i>Gapfilling</i>)	Génome assemblé avec <i>CeleraAssembler</i>
Nombre de séquences assemblées	27 247	27 247	26 906	25 194	343 357
Taille totale du génome assemblé (pb)	1 334 730 731	1 234 751 348	1 223 005 345	1 223 176 554	2 052 692 966
Taille moyenne des séquences assemblées (pb)	48 986	45 317	45 455	48 550	5 978
N50 (pb)	76 343	74 825	75 116	86 284	7 774
% N	0.00	0.00	0.00	0.01	0.00
% de <i>reads</i> mâles <i>Illumina</i> alignés	94.00	94.22	-	93.99	-
parmi ces <i>reads</i> alignés :					
% de <i>reads</i> alignés en paires "propres"	79.97	80.06	-	80.03	-
% de <i>reads</i> alignés appariés	96.85	97.01	-	96.97	-
% de <i>reads</i> alignés en singleton	3.15	2.99	-	3.03	-
Couverture <i>Illumina</i> par nucléotide	43.38x	43.83x	-	44.11x	-

L'étape de polissage n'a pas modifié le nombre de *contigs* présents mais a diminué d'environ 7.5% le nombre de nucléotides assemblés (99 979 383 nucléotides ont été retirés), ce qui se traduit aussi par une diminution de 7.5% de la taille moyenne des séquences assemblées, et une diminution de 2% du N50. L'impact de cette étape a donc été plus fort sur les séquences de plus petite taille (Tableau 7). Les statistiques d'alignement des *reads* sont restées stables, puisque 94.22% des *reads* ont pu être alignés sur ce génome, dont 80.06% sous la forme de paires "propres" (contre respectivement 94% et 79.97% pour la version précédente du génome, Tableau 7).

L'étape de décontamination a permis d'éliminer 341 séquences (qui représentaient 11 746 003 nucléotides, soit 1% des nucléotides) sur la base de leur taux de GC et de leur couverture respective (Figure 25). Les séquences présentant des *hits* avec des séquences n'étant pas taxonomiquement assignées ont été éliminées (143 séquences, soit 4.86 Mb), ainsi que les séquences de bactéries hors *Wolbachia* (41 séquences pour un total de 0.88 Mb) et les séquences assignées aux plantes et aux champignons (157 séquences, d'environ 6 Mb). Le faible nombre de séquences assignées aux bactéries ainsi qu'aux plantes et champignons traduit que l'exclusion des caeca et du tube digestif a été réalisée avec succès et efficacité. L'ensemble des séquences éliminées étant de petite taille mais en faible nombre, leur élimination a permis une légère augmentation de la taille moyenne des séquences assemblées et du N50 (0.3% en plus pour ces deux paramètres). Les 4 séquences identifiées comme d'origine virale ont été conservées, puisqu'il pouvait s'agir de séquences virales intégrées dans le génome d'*A. nasatum*. Dix séquences assignées aux bactéries ont été retenues puisqu'elles présentaient des analogies avec des séquences connues de *Wolbachia*, et peuvent donc correspondre à des événements de transferts horizontaux entre la bactérie endosymbiotique et l'hôte *A. nasatum* (voir Chapitre 1, page 95). Sur la Figure 25, il est possible d'observer un ensemble de 123 séquences assignées à des séquences eucaryotes présentant un taux de GC supérieur ou égal à 50%. Il s'agit de séquences présentant un fort taux de séquences microsatellites, contenant majoritairement des G ou des C.

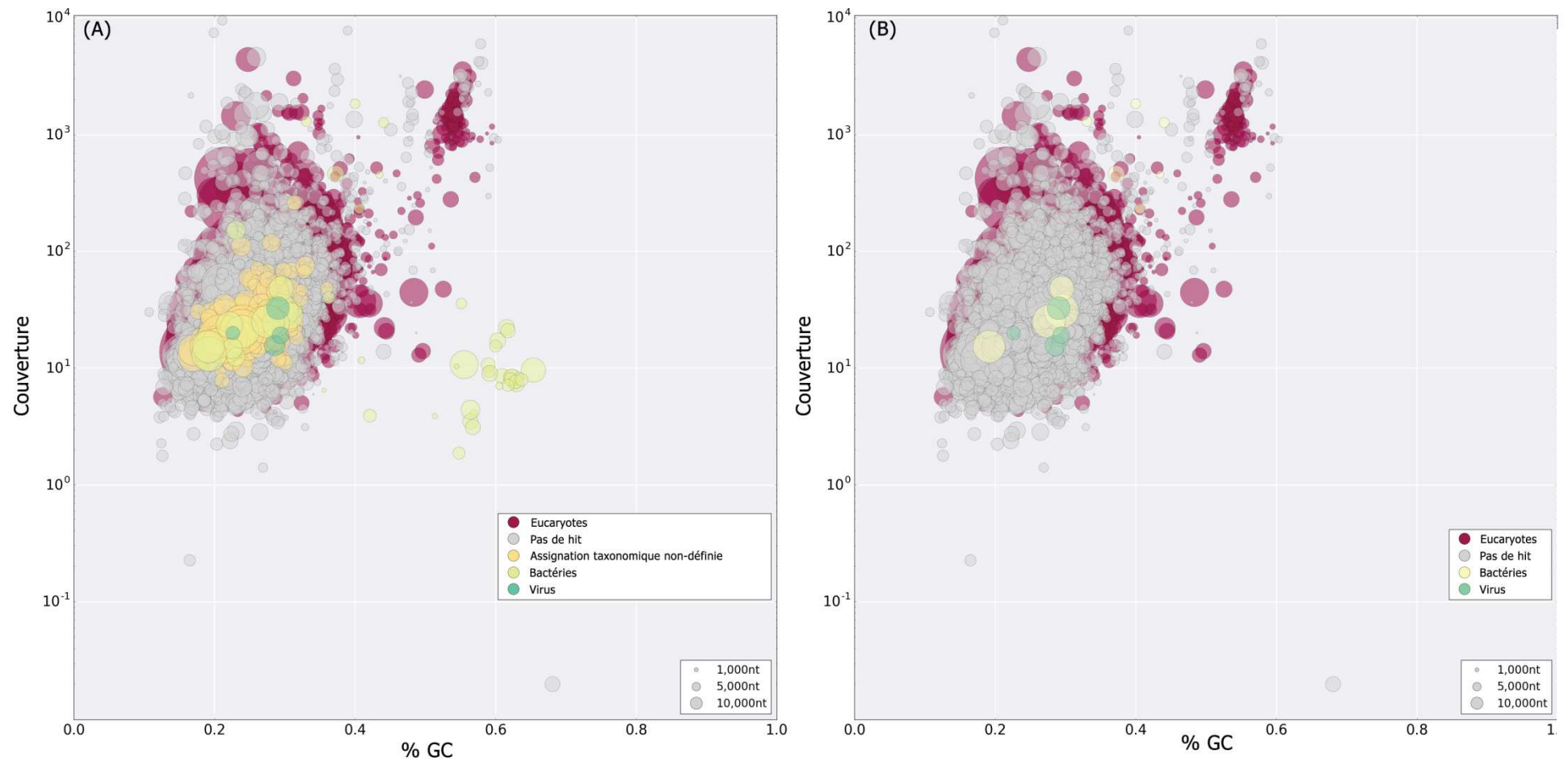


Figure 25 : Graphiques *Taxa-annotated-GC-Coverage* du génome assemblé avec *DBG2OLC* (A) avant et (B) après décontamination (généralisé avec *Blobtools*). Chaque cercle correspond à une séquence de l'assemblage, et la taille de ces cercles est proportionnelle à la taille des séquences. Chaque séquence est colorée selon son assignation taxonomique (Eucaryotes, pas de *hit*, assignation taxonomique non définie, bactéries et virus). 27 247 séquences sont présentes avant décontamination (A), et 26 906 le sont après décontamination (B).

La dernière phase de l'assemblage a été d'assembler les *contigs* en *scaffolds* à partir des données de séquençage ARN. Ainsi, un transcriptome d'*A. nasatum* a été généré à partir de données publiques (Romiguier et al., 2014). Celui-ci est composé de 58 009 *contigs*, pour une taille totale de 43 229 630 nucléotides (taille moyenne des *contigs* : 745 pb) et un N50 de 1 208 pb. L'assemblage final du génome de *A. nasatum* ne compte que 171 209 pb supplémentaires (soit une augmentation de 0.01%) qui sont en réalité des nucléotides indéterminés (N). L'assemblage final ne compte plus que 25 194 séquences (6% de séquences en moins), qui ont une longueur moyenne de 48 550 pb (7% plus grandes) et un N50 de 86 284 pb (soit une augmentation de 15%, Tableau 7). Cette étape constitue donc une véritable amélioration de la continuité de l'assemblage, sachant que les statistiques d'alignement des *reads* sont elles aussi restées stables (93.99% de *reads* alignés, dont 80.03% en paires "propres").

L'évaluation de l'exhaustivité fonctionnelle du génome a été réalisée en le comparant à une base de données de gènes fortement conservés spécifiques aux arthropodes (*BUSCO*). Cette analyse a révélé que sur les 1 066 gènes *BUSCO* recherchés, 974 sont identifiés comme complets dans l'assemblage d'*A. nasatum*, dont 932 en copie unique et 42 en multiples copies (Tableau 8). De plus, 27 gènes ont été identifiés comme fragmentés (Tableau 8). Ainsi, 93.9% des gènes *BUSCO* sont présents dans le génome d'*A. nasatum* et seulement 6.1% de ces gènes n'ont pas pu être identifiés. Ces résultats témoignent de la qualité fonctionnelle importante de l'assemblage, puisque la majorité des "*core genes*" attendus au sein d'un génome d'arthropode sont présents (Tableau 8).

Tableau 8 : Exhaustivité fonctionnelle de l'assemblage final évalué par *BUSCO*.

	Nombre de gènes <i>BUSCO</i> retrouvés dans l'assemblage final (%)
Nombre de gènes <i>BUSCO</i> recherchés	1 066
Gènes <i>BUSCO</i> complets	974 (91.4%)
Gènes <i>BUSCO</i> complets en copie unique	932 (87.4%)
Gènes <i>BUSCO</i> complets et dupliqués	42 (3.9%)
Gènes <i>BUSCO</i> fragmentés	27 (2.5%)
Gènes <i>BUSCO</i> manquants	65 (6.1%)

III. Annotation structurale et fonctionnelle du génome d'*A. nasatum* et caractérisation des éléments transposables

L'annotation structurale et fonctionnelle du génome, réalisée avec le logiciel *MAKER* (en complément des données issues du transcriptome d'*A. nasatum*), a permis de prédire la présence de 14 770 gènes (d'une taille moyenne de 9 435 pb en comptant le cadre de lecture ouvert et les introns) au sein du génome, représentant 11.39% de l'ensemble des nucléotides assemblés (Tableau 9). Les gènes d'*A. nasatum* sont composés en moyenne de 7 exons (de 199 pb) séparés par des introns d'environ 1 436 pb (Tableau 9). Les caractéristiques structurales des gènes annotés sont détaillées dans le Tableau 9. De ces gènes prédits, 11 927 domaines Pfam ont été identifiés ainsi que 13 245 termes GO associés à des fonctions biologiques.

Tableau 9 : Caractéristiques structurales des gènes présents dans le génome d'*A. nasatum*.

		Taille (nucléotides)	% du génome
Gènes	Taille totale des gènes annotés	139 351 398	11.40
	Nombre de gènes annotés	14 770	
	Taille moyenne des gènes	9 435	
Exons	Nombre total d'exons prédits	98 219	
	Taille totale de l'ensemble des exons	19 570 226	1.60
	Taille moyenne des exons	199	
	Nombre moyen d'exons par gène	7	
Introns	Nombre total d'introns prédits	83 448	
	Taille totale de l'ensemble des introns	119 781 172	9.80
	Taille moyenne des introns	1 436	
	Nombre moyen d'introns par gène	6	
CDS	Taille totale des régions codantes (CDS)	16 167 366	1.32
	Taille moyenne des régions codantes (CDS)	1 095	
UTR	Taille totale des régions transcrites non-traduites (UTR)	3 441 386	0.28
Autres	Densité moyenne des gènes (nucléotides/gène)	82 815	
	Distance moyenne entre chaque gène	73 380	

L'annotation des éléments répétés a permis de révéler que 63% du génome est composé d'éléments répétés (Tableau 10), et plus précisément de 38.97% d'ETs (représentant 476 691 017 pb) et de 23.79% de séquences répétées en tandem (correspondant à 290 959 664 pb). Les ETs les plus abondants identifiés sont les rétrotransposons non-LTR (LINEs) qui correspondent à 20.45% du génome total (250 188 430 pb) et les éléments à ADN, aussi appelés transposons, qui constituent 12.36% du génome (151 153 496 pb) (Tableau 10).

Tableau 10 : Statistiques des différents éléments répétés identifiés au sein du génome d'*A. nasatum*.

Type d'éléments répétés	Total en nucléotides	% du génome
SINEs	4 864 228	0.40
LINEs	250 188 430	20.45
Eléments LTR	60 934 808	4.98
Eléments à ADN (transposons)	151 153 496	12.36
Eléments non-classés	9 550 055	0.78
Total des éléments transposables	476 691 017	38.97
ARNs structuraux	441 287	0.04
Séquences satellites	2 532 935	0.21
Séquences répétées en tandem	290 959 664	23.79
Total	770 624 903	63.00

Les analyses de divergence entre les multiples copies d'ETs et leurs séquences consensus respectives (Figure 26) montrent que la majorité des ETs identifiés ont entre 6 et 12% de divergence avec leur séquence consensus respectives (Figure 26, B), tandis que l'activité de transposition semble avoir ralenti (sans pour autant être arrêtée), puisqu'environ 7% du génome provient d'ETs ayant moins de 6% de divergence vis-à-vis de leur séquence consensus (Figure 26, B). A l'aide d'analyses de datation moléculaire, il est possible d'estimer la période pendant laquelle les ETs ont été actifs. Cette estimation peut se faire en additionnant l'âge médian de chaque branche de l'arbre phylogénétique présenté dans le Chapitre 4 (âge estimé avec *BEAST*, Becking et al., 2017; Chapitre 4), et en le divisant par la somme des pourcentages de divergence de chaque branche (divergence estimée avec *RAxML*, Becking et al., 2017; Chapitre 4). Ainsi 1% de divergence

correspond à 5.61 Millions d'années (Ma). A partir de ces estimations, il est possible de déduire que la plus forte activité de transposition s'est déroulée entre -25 Ma et -50 Ma (Figure 26, B). Enfin, si les ETs les plus actifs, notamment récemment, appartiennent à la superfamille des LINE/CR1 (qui constituent à eux seuls 12.8% du génome d'*A. nasatum*), l'activité la plus récente (proche de 0% de divergence) semble surtout être attribuable à des transposons ADN.

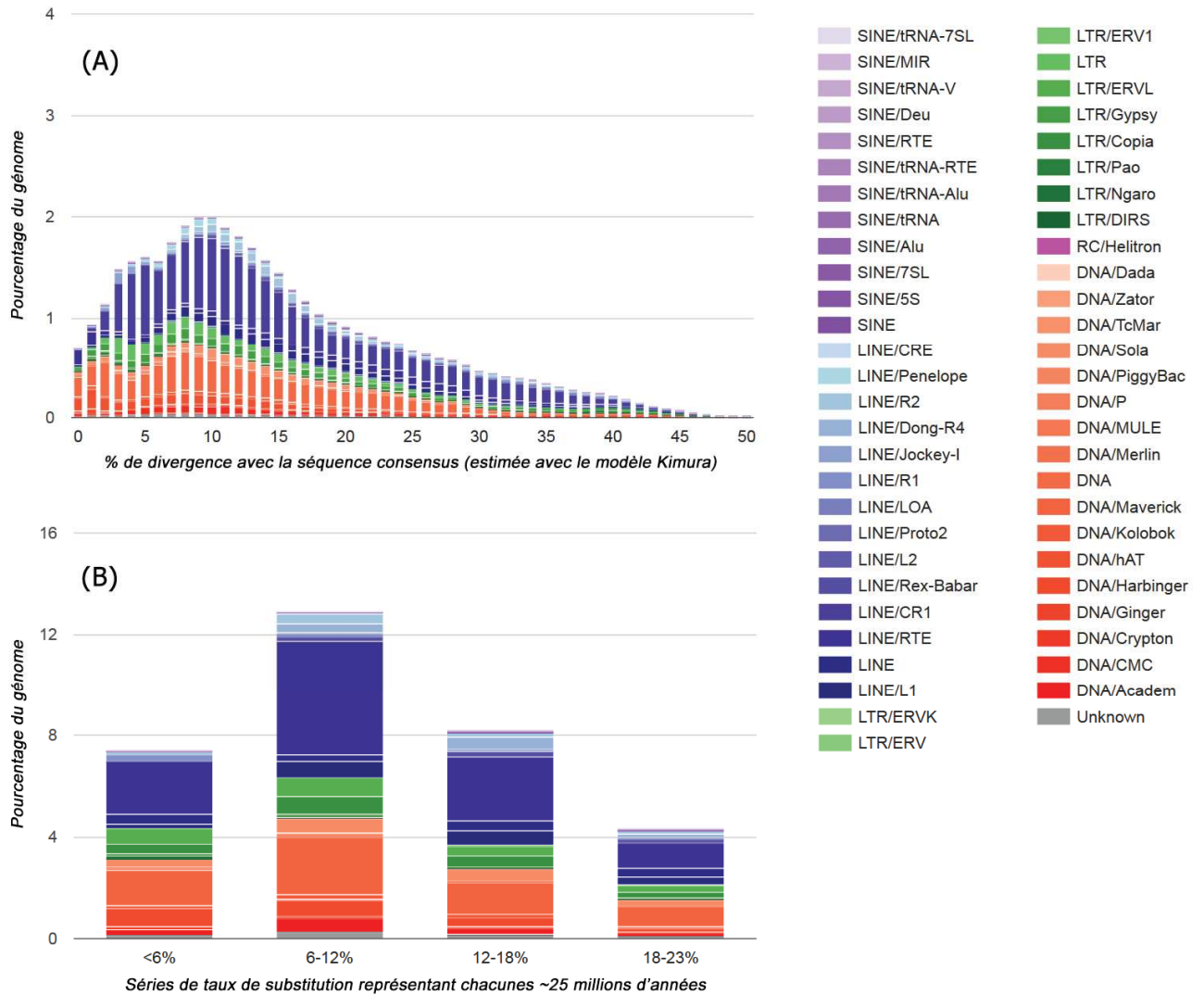


Figure 26 : Représentation de la divergence entre les ETs identifiés au sein du génome d'*A. nasatum* et leurs séquences consensus. Chaque couleur est associée à une classe d'ETs. (A) Contribution de chaque classe d'ETs à la composition du génome, représentée en fonction de leur divergence avec leurs séquences consensus respectives (rapportée au taux de substitution de Kimura calculé par *RepeatMasker*). (B) Contribution de chaque classe d'ETs à la composition du génome, représentée par gammes de 25 Ma (représentant environ 6% de divergence, Becking et al., 2017). Ces temps de divergence ont été estimés en additionnant l'âge médian de chaque branche de l'arbre phylogénétique présenté en partie 3 (âge estimé avec *BEAST*, Chapitre 4), et en le divisant par la somme des pourcentages de divergence de chaque branche (divergence estimée avec *RAxML*, Chapitre 4). Ainsi 1% de divergence correspond à 5.61 Millions d'années (Ma).

IV. Caractérisation des insertions de *Wolbachia* dans le génome d'*A. nasatum*

La caractérisation des insertions du génome de *Wolbachia* dans celui d'*A. nasatum* par *BLASTn* a révélé que 16 *contigs/scaffolds* étaient porteurs de 19 insertions, représentant un total de 6 064 pb. Les 10 *contigs/scaffolds* assignés à *Wolbachia* lors de la phase d'élimination des séquences contaminantes (Figure 25) sont par ailleurs inclus dans les 16 *contigs/scaffolds* identifiés par *BLASTn*. Les 6 séquences nouvellement identifiées ne correspondent qu'à des insertions de moins de 400 pb, et n'ont donc pas pu être caractérisées précédemment. Dans le cas d'*A. vulgare* (suite à la même approche de détection), 138 insertions de *Wolbachia* réparties dans 70 *contigs/scaffolds* du génome d'*A. vulgare* ont pu être caractérisées, totalisant 147 159 pb (Chebbi et al, en préparation). Les analyses réalisées par *BLASTp* ont permis d'identifier 5 gènes de *Wolbachia* intégrés dans le génome d'*A. nasatum*, ces gènes étant présents au sein de 3 *contigs/scaffolds* différents. Deux de ces gènes codent pour des facteurs d'élongation (EF4), deux autres gènes codent pour une même Pyridoxine-phosphate-synthase et le dernier gène identifié code pour une protéine à domaine ankyrine (ANK3). Des analyses similaires réalisées sur le génome d'*A. vulgare* ont permis de mettre en évidence l'insertion de 35 gènes de *Wolbachia* répartis dans 10 *scaffold/contigs*.

Seulement 2 des 19 insertions ont pu être alignées contre l'ensemble des génomes de *Wolbachia* afin de caractériser leur souche originelle. Seulement 3 des 17 autres insertions ont pu être alignés contre 3 à 5 génomes de *Wolbachia* (ce qui ne permet cependant pas une caractérisation fiable de l'origine de ces insertions), tandis que les 14 dernières insertions n'ont pu être alignées contre aucun génome de *Wolbachia*. Les insertions portées par le Contig6190 et par le Contig22683 présentent respectivement 4.6% (Figure 27, A) et 8.1% (Figure 27, B) de divergence avec la souche *wVulC* (connues pour infecter naturellement *A. vulgare*), alors qu'elles comptent respectivement 7.4% (Figure 27, A) et 10.1% (Figure 27, B) de divergence avec la seconde souche la plus proche (souche naturelle de l'hémiptère *Diaphorina citri*). Les 2 insertions caractérisées sont proches phylogénétiquement de la souche *wVulC* connue pour infecter *A. vulgare* (Figure 27).

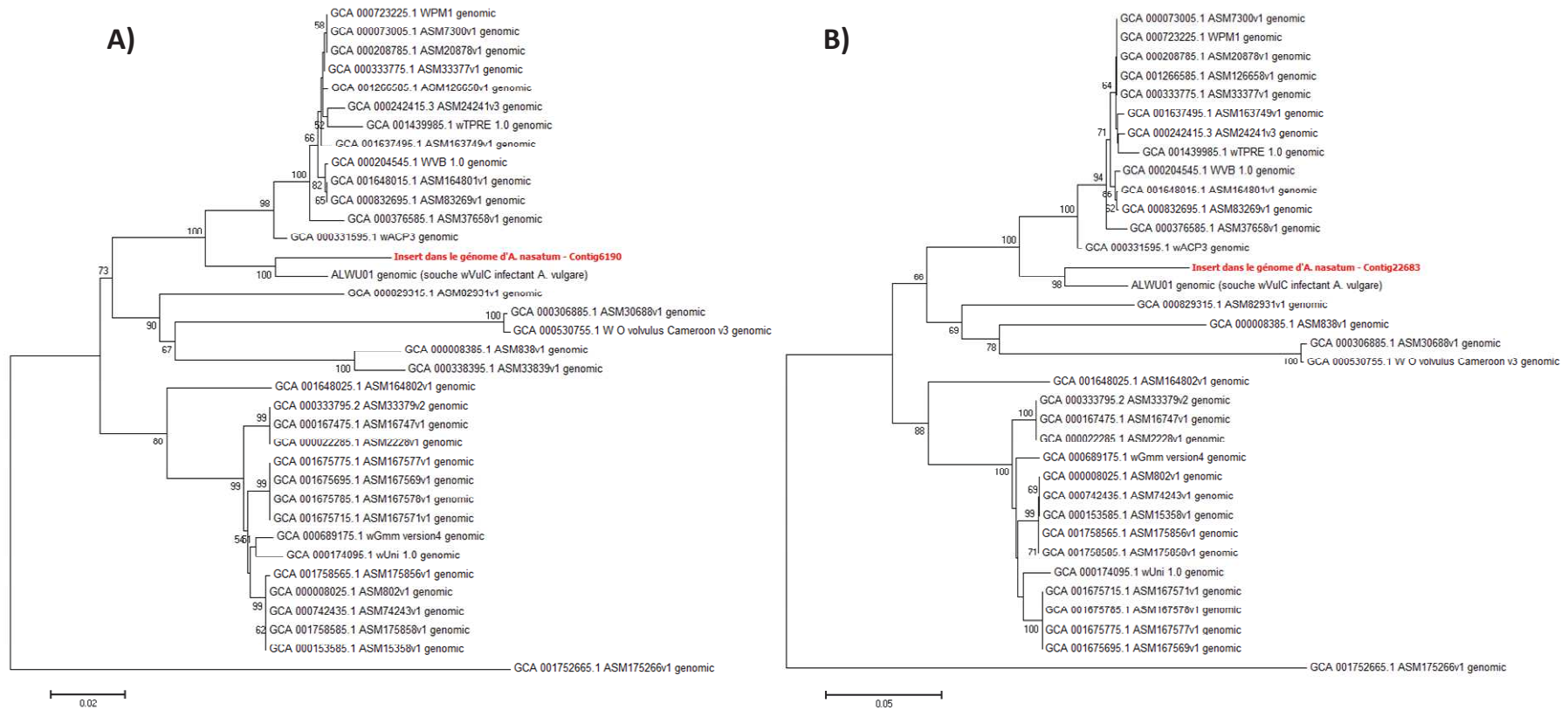


Figure 27 : Relations phylogénétiques de deux insertions de *Wolbachia* dans le génome d'*A. nasatum*. A) Analyse phylogénétique réalisée pour l'insertion portée par le *Contig6190* à partir d'un alignement de 1 199 pb. B) Analyse phylogénétique réalisée pour l'insertion portée par le *Contig22683* à partir d'un alignement de 1 125 pb. Pour les deux arbres, la longueur des branches est proportionnelle à la distance évolutive. Les nombres indiqués aux différents nœuds correspondent aux scores de *bootstrap* obtenus (les valeurs inférieures à 50 ne sont pas indiquées).

V. Comparaison de la structure des génomes d'*A. nasatum* et d'*A. vulgare*

Parallèlement à l'assemblage du génome d'*A. nasatum*, le génome d'un autre isopode terrestre apparenté, *A. vulgare* (ces deux espèces présentent environ 6% de divergence nucléaire, correspondant à un temps de divergence d'environ 25 Ma; Becking et al., 2017; Figure 63, Chapitre 4, page 211), a été assemblé au laboratoire par Amine Chebbi (Chebbi et al., en préparation). La même stratégie a été utilisée pour assembler les deux génomes, permettant ainsi une comparaison entre les deux espèces. Les statistiques de l'assemblage obtenu pour *A. vulgare*, ainsi que le compte rendu des analyses *BUSCO* effectuées sont détaillés dans le Tableau 11.

Tableau 11 : Statistiques descriptives de l'assemblage du génome d'*A. vulgare* (Chebbi et al., en préparation) et statistiques des analyses *BUSCO* réalisées.

	Assemblage du génome d' <i>A. vulgare</i>
<i>Statistiques générales</i>	
Nombre de séquences assemblées	43 548
Taille totale du génome assemblé (pb)	1 725 110 628
Taille moyenne des séquences assemblées (pb)	39 614
N50 (pb)	51 088
% N	0.43%
<i>Analyses BUSCO</i>	
Gènes <i>BUSCO</i> complets (% de gènes <i>BUSCO</i> recherché)	937 (87.9%)
Gènes <i>BUSCO</i> complets en copie unique (% de gènes <i>BUSCO</i> recherché)	879 (82.5%)
Gènes <i>BUSCO</i> complets et dupliqués (% de gènes <i>BUSCO</i> recherché)	58 (5.4%)
Gènes <i>BUSCO</i> fragmentés (% de gènes <i>BUSCO</i> recherché)	44 (4.1%)
Gènes <i>BUSCO</i> manquants (% de gènes <i>BUSCO</i> recherché)	85 (8.0%)
Nombre de gènes <i>BUSCO</i> recherchés	1 066

Par rapport à *A. nasatum*, le génome d'*A. vulgare* est plus grand de 501 934 074 pb (soit 41% de données en plus de celles assemblées dans le génome

d'*A. nasatum*). Cet assemblage contient 18 354 *scaffolds/contigs* supplémentaires et son N50 est inférieur de 35 196 pb. Le génome d'*A. vulgare* présente donc plus de séquences de petite taille en comparaison avec *A. nasatum*, probablement en relation avec les séquences répétées difficiles à assembler. Les résultats de l'analyse *BUSCO* sont en revanche très proches de ceux obtenus précédemment, puisque 92.0% (981 gènes sur 1 066) des gènes *BUSCO* ont été identifiés (contre 93.9% chez *A. nasatum*).

L'annotation structurale a permis d'identifier 19 208 gènes chez *A. vulgare* (contre 14 770 chez *A. nasatum*), ce qui correspond à 4 438 gènes supplémentaires présents chez *A. vulgare* (Tableau 12). L'ensemble des protéines annotées issues des génomes d'*A. nasatum* et d'*A. vulgare* ont été classées par *OrthoMCL* dans 9 140 groupes de protéines orthologues. Parmi ces 9 140 groupes d'orthologues, 8 036 contiennent des séquences appartenant à la fois à *A. nasatum* et à *A. vulgare*, totalisant respectivement 9 496 et 10 894 protéines (représentant donc 1 398 protéines supplémentaires dans le génome d'*A. vulgare*) (Figure 28).

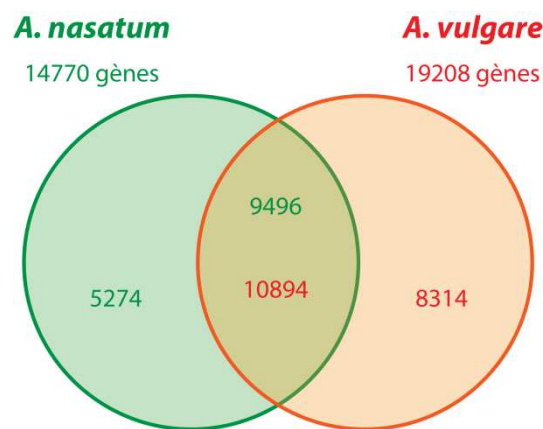


Figure 28 : Diagramme de Venn représentant l'assignation des gènes orthologues entre *A. nasatum* (en vert) et *A. vulgare* (en rouge). Les protéines communes aux deux génomes (9 496 pour *A. nasatum* et 10 894 pour *A. vulgare*) correspondent à 8 036 groupes de gènes orthologues.

L'analyse réalisée avec *OrthoMCL* a identifié 5 274 gènes propres à *A. nasatum* (Figure 28), dont 4 226 sont présents en copie unique au sein du génome (Tableau 12). Enfin, 8 314 gènes ont été identifiés comme étant propres au génome d'*A. vulgare* (Figure 28), dont 6 390 sont présents en copie unique. Il est notable que la majorité des protéines propres à chaque génome a été annotée comme n'ayant pas de fonction connues (Tableau 12). La forte variation en

nombre de gènes entre ces deux génomes ne se traduit pas par une forte augmentation en nombre de bases, puisque l'ensemble des gènes caractérisés représente 139 351 398 pb chez *A. nasatum* et 165 698 054 pb chez *A. vulgare*. Cela sous-entend que le plus grand nombre de gènes identifiés chez *A. vulgare* ne permet pas d'expliquer la différence de taille de génome entre ces deux espèces. D'ailleurs, la densité de gènes entre les deux génomes est similaire, puisque le génome d'*A. nasatum* compte environ 12 gènes par Mb, tandis que le génome d'*A. vulgare* près de 11 gènes par Mb. Enfin, l'annotation fonctionnelle réalisée a permis l'identification de 10 524 domaines Pfam et de 12 031 termes GO, ce qui est là aussi similaire avec les données du génome de référence d'*A. nasatum* (11 927 domaines *Pfam* identifiés ainsi que 13 245 termes GO).

Tableau 12 : Bilan des analyses d'orthologie réalisées avec *OrthoMCL* entre les génomes d'*A. nasatum* et d'*A. vulgare*. Le nombre de protéines supplémentaires présentes dans le génome d'*A. vulgare* ne tient pas compte des relations d'orthologies entre les séquences issues des deux génomes.

	<i>A. nasatum</i>		<i>A. vulgare</i>		Nombre de protéines supplémentaires retrouvées chez <i>A. vulgare</i>
	Nombre de protéines annotées	Nombre de protéines (%) annotées de fonctions inconnues	Nombre de protéines annotées	Nombre de protéines (%) annotées de fonctions inconnues	
Nombre de gènes annotés	14 770	5 402 (37)	19 208	8 659 (45)	4 438
Nombre de gènes classés dans des groupes de séquences orthologues communs aux deux espèces (correspondant à 8036 orthogroupes)	9 496	2 664 (28)	10 894	3 510 (32)	1 398
Nombre de gènes classés dans des groupes de séquences orthologues propres à chaque espèce (correspondant à 1104 orthogroupes)	1 048	617 (59)	1 924	1 291 (67)	876
Nombre de gènes classés en copie unique propres à chaque espèce	4 226	2 421 (57)	6 390	3 858 (60)	2 164

Les proportions d'éléments répétés présents sont également similaires entre ces deux génomes. De plus, les génomes d'*A. nasatum* et d'*A. vulgare* sont composés respectivement de 37% (environ 453 Mb) et de 31.2% (environ 538 Mb) de régions géniques (introns et exons) et intergéniques (hors répétitions) n'ayant pas été masquées suite à l'identification des séquences répétées (Figure 29).

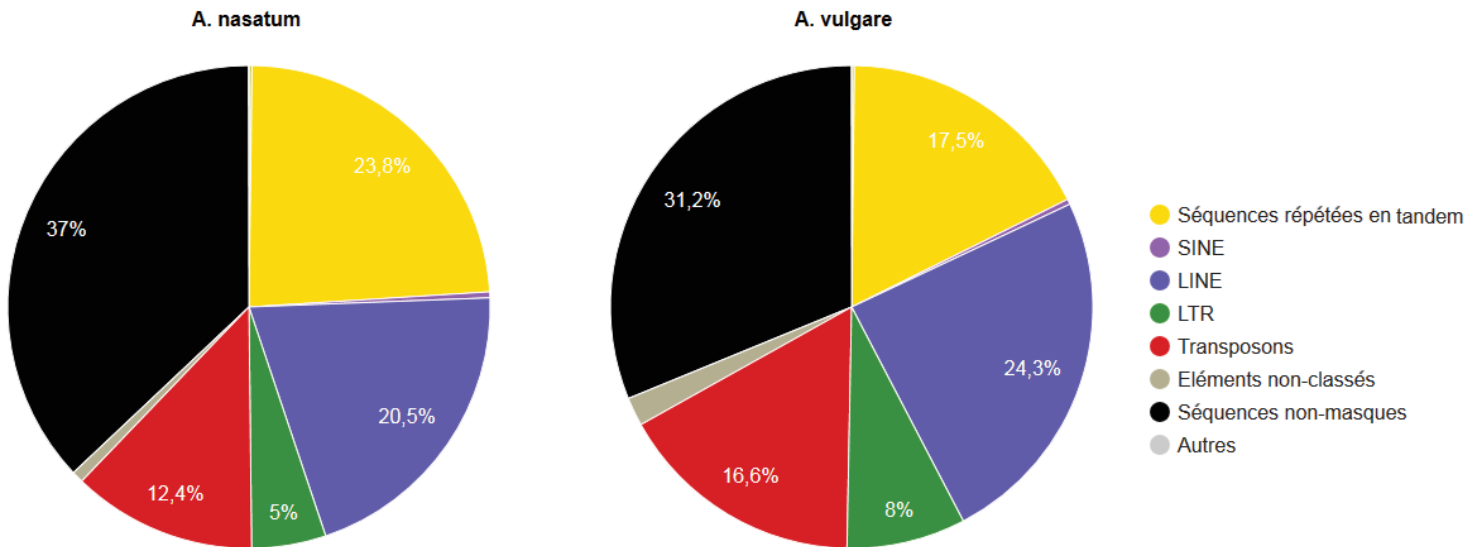


Figure 29 : Représentation de la proportion d'éléments répétés par rapport au génome d'*A. nasatum* et d'*A. vulgare*. Les 4 catégories d'éléments répétés les plus présents sont indiquées avec leur part de génome en pourcentage. Les séquences non-masquées correspondent à la part du génome n'étant pas considérée comme étant composée d'éléments répétés. Pour *A. nasatum*, les éléments non-classés, les SINEs, et les autres séquences (séquences satellites et ARNs structuraux) correspondent respectivement à 0.8%, 0.4% et 0.2% du génome. Pour *A. vulgare*, les éléments non-classés, les SINEs, et les autres séquences (séquences satellites et ARNs structuraux) correspondent respectivement à 1.9%, 0.3% et 0.2% du génome.

Ces deux génomes ont aussi une quantité d'éléments répétés en tandem similaire (respectivement 290 Mb et 301 Mb, qui représentent 23.8% et 17.5% de chaque génome). La différence de taille entre les deux génomes s'explique ainsi essentiellement par la présence d'un plus grand nombre d'ETs au sein du génome d'*A. vulgare* (Figure 30). Le génome d'*A. vulgare* compte en effet 872 078 406 pb issues d'ETs, tandis que le génome d'*A. nasatum* n'en compte que 476 691 017 pb. Ces 395 387 389 pb issues d'ETs présentes en plus chez *A. vulgare* constituent 78.8% de la différence de taille observée par rapport au génome d'*A. nasatum*. Sur les 53 superfamilles d'ETs détectées (appartenant aux SINEs, LINEs, élément LRTs ou aux transposons), 51 sont communes aux deux espèces. Seuls les éléments tRNA-CR1 et U (SINEs) sont propres à *A. vulgare* et les éléments MIR et RTE (SINEs) sont propres à *A. nasatum* (Figure 30). La différence de taille

entre les deux génomes pourrait donc s'expliquer par une activité de transposition plus forte chez *A. vulgare* que chez *A. nasatum* et/ou par un taux de délétion plus fort chez *A. nasatum*. Il est aussi notable que le pic d'activité des ETs chez *A. nasatum* correspond aux copies d'ETs ayant 10% de divergence avec leur séquence consensus, tandis que le pic d'activité des ETs chez *A. vulgare* correspond aux copies d'ETs ayant 6% de divergence avec leur séquence consensus (Figure 30). Le profil décrit par la Figure 30 semble indiquer que le génome d'*A. vulgare* a subi un pic d'activité de transposition peu après la divergence entre *A. vulgare* et *A. nasatum*, ce qui peut alors expliquer la différence d'abondance d'ETs entre les deux génomes.

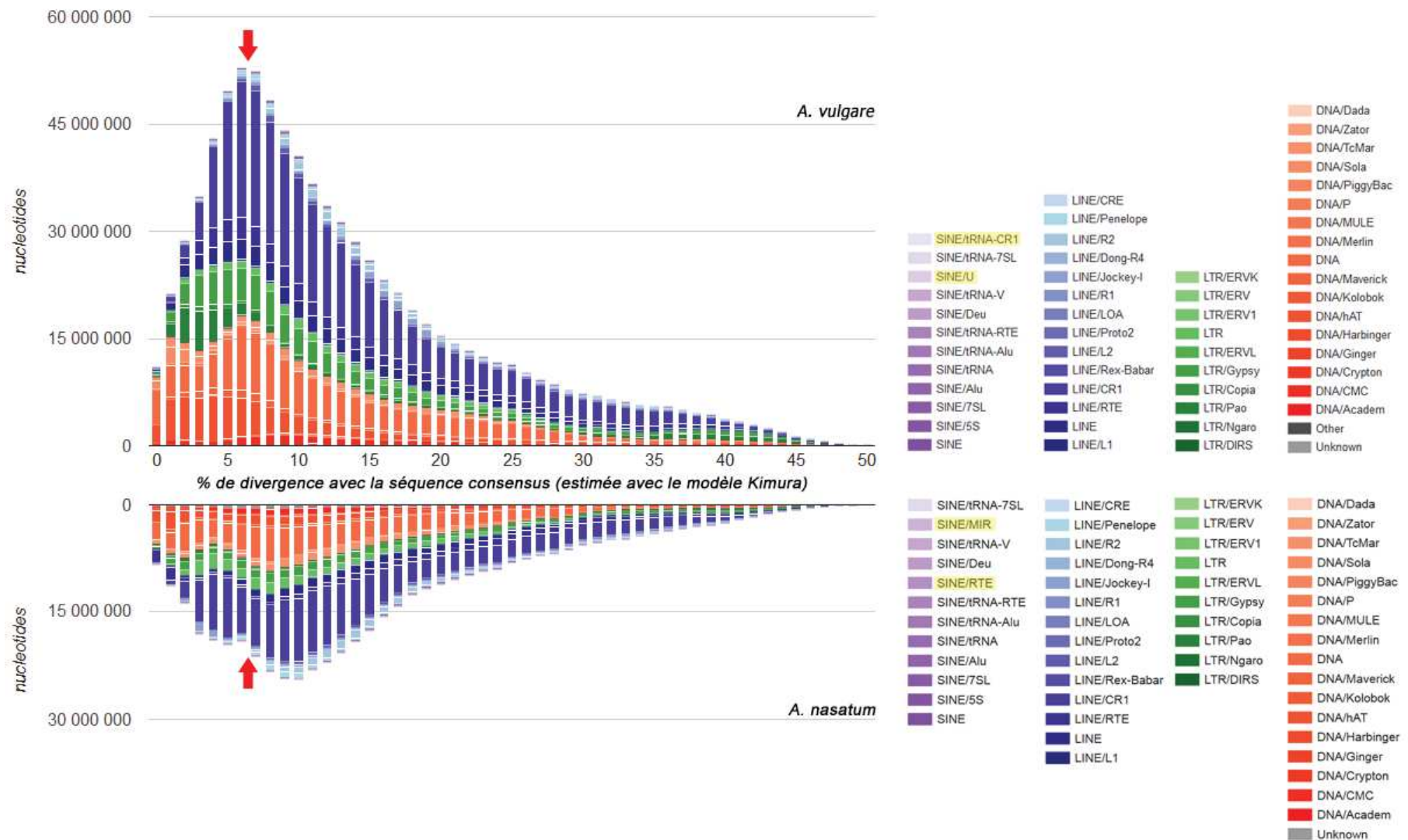


Figure 30 : Comparaison de la divergence entre les ETs identifiés au sein des génomes d'*A. vulgare* et d'*A. nasatum* et leurs séquences consensus respectives. La présence de chaque ETs est représentée en nombre de nucléotides dans le génome en fonction de leur divergence avec leurs séquences consensus respectives (rapportée au taux de substitution de Kimura calculé par *RepeatMasker*). Les ETs identifiés dans le génome d'*A. vulgare* sont indiqués dans la partie supérieure du graphique, et ceux identifiés dans le génome d'*A. nasatum* dans la partie inférieure. Les légendes correspondant aux superfamilles spécifiques à chaque espèce sont surlignées en jaune. Les flèches rouges correspondent au temps de divergence estimé entre *A. nasatum* et *A. vulgare* (soit il y a environ 25 Ma, Becking et al., 2017; Chapitre 4).

Discussion

I. Validité de l'assemblage combinant *reads* courts et longs

L'objectif de cette première partie de thèse était d'obtenir un génome de référence pour l'isopode terrestre *A. nasatum*. En plus de celui généré pour l'espèce apparentée *A. vulgare* (Chebbi et al., en préparation), il s'agit des deux premiers grands génomes de crustacés terrestres assemblés *de novo* à partir de données de séquençage issues à la fois des technologies *Illumina* (*reads* courts) et *PacBio* (*reads* longs). La complexité du génome d'*A. nasatum*, caractérisée par un large contenu en séquences répétées (63% du génome), a rendu cet assemblage particulièrement difficile à réaliser. Cette complexité est illustrée par la première version de l'assemblage, réalisée uniquement à partir des *reads Illumina*. Celle-ci était fortement fragmentée (N50 = 407 pb après *Gapfilling*, pour plus de 5 millions de *contigs/scaffolds*) et contenait une part de nucléotides indéterminés de 1.87 %. Une première approche hybride, combinant à la fois des *reads* courts et longs, a déjà permis une nette amélioration de la continuité de l'assemblage (N50 = 7 774 pb correspondant à 343 357 *contigs*). Une seconde approche, toujours basée sur la combinaison des *reads Illumina* et *PacBio*, a quant à elle permis une amélioration supplémentaire de l'assemblage, celui-ci n'étant plus composé que de 25 194 *contigs/scaffolds*, ayant un N50 supérieur à 86 kb et contenant seulement 0.01% de nucléotides indéterminés. L'apport de données transcriptomiques s'avère être un atout important dans l'amélioration de l'assemblage, permettant de connecter des *contigs* grâce à l'information structurale des gènes (Denton et al., 2014). En plus d'améliorer la connexion des séquences assemblées, les *reads* issus du transcriptome permettent d'éliminer des gènes créés artificiellement à cause de leur fragmentation (Denton et al., 2014). Ainsi dans le cas du génome d'*A. nasatum*, l'information concernant la structure des gènes a permis d'augmenter le N50 de 75 kb à 86 kb.

La stratégie basée sur la combinaison de données issues de *reads* courts (génomiques et transcriptomiques) et de *reads* longs semble donc être une

excellente alternative pour assembler *de novo* des génomes complexes, sans pour autant avoir recours à de très fortes profondeurs de séquençage, notamment concernant la génération de données *PacBio* qui reste encore une technologie de séquençage coûteuse (English et al., 2012; Rhoads & Au, 2015). A titre d'exemple, l'assemblage du génome du vers plat *Macrostomum lignano*, dont le taux de répétition estimé est plus faible qu'*A. nasatum* (environ 51% du génome), avait un N50 de seulement 64 kb malgré une profondeur de séquençage *PacBio* de 130x par nucléotide (contre environ 35x dans notre cas). Cette stratégie combinant la haute fiabilité et le faible coût par nucléotide des *reads Illumina* et les longs *reads PacBio* semble donc être une méthode plus abordable efficace pour assembler *de novo* des génomes complexes et fortement répétés.

II. Comparaison avec des génomes de crustacés déjà publiés

Jusqu'à présent, si plus de 60 000 espèces de crustacés ont été décrites (Zhang, 2013), seulement treize génomes sont disponibles dans les banques de données publiques (banques de données de NCBI consultée le 18/09/2017). De plus parmi ces treize génomes, seulement quatre d'entre eux ont fait l'objet de publications : (i) le génome de *Neocaridina denticulata* (Decapoda, Kenny et al., 2014), (ii) le génome d'*Eriocheir sinensis* (Song et al., 2016), (iii) le génome de *Parhyale hawaiiensis* (Amphipoda, Kao et al., 2016) et (iv) le génome de *Daphnia pulex* (Branchiopoda, Colbourne et al., 2011). Si le génome du branchiopode est de petite taille (environ 0.2 Gb), les trois autres génomes sont beaucoup plus grands (*N. denticulata* : 1.28 Gb ; *E. sinensis* : 1.12 Gb ; *P. hawaiiensis* : 3.63 Gb). Le premier génome de crustacé publié est le génome de l'organisme modèle *D. pulex* (Colbourne et al., 2011). Si cet assemblage est d'excellente qualité (N50 de 642 kb pour un total de 5 191 séquences assemblées, généré à partir de données de séquençage Sanger), il est cependant notable que près de 19.3% de cet assemblage est composé de nucléotides indéterminés (Colbourne et al., 2011). Le génome de *N. denticulata* a quant à lui été généré uniquement à partir de données *Illumina* (paired-end) avec une profondeur de séquençage par nucléotide estimée à 12x. Il en a résulté un assemblage fortement fragmenté en plus de 3.3 millions de *contigs*, avec un N50 égal à 400 pb (Kenny et al., 2014). Ces statistiques sont d'ailleurs parfaitement comparables à celles obtenues pour l'assemblage d'*A. nasatum* uniquement réalisé à partir des *reads Illumina*. Le génome d'*E. sinensis* a lui aussi été généré à partir de données *Illumina*, en associant à la fois des bibliothèques d'*insert* courts et longs (Song et al., 2016) avec une forte profondeur de séquençage (environ 231x en cumulant l'ensemble des données). L'assemblage réalisé est composé de près de 1.77 M de séquences, dont 17 553 d'entre-elles sont supérieures à 2 kb, avec un N50 estimé à 111.7 kb (Song et al., 2016). Cet assemblage contient donc un très grand nombre de séquences de petite taille malgré une valeur de N50 élevée, indiquant que la majorité des nucléotides assemblés sont contenus dans des séquences de grande taille. Néanmoins, cet assemblage contient plus de 12% de nucléotides indéterminés (soit un total d'environ 136 Mb), ce qui représente une part non négligeable du génome. Enfin, le génome de *P. hawaiiensis* a quant à lui été généré à partir de données *Illumina*

paired-end et *mate-pair* avec une forte profondeur de séquençage (estimée à 115x en cumulant les données *paired-end* et *mate-pair*). Le génome ainsi assemblé est composé de 133 035 séquences, avec un N50 de 81 kb (Kao et al., 2016). Si cet assemblage présente plus de 5 fois plus de séquences que l'assemblage final d'*A. nasatum*, le N50 reste similaire entre ces deux génomes. En revanche, comme dans le cas du génome d'*E. sinensis*, le génome assemblé de *P. hawaiiensis* est composé d'une importante part de nucléotides indéterminés (30% des nucléotides du génome, soit environ 1.1 Gb). Ceci indique que malgré un assemblage peu fragmenté avec seulement des *reads Illumina* pour les deux génomes et ce à l'aide d'importantes profondeurs de séquençage, une large part de ces génomes demeure totalement inconnue et n'a pas pu être déterminée uniquement à partir de *reads Illumina*. Une des grandes difficultés de ces différents génomes publiés est la large part de nucléotides indéterminés présents dans chacun d'eux, malgré parfois de très fortes profondeurs de séquençage (*D. pulex* : 19.3% de N; *N. dentriculata* : nombre de N inconnu mais forte fragmentation; *E. sinensis* : 12.2 % de N; *P. hawaiiensis* : 30% de N). Ainsi, les génomes d'*A. nasatum* et d'*A. vulgare* (Chebbi et al., en préparation) représentent des ressources génomiques de haute qualité (en se basant à la fois sur leur N50 et leur nombre de nucléotides indéterminés, dans les deux cas inférieurs à 0.5%) pour l'ensemble des crustacés.

III. Comparaison des génomes d'*A. nasatum* et d'*A. vulgare*

Si les génomes d'*A. nasatum* et d'*A. vulgare* présentent un contenu en "core genes" d'arthropodes très similaires, le génome d'*A. vulgare* compte en revanche près de 500 Mb supplémentaires par rapport à celui d'*A. nasatum*. Seulement 5.4% de cette différence de taille (27 Mb) s'explique par un contenu en gènes différent. La taille moyenne des gènes prédits (introns et exons) est assez proche entre les deux espèces (8.6 kb pour *A. vulgare* et 9.4 kb pour *A. nasatum*), et est plus élevée que la taille moyenne des gènes de *D. pulex* (2 kb, Colbourne et al., 2011) mais aussi plus faible que la taille moyenne des gènes de *P. hawaiiensis* (20 kb, Kao et al., 2016). Cette différence s'explique par la taille des introns, puisque si les deux génomes d'isopodes présentent une taille d'intron moyenne similaire (1.5 kb et 1.4 kb pour respectivement *A. vulgare* et *A. nasatum*), elle est aussi plus élevée que celle observée chez *D. pulex* (0.3 kb, Colbourne et al., 2011) et plus faible que celle mesurée chez *P. hawaiiensis* (5.4 kb, Kao et al., 2016). Une hypothèse pouvant expliquer la forte variation de taille de ces introns peut résider dans des taux de recombinaison fortement variables entre ces espèces. Les introns sont généralement plus nombreux et plus longs dans les régions génomiques les moins recombinantes (Lynch, 2002), car les processus de sélection y sont moins efficaces (Carvalho & Clark, 1999). Si le taux de recombinaison de *P. hawaiiensis* est inconnu, celui de *D. pulex* est connu pour être particulièrement élevé (Haag et al., 2009), ce qui pourrait expliquer les introns de petite taille observés chez cette espèce. Selon cette hypothèse, les taux de recombinaison d'*A. vulgare* et d'*A. nasatum* seraient donc assez proches puisque leur taille moyenne d'intron ne diffère que très peu. De plus, les tailles d'introns similaires entre *A. vulgare* et *A. nasatum* suggèrent que les ETs ayant transposés après la divergence de ces deux espèces auraient transposé préférentiellement dans des régions ne correspondant pas à des introns.

Néanmoins, comme évoqué précédemment, le contenu en gènes d'*A. nasatum* et d'*A. vulgare* n'est pas suffisant pour expliquer la différence de taille de génome observée. En effet, la majeure partie des nucléotides supplémentaires présents chez *A. vulgare* provient d'éléments répétés. En revanche, le génome d'*A. vulgare* compte près de 4 500 gènes supplémentaires par rapport au génome d'*A.*

nasatum. Il est tout d'abord possible qu'une part des gènes supplémentaires détectés chez *A. vulgare* soit un artéfact du processus d'annotation. En effet, de nombreux gènes dont la fonction est inconnue sont dans la différence observée entre ces deux génomes. Certaines copies ont pu être mal identifiées et attribuées à tort à des gènes non ETs (mais sans fonction connue), et puisque le génome d'*A. vulgare* contient plus d'ETs que celui d'*A. nasatum*, ce biais peut être plus accentué chez *A. vulgare*, augmentant artificiellement le nombre de gènes identifiés. Néanmoins, il apparaît plausible que le génome d'*A. vulgare* ait subi de plus nombreuses duplications de gènes que celui d'*A. nasatum*. En effet, les deux génomes partagent 8 036 groupes de gènes orthologues, qui représentent 10 894 gènes pour *A. vulgare* et 9 496 gènes pour *A. nasatum*, soient 1398 gènes supplémentaires présents dans le génome d'*A. vulgare*. De plus ces gènes comptent une faible part de protéines dont la fonction est inconnue (Tableau 12). Ainsi, il est plausible qu'après la divergence entre les deux isopodes terrestres, le génome d'*A. vulgare* ait subi davantage de duplication de gènes que celui d'*A. nasatum*. Ces duplications de gènes sont connues pour potentiellement améliorer la plasticité phénotypique des espèces (Magadum et al., 2013). Celle-ci est d'ailleurs susceptible de conférer un meilleur potentiel adaptatif, notamment lors de la colonisation de nouveaux milieux, comme cela a été encore récemment démontré chez le puceron *Myzus persicae* (Mathers et al., 2017). Néanmoins, l'hypothèse d'une meilleure plasticité phénotypique d'*A. vulgare* comparée à celle d'*A. nasatum* ne semble pas à privilégier, car ces deux espèces sont connues pour partager les mêmes aires de distribution, et sont donc soumises potentiellement aux mêmes contraintes environnementales (Vandel, 1962).

Plus généralement, les duplications de gènes constituent une force majeure créatrice d'innovations biologiques, et sont souvent associées à des phénomènes de spéciation (Taylor et al., 2001). Cette force innovatrice a d'ailleurs été de nombreuses fois caractérisée au sein du règne végétal (Holub, 2001; Lespinet et al., 2002; Kim et al., 2004; Carretero-Paulet & Fares, 2012) comme animal (Otto & Whitton, 2000; Hoegg et al., 2004). Ainsi, ce large écart en nombre de gènes a pu contribuer à la divergence entre *A. vulgare* et *A. nasatum*. Quoi qu'il en soit, de plus amples investigations sont nécessaires afin de mieux comprendre l'histoire évolutive des gènes présents chez ces deux espèces d'isopodes terrestres.

Comme évoqué plus haut, le large écart de gènes identifiés entre *A. nasatum* et *A. vulgare* ne permet pas pour autant d'expliquer la différence de taille de génome entre ces deux espèces. Cette différence ne peut pas non plus être expliquée par un contenu en séquences satellites différent, puisque les génomes d'*A. vulgare* et d'*A. nasatum* comptent une quantité d'éléments répétés en tandem similaire (respectivement 301 Mb et 290 Mb), mais bien par un contenu en ETs différents (respectivement 872 Mb et 477 Mb).

L'annotation fine réalisée sur les génomes d'*A. nasatum* et d'*A. vulgare* a permis d'établir que sur les 53 superfamilles détectées dans ces deux génomes, 51 sont communes à *A. nasatum* et *A. vulgare* (Figure 30). Ceci est cohérent avec le temps de divergence estimé entre ces deux espèces (environ 25 Ma, correspondant à près de 6% de divergence), puisque l'ensemble des 51 familles communes contiennent des copies d'ETs ayant un pourcentage de divergence avec leur séquence consensus supérieur à 6%. Ces familles d'ETs étaient donc déjà présentes au sein de l'ancêtre commun à ces deux espèces. Afin de confirmer cette hypothèse, il apparaît nécessaire d'identifier des copies d'ETs orthologues aux deux espèces. Selon nos analyses, les ETs les plus représentés de ces deux espèces appartiennent à la catégorie des LINEs (respectivement 20.5% et 24.3% de chaque génome). Ces LINEs sont répartis dans 14 familles communes aux deux espèces. Les LINEs sont en revanche bien moins représentés chez *D. pulex* et *P. hawaiiensis*, puisqu'ils ne représentent respectivement que 0.63% et 7.04% de ces génomes (Colbourne et al., 2011; Kao et al., 2016). Les transposons à ADN représentent 12.4% du génome d'*A. nasatum* et 16.6% du génome d'*A. vulgare* (et appartiennent aux 17 mêmes familles dans les deux génomes) ce qui est là aussi bien supérieur à leur proportion dans les génomes de *D. pulex* (0.07%, Colbourne et al., 2011) et de *P. hawaiiensis* (4.39%, Kao et al., 2016). La dernière catégorie d'ET fortement présente dans les génomes d'*A. nasatum* et d'*A. vulgare* sont les rétrotransposons à LTR qui représentent respectivement 5% et 8% de chaque génome (répartis dans les 10 mêmes familles pour les deux espèces). Ces proportions sont assez proches de celles observées chez *D. pulex* (7.97%, Colbourne et al., 2011) mais là encore bien supérieures à celles observées chez *P. hawaiiensis* (0.5%, Kao et al., 2016). Enfin, 12 familles de SINEs ont été identifiées chez *A. nasatum* et *A. vulgare*, dont 10 sont communes aux deux

espèces. Ceci suggère une introduction des familles de SINEs postérieure à la divergence entre *A. nasatum* et *A. vulgare*.

Si les larges écarts de proportions d'ETs présents chez *A. nasatum* et *A. vulgare* et ceux présents chez *D. pulex* ne sont guère surprenants compte tenu de la petite taille du génome du branchiopode (environ 200 Mb, Colbourne et al., 2011), les différences observées avec le génome de *P. hawaiiensis* sont plus surprenantes. Malgré une taille de génome assemblé d'environ 3.6 Gb, les ETs identifiés ne représentent que 15% du génome de *P. hawaiiensis*. Il faut aussi noter que les ETs détectés mais n'ayant pas pu être attribués à une famille d'ETs constituent 22% du génome assemblé de l'amphipode. Des biais méthodologiques peuvent constituer une première explication aux écarts de proportion d'ETs détectés entre les génomes d'isopodes terrestres et celui de *P. hawaiiensis*. En effet, l'annotation des ETs de *P. hawaiiensis* n'a été réalisée qu'à partir d'une approche par homologie (Kao et al., 2016), ce qui a pu diminuer fortement le nombre de séquences identifiées comme ETs en comparaison avec l'approche utilisée pour les génomes d'*A. nasatum* et d'*A. vulgare* (basée en plus sur une identification prédictive avec *TE-class*; Abrusán et al., 2009). De plus, le génome de l'amphipode est composé de plus de 30% de nucléotides indéterminés, ce qui biaise probablement la contribution de chaque classe d'ETs pour le génome de *P. hawaiiensis*. Enfin, la plus forte fragmentation du génome de l'amphipode constitue là aussi un biais dans la détection d'ETs, pouvant contribuer à une sous-estimation du nombre d'ETs identifiés (Tang, 2007; Lerat, 2010).

La différence de taille des génomes d'*A. nasatum* et d'*A. vulgare* s'explique donc principalement par une plus forte abondance en ETs présents dans le génome d'*A. vulgare*. Une telle différence peut s'expliquer soit par une plus forte activité de transposition au sein du génome d'*A. vulgare*, soit par un taux de délétion plus fort dans le génome d'*A. nasatum*. Cependant, si le taux de délétion était sensiblement plus fort chez *A. nasatum* que chez *A. vulgare*, alors il serait attendu d'observer un même taux de délétion au sein des éléments répétés en tandem, ce qui n'est pas le cas (Figure 29). De plus, la taille des introns d'*A. nasatum* est très proche de la taille des introns d'*A. vulgare*. Un taux de délétion plus fort chez *A. nasatum* aurait aussi pour conséquence de réduire la taille de ces introns, ce qui n'est pas non plus le cas. Enfin, il est notable que le pic de transposition observé

chez *A. vulgare* est légèrement plus récent que celui observé chez *A. nasatum*, et semble postérieur au temps de divergence estimé entre ces deux espèces (Figure 30). Ces derniers éléments semblent en accord avec l'hypothèse d'une expansion de l'activité de transposition postérieure à la divergence entre *A. vulgare* et *A. nasatum*. Afin de tester ces hypothèses, il sera nécessaire de comparer les taux de délétion des deux génomes et d'établir les relations d'orthologie entre les différentes copies issues de chaque génome.

Enfin, seulement 19 insertions du génome de *Wolbachia* ont été détectées dans le génome d'*A. nasatum*, tandis que 138 ont pu être décelées dans le génome d'*A. vulgare*. Seulement 2 de ces 19 insertions ont pu clairement être caractérisées phylogénétiquement comme étant proches du génome de la souche *wVulC* (qui est connue pour infecter naturellement les populations d'*A. vulgare*, Cordaux et al., 2004). Les insertions ayant pu être caractérisées dans le génome d'*A. vulgare* ont elles aussi pu être rattachées à la souche *wVulC* (Chebbi et al, en préparation). Cependant, les insertions détectées dans les deux génomes ne se sont pas avérées être orthologues. De plus, la souche naturelle infectant *A. nasatum*, à savoir *wNas* (Juchault & Legrand, 1989), est apparentée phylogénétiquement à la souche *wVulC* (Cordaux et al., 2012), dont le génome est disponible en interne au laboratoire (contrairement à la souche *wNas*). Il est donc possible que les insertions détectées dans le génome d'*A. nasatum* soient en réalité non pas des insertions du génome de *wVulC* mais des insertions du génome de *wNas*. Un argument appuyant cette hypothèse est que les deux insertions de *Wolbachia* caractérisées chez *A. nasatum* présentent 4.6% et 8.1% de divergence avec la souche *wVulC*, tandis que les insertions de *Wolbachia* détectées dans le génomes d'*A. vulgare* ne présentent qu'en moyenne 2.7% de divergence avec *wVulC* (Chebbi et al., en préparation). Le faible nombre d'insertions détectées dans le génome d'*A. nasatum* en comparaison du nombre d'insertions détectées chez *A. vulgare* peut tout d'abord s'expliquer par un taux de délétions supérieur dans le génome d'*A. nasatum* en comparaison à *A. vulgare*, éliminant ainsi plus efficacement les insertions du génome bactérien. Une autre hypothèse serait que la dérive génétique est plus forte chez *A. vulgare* que chez *A. nasatum*, cependant les tailles de populations naturelles de ces deux espèces sont encore méconnues. Enfin, une autre explication réside dans la stabilité de l'infection d'*A. nasatum* à

wNas en comparaison à celle de *wVulC*. En effet, la souche *wNas* est connue pour être instable au sein des populations naturelles d'*A. nasatum* (Juchault & Legrand, 1989), et cette instabilité a pu limiter le nombre d'insertions du génome de *Wolbachia* dans le génome d'*A. nasatum*. L'instabilité de la souche *wNas* est discutée en détail dans le Chapitre 2 (page 113). Plus globalement, la présence d'insertion de *Wolbachia* dans les génomes d'*A. nasatum* et d'*A. vulgare* n'est pas surprenante car les transferts horizontaux de séquences de *Wolbachia* vers leurs hôtes sont relativement fréquents et répandus, à tel point que des gènes de *Wolbachia* sont présents dans près de 70% des génomes d'arthropodes et de nématodes (Dunning Hotopp et al., 2007; Dunning Hotopp, 2011; Robinson et al., 2013). D'ailleurs, ces insertions de *Wolbachia* dans les génomes hôtes peuvent avoir un impact phénotypique conséquent chez l'hôte, notamment concernant le déterminisme du sexe de l'hôte, comme démontré avec l'élément *f* chez certaines populations d'*A. vulgare* (Leclercq et al., 2016). Dans cet exemple, l'autosome porteur de cette insertion est devenu un nouveau chromosome sexuel semblable à un chromosome W (Leclercq et al., 2016).

IV. Perspectives de recherche

Le génome de référence annoté d'*A. nasatum* constitue une véritable opportunité pour les recherches à venir, et ce dans de nombreux domaines de la biologie évolutive. La richesse en ETs décelée chez les deux génomes d'isopodes terrestres constitue par exemple un terrain d'étude prometteur pour la caractérisation de transferts horizontaux entre ces deux espèces. La majorité de ces transferts connus d'un organisme eucaryote à un autre sont des transferts d'ETs (Schaack et al., 2010). Les transposons ainsi que les rétrotransposons à LTR, de par leurs mécanismes de réplication, sont les ETs le plus souvent impliqués dans des événements de transferts horizontaux (Wallau et al., 2012). Or ces deux catégories d'ET constituent une large part des deux génomes d'isopodes terrestres, puisqu'ils représentent respectivement 17.4% (5% de rétrotransposons à LTR + 12.4% de transposons à ADN) et 24.6% (8% de rétrotransposons à LTR + 16.6% de transposons à ADN) des génomes d'*A. nasatum* et d'*A. vulgare*. Des transferts horizontaux d'ETs ont d'ailleurs déjà été mis en évidence chez les isopodes par le passé (Dupeyron et al., 2014). Cette étude s'était focalisée sur deux transposons à ADN appartenant à la superfamille *Mariner*, caractérisés tout d'abord chez *A. vulgare* puis chez plusieurs autres espèces d'isopodes (dont *A. nasatum*). Les deux familles d'ETs étudiées sont d'ailleurs considérées comme récemment actives, puisque certaines copies n'étaient pas présentes dans toutes les populations d'*A. vulgare* étudiées (Dupeyron et al., 2014). Ceci se révèle en accord avec nos résultats, puisque les ETs présentant l'activité la plus récente identifiés dans les génomes d'*A. nasatum* et d'*A. vulgare* sont majoritairement des transposons à ADN (Figure 30). Ainsi, un polymorphisme de présence/absence, associé à une activité récente de transposition, suggère que de nombreux autres transferts horizontaux d'ETs ont pu avoir lieu au cours de l'évolution des isopodes terrestres (Dupeyron et al., 2014). D'ailleurs, une étude récente à large échelle, menée sur 195 génomes d'hexapodes, a caractérisé plus de 2 200 événements de transferts horizontaux d'ET qui se sont produits il y a tout au plus 10 millions d'années (Peccoud et al., 2017). Avec les génomes d'*A. nasatum* et d'*A. vulgare*, il pourrait être possible d'évaluer la quantité de transferts entre ces deux espèces mais aussi entre ces deux isopodes terrestres et les hexapodes.

Les assemblages de références d'*A. nasatum* et d'*A. vulgare* constituent aussi des ressources de choix pour la caractérisation de virus endogènes (EVE pour *Endogenous Viral Elements*) présents dans ces génomes. Depuis plusieurs années, il est devenu clair que de nombreux génomes viraux se sont intégrés dans les génomes de leurs hôtes eucaryotes, modifiant par ailleurs notre compréhension et notre vision de l'évolution des virus et de l'évolution des interactions hôtes/virus (Katzourakis & Gifford, 2010; Feschotte & Gilbert, 2012). Chez les isopodes terrestres, une première étude effectuée sur *A. vulgare* avait révélé une forte diversité de virus endogènes, puisque les EVEs détectés ont été assignés à cinq groupes viraux différents (Thézé et al., 2014). Une seconde étude a permis de confirmer cette forte diversité virale rencontrée chez les isopodes terrestres puisque les mêmes groupes viraux ont été identifiés chez *A. nasatum* (Metegnier et al., 2015). Il serait donc intéressant d'appliquer la même procédure d'identification que celle réalisée lors des études précédentes (Thézé et al., 2014; Metegnier et al., 2015) sur les nouveaux assemblages d'isopodes terrestres, et même de l'élargir aux nouveaux génomes de crustacés publiés depuis tels que l'amphipode *P. hawaiiensis* (Kao et al., 2016) ou que le décapode *E. sinensis* (Song et al., 2016). Les ressources génomiques générées au cours de cette première partie de thèse constituent une opportunité pour mieux comprendre les patterns d'endogénisation virales, notamment au sein d'un groupe disposant de peu de ressources génomiques tel que les crustacés. Une telle démarche pourrait d'ailleurs permettre de caractériser de nouveaux virus exogènes pathogènes de crustacés (Metegnier et al., 2015).

Les génomes assemblés d'*A. nasatum* et d'*A. vulgare* représentent aussi une véritable opportunité pour caractériser génétiquement l'un des événements écologiques majeurs de l'histoire du Vivant, à savoir une transition du milieu marin vers le milieu terrestre. En effet, l'adaptation à un environnement terrestre nécessite des innovations à la fois morphologiques et physiologiques, telles que réduire la dépendance à l'eau ou encore résister aux plus fortes variations physicochimiques de l'environnement (Selden & Edwards, 1989). L'ordre des isopodes, qui contient à la fois des espèces aquatiques et terrestres, représente donc un modèle important pour comprendre les mécanismes impliqués dans ce processus de sortie des eaux (Broly et al., 2013). Si les adaptations physiologiques

et morphologiques à la vie terrestre ont été largement étudiées chez les cloportes (Warburg, 1993; Hornung, 2011), ces adaptations n'ont jamais pu être caractérisées génétiquement. De plus, les isopodes terrestres présentent une vaste gradation des adaptations (cuticule perméable, structure des pléopodes, comportements d'agrégation) permettant d'étudier les mécanismes de terrestrialisation (Vandel, 1943; Hornung, 2011; Broly et al., 2013). Ainsi l'apport de ces deux génomes constitue là encore une véritable opportunité de caractériser les gènes potentiellement impliqués dans les processus de terrestrialisation. Une première piste serait par exemple d'étudier le répertoire de gènes liés aux récepteurs olfactifs au sein des crustacés aquatiques et terrestres. En effet, la terrestrialisation est associée avec de profonds changements de gènes impliqués dans la synthèse des récepteurs olfactifs, car les molécules olfactives hydrosolubles et aériennes diffèrent fortement et ne peuvent être liées efficacement aux mêmes récepteurs (Niimura & Nei, 2005; Niimura, 2009; Vieira & Rozas, 2011; Lozano-Fernandez et al., 2016).

Enfin, ces données constituent aussi une base pour l'étude d'un autre aspect important de la biologie évolutive, à savoir l'évolution des mécanismes de déterminisme du sexe. Chez ces deux espèces, le sexe est déterminé par la présence de chromosomes sexuels. Ainsi, *A. nasatum* est connu pour être une espèce à hétérogamétie mâle (Juchault & Legrand, 1979) tandis qu'*A. vulgare* est une espèce à hétérogamétie femelle (Juchault & Legrand, 1972). Ces ressources génomiques ont donc servi de base pour l'étude à l'échelle moléculaire des mécanismes impliqués dans les transitions de système d'hétérogamétie, ce qui fait l'objet du chapitre suivant.

Chapitre 2

Identification de séquences Y- spécifiques et caractérisation du chromosome Y d'*A. nasatum*

Introduction

L'étude à l'échelle génomique des hétérochromosomes (qu'il s'agisse du chromosome Y ou du chromosome W) représente un véritable défi, notamment en raison des difficultés rencontrées lors du séquençage, de l'assemblage et de l'identification de ces séquences dans des génomes. Ils sont encore peu étudiés et décrits (en dehors de certains organismes modèles) et pourtant, ces chromosomes sont susceptibles d'être impliqués dans de nombreuses fonctions vitales pour l'organisme (Tomaszkiewicz et al., 2017). Le chromosome Y des mammifères contient par exemple le gène *SRY*, soit le gène maître du déterminisme du sexe, mais aussi de nombreux gènes impliqués dans la spermatogenèse (Vallender & Lahn, 2004). Dans le cas de *D. melanogaster*, des variations génétiques observées sur le chromosome Y peuvent affecter l'expression de milliers de gènes, tant situés sur le chromosome X que sur les autosomes (Jiang et al., 2010; Lemos et al., 2010). Il apparaît clair que la connaissance de ces séquences, notamment issues des organismes non-modèles, est nécessaire pour la compréhension de la complexité des génomes et de leur évolution.

Pour faciliter l'identification de ces séquences issues d'organismes dont le génome n'est pas connu, plusieurs méthodes basées sur l'enrichissement en chromosome Y (ou W) ont été développées. Elles peuvent concerner des microdissections de chromosomes suivies d'une amplification de l'hétérochromosome ainsi sélectionné (Zhou & Hu, 2007), ou encore un tri des chromosomes par cytométrie en flux (Doležel et al., 2012), voire la capture ciblée de séquences à partir de sondes mises au point sur des espèces apparentées à l'espèce d'intérêt, dont les séquences de l'hétérochromosome sont connues (Hallast et al., 2016). Cependant, toutes ces méthodes nécessitent soit une connaissance préalable du génome d'espèces apparentées, soit que les chromosomes sexuels soient morphologiquement différenciés. Dans le cas de l'isopode terrestre *A. nasatum*, de telles approches apparaissent inefficaces puisque d'une part, si le génome d'une espèce apparentée a été généré en parallèle de celui d'*A. nasatum*, à savoir le génome d'*A. vulgare*, il s'agit là d'une espèce à hétérogamétie femelle

(Juchault & Legrand, 1972) contrairement à *A. nasatum*. D'autre part, les chromosomes sexuels d'*A. nasatum* (X et Y) ne sont pas morphologiquement distinguables (Juchault & Legrand, 1979; Juchault & Rigaud, 1995).

Une stratégie alternative permettant une identification de ces séquences liées au chromosome Y est basée sur la comparaison entre les données de séquençage générées indépendamment pour des mâles et des femelles. Cette méthode nécessite tout d'abord d'assembler le génome du sexe hétérogamétique (cf : Chapitre 1, page 53), puis de le comparer aux données de séquençage du sexe homogamétique. En effet, les séquences issues de l'hétérochromosome ont un ratio de copies différent entre les mâles et les femelles : les séquences spécifiques au chromosome Y sont hémizygotiques chez les mâles tandis qu'elles sont totalement absentes chez les femelles. Il est alors possible d'utiliser la profondeur de séquençage (nombre de fois qu'un même nucléotide a été séquencé) entre les deux sexes pour caractériser ces séquences. C'est ce que proposent les approches basées sur les ratios de couverture (ratio entre la couverture des *reads* d'un sexe sur le génome du sexe hétérozygote par rapport à la couverture des *reads* de l'autre sexe sur ce même génome), telles que le *Chromosome Quotient* (CQ, Hall et al., 2013). Ce type d'approche a d'ailleurs été éprouvé sur de nombreux modèles, tels que des diptères (Vicoso & Bachtrog, 2015; Hall et al., 2016), des serpents (Vicoso et al., 2013a), des oiseaux (Chen et al., 2012; Zhou et al., 2014; Smeds et al., 2015) ou encore des mammifères (Bidon et al., 2015). Une approche complémentaire, nommée *Y chromosome Genome Scan* (YGS, Carvalho & Clark, 2013) est quant à elle basée sur un principe de soustraction. Elle consiste à comparer les *k-mers* (séquences nucléotidiques de longueur *k*) présents au sein du génome assemblé du sexe hétérogamétique et de les comparer avec les *k-mers* issus des *reads* du sexe homogamétique. Les séquences qui présentent une majorité de *k-mers* propres au génome hétérogamétique sont ainsi assignées par cette méthode à des séquences spécifiques du chromosome Y. Cette approche soustractive a d'ailleurs été appliquée avec succès chez *D. melanogaster*, *Homo sapiens* (Carvalho & Clark, 2013) ou encore le coléoptère *Rhodnius prolixus* (Koerich et al., 2016). Si ces méthodes ont démontré leur efficacité dans le cas de chromosomes sexuels morphologiquement différenciés, elles n'ont encore été que peu éprouvées dans le

cas d'espèces aux chromosomes sexuels faiblement différenciés, tels que ceux d'*A. nasatum*.

Comme indiqué lors de l'introduction générale de ce manuscrit (voir page 41), la majorité des espèces appartenant au groupe des Oniscidea sont potentiellement infectées par des souches de *Wolbachia* féminisantes (Bouchon et al., 1998, 2008; Cordaux et al., 2012). Ceci est notamment le cas de l'espèce *A. nasatum*, qui est connue pour être potentiellement infectée par la souche féminisante *wNas* (Juchault & Legrand, 1989). Selon le scénario évolutif décrit en introduction générale (voir page 52), les conflits génétiques entre l'hôte et l'endosymbiote féminisant, qui induisent des changements récurrents de positions des allèles déterminant le sexe, contraignent à l'homomorphie des chromosomes sexuels (Charlesworth, 1991, 1996). D'ailleurs dans le cas d'espèces dont le déterminisme du sexe suit un système XX/XY, des chromosomes sexuels hétéromorphes pourraient être délétères car la présence de *Wolbachia* féminisantes tend à éliminer le chromosome sexuel femelle (soit le chromosome W dans un système ZZ/ZW ou le chromosome X dans un système XX/XY) (Taylor, 1990). Ainsi à l'équilibre, il est prédit que les individus infectés tendent à avoir un déterminant chromosomique homozygote mâle, et donc être des individus ZZ ou YY selon le système d'hétérogamétie considéré (Rigaud, 1997). L'élimination du chromosome sexuel "femelle" (X ou W) au sein des populations infectées relève ainsi d'une influence directe de *Wolbachia* sur l'évolution du déterminisme du sexe de son hôte. La mise au point de marqueurs génétiques liés au chromosome Y pourrait donc permettre de suivre le pattern de transmission du chromosome Y en relation avec la présence de *Wolbachia* féminisante de façon empirique. Dans le cas d'une espèce aux chromosomes sexuels indifférenciés comme *A. nasatum* (Juchault & Rigaud, 1995), il est donc attendu que le chromosome X disparaisse des populations infectées par *Wolbachia*, permettant l'introduction d'individus YY viables. Un tel suivi permettrait aussi de mieux caractériser le niveau d'homomorphie fonctionnelle des chromosomes X et Y.

L'objectif de ce second chapitre de thèse était donc d'identifier et de caractériser des régions spécifiques au chromosome Y d'*A. nasatum*. Cette recherche a été effectuée à partir d'une approche de génomique comparative entre les différents assemblages du génome mâle d'*A. nasatum* et des données issues du

séquençage *Illumina* d'individus frères et sœurs appartenant à l'espèce *A. nasatum*. Une première phase exploratoire de recherche de séquence candidates Y-spécifiques a tout d'abord été réalisée sur le génome fortement fragmenté, assemblé uniquement à partir des données *Illumina*. Cette première recherche s'est révélée infructueuse car elle a conduit à la sélection de très nombreuses séquences candidates de petite taille. Ainsi, suite au séquençage de données *PacBio* et à un assemblage du génome moins fragmenté, de nouvelles recherches ont été effectuées afin de déterminer des séquences appartenant aux régions mâle-spécifiques du chromosome Y d'*A. nasatum*. Ces nouvelles analyses ont été effectuées selon deux approches. Tout d'abord l'approche du *Chromosome Quotient (CQ)* a été appliquée en utilisant les deux génomes assemblés à partir des données *Illumina* et *PacBio* (Tableau 13). Ce premier filtre de sélection a ensuite été complété par la recherche de séquences Y-spécifiques selon la méthode soustractive *YGS* (Carvalho & Clark, 2013). L'ensemble des analyses *in silico* réalisées afin de déterminer des séquences Y-spécifiques est résumée dans le Tableau 13. Les séquences candidates retenues par ces filtres *in silico* ont ensuite été testées par PCR afin de confirmer leur association au sexe mâle. Les *loci* candidats ont été testés par PCR tout d'abord sur des individus partageant les mêmes grands-parents que les individus séquencés en *PacBio*, puis sur des individus issus de lignées de laboratoire provenant de la même population naturelle, et enfin sur des individus issus de populations naturelles différentes des individus séquencés (Beauvoir-sur-Niort en France et Piriápolis en Uruguay). Enfin, les marqueurs Y-spécifiques mis au point ont permis le suivi du pattern de transmission du chromosome Y au sein de lignées infectées par la bactérie *Wolbachia*. Les analyses ont été réalisées sur des portées dont la mère était infectée par des *Wolbachia* féminisantes, et dont les descendants étaient exclusivement composés de femelles ou de mâles.

Tableau 13 : Résumé des approches utilisées pour la recherche de séquences Y-spécifiques chez *A. nasatum*. La première approche s'est donc focalisée sur le génome mâle d'*A. nasatum* uniquement assemblé à partir des données *Illumina*. Une analyse de ratio de profondeur d'alignement entre les données mâles et femelles a été réalisée. Une seconde version du génome a été générée à partir des données *PacBio*, en association avec les données *Illumina*, afin de générer un nouvel assemblage du génome avec *CeleraAssembler*. A partir de cet alignement, une recherche de séquences candidates Y-spécifiques a été réalisée à partir d'une approche d'alignement de *reads* (*CQ*) et d'une approche par comparaison de *k-mer* (*YGS*, testée selon deux paramètres). Enfin, une dernière version du génome a été générée à partir des données *PacBio* et *Illumina*, permettant de générer l'assemblage de référence du génome d'*A. nasatum*. Des analyses de *CQ* et d'*YGS* (18-mer) ont enfin été réalisées sur cette version finale du génome.

Type de données utilisées pour l'assemblage	Méthode d'assemblage	Analyses effectuées
<i>Illumina</i>	<i>SOAPdenovo</i>	- ratio de profondeur d'alignement
<i>Illumina</i> + <i>PacBio</i>	<i>CeleraAssembler</i>	- <i>CQ</i> - <i>YGS</i> (16-mer) - <i>YGS</i> (18-mer)
<i>Illumina</i> + <i>PacBio</i>	<i>DBG2OLC</i>	- <i>CQ</i> - <i>YGS</i> (18-mer)

Matériels et Méthodes

I. Recherche de séquences spécifiques au chromosome Y dans le génome d'*A. nasatum*

1. Recherche *in silico* par étude de ratio de couverture sur le génome assemblé avec les données *Illumina*

La première recherche *in silico* de séquences Y-spécifiques a été réalisée par comparaison de données de séquençage d'individus frères et sœurs (Figure 20) générées indépendamment. Elle consiste en une approche par un alignement (ou *mapping*) des *reads* des deux sexes sur un génome d'un sexe unique. Le ratio entre la quantité de *reads* de mâles et la quantité de *reads* de femelles alignés pour un *locus* donné a servi de base pour l'identification de séquences spécifiques au chromosome Y (Figure 31). En effet, le chromosome Y est absent du génome femelle et les *reads* de femelles ne devraient pas pouvoir être alignées sur des séquences assemblées mâle-spécifiques du chromosome Y. De plus, la région Y-spécifique étant supposée hémizygote, les *reads* formant des séquences Y-spécifiques devraient être deux fois moins abondantes que les *reads* formant un *contig/scaffold* autosomal diploïde. Cette première approche a été testée sur la première version du génome assemblé uniquement à partir des données *Illumina*.

Ainsi, les données *Illumina* de mâles et de femelles ont été alignées indépendamment sur le génome mâle avec *Bowtie2* (version 2.2.9, Langmead & Salzberg, 2012) en utilisant le paramètre d'alignement local "*Local-very-sensitive*" afin de maximiser la quantité de *reads* alignés. La profondeur moyenne de ces données sur le génome a été évaluée avec *bamtools* version 2.2.2, Barnett et al., 2011). Elle est de 39.33x pour les données femelles, et de 37.48x pour les données mâles. Une première sélection de séquences candidates Y-spécifiques a été effectuée en ne conservant que celles présentant au moins 100 pb contigües non-alignées par les données femelles. La stratégie d'alignement étant très permissive et pouvant mener à des alignements de *reads* erronés, toutes les séquences d'au

moins 100 pb présentant une couverture de 0 ou de 1 par nucléotide ont ainsi été conservées.

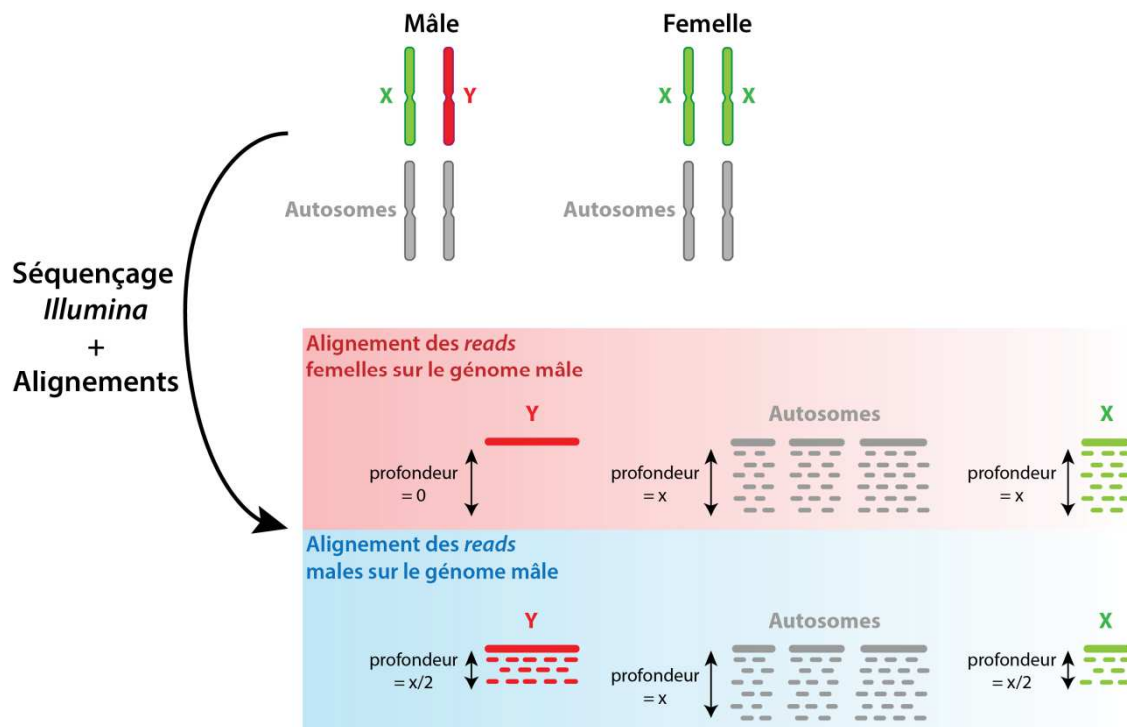


Figure 31 : Principe de l'étude par ratio de profondeur d'alignement pour la recherche de séquences Y-spécifiques. L'ADN génomique d'individus mâles et femelles est séquençé indépendamment. Le génome mâle a ensuite été assemblé, puis les *reads* Illumina mâles et femelles ont été alignés sur ce génome mâle. Une séquence candidate Y-spécifique (en rouge) est définie comme ne pouvant pas être alignée contre les *reads* femelles et doit présenter une profondeur d'alignement deux fois inférieure à celle d'un *locus* autosomale diploïde (ratio de profondeur de *reads* = $x/2$).

Les séquences d'origine bactérienne ont ensuite été filtrées par comparaison avec la base de données protéiques non-redondantes de NCBI (version de Décembre 2014) par *BLASTx* (version 2.2.23). L'assignation taxonomique des *hits* de *BLASTx* a été visualisée à partir du logiciel *MEGAN* (version 5.7.0, Huson et al., 2011). Toutes les séquences assignées à une origine procaryote ou fongique ont été éliminées si leur *e-value* était inférieure ou égale à 10^{-4} et ce sur au moins 80% de la longueur du *hit*.

Si la région spécifique au chromosome Y est absente du génome femelle, elle est aussi haploïde dans le génome d'un mâle XY. Cette haploïdie correspond à un ratio de profondeur d'alignement de $x/2$ où x correspond à la profondeur d'alignement d'un *locus* autosomique diploïde. La demie-couverture attendue des séquences décontaminées a été déterminée par une approche probabiliste à l'aide

de la suite *mixtools* (version 1.0.2, Benaglia et al., 2009) utilisée avec le logiciel R (version 3.1.1, R Core Team, 2014).

2. Recherche *in silico* par comparaison de ratio de couverture sur les génomes *PacBio* - Méthode du *Chromosome Quotient (CQ)*

Une approche analogue à celle décrite précédemment a été utilisée tout d'abord sur la version du génome *PacBio* assemblée avec *CeleraAssembler*, puis dans un second temps sur le génome de référence assemblé avec *DBG2OLC*. Cette méthode, appelée *Chromosome Quotient (CQ)*, se base sur le ratio entre le nombre de *reads* de mâles et le nombre de *reads* de femelles alignés sur un *locus* donné (ratio inverse de la méthode précédente) (Hall et al., 2013). Il se calcule de la façon suivante :

$$CQ_{(Si)} = \frac{F_{(Si)}}{M_{(Si)}}$$

avec :

$F_{(Si)}$ = nombre de *reads* femelles alignés pour une séquence (Si)

$M_{(Si)}$ = nombre de *reads* mâles alignés pour une même séquence (Si)

La différence principale réside dans la stratégie de cartographie, qui est non plus effectuée de façon locale, mais de façon stricte (*end-to-end*). L'objectif ici est de maximiser la précision de cette cartographie, et de s'assurer que les *reads* soient parfaitement cartographiées sur le génome mâle. La cartographie a une nouvelle fois été réalisée avec le logiciel *Bowtie2* (version 2.2.9, Langmead & Salzberg, 2012) avec l'option "*very-fast*". Seuls les *reads* ayant été cartographiées parfaitement (aucune inadéquation tolérée entre le *read* aligné et la séquence génomique assemblée) ont été conservés pour la suite des analyses. Les *reads* présentant des erreurs d'alignement ont été éliminés avec le logiciel *bamtools* (version 2.2.2, Barnett et al., 2011). La méthode *CQ* prédit trois catégories de résultats : (i) les séquences autosomales ont un score de *CQ* centré sur 1, (ii) les séquences spécifiques au chromosome X ont un score de *CQ* centré sur 2 (car la

ou les régions spécifiques au chromosome X est hémizygote chez les mâles XY), et (iii) les séquences Y-spécifiques ont un score de CQ centré sur 0 (Figure 32).

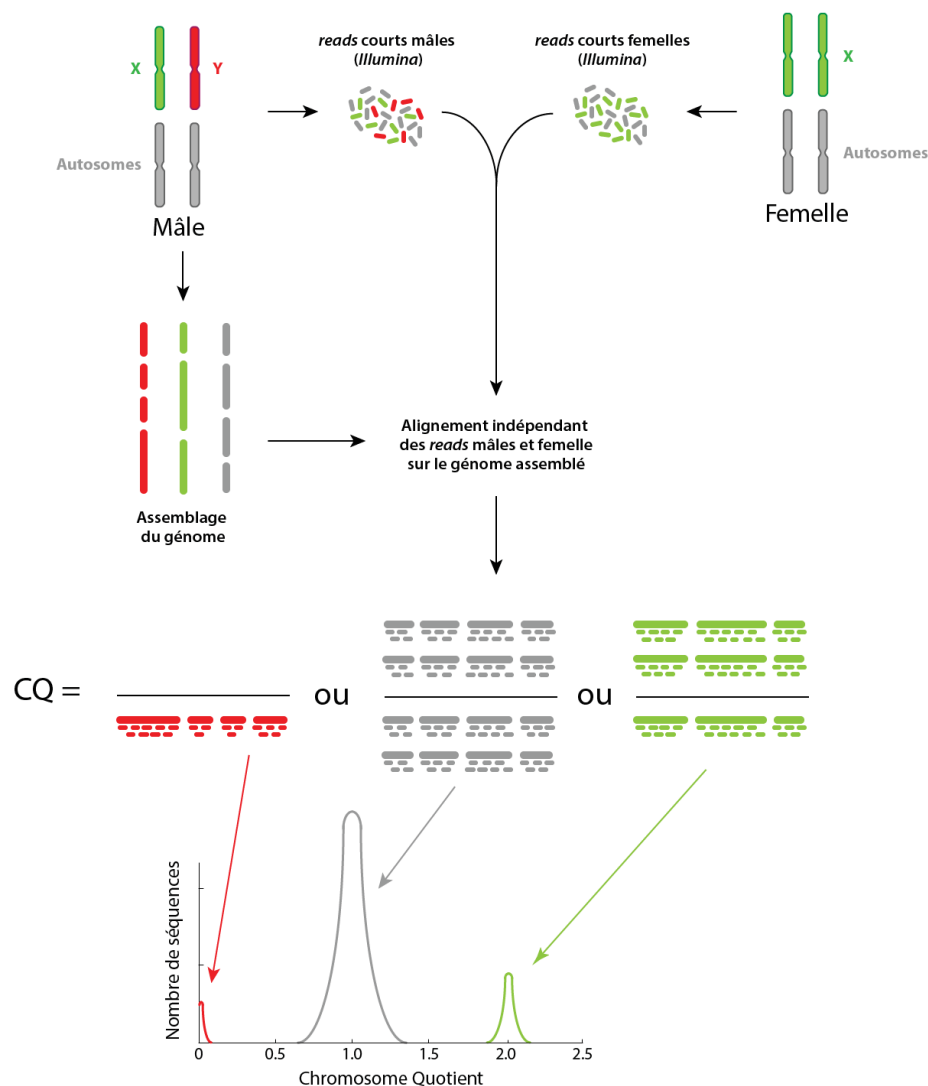


Figure 32 : Principe de la méthode de détection de séquences sexe-spécifiques CQ . Les reads *Illumina* de mâles et de femelles sont alignés sur le génome mâle avec une stratégie d'alignement stricte. Les séquences autosomales auront un score de CQ centré autour de la valeur 1, puisqu'elles sont autant représentées dans le génome mâle que dans le génome femelle (gris). Le chromosome X étant hémizygote chez le sexe mâle, les valeurs de CQ pour ce chromosome sont centrées autour de 2 (vert). Enfin, les reads femelles ne pouvant être alignées contre des séquences assemblées mâles (Y-spécifiques), ces contigs/scaffolds auront un score de CQ centré autour de 0 (rouge).

Les séquences ayant un score de *CQ* inférieur ou égal à 0.35 (seuil légèrement moins strict que celui proposé par les concepteur de cette approche, qui était de 0.3) ont été retenues comme candidates Y-spécifiques et soumises à d'autres analyses.

3. Recherche *in silico* par comparaison de *k-mer* sur les génomes *PacBio* - Méthode du *Y chromosome Genome Scanning (YGS)*

Une approche complémentaire a été utilisée pour détecter des séquences liées au chromosome Y. Cette méthode, appelée *YGS* pour *Y chromosome Genome Scanning* (Carvalho & Clark, 2013), se base sur une comparaison des *k-mers* uniques présents au sein du génome mâle et ceux présents au sein des *reads* de femelles (*Illumina*). Ainsi, les *contigs/scaffolds* candidats Y-spécifiques sont ceux pour lesquels les *k-mers* des *reads* de femelles ne correspondent pas aux *k-mers* présents au sein des séquences mâles assemblées (Figure 33).

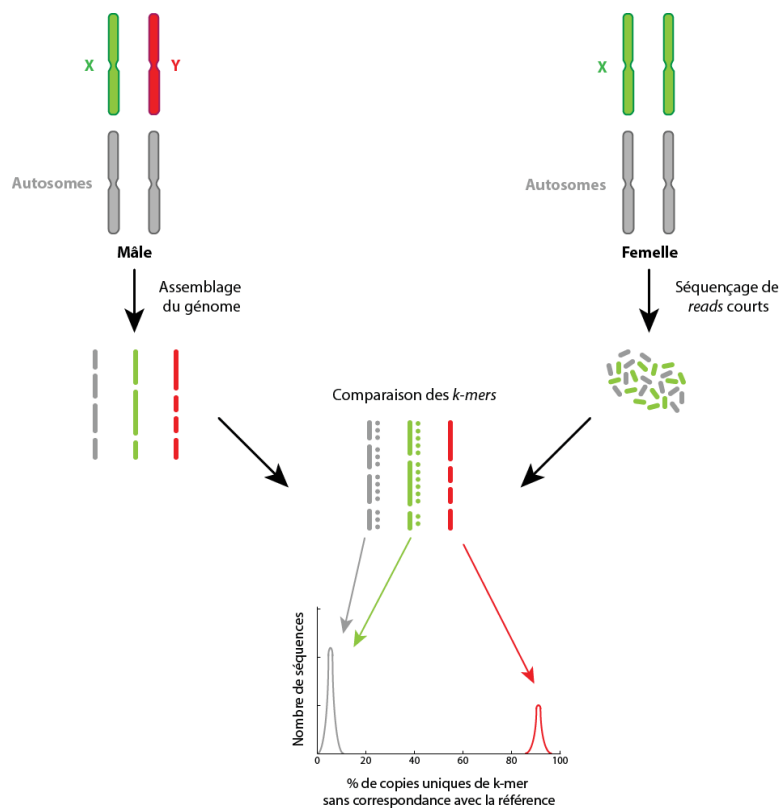


Figure 33 : Schéma de la méthode *Y chromosome Genome Scanning (YGS)*. Les séquences liées au Y sont identifiées par la comparaison des *k-mers* présents au sein du génome mâle et ceux présents au sein des *reads* femelles (*Illumina*). Les *k-mers* des séquences Y-spécifiques ne devraient pas correspondre aux *k-mers* des *reads* de femelles (pic rouge sur le graphique), tandis que les séquences autosomales ou liées au X sont totalement couvertes par les *k-mers* des *reads* de femelles (pic gris sur le graphique).

Selon les recommandations des concepteurs de la méthode *YGS*, les génomes de très grande taille (comme celui d'*A. nasatum*) sont analysés plus précisément (diminution des faux-positifs) en comparant des *k-mers* de plus grande taille (Carvalho & Clark, 2013). Ainsi, afin de déterminer le paramètre de *k-mer* idéal pour la suite des analyses *YGS*, une comparaison des *16-mers* et des *18-mers* a été effectuée sur le génome *PacBio* assemblé avec *CeleraAssembler*. Seule une comparaison des *18-mers* a été effectuée sur l'assemblage réalisé avec *DBG2OLC* puisque ce paramètre a été considéré comme optimal vis-à-vis de l'assemblage. Cependant, des risques de faux-positifs peuvent apparaître lorsqu'un faible nombre de *k-mer* uniques sont détectés au sein des *contigs/scaffolds* mâles, ainsi le nombre minimal de *k-mer* uniques présents par séquence a été fixé à au moins 20 avec *Jellyfish* (version 2.2.6, Marçais & Kingsford, 2011). Le seuil retenu pour considérer des séquences candidates était un score d'*YGS* d'au moins 35%.

II. Tests de validation *in vivo* des séquences candidates Y-spécifiques

Les séquences candidates Y-spécifiques identifiées *in silico* ont été testées par PCR sur l'ADN de mâles et de femelles *A. nasatum* asymbiotiques. Les extractions d'ADN ont été réalisées avec le kit d'extraction sur colonne *DNeasy Blood and Tissue* (Qiagen) en suivant le protocole fourni. La qualité des ADN extraits a été confirmée par l'amplification du gène mitochondrial *col* (Folmer et al., 1994). L'absence de *Wolbachia* a été vérifiée par PCR avec les amorces *wsp* et *ftsZ* (Werren et al., 1995; Braig et al., 1998).

Pour chaque séquence Y-spécifique candidate, différents tests PCR ont été réalisés afin de confirmer ou non l'association des *loci* candidats avec le sexe mâle. Les amorces ont été dessinées avec le logiciel en ligne *Primer-BLAST* (J. Ye et al., 2012). Leur spécificité a été vérifiée par *BLASTn* (version 2.2.30+, Camacho et al., 2009) en comparant les amorces dessinées avec le génome d'*A. nasatum*, de sorte que chaque couple d'amorces ne puissent s'hybrider qu'au

niveau du *locus* à tester. Les tests PCR ont ensuite été effectués selon trois phases de tests successives :

(i) test sur 6 mâles et 8 femelles issus de deux portées différentes, partageant les mêmes grands-parents que les mâles utilisés pour le séquençage *PacBio* (sachant que ces grands-parents sont aussi les parents des individus séquencés en *Illumina*, Figure 34).

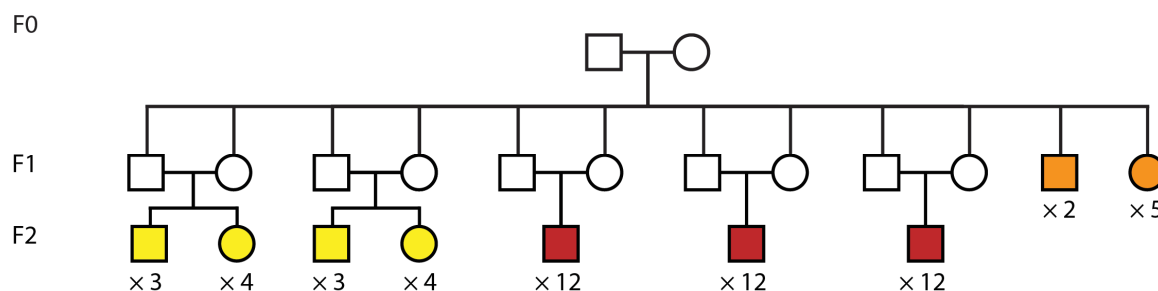


Figure 34 : Représentation des relations de parenté entre tous les individus séquencés et les individus utilisés pour la première phase de test des séquences candidates Y-spécifiques. Les carrés représentent les mâles, et les cercles les femelles. Les individus colorés en orange correspondent aux deux mâles et aux cinq femelles utilisés pour le séquençage *Illumina*. Les mâles colorés en rouge correspondent aux individus utilisés pour le séquençage *PacBio*. Les individus colorés en jaune correspondent aux individus utilisés pour la première phase de test des amorces.

(ii) test des *loci* validés lors de la phase précédente sur l'ADN de 6 mâles et de 12 femelles issus de lignées de laboratoire. Ces lignées sont dérivées de la même population que les individus séquencés (Thuré, 79, France) mais n'ayant pas de lien de parenté connu.

(iii) test des *loci* validés lors de la phase précédente sur l'ADN d'individus issus de deux populations naturelles additionnelles : 2 mâles et 1 pool de 3 femelles d'une population de Beauvoir-Sur-Niort (79, France) ainsi que 2 mâles et 1 pool de 3 femelles d'une population de Piriápolis (Uruguay).

Lors de chacune des trois phases, les amorces issues des séquences candidates Y-spécifiques amplifiant les ADN de mâles mais pas ceux des femelles ont été conservées pour les tests suivants. Si les amorces n'amplifiaient le *locus* ciblé chez aucun sexe ou l'amplifiaient chez une femelle, elles étaient éliminées de la suite des tests. A l'issue de la phase (iii), les *loci* restants ont été considérés

comme Y-spécifiques. Le programme PCR standard utilisé est indiqué en Annexe II. Les séquences des amorces testées sont listées en Annexe III.

Les amorces Y-spécifiques ont ensuite été testées sur des individus porteurs de la bactérie *Wolbachia*. Les individus testés sont issus de familles provenant d'une lignée de laboratoire, dont la population originelle est la même que celle des individus séquencés (Thuré, 79, France). Les tests ont été effectués sur des portées composées exclusivement de mâles ou de femelles. Ainsi, la totalité des portées immatriculées NASw39 (composée de 28 mâles) et NASw40 (composée de 42 femelles) ont été analysées. Les mères de ces portées sont sœurs et sont issues de la portée NASw23 (composée à l'origine de 60 femelles) (A, Figure 35). Les mêmes tests ont été effectués sur les portées immatriculées NASw31 (22 femelles) et NASw32 (32 mâles), dont les mères respectives sont deux sœurs issues de la portée NASw14 (composée à l'origine de 17 femelles). Cette dernière portée a aussi été analysée (B, Figure 35), permettant d'établir l'étude génétique sur deux générations. L'ADN des individus a été extrait par résine Chelex100 (Biorad) à partir d'une patte. Le protocole d'extraction détaillé est indiqué en Annexe II. La qualité des ADN extraits a été évaluée par l'amplification du gène mitochondrial *coI* (Folmer et al., 1994) et le statut d'infection des individus a été vérifié par PCR avec les amorces *wsp* et *ftsZ* (Werren et al., 1995; Braig et al., 1998). Afin de tester la présence du chromosome Y, tous les individus ont été testés avec trois marqueurs Y-spécifiques identifiés ci-dessus (correspondant aux identifiants scf7180001703882, scf7180001797263 et scf7180001668770).

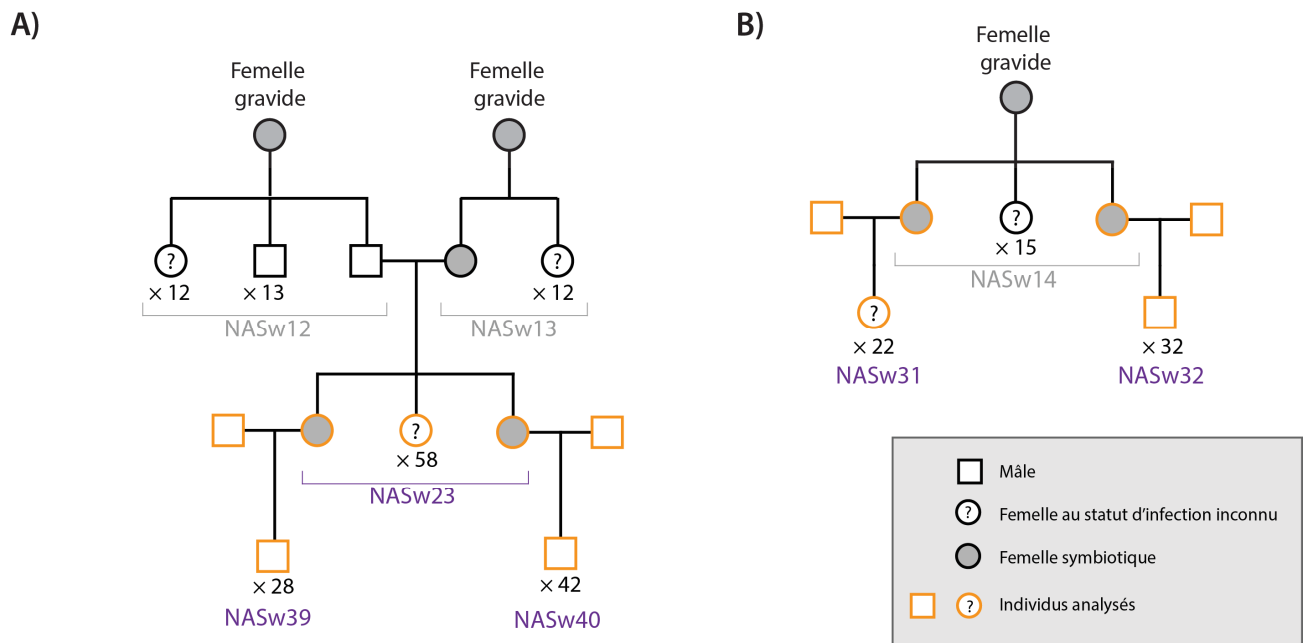


Figure 35 : Généalogie des individus issus de lignées symbiotiques (NASw) analysés avec les marqueurs Y-spécifiques. A) Relations de parenté entre les portées de matricule NASw23, NASw39 et NASw40. B) Relations de parenté entre les portées de matricule NASw31 et NASw32. Les carrés représentent des mâles, et les cercles correspondent aux femelles. Les portées dont le nom est indiqué en violet sont les portées analysées. Les femelles indiquées en gris correspondent à des femelles porteuses de *Wolbachia* dont le statut avait été établi avant croisement par le laboratoire.

Résultats

I. Recherche de séquences spécifiques au chromosome Y dans le génome d'*A. nasatum*

1. Recherche *in silico* par étude du ratio de couverture sur le génome assemblé avec les données *Illumina*

L'alignement des *reads Illumina* mâles et femelles sur les *contigs/scaffolds* mâles assemblés avec les données *Illumina* seules a été effectué indépendamment pour chaque jeu de données, en choisissant un paramètre d'alignement relâché (alignement local des *reads*). Plus de 98% de chacun des deux ensembles de *reads* (mâles et femelles) ont pu être alignés sur le génome mâle assemblé (Tableau 14). Dans les deux cas, plus de 90% de l'ensemble du génome a pu être aligné avec ces *reads*, et la couverture moyenne (par nucléotide) obtenue pour les données mâles et femelles était respectivement de 37.48x et de 39.33x (Tableau 14).

Tableau 14 : Statistiques de l'alignement des *reads* de mâles et femelles sur le génome mâle assemblé à partir des données *Illumina*. La couverture moyenne indiquée correspond au nombre de *reads* alignés sur un même nucléotide du génome assemblé. Les pourcentages affichés sont indiqués par rapport au nombre total de *reads*.

	Alignement des données mâles		Alignement des données femelles	
Nombre total de <i>reads</i> alignés	605 520 072	98.89%	624 036 328	98.84%
<i>reads</i> alignés appariés	600 803 328	98.12%	619 085 896	98.06%
<i>reads</i> alignés en "paires propres" ¹	527 547 910	86.16%	568 693 446	90.08%
% du génome aligné par les <i>reads</i>	92.53%		91.49%	
couverture moyenne par nucléotide	37.48x		39.33x	

Une série de filtres *in silico* a permis de sélectionner des régions candidates Y-spécifiques en se basant sur les ratios d'alignement des données mâles et femelles sur le génome total. La première étape consistait à ne

¹ Les *reads* alignés en "paires propres" sont des *reads* alignés par paires, dans le bon sens de lecture avec une taille d'*insert* cohérente.

sélectionner que les régions génomiques non-couvertes par les *reads* de femelles (sur au moins 100 pb contigües). Ainsi, 204 575 séquences ont été sélectionnées par cette approche (pour une taille totale de 27.7 Mb, Figure 36). Parmi ces séquences, 2 494 (représentant 0.47 Mb) ont été éliminées car elle étaient assignées à des taxa procaryotes ou fongiques (Figure 36).

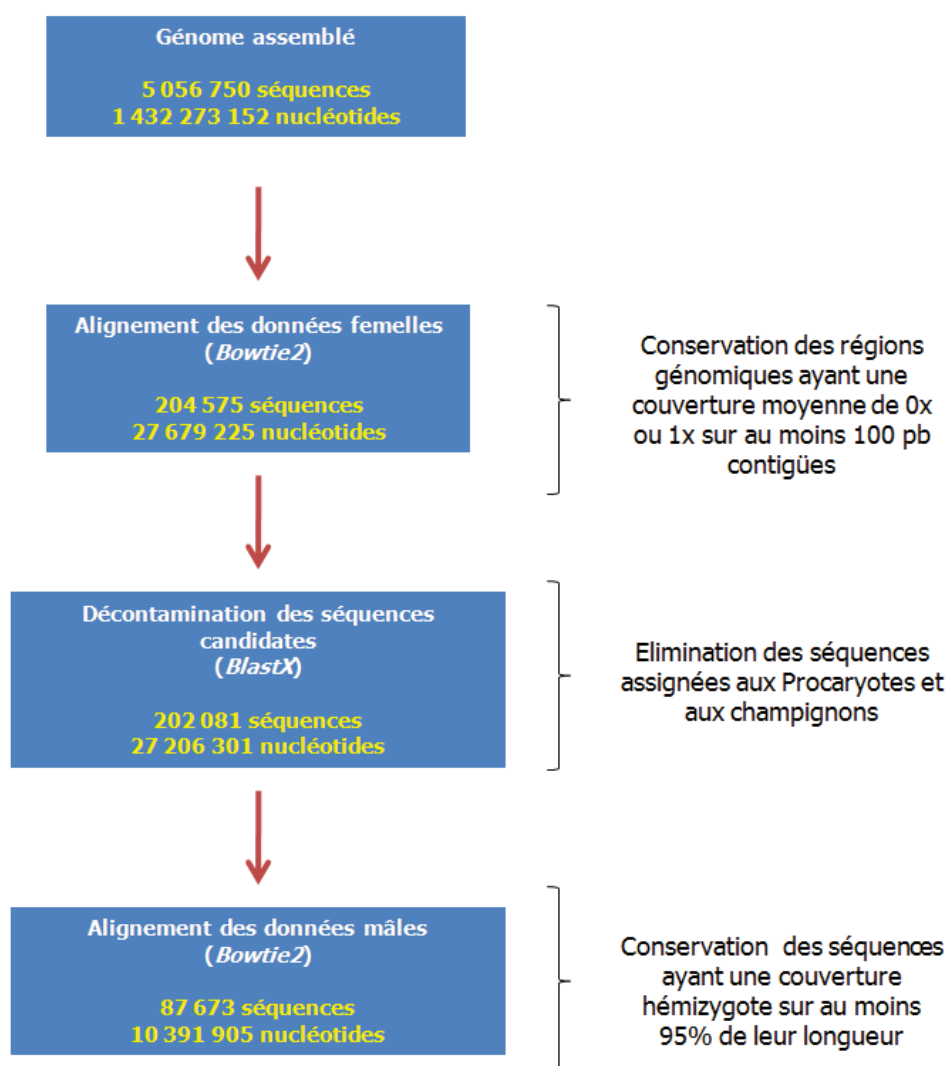


Figure 36 : Filtres *in silico* de détection de séquences Y-spécifiques par une approche basée sur les ratios de profondeur d'alignement des données mâles et femelles.

Enfin, les séquences Y-spécifiques sont normalement couvertes par les données mâles avec une profondeur de couverture deux fois inférieure à celle attendue d'un *locus* autosomique diploïde. La demi-couverture des *reads* de mâles sur les séquences sélectionnées a donc été établie de façon probabiliste avec la suite *mixtools* implémentée dans le logiciel *R*, et estimée à 2.56x. Cette valeur très faible s'explique du fait que les séquences sélectionnées restantes avaient elles-mêmes une très faible couverture moyenne par les *reads* de mâles. Ainsi toutes les séquences présentant une couverture de 3x sur au moins 95% de leur longueur ont été conservées. La totalité des séquences candidates Y-spécifiques était donc de 87 673 régions (d'une longueur moyenne de 119 pb) pour un total de 10.4 Mb (Figure 36). Ce nombre de séquences candidates étant très élevé, aucune d'entre-elles n'a été testée par PCR, et la stratégie de recherche s'est ensuite portée sur la génération de *reads* de grande taille (technologie *Pacific Biosciences*) afin d'assembler un génome moins fragmenté.

2. Recherche *in silico* par la combinaison des approches CQ et YGS sur le génome assemblé avec *CeleraAssembler*

a) Approche par ratio de couverture - méthode du Chromosome Quotient (CQ)

La recherche de séquences liées au chromosome Y a été effectuée à nouveau en utilisant le génome mâle d'*A. nasatum* assemblé à partir des *reads* *PacBio* avec le logiciel *CeleraAssembler*. L'approche choisie ici était de maximiser la fiabilité de l'alignement des *reads* avec une stratégie d'alignement en *end-to-end*, implémentée dans la méthode CQ (Hall et al., 2013). Cet alignement étant effectué avec des paramètres plus stricts (*end-to-end* sans erreur autorisée) sur un génome pouvant contenir encore des erreurs causées par la technologie de séquençage *PacBio*, moins de 60% des *reads* (mâles et femelles) ont pu être alignés (Tableau 15). Ainsi, seulement 57% du génome est couvert par les *reads* de mâles et 55.5% par les *reads* de femelles, avec une profondeur moyenne de 32.9x et 34.9x, respectivement (Tableau 15).

Tableau 15 : Statistiques de l'alignement des *reads* de mâles et femelles sur le génome mâle assemblé à partir des données *PacBio* par le logiciel *CeleraAssembler*. La profondeur moyenne indiquée correspond au nombre de reads alignées sur un même nucléotide du génome assemblé. Les pourcentages affichés sont indiqués par rapport au nombre total de *reads* (issus des données *Illumina*).

	Alignement des <i>reads</i> de mâles		Alignement des <i>reads</i> de femelles	
Nombre total de <i>reads</i> alignés	428 827 260	70.03%	438 740 181	69.49%
<i>reads</i> alignés appariés	389 956 220	63.66%	400 887 956	63.50%
<i>reads</i> alignés en "paires propres"	369 715 760	60.38%	382 546 794	60.59%
% du génome aligné par les <i>reads</i>	56.97%		55.51%	
profondeur moyenne par nucléotide	32.90x		34.89x	

Afin de limiter le biais causé par la différence de quantité de *reads* de mâles et femelles alignés sur le génome, les valeurs d'alignement des *reads* de femelles ont été normalisées par le rapport entre le nombre de *reads* de mâles alignés sur le nombre de *reads* de femelles alignés. Les alignements réalisés avec les *reads* de mâles et de femelles semblent confirmer que les chromosomes X et Y partagent une large fraction homomorphe (Figure 37, A et B). Les *contigs* associés à la région spécifique au chromosome Y sont sensés présenter une faible profondeur de *reads* de femelles ainsi qu'une profondeur de *reads* de mâles hémizygote (correspondant aux points rouges sur la Figure 37, A). Or il est notable que très peu de *contigs* présentent un tel pattern de profondeur d'alignement (Figure 37, B). D'ailleurs, les profondeurs moyennes par *contig* des *reads* de mâles et de femelles sont fortement corrélées (test de régression linéaire : R^2 ajusté = 0.9726, p-value < $2.2 \cdot 10^{-16}$ avec un coefficient directeur de 1.022).

L'analyse *CQ* présente une valeur moyenne de 1.06 (*CQ* médian = 1.024), indiquant que les *reads* de mâles et de femelles ont globalement été alignés de façon très similaire sur le génome (Figure 37, C). En se basant sur une valeur seuil de score de *CQ* fixée comme inférieure ou égale à 0.35 (Hall et al., 2013), nous avons néanmoins pu sélectionner 2 144 séquences Y-spécifiques candidates (représentant 12 039 722 pb, pour une taille moyenne de 5 616 pb par *contig*).

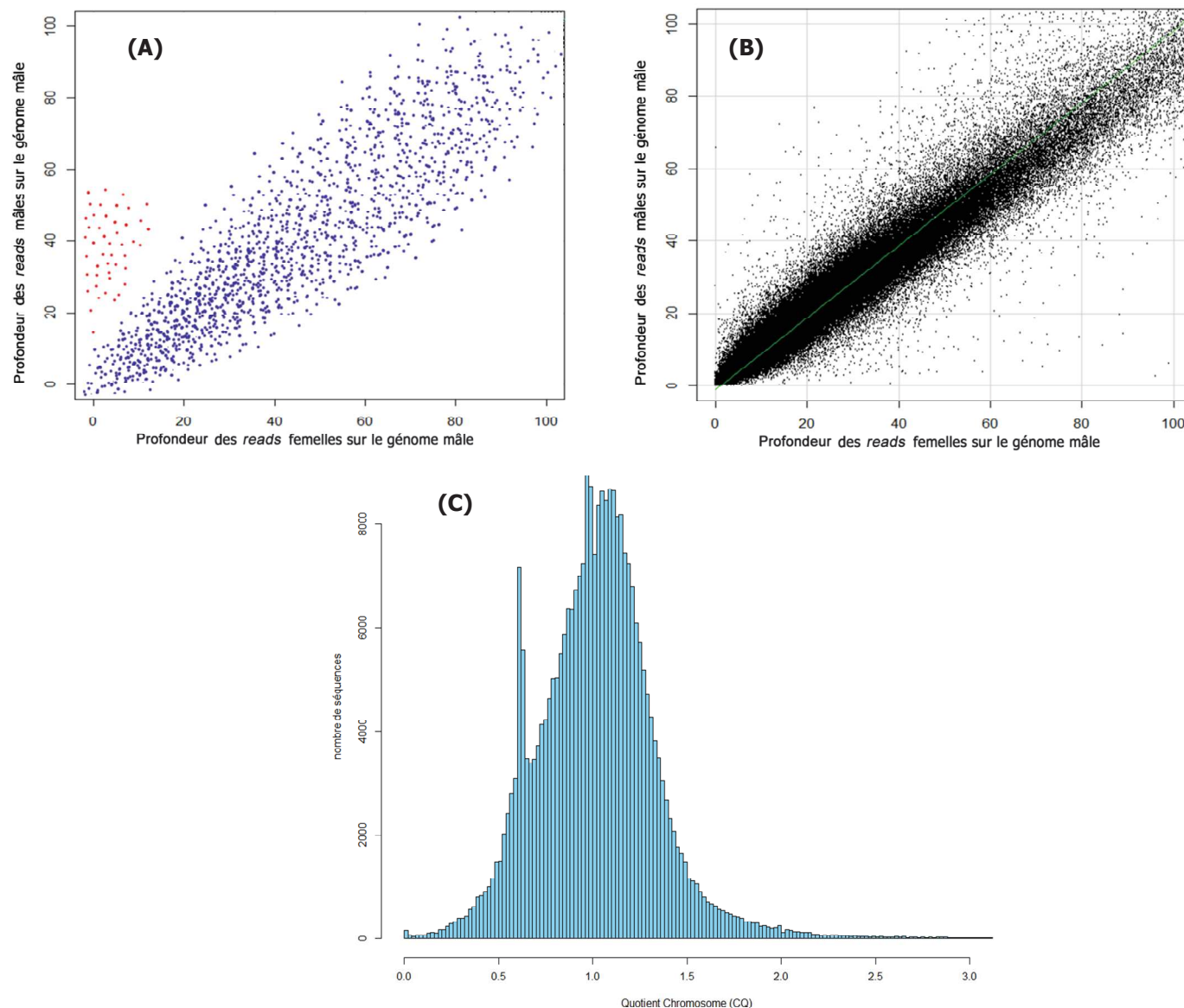


Figure 37 : Résultats des analyses de ratio de couverture effectuées sur le génome assemblé avec *CeleraAssembler*. (A) Graphique indiquant la distribution théorique de la profondeur moyenne de chaque *contig* par les *reads* de mâles en fonction de leur profondeur moyenne par les *reads* de femelles. Les points indiqués en bleu correspondent à des contigs autosomaux, tandis que les points indiqués en rouge correspondent à des contigs associés au chromosome Y. (B) Graphique présentant la profondeur moyenne de chaque *contig* par les *reads* de mâles en fonction de leur profondeur moyenne par les *reads* de femelles. La ligne verte correspond à la droite de régression linéaire entre les deux alignements (coefficient directeur = 1.022, R^2 ajusté = 0.9726, p-value < $2.2 \cdot 10^{-16}$). (C) Histogramme représentant la valeur de CQ obtenue pour chaque *contig*. La moyenne des valeurs de CQ est de 1.06 (pour une valeur médiane de 1.024).

b) Approche par comparaison de k-mer - méthode du Y chromosome Genome Scanning (YGS) et influence de la taille de k-mer lors de l'analyse

Afin de réduire le nombre de faux-positifs possiblement causés par des biais d'alignement dus à la présence de nombreuses séquences répétées (cf : Résultats du Chapitre 1), une approche d'analyse par *k-mer*, appelée *YGS* (Carvalho & Clark, 2013), a été réalisée sur les mêmes données que précédemment. Cette méthode se base sur la comparaison des *k-mers* présents au sein du génome mâle mais absents des *reads* de femelles afin d'identifier des séquences Y-spécifiques candidates. En suivant les recommandations des concepteurs de cette approche, deux tailles de *k-mer* ont été évaluées : l'analyse a tout d'abord été effectuée en 16-mer car il s'agissait dans un premier temps du paramètre de *k-mer* le plus élevé utilisable selon nos conditions techniques (plus adaptée pour les génomes assemblés de moins de 500 Mb). L'analyse a ensuite été répétée avec un paramètre de 18-mer, qui représente un paramètre mieux adapté à la taille de notre assemblage (Carvalho & Clark, 2013). Dans les deux cas, les séquences du génome mâle n'ont été conservées dans l'analyse que si elles disposaient d'au moins 20 *k-mers* en copie unique. La valeur d'*YGS* moyenne obtenue avec le paramètre 16-mer est de 36.85% (pour une valeur médiane de 38.20%, Figure 38 A, D), indiquant une forte imprécision quant à la comparaison des *k-mers* entre mâle et femelle puisque les valeurs obtenues correspondant aux autosomes doivent se rapprocher de 0. Or dans la Figure 38 (A), la répartition des valeur d'*YGS* est relativement étalée.

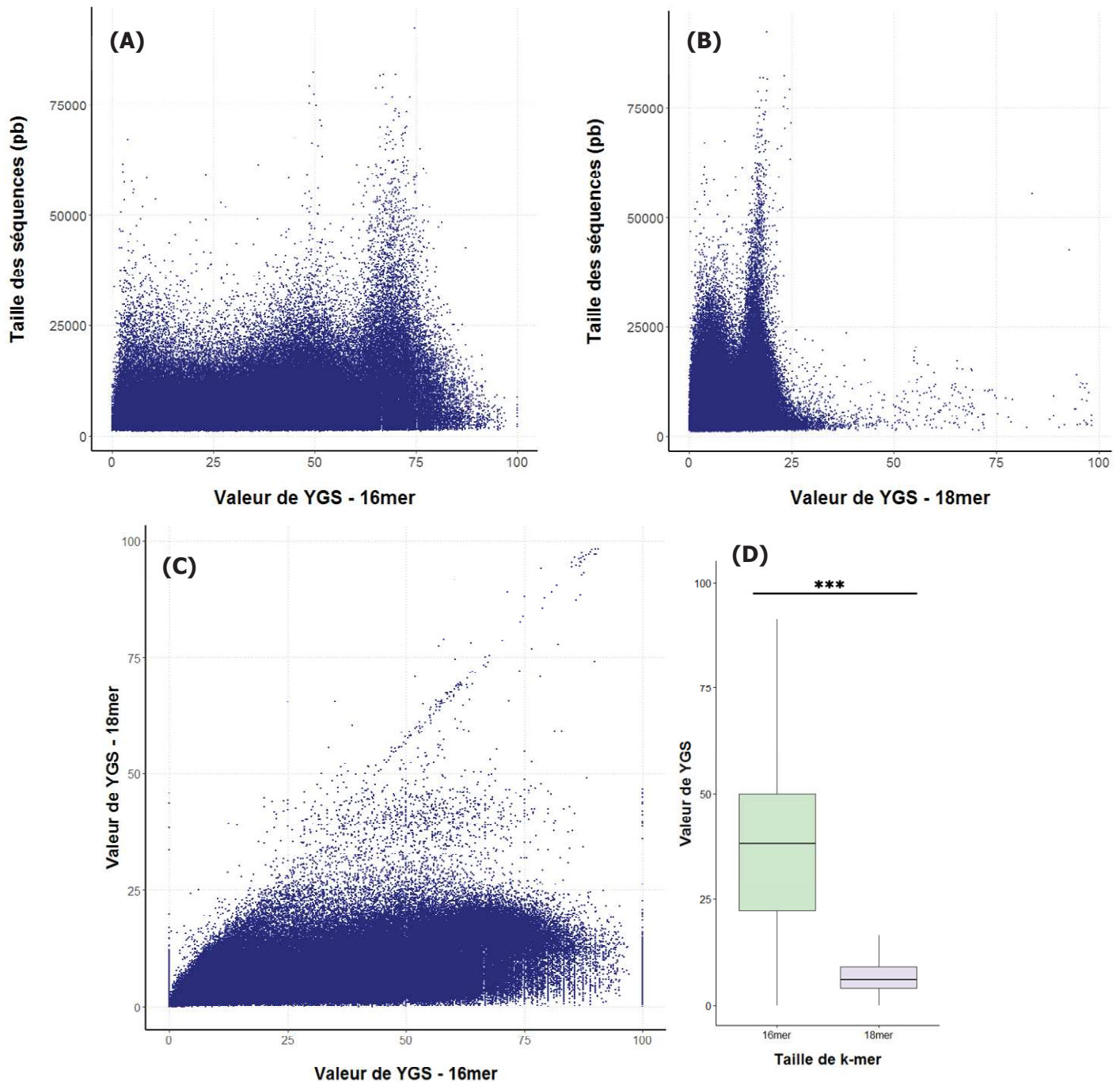


Figure 38 : Comparaison des analyses YGS réalisées avec les paramètres 16-mer et 18-mer. Chaque point correspond à une séquence assemblée. (A) Représentation de la valeur YGS selon le paramètre 16-mer pour les séquences présentes dans l'assemblage d'*A. nasatum* avec *CeleraAssembler*. (B) Représentation de la valeur YGS selon le paramètre 18-mer pour les séquences présentes dans l'assemblage d'*A. nasatum* avec *CeleraAssembler*. (C) Représentation de la valeur d'YGS obtenue avec les paramètres 16-mer et 18-mer pour les différentes séquences de l'assemblage. (D) *Boxplot* correspondant à l'ensemble des valeurs d'YGS obtenus par *contig* pour les paramètres 16-mer (vert) et 18-mer (rose). Les valeurs obtenues pour l'analyse en 16-mer sont statistiquement plus élevées que celles obtenues pour l'analyse en 18-mer (test t de Student, $t = 989.82$, $df = 343\,010$, $p\text{-value} < 2.2 \cdot 10^{-16}$).

Les valeurs d'*YGS* mesurées avec le paramètre 18-mer s'avèrent statistiquement plus faibles que celles mesurées en 16-mer (Figure 38, B, D), avec une valeur moyenne de 7.21% (valeur médiane de 5.90%). La répartition des valeurs de *YGS* mesurées en 18-mer s'avère donc plus proche de 0, ainsi l'augmentation du paramètre de *k-mer* utilisé lors de l'analyse *YGS* a pour effet de nettement diminuer la mesure du *YGS* global par *contig* (Figure 38, C, D). La ségrégation des contigs selon le paramètre *YGS* étant moins efficace avec le paramètre 16-mer qu'avec le paramètre 18-mer, c'est donc ce dernier qui a été retenu pour la sélection des séquences Y-spécifiques candidates en combinaison avec le paramètre *CQ*. Les séquences candidates ont été retenues si leur valeur d'*YGS* était supérieure ou égale à 35%. Ainsi, 341 *contigs* candidats ont été retenus, mesurant 5 983 pb en moyenne (pour une taille totale de 2 040 305 pb).

c) *Combinaison des approches CQ et YGS - sélection et test des candidats Y-spécifiques*

Les *contigs* sélectionnés à la fois selon le critère $CQ \leq 0.35$ et le critère $YGS \geq 35\%$ étaient au nombre de 111 (représentant un total de 522 214 pb, Tableau 16). Le génome assemblé avec *CeleraAssembler* n'ayant pas été décontaminé des séquences potentiellement bactériennes, un *BLASTx* effectué contre la base de données protéiques de NCBI a révélé que 26 séquences étaient assignées à des séquences bactériennes connues, diminuant le nombre de *contigs* candidats potentiels à 85 séquences (représentant un total de 409 388 pb, Tableau 16). Le détail de l'assignation des séquences en rapport avec leur score de *CQ* et d'*YGS* est indiqué dans la Figure 39. Ces 85 séquences ne représentent que 0.025% des séquences de l'assemblage total, et ne correspondent qu'à 0.02% de l'ensemble des nucléotides assemblé dans le génome. Ces proportions semblent ainsi confirmer l'hypothèse que les chromosomes X et Y sont peu divergents et donc encore fortement recombinants. Enfin, sur les 85 séquences candidates, 59 ont pu être testées par PCR afin de valider leur liaison au sexe mâle (les séquences des amorces sont listées dans l'Annexe III). Les 26 autres séquences candidates n'ont pas été testées car aucun couple d'amorces n'a pu être dessiné dans une zone ne présentant pas d'homologie avec d'autres séquences du génome.

Tableau 16 : Résumé des séquences candidates Y-spécifiques définies à partir des méthodes *CQ* et *YGS*. Ces analyses ont été réalisées sur le génome assemblé avec *CeleraAssembler*.

	Nombre de <i>contigs</i>	Taille totale des <i>contigs</i> (pb)
$CQ \leq 0.35$	2 144	12 039 722
$YGS \geq 35\%$	341	2 040 305
$CQ \leq 0.35 \cap YGS \geq 35\%$	111	522 214
Décontamination des séquences assignées aux bactéries	85	409 388
Nombre de candidats testés par PCR	59	241 874

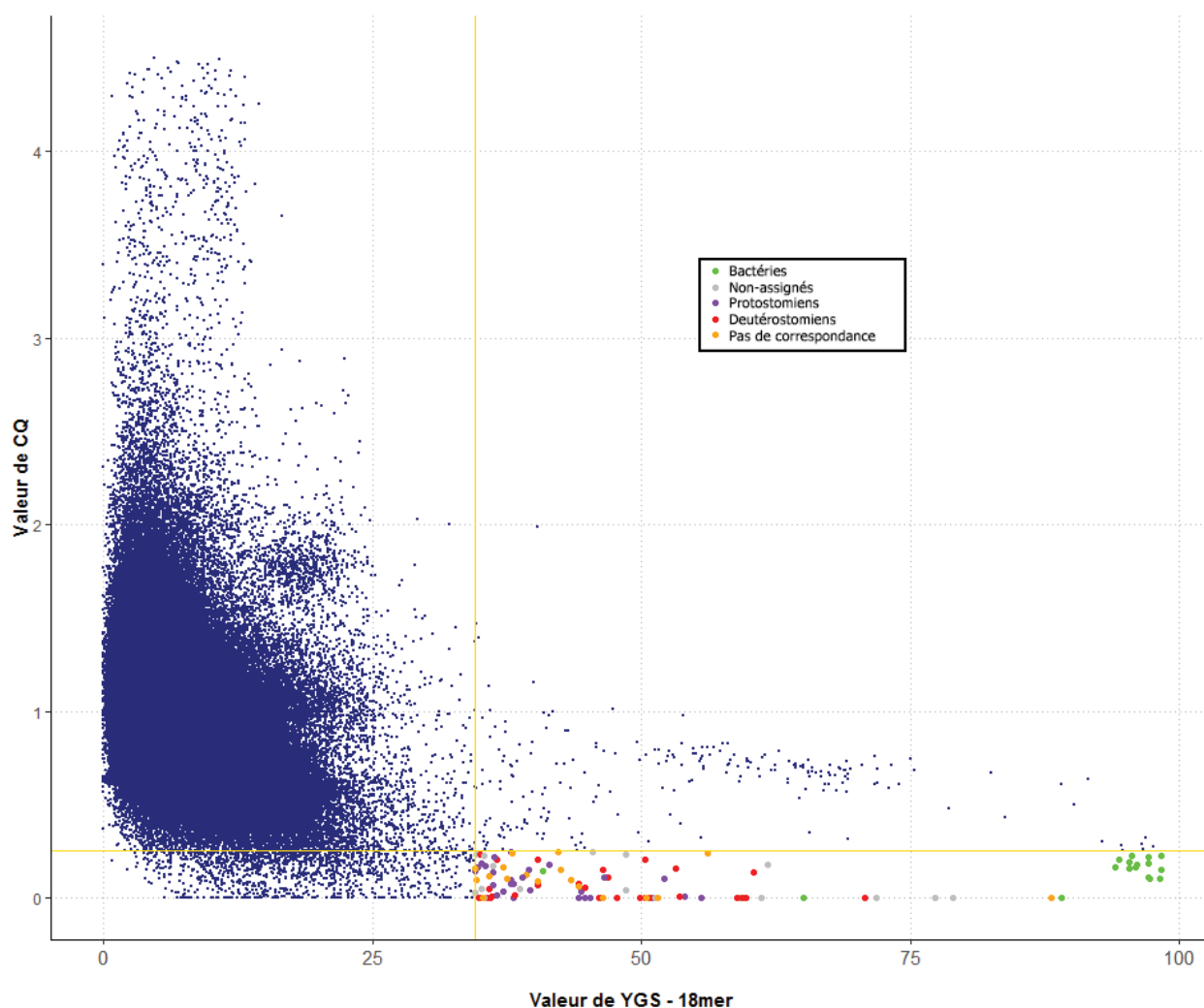


Figure 39 : Représentation de la valeur de *CQ* obtenue pour chaque *contig* en fonction de leur valeur d'*YGS* selon un paramètre de 18-mer. Chaque point correspond à une séquence assemblée. Les valeurs de *CQ* au-delà de 4.5 ne sont pas représentées. La ligne jaune verticale correspond au seuil d'*YGS* de 35%, tandis que la ligne jaune horizontale indique le seuil de *CQ* fixé à 0.35. Les 111 séquences candidates issues de ces deux analyses ont été colorées selon leur assignation taxonomique (par *BLASTx*).

Sur les 59 séquences testées en PCR, 6 ont montré des amplifications spécifiques au sexe mâle au cours des trois phases successives de tests PCR effectués (Tableau 17).

Tableau 17 : Résumé des séquences candidates Y-spécifiques restantes après les différentes phases de tests PCR effectués.

Filtres de sélection par tests PCR	Nombre de séquences candidates restantes
nombre de séquences candidates	85
nombre de séquences testées en phase (i) : sur des individus issus des mêmes familles que les individus séquencés	59
nombre de séquences testées en phase (ii) : sur des individus de familles différentes (mais partageant les mêmes grands-parents)	53
nombre de séquences testées en phase (iii) : sur des individus de populations différentes	9
Nombre de séquences spécifiques au sexe mâle à l'issue de la phase (iii)	6

La grande majorité des séquences a passé avec succès le premier crible de tests (53 sur 59, soit 90%), confirmant l'efficacité des analyses *in silico* vis-à-vis d'individus fortement apparentés à ceux séquencés. En revanche, seulement 9 des 53 séquences restantes ont présenté des amplifications spécifiques au sexe mâle sur l'ADN d'individus de fratries différentes mais issues des mêmes grands-parents. Enfin, 6 des 9 dernières séquences ont montré des amplifications spécifiques au sexe mâle sur des individus issus de différentes populations. Les 6 séquences restantes, qui constituent de très robustes candidats pour une liaison au sexe mâle, ne représentent que 26 098 pb (correspondant à 0.0013% du génome assemblé), soulignant encore la très faible différenciation des chromosomes sexuels chez *A. nasatum*. Parmi ces 6 séquences, 4 ne présentaient pas de similarité avec des séquences connues dans les banques de données de NCBI (*e-value* seuil fixée à $\leq 10^{-10}$), tandis que les 2 autres séquences présentaient des similarités avec des séquences protéiques connues de Deutérostomiens (49% d'identité avec protéine non-caractérisée et 71% d'identité avec une protéine homologue au facteur NUBP2 impliqué dans le processus de maturation protéique).

3. Recherche *in silico* par la combinaison des approches CQ et YGS sur le génome de référence assemblé avec DBG2OLC

a) Réutilisation des approches CQ et YGS pour la détermination *in silico* de séquences Y-spécifiques

Les analyses CQ et YGS ont été répétées sur la dernière version de l'assemblage du génome d'*A. nasatum*, considéré comme étant l'assemblage le moins fragmenté et donc le plus susceptible de donner des résultats robustes. Les alignements effectués dans le cadre de la méthode CQ présentent de meilleurs résultats que ceux réalisés avec la version précédente du génome (Tableau 18), témoignant d'une meilleure qualité d'assemblage. Ainsi, plus de 87% des *reads* de mâles et de femelles ont pu être alignés sur le génome de référence. Près de 86.9% du génome est couvert par les *reads* de mâles et 85.8% par les *reads* de femelles, couvrant en moyenne chaque nucléotide 48 et 50 fois, respectivement (Tableau 18). Les valeurs d'alignement des *reads* de femelles ont ensuite été normalisées par le rapport entre le nombre de *reads* de mâles alignés sur le nombre de *reads* de femelles alignés afin de limiter le biais causé par la différence de quantité de *reads* de mâles et de femelles alignés sur le génome.

Tableau 18 : Statistiques de l'alignement des *reads* de mâles et de femelles sur le génome de référence. La couverture moyenne indiquée correspond au nombre de *reads* alignées sur un même nucléotide du génome assemblé. Les pourcentages affichés sont indiqués par rapport au nombre total de *reads*.

	Alignement des <i>reads</i> de mâles		Alignement des <i>reads</i> de femelles	
Nombre total de <i>reads</i> alignés	537 742 389	87.82%	552 351 165	87.49%
<i>reads</i> alignés appariés	508 129 090	82.98%	523 510 288	82.92%
<i>reads</i> alignés en "paires propres"	417 290 270	68.15%	430 593 342	68.20%
% du génome aligné par les <i>reads</i>	86.93%		85.77%	
couverture moyenne par nucléotide	47.83x		50.38x	

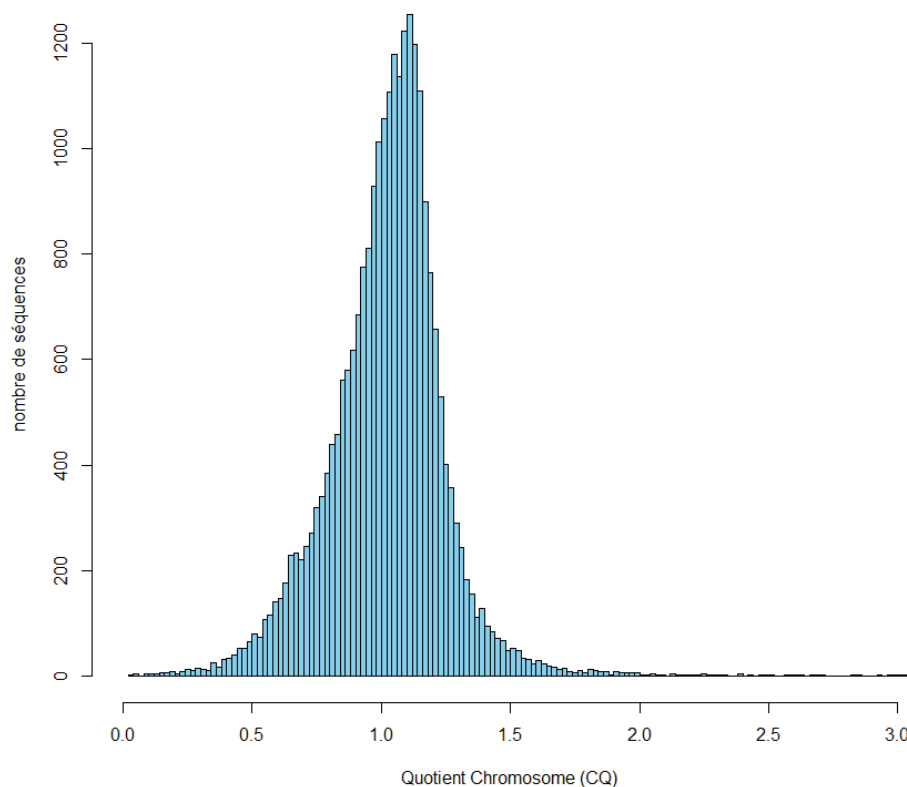


Figure 40 : Histogramme représentant la valeur de CQ obtenu pour chaque séquence assemblée. La valeur moyenne de CQ obtenue est de 1.024 (le CQ médian est de 1.041).

Comme pour l'analyse de CQ précédente, les *reads* de mâles et de femelles ont globalement été alignés de façon très similaire sur le génome, puisque la valeur moyenne de CQ par séquence assemblée était de 1.016 (valeur médiane : 1.036, Figure 40).

Cette analyse CQ a été complétée par une analyse YGS , avec le paramètre 18-mer considéré comme optimal suite aux tests précédemment réalisés. La valeur moyenne de score d' YGS mesurée par séquence assemblée était de 14.45% (Figure 41); elle est bien supérieure à la moyenne d' YGS mesurée sur le génome assemblé par *CeleraAssembler* (qui était de 7.21%). Ceci peut s'expliquer par le fait que les séquences assemblées du génome de référence étant de plus grande taille, il en résulte que plus de 18-mers uniques existent au sein de chaque séquence, facilitant ainsi la discrimination entre séquences mâles et femelles.



Figure 41 : Analyse YGS réalisée avec le paramètre 18-mer sur le génome de référence. Chaque point affiché correspond à une séquence assemblée.

Les mêmes filtres de sélection *in silico* que ceux utilisés précédemment ont été appliqués à ces nouvelles analyses. Ainsi, 108 *contigs/scaffolds* présentaient un score de $CQ \leq 0.35$, représentant un total de 4 114 839 pb (Tableau 19). De plus, 1 500 séquences avaient une valeur d'YGS supérieure ou égale à 35% (pour un total de 45 710 646 pb, Tableau 19). Ces deux analyses combinées font ressortir un ensemble de seulement 78 séquences potentiellement Y-spécifiques (d'une taille totale de 2 956 542 pb, Tableau 19). Ces séquences seront d'ailleurs prochainement testées par PCR.

Tableau 19 : Résumé des séquences candidates Y-spécifiques définies à partir des méthodes CQ et YGS. Ces analyses ont été réalisées sur le génome de référence d'*A. nasatum*.

	Nombre de <i>contigs</i>	Taille totale des séquences (pb)
$CQ \leq 0.35$	108	4 114 839
$YGS \geq 35\%$	1 500	45 710 646
$CQ \leq 0.35 \cap YGS \geq 35\%$	78	2 956 542

b) Comparaison entre les analyses *in silico* réalisées sur le génome de référence avec celles effectuées sur le génome assemblé avec *CeleraAssembler*

Les analyses *in silico* n'ont pas montré la même efficacité selon le génome sur lequel elles ont été réalisées. L'analyse *CQ* réalisée sur le génome de référence s'est avérée plus discriminante que celle réalisée sur le génome assemblé avec *CeleraAssembler*. En effet, seulement 108 séquences ayant un score de $CQ \leq 0.35$ ont été identifiées sur le génome de référence, alors que 2 144 avait été isolées à partir de la première analyse réalisée. Le nombre plus faible de candidats obtenus à partir du génome de référence témoigne probablement d'un alignement plus fiable et plus complet des *reads* sur le génome (Tableau 18). En revanche, l'analyse *YGS* réalisée sur le génome de référence a isolé bien plus de séquences candidates (1 500) que celle réalisée sur la version précédente du génome (341). Pour les deux génomes, la combinaison des deux approches a permis d'isoler un faible nombre de séquences candidates (85 après élimination des séquences bactériennes pour l'assemblage réalisé avec *CeleraAssembler* et 78 pour l'assemblage de référence du génome), ce qui souligne l'efficacité complémentaire des approches *YGS* et *CQ*.

Afin de mesurer le recouvrement entre les *loci* candidats identifiés dans les deux assemblages, les amplicons de PCR (ainsi que leurs amorces correspondantes) des 59 *loci* testés suite aux analyses réalisées sur le génome assemblé avec *CeleraAssembler* ont été alignés par *BLASTn* contre la totalité du génome assemblé avec *DBG2OLC*. Cette analyse n'a pas été effectuée sur la totalité des 85 candidats issus du génome assemblé avec *CeleraAssembler* car une partie d'entre eux était fortement répétées. Ainsi cette procédure n'a été effectuée que sur les 59 *loci* testés en PCR et uniques au sein du génome *CeleraAssembler*. Les séquences de l'assemblage *DBG2OLC* ont été conservées si l'amplicon et les amorces dessinés s'alignaient parfaitement contre ces dernières. Lorsque deux séquences pouvaient correspondre au même amplicon (ainsi qu'à ses amorces), alors les deux ont été conservées. Ainsi, les 59 *contigs* sélectionnés issus de l'assemblage réalisé avec *CeleraAssembler* et testés par PCR correspondaient à 49 séquences de l'assemblage réalisé avec *DBG2OLC* (une même séquence de l'assemblage réalisé avec *DBG2OLC* pouvant réunir plusieurs *contigs* de l'assemblage *CeleraAssembler*). De ces 49 séquences identifiées, seulement 2 répondaient toujours aux critères $CQ \leq 0.35$ et $YGS \geq 35\%$, et étaient donc communes aux 59 candidats testés par PCR au sein de l'assemblage réalisé avec

CeleraAssembler et aux 78 séquences candidates identifiés au sein du génome assemblé avec *DBG2OLC* (Figure 42).

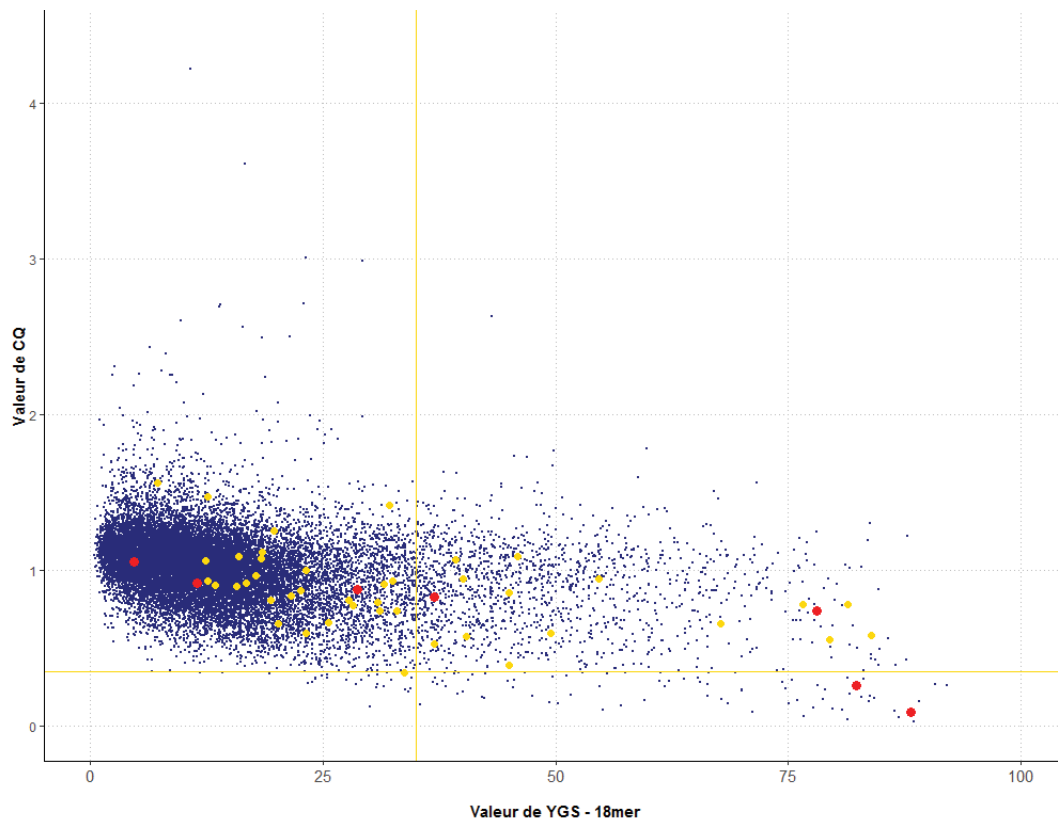


Figure 42 : Combinaison des résultats d'analyses CQ et YGS obtenus pour chaque séquence de l'assemblage réalisé avec *DBG2OLC*. Chaque point affiché correspond à une séquence assemblée. La ligne jaune verticale correspond au seuil d'YGS de 35, tandis que la ligne jaune horizontale indique le seuil de CQ fixé à 0.35. Les points rouges correspondent aux *contigs/scaffolds* porteurs des *loci* Y-spécifiques identifiés à partir du génome assemblé avec *CeleraAssembler*. Parmi les 25194 séquences du génome assemblé avec *DBG2OLC*, 78 correspondent à la fois au critère de sélection de l'analyse YGS et à celui de l'analyse CQ. Sur ces 78 séquences, seulement 2 correspondent à des séquences sélectionnées suite aux analyses réalisées sur le génome assemblé avec *CeleraAssembler*. Les points jaunes correspondent aux 49 *contigs/scaffolds* du génome *DBG2OLC* porteur des *loci* testés par PCR suite aux analyses réalisées sur le génome *CeleraAssembler*.

Les premières analyses *in silico* effectuées avait permis l'identification de 6 *loci* Y-spécifiques. Si les 6 séquences portant ces *loci* Y-spécifiques remplissaient les conditions $CQ \leq 0.35$ et $YGS \geq 35\%$ lors de la première série d'analyses, seulement 2 d'entre elles correspondent toujours à ces mêmes conditions (Figure 42), à savoir le Contig12740 et le Contig18908. Les méthodes *YGS* et *CQ* sont donc toujours efficaces pour discriminer ces séquences candidates Y-spécifiques, et ce probablement car ces deux *contigs* contiennent relativement peu de répétitions (72% de répétitions sur l'ensemble du Contig12740 et 66% pour le Contig18908, Tableau 20). Néanmoins, il est aussi notable que le Scaffold1082 n'est

sélectionné par aucun filtre (*CQ* ou *YGS*) alors qu'il contient lui aussi 72% de séquences répétées. Les marqueurs identifiés sur les Contig3469 et Contig21476 sont toujours discriminés d'après le filtre $YGS \geq 35\%$ mais ont des valeurs de *CQ* plus proches de 1. Ceci suggère que la présence de davantage de répétitions (respectivement 82% et 81% de leur longueur) pourrait empêcher d'obtenir une mesure de *CQ* discriminante. Enfin, le Contig4747 ainsi que le Contig3242 ne sont pas sélectionnés par les filtres d'analyses *in silico* appliqués sur le génome assemblé avec *DBG2OLC*. Dans le cas du Contig4747, le taux de répétitions (91%) est possiblement trop important pour que cette séquence soit discriminée efficacement par les méthodes *CQ* et *YGS*.

Tableau 20 : Comparaison des *contigs/scaffolds* porteurs des *loci* Y-spécifique identifiés à partir de l'assemblage réalisé avec *CeleraAssembler* et leurs homologues identifiés dans le génome assemblé avec *DBG2OLC*.

Génome assemblé avec <i>CeleraAssembler</i>				Génome assemblé avec <i>DBG2OLC</i>					
Identifiant	Taille (pb)	% de GC	% de répétitions	Identifiant	Taille (pb)	% de GC	% de répétitions	Score de <i>CQ</i>	Score d' <i>YGS</i>
scf7180001668770	6691	24.44	21.78	Contig21476	29141	30.23	81.34	0.74	78.1
scf7180001797263	6713	27.79	29.00	Contig12740	11489	30.49	65.60	0.26	82.3
				Contig3242	15041	31.92	79.64	1.05	4.7
scf7180001685477	3806	35.74	53.76	Contig18908	27972	27.70	72.22	0.09	88.2
				Scaffold1082	324134	27.69	72.01	0.92	11.4
scf7180001705172	2509	31.81	22.16	Contig3469	69463	29.24	81.86	0.83	37.0
scf7180001703882	2611	35.46	80.70	Contig4747	23585	30.17	91.16	0.88	28.7
scf7180001550276	3768	36.43	74.12	Scaffold1082	324134	27.69	72.01	0.92	11.4

L'objectif d'identifier les *contigs/scaffolds* porteurs des *loci* Y-spécifiques établis à partir la version du génome assemblé avec *DBG2OLC* était d'obtenir des séquences de plus grandes tailles, potentiellement porteuses de gènes pouvant constituer de potentiels candidats impliqués dans le déterminisme du sexe d'*A. nasatum*. Cependant, parmi les 7 séquences issues du génome assemblé avec *DBG2OLC* déduites à partir des 6 *loci* vérifiés par PCR, 4 ne présentent aucun gène annoté (Contig3242, Contig18908, Contig3469 et Contig4747, Tableau 21). Parmi les différents gènes identifiés dans le Contig12740 et dans le Scaffold1082 et ayant une fonction établie, aucun n'est connu pour être potentiellement impliqué dans le déterminisme du sexe (Tableau 21). Par ailleurs, 18 gènes ont été annotés dans les 73 séquences candidates issues des analyses réalisées sur le génome assemblé avec *DBG2OLC*.

Parmi ces 18 gènes, 8 codent des protéines dont la fonction est inconnue, et 10 gènes codent pour des protéines n'étant pas connues pour être impliquées dans les mécanismes de déterminisme du sexe.

Tableau 21 : Comparaison des annotations des marqueurs Y-spécifiques selon la version du génome utilisée. Les séquences annotées avec le symbole "*" disposent de la même annotation.

Génome assemblé avec <i>CeleraAssembler</i>		Génome assemblé avec <i>DBG2OLC</i>	
Identifiant	Annotations disponibles	Identifiant	Annotations disponibles
scf7180001668770	non annotée	Contig21476	protéine de fonction inconnue protéine de fonction inconnue protéine de fonction inconnue
scf7180001797263	protéine homologue au facteur NUBP2	Contig12740	protéine homologue au facteur NUBP2
		Contig3242	non annotée
scf7180001685477	non annotée	Contig18908	non annotée
		Scaffold1082*	(annotation listées plus bas)
scf7180001705172	non annotée	Contig3469	non annotée
scf7180001703882	non annotée	Contig4747	non annotée
scf7180001550276	protéine de fonction inconnue	Scaffold1082*	protéine de fonction inconnue protéine de fonction inconnue protéine de fonction inconnue protéine de fonction inconnue protéine homologue à la familles des TBC1 protéine homologue à protéine kinase Srms protéine homologue à récepteur insuline-like protéine homologue au facteur de transcription RE1 protéine à domaine BTB/POZ protéine à bromodomaine protéine porteuse de doigts de zinc HIT protéine benzolquinol méthylase protéine impliquée dans les interactions Sérine/Thréonine/Tyzosine

4. Tests des marqueurs Y-spécifiques sur des lignées d'individus infectés par des *Wolbachia* féminisantes

Les lignées d'*A. nasatum* infectées par *Wolbachia* produisent régulièrement des portées dont le sex-ratio peut varier de 100% de femelles (ce qui peut être attendu dans le cadre d'une infection par des *Wolbachia* féminisantes) à 100% de mâles (Juchault & Legrand, 1979, 1989). Les marqueurs Y-spécifiques développés précédemment ont donc été utilisés afin d'étudier le pattern de transmission du chromosome Y au sein de lignées infectées par la souche de *Wolbachia* féminisante *wNas*. Le premier pedigree analysé réunissait les portées NASw39 et NASw40 (Figure 43), composées respectivement de 28 et de 42 mâles, alors que leurs mères respectives sont issues d'une portée (NASw23) composée uniquement de femelles (60 femelles). Tous les individus des portées NASw39 et NASw40 ont présenté des amplifications pour les marqueurs Y utilisés, confirmant leur statut de mâles XY ou YY (leur statut asymbiotique a aussi été confirmé). Les mères de ces deux portées ont aussi été testées positives aux marqueurs Y (en plus d'être infectées par *Wolbachia*). Les pères des portées NASw39 et NASw40 sont quant à eux issus de portées asymbiotiques, ce qui sous-entend qu'ils sont nécessairement XY (la présence du chromosome Y a été validée par PCR). Par conséquent, les mères des portées NASw39 et NASw40 ne peuvent être que des femelles YY, car s'il s'agissait de mères XY, des individus XX auraient été attendus dans les portées NASw39 et NASw40, ce qui n'est pas le cas (Figure 43). Dans un second temps, la présence du chromosome Y a été testée par PCR chez toutes les femelles de la portée NASw23, qui se sont toutes avérées être des mâles génétiques (XY ou YY) féminisés par *Wolbachia* (dont la présence a aussi été confirmée par PCR, Figure 43). Ainsi, la portée NASw23 était nécessairement issue de : (i) une femelle infectée par *Wolbachia*, et (ii) un père ou une mère YY. Cela permet d'expliquer pourquoi la portée NASw23 est composée d'individus ayant les génotypes XY ou YY mais présentent tous un phénotype femelle (Figure 43).

Ce premier pedigree explique comment la présence de *Wolbachia* féminisantes dans les populations d'*A. nasatum* peut conduire à des portées ayant des sex-ratio allant de 0 à 100% de femelles, c'est-à-dire par la production de

femelles YY. De plus, il faut noter que le taux de transmission de *Wolbachia* à la descendance est fortement instable, puisque soit 100% des individus issus d'une portée sont symbiotiques, soit aucun des individus ne l'est.

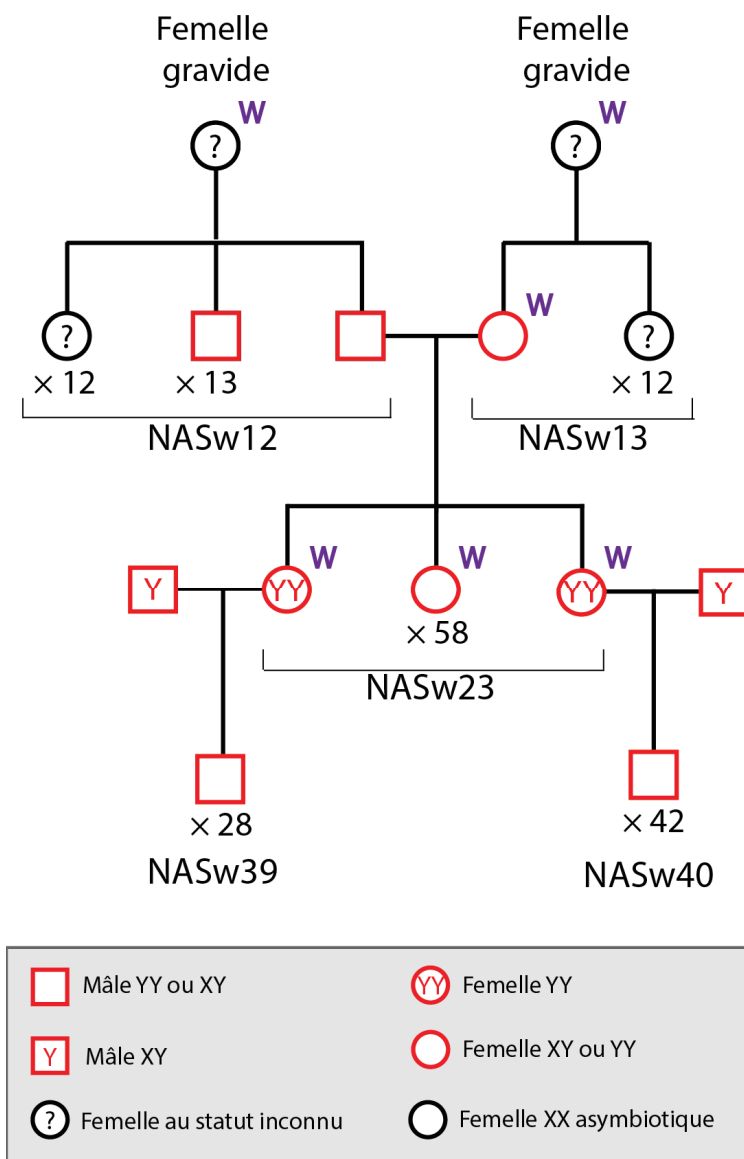


Figure 43 : Pédigree des portées *A. nasatum* symbiotiques NASw39 et NASw40. Les carrés représentent des mâles, et les cercles correspondent aux femelles. Les individus indiqués en rouge ont soit un génotype XY, soit YY. Les mères des portées NASw12 et NASw13 (notée "femelles gravides") étaient des femelles gravides (c'est à dire porteuses d'embryons en développement) issues de la nature, dont le statut d'infection avait été confirmé. Les individus annotés avec un "W" violet sont porteurs de *Wolbachia*. Les chiffres indiqués sous les symboles des mâles ou femelles correspondent au nombre d'individus.

Le second pedigree analysé rassemblait les portées NASw31 (22 femelles) et NASw32 (32 mâles), issus de deux sœurs appartenant à la portée NASw14 (qui n'était composée que de femelles). Tous les individus de la portée NASw32 ont

présenté des amplifications pour les marqueurs Y utilisés, confirmant leur statut de mâles XY ou YY. Le père de la portée NASw32 est issu d'une portée asymbiotique, ce qui sous-entend que son génotype est XY. Par conséquent la mère de la portée NASw32 est une femelle YY (elle ne peut pas être XY car tous les individus de la portée NASw32 sont des mâles XY ou YY) (Figure 44). Concernant la portée NASw31 (composée de 22 femelles porteuses de *Wolbachia*), 11 femelles sur 22 ont présenté des amplifications pour les marqueurs Y, ce qui signifie que 11 individus ont le génotype XY ou YY, tandis que les 11 autres individus qui n'ont pas présenté d'amplification ont donc le génotype XX. La mère de la portée NASw31 ne présentait pas d'amplification pour le chromosome Y, ce qui indique que son génotype est XX (il s'agit donc d'une femelle XX infectée par *Wolbachia*). Le père de la portée NASw31 est issu d'une portée asymbiotique, ce qui indique que son génotype est XY. Par conséquent, les 11 individus de la portée NASw31 présentant une amplification pour les marqueurs du chromosome Y ont le génotype XY, car ils sont issus d'un croisement entre un mâle XY et une femelle symbiotique XX (Figure 44).

Comme indiqué précédemment, les mères des portées NASw31 et NASw32 ont respectivement le génotype XX et YY. Puisqu'il s'agit de deux femelles sœurs (issues de la portée NASw14), il est possible de déduire que la mère de la portée NASw14 a un génotype XY, puisqu'elle a donné naissance à un individu XX (mère de la portée NASw31) et à un individu YY (mère de la portée NASw32). En effet, le seul croisement possible pouvant conduire à la fois à des individus XX et YY dans une même portée est un croisement entre un mâle XY et une femelle XY.

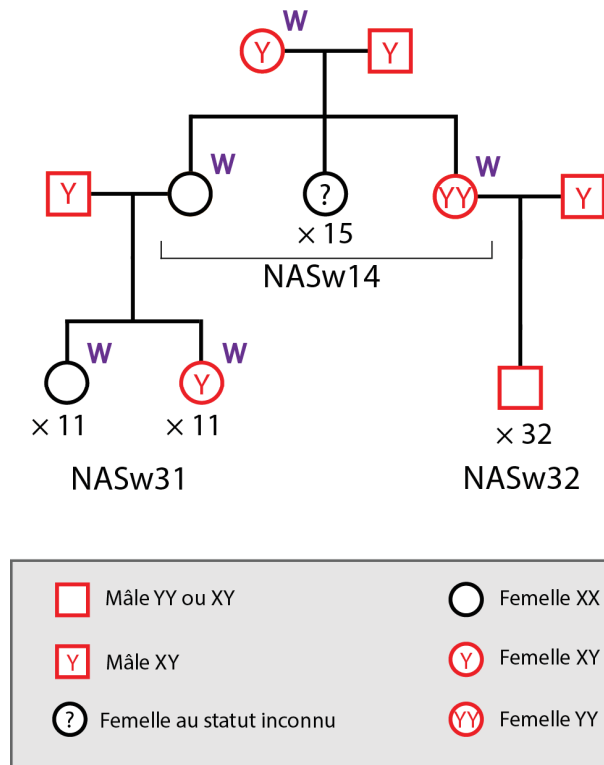


Figure 44 : Généalogie des portées *A. nasatum* symbiotiques NASw31 et NASw32. Les carrés représentent des mâles, et les cercles correspondent aux femelles. Les individus grisés sont porteurs de *Wolbachia*. Les chiffres indiqués sous les symboles des mâles ou femelles correspondent au nombre d'individus.

Ces exemples illustrent et confirment le taux de transmission fortement variable de *Wolbachia* à la descendance (0% ou 100% de transmission), tout particulièrement chez les individus YY féminisés (cf : portées NASw32, NASw39 et NASw40, Figure 43 et Figure 44). Il confirme aussi l'obtention de sex-ratios extrêmes, causés essentiellement par la transmission ou non de *Wolbachia* à la descendance, possiblement en lien avec le génotype de la mère.

Enfin, ces exemples soutiennent la très forte homomorphie des chromosomes X et Y, puisque non seulement des femelles génétiques (XX) peuvent être transformées en mâles phénotypiques par l'injection de glandes androgènes (Juchault & Legrand, 1979; Becking et al., 2017 ; Chapitre 4), mais des mâles génétiques YY, en plus d'être viables, peuvent aussi se développer en femelles phénotypiques fonctionnelles en cas d'infection avec des *Wolbachia* féminisantes (Juchault & Legrand, 1979).

Discussion

I. Recherche de séquences liées au chromosome Y d'*A. nasatum*

1. Analyse des assemblages du génome d'*A. nasatum*

La première phase de recherche de séquences Y-spécifiques candidates a été réalisée sur le génome d'*A. nasatum* assemblé *de novo* uniquement à partir des données *Illumina*. Les séquences candidates sélectionnées devaient non seulement ne pas être couvertes par les *reads* de femelles, mais devaient aussi présenter une couverture hémizygote des *reads* de mâles. Cependant, la haute fragmentation du génome (plus de 5 millions de séquences assemblées) en séquences de petite taille (N50 de 407 pb) constituait un biais potentiel pour l'alignement des *reads*, conduisant à de faibles profondeurs d'alignement par nucléotide au sein des contigs de petites tailles (Chen et al., 2012). Ce qui explique d'ailleurs pourquoi les régions non-couvertes par les *reads* de femelles (couverture moyenne de 0 ou 1 sur au moins 100 pb contigües) avaient aussi une très faible couverture par les données mâles. Cette approche effectuée sur le génome *Illumina* a conduit à un très grand nombre de séquences candidates (plus de 87 000 séquences), soulignant l'inefficacité de cette méthode sur un génome aussi fragmenté. L'utilisation de *reads PacBio* pour améliorer la qualité de l'assemblage, en plus de l'utilisation d'un double filtre de sélection *in silico* (*CQ* et *YGS*), a permis la sélection de seulement 85 séquences candidates issues du génome assemblé avec *CeleraAssembler*, dont 59 ont pu être testés par PCR. Sur ces 59 marqueurs testés (phase de test (i), Tableau 17), 53 ont montré une liaison au sexe mâle chez les frères des individus séquencés. Sur ces 53 marqueurs, seulement 9 se sont avérés spécifiques au sexe mâle chez les cousins des individus séquencés (à savoir des individus qui partagent les mêmes grands-parents que les individus séquencés en *Illumina*, phase de test (ii), Tableau 17). Enfin au total, seulement 10% des marqueurs testés par PCR se sont avérés être totalement liés au sexe mâle lorsque

les tests ont été étendus à des individus issus d'autres populations (6 séquences sur 59, phase de test (iii), Tableau 17).

La majorité des séquences testées est donc spécifique au sexe mâle, dès lors que les tests ont été effectués sur des individus appartenant à la même famille que celle des individus séquencés (53 marqueurs sur 59). Une première explication de ces résultats serait qu'une partie de ces marqueurs testés soient seulement des variants alléliques entre mâles et femelles issues de la même portée (Hall et al., 2013). Il s'agirait donc de faux-positifs causés par un biais de structure populationnelle lié aux conditions de l'élevage, et notamment au fort apparentement entre les individus séquencés et leurs frères et sœurs utilisés pour réaliser la phase de test (i). Le moyen d'éviter ce type de faux-positifs est d'augmenter la variabilité génétique des *reads* de mâles et de femelles utilisés pour l'alignement et la comparaison de *k-mers*, afin de maximiser le nombre d'allèles commun entre mâles et femelles. Cette stratégie a d'ailleurs été utilisée tout d'abord pour la recherche de séquences W-spécifiques chez le gobe-mouche à collier (Smeds et al., 2015), ainsi que pour la recherche de séquences W-spécifiques chez *A. vulgare* (Chebbi et al., en préparation). Une autre explication consiste en une liaison partielle au sexe de ces marqueurs. Cette explication implique que la taille de la zone non-recombinante du chromosome Y est variable (y compris d'une famille à l'autre), et sous-entend que des événements de recombinaison occasionnels peuvent encore intervenir entre le chromosome X et la région spécifique au chromosome Y (Charlesworth, 2013). Cette hypothèse, déjà démontrée notamment chez le poisson *Nothobranchius furzeri* (Reichwald et al., 2015), soutient le fait que si les chromosomes X et Y recombinent encore (même rarement) au niveau de la région Y-spécifique, c'est donc que ces deux chromosomes sont encore largement homomorphiques, ce qui est également corroboré par nos propres résultats chez *A. nasatum*.

Malgré tout, suite aux tests PCR réalisés, six *contigs* du génome d'*A. nasatum* ont été identifiés comme Y-spécifiques à l'issue de cette première analyse combinant *CQ* et *YGS*. Parmi ces séquences, seulement deux présentaient des similarités avec des séquences issues de la base de données non-redondante de NCBI. Une des deux séquences code pour une protéine de fonction inconnue,

tandis que l'autre séquence est similaire à une protéine homologue au facteur NUBP2. Ce facteur est fortement conservé au sein des eucaryotes, et est impliqué dans les mécanismes de régulation de la duplication du centrosome chez la levure *S. cerevisiae* (Christodoulou et al., 2006). Ainsi aucun des *contigs* liés au chromosome Y ne semble être porteur d'un gène pouvant intervenir dans la cascade de déterminisme du sexe d'*A. nasatum*.

Les mêmes analyses *in silico* ont été répétées indépendamment sur le génome de référence d'*A. nasatum* généré avec *DBG2OLC* (N50 de 86.3 kb pour 25 194 *contigs/scaffolds*). Les statistiques d'alignement des *reads* de mâles et de femelles, meilleures que celles issues des alignements réalisés sur le génome précédent (plus de 86% de *reads* de mâles et de femelles alignés), témoignent d'un assemblage contenant moins d'erreurs que celui généré avec *CeleraAssembler*. En associant les approches *CQ* et *YGS* sur cet assemblage, 78 séquences Y-spécifiques candidates ont été retenues. Si le nombre de candidats est similaire au nombre de candidats obtenus à partir de la version précédente du génome, très peu de *contigs/scaffolds* ont été sélectionnés à la fois dans le génome assemblé avec *CeleraAssembler* et dans le génome de référence. En effet, seulement 2 séquences candidates sont sélectionnées par les analyses réalisées sur les deux versions du génome (ces deux *loci* ont été confirmés par les tests PCR comme étant liés au sexe mâle). A titre d'exemple, le marqueur scf7180001703882 (génome *CeleraAssembler*, Tableau 20) était sélectionné comme candidat Y-spécifique dans le premier génome, alors que son homologue issu du génome de référence (Contig4747, Tableau 20) n'est sélectionné par aucune des deux approches utilisées. Ceci peut être en partie causé par la présence importante de répétitions au sein des séquences génomiques assemblées. En effet, le Contig18908 et le Contig12740 ont été sélectionnés par les approches *CQ* et *YGS* sur les deux versions du génome, alors que ces séquences comptent près de 70% de répétitions, ce qui est relativement proche de la densité moyenne pour l'ensemble du génome (63%). En revanche, le Contig21476 et le Contig3469 ont été sélectionnés par *YGS* et *CQ* sur la première version du génome, et seulement par *YGS* sur le génome de référence, alors que ces séquences comptent plus de 80% de répétitions. Comme indiqué plus haut, le Contig4747 a bien été retenu par les approches *YGS* et *CQ* sur la première version du génome alors qu'il ne l'a pas été

sur le génome de référence. Or, ce *contig* compte plus de 90% de répétitions. Ainsi la très forte présence de répétitions a pu impacter défavorablement le signal Y-spécifique du Contig4747, alors que son homologue dans la version précédente du génome, plus court et contenant moins de répétitions, a une signature Y-spécifique détectable par *CQ* et *YGS* (Tableau 20).

Enfin, le Scaffold1082 n'est détecté par aucune des approches *CQ* ou *YGS*, alors que son contenu en séquences répétées est proche de 70%. Néanmoins, le signal Y-spécifique a pu être dilué par la présence de plusieurs copies paralogues des gènes présents sur le Scaffold1082 dispersées au sein du génome. A titre d'exemple, le Scaffold1082 contient une protéine porteuse du domaine BTB/POZ, or il s'avère que 17 autres gènes codant pour des protéines contenant le même domaine protéique sont présents dans le reste du génome. Ainsi, dans le cas du Scaffold1082, ce n'est pas uniquement la présence de nombreuses répétitions (ETs, répétitions en tandem) mais aussi la présence de gènes répétés à d'autres positions dans le génome qui peut être responsable de la perte du signal Y-spécifique.

Ces résultats suggèrent que l'approche par alignement de *reads* (*CQ*) semble plus sensible à la présence de séquences répétées que l'approche par comparaison de *k-mers* (*YGS*). En effet, *YGS* ne comptabilise et ne compare que les copies uniques de *k-mers* présents dans les jeux de données, les *k-mers* répétés étant exclus des analyses (Carvalho & Clark, 2013). Néanmoins, ces deux méthodes sont connues pour être sensibles au taux de répétition des séquences. Ainsi, les régions fortement répétées ne sont généralement pas retenues par de telles approches, tout comme les PARs (Tomaszkiewicz et al., 2017). Or il s'avère que les répétitions sont généralement abondantes au niveau des régions non recombinantes des hétérochromosomes (Charlesworth & Charlesworth, 2000). Un exemple spectaculaire relevé est celui du gobe-mouche à collier, qui dispose d'un génome en moyenne faiblement répété (5.9% de répétitions sur les autosomes, 8.8% sur le chromosomes Z) mais dont le chromosome W compte près de 48.5% de régions répétées, soit plus de 8 fois plus que dans les régions autosomiques (Smeds et al., 2015). Dans le cas d'*A. nasatum*, les 7 séquences identifiées comme Y-spécifiques issues de l'assemblage réalisé avec *DBG2OLC* comptent en

moyenne 80% de répétitions (jusqu'à 90% pour le Contig4747), alors que le taux de répétitions du génome total est de l'ordre de 63% (Chapitre 1), ce qui suggère là aussi une accumulation d'éléments répétés dans cette région du génome.

Une étude phylogénétique basée sur une datation moléculaire a estimé que le chromosome Y d'*A. nasatum* aurait évolué au cours des 25 derniers Ma (Chapitre 4, Figure 63, Becking et al., 2017), ce qui suggère que ce chromosome est assez récent à l'échelle évolutive. Ainsi, le processus d'accumulation d'éléments répétés dans la région Y-spécifique semble être un mécanisme rapide et précoce dans l'évolution du Y. Ce mécanisme précoce d'accumulation d'éléments répétés est d'ailleurs aussi retrouvé dans l'exemple du néo-chromosome Y de *Drosophila miranda* (Steinemann & Steinemann, 1992; Bachtrog, 2003, 2013). Dans cet exemple, tandis que le taux de répétitions du génome est d'environ 5%, il est en revanche estimé à près de 50% au sein de la région spécifique au chromosome Y (Steinemann & Steinemann, 1992; Bachtrog, 2003, 2013), dont l'âge est estimé à 1 Ma (Bachtrog & Charlesworth, 2002). Le même pattern d'accumulation d'éléments répétés est également observé au sein du chromosome Y de la papaye : le chromosome Y est âgé de moins de 10 Ma, mais il est composé de près de 80% de séquences répétées (alors que le génome n'en compte en moyenne que 52%) (Wang et al., 2012). Ces différents exemples, tout comme le chromosome Y d'*A. nasatum*, sont en accord avec des travaux théoriques indiquant que l'accumulation d'éléments répétés au niveau de la région Y-spécifique est un mécanisme particulièrement précoce impliqué dans la dégénérescence du chromosome Y (Charlesworth et al., 2005; Bachtrog, 2013).

2. Perspectives concernant la caractérisation de nouvelles séquences Y-spécifiques

Au delà des caractéristiques intrinsèques du génome d'*A. nasatum* (et notamment la proportion élevée d'éléments répétés), la composition même des régions spécifiques à l'hétérochromosome les rend difficiles à détecter (Tomaszkiewicz et al., 2017). Une meilleure caractérisation des séquences Y-spécifiques d'*A. nasatum* pourrait permettre d'identifier des gènes candidats du

déterminisme du sexe chez *A. nasatum*. A ce titre, l'analyse des 73 *contigs/scaffolds* candidats Y-spécifiques encore non-testés, issus du génome assemblé avec *DBG2OLC*, constitue une première piste prometteuse pour une meilleure compréhension de la structure génomique du chromosome Y. Ainsi, parmi ces *contigs/scaffolds*, 18 gènes ont été annotés (répartis dans 11 des 73 séquences) dont 8 gènes ayant une fonction inconnue. L'analyse de ces candidats offre ainsi de nouvelles perspectives pour la découverte de gènes potentiellement impliqués dans le déterminisme du sexe.

Afin d'affiner notre recherche de séquences liées au chromosome Y, une première solution serait de masquer les répétitions du génome et de reconduire ensuite les approches *CQ* et *YGS*. L'élimination des séquences répétées pourrait permettre des analyses moins biaisées (et notamment *CQ* qui semble fortement influencé par les séquences répétées) au détriment de la continuité de l'assemblage (Hall et al., 2013). Il pourrait s'agir d'une alternative pertinente pour la caractérisation du chromosome Y, et plus spécialement dans le cas de chromosomes sexuels homomorphes disposant probablement d'une zone non-recombinante propre au Y de petite taille. Cependant, l'élimination des séquences répétées risque de fragmenter virtuellement le génome, pouvant biaiser l'alignement des *reads Illumina* sur le génome dépourvu de répétitions.

Une autre approche basée sur la détection de polymorphismes nucléotidiques entre les mâles et les femelles peut aussi constituer une stratégie alternative pour la recherche de séquences spécifiques aux hétérochromosomes. Cette approche de génomique des populations, basée sur la densité de SNPs (*Single Nucleotide Polymorphism*) entre deux chromosomes sexuels faiblement différenciés, peut permettre de détecter des différences nucléotidiques fines entre les régions spécifiques à l'hétérochromosome même si elles présentent de fortes homologies avec le chromosome X (ou Z) (Tomaszkiewicz et al., 2017). Cette méthode a déjà été éprouvée avec succès sur le néo-chromosome Y de l'espèce d'épinoche japonaise *Gasterosteus nipponicus* (Yoshida et al., 2017), qui s'est formé suite à une fusion entre un chromosome Y ancestral et un autosome au cours des 2 derniers millions d'années (Kitano et al., 2009; Natri et al., 2013). Elle a aussi été utilisée pour la caractérisation d'un ensemble de gènes condensés se

comportant tel un hétérochromosome chez le bruant à gorge blanche *Zonotrichia albicollis* (Tuttle et al., 2016). Cependant, une telle approche nécessite le plus souvent le séquençage d'un grand nombre d'individus ce qui peut rendre la production des données très onéreuse (Gautier, 2014), en particulier pour des espèces dont le génome est très grand, comme c'est le cas d'*A. nasatum*.

Une méthode alternative, basée uniquement sur l'utilisation de données issues de séquençage d'ARN, permet l'identification de gènes appartenant à l'hétérochromosome. Tout d'abord mise au point à partir de données issues d'individus fortement apparentés (et donc présentant une faible variabilité génétique, Muyle et al., 2012), cette approche est basée sur la détection de SNPs entre les gènes issus du chromosome Y (ou W) et ceux issus du chromosome X (ou Z) (Muyle et al., 2016). Plusieurs centaines de gènes spécifiques à l'hétérochromosome ont ainsi été identifiés par cette approche chez des espèces pour lesquelles des génomes séquencés n'étaient pas disponibles tels que chez la plante *Silene latifolia* (Muyle et al., 2016), ou encore récemment la plante *Mercurialis annua* (Ridout et al., 2017). Ce pipeline, nommé *SEX-DETECTOR*, est d'ailleurs plus adapté aux espèces ayant des chromosomes sexuels faiblement différenciés. Cette méthode implique le séquençage d'individus d'une même famille sur plusieurs générations afin d'étudier le patron de ségrégation des allèles, et d'en déduire les gènes potentiellement liés au sexe (en plus de permettre une estimation du niveau d'expression de ces gènes dans les conditions expérimentales) (Muyle et al., 2016). Une telle approche constituerait un véritable atout pour la recherche de gènes impliqués dans la cascade de déterminisme du sexe, notamment grâce à l'utilisation de données de séquençage d'ARN issues d'individus en cours de différenciation sexuelle. Cette approche implique que le chromosome sexuel doit disposer de suffisamment de régions exprimées pour être détectées. Plus important encore, il apparaît aussi nécessaire de séquencer les individus lorsque les gènes liés au sexe sont exprimés.

L'usage de telles approches de détection de séquences liés au sexe basées sur les SNPs chez *A. nasatum* nécessiteraient de générer de nouvelles données de séquençage à haut débit. Cependant, à plus court terme, les dernières analyses *YGS* et *CQ*, réalisées sur le génome de référence d'*A. nasatum*, ont identifié 73

nouveaux *contigs/scaffolds* candidats. Des tests PCR seront réalisés prochainement, mais cette opportunité laisse déjà présager une meilleure caractérisation à venir des séquences spécifiques au chromosome Y d'*A. nasatum*. Une meilleure caractérisation des séquences propres au chromosome Y chez *A. nasatum* et au chromosome W chez *A. vulgare* (Chebbi et al., en préparation) permettrait de mieux comprendre l'histoire évolutive de ces chromosomes sexuels, et notamment de décrire la transition de système d'hétérogamétie observée entre *A. nasatum* et *A. vulgare*. La comparaison des séquences Y-spécifiques (*A. nasatum*) et W-spécifiques (*A. vulgare*) entre les deux génomes pourra indiquer si ces séquences sont situées dans des régions génomiques homologues, et donc si la transition de système d'hétérogamétie entre ces deux espèces est homologue (car ces séquences sont situées sur le même chromosome) ou non (Sarre et al., 2011). Dans cette perspective, une première analyse a été réalisée en recherchant les 6 *loci* Y-spécifiques identifiés chez *A. nasatum* dans le génome d'*A. vulgare* par *BLASTn* entre les séquences des amplicons de PCR des *loci* Y-spécifiques et l'ensemble du génome d'*A. vulgare* (Tableau 22).

Tableau 22 : Caractéristiques des séquences porteuses des *loci* Y-spécifiques d'*A. nasatum* identifiées chez *A. vulgare*. Le pourcentage d'identité indiqué correspond au % d'identité entre l'amplicon PCR des marqueurs Y-spécifiques identifiés chez *A. nasatum* et la séquence correspondante identifiée dans le génome d'*A. vulgare*.

Marqueurs Y-spécifiques (<i>A. nasatum</i>)		Séquences correspondantes identifiées chez <i>A. vulgare</i>				
Identifiant	Identifiant	% d'identité	% de répétitions	% de GC	Score d' <i>YGS</i>	Score de <i>CQ</i>
scf7180001668770	Contig28598	89%	54.58	27.84	1.3%	1.06
	Contig46097	85%	58.06	34.54	24.1%	0.81
scf7180001797263	Contig27594	86%	64.94	28.84	4.2%	1.06
scf7180001685477	aucun	-	-	-	-	-
scf7180001705172	Contig40169	91%	65.81	27.64	2.6%	0.99
scf7180001703882	Contig31602	89%	85.74	29.06	3.9%	0.9
scf7180001550276	aucun	-	-	-	-	-

Quatre des 6 *loci* Y-spécifiques d'*A. nasatum* sont présents dans le génome d'*A. vulgare*, dont 3 en copie unique et 1 dupliqué (Tableau 10). Les séquences identifiées chez *A. vulgare* présentent des scores d'*YGS* et de *CQ* autosomaux, ce qui semble indiquer que les *loci* Y-spécifiques d'*A. nasatum* ne correspondent pas

à une région W-spécifique ou Z-spécifique (les régions Z-spécifiques ont normalement un score de *CQ* proche de 2, Figure 32) lorsqu'ils sont présents chez *A. vulgare*. Bien que ces résultats soient préliminaires, ils pourraient être en accord avec une transition hétérogamétique non-homologue entre les deux espèces, selon la définition de Sarre et al. (2011). En effet, des chromosomes sexuels non-homologues peuvent évoluer suite à : (i) la duplication du *locus* déterminant le sexe et à sa transposition sur un autosome, comme décrit chez le poisson médaka (Tanaka et al., 2007) ou (ii) l'apparition d'un nouveau gène déterminant le sexe indépendamment des chromosomes sexuels préexistants (Sarre et al., 2011), comme dans le cas des reptiles *Ctenophorus fordi* et *Pogona vitticeps* (Ezaz et al., 2009). En revanche, si les séquences identifiées chez *A. vulgare* sont situées dans des régions pseudo-autosomales, alors la transition de système d'hétérogamétie entre *A. nasatum* et *A. vulgare* serait une transition homologue (Sarre et al., 2011). Quoi qu'il en soit, l'approche génomique utilisée dans les cas d'*A. nasatum* et d'*A. vulgare* constitue une première piste pour caractériser la transition de systèmes d'hétérogamétie entre ces deux espèces, et elle pourrait être complétée par des approches de peintures chromosomiques par hybridation *in situ* en fluorescence (Shetty et al., 1999).

II. Relations entre *A. nasatum* et sa souche de *Wolbachia* féminisante : *wNas*

Les résultats sur les lignées symbiotiques d'*A. nasatum* ont révélé des taux de transmission de *Wolbachia* fortement instables, étant soit de 0% (pour les portées NASw39, NASw40 et NASw32), soit de 100% (pour les portées NASw23 et NASw31) (Figure 43, Figure 44). Contrairement à *A. vulgare*, *A. nasatum* n'est connu pour être infecté naturellement que par une seule souche de *Wolbachia* nommée *wNas* (Juchault & Legrand, 1989; Bouchon et al., 1998; Cordaux et al., 2001, 2012). Dès sa découverte, les taux de transmission de cette souche de *Wolbachia* plus variables dans les populations d'*A. nasatum* que les souches de *Wolbachia* infectant les populations d'*A. vulgare* ont été relevés (Juchault & Legrand, 1989). La souche *wNas* est phylogénétiquement proche de la souche *wVulC* d'*A. vulgare* (Cordaux et al., 2004, 2012). Des expériences d'infections croisées (*wNas* transférée chez l'hôte *A. vulgare* et *wVulC* transférée chez l'hôte *A. nasatum*) ont permis de démontrer que la souche *wNas* se transmettait moins souvent à la descendance que la souche *wVulC*, peu importe que *wNas* infecte des individus *A. nasatum* ou *A. vulgare* (Rigaud et al., 2001). A l'époque, les auteurs suspectaient une plus forte spécialisation de *wNas* envers son hôte (*A. nasatum*) que de *wVulC* pour *A. vulgare* (Rigaud et al., 2001), sans pour autant parvenir à expliquer un tel écart de transmission entre les deux souches. Ainsi d'après cette étude, les fluctuations du taux de transmission observé chez *wNas* ne semblent pas s'expliquer par un effet causé par l'hôte mais semblent bien être une caractéristique inhérente de cette souche (Rigaud et al., 2001).

Le génotype des mères des portées NASw39, NASw40 et NASw32 a été inféré comme étant YY (Figure 43, Figure 44), et aucune de ces mères n'a transmis la bactérie à sa descendance. Le génotype de la femelle ayant donné naissance à la portée NASw31 (100% de transmission de *Wolbachia*) est considéré comme étant XX (Figure 44). Enfin, le génotype de la femelle à l'origine de la portée NASw23 est soit XY, soit YY (la portée NASw23 n'étant composée que de d'individus XY ou YY, Figure 43). Dans le cas où il s'agirait d'une femelle YY, il semblerait alors que ce génotype présente une forme d'incompatibilité à la transmission de *Wolbachia*, et non pas une forme de

résistance à l'effet féminisant de la bactérie puisque celle-ci est absente des descendants des femelles YY. Ce résultat est en désaccord avec les premières observations réalisées par Rigaud et al., (2001). Une telle résistance à la transmission de *Wolbachia* a d'ailleurs déjà été observé chez *A. vulgare* à travers l'émergence d'un système polygénique limitant la transmission de la bactérie à la descendance (Rigaud & Juchault, 1992). Les auteurs ont proposé que ce mécanisme polygénique pouvait soit : (i) limiter la densité de *Wolbachia* dans les tissus hôtes et limiter ainsi sa probabilité de transmission (comme cela a déjà été observé chez le coléoptère *Sitophilus oryzae*; Nardon & Grenier, 1989), soit (ii) diminuer directement le taux de transmission de *Wolbachia* (Rigaud & Juchault, 1992). Quel que soit le mécanisme impliqué dans ce système, ce contrôle génétique correspond à la définition de "gènes de résistance" proposée par Werren (1987), qui considère que ces gènes influencent indirectement le sex-ratio par leur effet sur les facteurs cytoplasmiques (à savoir *Wolbachia*).

Des travaux récents dans notre laboratoire ont permis de caractériser la présence de *Wolbachia* directement au sein des ovocytes des femelles d'*A. vulgare* infectées (Genty et al., 2014), offrant de nouvelles perspectives quant à la compréhension de la transmission de cet endosymbiote. Cette étude a démontré que la présence de la bactérie suivait la maturation des ovocytes : seulement 30% des ovocytes les plus récents étaient infectés, tandis que près de 90% des ovocytes matures portaient *Wolbachia* (Genty et al., 2014), ce qui correspond au taux de transmission de *Wolbachia* généralement observé dans cette espèce. Les auteurs ont proposé deux hypothèses non-exclusives pouvant expliquer ce pattern de colonisation des ovocytes : la bactérie est acquise secondairement par les ovocytes lors des derniers instants de la maturation, et/ou seuls les ovocytes infectés par *Wolbachia* sont sélectionnés pour être transmis (Genty et al., 2014). Il serait alors pertinent de réaliser un suivi *in situ* similaire pour les individus *A. nasatum* infectés et plus spécifiquement sur les femelles YY, afin de quantifier la présence de *Wolbachia* dans les ovocytes et de vérifier si les ovocytes matures de ces femelles sont colonisés de la même façon que des ovocytes matures issus de femelles XY ou XX porteuses de la bactérie.

Quoi qu'il en soit, contrairement à l'infection d'*A. vulgare* par *Wolbachia* qui conduit à terme à l'élimination du chromosome W dans les populations (suite à la disparition des femelles génétiques ZW; Taylor, 1990; Ferdy et al., 2016), l'instabilité de transmission de *wNas*, qui peut être due à la possible incompatibilité entre *wNas* et le génotype YY, semble empêcher la disparition du chromosome X (Juchault & Legrand, 1989). De plus amples investigations sont nécessaires pour mieux comprendre les patterns de transmissions de *wNas*, d'autant plus que les analyses ont été réalisées sur un faible nombre d'individus. Le cas d'*A. nasatum* pourrait constituer une nouvelle source d'information pour la compréhension de la dynamique de transmission et de diffusion de *Wolbachia* au sein des populations hôtes.

III. Caractérisation de l'homomorphie des chromosomes sexuels d'*A. nasatum*

Les premières études focalisées sur les caryotypes d'isopodes terrestres ont révélé qu'en plus d'avoir un grand nombre de chromosomes de petite taille, les chromosomes sexuels de ces espèces sont généralement indifférenciés et donc indiscernables (Juchault & Rigaud, 1995; Rigaud, 1997). Les chromosomes X et Y d'*A. nasatum* ne font pas exception et sont morphologiquement non identifiables (Juchault & Rigaud, 1995; Rigaud et al., 1997). Ils sont aussi fonctionnellement semblables puisque des individus présentant un phénotype mâles malgré un génotype XX peuvent être obtenus suite à une implantation de glande androgène lors de la phase de différenciation sexuelle (Juchault & Legrand, 1979; Becking et al., 2017). Ainsi le chromosome Y ne semble pas porter de gène essentiel au développement du phénotype mâle. De plus, la présence de bactéries *Wolbachia* féminisantes dans certaines populations d'*A. nasatum* induit la présence de femelle phénotypiques et fonctionnelles malgré leur génotype XY (Juchault & Legrand, 1989). Le génotype XX n'est donc pas indispensable à l'établissement d'un phénotype femelle. Enfin, les tests PCR réalisés sur les lignées d'*A. nasatum* infectées par *Wolbachia* ont permis de confirmer la viabilité fonctionnelle des mâles porteurs du génotype YY, mais aussi de confirmer l'existence d'individus YY féminisés par *Wolbachia* en femelles phénotypiques viables.

Ces résultats sont parfaitement en accord avec les analyses *in silico* réalisées sur le génome d'*A. nasatum*. En effet, l'alignement des *reads* de mâles et de femelles sur le génome assemblé avec *CeleraAssembler* a indiqué que les couvertures des deux ensembles de *reads* sont fortement corrélées (Figure 37), impliquant que les génomes mâles et femelles d'*A. nasatum* sont très fortement similaires. De plus, le faible nombre et la petite taille globale des séquences candidates retenues suite aux analyses *YGS* et *CQ* sur les différentes versions du génome d'*A. nasatum* reflètent la très forte homomorphie des chromosomes X et Y. Tous ces éléments sont en adéquation avec la présence de chromosomes sexuels morphologiquement très homomorphes, mais aussi fonctionnellement très similaires, puisque les individus porteurs des génotypes XX, XY ou YY peuvent

appartenir au sexe mâle ou femelle et être parfaitement viables. Les chromosomes X et Y ne diffèrent donc possiblement que sur la présence du signal primaire de la cascade de déterminisme du sexe, et l'ensemble des gènes impliqués en aval de cette cascade pourraient être autosomaux ou pseudo-autosomaux (Legrand et al., 1987).

L'homomorphie des chromosomes sexuels est une caractéristique assez commune chez de nombreuses espèces. Elle a été constatée chez des espèces de serpents (Vicoso et al., 2013a), d'oiseaux (Mank & Ellegren, 2007; Vicoso et al., 2013b), d'amphibiens (Stöck et al., 2011), de poissons (Kamiya et al., 2012) ou encore de diptères (McClelland, 1962; Newton et al., 1974; Traut, 2010; Hall et al., 2014). Le premier facteur pouvant expliquer cette homomorphie est l'âge récent des chromosomes sexuels. En effet, des chromosomes sexuels récents (correspondant à une acquisition récente d'un *locus* déterminant le sexe sur l'un des deux chromosomes d'une paire) peuvent ne pas avoir eu assez de temps pour être morphologiquement différenciés (Bachtrog, 2013). Cet argument peut en revanche s'avérer trompeur, puisqu'il a été démontré plusieurs fois que des chromosomes sexuels homomorphes n'étaient pas nécessairement récents et pouvaient même être assez anciens. En effet, les gonosomes de certaines espèces de pythons sont âgés d'environ 120 millions d'années alors qu'ils ne présentent pas de différence morphologique (Vicoso et al., 2013a). Il en va de même pour les chromosomes sexuels d'oiseaux ratites (tels que les autruches ou les émeus) qui sont âgés d'environ 140 millions d'années alors qu'ils sont eux aussi homomorphes (Vicoso et al., 2013b). Ces âges sont similaires à ceux des chromosomes sexuels de mammifères (âgés d'environ 180 millions d'années, Cortez et al., 2014), dont le chromosome Y est pourtant connu pour être fortement dégénéré (voir la revue de Graves, 2006).

Dans le cas d'*A. nasatum*, l'hypothèse d'un âge récent des chromosomes sexuels est une explication pertinente puisque, comme évoqué précédemment, l'âge du chromosome Y d'*A. nasatum* est estimé à moins de 25 Ma (Chapitre 4, Figure 63, Becking et al., 2017). Afin d'essayer de préciser l'âge du chromosome Y d'*A. nasatum*, nous avons testé si *Armadillidium tunisiense* possède le même chromosome sexuel Y qu'*A. nasatum*. *A. tunisiense* a été choisi parce qu'il s'agit à

notre connaissance de l'espèce phylogénétiquement la plus apparentée à *A. nasatum*, dont la divergence est estimée à environ 15 Ma (Chapitre 4, Figure 63, Becking et al., 2017). Les 6 marqueurs Y-spécifiques d'*A. nasatum* ont été testés individuellement sur 6 femelles et 4 mâles *A. tunisiense* issus de notre unique population de laboratoire. Trois de ces 6 marqueurs n'ont montré aucune amplification, 2 marqueurs ont montré des amplifications aspécifiques (confirmées par séquençage) à la fois chez des mâles et des femelles. Enfin, L'amplification du dernier marqueur a généré un fragment de la taille attendue (spécificité de l'amplicavation confirmée par séquençage) chez tous les mâles et toutes les femelles *A. tunisiense*. Aucun marqueur ne s'est donc avéré sexe-spécifique chez *A. tunisiense*, suggérant que cette espèce ne dispose pas du même chromosome Y qu'*A. nasatum*. Compte tenu de ces résultats, le chromosome Y d'*A. nasatum* apparaîtrait donc plus récent que la divergence entre *A. nasatum* et *A. tunisiense* (à savoir environ 15 Ma, Chapitre 4, Figure 63, Becking et al., 2017), suggérant que ce chromosome Y aurait bien un âge récent. Néanmoins, d'autres analyses sont cependant nécessaires pour totalement valider cette conclusion.

Un autre facteur pouvant expliquer le maintien de l'homomorphie est la présence d'une longue phase haploïde dans le cycle de vie de l'animal (Bachtrog, 2013). Une longue phase haploïde (et donc une expression des gènes de l'hétérochromosome au cours de cette phase) peut résulter en une forte pression de sélection purifiante pour assurer le maintien des gènes présents sur le chromosome Y ou le W. Si de tels cas ont été constatés chez des plantes (Chibalina & Filatov, 2011; Ahmed et al., 2014; Bergero et al., 2015) avec notamment une forte expression génique au sein de grains de pollens haploïdes (Schäfer et al., 1995), ceci ne peut être valide dans le cas des isopodes terrestres puisque la phase haploïde se résume uniquement aux gamètes.

Une autre explication pour le maintien de l'homomorphie des gonosomes repose sur un phénomène de "fontaine de jouvence", grâce à la présence en populations naturelles d'individus dont le sexe phénotypique diffère du sexe génétique (Perrin, 2009). Chez de nombreuses espèces, les autosomes (qui sont partagés par les deux sexes) montrent des patterns de recombinaison sexe-

spécifique (on parle d'hétérochiasmie) qui touchent à la fois le nombre et la localisation des phénomènes d'enjambement (Lenormand, 2003; Lenormand & Dutheil, 2005). En effet, le ratio des taux de recombinaison femelle/mâle varie fortement selon l'espèce considérée (de 0.14 à 14, Coimbra et al., 2003; Berset-Brändli et al., 2008), sans compter l'exemple des espèces qui ne recombinent que chez un seul sexe comme chez *D. melanogaster* (Morgan, 1914) ou chez *B. mori* (Tanaka, 1914). L'hétérochiasmie entre mâle et femelle touche aussi des espèces avec ESD qui sont pourtant génétiquement identiques, comme dans l'exemple de *Crocodylus porosus* (Isberg et al., 2006). Cette hétérochiasmie dépendante du sexe phénotypique est d'ailleurs confirmée dans certaines espèces chez des individus dont le sexe a été réversé expérimentalement. Ceci est illustré par l'exemple du triton crête *Triturus cristatus* chez qui les femelles XY montrent le même pattern de recombinaison que les femelles XX (Wallace et al., 1997). C'est également le cas chez le médaka *O. latipes* (Kondo et al., 2001). La recombinaison entre le X et le Y a pour effet de contrer le cliquet de Müller et ainsi d'empêcher la dégradation du Y, ce qui a été démontré notamment chez plusieurs espèces d'amphibiens (Stöck et al., 2011, 2013; Dufresnes et al., 2015). Par analogie, dans le cas d'*A. nasatum*, les femelles XY infectées par *Wolbachia* pourraient présenter le même pattern de recombinaison que les femelles XX et donc permettre la recombinaison occasionnelle entre le X et le Y chez les femelles (alors que ce n'est pas le cas chez les mâles), et maintenir ainsi l'homomorphie des chromosomes sexuels. Pour tester cette hypothèse, il apparaît utile de déterminer si le taux de recombinaison entre les mâles et les femelles phénotypiques est différent chez *A. nasatum*, y compris chez des individus sexe-réversés (femelles XY infectées par *Wolbachia* et mâles XX après implantation de glande androgène au cours du développement sexuel de l'individu). De plus, les résultats obtenus concernant les tests des marqueurs Y-spécifiques effectués sur les pédigrées des individus porteurs de *Wolbachia* ont permis de démontrer la présence d'individus YY (mâles ou femelles). Il est donc probable que la recombinaison soit toujours effective au niveau de la région Y-spécifique chez ces individus, induisant là aussi un maintien de l'état homomorphe du chromosome Y d'*A. nasatum*. Quoi qu'il en soit, la présence de *Wolbachia* pourrait donc permettre le maintien des phénomènes de recombinaison au niveau de la région Y-spécifique d'*A. nasatum*.

Enfin, il est connu que les conflits génétiques (correspondant à la situation où un élément génétique sélectionné conduit à la diminution de la valeur adaptative d'un autre élément) jouent un rôle majeur dans l'évolution morphologique des chromosomes sexuels (voir la revue de Bachtrog, 2013). Les modèles théoriques prédisent que plus des conflits génétiques sont présents et importants, plus l'expansion de la zone non-recombinante de l'hétérochromosome est rapide suite à l'accumulation sur l'hétérochromosome de gènes bénéfiques au sexe hétérogamétique (Charlesworth & Mank, 2010). Ces conflits génétiques intragénomiques tendent ainsi à accélérer le processus de dégradation de l'hétérochromosome. Cependant, ces conflits génétiques peuvent aussi être responsables de turnovers de chromosomes sexuels, créant ainsi de nouveaux gonosomes, et réinitialisant les processus de dégénérescence de l'hétérochromosome (van Doorn & Kirkpatrick, 2010). Un tel turnover contraint les chromosomes sexuels à un faible degré d'hétéromorphie (à la fois morphologiquement et fonctionnellement), car des changements de position réguliers des allèles déterminant le sexe limitent l'accumulation de différences entre les chromosomes X et Y (ou Z et W), ces derniers demeurant largement recombinants (Charlesworth, 1991, 1996). Dans le cas d'*A. nasatum*, ces conflits génétiques se produiraient entre le/les déterminant(s) du sexe portés par le chromosome Y à effet masculinisant et la présence de *Wolbachia* comme perturbateur du sex-ratio en faveur des femelles. Comme décrit dans l'introduction générale, le conflit entre *Wolbachia* féminisante et déterminant du sexe masculinisant pourrait induire des changements récurrents de chromosomes sexuels, assurant ainsi le maintien de l'homomorphie des gonosomes (Juchault & Rigaud, 1995; Rigaud et al., 1997).

En conclusion, l'homomorphie extrême des chromosomes sexuels d'*A. nasatum* peut s'expliquer principalement par trois facteurs mutuellement non exclusifs : (i) l'existence d'individus XY femelles en raison de la présence de *Wolbachia* dans des populations naturelles qui pourrait permettre la recombinaison du chromosome Y, (ii) l'existence d'individus YY femelles, chez qui la recombinaison au niveau de la région Y-spécifique est toujours possible et

(iii) des conflits génétiques entre les déterminants du sexe d'*A. nasatum* et la présence de *Wolbachia* féminisantes, pouvant être responsable de l'âge récent du chromosome Y (moins de 15 Ma). Tous ces facteurs ont pour dénominateur commun *Wolbachia*, qui pourrait ainsi jouer un rôle majeur dans l'évolution des chromosomes sexuels d'*A. nasatum* (Juchault & Rigaud, 1995; Rigaud et al., 1997; Cordaux et al., 2011; Becking et al., 2017).

Chapitre 3

**Caractérisation de gènes candidats
du déterminisme du sexe et mesure
de leur expression au cours du
développement**

Introduction

La famille des gènes à domaine DM est une des familles de gènes les plus étudiées, notamment car elle contient de nombreux effecteurs majeurs impliqués dans plusieurs fonctions développementales, dont le déterminisme ou la différenciation sexuelle (Picard et al., 2015). C'est le cas par exemple du médaka *Oryzias latipes* (Matsuda et al., 2007), du xénope *Xenopus laevis* (Yoshimoto et al., 2008), du planaire *Schmidtea mediterranea* (Chong et al., 2013) ou encore du branchiopode *D. magna* (Kato et al., 2011). Ces gènes ont tout d'abord été décrits chez *D. melanogaster* (Hildreth, 1965), chez qui le gène *Doublesex* présente des variants d'épissage alternatif spécifiques à chaque sexe (le variant femelle-spécifique ayant un rôle d'activateur de gènes cibles, tandis que le variant mâle a un rôle répresseur) (Burtis & Baker, 1989). L'existence de ces gènes chez des organismes non-bilatériens (tels que les coraux ou les placozoaires) suggère que leur origine est antérieure à l'apparition des eumétazoaires (Wexler et al., 2014). Les gènes à domaine DM sont des facteurs de transcription caractérisés par un motif en double doigts de zinc appelé le domaine DM (Erdman & Burtis, 1993; Zhu et al., 2000), formé de six histidines et de deux cystéines formant le motif de liaison à l'ADN (Zhu et al., 2000).

Il est généralement admis qu'un gène peut être considéré comme un déterminant (ou un différenciant) du sexe lorsque l'altération de son expression induit une réversion du sexe (Picard et al., 2015). Cependant, l'identification d'un gène comme étant impliqué dans les cascades de déterminisme/différenciation peut aussi prendre en compte la fenêtre temporelle d'expression de ce gène (si elle correspond à la fenêtre temporelle de différenciation sexuelle), ou encore la stricte nécessité de ce gène à l'établissement d'un phénotype sexuel donné (Picard et al., 2015). Bien que les gènes à domaine DM soient présents chez la grande majorité des eumétazoaires, la fonction (liée au déterminisme du sexe ou non) de la plupart d'entre eux demeure encore inconnue (Picard et al., 2015).

D'autres acteurs moléculaires essentiels à la cascade de déterminisme du sexe des arthropodes, identifiés tout d'abord chez les insectes, apparaissent

fortement conservés au cours de l'évolution : les gènes *Transformer* (*Tra*) et *Transformer-2* (*Tra-2*) (voir la revue de Verhulst et al., 2010). Chez les insectes, ces gènes sont largement répandus et sont connus pour agir en amont du gène *Doublesex* dans la cascade du déterminisme du sexe (Sánchez, 2008; Verhulst et al., 2010). Ces deux gènes en association servent de régulateur de l'épissage alternatif de *Doublesex*, responsable de la formation de variants spécifiques aux mâles ou de variants spécifiques aux femelles (Verhulst et al., 2010). Les gènes *Tra* et *Tra-2* ont aussi été découverts chez différentes espèces de crustacés, sans pour autant qu'un lien avec le déterminisme du sexe n'ait pu être clairement démontré (Li et al., 2012; P. Chen et al., 2014; Kong et al., 2015; Sellars et al., 2015).

La recherche de tels gènes chez *A. nasatum* pourrait permettre de commencer à caractériser la cascade de déterminisme du sexe chez cet isopode terrestre, celle-ci étant encore totalement inconnue. L'assemblage du génome d'*A. nasatum* (Chapitre 1, page 58) ainsi que son annotation structurale et fonctionnelle constituent des ressources de choix pour la caractérisation des gènes DM et *Tra/Tra-2* chez cette espèce. Afin d'évaluer s'ils ont un rôle dans le déterminisme du sexe, un premier indice réside dans la mesure de leur expression au cours de la différenciation sexuelle. En effet, un pattern d'expression spécifique à l'un des deux sexes serait cohérent avec une potentielle implication dans le déterminisme ou la différenciation du sexe. Une correspondance entre une variation temporelle de l'expression de ces gènes et la fenêtre temporelle de la différenciation sexuelle constituerait une indication supplémentaire du rôle potentiel de ces gènes dans le déterminisme ou dans la différenciation sexuelle. A terme, l'assemblage et l'annotation du génome de l'espèce apparentée à hétérogamétie femelle *A. vulgare* (Chebbi et al, en préparation) représentent une opportunité rare de comparer l'impact d'une transition entre deux systèmes d'hétérogamétie (XY/XX pour *A. nasatum* et ZZ/ZW pour *A. vulgare*) sur l'expression de ces gènes, et ainsi d'évaluer les conséquences d'une telle transition sur les mécanismes moléculaires du déterminisme et/ou de la différenciation du sexe.

L'approche par gènes candidats utilisée dans ce chapitre est complémentaire de l'approche sans *a priori* réalisée dans le Chapitre 2. Il s'agit

d'une opportunité supplémentaire d'identifier des acteurs moléculaires potentiellement impliqués dans le déterminisme du sexe. Ainsi, le premier objectif de ce troisième chapitre de thèse était d'identifier et de caractériser les gènes à domaines DM ainsi que les gènes *Tra/Tra-2* présents dans les génomes d'*A. nasatum* et d'*A. vulgare*. Suite à leur caractérisation, le second objectif était de quantifier l'expression de ces gènes au cours du développement des mâles et des femelles, en suivant les différents stades de mue post-nataux. Comme décrit dans l'introduction générale (voir page 40), les isopodes terrestres ne se différencient sexuellement qu'après plusieurs stades de mue post-naissance (Katakura, 1984; Suzuki & Yamasaki, 1995; Badawi et al., 2015). Ainsi il est impossible de connaître le sexe d'un individu par la morphologie avant sa différenciation sexuelle, ce qui complique l'étude des gènes dans le contexte du déterminisme du sexe. Les juvéniles *A. nasatum* ont donc été sexés grâce aux marqueurs moléculaires développés lors du second chapitre de thèse, et le suivi de l'expression des gènes caractérisés a été effectué du second stade de mue au stade pleinement différencié (stade 6) ainsi qu'au stade adulte. Les marqueurs du chromosome W d'*A. vulgare* n'étant pas encore disponibles au moment de cette étude, l'expression des gènes à domaine DM et *Tra/Tra-2* caractérisés a été effectuée seulement pour le sixième stade de mue et le stade adulte chez cette espèce.

Matériels et Méthodes

I. Caractérisation *in silico* de gènes potentiellement impliqués dans le déterminisme du sexe

La recherche de gènes potentiellement impliqués dans le déterminisme du sexe a tout d'abord été effectuée à partir des fichiers d'annotation des génomes d'*A. nasatum* et d'*A. vulgare* (ce dernier généré dans le cadre de la thèse d'Amine Chebbi). Ainsi, toutes les séquences dont l'annotation indiquait des similarités avec des séquences connues de type *Doublesex*, *DMRT*, *Transformer*, ou encore *Transformer-2* ont été relevées. Cette recherche a été complétée par des recherches de similarité avec *BLASTx* (version 2.2.23) entre les séquences en nucléotides des deux génomes et une base de données protéiques. Celle-ci était composée de 1 058 protéines à domaine DM (*Doublesex* et *DMRT*) et de 1 164 protéines *Transformer* (et *Transformer-2*) (les numéros d'accession composant cette base de données sont listés dans l'Annexes IV).

Les gènes identifiés ont été positionnés dans un contexte phylogénétique afin de caractériser plus précisément la famille à laquelle ces gènes appartiennent. Par exemple, les gènes à domaine DM sont répartis dans plusieurs familles géniques, dans lesquelles sont retrouvés les homologues des gènes *Doublesex* ou encore les homologues des différents gènes *DMRT*. Les arbres phylogénétiques ont été générés à partir de protéines à domaine DM et de gènes *Transformer* connus (cf : Annexe V et Annexe VI pour la liste des numéros d'accession) ainsi qu'à partir des séquences protéiques prédites identifiées dans les génomes d'*A. nasatum* et *A. vulgare*. Les alignements multiples ayant permis la réalisation de ces arbres ont été effectués en utilisant les paramètres par défaut de la fonction d'alignement native du logiciel *Geneious* (version 7.0.6, Kearse et al., 2012). Les régions présentant des ambiguïtés d'alignement ont été retirées avec le logiciel *GBLOCKS* (version 0.91b ; Talavera & Castresana, 2007) en utilisant les paramètres par défaut. Les arbres ont été reconstruits avec la méthode du maximum de vraisemblance (*Maximum Likelihood*) implémentée dans le logiciel

RAxML (version 7.4.6 ; Stamatakis, 2006). Chaque arbre a été généré à partir de 100 topologies de départ différentes, et la robustesse des arbres finaux a été évaluée par 1 000 tests de *bootstrap*. Cette procédure s'appuie sur des alignements ré-échantillonnés de même taille que l'alignement initial. Enfin, les modèles d'évolution servant à la construction de chaque arbre ont été définis avec *Protest* (Version 3.4, Darriba et al., 2011). Selon les différents critères d'information donnés par *Protest*, le modèle JTT+G était le meilleur modèle évolutif pour l'arbre des gènes à domaine DM, et le modèle LG+G+F était le modèle le plus adéquat pour l'arbre des gènes *Transformer-2*.

II. Mesure de l'expression des gènes candidats au cours du développement

1. Echantillonnage de juvéniles pour étudier l'expression des gènes candidats au cours du développement

Pour *A. nasatum*, les individus analysés étaient les descendants des individus utilisés pour le séquençage *PacBio* (dont la population originelle provient de Thuré, 79, France). Dans le cas d'*A. vulgare*, les individus prélevés étaient issus de croisements entre deux lignées asymbiotiques du laboratoire : la lignée BF en provenance de Nice (06, France) et la lignée ZM en provenance d'Héraklion (Grèce). Pour les deux espèces, les juvéniles ont été collectés individuellement deux à quatre jours après chaque mue, et ce de la première mue correspondant à des individus indifférenciés (individus de stade 1) à la sixième mue correspondant à des individus totalement différenciés (stade 6), puis congelés dans de l'azote liquide. Des individus adultes ont aussi été analysés, issus des mêmes lignées que les juvéniles.

2. Sexage des juvéniles *A. nasatum*

Les isopodes terrestres étant sexuellement indifférenciés au cours de leurs premiers stades de développement post-naissance (Katakura, 1984; Suzuki & Yamasaki, 1995; Badawi et al., 2015), il est nécessaire de procéder à un sexage moléculaire des individus si l'on veut étudier l'expression sexe-spécifique des gènes identifiés *in silico* au cours du développement. Afin d'obtenir des pools d'ARN de mâles et de femelles séparément, il était nécessaire d'extraire indépendamment l'ADN de chaque individu (pour le sexage) et dans un second temps l'ARN de pools d'individus sexés (pour étudier l'expression des gènes). Dans le but de limiter la perte de matériel biologique par individu (c'est-à-dire dans l'objectif de maximiser le rendement de l'extraction ARN), un protocole d'extraction d'ADN particulier a été mis au point, adapté de la méthode décrite par Carvalho et al. (2009). Une à deux pattes de chaque individu ont été prélevées et placées directement dans 10 μ L de protéinase K (20 mg.mL⁻¹). Les pattes ont

ensuite été incubées durant 1 h à 56°C, puis 2 minutes à 95°C afin d'inactiver la protéinase K. Le sexage moléculaire n'a pu être réalisé que pour l'espèce *A. nasatum*, les marqueurs du chromosome W d'*A. vulgare* n'étant pas encore disponibles au moment de ces expériences. La qualité des ADN extraits a été vérifiée à l'aide d'amorces génomiques ciblant le gène *ARNr 18S* (la liste des amorces générales est disponible dans l'Annexe VII). Les individus d'*A. nasatum* ont été sexés en utilisant trois marqueurs Y-spécifiques validés (scf7180001685477, scf7180001550276 et scf7180001797263).

3. Quantification de l'expression des gènes candidats au cours du développement des mâles et femelles *A. nasatum* et *A. vulgare*

Afin d'obtenir suffisamment de matériel pour les analyses d'expression en RT-qPCR, les extractions d'ARN ont été réalisées sur des pools d'individus sexés. Pour chacun des deux sexes, dix juvéniles de stade 2 ont été poolés, neuf pour le stade 3, six pour le stade 4, quatre pour le stade 5 et trois pour le stade 6 et les adultes. Ces pools ont été effectués en trois répliques par sexe et par stade de développement. L'ARN des échantillons a été extrait à partir du kit d'extraction ARN *TRIzol* (ThermoFisher Scientific) en suivant le protocole fourni. La concentration d'ARN a été mesurée avec un *Qubit 2.0* (Invitrogen). Afin de s'assurer de l'homogénéité de dégradation des différents ARN extraits, leur qualité a été évaluée avec un système de migration par électrophorèse *MultiNA* (Shimadzu).

L'ARN des échantillons extraits a reçu un traitement DNase (*RNase-Free DNase Set*, Qiagen) avant de réaliser une transcription inverse (kit *SuperScript II Reverse Transcriptase*, ThermoFisher), à partir de 250 ng d'ARN issus de chaque échantillon (Annexe II pour la description détaillée du protocole). L'expression des gènes candidats a tout d'abord été testée par PCR non quantitative sur les ADN complémentaires (ADNc) obtenus après l'étape de transcription inverse, suivie d'une migration sur gel d'agarose afin de s'assurer de la spécificité des amplifications (Annexe II pour la description détaillée du protocole).

Des amorces de PCR quantitative ont été dessinées pour les gènes d'intérêt identifiés *in silico*. Elles ont été réalisées de sorte qu'elles puissent amplifier l'ADN complémentaire à la fois d'*A. vulgare* et celui d'*A. nasatum*. Ces amorces ont été dessinées à l'aide du logiciel *Primer3Plus* (version 2.4.0, Untergasser et al., 2012) réglé sur l'option qPCR. La liste des amorces utilisées est indiquée dans le Tableau 23. De plus, la présence de deux copies d'un homologue du gène *Dsx*, identifiées chez *A. nasatum*, a aussi été vérifiée par PCR (Tableau 23), et ce au sein des deux sexes. La spécificité des amorces a été vérifiée par *BLASTn* en alignant les amorces dessinées contre les génomes d'*A. nasatum* et d'*A. vulgare*.

Tableau 23 : Liste des amorces utilisées pour l'analyse de l'expression de gènes candidats impliqués dans le déterminisme du sexe chez *A. vulgare* et *A. nasatum*. Les amorces pour les gènes de référence *Elongation factor 2* et *Rbl8* sont issues de Chevalier et al. (2012).

Locus	Séquences des amorces		Taille du produit de PCR (nucléotides)	Température de fusion
	Forward	Reverse		
<u>gènes candidats</u>				
DSX-like	CCCCAGGGTGAAATGTTTGATG	TCCAGAACTTG TAGAGGCTTCG	126	60°C
DMRT99B-like	TCAAAGGACACCCAAATGCG	AGCGTTTATGTCCCTTCAGC	71	60°C
DMRT93B-like	ACAGTAGCTTACCCGACAACG	CATTCCGTGATTTCGACACCTC	72	60°C
Transformer2-A-like	GCTCTGGGATGGA AATTGACG	TGGAGTGTGTGGTCTTTCG	72	60°C
Transformer2-B-like	TCACCAAAACGAAGCAGGAG	TCCGTGTGTAGAGACTGAGAC	143	60°C
<u>gènes de référence</u>				
Elongation Factor 2	GTCAGATCATTCCACAACA	TTTCCTCAAATACCACTCCAC	164	60°C
Rbl8	AACTGGAGATAGAGGCAAA	CCACCAGCAGCAATTC	159	60°C
<u>Test de la présence de deux copies de DSX chez A. nasatum</u>				
5150/21795-DSX	CCTTATCGTGATTGCAACTG	CACGTCGTAAAGGAGACTTG	414	52°C

L'expression des gènes candidats potentiellement impliqués dans le déterminisme du sexe a ensuite été quantifiée selon le sexe au cours du développement par PCR quantitative (*LightCycler 480*, Roche ; Annexe II pour la description détaillée du protocole). Pour chaque condition (sexe, stade de développement), les mesures ont été réalisées en triplicatas biologiques, eux-mêmes réalisés en duplicatas techniques (donc 6 mesures effectuées par condition). Les mesures ont été réalisées pour les stades 2, 3, 4, 5, 6 et adultes pour *A. nasatum*. En revanche, puisqu'il n'a pas été possible de sexer à l'aide de marqueurs moléculaires les juvéniles *A. vulgare*, les mesures ont été effectuées

seulement pour les stades 6 et adultes chez cette espèce (seuls stades pouvant être sexés phénotypiquement avec certitude).

L'expression relative des gènes se calcule en fonction de l'expression de gènes de référence, dont l'expression est stable d'une condition à l'autre (ici une expression stable entre les sexes, et au cours du développement). Les deux gènes de référence utilisés sont issus de précédents travaux réalisés au laboratoire sur *A. vulgare*, à savoir les gènes *Rbl8* et *Elongation Factor 2* (Chevalier et al., 2012). Afin de confirmer ces gènes de référence chez *A. nasatum* et selon nos conditions d'étude, la stabilité d'expression de ces gènes, le logiciel *geNorm* a été utilisé (Vandesompele et al., 2002). Les auteurs de cette étude préconisent que le facteur de stabilité (appelé *M*) des gènes de référence doit être ≤ 0.7 , sachant que plus ce facteur est élevé, moins les gènes étudiés sont exprimés de façon stable entre les différentes conditions. Dans notre cas, l'utilisation des gènes *Rbl8* et *Elongation factor 2* présente une valeur $M = 0.440$. L'expression des deux gènes de référence utilisés en fonction des stades de développement est représentée en Figure 45.

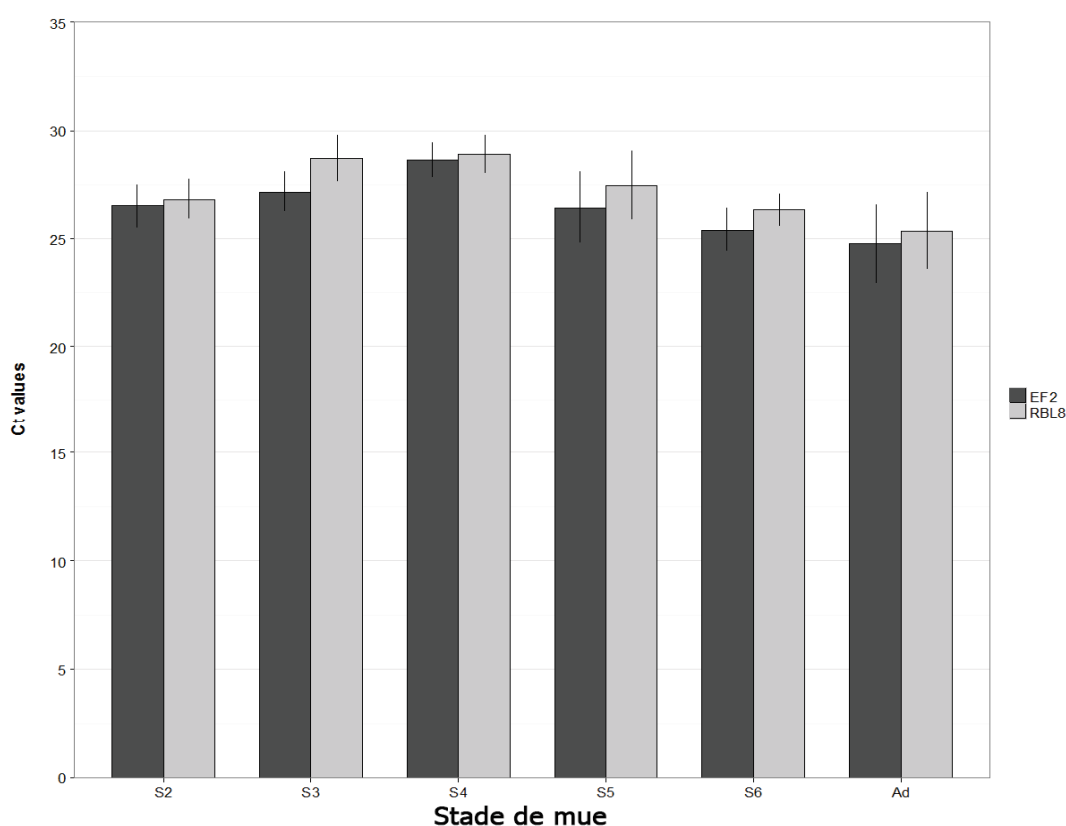


Figure 45 : Expression des gènes de référence *Rbl8* et *Elongation Factor 2* en fonction des stades de développement. Les histogrammes correspondent à la valeur moyenne des Ct mesurés pour chaque stade de développement et les barres verticales au sommet des histogrammes représentent l'écart-type.

Le calcul de l'expression relative des gènes candidats a été réalisé en suivant la méthode du $2^{-\Delta\Delta Ct}$ (Livak & Schmittgen, 2001), c'est à dire que l'expression d'un gène normalisée (par les gènes de référence) est indiquée en fonction de l'expression de ce même gène normalisé à un stade donné, et ce indépendamment pour chaque sexe. Dans le cas d'*A. nasatum*, l'expression a été normalisée par rapport au stade 2 (car il correspond à un stade où la différenciation sexuelle n'a pas encore débuté) tandis que pour *A. vulgare*, ce ratio a été calculé par rapport au stade adulte. Le calcul se décompose de la façon suivante :

$$-\Delta\Delta Ct = -(\Delta Ct_{\text{stade } x} - \Delta Ct_{\text{stade 2}})$$

avec :

$$\Delta Ct_{\text{stade } x} = \text{Moyenne}(Ct_{\text{gène cible, stade } x}) - \text{Moyenne}(Ct_{\text{gène référence, stade } x})$$

$$\Delta Ct_{\text{stade 2}} = \text{Moyenne}(Ct_{\text{gène cible, stade 2}}) - \text{Moyenne}(Ct_{\text{gène référence, stade 2}})$$

Ct = Cycle threshold (cycle à partir duquel l'amplification PCR est mesurable)

Afin d'obtenir une valeur correspondant au ratio d'expression d'un gène donné par rapport au stade 2, la formule donne enfin :

$$\text{Ratio calibré} = 2^{-\Delta\Delta Ct}$$

Afin de pouvoir comparer l'expression des gènes cibles entre *A. nasatum* et *A. vulgare*, les résultats obtenus pour *A. nasatum* ont aussi été calibrés par rapport au stade adulte.

4. Analyses statistiques

Pour chaque condition (sexe, stade de développement), les $\Delta\Delta Ct$ ont été testés pour la normalité (test de Shapiro-Wilk) et pour l'homoscédasticité des variances (test de Bartlett). Concernant les comparaisons intra-sexe de la variation de l'expression des gènes au cours du développement, les $\Delta\Delta Ct$ des gènes respectant les critères de normalité et d'homoscédasticité ont été analysés par des tests de type ANOVA (1 facteur). Lorsque l'hypothèse nulle (H_0 : pas de différence entre conditions) a été rejetée, la significativité des différences a été vérifiée par des tests de Tukey HSD. Dans le cas des comparaisons intra et inter-sexe, l'expression des gènes respectant les critères de normalité et d'homoscédasticité a été analysée par des tests de type ANOVA à 2 facteurs (sexe et stade de développement). Lorsque l'hypothèse nulle a été rejetée, la significativité des différences a été vérifiée par l'intermédiaire de tests de différence significative minimale (LSD) de Fisher. Enfin, dans le cadre des comparaisons intra ou inter-sexe ne respectant pas ces critères de normalité et/ou d'homoscédasticité des variances, des tests de Kruskal-Wallis, suivis de tests de comparaisons multiples de Dunn ont été réalisés. L'ensemble de ces tests statistiques a été réalisé sous *R* (version 3.3.0, R Core Team, 2014)

Résultats

I. Caractérisation *in silico* de gènes candidats impliqués dans le déterminisme du sexe

La recherche de gènes potentiellement impliqués dans le déterminisme du sexe chez *A. nasatum* et *A. vulgare* a été effectuée à partir des fichiers d'annotation issus du génome de référence des deux espèces (Chapitre 1, page 58 et Chebbi et al., en préparation), en plus d'une recherche par homologie de séquence vis-à-vis d'une base de données de gènes à domaines DM et de gènes homologues de *Transformer*. Ainsi, 3 séquences porteuses d'un domaine DM ont été identifiées dans le génome d'*A. vulgare*, et 4 dans le génome d'*A. nasatum*. De plus, si aucun homologue de *Transformer* n'a pu être identifié dans aucun des deux génomes, 2 séquences homologues au gène *Transformer-2* ont été identifiées dans les génomes d'*A. nasatum* et d'*A. vulgare*.

Afin de caractériser les familles auxquelles appartiennent ces gènes, des phylogénies ont été établies à partir des séquences protéiques pour les deux catégories de gènes identifiées. Un premier arbre phylogénétique a été réalisé à partir d'un alignement de 48 acides aminés incluant 44 domaines DM connus (les numéros d'accession sont précisés dans l'Annexe V), en plus de ceux identifiés chez *A. nasatum* et *A. vulgare* (Figure 46). La topologie résultante montre trois groupes monophylétiques (*Doublesex*, *DMRT1*, et *DMRT2/11*) et quatre groupes paraphylétiques (*DMRT99B*, *DMRT4*, *DMRT5* et *DMRT3/93B*). Même si la topologie de l'arbre n'est pas parfaitement résolue, elle permet malgré tout d'identifier clairement trois gènes appartenant à la famille des gènes *Doublesex* : le gène AvulDsx (*A. vulgare*), le gène AnasDsx1 (*A. nasatum*) et le gène AnasDsx2 (*A. nasatum*). Le gène *Dsx* a donc subi un événement de duplication chez *A. nasatum* ou une délétion chez *A. vulgare*, dans les deux cas postérieur à la divergence entre *A. vulgare* et *A. nasatum*.

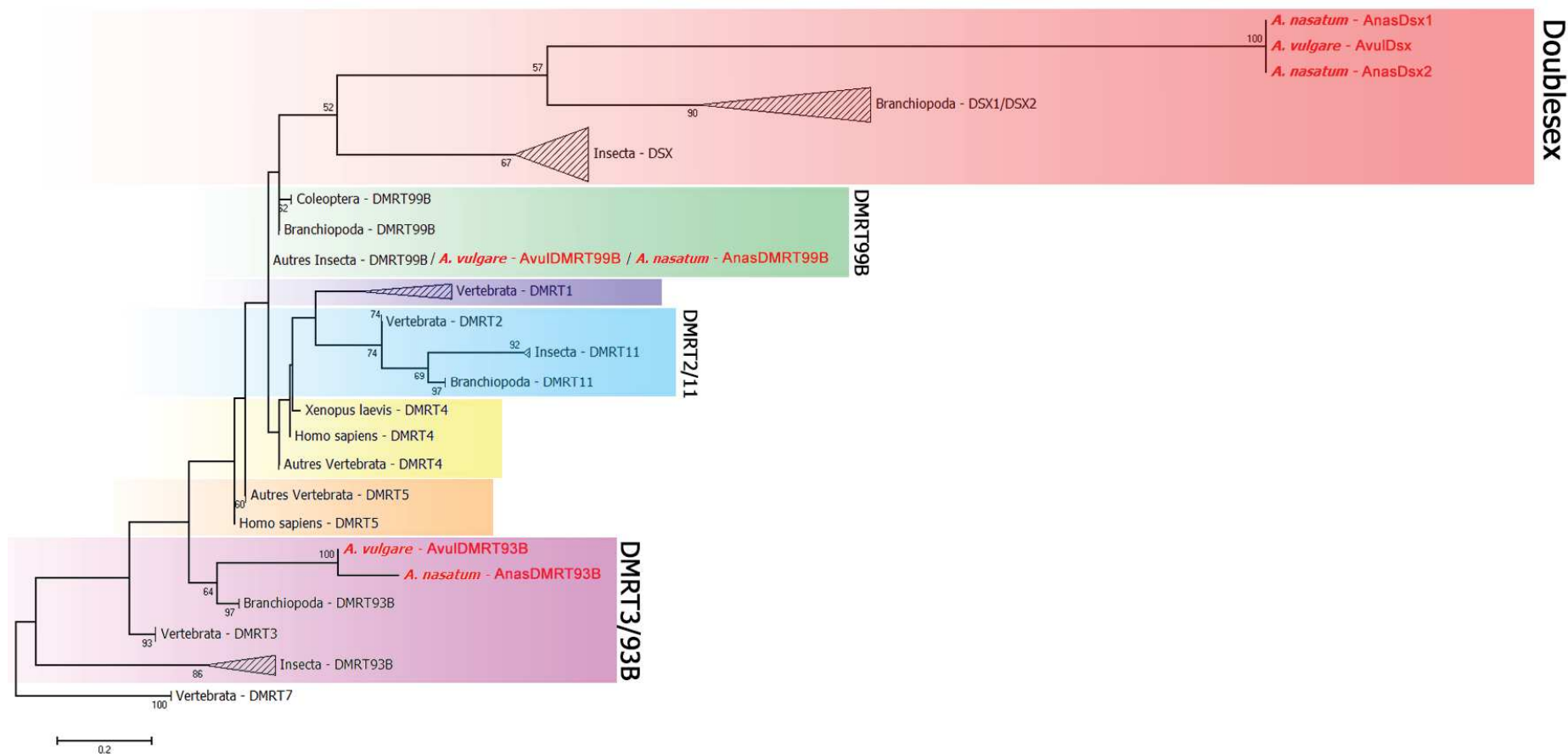


Figure 46 : Arbre phylogénétique de différents gènes à domaines DM connus et des domaines DM identifiés chez *A. nasatum* et *A. vulgare*. La longueur des branches de l'arbre est proportionnelle à la distance évolutive. Les nombres indiqués aux différents nœuds correspondent aux scores de *bootstrap* obtenus (les valeurs inférieures à 50 ne sont pas indiquées). L'arbre a été enraciné avec les domaines DM des gènes *DMRT7* appartenant à la souris (NP_082008) et à la vache (ABQ14481).

L'hypothèse d'une duplication artificielle issue de l'assemblage est à exclure, puisque la présence génomique des deux copies chez *A. nasatum* a été confirmée par PCR dans un panel de 3 mâles et de 3 femelles. Les alignements réalisés sur la zone commune au trois *contigs* porteurs des gènes *Dsx* chez *A. nasatum* et *A. vulgare* (Figure 47, A) suggèrent que le gène *AnasDsx1* serait orthologue au gène *AvulDsx*. En effet, l'alignement d'environ 6.9 kb commun au trois *contigs* révèle que le Contig5150 (porteur du gène *AnasDsx1*) présente 89.7% d'identité nucléotidique avec le Scaffold1990 (porteur du gène *AvulDsx*), et seulement 81.3% avec le Contig21759 (porteur du gène *AnasDsx2*) (Figure 47, B).

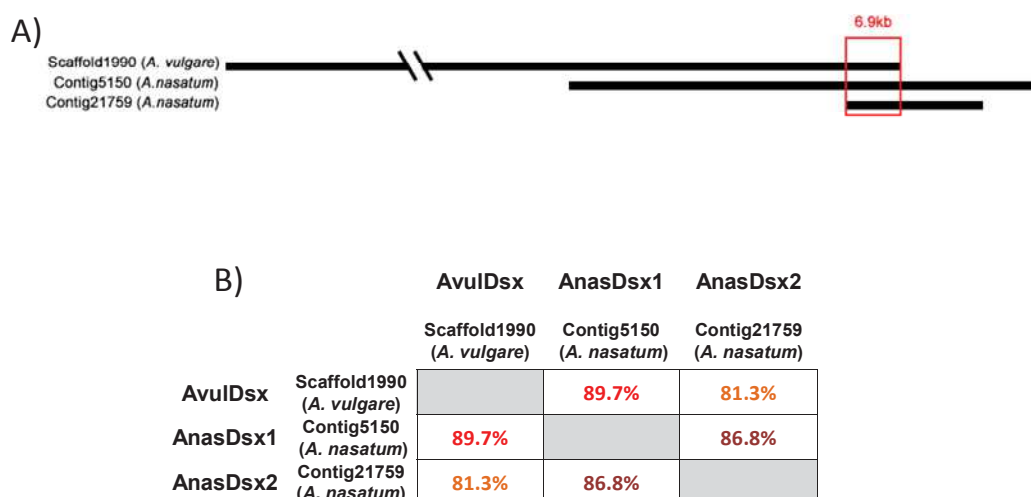


Figure 47 : Alignement des séquences porteuses des gènes *Dsx* chez *A. nasatum* et *A. vulgare*. A) Schéma de l'alignement montrant la zone commune (de 6.9 kb) au trois *contigs/scaffolds* porteurs des gènes *Dsx* chez *A. nasatum* et *A. vulgare*. Cet alignement a été réalisé avec l'algorithme d'alignement natif du logiciel *Geneious* (version 7.0.6, Kearse et al., 2012). Le Scaffold1990 (*A. vulgare*) a été tronqué par souci de lisibilité. B) Matrice d'identité nucléotidique issue de l'alignement de la région commune aux trois séquences porteuses des gènes *Dsx* identifiées chez *A. nasatum* et *A. vulgare*. Les nucléotides présentant des ambiguïtés d'alignement ont été éliminés avec *GBLOCKS* (version 0.91b, Talavera & Castresana, 2007), résultant en un alignement final de 2 764 pb.

Par ailleurs, les analyses phylogénétiques ont aussi permis d'identifier un gène appartenant à la famille des *DMRT99B* chez *A. nasatum* et *A. vulgare* : à savoir respectivement les gènes *AnasDMRT99B* et *AvulDMRT99B* (Figure 46). Les séquences génomiques porteuses de ces gènes (le Contig1049 pour le gène *AnasDMRT99B* et Contig29695 pour *AvulDMRT99B*) ont été alignées sur

environ 46 kb. Après un traitement de l'alignement avec *GBLOCKS* (résultant en un alignement final d'environ 18.5 kb sans ambiguïtés d'alignement), les deux séquences génomiques sont identiques à 91.1%. Sachant que la divergence nucléotidique estimée entre *A. nasatum* et *A. vulgare* est d'environ 6.5% (divergence mesurée à partir de 88 gènes fortement conservés, Chapitre 4, Becking et al., 2017), la longueur d'alignement de ces séquences génomiques ainsi que leur faible divergence suggère que ces deux gènes sont situés dans des régions génomiques orthologues.

Enfin, les analyses phylogénétiques ont permis d'identifier un gène appartenant à la famille des *DMRT93B* chez *A. nasatum* et *A. vulgare* : à savoir respectivement les gènes *AnasDMRT93B* et *AvulDMRT93B* (Figure 46). Les séquences génomiques porteuses de ces gènes (le Contig35358 pour le gène *AnasDMRT93B* et Contig956 pour *AvulDMRT93B*) ont été alignées sur plus de 30 kb (à savoir la longueur totale du Contig35358 d'*A. nasatum*). Après un traitement de l'alignement avec *GBLOCKS* (résultant en un alignement final d'environ 25 kb sans ambiguïtés d'alignement), les deux séquences génomiques sont identiques à 90.9%. Une fois encore, la longueur d'alignement de ces séquences génomiques ainsi que leur faible divergence suggère que ces deux gènes sont situés dans des régions génomiques orthologues.

Un arbre phylogénétique a aussi été réalisé à partir de l'alignement de gènes *Transformer2* (*Tra-2*) connus et des séquences identifiées chez *A. nasatum* et *A. vulgare* (Figure 48). La longueur totale de l'alignement était de 177 acides aminés, correspondant à un alignement partiel des deux domaines RS (Arginine et Sérine) et du domaine RRM (*RNA Recognition Motif*). Les séquences identifiées chez *A. nasatum* et *A. vulgare* sont rattachées au clade formé exclusivement de séquences de crustacés (Copepoda, Branchiopoda et Decapoda), dont le nœud est soutenu par une valeur de *bootstrap* de 75.

Cette phylogénie soutient l'hypothèse que les gènes *AnasTra-2A* (*A. nasatum*) et *AvulTra-2A* (*A. vulgare*) sont des *loci* orthologues, tout comme les gènes *AnasTra-2B* (*A. nasatum*) et *AvulTra-2B* (*A. vulgare*). Les séquences génomiques porteuses des gènes *AnasTra-2A* (à savoir le Scaffold952) et *AvulTra-2A* (à savoir le Scaffold3463) ont été alignées sur plus de 45 kb. Après

un traitement de l'alignement avec *GBLOCKS* (résultant en un alignement final d'environ 17.6 kb sans ambiguïtés d'alignement), les deux séquences génomiques apparaissent identiques à 87.5%. Même si ce pourcentage d'identité est plus faible que ceux relevés pour les gènes précédents, ces séquences génomiques restent faiblement divergentes (sachant que l'alignement final regroupe près de 17.6 kb). De plus, les domaines RS et RRM des gènes *AnasTra-2A* et *AvulTra-2B* utilisés pour les analyses phylogénétiques (Figure 48) sont identiques. Ces arguments suggèrent que ces deux gènes sont situés dans des régions génomiques orthologues.

Enfin, les séquences génomiques porteuses des gènes *AnasTra-2B* (à savoir le Scaffold3200) et des gènes *AvulTra-2B* (le Contig7839) ont été alignées sur environ 139.8 kb. Après un traitement de l'alignement avec *GBLOCKS* (résultant en un alignement final d'environ 59.33 kb sans ambiguïtés d'alignement), les deux séquences génomiques sont identiques à 89%. Une fois encore, la longueur d'alignement de ces séquences génomiques ainsi que leur faible divergence suggère que ces deux gènes sont une fois encore situés dans des régions génomiques orthologues.

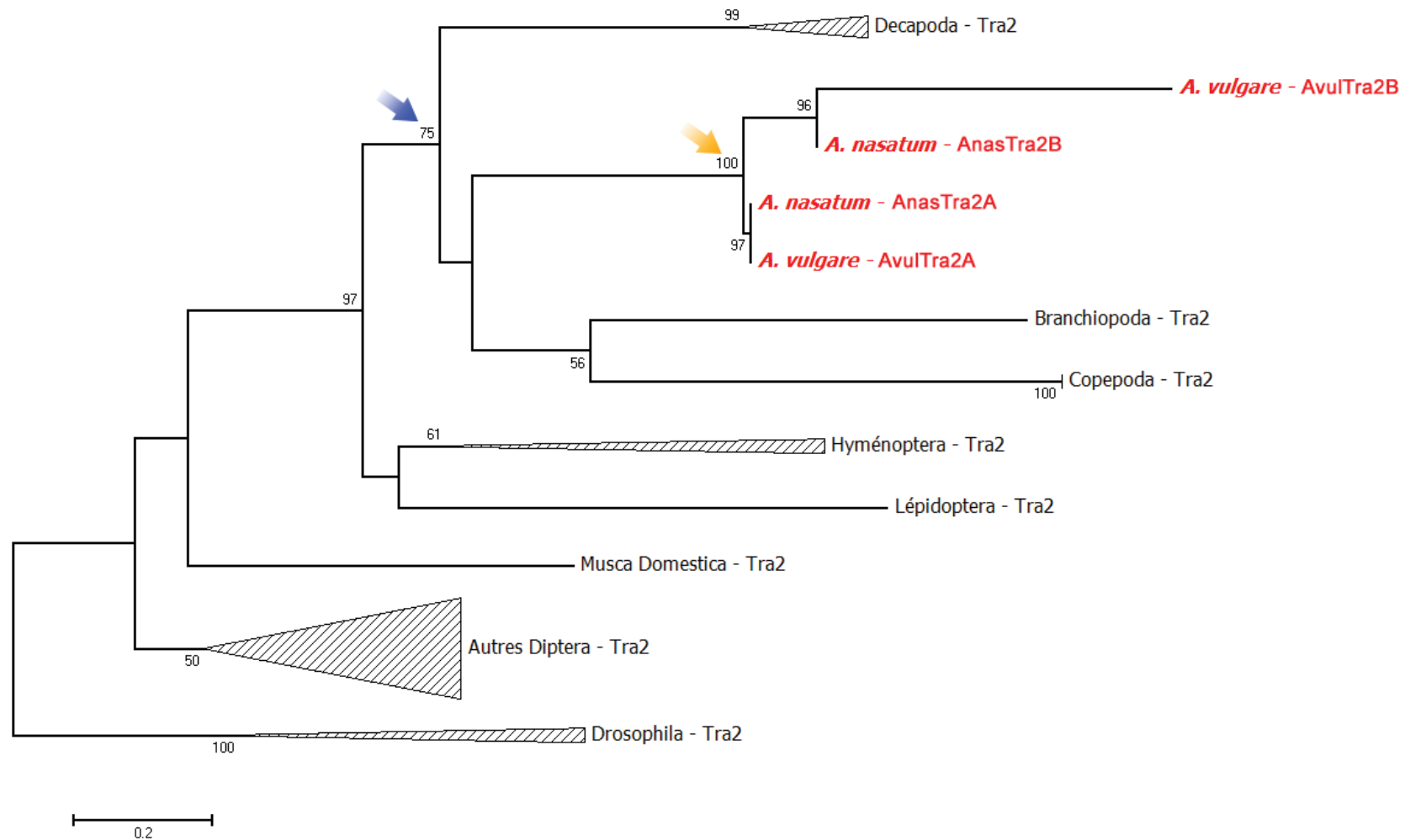


Figure 48 : Arbre phylogénétique de différents gènes *Transformer-2* connus et de ceux identifiés chez *A. nasatum* et *A. vulgare*. La longueur des branches de l'arbre est proportionnelle à la distance évolutive. Les nombres indiqués aux différents nœuds correspondent aux scores de *bootstrap* obtenus (les valeurs inférieures à 50 ne sont pas indiquées). La flèche bleue correspond au nœud réunissant l'ensemble des crustacés, tandis que la flèche orange correspond au nœud réunissant les séquences identifiées chez les isopodes terrestres *A. nasatum* et *A. vulgare*.

Les séquences à domaine DM et celles assignées aux gènes *Tra-2* identifiées chez *A. nasatum* et *A. vulgare* présentent une signature autosomale selon les analyses *YGS* et *CQ* réalisées précédemment (Tableau 24). En effet, aucune de ces séquences n'auraient été sélectionnée par un des deux filtres *in silico* établis précédemment ($CQ \leq 0.35$ et $YGS \geq 35\%$), laissant suggérer que soit ces séquences sont portées par des autosomes, soit elles sont situées dans des zones encore recombinantes des chromosomes sexuels.

Tableau 24 : Résumé des scores d'*YGS* et de *CQ* obtenus pour les différentes séquences portant les gènes à domaine DM et *Transformer2* identifiées chez *A. nasatum* et *A. vulgare*. Les scores d'*YGS* et de *CQ* mesurés chez *A. vulgare* correspondent à la moyenne des valeurs obtenues à partir de deux jeux de données, issus de deux lignées différentes (P34 et BF2875, Chebbi et al., en préparation).

Gènes	Identifiant génomique	Score de <i>CQ</i>	Score d' <i>YGS</i> (%)
<i>Armadillidium vulgare</i>			
AvulDMRT93B	Contig29695	0.98	2.1
AvulDMRT99B	Contig35358	1.03	5.2
AvulDsx	Scaffold1990	1.10	2.3
AvulTra-2A	Scaffold3463	1.01	6.4
AvulTra-2B	Scaffold3200	1.01	2.6
<i>Armadillidium nasatum</i>			
AnasDMRT93B	Contig1049	0.92	15.3
AnasDMRT99B	Contig956	1.16	7.1
AnasDsx1	Contig5150	0.87	20.4
AnasDsx2	Contig21759	0.87	31.2
AnasTra-2A	Scaffold932	0.87	5.6
AnasTra-2B	Contig7839	0.99	10.2

II. Description structurale des séquences candidates identifiées chez *A. nasatum* et *A. vulgare*

Le gène AvulDsx identifié chez *A. vulgare* code une protéine de 227 acides aminés, à partir d'un ARNm de 1 006 pb issu de deux exons (notés E sur les schémas, Figure 49). Le gène AnasDsx1 d'*A. nasatum* est composé de 3 exons codant également pour une protéine de 227 acides aminés (Figure 49). Le gène AnasDsx2 d'*A. nasatum* se situe à l'extrémité 5' de la séquence assemblée, ainsi seulement deux exons ont été détectés (le 3ème étant manquant). Les séquences nucléotidiques codantes (CDS, pour *Coding DNA Sequence*) des gènes AnasDsx1 et AnasDsx2 sont parfaitement identiques. Ces gènes diffèrent cependant par la taille de l'intron situé entre E2 et E3 (AnasDsx1) qui est plus petit que l'intron homologue situé entre E2' et E3' (AnasDsx2) (Figure 49).

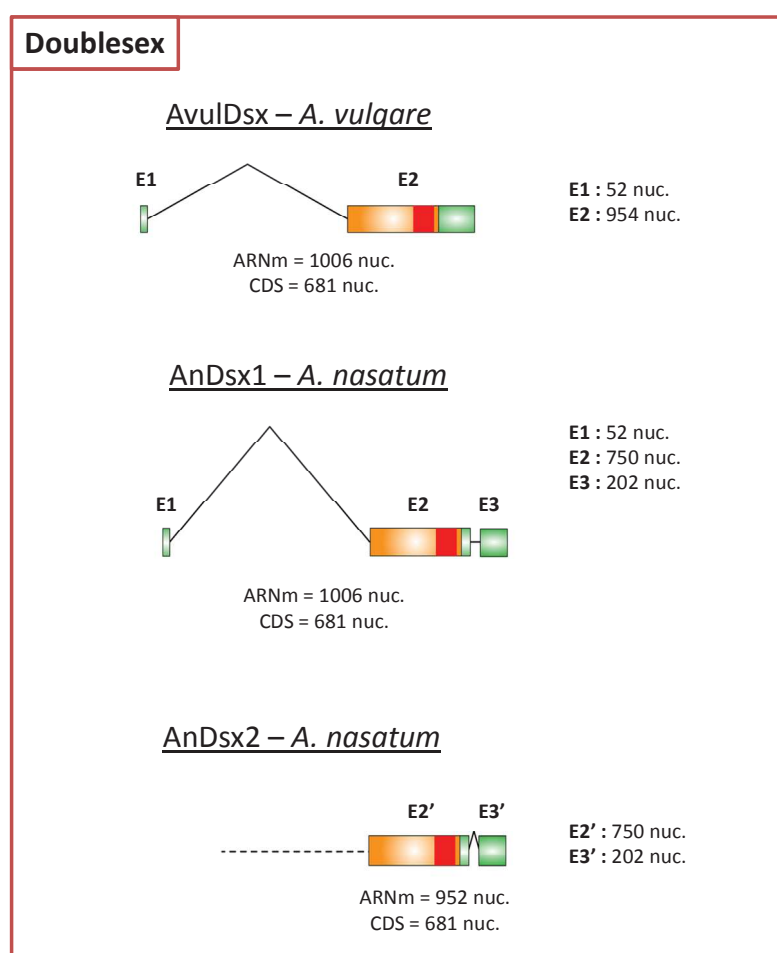


Figure 49 : Schémas de la structure des gènes *Dsx* identifiés chez *A. nasatum* et *A. vulgare*. Les séquences sont orientées dans le sens 5' vers 3'. Les rectangles correspondent aux exons (notés E), et les traits noirs représentent les introns. Les pointillés correspondent à la limite 5' du Contig21759 (porteur du gène AnasDsx2). Les parties orange représentent les CDS (*Coding DNA Sequences*), et les parties vertes correspondent aux UTR (*Untranslated Transcribed Region*). La partie rouge correspond au domaine DM. La longueur des exons/introns sur le schéma est proportionnelle à leur taille réelle.

En accord avec la phylogénie (Figure 46), les domaines DM de ces séquences montrent un plus fort niveau de similarité avec les gènes *Dsx* caractérisés chez *D. magna* (15 acides aminés identiques et 6 acides aminés ayant des propriétés physicochimiques similaires, pour un alignement de 56 acides aminés, Figure 50). Cependant, contrairement aux gènes *Dsx* d'insectes ou aux homologues de *D. magna*, aucun domaine d'oligomérisation n'a été détecté dans les séquences d'isopodes terrestres.



Figure 50 : Alignement de plusieurs domaines DM de gènes *Dsx* connus avec leurs homologues identifiés chez *A. nasatum* et *A. vulgare*. Les acides aminés identiques sont surlignés en noir, tandis que ceux ayant une propriété physicochimique similaire sont indiqués en rouge. Les positions des acides aminés CCHC et HCCC responsables des motifs en doigts de zinc sont respectivement indiquées avec des triangles rouges et orange. Les séquences utilisées pour cet alignement sont les séquences *Dsx* de : *D. magna* - *Dsx1* (BAJ78307), *D. magna* - *Dsx2* (BAJ78309), *T. castaneum* (KYB26747), *B. mori* (NP_001104815), *A. aegypti* (ABD96573) et *D. melanogaster* (NP_731197).

Les gènes *AnasDMRT99B* et *AvulDMRT99B* détectés respectivement chez *A. nasatum* et *A. vulgare* codent des protéines respectivement de 171 (à partir de deux exons) et de 196 acides aminés (à partir d'un exon) (Figure 51). Il n'a pas été possible de détecter *in silico* de séquence transcrite non-traduite (UTR pour *Untranslated Transcribed Region*) car aucun transcrit correspondant n'a pu être identifié dans les transcriptomes utilisés lors de l'annotation du génome (transcriptomes générés par Romiguier et al., 2014, Chapitre 1). En accord avec la phylogénie (Figure 46), les domaines DM de ces deux gènes montrent de très fortes similarités aux domaines DM des gènes *DMRT99B* d'invertébrés et des gènes *DMRT4* de vertébrés (57 acides aminés identiques et 1 acide aminé ayant une fonction similaire, sur 64 acides aminés analysés, Figure 52).

Les gènes *AnasDMRT93B* et *AvulDMRT93B* détectés respectivement chez *A. nasatum* et *A. vulgare* codent quant à eux des protéines respectivement de 116 (à partir de quatre exons) et de 87 acides aminés (à partir de deux exons) (Figure 51). Une fois encore, aucune séquence UTR n'a pu être mise en évidence. Toujours en accord avec la phylogénie (Figure 46), les domaines DM de ces deux gènes montrent de très fortes similarités aux domaines DM des gènes *DMRT3/93B* d'invertébrés et des gènes *DMRT3* de vertébrés (32 acides aminés identiques et 6 acides aminés ayant des propriétés physicochimiques similaires, sur 53 sites analysés, Figure 52).

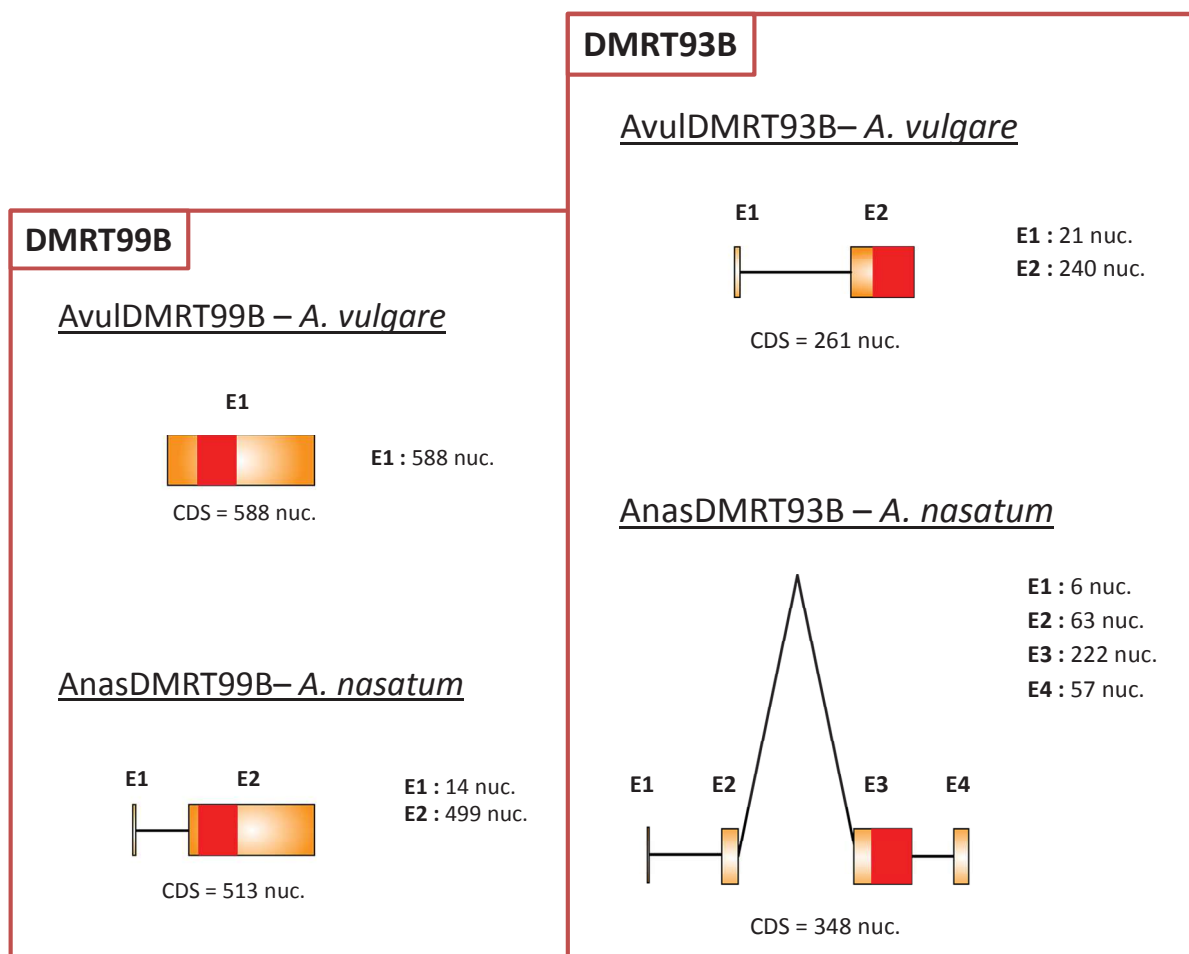


Figure 51 : Schémas de la structure des gènes *DMRT99B* et *DMRT93B* identifiés chez *A. nasatum* et *A. vulgare*. Les séquences sont orientées dans le sens 5' vers 3'. Les rectangles correspondent aux exons (notés E), et les traits noirs représentent les introns. Les parties orange représentent les CDS (*Coding DNA Sequences*), et les parties rouges indiquent les domaines DM. La longueur des exons/introns sur le schéma est proportionnelle à leur taille réelle.



Figure 52 : Alignements de plusieurs domaines DM de gènes *DMRT99B/4* et *DMRT93B/3* connus avec leurs homologues identifiés chez *A. nasatum* et *A. vulgare*. (A) Alignement des séquences *DMRT99B/4*. Les séquences utilisées pour cet alignement sont les séquences *DMRT99B/4* de : *D. pulex* (EFX84867), *D. magna* (BAG12873), *T. castaneum* (XP_975675), *D. melanogaster* (NP_524549), *T. rubripes* (NP_001033037), *X. laevis* (AAV66322), *O. latipes* (BAB63259) et *H. sapiens* (NP_071443). (B) Alignement des séquences *DMRT93B/3*. Les séquences utilisées pour cet alignement sont les séquences *DMRT93B/3* de : *D. pulex* (EFX89054), *D. magna* (BAG12872), *D. melanogaster* (NP_524428), *T. rubripes* (NP_001033034), *D. rerio* (AAU89440) et *H. sapiens* (NP_067063). Les acides aminés identiques sont surlignés en noir, tandis que ceux ayant une propriété physicochimique similaire sont indiqués en rouge. Les positions des acides aminés CCHC et HCCC responsables des motifs en doigts de zinc sont respectivement indiquées avec des triangles rouges et orange.

Les 4 copies homologues de *Tra-2* identifiées chez *A. nasatum* et *A. vulgare* disposent d'un domaine de reconnaissance de l'ARN (domaine RRM), de deux domaines riches en Arginine et en Sérine (domaines RS, impliqués dans les interactions protéines/protéines) et d'un domaine *Linker* (caractéristique des gènes *Tra-2*, Dauwalder et al., 1996) évitant à deux autres domaines proches d'interférer entre eux par encombrement allostérique. Pour *A. nasatum* (Figure 53), le gène *Tra-2* porté par le Contig7839 est composé de neuf exons et permet la production d'une protéine de 305 acides aminés. Les deux domaines RS sont composés respectivement de 42 Arginines et 35 Sérines (64.7% du domaine) et de 10 Arginines et 7 Sérines (68% du domaine), et le domaine RRM est composé de 72 acides aminés. Le gène *Tra-2* porté par le Scaffold932 est composé de huit exons et code une protéine de 312 acides aminés. Ses domaines RS sont composés respectivement de 50 Arginines et 42 Sérines (61.8% du domaine) et de 12 Arginines et 7 Sérines (67.9% du domaine), et son domaine RRM est aussi composé de 72 acides aminés.

Concernant les séquences *Tra-2* identifiées chez *A. vulgare* (Figure 54), le gène *Tra-2* porté par le Scaffold3200 est composé de huit exons et permet la production d'une protéine de 220 acides aminés. Le premier domaine RS n'est composé que de 9 Arginines et 8 Sérines (54.8% du domaine), tandis que son deuxième domaine RS comporte 12 arginines et 7 sérines (67.9% du domaine). Son domaine RRM apparaît lui aussi tronqué puisqu'il n'est composé que de 51 acides aminés contre 72 pour les autres domaines RRM (Figure 55). Enfin, le gène *Tra-2* porté par le Scaffold3463 est composé de huit exons et code une protéine de 385 acides aminés. Ses domaines RS sont composés respectivement de 50 arginines et 42 sérines (61.8% du domaine) et de 12 Arginines et 7 Sérines (67.9% du domaine), et son domaine RRM compte 72 acides aminés. Les régions RRM et *Linker* présentent de fortes similarités aux domaines homologues identifiés chez les insectes et les crustacés (32 acides aminés identiques et 11 acides aminés ayant des propriétés physicochimiques similaires, le tout sur 103 sites alignés, Figure 55).

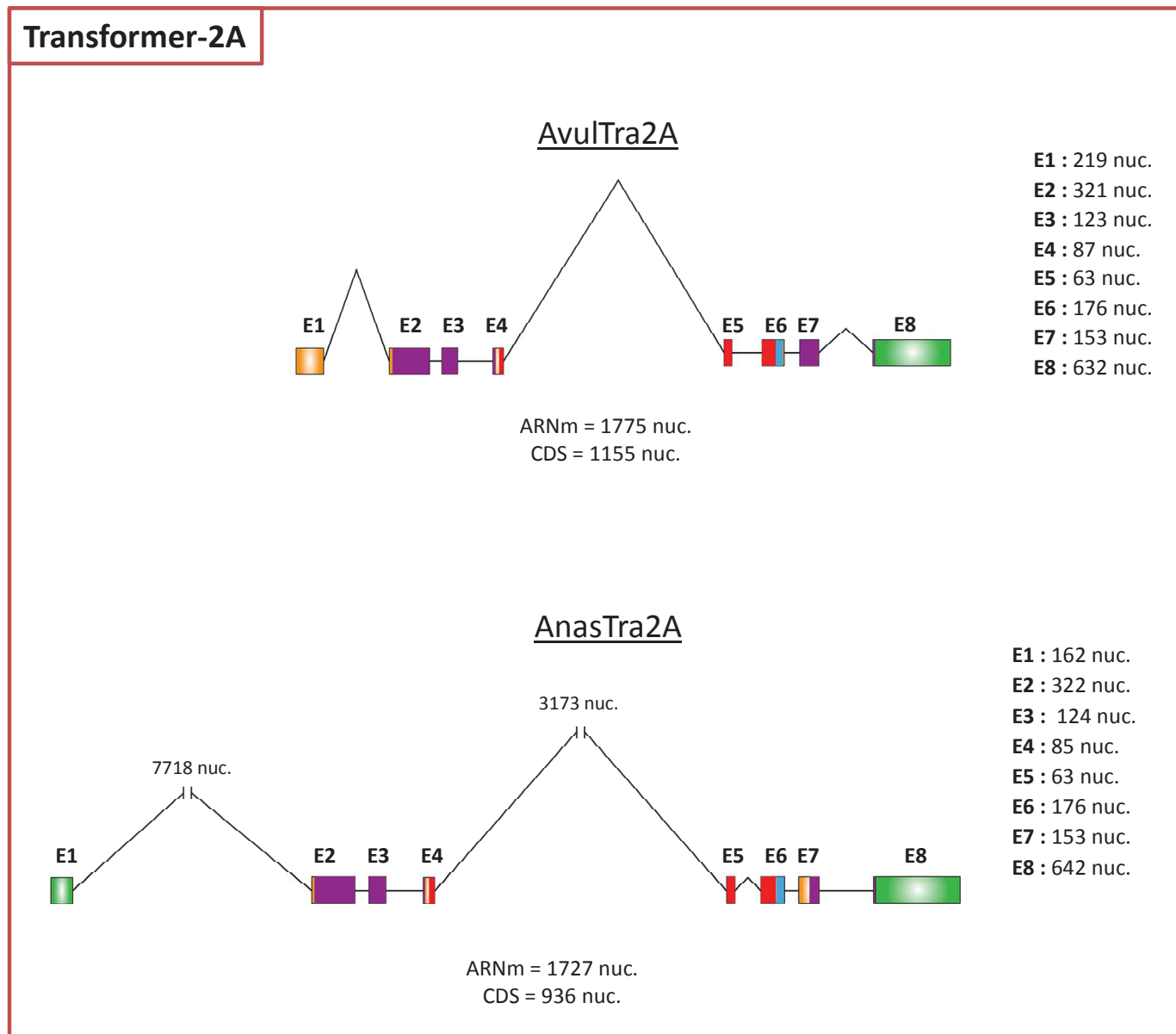


Figure 53 : Schémas de la structure des gènes *Transformer-2-A* identifiés chez *A. nasatum* et *A. vulgare*. Les séquences sont orientées dans le sens 5' vers 3'. Les rectangles correspondent aux exons (notés E), et les traits noirs représentent les introns. Les parties orange représentent les CDS (*Coding DNA Sequences*), les parties vertes correspondent aux UTR, les parties rouges indiquent les domaines RRM, les parties violettes sont les domaines RS et les régions bleues représentent les régions *Linker*. La longueur des exons/introns sur le schéma est proportionnelle à leur taille réelle. Les introns supérieurs à 3kb sont représentés tronqués pour plus de lisibilité.

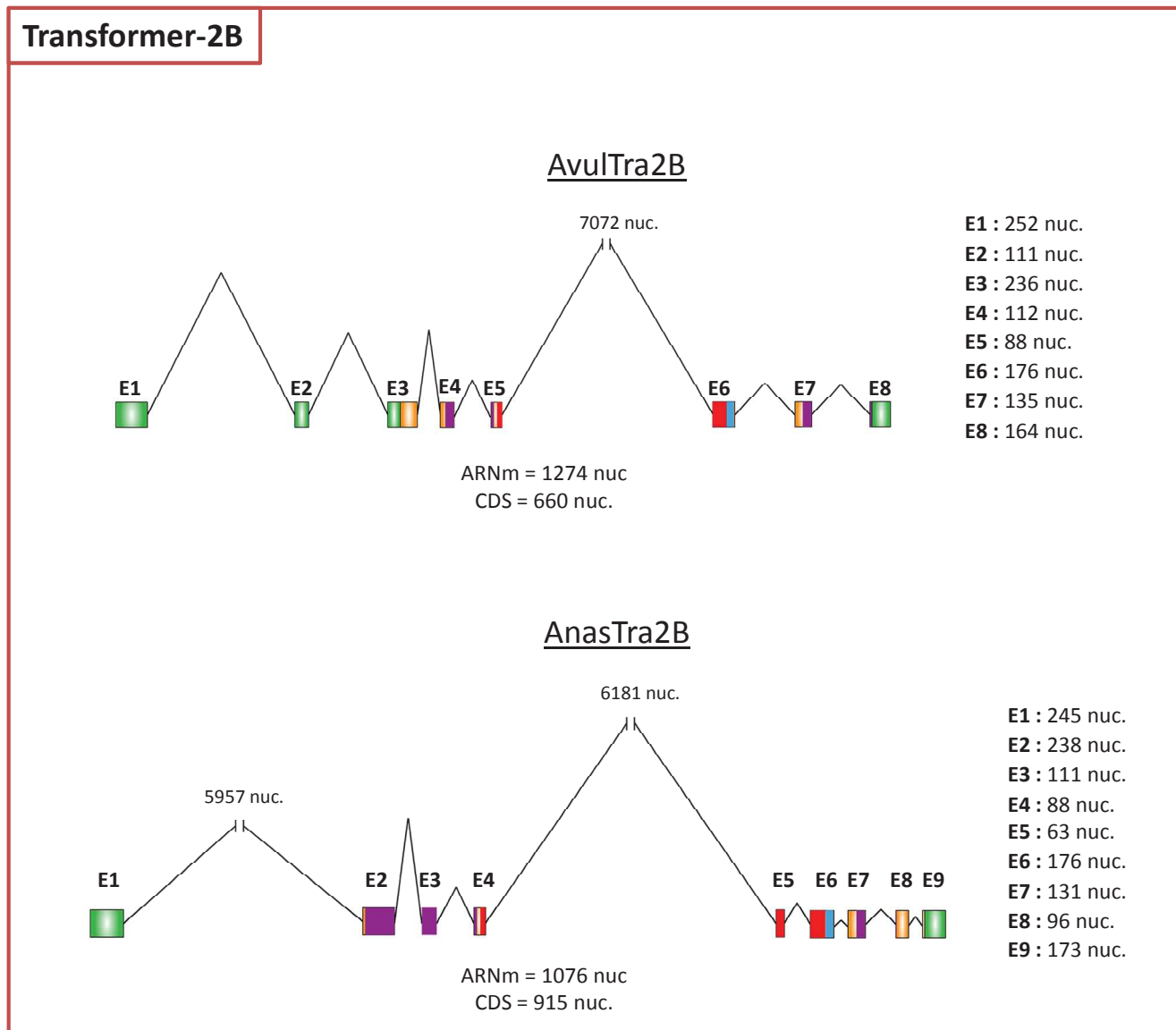


Figure 54 : Schémas de la structure des gènes *Transformer-2* identifiés chez *A. vulgare*. Les séquences sont orientées dans le sens 5' vers 3'. Les rectangles correspondent aux exons (notés E), et les traits noirs représentent les introns. Les parties orange représentent les CDS (*Coding DNA Sequences*), les parties vertes correspondent aux UTR, les parties rouges indiquent les domaines RRM, les parties violettes sont les domaines RS et les régions bleues représentent les régions *Linker*. La longueur des exons/introns sur le schéma est proportionnelle à leur taille réelle. Les introns supérieurs à 3kb sont représentés tronqués pour plus de lisibilité.



Figure 55 : Alignement de plusieurs domaines RRM et de régions *Linker* de gènes *Tra2* connus avec leurs homologues identifiés chez *A. nasatum* et *A. vulgare*. Les acides aminés identiques sont surlignés en noir, tandis que ceux ayant une fonction similaire sont indiqués en rouge. Les domaines RRM et *Linker* sont indiqués avec des rectangles orange et bleus. Les séquences utilisées pour cet alignement sont les séquences *Tra2* de : *D. pulex* (EFX90042), *P. monodon* (ACD13597), *E. sinensis* (APJ36536), *M. nipponense* (AGI50962), *M. domestica* (AY847518), *L. cuprina* (FJ461620), *B. mori* (NM_001126233), *A. mellifera* (NM_001265585) et *D. melanogaster* (M23633).

III. Comparaison de l'expression des gènes candidats entre mâles et femelles au cours du développement

1. Comparaison de l'expression des gènes candidats entre mâles et femelles chez *A. nasatum*

Suite aux analyses précédentes, six gènes candidats ont été identifiés comme potentiellement impliqués dans la cascade de déterminisme du sexe d'*A. nasatum* : *AnasDsx1*, *AnasDsx2*, *AnasDMRT99B*, *AnasDMRT93*, *AnasTra-2A* et *AnasTra-2B*. Comme indiqué précédemment, les *CDS* des deux copies des gènes *Dsx* étant identiques, il n'était pas possible de distinguer l'expression des deux gènes indépendamment. Les résultats de l'expression des gènes *Dsx* sont indiqués dans la Figure 56. L'expression de ces gènes est la plus forte au stade 4 (qui correspond au stade du début de la différenciation sexuelle) chez les mâles, puisque l'expression relevée à ce stade est en moyenne près de 18 fois supérieure à celle mesurée au stade 2. En revanche, l'expression des gènes *AnasDMRT99B*, *AnasDMRT93*, ainsi que *AnasTra-2A* et *AnasTra-2B* ne présentent pas de variation significativement différente entre mâles et femelles pour un stade donné (Figure 57, Figure 58, Figure 59 et Figure 60, respectivement).

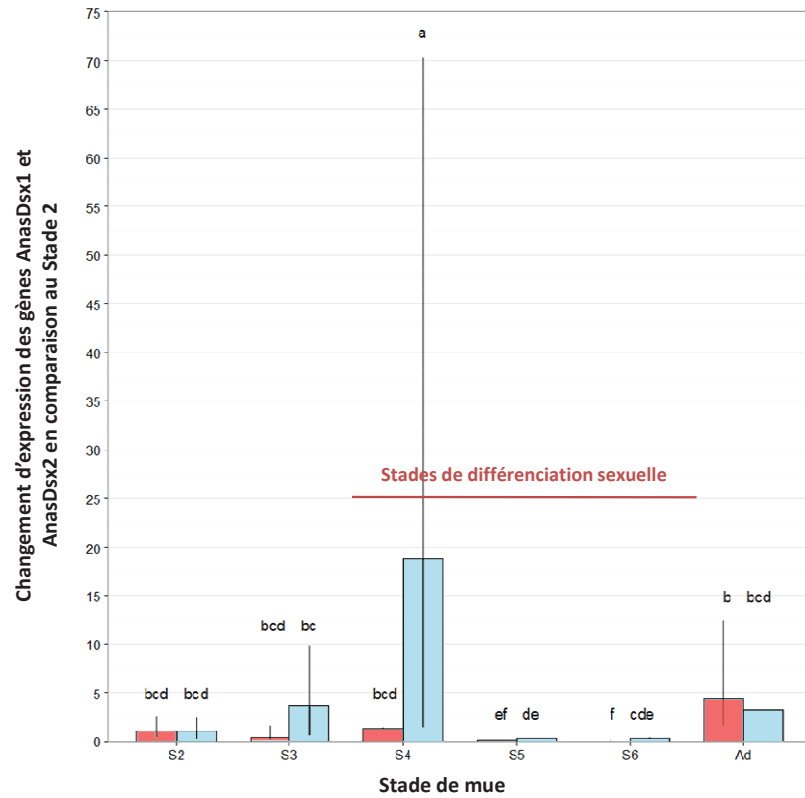


Figure 56 : Changement d'expression des gènes AnasDsx1 et AnasDsx2 par rapport au stade 2 entre mâles et femelles *A. nasatum*. Chaque lettre commune entre deux histogrammes indique une absence de différence significative (test statistique : ANOVA à deux facteurs : $F_{(\text{sexe})} = 38.02$, $p\text{-value}_{(\text{sexe})} = 8.66 \cdot 10^{-8}$; $F_{(\text{stade de mue})} = 28.98$, $p\text{-value}_{(\text{stade de mue})} = 2.82 \cdot 10^{-14}$; $F_{(\text{sexe:stade de mue})} = 5.13$, $p\text{-value}_{(\text{sexe:stade de mue})} = 6.25 \cdot 10^{-4}$). Les barres verticales correspondent aux écart-types. L'expression mesurée chez les mâles est affichée en bleu, tandis que celle relevée chez les femelles apparaît rouge. Ad correspond au stade adulte.

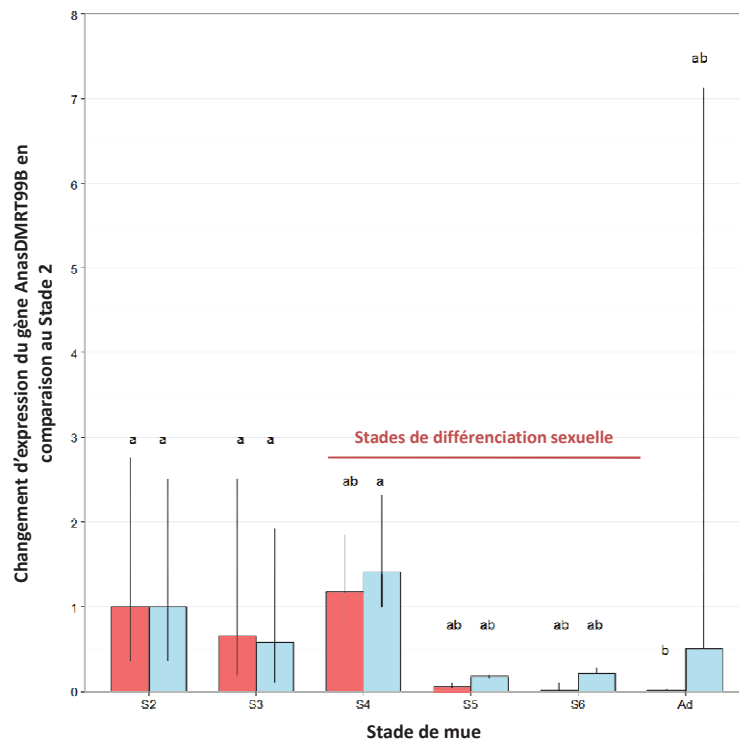


Figure 57 : Changement d'expression du gène AnasDMRT99B par rapport au stade 2 entre mâles et femelles *A. nasatum*. Chaque lettre commune entre deux histogrammes indique une absence de différence significative (test statistique : Kruskal-Wallis : $\chi = 41.58$, $p\text{-value} = 1.91 \cdot 10^{-05}$). Les barres verticales correspondent aux écart-types. L'expression mesurée chez les mâles est affichée en bleu, tandis que celle relevée chez les femelles apparaît rouge. Ad correspond au stade adulte.

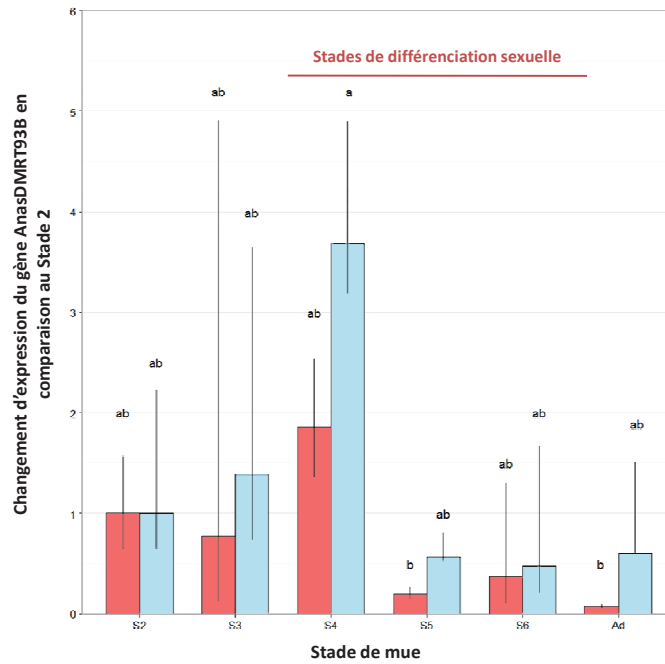


Figure 58 : Changement d'expression du gène AnasDMRT93B par rapport au stade 2 entre mâles et femelles *A. nasatum*. Chaque lettre commune entre deux histogrammes indique une absence de différence significative (test statistique : Kruskal-Wallis : $\chi = 31.25$, $p\text{-value} = 0.0010$). Les barres verticales correspondent aux écart-types. L'expression mesurée chez les mâles est affichée en bleu, tandis que celle relevée chez les femelles apparaît rouge. Ad correspond au stade adulte.

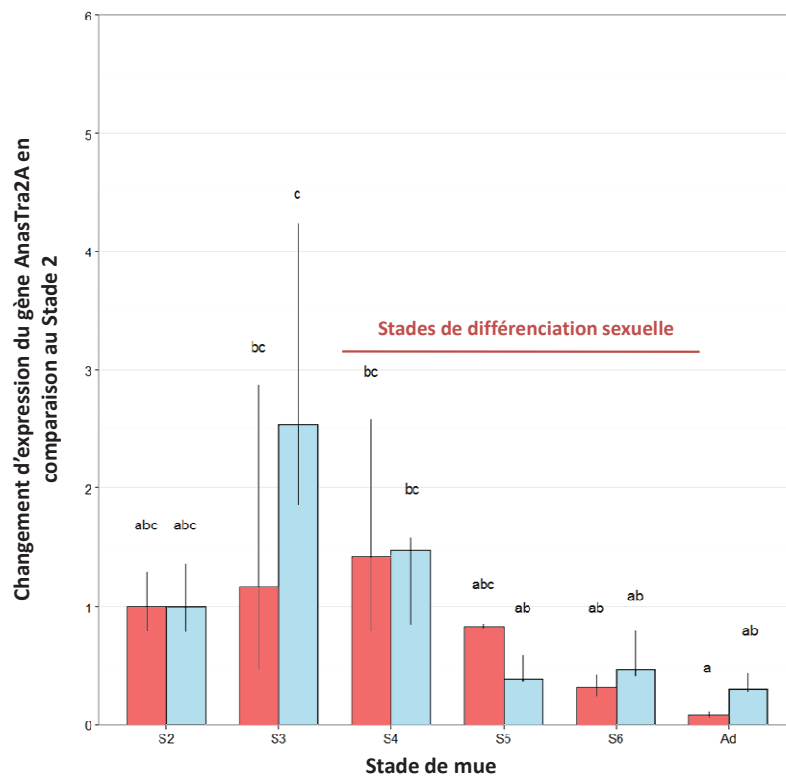


Figure 59 : Changement d'expression du gène AnasTra2A par rapport au stade 2 entre mâles et femelles *A. nasatum*. Chaque lettre commune entre deux histogrammes indique une absence de différence significative (test statistique : Kruskal-Wallis : $\chi = 52.14$, $p\text{-value} = 2.58.10^{-7}$). Les barres verticales correspondent aux écart-types. L'expression mesurée chez les mâles est affichée en bleu, tandis que celle relevée chez les femelles apparaît rouge. Ad correspond au stade adulte.

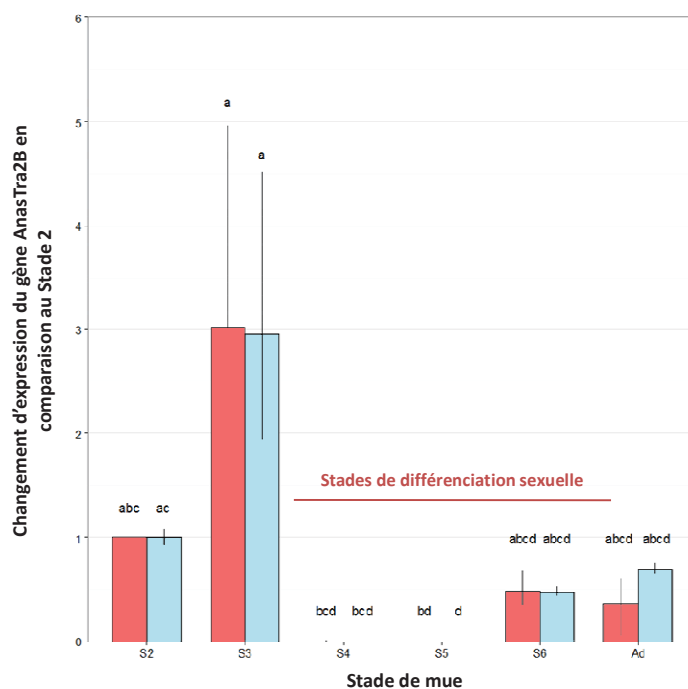


Figure 60 : Changement d'expression du gène AnasTra2B par rapport au stade 2 entre mâles et femelles *A. nasatum*. Chaque lettre commune entre deux histogrammes indique une absence de différence significative (test statistique : Kruskal-Wallis : $\chi = 62.48$, $p\text{-value} = 2.08.10^{-9}$). Les barres verticales correspondent aux écart-types. L'expression mesurée chez les mâles est affichée en bleu, tandis que celle relevée chez les femelles apparaît rouge. Ad correspond au stade adulte.

2. Comparaison de l'expression des gènes candidats entre *A. nasatum* et *A. vulgare*

Comme précisé dans la section Matériels et Méthodes de ce chapitre (voir page 182), il n'a pas été possible de sexer par des tests PCR les individus indifférenciés d'*A. vulgare* (stade 2 à 5 inclus). Ainsi, l'expression des gènes candidats n'a été mesurée qu'aux stades 6 et Adulte (le stade 5 pouvant présenter des incertitudes au sexage). L'expression des gènes a donc été normalisée par rapport au stade adulte, y compris pour *A. nasatum* afin de pouvoir comparer l'expression des gènes mesurée dans les deux espèces. Les tests statistiques ont été effectués indépendamment pour chaque gène. Ainsi, aucune différence d'expression significative n'a été relevée entre mâles et femelles des deux espèces aux stades 6 et Adulte (Figure 61). A la lumière des résultats obtenus pour l'ensemble des stades de mue d'*A. nasatum*, ceci était attendu puisqu'aucune différence n'était observée entre les sexes pour ces deux stades, d'autant plus que le stade 6 constitue le dernier stade de mue au cours duquel la différenciation sexuelle s'effectue.

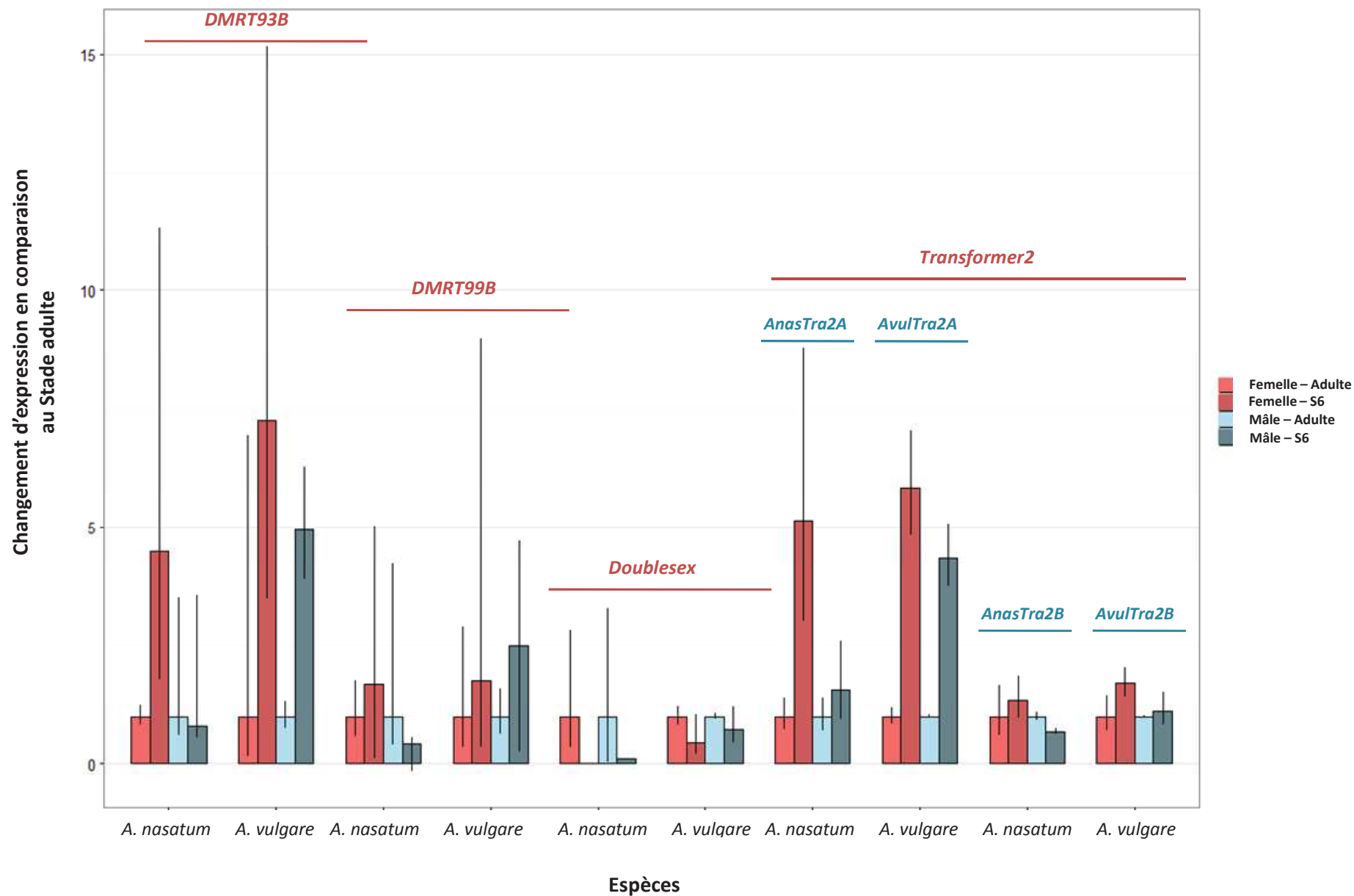


Figure 61 : Changements d'expression des gènes candidats identifiés chez *A. nasatum* et *A. vulgare* pour les stades 6 et Adulte, normalisés par rapport à leur expression respective au stade Adulte. Les barres verticales correspondent aux écart-types.

Discussion

I. Caractérisation des gènes à domaine DM et suivi de l'expression au cours du développement

A partir de l'annotation des génomes d'*A. nasatum* et d'*A. vulgare*, complétée par des recherches *in silico* par *BLASTx*, respectivement quatre et trois gènes à domaine DM ont été identifiés. Le domaine DM de ces gènes est caractérisé par une structure classique en double doigts de zinc formée par les motifs CCHC et HCCC (Erdman & Burtis, 1993; Zhu et al., 2000). La phylogénie de ces gènes (Figure 46) révèle que plusieurs familles sont paraphylétiques (les familles *DMRT99B*, *DMRT4*, *DMRT5* et *DMRT3/93B*), contrairement à ce que proposait Wexler et al. (2014) dans une analyse phylogénétique réalisée à l'échelle de l'ensemble des métazoaires. Néanmoins, les nœuds profonds de cette phylogénie n'étaient pas résolus, tout comme dans nos analyses. Ce manque de résolution s'explique principalement par la faible longueur de l'alignement utilisé dans les deux cas, à savoir 48 acides aminés dans cette étude et 50 dans l'analyse de Wexler et al. (2014). Cette faible longueur d'alignement est due au fait que ces gènes sont composés de séquences hautement divergentes (selon les taxa et les familles de gènes), à l'exception de leur domaine DM qui constitue la seule partie homologue et donc comparable (Volff et al., 2003).

La topologie obtenue suite à nos analyses permet malgré tout de classer sans ambiguïté les gènes identifiés. Ainsi, les génomes d'*A. vulgare* et d'*A. nasatum* disposent tous les deux d'un gène DM appartenant à la famille des gènes *DMRT3/93B* (AvulDMRT93B et AnasDMRT93B). Découverts tout d'abord chez les vertébrés sous la dénomination *DMRT3*, ces gènes sont connus pour être impliqués dans de nombreux processus développementaux tels que la formation des gonades mâles chez la souris (Kim et al., 2003) ou encore du système nerveux chez le médaka *Oryzias latipes* (Winkler et al., 2004) et le poisson zèbre *Danio rerio* (Li et al., 2008). Plus récemment, un homologue de ces gènes (nommé *DMRT93B*) a été identifié chez le branchiopode *D. magna* (Kato et al., 2008). Ce

gène ne s'exprime qu'au sein des testicules sans pour autant montrer une expression différentielle au cours du développement (Kato et al., 2011), suggérant ainsi un rôle dans le maintien de la fonction de reproduction des mâles. Dans le cas d'*A. nasatum* (stades 2 à adulte) et d'*A. vulgare* (stades 6 et adulte), aucune expression sexe-spécifique n'a été relevée et cette expression ne présente pas de forte variation selon les différents stades de développement (Figure 58, Figure 61). Ces premiers résultats ne semblent pas être en accord avec un éventuel rôle de ce gène dans le déterminisme du sexe. Cependant, nos mesures ont été effectuées à partir d'individus entiers et non à partir de tissus séparés. Il pourra donc être utile à l'avenir de répéter ces mêmes analyses sur des ovaires et des testicules isolés afin de mieux caractériser l'expression de ce gène et notamment d'évaluer sa potentielle implication dans le maintien des fonctions sexuelles, de la même façon que dans le cas de *D. magna* (Kato et al., 2008, 2011).

D'autre part, une copie appartenant à la famille *DMRT99B* a aussi été identifiée chez *A. nasatum* (AnasDMRT99B, Figure 46) et chez *A. vulgare* (AvulDMRT99B, Figure 46). Découverts tout d'abord chez *D. melanogaster* et chez *C. elegans*, les gènes de cette famille sont impliqués dans la formation de la vitellogénine (Coschigano & Wensink, 1993; Yi & Zarkower, 1999). Un orthologue du gène *DMRT99B* a aussi été identifié chez *D. magna* et est fortement exprimé dans les ovaires (mais pas dans les testicules; Kato et al., 2008), sans pour autant être exprimé différemment entre mâles et femelles au cours du développement, ce qui suggère que ce gène n'est pas impliqué dans le déterminisme du sexe mais peut-être plutôt dans le maintien des fonctions de reproduction des femelles (Kato et al., 2011). Cependant, de la même façon que pour le gène *DMRT93B*, l'expression du gène *DMRT99B* n'est pas spécifique à un des deux sexes et ne présente pas de forte variation au cours du développement chez *A. nasatum*. Une fois encore, il paraîtrait pertinent de mesurer l'expression de ces gènes à l'échelle tissulaire afin de mieux caractériser l'expression de *DMRT99B*, puisque ce gène ainsi que *DMRT93B* sont tous les deux connus pour agir sur de nombreux processus développementaux en dehors de la différenciation sexuelle (voir la revue de Picard et al., 2015).

Enfin, un gène orthologue du gène *Dsx* a été identifié dans le génome d'*A. vulgare* (AvulDsx, Figure 46), tandis que deux copies ont été détectées chez *A. nasatum* (AnasDsx1 et AnasDsx2, Figure 46). Ces gènes sont connus pour être des acteurs majeurs impliqués dans la cascade de déterminisme du sexe des insectes (Hildreth, 1965; Williams & Carroll, 2009), à partir d'isoformes mâle-spécifiques et femelle-spécifiques suite à l'épissage alternatif de l'ARNm primaire de ces gènes (Marín & Baker, 1998). Deux copies de ce gène ont aussi été identifiées chez *D. magna*, mais elles ne présentaient pas d'isoforme sexe-spécifique (Kato et al., 2011). En revanche, ces gènes sont uniquement exprimés chez les embryons mâles dans les premiers jours de développement, et suite à des tests fonctionnels (extinction post-transcriptionnelle par interférence ARN chez les embryons mâles, expression ectopique chez les embryons femelles), il a été démontré que seule une des deux copies jouait un rôle de gène régulateur clé du phénotype mâle (Kato et al., 2011). Ce profil d'expression est assez proche de celui observé chez *A. nasatum*, qui présente une forte surexpression de ce gène chez les embryons mâles au stade 4 (Figure 56). Il s'agit d'ailleurs du stade de mue correspondant à la première étape de différenciation des gonades chez *A. vulgare* (Katakura, 1984; Badawi et al., 2015). Cependant dans notre cas, il n'est pas possible de distinguer l'expression des deux copies du gène *Dsx* chez *A. nasatum* puisqu'elles ont exactement la même séquence codante. Ceci sous-entend d'ailleurs que cette duplication est probablement un événement très récent de l'histoire évolutive de cette espèce, d'autant plus qu'elle n'est pas retrouvée chez *A. vulgare* (suggérant ainsi que la duplication de ce gène est postérieure à la divergence entre *A. nasatum* et *A. vulgare*, soit environ 25 millions d'années (Becking et al., 2017; Chapitre 4, Figure 63). Enfin, les séquences identifiées chez *A. vulgare* et *A. nasatum* ne disposent pas de domaine d'oligomérisation (qui est généralement présent en position C-terminale, Geuverink & Beukeboom, 2014), ce qui sous-entend que la protéine finale n'agirait pas par l'intermédiaire d'un processus de dimérisation (ayant pour fonction de faciliter la reconnaissance de l'ADN, Bayrer et al., 2005).

Chez les insectes, le gène *Dsx* est connu pour présenter deux variants d'épissage alternatifs sexe-dépendants (Burtis & Baker, 1989). Chez l'ensemble des insectes étudiés jusqu'à présent, à l'exception de *B. mori* (Fujii & Shimada,

2007), la régulation de l'épissage sexe-spécifique de *Dsx* est assurée par le gène *Tra*. Les variants sexe-spécifiques ainsi générés agissant directement sur l'activité transcriptionnelle de nombreux gènes ayant une activité sexe-spécifique (Verhulst et al., 2010). Cependant, la présence de tels variants sexe-spécifiques n'a pas été relevée chez *D. magna*, alors que l'activité du gène *Dsx1* a été caractérisée pour son rôle comme déterminant/différenciant sexuel (Kato et al., 2011) et que le gène *Tra* a pourtant aussi été caractérisé chez cette espèce (Kato et al., 2010). Dans le cas d'*A. nasatum*, les gènes *Dsx* (*AnasDsx1* et *AnasDsx2*) sont surexprimés chez les mâles de stades 4, ce qui est assez proche de la situation observée chez *D. magna* (Kato et al., 2011). Ceci suggère que le mode d'action de ces gènes chez *A. nasatum* pourrait être homologue à la situation observée chez *D. magna*. Des analyses complémentaires sont néanmoins nécessaires pour caractériser le mode d'action de ces gènes chez *A. nasatum*.

L'ensemble des séquences contenant des gènes à domaine DM identifiées chez *A. nasatum* et *A. vulgare* présente une signature autosomale suite aux analyses *CQ* et *YGS* (Tableau 24). Ceci est d'ailleurs cohérent avec le fait que les gènes *DMRT93B*, *DMRT99B* et *Dsx* s'expriment à la fois chez les mâles et les femelles, impliquant la présence de ces gènes dans le génome des deux sexes. Cette situation est assez classique puisque même le gène *Dsx*, souvent impliqué dans le déterminisme du sexe chez les insectes, est présent en position autosomale, notamment chez les diptères (Baker et al., 1991) ou encore chez les lépidoptères (Kunte et al., 2014). Ces arguments soutiennent donc que même si le gène *Dsx* apparaît être un excellent gène candidat pour être impliqué dans le déterminisme ou la différenciation du sexe chez *A. nasatum*, il ne s'agit probablement pas pour autant du gène maître du déterminisme du sexe, mais d'un élément impliqué dans la cascade de déterminisme ou de différenciation sexuelle, à l'instar des insectes (Sánchez, 2008).

II. Caractérisation des gènes *Transformer-2* et suivi de l'expression au cours du développement

Chez les insectes, l'action des gènes à domaine DM (et principalement du gène *Doublesex*) est souvent associée à l'action des gènes *Transformer* (*Tra*) et *Transformer-2* (*Tra-2*) (voir la revue de Verhulst et al., 2010). Les gènes *Tra* et *Tra-2* jouent le rôle de régulateurs de l'épissage alternatif du transcrit *Doublesex*, et jouent un rôle central dans l'évolution des mécanismes de déterminisme du sexe chez les insectes (Verhulst et al., 2010). Aucun gène *Tra* n'a pu être identifié dans les génomes d'*A. nasatum* et d'*A. vulgare*. Ce gène n'a d'ailleurs jamais été caractérisé chez les crustacés en dehors des branchiopodes (Kato et al., 2010; P. Chen et al., 2014; Kong et al., 2015). Cependant, deux copies de gènes *Tra-2* ont été détectées dans chacun des deux génomes d'isopodes terrestres. Compte tenu de la topologie de l'arbre phylogénétique réalisé à partir de séquences de gène *Tra-2* (Figure 48), il semblerait que les gènes *AnasTra-2A* et *AvulTra-2A* soient orthologues, tout comme les gènes *AnasTra-2B* et *AvulTra-2B* (et ce d'autant plus que les nœuds incluant ces gènes sont très fortement soutenus, Figure 48). Cette topologie suggère que la duplication du gène *Tra-2* est antérieure à l'évènement de divergence entre *A. nasatum* et *A. vulgare*. Les *contigs/scaffolds* portant des gènes *Tra-2* ont tous une signature autosomale suite aux analyses *YGS* et *CQ* (Tableau 24), ce qui est congruent avec la position autosomale de *Tra-2* chez *D. melanogaster* notamment (Watanabe, 1975). C'est d'ailleurs chez cet organisme modèle que ce gène a été découvert pour la première fois. Chez cette espèce, il a été montré que le gène *Tra-2* a une expression constitutive aux sein des mâles et des femelles (Amrein et al., 1988), mais en agissant en interaction avec le variant femelle de *Tra*, il permet l'épissage alternatif du transcrit du gène *Doublesex* afin de produire le variant *Doublesex* spécifique aux femelles (Hedley & Maniatis, 1991; Hoshijima et al., 1991). Depuis quelques années, le gène *Tra-2* a été identifié chez différentes espèces de crustacés décapodes, notamment chez *Penaeus monodon* (Leelatanawit et al., 2009), *Penaeus japonicus* (Sellars et al., 2015) ou encore *Fenneropenaeus chinensis* (Li et al., 2012). Seul le dernier exemple a par ailleurs montré une expression biaisée de ce gène au sein des embryons de femelles *F. chinensis* (Li et al., 2012). Dans le cas d'*A. nasatum* et

d'*A. vulgare*, les gènes *Tra-2* identifiés ne semblent pas s'exprimer selon un pattern spécifique à un sexe (Figure 59, Figure 60, Figure 61). Néanmoins, il n'est pas exclu que ces gènes puissent jouer malgré tout un rôle dans le déterminisme du sexe, comme c'est le cas chez *D. melanogaster* (Amrein et al., 1988).

Au final, peu d'informations sont disponibles concernant le gène *Tra-2* et sur son implication dans la cascade de déterminisme du sexe chez les crustacés. Si un rôle dans le déterminisme du sexe femelle de *F. chinensis* peut être suspecté sur la base de l'expression de ce gène au cours du développement (Li et al., 2012), les informations disponibles sont encore parcellaires, et ce d'autant plus qu'aucun test fonctionnel n'a encore été effectué sur le gène *Tra-2* chez les crustacés.

III. Perspectives de recherche

Ces premières mesures d'expression de gènes réalisées ont permis de mettre en évidence que le gène *Doublesex* constitue un excellent gène candidat pour être impliqué dans la cascade de déterminisme du sexe d'*A. nasatum*. Il s'agit du seul gène testé présentant des variations d'expression sexe-spécifiques, étant fortement surexprimé chez les mâles de stade 4. Afin de confirmer son implication dans la cascade de déterminisme du sexe, il apparaît nécessaire d'effectuer une série de tests fonctionnels (par l'intermédiaire d'interférence ARN par exemple) et d'en évaluer les conséquences phénotypiques. De plus, comme évoqué précédemment avec l'exemple du gène *Tra-2* chez *D. melanogaster*, même un gène présentant une expression constitutive peut être impliqué dans la cascade de déterminisme du sexe (Amrein et al., 1988). Seuls des tests fonctionnels permettraient ainsi de mettre clairement en évidence les acteurs moléculaires impliqués dans le déterminisme du sexe. L'utilisation de protocoles tels que l'interférence par ARN ou encore de nouveaux outils moléculaires, tels que le système CRISPR/Cas9 (Barrangou & Doudna, 2016), pourrait permettre de bloquer indépendamment l'activité des différents gènes caractérisés chez *A. nasatum* et *A. vulgare*, et donc d'évaluer l'impact de chacun de ces gènes sur le phénotype sexuel de ces espèces. Ces tests fonctionnels pourraient permettre de faciliter de nouvelles avancées afin de mieux comprendre comment les gènes à domaine DM agissent sur le développement sexuel.

Afin de compléter ces premières données d'expression, il apparaît nécessaire de finaliser l'étude de l'expression des gènes à domaine DM et des gènes *Tra-2* identifiés chez *A. vulgare* au cours du développement. Cette analyse permettrait de savoir de façon directe si la transition de système d'hétérogamétie entre *A. nasatum* (XY) et *A. vulgare* (ZW) a modifié le patron d'expression de ces gènes. De plus, de nouvelles mesures effectuées sur des individus infectés par des *Wolbachia* féminisantes permettraient aussi non seulement de mieux comprendre le mécanisme d'action de l'endosymbiote sur le phénotype sexuel de son hôte, mais aussi pourrait aider à confirmer l'implication des gènes étudiés (et tout

particulièrement *Dsx*) dans le déterminisme du sexe. En effet, le mécanisme de la féminisation est aujourd'hui encore inconnu. Ainsi la mesure de l'activité des gènes à domaine DM et des gènes *Tra-2* au sein de lignées infectées permettrait de mieux caractériser la cible du mécanisme de féminisation de *Wolbachia* chez les isopodes terrestres, et donc de savoir si l'effet de *Wolbachia* agit en amont ou en aval de l'expression de ces gènes. Ces mesures pourraient aussi être étendues aux populations d'*A. vulgare* porteuses de l'élément f (pour rappel, il s'agit d'un large insert du génome de *Wolbachia* dans un autosome d'*A. vulgare*, ce dernier se comportant alors comme un chromosome sexuel W; Juchault & Mocquard, 1993; Cordaux et al., 2011; Leclercq et al., 2016), afin de comprendre l'impact moléculaire de cet insert bactérien sur les mécanismes de déterminisme du sexe de l'hôte. Plus globalement, le développement d'outils de sexage moléculaire associé à l'identification d'effecteurs impliqués dans le développement de phénotypes sexuels (mâles ou femelles) représentent des éléments essentiels pour mieux comprendre non seulement la cascade de déterminisme du sexe des isopodes terrestres mais aussi l'action moléculaire de *Wolbachia* conduisant à la féminisation de son hôte.

Chapitre 4

Diversité et évolution des systèmes de déterminisme du sexe chez les isopodes terrestres

Avant-propos

Les chapitres précédents se sont focalisés sur l'étude de deux espèces d'isopodes terrestres ayant chacune un système d'hétérogamétie différent : *A. nasatum* (XX/XY) et *A. vulgare* (ZZ/ZW). Il faut ajouter que, comme indiqué lors de l'introduction générale, les espèces d'Oniscidea dont le mécanisme de déterminisme du sexe est connu suivent un déterminisme hétérochromosomique du sexe (Rigaud et al., 1997). Ainsi parmi les neuf espèces d'isopodes terrestres dont le système d'hétérogamétie était connu avant notre étude, trois espèces présentent une hétérogamétie mâle (*A. nasatum* - Juchault & Legrand, 1979; *Porcellio dilatatus dilatatus* et *Helleria brevicornis* - Juchault & Legrand, 1964) et six espèces une hétérogamétie femelle (*A. vulgare* - Juchault & Legrand, 1972; *Porcellio dilatatus petiti* - Legrand et al., 1974; *Trachelipus rathkei* - Mittal & Pahwa, 1980; *Porcellio laevis* - Mittal & Pahwa, 1981; *Eluma purpurascens/caelatum* et *Oniscus asellus* - Juchault & Rigaud, 1995). Ces informations, bien que fragmentaires, suggèrent qu'une grande variabilité de systèmes d'hétérogamétie peut être rencontrée à l'échelle des isopodes terrestres. L'objectif de l'étude présentée dans le chapitre 4 était tout d'abord de caractériser les systèmes d'hétérogamétie de nouvelles espèces d'Oniscidea. L'établissement d'une phylogénie moléculaire robuste des Oniscidae a ensuite permis de déterminer le nombre et la direction des transitions de systèmes d'hétérogaméties rencontrés chez les espèces d'Oniscidae étudiées.

Cette étude a fait l'objet d'une publication par Becking et al. dans la revue *Scientific Reports* en avril 2017, qui est reproduite ci-après. Cet article est publié sous licence CC BY (*Creative Commons Attribution 4.0 International License*, <https://creativecommons.org/licenses/by/4.0/>) permettant sa libre diffusion. Par souci de lisibilité, chacune des figures de cette publication a été ajoutée à la fin de l'article en format pleine page. Les informations supplémentaires publiées avec cet article ont été ajoutées dans l'Annexe VII.

SCIENTIFIC REPORTS

OPEN

Diversity and evolution of sex determination systems in terrestrial isopods

Thomas Becking¹, Isabelle Giraud¹, Maryline Raimond¹, Bouziane Moumen¹, Christopher Chandler², Richard Cordaux¹ & Clément Gilbert¹

Received: 23 January 2017

Accepted: 27 March 2017

Published online: 24 April 2017

Sex determination systems are highly variable in many taxa, sometimes even between closely related species. Yet the number and direction of transitions between these systems have seldom been characterized, and the underlying mechanisms are still poorly understood. Here we generated transcriptomes for 19 species of terrestrial isopod crustaceans, many of which are infected by *Wolbachia* bacterial endosymbionts. Using 88 single-copy orthologous genes, we reconstructed a fully resolved and dated phylogeny of terrestrial isopods. An original approach involving crossings of sex-reversed individuals allowed us to characterize the heterogametic systems of five species (one XY/XX and four ZW/ZZ). Mapping of these and previously known heterogametic systems onto the terrestrial isopod phylogeny revealed between 3 and 13 transitions of sex determination systems during the evolution of these taxa, most frequently from female to male heterogamety. Our results support that WW individuals are viable in many species, suggesting sex chromosomes are at an incipient stage of their evolution. Together, these data are consistent with the hypothesis that nucleo-cytoplasmic conflicts generated by *Wolbachia* endosymbionts triggered recurrent turnovers of sex determination systems in terrestrial isopods. They further establish terrestrial isopods as a model to study evolutionary transitions in sex determination systems and pave the way to molecularly characterize these systems.

In metazoans, sex determination enables the development of individuals as males or females. Sex determination is widespread and, thus, could be expected to be governed by highly conserved mechanisms throughout metazoan evolution. While the bottom ends of sex determination pathways indeed tend to be conserved¹, the higher steps show extensive evolutionary flexibility, as they can be triggered by a variety of mechanisms². Genetic Sex Determination (GSD), in which the signal inducing sex differentiation is carried by sex chromosomes, is the most common mechanism in metazoans. In GSD, sex is determined by a locus that is heterozygous in one sex (the heterogametic sex) and homozygous in the other sex (the homogametic sex)². A variety of GSD systems exist and several approaches have been developed to characterize them directly or indirectly. Most of these approaches rely on the presence of differences in size and/or morphology in the sex chromosomes of the heterogametic sex (heteromorphy). Such differences are expected to result from the degeneration of the sex chromosome present only in the heterogametic sex, which, after gaining a master sex determination gene, ceases to recombine at meiosis³. Many studies have shown that the primary signal of sex determination can be subject to quick and frequent changes during evolution^{4–6}. Moreover, in several taxa, sex determination mechanisms do not match phylogenetic relationships, sometimes even in groups of closely related species^{7–9}. This indicates that sex chromosome turnovers can take place recurrently in some lineages.

Several evolutionary forces have been proposed to explain turnovers of sex determination mechanisms: sexually antagonistic selection¹⁰, accumulation of deleterious mutations in the heterogametic sex chromosome¹¹, the combination of these two effects known as the “hot potato model”¹², meiotic drive¹³, heterozygote advantage¹⁴ and sex ratio selection^{15, 16}. In particular, sex ratio selection occurs to restore balanced sex ratios in animals producing progenies that are skewed towards one sex¹⁷. Sex ratio biases can be caused by sex ratio distorters, which

¹Université de Poitiers, UMR CNRS 7267 Ecologie et Biologie des Interactions, Equipe Ecologie Evolution Symbiose, TSA 51106, 86073, Poitiers Cedex 9, France. ²Department of Biological Sciences, SUNY Oswego, Oswego, New York, 13126, USA. Richard Cordaux and Clément Gilbert contributed equally to this work. Correspondence and requests for materials should be addressed to R.C. (email: richard.cordaux@univ-poitiers.fr) or C.G. (email: clement.gilbert@univ-poitiers.fr)

are selfish genetic elements with sex-biased inheritance that enhance their own transmission by biasing sex ratios to their advantage.

One example of sex ratio distorters are *Wolbachia* bacterial endosymbionts. *Wolbachia* are strictly intracellular, maternally-inherited, α -proteobacteria that are widespread in arthropods and some nematodes¹⁸. In arthropods, *Wolbachia* often enhances its own transmission by manipulating host reproduction in favor of infected females, which can have important consequences on the evolution of host sex determination mechanisms¹⁹. This is well illustrated by the common pillbug *Armadillidium vulgare*^{19,20}. In this terrestrial isopod species, GSD follows female heterogamety (ZW females and ZZ males). However, some females have been shown to produce progenies that are highly biased towards females. This is due to the presence of feminizing *Wolbachia*, which induce feminization of (non-transmitting) genetic males into (transmitting) phenotypic females^{21,22}. In *A. vulgare* lines in which *Wolbachia* are present, the W sex chromosome ends up being lost. As a result, all individuals are ZZ genetic males, those carrying *Wolbachia* develop as females whereas uninfected individuals develop as males. As the *Wolbachia* transmission rate from mother to offspring is ~90%, progenies are composed of ~90% females and ~10% males. This is a perfect example of cytoplasmic sex determination^{19,20}.

Wolbachia-mediated sex ratio distortion induces nucleo-cytoplasmic conflicts, leading to strong selective pressures promoting the evolution of nuclear repressors of feminization that can restore balanced sex-ratios, i.e. masculinizing genes²³. Such nuclear repressors may evolve as new male sex-determining genes, thereby establishing new male heterogametic systems XY/XX²⁴. In addition, we recently showed that *Wolbachia* endosymbionts triggered the evolution of a new W sex chromosome in *A. vulgare* by horizontal genome transfer and nuclear incorporation as sex-determining factor²⁵. Then, *A. vulgare* appears to experience frequent turnovers of sex-determining factors ultimately caused by feminizing *Wolbachia* endosymbionts^{19,26}. In this context, the widespread distribution of *Wolbachia* infection in terrestrial isopods^{27,28} raises the possibility that *Wolbachia* endosymbionts may have impacted sex chromosome evolution in terrestrial isopods in general.

If *Wolbachia* endosymbionts triggered transitions between sex determination systems, it can be predicted that various heterogametic systems should be observed in diverse species and they should display a patchy distribution in the terrestrial isopod phylogeny. Interestingly, both female (ZW/ZZ) and male (XY/XX) heterogametic systems have been reported in various terrestrial isopod species²⁰. However, transitions of sex chromosomes are difficult to establish due to the lack of a robust phylogenetic framework. Two studies have so far attempted to reconstruct molecular phylogenies of some terrestrial isopod species^{29,30}, but both yielded poorly resolved trees owing to the small number of markers used and the relatively low amount of phylogenetic signal contained in these markers. Here, we investigated the diversity and evolution of sex determination systems in terrestrial isopods in two steps: (i) we reconstructed the first fully-resolved molecular phylogeny of terrestrial isopods, and (ii) we assessed heterogametic systems in terrestrial isopod species using experimental procedures.

Classically, characterization of heterogametic systems can be achieved by combining cytogenetic methods with fluorescent *in situ* hybridization, possibly associated with chromosome mapping. However, cytogenetic analyses are often difficult in isopods and they may be uninformative in the case where sex chromosomes are homomorphic. Alternatively, comparative genomics approaches have recently been developed, often relying on mapping/coverage analysis of both male and female genomes^{31–35}. However, these methods require large amounts of sequence data and crustacean genomes are often very large³⁶. To circumvent these issues, we used an original experimental strategy based on a combination of sex reversals and crossings of individuals with identical sex chromosome genotypes³⁷. These crossings are possible because of two features characterizing the peculiar sexual differentiation of terrestrial isopods: (i) individuals are sexually undifferentiated at birth and sexual differentiation occurs in a specific time window after birth^{38,39}, and (ii) male sexual differentiation is under the control of the androgenic hormone secreted by the male-specific androgenic gland^{40,41}. Then, it is possible to induce genetic females to differentiate as phenotypic males (termed “neomales”) by experimental implantation of androgenic glands during early sexual differentiation. These neomales are then crossed with their sisters, which is equivalent to crossing individuals with the same sex chromosome genotype. Analysis of the sex ratios of the resulting progenies allows definitive characterization of heterogametic systems (Fig. 1). This strategy enabled us to map heterogametic systems onto a robust phylogenetic context and infer transitions of sex chromosome systems during terrestrial isopod evolution.

Results

Phylogenetic analyses. To reconstruct the phylogeny of terrestrial isopods, we generated transcriptome data for 19 species available in our laboratory and used the transcriptomes of 5 other relevant crustaceans that were publicly available (Supplementary Table S1). Our search for orthologous genes in the transcriptomes of 24 crustacean species yielded 88 markers (81 nuclear and 7 mitochondrial). Out of 22 isopod species, we found 76–88 markers in all species but *Helleria brevicornis* (67 markers) and *Asellus aquaticus* (66 markers). In addition, all 88 markers were found in the two outgroup species *Cherax quadricarinatus* (Australian freshwater crayfish) and *Talitrus saltator* (European sand hopper). When combined, these markers produced a 69,570 bp-long alignment with a very low amount of missing data (7%). Two subspecies for which the heterogametic system is known were not included in our RNA-seq experiment: *Porcellio dilatatus dilatatus* (XY/XX)⁴² and *Porcellio dilatatus petiti* (ZZ/ZW)⁴³. To include these species in our phylogeny at lower cost, we first performed phylogenetic analyses of each of the 88 markers independently. Among the 35 markers that independently produced a topology identical to the one retrieved with the combined alignment, we selected the 10 longest ones (5 nuclear and 5 mitochondrial) and Sanger-sequenced them in the two *P. dilatatus* subspecies. The 10 markers sequenced in *P. d. dilatatus* (total length: 3,054 bp) and *P. d. petiti* (total length: 3,284 bp) were analyzed both separately as a combined 10-marker alignment and together with the other 78 markers. The phylogenetic analysis based on the 88-marker alignment produced a fully resolved tree, with all nodes supported by bootstrap values of 100% and all Bayesian posterior probabilities equal to 1 (Fig. 2, Supplementary Figure S1). The 10-marker tree was fully congruent with

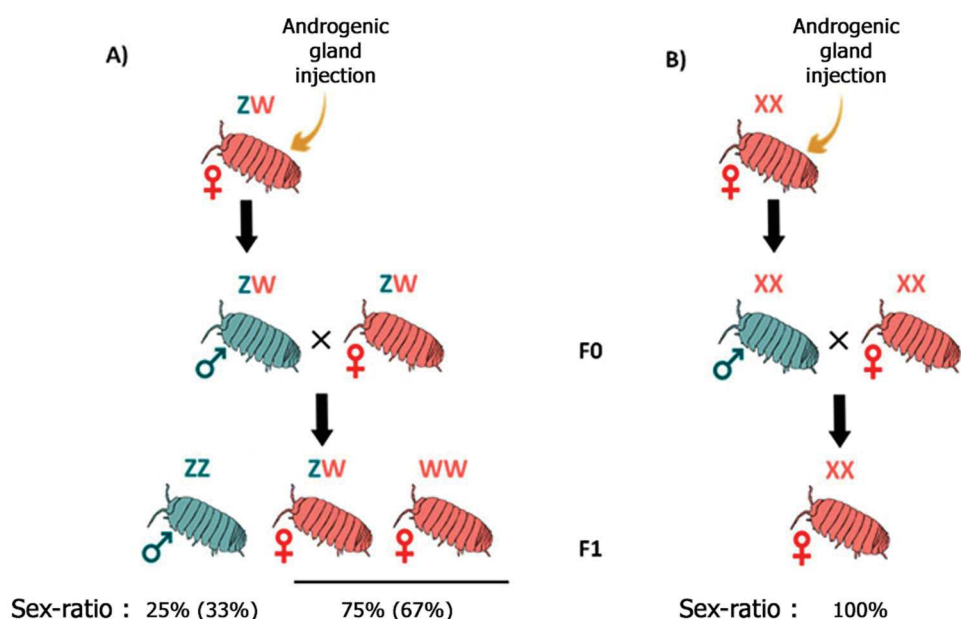


Figure 1. Principle of crossing experiments between neo-males (i.e. sex-reversed females) and genetic females. **(A)** Female heterogametic system (ZZ/ZW). The expected sex-ratio of the progeny of a neo-male and a genetic female is 25% males (ZZ) and 75% females (ZW and WW) if the WW genotype is viable, or 33% males (ZZ) and 67% females (ZW only) if the WW genotype is not viable. **(B)** Male heterogametic system (XX/XV). The expected sex-ratio of the progeny of a neo-male and a genetic female is 100% females (XX).

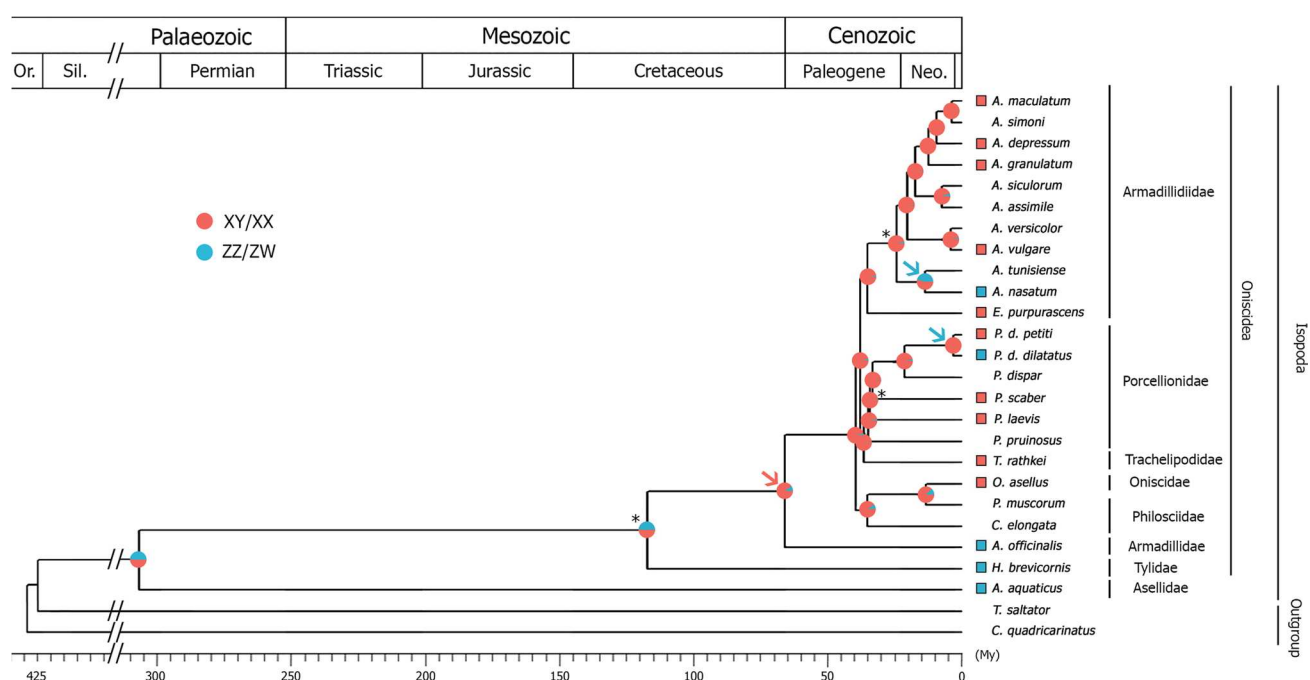


Figure 2. Ancestral state reconstruction of sex determination systems in terrestrial isopod crustaceans, based on a Bayesian phylogenetic analysis using 88 molecular markers. The reconstruction is based on the one-parameter equal rates model ("ER" model), considering the *A. assimile* sex determination system as unknown. Pie charts on nodes show the probability of the ancestral state at each node, calculated from 1,000 stochastic mappings. Branch length is scaled to time, with nodal ages corresponding to median posterior estimates. *Denote the calibration points. Red and blue squares at branch tips indicate a ZZ/ZW and XY/XX sex determination systems, respectively. Transitions of sex determination systems determined by parsimony ancestral state reconstruction are outlined with arrows (one red arrow for a transition from XY/XX system to ZZ/ZW, and two blue arrows for transitions from ZZ/ZW to XX/XY systems). Isopod family names and higher taxonomic ranks are shown on the right.

the 88-marker tree and strongly supported as well (all bootstrap values but one $\geq 95\%$, Supplementary Figure S1). We also compared the topology of the tree obtained using a combined alignment of the mitochondrial markers only (7 out of 88; alignment length: 5,847 bp) to that obtained with a combined alignment of the nuclear markers only (81 out of 88; alignment length: 63,723 bp). The two topologies were identical with each other and with the 88-marker and 10-marker trees (10 out of 88; alignment length: 10,773 bp), and they were both strongly supported (bootstrap scores for mitochondrial markers: all but two $\geq 95\%$; for nuclear markers: all but one $\geq 95\%$, Supplementary Figure S1).

Our phylogenetic analyses confirmed the monophyly of the Armadillidiidae family (all species from genus *Armadillidium* and *Eluma purpurascens*). The Porcellionidae family (all species from genus *Porcellio* and *Porcellionides pruinosus*) also appeared to be monophyletic, and was sister to the Trachelipodidae family (*Trachelipus rathkeii* in the phylogeny). By contrast, our analysis revealed the paraphyly of the Philosciidae family (*Philoscia muscorum* and *Chaetophiloscia elongata*), as it included *Oniscus asellus*, which belongs to the Oniscidae family. Finally, our analysis provided strong support for the position of the Tylidae family (*H. brevicornis* in the phylogeny) as sister to all other terrestrial isopods (Oniscidea suborder) (Fig. 2, Supplementary Figure S1).

Divergence times. Estimations of divergence times suggested that terrestrial isopods (Oniscidea) diverged from the aquatic isopod suborder Asellota (*A. aquaticus*) between the late Carboniferous (corresponding to the explosion of terrestrial arthropod diversity^{44,45}) and the early Permian (307 My, 95% Highest Posterior Density (HPD): 213–362 My, Fig. 2 and Supplementary Figure S2). As expected, the estimated age of the *Armadillidium* and *Porcellio* genera, and of the Oniscidea suborder fit with the age of their respective oldest known fossils used as calibration points (25 My, 95% HPD: 23–27 My; 35 My, 95% HPD: 34–37 My and 117 My, 95% HPD: 110–155 My, respectively, Fig. 2 and Supplementary Figure S2). According to our analyses, the Trachelipodidae family diverged from other terrestrial isopods during the late Eocene (38 My, 95% HPD: 35–42 My, Fig. 2 and Supplementary Figure S2). The *Oniscus* genus diverged from the *Philoscia* genus during the Miocene (15 My, 95% HPD: 12–19 My, Fig. 2 and Supplementary Figure S2). The paraphyletic family Philosciidae originated between the late Eocene and the early Oligocene (36 My, 95% HPD: 32–48 My, Fig. 2 and Supplementary Figure S2). Finally, the *Armadillo* genus diverged from other terrestrial isopods during the late Cretaceous (66 My, 95% HPD: 57–79 My, Fig. 2 and Supplementary Figure S2).

Substitution rates. Mitochondrial and nuclear codon alignments were used to estimate the rate of silent substitutions across isopods. When scaled to the fossil-calibrated chronogram, substitution rates in isopods were estimated at 3.89×10^{-9} and 4.17×10^{-9} silent substitutions per year for mitochondrial and nuclear sequences, respectively (Supplementary Figure S3). We also calculated the rate of non-synonymous/synonymous substitutions (dN/dS) for the two datasets. Both mitochondrial and nuclear dN/dS ratios were < 1 , thus indicating that our markers evolve under strong purifying selection (mitochondrial dN/dS = 0.28; nuclear dN/dS = 0.07).

Assessment of heterogametic systems. Injections of androgenic glands were attempted in 135 young females from 15 terrestrial isopod species, representing 4–18 females per species (Supplementary Table S2). We discarded 25 individuals for which sex reversal was not successful (from 0 to 4 individuals per species) because they developed as females. In addition, 60 individuals died during the experiments (from 0 to 8 individuals per species). The remaining 50 individuals from 13 species developed as putative neomales (none was obtained for *P. muscorum* and *C. elongata*). Each putative neomale was crossed with 2 to 3 sisters for a total of 125 crosses, representing 2 to 18 crosses per species. We obtained 43 progenies from 11 species (none was obtained for *Armadillidium tunisiense* and *P. pruinosus*), representing 1–8 progenies per species (Supplementary Table S2).

Among the 43 progenies, 5 were discarded because the mothers were infected by *Wolbachia* (Supplementary Table S3). We also discarded 11 progenies because observed sex ratios did not significantly differ from a balanced sex ratio. This may be explained by small progeny sizes resulting in lack of statistical power and/or by the fact that individuals selected for sex reversal were undifferentiated young males instead of females. As a result, crosses occurred between genetic males (instead of neomales) and genetic females, hence the balanced sex ratio in the progenies. In sum, we obtained 27 progenies showing unbalanced sex ratios in the absence of *Wolbachia* infection from 7 species, representing 1–6 progenies per species (Supplementary Table S3).

For each species for which we obtained at least two progenies, χ^2 tests indicated consistent sex ratio deviations among progenies. Therefore, we pooled data from these progenies and calculated overall sex ratios for each species (Fig. 3, Supplementary Table S4). Four species showed sex ratios indicating a female heterogametic system (ZZ/ZW): *Armadillidium depressum*, *Armadillidium granulatum*, *Armadillidium maculatum* and *Porcellio scaber*. Interestingly, for the three *Armadillidium* species, sex ratios were statistically different from 1/3♂:2/3♀ but not from 1/4♂:3/4♀, suggesting that the WW genotype is viable in these species (Supplementary Table S4). For *P. scaber*, the sex ratio was consistent with both 1/3♂:2/3♀ and 1/4♂:3/4♀, which prevented any conclusion to be drawn about viability of the WW genotype. This lack of resolution is probably attributable to the fact that we obtained only 30 offspring for this species. In addition, 2 species showed sex ratios indicating a male heterogametic system (XY/XX): *Armadillidium nasatum* and *Armadillo officinalis*. Our results for *A. nasatum* confirmed earlier findings⁴⁶ and showed our approach is robust and reproducible. Finally, the sex ratio observed in *Armadillidium assimile* (100%♂:0%♀) cannot readily be explained by our predictions (Fig. 1), and further investigations are required to fully understand these results.

A review of the literature indicated that the heterogametic system has previously been assessed in 6 other terrestrial isopod species using sex reversal experiments^{37,42,43,47} (Supplementary Table S3). To further analyze the data produced by these studies, we performed χ^2 tests, which were not systematically conducted at the time. All tests indicated consistent sex ratio deviations among progenies for each species. Therefore, we pooled data from these progenies and calculated overall sex ratios for each species (Fig. 3, Supplementary Table S4). We confirmed



Figure 3. Sex ratio of progenies of 13 terrestrial isopod species resulting from crosses between genetic and sex-reversed females (neomales). Red and blue colors indicate the percentages of females and males, respectively. Red dashed lines indicate a 1/4♂: 3/4♀ sex ratio (expected in ZZ/ZW systems) and blue dashed lines indicate a 0♂: 1♀ sex ratio (expected in XY/XX systems). *n* is the number of pooled individuals per sex, and the number of corresponding progenies are specified in brackets for each species.

the original conclusions of a female heterogametic system (ZZ/ZW) in 4 species: *O. asellus* and *E. purpurascens*³⁷, *P. d. petiti*⁴³ and *Armadillidium vulgare*⁴⁷. Our statistical results indicated that the WW genotype is apparently viable in all four species. This confirmed earlier findings established after crossing F1 females (WW) with genetic males (ZZ) and analyzing F2 sex ratios under the prediction that such crossings should entirely be composed of females (ZW)^{37,47}. The same approach demonstrated that *P. d. petiti* has a viable WW genotype⁴³. We also confirmed that *H. brevicornis* and *P. d. dilatatus* have a male heterogametic system (XX/XY).

Transitions of heterogametic systems in terrestrial isopods. In addition to the 13 terrestrial isopod species discussed in the previous section, the heterogametic system has been determined using cytogenetics in the terrestrial isopods *Porcellio laevis* (ZZ/ZW)⁴⁸ and *T. rathkei* (ZZ/ZW)⁴⁹, as well as in a few aquatic isopods, including *A. aquaticus* (XY/XX)⁵⁰. Mapping of sex determination systems onto the terrestrial isopod phylogeny revealed a heterogeneous distribution, which cannot readily be explained by a simple evolutionary scenario involving a single transition between male and female heterogametic systems. Despite missing information for 9 species (including *A. assimile* and 8 out of 24 ingroup species in our phylogenetic tree), the MESQUITE software⁵¹ parsimoniously identified that at least 3 transitions are required to explain the distribution of these sex determination systems (Fig. 2): one transition from male to female heterogamety (in the ancestor of a clade comprising the Armadillidiidae, Porcellionidae, Trachelipodidae, Oniscidae and Philoscidae families) and two transitions from female to male heterogamety (between the two subspecies *P. d. dilatatus* (XY/XX) and *P. d. petiti* (ZZ/ZW) and within the Armadillidiidae family). A fourth transition is required if *A. assimile* sex determination mechanism is considered as a new and independent system (in the ancestor of a clade grouping *Armadillidium siculorum* and *A. assimile* or later on the branch leading to *A. assimile* if *A. siculorum* is ZZ/ZW). In a separate analysis, a maximum likelihood-based reconstruction of ancestral states yielded between 11 and 13 transitions, depending on the model and the type of discrete states used (Supplementary Table S5). Both parsimony and maximum likelihood-based methods indicated that transitions from female to male heterogamety are twice as frequent as transitions from male to female heterogamety (Supplementary Table S5).

Discussion

Phylogenetic analyses. In this study, we used a large combination of 88 nuclear and mitochondrial molecular markers obtained by transcriptome sequencing to establish a fully resolved molecular phylogeny of terrestrial isopods (Supplementary Figure S1). We also identified a subset of 10 markers which produced the same topology with strong bootstrap support. We designed degenerate primers for these 10 loci and we showed they can be used to complement the phylogeny with additional species. As obtaining transcriptomes for a large number of species may be costly and/or time-consuming, we expect that the cheap and fast alternative 10-marker strategy we propose may enable further phylogenetic analyses of terrestrial isopods. Oniscidea being one of very few crustacean groups adapted to terrestrial lifestyle, their study provides unique research opportunities in fields such as ecology, ecophysiology and evolution⁵², which will be facilitated by the establishment of an adequately resolved phylogeny.

Our fully resolved phylogeny confirmed the monophyly of the Armadillidiidae and Porcellionidae families. These two taxa were considered as sister families²⁹. However, this earlier study did not include a member of the Trachelipodidae family, which appears to be sister to the Porcellionidae, according to our phylogeny. Our results also highlight the intriguing evolution of *H. brevicornis*. As the only recognized species of the Helleriinae subfamily (grouped with the *Tylos* genus in the Tylidae family), this endemic species from the North Tyrrhenian area⁵³ has been considered as a branch that diverged early during terrestrial isopod evolution based on morphological characters⁵⁴. Characterized by a high mitochondrial AT nucleotide content, the early divergence of *H. brevicornis* from the other Oniscidea was already suspected²⁹, without being able to solve the phylogenetic relationships of this species. Our work offers a definitive conclusion by confirming earlier molecular and morphological studies (Fig. 2). Our results also show that the Philosciidae family is paraphyletic, because of the clustering of *O. asellus* (Oniscidae) with *C. elongata* and *P. muscorum*. The close relationship between these two families has been suggested^{30,55} but their respective monophyly could not be tested because of insufficient species sampling. This close relationship is congruent with morphological analyses, as both families share 5 pairs of similar pleopodal lungs covered with a slightly wavy membrane, a key character in the evolution of respiratory structures⁵⁶. However, there is currently no obvious key morphological characteristic common to all Philosciidae species, which is a family presenting one of the highest species richness among terrestrial isopods and the richest family in terms of genera⁵⁵. The unambiguous paraphyly of Philosciidae reported here calls for a reevaluation of the taxonomy of this family, which could involve its subdivision in several smaller families.

Divergence times and estimates of substitution rates. Unlike insects, terrestrial isopods do not possess a waxy epicuticle, thus complicating the fossilization process⁵⁷. The terrestrial isopod fossil record is thus relatively poor, implying that the divergence times of these taxa must be interpreted with caution. Our results differ from those obtained in an earlier study⁵⁸, which included several terrestrial isopods as part of an investigation of the evolution of deep sea isopods. This discrepancy is likely due to the fact that they did not calibrate any node within terrestrial isopods⁵⁸. Thus, our analyses are more likely to be closer to the true divergence times for this clade.

Our study also allowed us to calculate absolute silent substitution rates (dS) in terrestrial isopods, which are very similar between mitochondrial and nuclear sequences (3.89×10^{-9} and 4.17×10^{-9} silent substitutions per years, respectively). In comparison, several studies based on mitochondrial COI sequences from crustacean species showed substantial dS variations depending on the studied order. Whereas mitochondrial dS rates similar to those reported here were found in copepods (9.3×10^{-9} , *Tigriopus* genus⁵⁹) and aquatic isopods (4.6×10^{-9} , *Stenasellus* genus⁶⁰), it has been established that this rate might be 20 times higher in decapods (86×10^{-9} in genus *Alpheus*⁶¹) for the same sequence (COI). More surprisingly, mitochondrial genes usually evolve at an elevated rate compared to nuclear genes, e.g. 11 to 62-fold higher in mammals and 1.4 to 18-fold higher in insects^{62,63}. By contrast, we did not observe such a variation in our datasets where mitochondrial and nuclear dS appear to be similar. To some extent, such a situation has already been reported in corals⁶⁴, fungi⁶⁵ and sponges⁶⁶. In these taxa, a specific mitochondrial DNA repair function could be responsible for the low levels of mitochondrial divergence. In spite of the availability of well assembled mitochondrial genomes, no such gene is known in terrestrial isopods^{67,68} and at present, we cannot explain why mitochondrial and nuclear dS are similar.

Diversity of sex determination systems. We used an original strategy combining sex reversals and crossings of individuals with the same sex chromosome genotype to characterize heterogametic systems in 5 species of terrestrial isopods, including 4 species showing female heterogamety (*A. depressum*, *A. granulatatum*, *A. maculatum* and *P. scaber*) and 1 species with male heterogamety (*A. officinalis*) (Fig. 3). In general, female heterogamety appears to be more frequent (10/16 species) than male heterogamety (5/16 species) in terrestrial isopods (Fig. 3 and Table 1). This does not necessarily imply that female heterogamety is the ancestral sex determination system of this clade. Indeed, the three most deeply branching species within isopods exhibit male heterogamety. Homology of the Y chromosomes between these species would constitute a strong argument for an ancestral male heterogamety in terrestrial isopods, but our approach does not allow to assess whether sex chromosomes are homologous or not. Nevertheless, sex chromosome turnovers apparently occur frequently in terrestrial isopods (see discussion below), which makes it possible that the Y chromosomes of deep-branching species are not homologous. Sequencing the genomes of *A. officinalis*, *H. brevicornis* and *A. aquaticus* will be useful to settle the issue of the ancestral heterogametic type in terrestrial isopods.

Investigation of heterogametic systems in *A. assimile* yielded surprising results. Indeed, all broods showed an unexpected and reproducible sex ratio bias exclusively composed of males, which cannot be readily explained by our predictions under standard male or female heterogamety (Fig. 1). We did not observe embryonic mortality in the 6 analyzed females, thus excluding a selective elimination of female embryos. Furthermore, it seems unlikely that sex reversal could affect the sex or the mortality of the progenies, especially since such an observation has never been reported for other sex reversal experiments in the literature. The coexistence of three sex chromosomes, such as Y, W and Z (as in the amphibian *Xenopus tropicalis*⁶⁹) or W, Y and X (as in the platyfish *Xiphophorus maculatus*⁷⁰) is also excluded because crossings of genetic females would necessarily produce females in the offspring whatever the combination of sex chromosomes. The hypothesis of a multi-locus sex determination system cannot be excluded, implying a combination of a major sex factor with numerous minor sex factors, as previously described in the marine isopod *Idotea balthica*⁴¹. Finally, it is noteworthy that male monogeny, as observed in *A. assimile*, also exists in the dipteran *Mayetiola destructor*⁷¹, in which sex determination is governed by the ratio between sex chromosomes and autosomes and involves maintenance/elimination of paternally derived sex chromosomes. Further experiments are needed to characterize the intriguing sex determination mechanism of *A. assimile*.

Suborder	Family	Species	Heterogametic type (Reference)	Known presence of <i>Wolbachia</i> (Reference)
Oniscidea	Armadillidiidae	<i>Armadillidium maculatum</i>	ZZ/ZW (This study)	No ²⁷
		<i>Armadillidium simoni</i>	Unknown	No (Unpublished)
		<i>Armadillidium depressum</i>	ZZ/ZW (This study)	Yes (This study, Supplementary Table S3)
		<i>Armadillidium granulatum</i>	ZZ/ZW (This study)	Yes ²⁸
		<i>Armadillidium siculorum</i>	Unknown	No (This study, Supplementary Table S3)
		<i>Armadillidium assimile</i>	Unknown* (This study)	No (This study, Supplementary Table S3)
		<i>Armadillidium versicolor</i>	Unknown	No (This study, Supplementary Table S3)
		<i>Armadillidium vulgare</i>	ZZ/ZW ⁴⁷	Yes ²²
		<i>Armadillidium tunisiense</i>	Unknown	Yes ¹⁵
		<i>Armadillidium nasatum</i>	XY/XX ⁴⁶	Yes ²⁷
		<i>Eluma purpurascens (caelatum)</i>	ZZ/ZW ³⁷	No ²⁷
	Porcellionidae	<i>Porcellio dilatatus petiti</i>	ZZ/ZW ⁴³	Yes ²⁷
		<i>Porcellio dilatatus dilatatus</i>	XY/XX ⁴²	Yes ¹⁶
		<i>Porcellio dispar</i>	Unknown	Yes ²⁷
		<i>Porcellio scaber</i>	ZZ/ZW (This study)	Yes ²⁷
		<i>Porcellio laevis</i>	ZZ/ZW ⁴⁸	Yes ¹⁵
		<i>Porcellionides pruinosus</i>	Unknown	Yes ¹⁷
	Trachelipodidae	<i>Trachelipus rathkei</i>	ZZ/ZW ⁴⁹	Yes ²⁸
	Oniscidae	<i>Oniscus asellus</i>	ZZ/ZW ³⁷	Yes ²⁷
	Philosciidae	<i>Philoscia muscorum</i>	Unknown	Yes ²⁷
		<i>Chaetophiloscia elongata</i>	Unknown	Yes ¹⁸
	Armadillidae	<i>Armadillo officinalis</i>	XY/XX (This study)	Yes ²⁷
	Tylidae	<i>Helleria brevicornis</i>	XY/XX ⁴²	Yes ²⁷
Asellota	Asellidae	<i>Asellus aquaticus</i>	XY/XX ⁵⁰	Yes ²⁷

Table 1. Heterogametic types found in isopod crustaceans. (*considered as unknown because crossings of sex reversed individuals yielded inconclusive results with respect to our predictions for XY/XX and ZZ/ZW systems).

Our results indicate that a wide variety of sex determination systems exists in terrestrial isopods, mainly in the form of heterogametic systems. It is noteworthy that an even larger diversity of systems is known in isopods. For example, protogynous hermaphroditism was reported in *Cyathura carinata* (Cymothoida suborder)⁷² and *Gnorimosphaeroma oregonense* (Sphaeromatidea suborder)⁷³, and protandric hermaphroditism in *Anilocra frontalis* (Cymothoida suborder)⁷⁴. While these marine taxa have never been included in a phylogenetic analysis, the suborders to which they belong seemingly fall outside of the isopod clade included in this study. Cases of parthenogenetic isopods have also been reported, with coexisting forms of sexual diploid and parthenogenetic triploid forms in *Trichoniscus pusillus pusillus*⁷⁵. The fact that this species falls within the terrestrial isopod clade included in this study implies that the true number of transitions between sex determination mechanisms in isopods is necessarily higher than that indicated by our data. Whether these other systems are as widespread as heterogametic systems in other parts of the isopod phylogeny will be worth investigating in the future.

Evolution of sex chromosome turnovers. We found that the distribution of sex determination mechanisms in the terrestrial isopod phylogeny requires at least three transitions between systems since the late Cretaceous (Fig. 2), but more probably between 11 and 13 according to our probabilistic analysis of reconstruction of ancestral states. Importantly, the number of transitions is necessarily underestimated, because our analysis is based on the assumption that all W chromosomes are homologous, and that all Y chromosomes are also homologous. Therefore, transitions of sex chromosomes preserving the heterogametic type cannot be taken into account in our analyses. Yet, such transitions are known to be possible in terrestrial isopods, as shown in *A. vulgare*. In some lines of this species, the presence of feminizing *Wolbachia* endosymbionts induced the loss of the native W sex chromosome under cytoplasmic sex determination^{19,20}, followed by the evolution of a new W sex chromosome by horizontal transfer of the *Wolbachia* genome in the *A. vulgare* genome²⁵. This effectively led to a turnover of sex chromosomes without changing the heterogametic type.

More generally, *Wolbachia* endosymbionts are known to be widespread in terrestrial isopods^{27,28}, suggesting that these bacteria may have been involved in sex chromosome transitions in species other than *A. vulgare*. Our results on the distribution of sex determination systems in the terrestrial isopod phylogeny are in agreement with the prediction that if *Wolbachia* endosymbionts triggered transitions between sex determination systems, various heterogametic systems should be expected in diverse species and they should display a patchy distribution in the terrestrial isopod phylogeny. Another prediction arising from this scenario is that repeated turnovers of sex chromosomes may prevent the degeneration of sex chromosomes¹⁰ because of the regular relocation of sex-determining genes on different autosomal pairs. As a result, one would expect terrestrial isopods to exhibit homomorphic sex chromosomes or, at least, to show limited heteromorphy. This is congruent with the fact that homomorphic sex chromosomes are generally observed in isopods^{19,37}. However, cytogenetically homomorphic sex chromosomes could differ substantially in terms of gene content. In this context, sex reversal experiments are particularly powerful because they indicate that a simple experimental implantation of androgenic gland is sufficient to reverse an undifferentiated female into a male. This clearly implies that both male and female isopods share most if not all genes required for differentiation of either sex, and that these genes are identical in both sexes^{20,41}. The apparently labile mechanism of sex differentiation observed in terrestrial isopods probably facilitated the occurrence of feminizing endosymbionts such as *Wolbachia*⁷⁶. Interestingly, our approach allows us to statistically infer whether WW genotypes may be viable. Our results are consistent with largely homomorphic sex chromosomes in most ZZ/ZW species³⁷. Indeed, WW genotypes are inferred to be viable in at least 7 out of 8 species and this conclusion has been experimentally confirmed in 4 of these species³⁷. Furthermore, the viability of YY genotypes has been experimentally confirmed in 3 out of the four XY/XX species⁴². This is in stark contrast with the situation in mammals and insects, where YY individuals are often sterile or inviable^{2,77}.

Conclusion

Our inference of between three and 13 transitions between heterogametic XX/XY and ZZ/ZW systems of sex determination in terrestrial isopods is quite remarkable because such transitions are believed to be rare. For example, while sex determination systems are known for thousands of insect species, only two such transitions have been inferred in these taxa, one in the ancestor of the superorder Amphiesmenoptera (Lepidoptera and Trichoptera), and one within the Tephritidae family of Diptera⁷⁸. The paucity of such transitions may be explained by the necessary production of offspring that are homozygous for the Y or W chromosome. Such offspring may often be inviable or sterile in the context of heteromorphy, where gene content differs between sex chromosomes^{2,77,78}. In terrestrial isopods, transitions between ZZ/ZW and XX/XY systems are likely facilitated by the fact that gene content is nearly identical between sex chromosomes, as indicated by the viability of WW individuals in most ZZ/ZW isopod species. Overall, the repeated and likely recent transitions of heterogametic systems in terrestrial isopods are consistent with the hypothesis that *Wolbachia* endosymbionts have had a strong impact on the evolution of sex determination mechanisms in these taxa. Terrestrial isopods thus provide a great opportunity to further understand the evolutionary forces driving transitions and early evolution of sex chromosomes.

The resulting turnovers also raise the question of the role of *Wolbachia* in host divergence and speciation. Sex chromosomes can play key roles in phenotypic evolution and speciation. For example, studies based on genetic mapping have shown that hybrid deficiencies, such as hybrid sterility, mostly map to sex chromosomes^{79,80}. Sex chromosomes often harbor genes linked to mate choice or behavioral isolation in a variety of species such as Poeciliid fishes⁸¹, flycatchers⁸², stickleback⁸³ or finches⁸⁴. Whether such sexually antagonistic genes have sufficient time to accumulate or to be expressed in a sex-biased fashion on otherwise homomorphic sex chromosomes in terrestrial isopods is an open question deserving further investigation. More generally, it will be interesting to assess the contribution that feminizing *Wolbachia* may have had as a trigger of sex chromosome turnovers in relation to speciation, in generating the large, worldwide diversity of terrestrial isopods^{55,85}.

Materials and Methods

RNA preparation, sequencing and *de novo* assembly. Transcriptome data were generated for 19 species of terrestrial isopods (Supplementary Table S1) available in our laboratory. Total RNA was extracted from one or two adult individuals frozen in liquid nitrogen and grinded with a mortar and pestle. The resulting powders were processed using the RNeasy extraction protocol (Qiagen). Quality and quantity of total RNA was examined using a MCE-202 MultiNA (Shimadzu) and a Qubit 2.0 (Invitrogen). For each species, 1–12 µg of total RNA were extracted at concentrations ranging from 15.5 ng µL⁻¹ to 240 ng µL⁻¹. Paired-end sequencing libraries were constructed and sequenced by Eurofins (Germany) on one multiplexed lane of an Illumina HiSeq2500 platform, with a read length of 125 bp. The total number of reads obtained ranged from 12,766,990 to 24,948,176 depending on the species (Supplementary Table S1). For *T. rathkei*, RNA was isolated from head, leg, and gonad tissues from one male and one female, both wild-caught at Rice Creek Field Station in Oswego, NY, using an SV Total RNA Isolation Kit (Promega). Libraries for each sample were prepared at the State University of New York at Buffalo Genomics Core Facility, barcoded, pooled, and sequenced in a single paired-end Illumina HiSeq2500 lane with a read length of 100 bp. Sequence data of the 19 transcriptomes generated in this study have been deposited in GenBank under accession numbers SRX2600476-SRX2600493 and SRR5198726-SRR5198727. Transcriptome data from five additional crustacean species were included in our analyses: *A. vulgare*⁸⁶, *A. nasatum*⁸⁶, *A. aquaticus*⁸⁷, *C. quadricarinatus*⁸⁸ and *T. saltator*⁸⁹ (all SRA accession numbers are provided in Supplementary Table S1).

Read quality was analyzed with FastQC (version 0.11.4, <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>). Removal of low quality reads and sequencing adaptors was performed with Trimmomatic (version 0.33⁹⁰), setting the ILLUMINACLIP palindrome clip threshold at 30 and the simple clip threshold at 10. The quality threshold was set to a minimum Phred score of 20. Filtered reads were assembled using the Trinity *de novo* Assembler (release 2014 07 17⁹¹). Assembly metrics for each transcriptome are provided in Supplementary Table S6.

Marker selection and phylogenetic inferences. To establish a set of orthologous genes to be used in phylogenetic analyses, we first translated each transcriptome using Virtual Ribosome (version 1.1⁹²). The first translation strategy consisted in keeping the longest contigs starting with a start codon among the 6 possible open reading frames (ORFs) (parameters used: -readingframe = all, -orf = any). The second strategy consisted in keeping the longest contigs ending with a stop codon among the 6 possible ORFs (parameters used: -readingframe = all, -orf = none). Of these two datasets of translated expected proteins, the longest sequences were kept (including sequences that had both a start and a stop codon, or sequences that had only a start or a stop codon).

Orthology relationships among all terrestrial isopod proteomes (i.e. all species except *T. saltator*, *C. quadricarinatus* and *A. aquaticus*) were inferred using the OrthoMCL pipeline⁹³. Among the 45,724 orthology groups returned, 2,168 contained at least 20 of the 21 taxa. We focused on orthogroups containing at most 3 paralogous sequences, corresponding to 90 orthogroups. Annotation of these orthogroups was performed using BlastX⁹⁴ against the non-redundant (NR) database of NCBI (version January 2016) (Supplementary Table S7). Two orthogroups showing no significant hit to any known protein were removed from this dataset, as they could not be confidently assigned to a terrestrial isopod origin (orthogroups 39 and 60). The orthologous sequences from *T. saltator*, *C. quadricarinatus* and *A. aquaticus* were identified using a one-to-one reciprocal best BlastP alignment approach on the 88 terrestrial isopod orthogroups. This search was performed after the definition of the terrestrial isopod orthogroups to avoid underestimating the amount of orthogroups determined by OrthoMCL due to the large genetic distances between terrestrial isopods and *T. saltator* and *C. quadricarinatus*, or because of a lack of data for *A. aquaticus*, which transcriptome contained 3 times fewer assembled sequences and 7 times fewer nucleotides than other transcriptomes (Supplementary Table S6). All 88 orthologous sequences were found in *T. saltator* and *C. quadricarinatus* and 66 orthologous sequences were found in *A. aquaticus*.

All sequences within each orthogroup were codon-aligned using default parameters in Geneious (version 7.0.6⁹⁵). Each multiple alignment was then trimmed with GBLOCKS (version 0.91b⁹⁶) to remove ambiguous regions. All alignments were finally concatenated into a single nexus interleaved file using SequenceMatrix (version 1.8⁹⁷). To determine the best-fit model of nucleotide substitution, we used Jmodeltest (version 2.1.7⁹⁸). The best substitution model was the GTR + G + I, according to all information criteria. Maximum-likelihood analyses were performed on the concatenated codon alignment using RAxML (version 7.4.6⁹⁹) with 100 independent replicates followed by 200 replicates of bootstrap resampling. Bootstrap values were subsequently mapped onto the optimal consensus tree obtained from the 100 independent searches. The two non-isopod species *T. saltator* (Amphipoda) and *C. quadricarinatus* (Decapoda) were used as outgroups.

To include two additional species for which the heterogametic system is known (*P. d. dilatatus* [XY/XX] and *P. d. petiti* [ZZ/ZW]), we selected ten markers (5 nuclear and 5 mitochondrial) on the basis of their length and phylogenetic signal and we Sanger-sequenced them. PCRs were performed using degenerate primers as follows: 3 min at 94 °C for the initial denaturing step, followed by 35 cycles of 30 s at 94 °C; 30 s at 54 °C, 56 °C, 57 °C or 58 °C (depending on the melting temperature of the primers) and 1 min at 72 °C. The final elongation step was 10 min at 72 °C. Purified PCR products were sequenced on an ABI PRISM 3130xl automated sequencer (Applied Biosystems). Primer sequences, PCR product sizes and associated melting temperatures are listed in Supplementary Table S8.

Divergence time analyses. Divergence time analyses were performed using BEAST (version 1.8.3¹⁰⁰) on the combined alignment including 88 markers for all species except *P. d. dilatatus* and *P. d. petiti* (10 markers) and *A. aquaticus* (66 markers). We used the RAxML topology tree as starting tree and conducted a random clock analysis. To obtain the Bayesian posterior probability for each node, the same analysis was performed with no constrained topology as starting tree. The rate parameter (*clock.rate*) was left to default prior distribution (CTMC Rate Reference with an initial value of 0.01¹⁰¹). A Poisson distribution was used for the number of local clocks (*rateChanges* parameter, mean 0.693) and an exponential prior distribution was set for the relative rates among the random local clock (*localClock.relativeRates*, mean 1.0, initial value 0.001). Base frequencies (*frequencies*) and proportion of invariant sites parameter (*pInv*) were left to their initial priors (uniform distribution with an initial value of 0.25, and uniform distribution with an initial value of 0.5, respectively). The gamma shape parameter (*alpha*) was also left to his initial prior (exponential distribution, initial value 0.5, mean 0.5). Three fossil calibration points were used to calibrate our molecular clock analysis. A first fossil found in amber from the early beginning of the Oligocene in Germany provides a minimum age of 23.03 million years (My) for the *Armadillidium* genus¹⁰². A second fossil specimen in amber from the Oligocene close to the Baltic Sea in Northern Europe indicates a minimum age of 33.9 My for the *Porcellio* genus¹⁰³. Third, the earliest terrestrial isopod fossils were found in amber from the Lower Cretaceous of Northern Spain, which allowed us to set the minimum age of terrestrial isopods (Oniscidea) at 110 My^{104, 105}. All calibrations were modeled following a gamma prior distribution (shape 2.0), with an offset value fixed at the oldest age bordering the time interval containing the related fossil (*Armadillidium*: Miocene 5.333 to 23.03 My; *Porcellio*: Oligocene 23.03 to 33.9 My).

Markov chain Monte Carlo (MCMC) were run for 50 million generations, and parameter values were sampled every 1,000 generations. The log file output was processed with LogCombiner (version 1.8.3, implemented in BEAST) with 10% burn-ins removed. The resulting log file was analyzed with Tracer (version 1.6, also implemented in BEAST) to confirm that the MCMC converged with an effective sample size greater than 200 for all priors. The maximum clade credibility tree was created with TreeAnnotator (part of BEAST suit), and the resulting tree was visualized with FigTree (version 1.4.2, <http://tree.bio.ed.ac.uk/software/figtree>) and R software (version 3.1.1¹⁰⁶), using the packages *strap*¹⁰⁷ and *phyloch*¹⁰⁸.

Substitution rates. Using the RAxML topology tree and a nuclear codon alignment of the 81 concatenated nuclear sequences, maximum likelihood estimates of branch-specific synonymous and non-synonymous

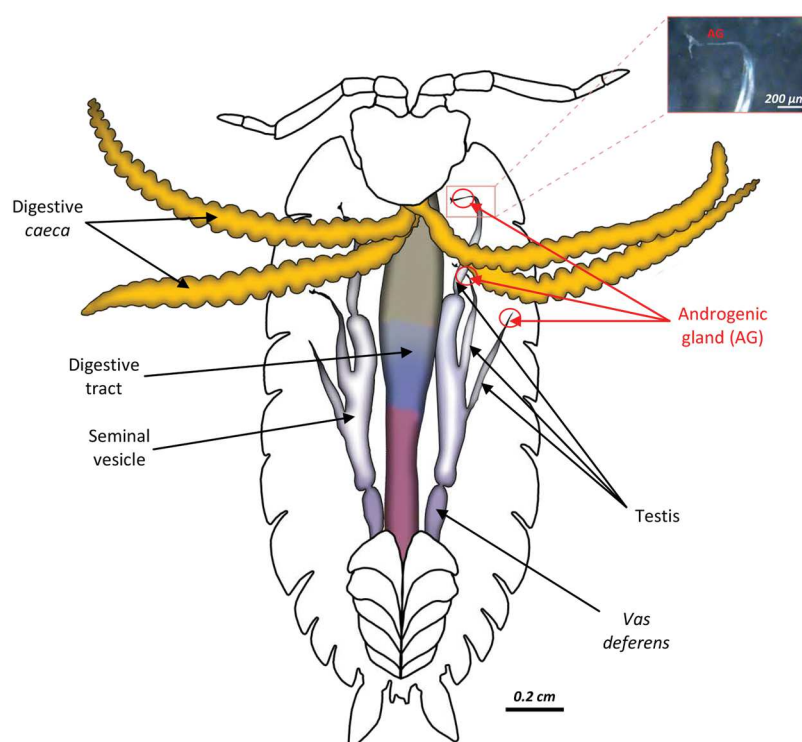


Figure 4. Schematics of the internal organs of a terrestrial isopod (*Oniscus asellus*) in ventral view. Androgenic glands, composed of a cellular mass excreting androgenic hormone, are located at the top of each testis, included in the suspensory filament. Depending on the species, its size is 250–700 μm in length and 20–150 μm in width.

substitution rates (dS and dN) were calculated with CodeML, a program from the PAML software package (version 4.9c^{109, 110}), with default parameters. The analysis was also conducted independently with a concatenated codon alignment of the 7 mitochondrial markers. To scale the substitution rate per site to absolute substitution rate, dS values calculated with CodeML were divided by the BEAST median estimate of branch age for each branch of the tree, in both nuclear and mitochondrial analyses. The global dS of Oniscidea was calculated by adding dS values of all branches, divided by the sum of ages calculated for all branches of the tree, for both nuclear and mitochondrial sequences.

Sex-reversal experiments. Sex reversal experiments were attempted on 15 terrestrial isopod species listed in Supplementary Table S2. Sex reversion of genetic females into functional phenotypic males (neomales) was performed by injection of androgenic glands into female embryos following the technique used in older study⁴². In crustaceans, androgenic glands are located at the end of each male testes (Fig. 4). They control androgenic hormone production, which itself triggers male sexual differentiation⁴⁰. For each species, fresh androgenic glands were first dissected under a magnifying binocular from three adult males placed in Ringer solution (NaCl: 394 mM, KCl: 2 mM, CaCl_2 : 2 mM, NaHCO_3 : 2 mM). Androgenic glands were collected and stored in Ringer solution. Then, they were injected through a hole pierced in the tergite of the 6th thoracic segment of juvenile females. From 1 to 3 androgenic glands were implanted in each developing female. Injected genetic females were between the 4th and 6th molt stage after birth, which corresponds to the molt stages during which sex differentiation takes place in *A. vulgare*³⁸. These young females were selected for injections by comparison with male siblings of the same size (between 3 and 5 mm in length depending on the species) which were at the beginning of sexual differentiation. Early differentiating males can morphologically be distinguished from early differentiating females by the development of two ventral copulating pleopods.

Genetic crosses and statistical analyses of F1 progenies. After full sexual differentiation (20–28 weeks post-injection), neomales were crossed with their sisters (genetic females) to produce F1 progenies. Analyses of the sex-ratios of these progenies allowed us to deduce the type of heterogametic system in the various species. In the case of a male heterogametic system (XY/XX), neomales (i.e. sex-reversed genetic females) are XX and crossing them with their XX sisters is expected to produce progenies comprising 100% XX females (Fig. 1). In the case of a female heterogametic system (ZZ/ZW), neomales are ZW and crossing them with their ZW sisters is expected to produce progenies comprising 25% ZZ males, 25% WW females and 50% ZW females. Thus, we predict a proportion of 3/4 females if the WW genotype is viable, and 2/3 females if the WW genotype is not viable (Fig. 1).

To rule out any confounding effect due to feminizing *Wolbachia* infection (which would also result in female-biased sex ratios), total DNA was extracted from the head of all mothers after they produced their progenies, using the Qiagen DNeasy Blood and Tissue kit according to the manufacturer's instructions. Success of DNA extractions was checked by PCR amplification of the mitochondrial cytochrome oxidase I marker¹¹¹. Next, absence of *Wolbachia* infection was checked by PCR using *Wolbachia*-specific primers for the *wsp* and *ftsZ* markers^{112, 113}.

Sex ratios of F1 progenies were analyzed using χ^2 tests. First, we assessed whether sex ratios were statistically different from a balanced sex ratio (1:1). If not, we concluded that the progeny originated from a cross between a genetic male (instead of a neomale) and a genetic female, implying that androgenic glands had been injected into an undifferentiated genetic male instead of a female during the sex reversal experiment. We also performed χ^2 tests to assess whether observed sex ratios were significantly different from 1/3♂:2/3♀, in which case the WW genotype would be deemed viable, or from 1/4♂:3/4♀, in which case the WW genotype would be deemed inviable. These statistical analyses were performed with the R software (version 3.1.1¹⁰⁶). The significance level was set at 0.05.

Identification of sex chromosome transitions. In the reconstruction of ancestral traits, we assumed all XY/XX chromosomes are homologous and all ZZ/ZW chromosomes are homologous. This constitutes a conservative hypothesis, limiting our estimates to a minimal number of transitions between heterogametic systems. To consider the possibility of a different and undetermined sex determination system in *A. assimile*, two kinds of discrete states were defined to conduct ancestral states analysis: (1) three discrete states were defined according to sex chromosome types: male heterogamety (XY/XX), female heterogamety (ZZ/ZW) and undetermined (in the case of *A. assimile*), and (2) only two discrete states were defined (XY/XX or ZZ/ZW) and *A. assimile* was treated as the other species for which no information on sex determination system is available. The evolution of character states was inferred using the Parsimony Ancestral States reconstruction of the MESQUITE software (version 3.04, build 725⁵¹). This reconstruction infers the most parsimonious (i.e. the lowest) number of possible transitions to explain the distribution of sex determination systems across the phylogeny. In addition, maximum likelihood-based ancestral state reconstructions analyses were performed using the R *Phytools* package¹¹⁴ to take into account possibly unobserved transitions between heterogametic systems. In both defined conditions of discrete states (1) and (2), equal probability between sex determination systems (XY/XX, ZZ/ZW and *A. assimile* system in (1)) was set for species with unknown sex determination mechanisms. Outgroup species were excluded from the reconstructions. Analyses were performed using the one-parameter equal rates model ("ER" model) and the symmetric model in which forward and reverse transitions between states are constrained to be equal ("SYM" model). All maximum likelihood simulations were conducted with 1,000 stochastic mappings.

References

- Wilkins, A. S. Moving up the hierarchy: a hypothesis on the evolution of a genetic sex determination pathway. *Bioessays* **17**, 71–77, doi:10.1002/bies.950170113 (1995).
- Bull, J. J. *Evolution of sex determining mechanisms*. (Benjamin/Cummings Pub. Co., 1983).
- Bachtrog, D. Y-chromosome evolution: emerging insights into processes of Y-chromosome degeneration. *Nat. Rev. Genet.* **14**, 113–124, doi:10.1038/nrg3366 (2013).
- de Bono, M. & Hodgkin, J. Evolution of sex determination in *Caenorhabditis*: unusually high divergence of *tra-1* and its functional consequences. *Genetics* **144**, 587–595 (1996).
- Lundrigan, B. L. & Tucker, P. K. Evidence for multiple functional copies of the male sex-determining locus, *Sry*, in African murine rodents. *J. Mol. Evol.* **45**, 60–65, doi:10.1007/PL00006202 (1997).
- Gamble, T. & Zarkower, D. Sex determination. *Curr. Biol.* **22**, R257–262, doi:10.1016/j.cub.2012.02.054 (2012).
- Marshall Graves, J. A. Weird animal genomes and the evolution of vertebrate sex and sex chromosomes. *Annu. Rev. Genet.* **42**, 565–586, doi:10.1146/annurev.genet.42.110807.091714 (2008).
- Sarre, S. D., Ezaz, T. & Georges, A. Transitions between sex-determining systems in reptiles and amphibians. *Annu Rev Genomics Hum Genet* **12**, 391–406, doi:10.1146/annurev-genom-082410-101518 (2011).
- Vicoso, B. & Bachtrog, D. Numerous Transitions of Sex Chromosomes in Diptera. *PLOS Biol* **13**, e1002078, doi:10.1371/journal.pbio.1002078 (2015).
- van Doorn, G. S. & Kirkpatrick, M. Turnover of sex chromosomes induced by sexual conflict. *Nature* **449**, 909–912, doi:10.1038/nature06178 (2007).
- Blaser, O., Grossen, C., Neuenschwander, S. & Perrin, N. Sex-Chromosome Turnovers Induced by Deleterious Mutation Load. *Evolution* **67**, 635–645, doi:10.1111/j.1558-5646.2012.01810.x (2013).
- Blaser, O., Neuenschwander, S. & Perrin, N. Sex-chromosome turnovers: the hot-potato model. *Am. Nat.* **183**, 140–146, doi:10.1086/674026 (2014).
- Yoshida, K. & Kitano, J. The Contribution of Female Meiotic Drive to the Evolution of Neo-Sex Chromosomes. *Evolution* **66**, 3198–3208, doi:10.1111/j.1558-5646.2012.01681.x (2012).
- Charlesworth, B. & Wall, J. D. Inbreeding, heterozygote advantage and the evolution of neo-X and neo-Y sex chromosomes. *Proceedings of the Royal Society of London B: Biological Sciences* **266**, 51–56, doi:10.1098/rspb.1999.0603 (1999).
- Werren, J. H. & Beukeboom, L. W. Sex Determination, Sex Ratios, and Genetic Conflict. *Annual Review of Ecology and Systematics* **29**, 233–261, doi:10.1146/annurev.ecolsys.29.1.233 (1998).
- Mank, J. E., Hosken, D. J. & Wedell, N. Conflict on the Sex Chromosomes: Cause, Effect, and Complexity. *Cold Spring Harb Perspect Biol* **6**, a017715–a017715, doi:10.1101/cshperspect.a017715 (2014).
- Beukeboom, L. W. & Perrin, N. *The Evolution of Sex Determination*. (Oxford University Press, 2014).
- Werren, J. H., Baldo, L. & Clark, M. E. *Wolbachia*: master manipulators of invertebrate biology. *Nat Rev Micro* **6**, 741–751, doi:10.1038/nrmicro1969 (2008).
- Cordaux, R., Bouchon, D. & Grève, P. The impact of endosymbionts on the evolution of host sex-determination mechanisms. *Trends Genet.* **27**, 332–341, doi:10.1016/j.tig.2011.05.002 (2011).
- Rigaud, T., Juchault, P. & Mocquard, J.-P. The evolution of sex determination in isopod crustaceans. *Bioessays* **19**, 409–416, doi:10.1002/(ISSN)1521-1878 (1997).

21. Martin, G., Juchault, P. & Legrand, J. J. Mise en évidence d'un micro-organisme intracytoplasmique symbiote de l'oniscoïde *Armadillidium vulgare* Latr. dont la présence accompagne l'intersexualité ou la féminisation totale des mâles génétiques de la lignée thelygène. *C. R. Acad. Sci. Paris* **276**, 2213–2216 (1973).
22. Cordaux, R., Michel-Salz, A., Frelon-Raimond, M., Rigaud, T. & Bouchon, D. Evidence for a new feminizing *Wolbachia* strain in the isopod *Armadillidium vulgare*: evolutionary implications. *Heredity* **93**, 78–84, doi:10.1038/sj.hdy.6800482 (2004).
23. Rigaud, T. & Juchault, P. Conflict between feminizing sex ratio distorters and an autosomal masculinizing gene in the terrestrial isopod *Armadillidium vulgare* Latr. *Genetics* **133**, 247–252 (1993).
24. Caubet, H. & Mocquard, R. Genetic conflict and changes in heterogametic mechanisms of sex determination. *Journal of Evolutionary Biology* **13**, 766–777, doi:10.1046/j.1420-9101.2000.00225.x (2000).
25. Leclercq, S. *et al.* Birth of a W sex chromosome by horizontal transfer of *Wolbachia* bacterial symbiont genome. *PNAS* **113**, 201608979, doi:10.1073/pnas.15036-15041 (2016).
26. Juchault, P. & Mocquard, J. P. Transfer of a parasitic sex factor to the nuclear genome of the host: A hypothesis on the evolution of sex-determining mechanisms in the terrestrial isopod *Armadillidium vulgare* Latr. *Journal of Evolutionary Biology* **6**, 511–528, doi:10.1046/j.1420-9101.1993.6040511.x (1993).
27. Bouchon, D., Rigaud, T. & Juchault, P. Evidence for widespread *Wolbachia* infection in isopod crustaceans: molecular identification and host feminization. *Proceedings of the Royal Society of London B: Biological Sciences* **265**, 1081–1090, doi:10.1098/rspb.1998.0402 (1998).
28. Cordaux, R. *et al.* Widespread *Wolbachia* infection in terrestrial isopods and other crustaceans. *Zookeys* **176**, 123–131, doi:10.3897/zookeys.176.2284 (2012).
29. Michel-Salz, A. & Bouchon, D. Phylogenetic analysis of mitochondrial LSU rRNA in oniscids, *C. R. Acad. Sci. III, Sci. Vie* **323**, 827–837, doi:10.1016/S0764-4469(00)01221-X (2000).
30. Mattern, D. & Schlegel, M. Molecular evolution of the small subunit ribosomal DNA in woodlice (Crustacea, Isopoda, Oniscidea) and implications for Oniscidea phylogeny. *Mol. Phylogenet. Evol.* **18**, 54–65, doi:10.1006/mpev.2000.0861 (2001).
31. Vicoso, B. & Bachtrog, D. Lack of Global Dosage Compensation in *Schistosoma mansoni*, a Female-Heterogametic Parasite. *Genome Biol. Evol.* **3**, 230–235, doi:10.1093/gbe/evr010 (2011).
32. Hall, A. B. *et al.* Six novel Y chromosome genes in *Anopheles* mosquitoes discovered by independently sequencing males and females. *BMC Genomics* **14**, 273, doi:10.1186/1471-2164-14-273 (2013).
33. Vicoso, B., Emerson, J. J., Zektser, Y., Mahajan, S. & Bachtrog, D. Comparative Sex Chromosome Genomics in Snakes: Differentiation, Evolutionary Strata, and Lack of Global Dosage Compensation. *PLOS Biol* **11**, e1001643, doi:10.1371/journal.pbio.1001643 (2013).
34. Cortez, D. *et al.* Origins and functional evolution of Y chromosomes across mammals. *Nature* **508**, 488–493, doi:10.1038/nature13151 (2014).
35. Akagi, T., Henry, I. M., Tao, R. & Comai, L. Plant genetics. A Y-chromosome-encoded small RNA acts as a sex determinant in persimmons. *Science* **346**, 646–650, doi:10.1126/science.1257225 (2014).
36. Jeffery, N. W. & Gregory, T. R. Genome size estimates for crustaceans using Feulgen image analysis densitometry of ethanol-preserved tissues. *Cytometry* **85**, 862–868, doi:10.1002/cyto.a.v85.10 (2014).
37. Juchault, P. & Rigaud, T. Evidence for female heterogamety in two terrestrial crustaceans and the problem of sex chromosome evolution in isopods. *Heredity* **75**, 466–471, doi:10.1038/hdy.1995.163 (1995).
38. Suzuki, S. & Yamasaki, K. Morphological studies on sexual differentiation in *Armadillidium vulgare* (Isopoda: Armadillidae): androgenic gland and male sexual characters. *Crustacean Research. December* Available at: <https://eurekamag.com/research/021/372/021372660.php> (Accessed: 18th October 2016) (1995).
39. Badawi, M., Grève, P. & Cordaux, R. Feminization of the isopod *Cylisticus convexus* after transinfection of the wVulC *Wolbachia* strain of *Armadillidium vulgare*. *PLoS One* **10**, e0128660, doi:10.1371/journal.pone.0128660 (2015).
40. Katakura, Y. Sex differentiation and androgenic gland hormone in the terrestrial isopod *Armadillidium vulgare*. in *Symposia of the Zoological Society of London* 127–142 (Cambridge University Press, 1984).
41. Legrand, J.-J., Legrand-Hamelin, E. & Juchault, P. Sex determination in Crustacea. *Biological Reviews* **62**, 439–447, doi:10.1111/brv.1987.62.issue-4 (1987).
42. Juchault, P. & Legrand, J.-J. Démonstration de l'homogamétie femelle par croisement de deux femelles génétiques, chez les Oniscoïdes *Porcellio dilatatus* et *Helleria brevicornis*. *Comptes Rendus de l'Académie des Sciences de Paris* **258**, 2685–2686 (1964).
43. Legrand, J. J., Juchault, P., Artault, J. C., Mocquard, J. P. & Picaud, J. L. Le statut systématique de la «forme» petit Vandel de *Porcellio dilatatus* Brandt, récoltée à l'île Saint-Honorat (Alpes-Maritimes). Critères morphologiques, génétiques et physiologiques. *Bull. Soc. Zool. Fr* **99**, 461–471 (1974).
44. Labandeira, C. C. & Sepkoski, J. J. Insect diversity in the fossil record. *Science* **261**, 310–315, doi:10.1126/science.11536548 (1993).
45. Labandeira, C. C. Insect mouthparts: Ascertaining the Paleobiology of Insect Feeding Strategies. *Annual Review of Ecology and Systematics* **28**, 153–193, doi:10.1146/annurev.ecolsys.28.1.153 (1997).
46. Juchault, P. & Legrand, J.-J. Analyse génétique et physiologique de la détermination du sexe dans une population du Crustacé isopode Oniscoïde *Armadillidium nasatum*. *Arch. Zool. exp. gén* **120**, 25–43 (1979).
47. Juchault, P. & Legrand, J.-J. Croisements de néo-mâles expérimentaux chez *Armadillidium vulgare* Latr. (Crustacé isopode Oniscoïde). Mise en évidence d'une hétérogamétie femelle. *C. R. Acad. Sc. Paris* **1387**–**1389**, 274–276 (1972).
48. Mittal, O. P. & Pahwa, S. Germ cell chromosomes in two species of terrestrial isopods from India. *Experientia* **37**, 343–345, doi:10.1007/BF01959853 (1981).
49. Mittal, O. P. & Pahwa, S. Cytogenetical studies on two terrestrial isopods (Crustacea), *Philoscia lodnensis* Ramakrishna and *Porcellio rathkei* Brandt. *Cytobios* **29**, 17–24 (1980).
50. Rocchi, A., Pranter, G., Lanza, V. & Castro, M. D. Incipient sex chromosome differentiation in an isopod crustacean species, *Asellus aquaticus*. *Chromosoma* **89**, 193–196, doi:10.1007/BF00294998 (1984).
51. Maddison, W. P. & Maddison, D. R. Mesquite: a modular system for evolutionary analysis. (2015).
52. Warburg, M. R. *Evolutionary Biology of Land Isopods*. (Springer Berlin Heidelberg, 1993).
53. Gentile, G., Campanaro, A., Carosi, M., Sbordoni, V. & Argano, R. Phylogeography of *Helleria brevicornis* Ebner 1868 (Crustacea, Oniscidea): Old and recent differentiations of an ancient lineage. *Molecular Phylogenetics and Evolution* **54**, 640–646, doi:10.1016/j.ympev.2009.10.005 (2010).
54. Vandel, A. Faune de France. Isopodes Terrestres (Première Partie). *Lechevalier, Paris* (1960).
55. Sfenthourakis, S. & Taiti, S. Patterns of taxonomic diversity among terrestrial isopods. *Zookeys* **515**, 13–25, doi:10.3897/zookeys.515.9332 (2015).
56. Schmalfuss, H. D. L.-I. (Oniscidea) Griechenlands. 12. Beitrag: Gattungen *Lepidoniscus* und *Philoscia* (Philosciidae). *Stuttgarter Beiträge zur Naturkunde Serie A (Biologie)* **448**, 1–11 (1990).
57. Hild, S., Marti, O. & Ziegler, A. Spatial distribution of calcite and amorphous calcium carbonate in the cuticle of the terrestrial crustaceans *Porcellio scaber* and *Armadillidium vulgare*. *J. Struct. Biol.* **163**, 100–108, doi:10.1016/j.jsb.2008.04.010 (2008).
58. Lins, L. S. F., Ho, S. Y. W., Wilson, G. D. F. & Lo, N. Evidence for Permo-Triassic colonization of the deep sea by isopods. *Biol. Lett.* **8**, 979–982, doi:10.1098/rsbl.2012.0774 (2012).
59. Willett, C. S. Quantifying the elevation of mitochondrial DNA evolutionary substitution rates over nuclear rates in the intertidal copepod *Tigriopus californicus*. *J. Mol. Evol.* **74**, 310–318, doi:10.1007/s00239-012-9508-1 (2012).

60. Ketmaier, V., Argano, R. & Caccone, A. Phylogeography and molecular rates of subterranean aquatic Stenaseiid Isopods with a peri-Tyrrhenian distribution. *Molecular Ecology* **12**, 547–555, doi:10.1046/j.1365-294X.2003.01734.x (2003).
61. Knowlton, N., Weigt, L. A., Solórzano, L. A., Mills, D. K. & Bermingham, E. Divergence in proteins, mitochondrial DNA, and reproductive compatibility across the isthmus of Panama. *Science* **260**, 1629–1632, doi:10.1126/science.8503007 (1993).
62. Lynch, M., Koskella, B. & Schaack, S. Mutation pressure and the evolution of organelle genomic architecture. *Science* **311**, 1727–1730, doi:10.1126/science.1118884 (2006).
63. Lynch, M. & Walsh, B. *The origins of genome architecture*. **98** (Sinauer Associates Sunderland, 2007).
64. Hellberg, M. E. No variation and low synonymous substitution rates in coral mtDNA despite high nuclear variation. *BMC Evolutionary Biology* **6**, 24, doi:10.1186/1471-2148-6-24 (2006).
65. Clark-Walker, G. D. Contrasting mutation rates in mitochondrial and nuclear genes of yeasts versus mammals. *Curr. Genet.* **20**, 195–198, doi:10.1007/BF00326232 (1991).
66. Schröder, H. C. *et al.* Molecular phylogeny of the freshwater sponges in Lake Baikal. *Journal of Zoological Systematics and Evolutionary Research* **41**, 80–86, doi:10.1046/j.1439-0469.2003.00199.x (2003).
67. Chandler, C. H., Badawi, M., Moumen, B., Grève, P. & Cordaux, R. Multiple Conserved Heteroplasmic Sites in tRNA Genes in the Mitochondrial of Terrestrial Isopods (Oniscidea). *G3* **5**, 1317–1322, doi:10.1534/g3.115.018283 (2015).
68. Marcadé, I. *et al.* Structure and Evolution of the Atypical Mitochondrial Genome of Armadillidium vulgare (Isopoda, Crustacea). *J Mol Evol* **65**, 651–659, doi:10.1007/s00239-007-9037-5 (2007).
69. Roco, Á. S. *et al.* Coexistence of Y, W, and Z sex chromosomes in *Xenopus tropicalis*. *PNAS* **112**, E4752–E4761, doi:10.1073/pnas.1505291112 (2015).
70. Volff, J.-N. & Schartl, M. Sex determination and sex chromosome evolution in the medaka, *Oryzias latipes*, and the platyfish, *Xiphophorus maculatus*. *Cytogenet. Genome Res.* **99**, 170–177 (2002).
71. Stuart, J. J. & Hatchett, J. H. Genetics of Sex Determination in the Hessian Fly, *Mayetiola destructor*. *J Hered* **82**, 43–52, doi:10.1093/jhered/82.1.43 (1991).
72. Legrand, J.-J. & Juchault, P. Mise en évidence d'un hermaphrodisme protogynique fonctionnel chez l'isopode Anthuridé *Cyathura carinata* (Krøyer) et étude du mécanisme de l'inversion sexuelle. *C. R. Acad. Sc. Paris* **256**, 2931–2933 (1963).
73. Brook, H. J., Rawlings, T. A. & Davies, R. W. Protogynous Sex Change in the Intertidal Isopod *Gnorimosphaeroma oregonense* (Crustacea: Isopoda). *The Biological Bulletin* **187**, 99–111, doi:10.2307/1542169 (1994).
74. Legrand, J. J. & Juchault, P. Contrôle de la sexualité chez les crustacés isopodes gonochoriques et hermaphrodites. *Bull. Soc. Zool. Fr.* **95**, 551–563 (1970).
75. Christensen, B. Genetic variation in coexisting sexual diploid and parthenogenetic triploid *Trichoniscus pusillus* (Isopoda, Crustacea). *Hereditas* **98**, 201–207, doi:10.1111/j.1601-5223.1983.tb00594.x (1983).
76. Rigaud, T. In Influential passengers: inherited microorganisms and arthropod reproduction 81–101 (O'Neill S. L., Hoffmann A. A., Werren J. H., 1997).
77. Bachtrog, D. *et al.* Sex determination: why so many ways of doing it? *PLoS Biol.* **12**, e1001899, doi:10.1371/journal.pbio.1001899 (2014).
78. Blackmon, H., Ross, L. & Bachtrog, D. Sex Determination, Sex Chromosomes, and Karyotype Evolution in Insects. *J Hered* **108**, 78–93, doi:10.1093/jhered/esw047 (2017).
79. Coyne, J. A. & Orr, H. A. *Speciation*. (Sinauer Associates, 2004).
80. Presgraves, D. C. Sex chromosomes and speciation in *Drosophila*. *Trends Genet.* **24**, 336–343, doi:10.1016/j.tig.2008.04.007 (2008).
81. Lindholm, A. & Breden, F. Sex chromosomes and sexual selection in poeciliid fishes. *Am. Nat.* **160** Suppl 6, S214–224, doi:10.1086/342898 (2002).
82. Saether, S. A. *et al.* Sex chromosome-linked species recognition and evolution of reproductive isolation in flycatchers. *Science* **318**, 95–97, doi:10.1126/science.1141506 (2007).
83. Kitano, J. *et al.* A role for a neo-sex chromosome in stickleback speciation. *Nature* **461**, 1079–1083, doi:10.1038/nature08441 (2009).
84. Pryke, S. R. Sex chromosome linkage of mate preference and color signal maintains assortative mating between interbreeding finch morphs. *Evolution* **64**, 1301–1310, doi:10.1111/j.1558-5646.2009.00897.x (2010).
85. Schmidt, C. Phylogeny of the terrestrial Isopoda (Oniscidea): a review. *Arthropod Systematics & Phylogeny* **66**, 191–226 (2008).
86. Romiguier, J. *et al.* Comparative population genomics in animals uncovers the determinants of genetic diversity. *Nature* **515**, 261–263, doi:10.1038/nature13685 (2014).
87. Speiser, D. I. *et al.* Using phylogenetically-informed annotation (PIA) to search for light-interacting genes in transcriptomes from non-model organisms. *BMC Bioinformatics* **15**, 350, doi:10.1186/s12859-014-0350-x (2014).
88. Tan, M. H. *et al.* First comprehensive multi-tissue transcriptome of *Cherax quadricarinatus* (Decapoda: Parastacidae) reveals unexpected diversity of endogenous cellulase. *Org Divers Evol* **16**, 185–200, doi:10.1007/s13127-015-0237-3 (2016).
89. O'Grady, J. F., Hoelters, L. S., Swain, M. T. & Wilcockson, D. C. Identification and temporal expression of putative circadian clock transcripts in the amphipod crustacean *Talitrus saltator*. *PeerJ* **4**, e2555, doi:10.7717/peerj.2555 (2016).
90. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120, doi:10.1093/bioinformatics/btu170 (2014).
91. Haas, B. J. *et al.* De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat Protoc* **8**, 1494–1512, doi:10.1038/nprot.2013.084 (2013).
92. Wernersson, R. Virtual Ribosome—a comprehensive DNA translation tool with support for integration of sequence feature annotation. *Nucleic Acids Res.* **34**, W385–388, doi:10.1093/nar/gkl252 (2006).
93. Li, L., Stoeckert, C. J. & Roos, D. S. OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res.* **13**, 2178–2189, doi:10.1101/gr.1224503 (2003).
94. Madden, T. *The BLAST Sequence Analysis Tool*. (National Center for Biotechnology Information (US), 2003).
95. Kearse, M. *et al.* Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28**, 1647–1649, doi:10.1093/bioinformatics/bts199 (2012).
96. Talavera, G. & Castresana, J. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Syst. Biol.* **56**, 564–577, doi:10.1080/10635150701472164 (2007).
97. Vaidya, G., Lohman, D. J. & Meier, R. SequenceMatrix: concatenation software for the fast assembly of multi-gene datasets with character set and codon information. *Cladistics* **27**, 171–180, doi:10.1111/cla.2011.27.issue-2 (2011).
98. Darriba, D., Taboada, G. L., Doallo, R. & Posada, D. jModelTest 2: more models, new heuristics and parallel computing. *Nat Meth* **9**, 772–772, doi:10.1038/nmeth.2109 (2012).
99. Stamatakis, A. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22**, 2688–2690, doi:10.1093/bioinformatics/btl446 (2006).
100. Drummond, A. J., Suchard, M. A., Xie, D. & Rambaut, A. Bayesian Phylogenetics with BEAUti and the BEAST 1.7. *Mol Biol Evol* **29**, 1969–1973, doi:10.1093/molbev/mss075 (2012).
101. Ferreira, M. A. R. & Suchard, M. A. Bayesian analysis of elapsed times in continuous-time Markov chains. *Can J Statistics* **36**, 355–368, doi:10.1002/cjs.v36.3 (2008).
102. Spahr, U. Ergänzungen und Berichtigungen zu R. Keilbachs Bibliographie und Liste der Bernstein fossilen—Verschiedene Tiergruppen, ausgenommen Insecta und Araneae. *Stutt Beitr Naturk Ser B* **194**, 1–77 (1993).

103. Weitschat, W. & Wichard, W. In *Biodiversity of Fossils in Amber from the Major World Deposits* 80–115 (Siri Scientific Press, 2010).
104. Alonso, J. *et al.* A new fossil resin with biological inclusions in lower cretaceous deposits from Álava (northern Spain, Basque-Cantabrian basin). *Journal of Paleontology* **74**, 158–178, doi:10.1017/S0022336000031334 (2000).
105. Delclòs, X. *et al.* Fossiliferous amber deposits from the Cretaceous (Albian) of Spain. *Comptes Rendus Palevol* **6**, 135–149, doi:10.1016/j.crpv.2006.09.003 (2007).
106. R Core Team. *R: A language and environment for statistical computing*. (R Foundation for Statistical Computing, 2014).
107. Bell, M. A. & Lloyd, G. T. strap: an R package for plotting phylogenies against stratigraphy and assessing their stratigraphic congruence. *Palaeontology* **58**, 379–389, doi:10.1111/pala.12142 (2015).
108. Heibl, C. PHYLOCH: R language tree plotting tools and interfaces to diverse phylogenetic software packages. (2008).
109. Yang, Z. PAML 4: Phylogenetic Analysis by Maximum Likelihood. *Mol Biol Evol* **24**, 1586–1591, doi:10.1093/molbev/mst179 (2007).
110. Xu, B. & Yang, Z. pamlX: A Graphical User Interface for PAML. *Mol Biol Evol* **30**, 2723–2724, doi:10.1093/molbev/mst179 (2013).
111. Folmer, O., Black, M., Hoeh, W., Lutz, R. & Vrijenhoek, R. DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Mol. Marine Biol. Biotechnol.* **3**, 294–299 (1994).
112. Werren, J. H., Zhang, W. & Guo, L. R. Evolution and phylogeny of Wolbachia: reproductive parasites of arthropods. *Proc. Biol. Sci.* **261**, 55–63, doi:10.1098/rspb.1995.0117 (1995).
113. Braig, H. R., Zhou, W., Dobson, S. L. & O'Neill, S. L. Cloning and characterization of a gene encoding the major surface protein of the bacterial endosymbiont Wolbachia pipientis. *J. Bacteriol.* **180**, 2373–2378 (1998).
114. Revell, L. J. phytools: an R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution* **3**, 217–223, doi:10.1111/j.2041-210X.2011.00169.x (2012).
115. Bouchon, D., Cordaux, R. & Grève, P. In *Insect Symbiosis, Volume 3*, 273–294 (CRC Press, 2008).
116. Sicard, M. *et al.* Bidirectional cytoplasmic incompatibility caused by Wolbachia in the terrestrial isopod Porcellio dilatatus. *Journal of Invertebrate Pathology* **121**, 28–36, doi:10.1016/j.jip.2014.06.007 (2014).
117. Michel-Salzat, A., Cordaux, R. & Bouchon, D. Wolbachia diversity in the Porcellionides pruinosus complex of species (Crustacea: Oniscidea): evidence for host-dependent patterns of infection. *Heredity* **87**, 428–434, doi:10.1046/j.1365-2540.2001.00920.x (2001).
118. Juchault, P., Frelon, M., Bouchon, D. & Rigaud, T. New evidence for feminizing bacteria in terrestrial isopods: evolutionary implications. *C. R. Acad. Sci. III, Sci. Vie* **317**, 225–230 (1994).

Acknowledgements

We thank Alexandra Lafitte for assistance with species morphological identification and animal rearing, and Daniel Guyonnet for technical assistance on Sanger sequencing. This work was funded by Agence Nationale de la Recherche Grant ANR-15-CE32-0006-01 (CytoSexDet) to R.C., the 2015–2020 State-Region Planning Contract and European Regional Development Fund, and intramural funds from the Centre National de la Recherche Scientifique and the University of Poitiers. C.C. and the *T. rathkei* transcriptome data were funded by NSF DEB-1453298. T.B. was supported by a PhD fellowship from the French Ministère de l'Éducation Nationale, de l'Enseignement Supérieur et de la Recherche.

Author Contributions

T.B., R.C. and C.G. conceived and designed the experiments. T.B., M.R. and I.G. performed the experiments. T.B., C.C., R.C. and C.G. analyzed the data. B.M. provided bioinformatics support. C.C., R.C. and C.G. contributed reagents/materials. T.B., R.C. and C.G. wrote the manuscript. All authors read and approved the manuscript.

Additional Information

Supplementary information accompanies this paper at doi:10.1038/s41598-017-01195-4

Competing Interests: The authors declare that they have no competing interests.

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



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2017

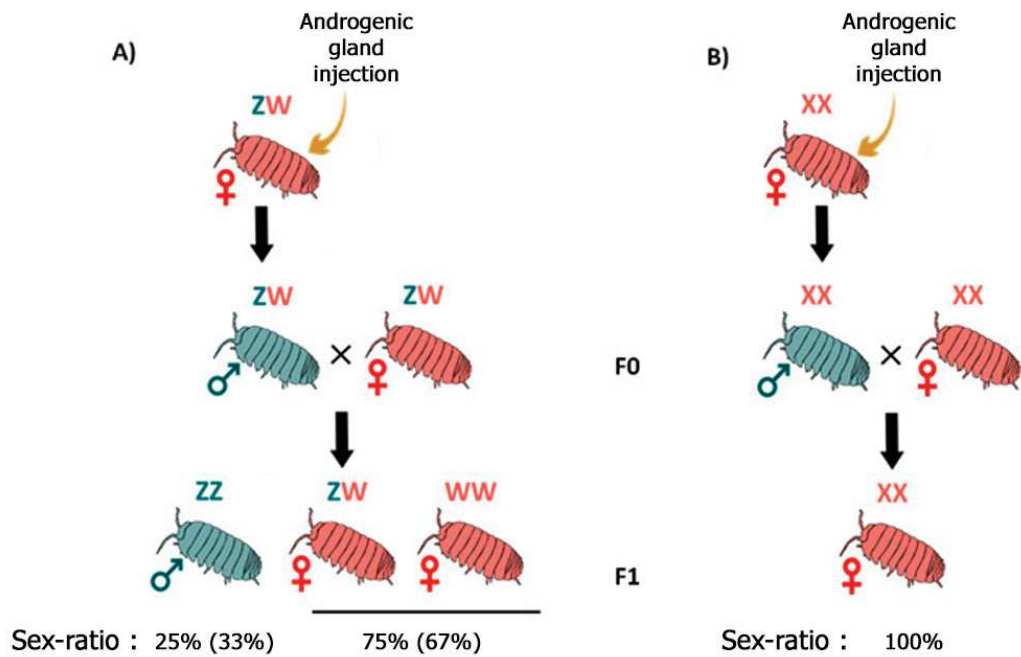


Figure 62 : Principle of crossing experiments between neo-males (i.e. sex-reversed females) and genetic females. A) Female heterogametic system (ZZ/ZW). The expected sex-ratio of the progeny of a neo-male and a genetic female is 25% males (ZZ) and 75% females (ZW and WW) if the WW genotype is viable, or 33% males (ZZ) and 67% females (ZW only). B) Male heterogametic system (XX/XY). The expected sex-ratio of the progeny of a neo-male and a genetic female is 100% females (XX).

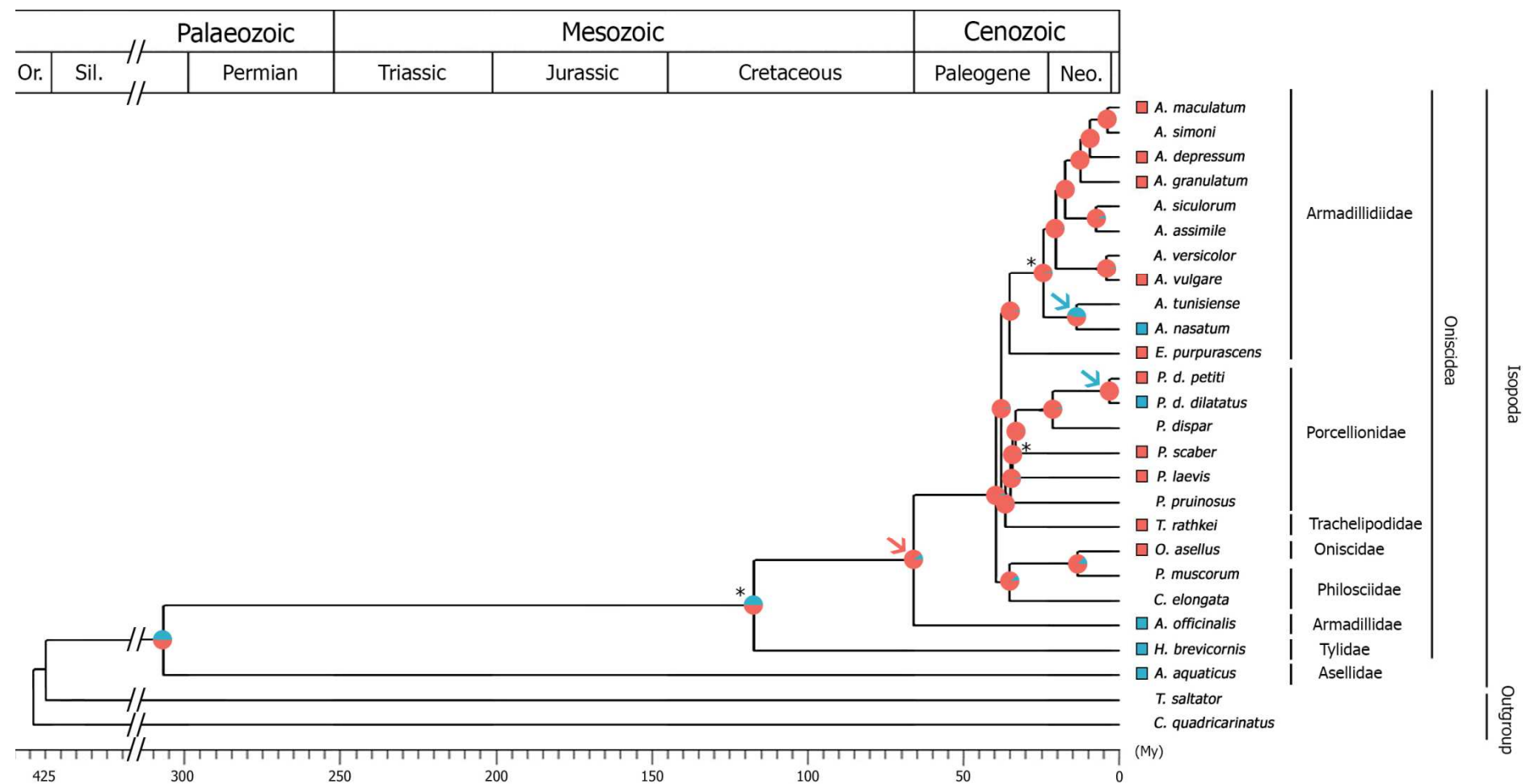


Figure 63 : Ancestral state reconstruction of sex determination systems in terrestrial isopod crustaceans, based on a Bayesian phylogenetic analysis using 88 molecular markers. The reconstruction is based on the one-parameter equal rates model ("ER" model), considering the *A. assimile* sex determination system as unknown. Pie charts on nodes show the probability of the ancestral state at each node, calculated from 1,000 stochastic mappings. Branch length is scaled to time, with nodal ages corresponding to median posterior estimates. * denote the calibration points. Red and blue squares at branch tips indicate a ZZ/ZW and XY/XX sex determination systems, respectively. Transitions of sex determination systems determined by parsimony ancestral state reconstruction are outlined with arrows (one red arrow for a transition from XY/XX system to ZZ/ZW, and two blue arrows for transitions from ZZ/ZW to XX/XY systems). Isopod family names and higher taxonomic ranks are shown on the right.



Figure 64 : Sex ratio of progenies of 13 terrestrial isopod species resulting from crosses between genetic and sex-reversed females (neomales). Red and blue colors indicate the percentages of females and males, respectively. Red dashed lines indicate a 1/4 ♂: 3/4 ♀ sex ratio (expected in ZZ/ZW systems), blue dashed lines indicate a 0 ♂: 1 ♀ sex ratio (expected in XY/XX systems), and green dashed lines indicate a 1 ♂: 0 ♀ sex ratio. *n* is the number of pooled individuals per sex, and the number of corresponding progenies are specified in brackets for each species.

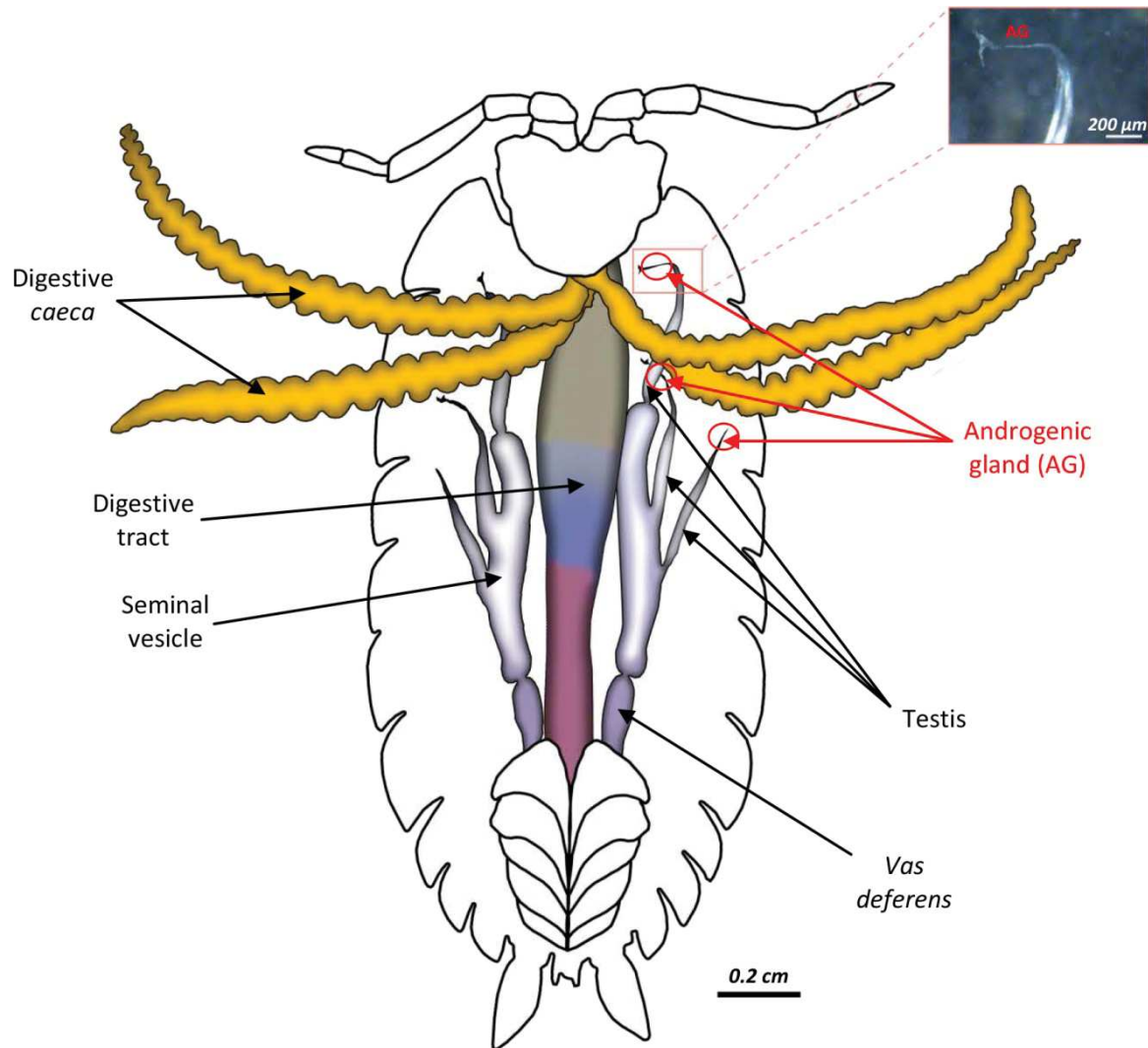


Figure 65 : Schematics of the internal organs of a terrestrial isopod (*Oniscus asellus*) in ventral view. Androgenic glands, composed of a cellular mass excreting androgenic hormone, are located at the top of each testis, included in the suspensory filament. Depending on the species, its size is 250-700 μm in length and 20-150 μm in width.

Perspectives

Cette étude a révélé de nombreuses transitions de systèmes d'hétérogamétie au cours de l'évolution des isopodes terrestres, soulignant la plasticité évolutive des mécanismes de déterminisme du sexe dans ce groupe d'animaux. Néanmoins, les résultats des expériences de réversion de sexe obtenus pour l'espèce *Armadillidium assimile* ne peuvent être expliqués par des mécanismes de déterminisme hétérochromosomiques classiques (voir Figure 3 dans Becking et al. 2017). Ainsi, ces résultats ouvrent la perspective qu'un autre mécanisme de déterminisme du sexe puisse être à l'œuvre dans cette espèce.

Très peu d'informations sont disponibles concernant cette espèce, si ce n'est qu'elle serait originaire du Sud de la France et que son aire de répartition s'étendrait peu à peu vers le Nord de la France (Vandel, 1962). Aucune information n'est cependant disponible concernant les sex-ratios naturels de *A. assimile*, et la présence de *Wolbachia* n'a pour le moment jamais été mise en évidence chez cette espèce (Becking et al., 2017). Concernant les 6 portées composées uniquement de mâles obtenues suite à des croisements entre des femelles génétiques et des femelles génétiques reversées expérimentalement en mâles phénotypiques (voir Figure 3 issue de Becking et al., 2017), des analyses moléculaires ont confirmé l'absence de *Wolbachia*. De plus, le séquençage du marqueur mitochondrial COI et du marqueur nucléaire ARNr 18S a également confirmé l'assignation taxonomique des 6 femelles ayant produit les portées, ainsi que leur fort degré d'apparentement (elles sont toutes soeurs), conformément au design expérimental décrit par Becking et al. (2017). Ces analyses complémentaires suggèrent que les résultats obtenus par Becking et al. (2017) ne sont pas dus à des erreurs de nature expérimentale en relation avec les expériences de réversion de sexe.

Afin d'en savoir plus sur le déterminisme du sexe dans cette espèce, nous avons d'abord obtenu des informations sur le sexe ratio naturel (croisements

n'impliquant pas d'individus sex-reversés) pouvant être rencontré dans les portées de cette espèce. Dans un premier temps, 5 femelles gravides ont été prélevées dans notre unique population de laboratoire d'*A. assimile* établie en 2014 à partir de 40 femelles porteuses d'embryons en développement (originaires de Saint-Maixent l'Ecole, 79) afin d'en analyser le sexe ratio. L'absence de *Wolbachia* a été confirmée par PCR pour les 5 femelles gravides prélevées. Les portées F1 issues de ces 5 femelles gravides ont présenté de fortes variations de sex-ratio, allant de 0% à 100% de mâles selon la femelle considérée (Tableau 25). Les sex-ratios des portées F1 issues de ces 5 femelles gravides nous ont permis de classer les différentes mères selon 3 catégories :

- Thélygène (T) : Une femelle est considérée comme thélygène si le sex-ratio de sa descendance est significativement biaisé en faveur des femelles (femelles I et III, Tableau 25).
- Amphogène (AM) : Une femelle est considérée comme amphogène si le sex-ratio de sa descendance n'est pas significativement différent d'un sex-ratio équilibré (femelles II et V, Tableau 25).
- Arrhénogène (AR) : Une femelle est considérée comme arrhénogène si le sex-ratio de sa descendance est significativement biaisé en faveur des mâles (femelle IV, Tableau 25).

Tableau 25 : Sex-ratio de portées F1 issues de 5 femelles gravides prélevées au sein de la population d'*A. assimile* maintenue au laboratoire. Les 5 mères ont été testées négatives à la présence de *Wolbachia*. Des tests de χ^2 ont été réalisés afin de tester si les sex-ratios des portées obtenues étaient différents d'un sex-ratio équilibré (le seuil de significativité a été fixé à 0.05).

Matricule de la portée F1	Taille de la portée	Proportion de mâles	Proportion de mâles $\neq$ 50% (<i>p-value</i>)	Type de femelle
I	24	0.46	0.68	Amphogène (AM)
II	23	0	1.62E-06	Thélygène (T)
III	23	0.57	0.53	Amphogène (AM)
IV	32	1	1.54E-08	Arrhénogène (AR)
V	24	0.04	7.10E-06	Thélygène (T)

Afin de tester si ces sex-ratios observés étaient dus à nos conditions d'élevage (population de laboratoire fondée en 2014 à partir de 40 femelles gravides), 32 femelles gravides issues de la même population naturelle (Saint

Maixent l'Ecole, 79) ont été échantillonnées sur le terrain en avril 2017 par Nicolas Bech et Frédéric Grandjean (UMR-CNRS 7267, Poitiers). Parmi ces 32 femelles gravides, 24 portées ont été obtenues, et 10 d'entre elles ont été sexées à ce jour (Tableau 26). L'interprétation des sex-ratios obtenus doit être faite avec précaution, car la présence de *Wolbachia* n'a pas encore été testée chez les mères de ces portées. Cependant, en l'état, ces résultats confirment le pattern observé à partir des femelles issues de notre population de laboratoire, puisque les portées présentent des variations de sex-ratio allant de 0% à 100% de mâles (Tableau 26). Même si le sex-ratio global de ces portées est à l'équilibre, les 3 catégories de femelles précédemment identifiées (T, AM et AR) sont bel et bien présentes dans la nature (Tableau 26). Les observations faites ne relèvent donc pas d'un biais lié aux conditions d'élevage en laboratoire.

Tableau 26 : Sex-ratio des 10 portées issues de femelles gravides échantillonnées dans la population naturelle. Des tests de χ^2 ont été réalisés afin de tester si les sex-ratios des portées obtenues étaient différents d'un sex-ratio équilibré (le seuil de significativité a été fixé à 0.05).

Identifiant de la portée	Taille de la portée	Proportion de mâles	Proportion de mâles $\neq$ 50% (<i>p-value</i>)	Type de portée
N1	20	0.75	0.02	AR
N2	39	0.82	6.25E-05	AR
N4	19	0	1.31E-05	T
N5	15	0.93	7.89E-04	AR
N7	15	0	1.08E-04	T
N8	17	0	3.74E-05	T
N9	11	0.18	0.035	T
N11	24	0.75	0.014	AR
N12	19	1	1.31E-05	AR
N13	18	0.67	0.16	AM
Global	197	0.57	0.06	-

Les sex-ratios observés parmi les 15 portées analysées (Tableaux 25 et 26) laissent apparaître un grand nombre de portées extrêmement biaisées en faveur des femelles (5 portées avec >95% de femelles) ou des mâles (3 portées avec >90% de mâles). Cela pourrait s'expliquer par la mort sélective des embryons mâles ou femelles dans les portées des femelles T et AR, respectivement. Dans ce cas, on prédit que la taille des portées des femelles T et AR devrait être réduite de moitié par rapport à celles des femelles AM. La comparaison des tailles des 15

portées des femelles AM, AR et T n'a cependant pas révélé de différences significatives (Test de Kruskal-wallis : $\chi = 2.27$, $p\text{-value} = 0.32$, Figure 66). Les différences de sex-ratios observées entre les portées d'*A. assimile* ne sont donc pas attribuables à une mort sélective des embryons d'un sexe ou l'autre.

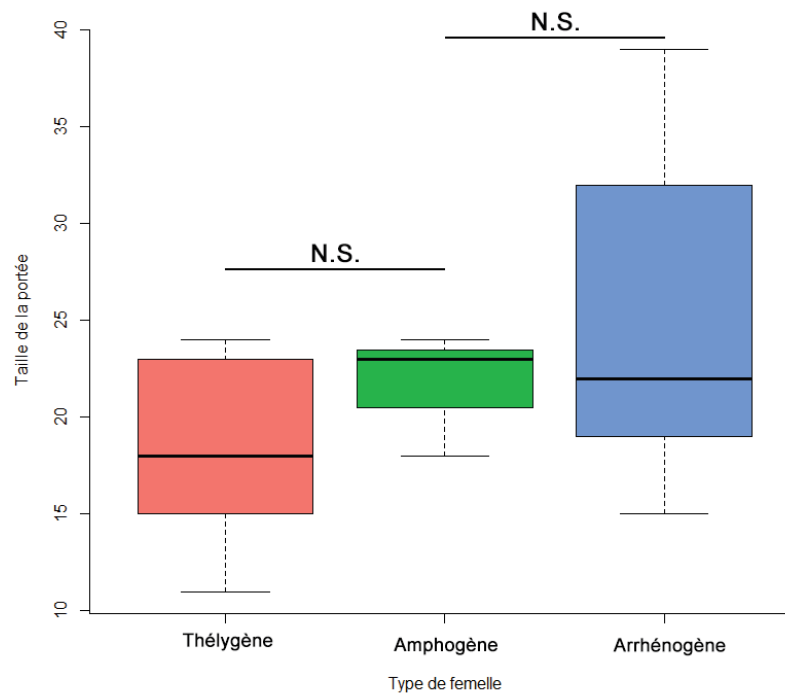


Figure 66 : Boxplot de la taille des portées issues des 3 différents types de femelles (Thélygène, Amphogène et Arrhénogène).

Afin de mieux décrire et comprendre les patterns d'héritabilité des sex-ratios observés au sein des portées d'*A. assimile*, les descendants F1 des 5 femelles issues de la population de laboratoire (Tableau 25), ainsi que les descendants F1 (tous mâles) issus des 6 portées des expériences de réversion de sexe (dénommées SR par la suite) décrites dans Becking et al. (2017) ont été utilisés pour réaliser 58 croisements contrôlés selon six combinaisons (Figure 67) :

- 10 croisements mâle AR x femelle AM : pour chaque croisement, un mâle issu d'une mère AR (femelle IV) a été croisé avec une femelle issue d'une mère AM (femelles I et III) (Figure 67, A).
- 10 croisements mâle AR x femelle T : pour chaque croisement, un mâle issu d'une mère AR (femelle IV) a été croisé avec une femelle issue d'une mère T (femelles II et V) (Figure 67, B).
- 7 croisements mâle AM x femelle AM : pour chaque croisement, un mâle issu d'une mère AM (femelles I et III) a été croisé avec une femelle issue d'une mère AM (femelles I et III) (Figure 67, C).
- 10 croisements mâle AM x femelle T : pour chaque croisement, un mâle issu d'une mère AM (femelles I et III) a été croisé avec une femelle issue d'une mère T (femelles II et V) (Figure 67, D).
- 9 croisements mâle SR x femelle AM : pour chaque croisement, un mâle issu d'une portée SR a été croisé avec une femelle issue d'une mère AM (femelles I et III) (Figure 67, E).
- 12 croisements mâle SR x femelle T : pour chaque croisement, un mâle issu d'une portée SR a été croisé avec une femelle issue d'une mère T (femelles II et V) (Figure 67, F).

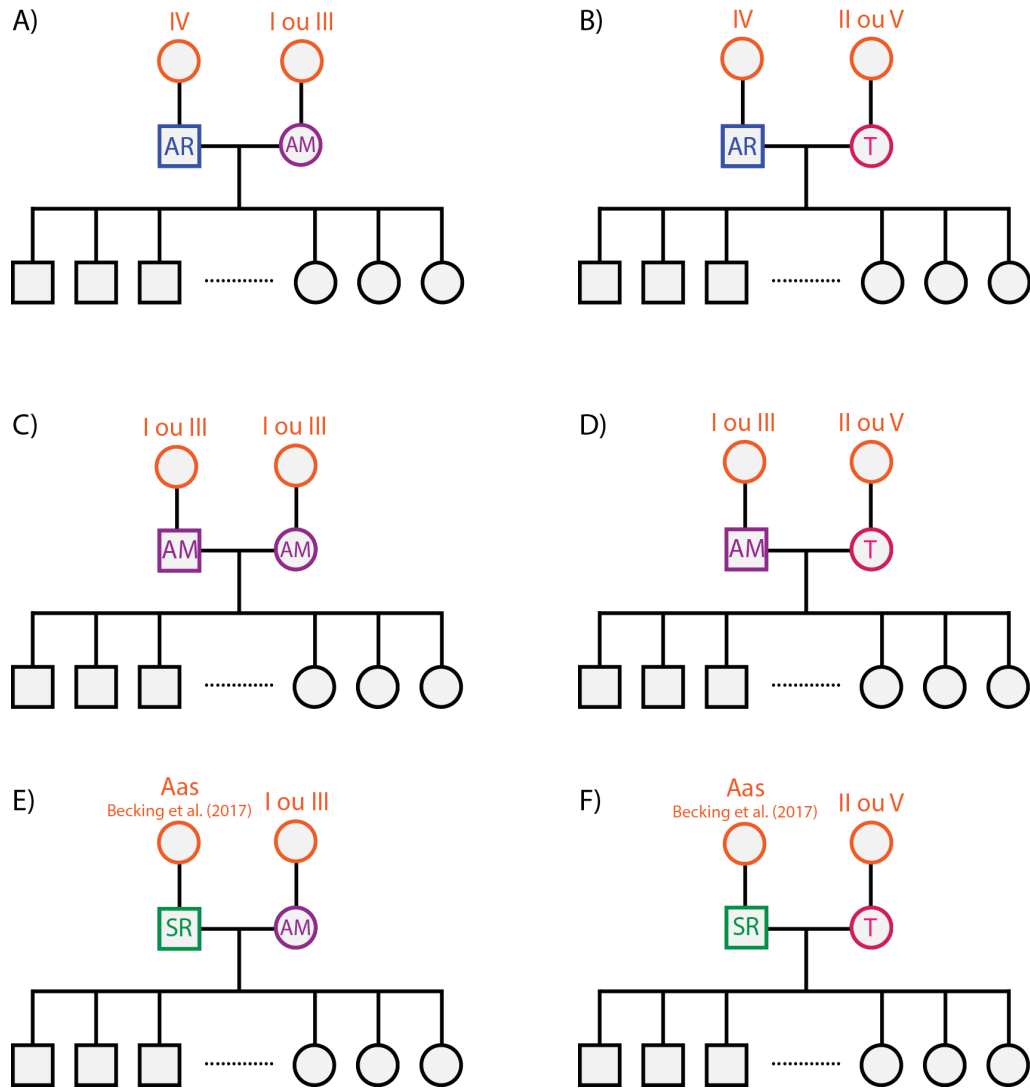


Figure 67 : Design des croisements réalisés à partir des descendants F1 des 5 femelles *A. assimile* prélevées issues de notre population de laboratoire. Les cercles correspondent aux femelles, tandis que les carrés correspondent aux mâles. Les femelles indiquées en orange correspondent à la génération F0. A) Croisements entre des mâles issus de la mère arrhénogène (IV, AR) et des femelles issues des mères amphogènes (I ou III, AM). B) Croisements entre des mâles issus de la mère arrhénogène (IV, AR) et des femelles issues des mères thélygènes (II ou V, T). C) Croisements entre des mâles issus des mères amphogènes (I ou III, AM) et des femelles issues des mères amphogènes (I ou III, AM). D) Croisements entre des mâles issus des mères amphogènes (AM) et des femelles issues des mères thélygènes (II ou V, T). E) Croisements entre des mâles issus des portées des expériences de réversion de sexe (Aas, SR) et des femelles issues des mères amphogènes (I ou III, AM). F) Croisements entre des mâles issus des portées des expériences de réversion de sexe (Aas, SR) et des femelles issues des mères thélygènes (II ou V, T).

Suite à ces croisements, 43 portées F2 comportant chacune au moins 6 individus ont été obtenues (Tableau 27) :

- 6 portées issues de mâle AR $\times$ femelle AM. Il en résulte que 2 portées (1 et 2) sont de type AR, 3 portées (3, 5 et 6) sont de type T, et 1 portée (4) est de type AM.
- 8 portées issues de mâle AR $\times$ femelle T. Il en résulte que 5 portées (9 à 13) sont de type AR et 3 portées (7, 8 et 14) sont de type T.
- 4 portées issues de mâle AM $\times$ femelle AM. Il en résulte que ces 5 portées sont de type AR (portées 15 à 18).
- 10 portées issues de mâle AM $\times$ femelle T. Il en résulte que 7 portées sont de type T (portées 19 et 21 à 26), 1 portée est de type AR (portée 27) et 2 portées sont de type AM (portées 20 et 28).
- 7 portées issues de mâle SR $\times$ femelle AM. Il en résulte que 4 portées sont de type AR (portées 29, 30, 32 et 34), 1 portée est de type T (portée 31) et 2 portées sont de type AM (portées 33 et 35).
- 8 portées issues de mâles SR $\times$ femelle T. Il en résulte que 6 de ces portées sont de type T (portées 36, 37, 39 à 41 et 43), 1 portée est de type AR (portée AR) et 1 portée est de type AM (portée 42).

Tableau 27: Bilan des croisements effectués à partir des descendants F1 des 5 femelles gravides issues de notre population de laboratoire et des mâles F1 issus des portées sexe-réversées (Becking et al., 2017). Les matricules indiqués correspondent aux différentes familles d'individus croisés. Des tests de χ^2 ont été réalisés afin de tester si les sex-ratios des portées F2 obtenues étaient différents d'un sex-ratio équilibré, afin de définir le type de portée (le seuil de significativité a été fixé à 0.05). Les *p-values* indiquées sous les proportions de mâles par type de croisement et sous la proportion globale de mâles sont issues de tests de χ^2 afin d'estimer statistiquement un biais de sex-ratio vis-à-vis d'un sex-ratio équilibré. Les croisements impliquant des mâles SR (29 à 43) ont été exclus de la proportion globale de mâles car les pères de portées SR ne peuvent pas exister naturellement (néo-mâles issus de Becking et al., 2017).

	Numéro de la portée F2	Matricule de la mère	Matricule du père	Nombre de descendants	Proportion de mâles	Type de portée	Proportion globale de mâles par type de croisement	Proportion globale de mâles
Mâle AR × Femelle AM	1	I	IV	39	1	AR	0.53 (0.37)	0.51 (0.47)
	2	III	IV	26	0.96	AR		
	3	III	IV	39	0.25	T		
	4	I	IV	23	0.57	AM		
	5	I	IV	22	0	T		
	6	III	IV	33	0.3	T		
Mâle AR × Femelle T	7	II	IV	25	0.04	T	0.58 (0.04)	
	8	II	IV	31	0	T		
	9	V	IV	23	1	AR		
	10	V	IV	34	1	AR		
	11	II	IV	6	1	AR		
	12	V	IV	10	1	AR		
	13	V	IV	14	1	AR		
	14	II	IV	23	0.35	T		
Mâle AM × Femelle AM	15	III	I	27	0.96	AR	0.97 ($< 2.2 \cdot 10^{-16}$)	
	16	I	I	29	1	AR		
	17	I	III	27	0.93	AR		
	18	III	III	10	1	AR		
Mâle AM × Femelle T	19	V	I	31	0	T	0.21 ($2.58 \cdot 10^{-15}$)	
	20	II	I	19	0.47	AM		
	21	II	I	6	0	T		
	22	V	III	21	0	T		
	23	II	I	21	0	T		
	24	V	I	18	0.06	T		
	25	V	III	17	0	T		
	26	II	III	13	0	T		
	27	II	III	21	0.81	AR		
	28	V	III	16	0.69	AM		
Mâle SR × Femelle AM	29	I	Aas1f2	15	1	AR	0.67 ($8.02 \cdot 10^{-6}$)	
	30	III	Aas1f1	24	0.71	AR		
	31	III	Aas3f1	23	0.17	T		
	32	III	Aas3f1	28	0.96	AR		
	33	I	Aas8f1	19	0.47	AM		
	34	III	Aas9f1	20	1	AR		
	35	I	Aas8f2	34	0.53	AM		
Mâle SR × Femelle T	36	II	Aas1f2	36	0	T	0.09 ($< 2.2 \cdot 10^{-16}$)	
	37	II	Aas1f1	33	0.06	T		
	38	V	Aas3f1	6	1	AR		
	39	II	Aas8f1	32	0	T		
	40	V	Aas8f1	26	0	T		
	41	II	Aas9f1	9	0	T		
	42	II	Aas8f2	15	0.6	AM		
	43	V	Aas8f2	24	0	T		

Malgré les fortes variations de sex-ratio observées parmi les portées F2 issues de nos croisements contrôlés, le sex-ratio global (basé sur 28 croisements) demeure à l'équilibre (0.51, Tableau 27). Afin d'évaluer l'héritabilité de ces variations de sex-ratio, une analyse en composante principale (ACP) a été réalisée, en prenant en compte le sex-ratio de chaque portée F2, mais aussi le sex-ratio des portées F1 dont sont issus les mères et les pères utilisés dans les différentes séries de croisements (ACP réalisée avec le logiciel R, version 3.3.0, R Core Team, 2014) (Figure 68).

L'axe horizontal, qui totalise 48.5% de l'information de l'ACP, est composé majoritairement des variables "Sex-ratio F2" et "Sex-ratio F1 maternelle" (Figure 68, A). L'axe vertical, qui représente 31.9% de l'information de l'ACP, est composé majoritairement de la variable "Sex-ratio F1 paternelle". Cela est confirmé par la séparation selon l'axe vertical des portées F2 selon le sex-ratio dont sont issus leurs pères (Figure 68, B), et selon l'axe horizontal de ces mêmes portées F2 selon le sex-ratio dont sont issues leurs mères (Figure 68, C). Le résultat le plus notable est que le sex-ratio des portées F2 semble fortement corrélé au sex-ratio des portées F1 dont sont issues leurs mères, car ces deux variables suivent globalement la même orientation sur le cercle des corrélations (Figure 68, A).

Le sex-ratio des portées F1 dont sont issues les mères semble donc avoir une influence sur le sex-ratio des portées F2 que produisent ces mères, ce qui est d'ailleurs statistiquement fortement soutenu, comme le montre un modèle linéaire généralisé (GLM : $\chi^2 = 52.98$, $df = 1$, $p\text{-value} = 3.4.10^{-13}$). Par contre, le sex-ratio des portées F1 dont sont issus les pères ne semble pas avoir d'influence sur le sex-ratio des portées F2 produites par ces pères (GLM : $\chi^2 = 1.14$, $df = 1$, $p\text{-value} = 0.29$).

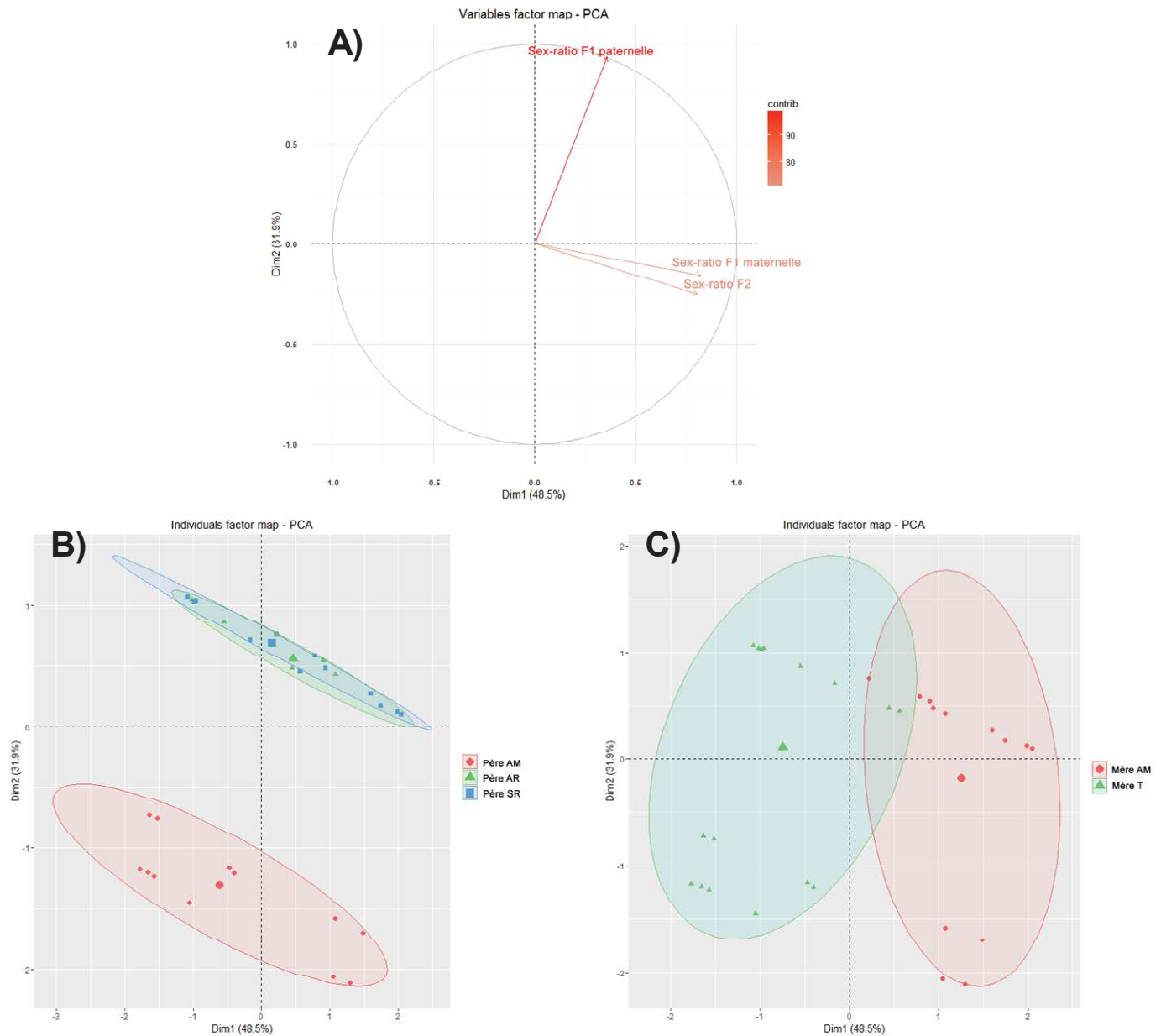


Figure 68 : Analyse en Composantes Principales (ACP) réalisée à partir des sex-ratios des portées issues de la série de croisements contrôlés, et des sex-ratios des portées des parents des croisements contrôlés. A) Cercle des corrélations décrivant la projection des variables sur le plan des composantes principales. L'axe horizontal (48.5% de l'information) est principalement composé du sex-ratio des portées issues de croisements contrôlés (noté "Sex-ratio F1") et du sex-ratio des portées dont sont issues les mères de chaque croisement contrôlé. L'axe vertical (31.9% de l'information) est principalement composé du sex-ratio des portées dont sont issues les pères de chaque croisement contrôlé. B) Représentation des portées contrôlées groupées par le type de père selon les axes de l'ACP (les portées groupées en rouge ont un père AM, les portées groupées en vert sont issues d'un père AR et les portées en bleu sont issues d'un père SR). C) Représentation des portées contrôlées groupées par le type de mère selon les axes de l'ACP (les portées groupées en rouge ont une mère AM, les portées groupées en vert sont issues d'une mère T).

Comment expliquer les résultats que nous avons obtenus ? Différentes catégories de mécanismes connus pour être impliqués dans le déterminisme du sexe ont été résumées dans une revue rédigée par Werren et Beukeboom (1998). L'ensemble des mécanismes décrits ont été classés selon leur niveau d'action : (i) les gènes de sex-ratio parentaux agissant chez les parents, (ii) les effecteurs parentaux déterminant le sexe au niveau des gamètes produits et (iii) les effecteurs agissant directement post-fécondation au niveau du zygote. Ces derniers ont d'ailleurs été exclus pour *A. assimile* car un système faisant intervenir des chromosomes sexuels classiques a été écarté par les résultats de croisement obtenus suite aux expériences de réversion de sexe (Becking et al., 2017). Les mécanismes décrits en (i) et en (ii) permettent de différencier l'effet paternel de l'effet maternel. Dans le cas d'*A. assimile*, nos analyses statistiques (Figure 68) indiquent que le mécanisme impliqué est dépendant de la mère et non du père. Ainsi tous des mécanismes basés sur un effet paternel ont été exclus. Les mécanismes des catégories (i) et (ii) à effet maternel décrits par Werren et Beukeboom (1998) sont discutés ci-dessous :

- *Sex chromosome drivers* : Ce phénomène est caractérisé par une transmission inégale des hétérochromosomes d'un individu du sexe hétérogamétique (Jaenike, 2001). Ce biais de ségrégation chromosomique mène à d'importants biais de sex-ratio au sein des populations. Dans le cas d'*A. assimile*, un tel biais à l'échelle populationnelle est inconnu, car peu d'informations sont disponibles sur cette espèce. L'ensemble des croisements effectués a révélé un sex-ratio global équilibré (proportion globale de mâles au sein des 28 croisements : 0.51, Tableau 27). Ces phénomènes de "*drive*" induisent d'ailleurs des biais de sex-ratio uniquement mâles ou femelles. Il n'est donc pas possible d'expliquer nos résultats avec cette hypothèse (puisque des portées AR et T ont été obtenues), sauf en cas d'intervention d'un mécanisme répresseur, qui implique un niveau de complexité supplémentaire à cette explication. Enfin, un tel phénomène de biais de ségrégation de l'hétérochromosome ne peut expliquer les sex-ratios observés lors des croisements effectués entre femelles et néo-mâles (Becking et al., 2017, Figure 64).

- *Cytoplasmic Sex-Ratio elements* : L'élément cytoplasmique perturbateur du sex-ratio connu chez les isopodes terrestres est *Wolbachia*, cependant sa présence induit des biais de sex-ratio en faveur des femelles (expliquant l'existence des femelles T) mais pas en faveur des mâles (*Wolbachia* ne peut donc pas expliquer l'existence de femelles AR). De plus, sa présence ayant été exclue de l'ensemble de nos croisements, cette possibilité est aussi à exclure.
- *Sex-Ratio behavior + resource allocation* : En allouant plus de ressources à un seul sexe de la descendance, la mère peut influencer directement le sex-ratio de sa progéniture. Une telle situation est possible si la valeur adaptative d'un sexe est supérieure à l'autre sexe pour des conditions environnementales données (Trivers & Willard, 1973; Bull, 1981). Dans le cas d'*A. assimile*, toutes les portées ont été élevées dans les mêmes conditions environnementales. Il serait donc attendu d'observer un unique pattern de sex-ratio (par exemple uniquement des portées T ou AR). Cependant malgré les conditions d'élevage identiques, 3 patterns de sex-ratio ont été observés (T, AR et AM), ce qui rejette cette hypothèse d'allocation différentielle des ressources en fonction des variations environnementales.
- *Gametic imprinting* : Ce mécanisme implique des phénomènes épigénétiques (tels que des méthylations) agissant sur les déterminants génétiques du sexe. Cette catégorie de mécanismes est discutée plus en détail à la fin des perspectives.
- *Maternal effect sex determiner* : La majorité des effets maternels sont causés par des produits maternels, tels que des ARNm ou des protéines, présents au sein des embryons au cours du développement. Ces effets maternels sont particulièrement importants au cours des premiers stades du développement car chez la majorité des organismes, le génotype du zygote n'est pas exprimé au cours des premières divisions mitotiques. Les premières étapes du développement sont donc dépendantes des produits présents dans les embryons. Le produit de ces gènes maternels dans les embryons peut avoir un effet majeur sur le mécanisme de déterminisme du sexe de la descendance. Dans le cas d'*A. assimile*, ces effecteurs maternels

seraient donc soit maintenus après l'émergence des juvéniles, soit permettraient le déclenchement de la cascade de déterminisme du sexe, bien avant la différenciation sexuelle phénotypique, qui débute classiquement après 2 à 4 stades de mue post-naissance (Juchault, 1966; Katakura, 1984; Suzuki & Yamasaki, 1995; Badawi et al., 2015), ce qui correspond au moins à 8 semaines post-naissance dans l'exemple d'*A. vulgare* (Suzuki & Yamasaki, 1995b; Badawi et al., 2015).

L'ensemble des scénarios évoqués ci-dessus ne parviennent donc pas à expliquer totalement nos résultats, à moins de faire intervenir des éléments complexes et difficiles à tester. Les analyses effectuées suggèrent non pas l'implication d'un gène déterminant le sexe mais plutôt un gène de sex-ratio dont la composition allélique chez la mère influencerait le sex-ratio de sa descendance. Le modèle le plus simple serait composé d'un unique *locus* à deux allèles sous relation de dominance (l'allèle F provoquant un sex-ratio fortement biaisé en faveur des femelles, et l'allèle M responsable d'un sex-ratio fortement biaisé en faveur des mâles). Un tel modèle peut expliquer l'existence des portées T et AR, mais pas des portées AM. Cette hypothèse peut être complétée sur la base d'une relation de codominance entre les allèles F et M. Dans ce cas, une femelle peut donc être : (i) homozygote FF et produire majoritairement des femelles, correspondant aux portées T, (ii) hétérozygote FM et produire autant de mâles que de femelles, correspondant aux portées AM, (iii) ou bien homozygote MM et produire majoritairement des mâles, correspondant aux portées AR. Une telle configuration permet d'expliquer les sex-ratios de 36 des 43 portées issues de nos croisements contrôlés. Elle permet aussi d'expliquer pourquoi seulement des mâles ont été produits suite aux expériences de réversion de sexe (Becking et al., 2017, Figure 64), en partant de l'hypothèse que les 6 portées ont été produites à partir de femelles MM. Malgré son bon pouvoir prédictif (84% des portées expliquées), cette hypothèse ne permet cependant pas d'expliquer les sex-ratios des 7 portées AR issues de mères T (portées 9 à 13, 27 et 38, Tableau 28).

Tableau 28 : Génotypes possibles des mères ayant produit l'ensemble des portées issues de croisements contrôlés (les croisements entre un néo-mâle et une femelle génétique, décrits dans Becking et al., 2017, ont aussi été ajoutés). Le symbole "=" décrit une relation de codominance allélique, tandis que le symbole ">" traduit d'une relation de dominance allélique. Les croisements pour lesquels aucun génotype ne peut expliquer le sex-ratio obtenu sont indiqués avec un symbole "-".

	Numéro de la portée	Proportion de mâles	Type de portée	1 locus à 2 allèles codominants (F = production de femelles, M = production de mâles)	1 locus à 3 allèles (F = production de femelles, M = production de mâles, E= sex-ratio équilibré)		
					M = F	F > M > E	M > F > E (M = F) > E
Portées SR issues de Becking et al. (2017)	-	1	AR	MM	MM/ME	MF/ME	MM/ME
	-	1	AR	MM	MM/ME	MF/ME	MM/ME
	-	1	AR	MM	MM/ME	MF/ME	MM/ME
	-	1	AR	MM	MM/ME	MF/ME	MM/ME
	-	1	AR	MM	MM/ME	MF/ME	MM/ME
	-	1	AR	MM	MM/ME	MF/ME	MM/ME
Mâle AR × Femelle AM	1	1	AR	MM	MM/ME	MF/ME	EM/MM
	2	0.96	AR	MM	MM/ME	MF/ME	EM/MM
	3	0.25	T	FF	FF/MF/EF	FF/EF	EF/FF
	4	0.57	AM	MF	EE	EE	EE/MF
	5	0	T	FF	FF/MF/EF	FF/EF	EF/FF
	6	0.3	T	FF	FF/MF/EF	FF/EF	EF/FF
Mâle AR × Femelle T	7	0.04	T	FF	FF/MF/EF	FF/EF	EF/FF
	8	0	T	FF	FF/MF/EF	FF/EF	EF/FF
	9	1	AR	-	ME	MF/EM	EM
	10	1	AR	-	ME	MF/EM	EM
	11	1	AR	-	ME	MF/EM	EM
	12	1	AR	-	ME	MF/EM	EM
	13	1	AR	-	ME	MF/EM	EM
	14	0.35	T	FF	FF/MF/EF	FF/EF	EF/FF
Mâle AM × Femelle AM	15	0.96	AR	MM	MM/ME	MF/ME	ME/MM
	16	1	AR	MM	MM/ME	MF/ME	ME/MM
	17	0.93	AR	MM	MM/ME	MF/ME	ME/MM
	18	1	AR	MM	MM/ME	MF/ME	ME/MM
Mâle AM × Femelle T	19	0	T	FF	FF/MF/EF	FF/EF	EF/FF
	20	0.47	AM	MF	EE	EE	EE/MF
	21	0	T	FF	FF/MF/EF	FF/EF	EF/FF
	22	0	T	FF	FF/MF/EF	FF/EF	EF/FF
	23	0	T	FF	FF/MF/EF	FF/EF	EF/FF
	24	0.06	T	FF	FF/MF/EF	FF/EF	EF/FF
	25	0	T	FF	FF/MF/EF	FF/EF	EF/FF
	26	0	T	FF	FF/MF/EF	FF/EF	EF/FF
	27	0.81	AR	-	ME	MF/EM	EM
	28	0.69	AM	MF	EE	EE	EE/MF
Mâle SR × Femelle AM	29	1	AR	MM	MM/ME	MF/ME	MM/ME
	30	0.71	AR	MM	MM/ME	MF/ME	MM/ME
	31	0.17	T	FF	FF/MF/EF	FF/EF	EF/FF
	32	0.96	AR	MM	MM/ME	MF/ME	MM/ME
	33	0.47	AM	MF	EE	EE	EE/MF
	34	1	AR	MM	MM/ME	MF/ME	MM/ME
	35	0.53	AM	MF	EE	EE	EE/MF
Mâle SR × Femelle T	36	0	T	FF	FF/MF/EF	FF/EF	EF/FF
	37	0.06	T	FF	FF/MF/EF	FF/EF	EF/FF
	38	1	AR	-	ME	MF/EM	EM
	39	0	T	FF	FF/MF/EF	FF/EF	EF/FF
	40	0	T	FF	FF/MF/EF	FF/EF	EF/FF
	41	0	T	FF	FF/MF/EF	FF/EF	EF/FF
	42	0.6	AM	MF	EE	EE	EE/MF
	43	0	T	FF	FF/MF/EF	FF/EF	EF/FF

Néanmoins, l'introduction d'un troisième allèle à ce modèle permet d'expliquer la totalité de nos résultats (Tableau 28) : l'allèle E, qui induit un sex-ratio équilibré et est récessif sur les allèles M et F (car les portées AM sont minoritaires, seulement 6 portées sur les 43 présentes dans le Tableau 28). Dans ce modèle à trois allèles, les allèles M et F peuvent être soit soumis à une relation de codominance comme évoqué précédemment, soit à une relation de dominance (Figure 69). La totalité des sex-ratios observés peut ainsi être expliquée par au moins un génotype maternel possible (Tableau 28), quelle que soit la relation de dominance/codominance entre les allèles M et F. Ces modèles constituent ainsi les premières bases à tester pour de futurs croisements, et notamment la génération d'une F3.

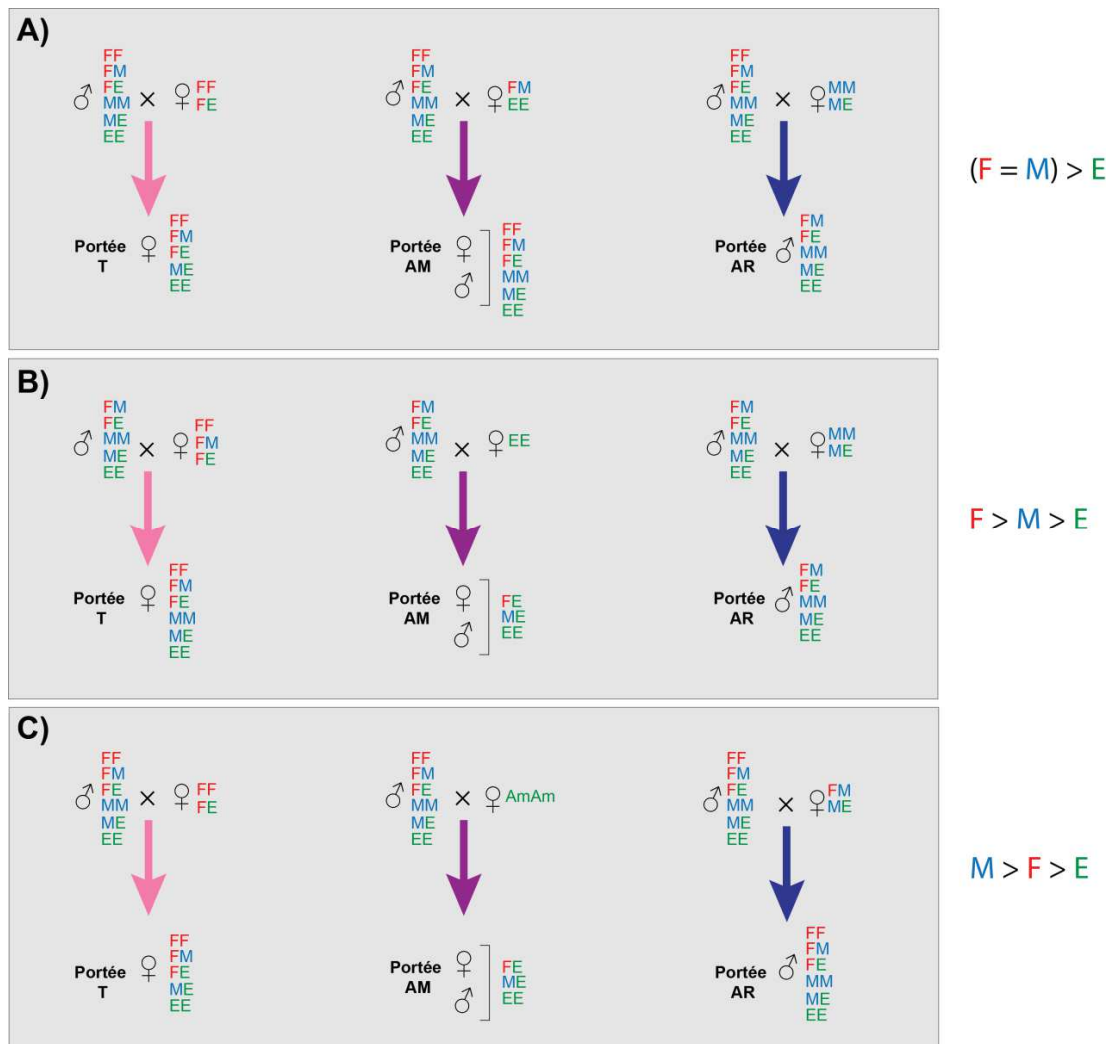


Figure 69 : Hypothèses d'un gène à 3 allèles responsables de biais de sex-ratio chez *A. assimile*. A) Hypothèse des allèles F et M codominants. Une femelle FF ou FE produit une portée T. Une femelle FM ou EE produit quant à elle une portée AM, tandis qu'une femelle MM ou ME produit une portée AR. B) Hypothèse de l'allèle F dominant l'allèle M. Une femelle FF, FM ou FE produit une portée T. Une femelle EE produit une portée AM, tandis qu'une femelle MM ou ME produit une portée AR. C) Hypothèse de l'allèle M dominant l'allèle F. Une femelle FF ou FE produit une descendance T. Une femelle EE produit une descendance AM, tandis qu'une femelle FM ou ME produit une descendance AR.

De telles variations de sex-ratio ont déjà été observées chez d'autres organismes. Ainsi, dans l'exemple du diptère *Chrysomya rufifacies*, des portées composées uniquement de mâles ou uniquement de femelles ont été observées (Ullerich, 1963, 1984). Pour expliquer ce pattern, il a été proposé que le sex-ratio est déterminé maternellement par un *locus* unique à 2 allèles F (dominant) et f (récessif) : les femelles F/f produisant des portées uniquement composées de femelles, et les femelles f/f produisant des portées uniquement composées de mâles (Ullerich, 1963, 1984; Scott et al., 2014). Si ce système est analogue à celui que nous proposons pour *A. assimile*, il ne permet cependant pas d'expliquer des sex-ratios autres que 100% de mâles ou 100% de femelles. Or dans le cas d'*A. assimile*, 19 portées sur 43 présentent à la fois des mâles et des femelles (Tableau 27), et ne peuvent donc être expliqués par le même système que celui proposé pour *C. rufifacies*. Des phénomènes proches ont aussi été observés chez le gastéropode *Pomacea canaliculata*, chez qui une corrélation directe a été observée entre le sex-ratio de la portée F0 maternelle et le sex-ratio de la F1 (Yusa, 2005). Une explication génétique a ensuite été proposée pour expliquer ces patterns de sex-ratio, basée sur l'intervention d'au moins 4 gènes déterminant le sexe chez les zygotes (Yusa, 2007). Des sex-ratios fortement variables ont aussi été observés chez l'ortie *Urtica dioica*, associés là encore à un fort effet maternel (Glawe & de Jong, 2007). Les auteurs de cette étude ont proposé un modèle de déterminisme du sexe chez le zygote basé sur un unique *locus* à 4 allèles, cependant ce modèle ne parvenait pas à expliquer pleinement leurs observations (Glawe & de Jong, 2009). Les deux cas exposés précédemment rapportent néanmoins peu de cas de portées composées uniquement de femelles ou uniquement de mâles, alors qu'il s'agit de la situation la plus souvent rencontrée suite à nos croisements impliquant *A. assimile* (24 portées sur 43, Tableau 27).

Dans notre cas, il est donc possible d'expliquer les variations de sex-ratio observés chez *A. assimile* en invoquant une base génétique pure et simple (1 *locus* à 3 allèles). Néanmoins, il n'est pas exclu que des facteurs épigénétiques puissent influencer le sex-ratio des portées d'*A. assimile*. Par exemple, une étude récente réalisée sur le poisson *Cynoglossus semilaevis* (ayant un déterminisme chromosomique du sexe de type ZZ/ZW), a permis de révéler l'influence du pattern de méthylation du chromosome Z sur le sexe phénotypique de la

descendance (S. Chen et al., 2014). Suite à la réversion de femelles génétiques ZW en mâles phénotypiques (après un traitement par la température), ces néo-mâles ZW ont été croisés avec des femelles génétiques. Les individus ZW ayant hérité du chromosome Z des néo-mâles se sont très majoritairement développés en mâles ZW. Les auteurs de cette étude ont démontré que le pattern de méthylation du Z a été modifié suite à la réversion des femelles ZW en néo-mâles, et que ce pattern de a été conservé chez les descendants de ces néo-mâles ayant hérité du Z de leur père (S. Chen et al., 2014). Dans le cas d'*A. assimile*, une telle étude pose tout d'abord la question de l'existence de chromosomes sexuels, mais aussi de l'influence de la réversion de sexe sur l'épigénome des néo-mâles, et enfin sur l'héritabilité trans-générationnelle des patrons de méthylation à la descendance (d'autant plus que les phénomènes de méthylation ont déjà été observés chez les isopodes, Barzotti et al., 2006).

En conclusion, ces premières investigations réalisées chez *A. assimile* mettent en évidence un fort effet maternel sur le sex-ratio de la descendance. Il faut noter que ces travaux sont encore préliminaires, et que les résultats et interprétations qui en découlent doivent être considérés avec précaution. Dans l'état actuel de nos connaissances, un mécanisme de déterminisme du sexe suivant une hétérogamétie chromosomique n'est pas suffisant pour expliquer les sex-ratios obtenus suite aux croisements entre néo-mâles et femelles génétiques (Becking et al., 2017, Figure 64) ainsi que les sex-ratios obtenus suite à nos croisements contrôlés (Tableau 27). Malgré les biais particulièrement forts de sex-ratio observés (de 100% de femelles à 100% de mâles), le sex-ratio global reste équilibré, Tableau 27). Le principe de Fisher prédit un sex-ratio équilibré à l'échelle populationnelle (Fisher, 1930). Un tel équilibre peut être atteint si les sex-ratios de chaque portée sont globalement équilibrés, ou alors si les sex-ratios de chaque portée sont fortement variables, mais que l'ensemble de ces variations permette un sex-ratio populationnel équilibré, ce qui semblerait être le cas d'*A. assimile*. De plus, ces résultats suggèrent que des mécanismes de déterminisme du sex-ratio différents des mécanismes de déterminisme chromosomique du sexe pourraient exister chez les isopodes terrestres.

Une telle observation soulève la question de l'influence de *Wolbachia* féminisantes sur l'évolution des mécanismes de déterminisme du sexe chez les Oniscidea. Si *Wolbachia* est responsable des turnovers de systèmes d'hétérogamétie observés chez les isopodes terrestres (Becking et al., 2017), l'exemple d'*A. assimile* soulève la question de savoir si *Wolbachia* pourrait être responsable de l'émergence de processus impliqués dans le déterminisme ou la différenciation sexuelle différents des systèmes hétérogamétiques. La présence de *Wolbachia* n'a pas été établie pour le moment chez *A. assimile*, cependant il est possible que la bactérie soit présente dans des populations naturelles autres que celle que nous avons étudiée, ou alors que *Wolbachia* ait été associée à *A. assimile* par le passé et que l'infection ait été perdue au cours de l'évolution. A l'inverse, si *Wolbachia* n'a jamais été associée avec *A. assimile*, alors le mécanisme impliqué dans le déterminisme du sexe/sex-ratio de cette espèce pourrait constituer le processus de déterminisme ancestral des isopodes terrestres, l'émergence de chromosomes sexuels constituant alors l'état dérivé d'une espèce soumise à une infection à *Wolbachia*.

L'hypothèse selon laquelle *Wolbachia* (ou un facteur cytoplasmique féminisant équivalent) serait responsable de l'évolution du gonochorisme chez les crustacés a été proposée par Juchault (1999). En effet, il a été proposé que l'hermaphrodisme simultané (un individu dispose à la fois des caractères mâles et femelles) constitue le mode de reproduction ancestral des crustacés (Charniaux-Cotton et al., 1992), comme rencontré chez des espèces appartenant aux classes de crustacés telles que les Cephalocorida (Hessler et al., 1995) et les Remipedia (Yager, 1989). Dans son modèle, Juchault propose que chez les premiers crustacés hermaphrodites, l'introduction d'un facteur cytoplasmique (donc à transmission ovocytaire) féminisant aurait permis de générer des individus uniquement femelles. L'effet féminisant aurait été sélectionné par ce facteur parasitique afin d'accroître sa transmission à la descendance (car la fonction mâle d'un individu hermaphrodite serait une impasse évolutive pour ce facteur cytoplasmique), cette dernière étant composée de femelles symbiotiques qui se seraient accouplées avec les individus asymbiotiques hermaphrodites. Puis des phénomènes de sélection auraient conduit à la disparition de la fonction femelle chez les hermaphrodites, qui auraient peu à peu évolué pour devenir des mâles gonochoriques. Il faut noter

que la transition vers l'évolution de chromosomes sexuels implique aussi l'intégration de la fonction féminisante du facteur cytoplasmique dans le génome de son hôte, conduisant à la formation de femelles génétiques (comme cela a pu être démontré récemment avec l'exemple de l'intégration de l'élément f chez *A. vulgare*, Leclercq et al., 2016). Selon cette hypothèse, si *A. assimile* a évolué totalement indépendamment de *Wolbachia*, alors cette espèce a pu conserver le potentiel que chaque individu soit hermaphrodite, et le déterminisme ou la différenciation sexuelle serait directement sous l'effet d'effecteurs d'origine maternels.

Même si ces hypothèses sont issues de travaux encore préliminaires, l'exemple que constitue *A. assimile* met en exergue la labilité des mécanismes menant à la différenciation d'individus sexués chez les isopodes terrestres. Cet exemple suggère aussi que la diversité de ces mécanismes rencontrés chez les isopodes terrestres ne se limite peut-être pas uniquement à de l'hétérogamétie chromosomique. Enfin, les investigations réalisées chez *A. assimile* illustrent une fois encore le formidable potentiel des isopodes terrestres pour l'étude de l'évolution du déterminisme du sexe.

Synthèse générale

L'association endosymbiotique entre les isopodes terrestres et *Wolbachia* permet d'illustrer l'impact fort de l'endosymbiose sur l'histoire évolutive de l'hôte. A l'échelle individuelle, *Wolbachia* est capable de féminiser des mâles génétiques en femelles fonctionnelles (Martin et al., 1973). Si l'effet de *Wolbachia* sur le phénotype de son hôte est avéré depuis longtemps (changement du sexe phénotypique), les différents travaux réalisés au cours de cette thèse suggèrent que ces effets ont également des conséquences à l'échelle de l'évolution d'un groupe taxonomique entier (à savoir ici les Oniscidea).

Dans un premier temps, cette thèse a permis de générer un génome mâle de l'espèce *A. nasatum* de qualité supérieure à tous les génomes de crustacés disponibles jusqu'à présent dans les banques de données publiques (Taille: 1.22 Gb dont 0.01% de N, *contigs/scaffolds* : 25 194, N50: 86.3 kb). La production d'un génome de bonne qualité a été possible grâce à l'essor des technologies de séquençage *long reads* au cours des dernières années, et au développement d'outils de traitement de ces nouvelles données (Goodwin et al., 2016). L'annotation du génome d'*A. nasatum* a permis de mettre en évidence que près de 63% du génome est composé d'éléments répétés (ETs, répétitions en tandem), expliquant pourquoi le premier assemblage réalisé uniquement avec les données *Illumina* était extrêmement fragmenté. L'annotation a aussi permis d'identifier plusieurs fragments du génome de *Wolbachia* insérés dans le génome d'*A. nasatum*. Les transferts horizontaux de matériel génétique issu de *Wolbachia* vers son hôte sont fréquents et répandus (Dunning Hotopp et al., 2007; Dunning Hotopp, 2011; Robinson et al., 2013), et sont d'ailleurs facilités par la relation d'endosymbiose entre la bactérie et ses hôtes (Cordaux & Gilbert, 2017). Ces insertions illustrent ainsi l'impact direct de *Wolbachia* sur la structuration du génome de son hôte, puisque l'impact de *Wolbachia* sur *A. nasatum* (mais aussi sur *A. vulgare*) affecte directement le phénotype de l'hôte (par la féminisation), mais aussi le génotype de l'hôte. De plus en plus de transferts horizontaux de gènes d'origine bactérienne vers des génomes eucaryotes ont été caractérisés au cours des dernières années. Ces intégrations peuvent d'ailleurs avoir une incidence sur le potentiel adaptatif de l'hôte, comme par exemple à travers l'acquisition de toxines au rôle antibactérien chez de nombreux organismes (Metcalf et al., 2014;

Chou et al., 2015), d'enzymes impliquées dans la phytophagie (Danchin et al., 2010; Wybouw et al., 2016) ou encore la synthèse de l'hème (Wu et al., 2013) chez plusieurs arthropodes et nématodes. De plus, il a été démontré récemment que des intégrations de *Wolbachia* pouvaient directement influencer le déterminisme du sexe de l'hôte. C'est le cas de l'élément f, qui est une insertion près de 80% du génome de *Wolbachia* dans un autosome d'*A. vulgare*, ce dernier agissant tel un nouveau chromosome W (Leclercq et al., 2016).

Enfin, le génome de haute qualité d'*A. nasatum*, couplé à celui d'*A. vulgare* généré dans le cadre de la thèse d'Amine Chebbi, constituent de précieuses ressources qui permettront d'aborder plusieurs questions dans le domaine de la biologie évolutive, telles que la caractérisation des acteurs moléculaires impliqués dans la sortie des eaux chez les crustacés, ou encore l'étude de l'évolution même des crustacés, encore méconnue à ce jour. Plus généralement, ces génomes peuvent aussi contribuer à mieux comprendre l'évolution des arthropodes, et plus précisément les relations phylogénétiques entre crustacés et insectes (Kao et al., 2016). Bien entendu, ces ressources sont aussi de grande valeur pour l'étude de l'évolution du déterminisme du sexe, comme décrit dans le Chapitre 2.

La génome réalisé au cours du premier chapitre de cette thèse a déjà permis de caractériser les séquences associées au chromosome Y d'*A. nasatum*. Comme pour de nombreux isopodes terrestres (Juchault & Rigaud, 1995; Rigaud et al., 1997), les chromosomes sexuels d'*A. nasatum* sont homomorphes, il est donc impossible de différencier le chromosome X du chromosome Y sur une base cytogénétique. La stratégie de recherche des séquences Y-spécifiques a été réalisée sans *a priori*. L'identification de plusieurs séquences liées au sexe mâle constitue un premier pas vers la caractérisation de la région Y-spécifique d'*A. nasatum*. Ces analyses ont permis de confirmer *in silico* l'homomorphie des chromosomes sexuels, et donc que la région Y-spécifique représente une très faible fraction du génome d'*A. nasatum*. Cette homomorphie a aussi été fonctionnellement confirmée puisque des individus femelles et mâles YY viables ont pu être mis en évidence. Les résultats obtenus au cours de ce second chapitre suggèrent que l'homomorphie des chromosomes sexuels d'*A. nasatum* pourrait

être causée par la présence de *Wolbachia* féminisantes. En effet, ces bactéries induisent l'apparition d'individus YY au sein des populations (par exemple suite au croisement entre un mâles phénotypique XY et une femelle XY féminisée par la bactérie), chez lesquels les deux chromosomes Y sont susceptibles de recombiner, prévenant ainsi leur dégénérescence. Il a été démontré chez certaines espèces que la recombinaison pouvait avoir lieu selon des patterns spécifiques de chaque sexe phénotypique (hétérochiasmie) (Isberg et al., 2006; Berset-Brändli et al., 2008). Dans le cas d'*A. nasatum*, si la recombinaison est plus forte chez les femelles phénotypiques, les femelles XY infectées par *Wolbachia* pourraient présenter le même pattern de recombinaison que les femelles XX et donc permettre la recombinaison occasionnelle entre le X et le Y, maintenant l'homomorphie des chromosomes sexuels. Selon cette hypothèse, l'existence d'individus XY féminisés dans les populations constitue une "fontaine de jouvence" pour le chromosome Y (Perrin, 2009), empêchant ainsi sa dégénérescence. Enfin, une autre explication est que les conflits génétiques entre les déterminants du sexe d'*A. nasatum* et les *Wolbachia* féminisantes soient responsables d'un turnover des chromosomes sexuels, créant régulièrement de nouveaux gonosomes, et réinitialisant le processus de dégénérescence de l'hétérochromosome (van Doorn & Kirkpatrick, 2010). Cette hypothèse sous-entend que le chromosome Y d'*A. nasatum* est un chromosome sexuel récent. Les tests réalisés au cours de ce second chapitre suggèrent un âge de moins de 15 Ma pour ce chromosome, mais des analyses plus poussées devront être effectuées pour affiner cette estimation. En effet, les résultats du Chapitre 2 suggèrent que le chromosome Y est plus récent que la divergence entre *A. nasatum* et l'espèce qui lui est le plus apparentée selon nos analyses, à savoir *A. tunisiense* (soit environ 15 Ma, Chapitre 4, Figure 63, Becking et al., 2017). Cependant, il est possible que le chromosome Y d'*A. nasatum* soit bien plus jeune encore. En effet, seulement 11 espèces appartenant à la famille des Armadillidiidae sont incluses dans la phylogénie établie dans Becking et al. (2017) (Figure 63), alors que plus de 250 espèces appartenant à cette famille sont connues (Sfenthourakis & Taiti, 2015). Il est donc probable que des espèces plus apparentées à *A. nasatum* que l'espèce *A. tunisiense* existent, et la connaissance de ces espèces nous permettrait d'évaluer plus précisément l'âge du chromosome Y d'*A. nasatum*.

Le troisième chapitre de cette thèse s'est focalisé sur l'étude de gènes candidats potentiellement impliqués dans la cascade de déterminisme/différenciation du sexe d'*A. nasatum* (et dans une moindre mesure d'*A. vulgare*). Les gènes à domaine DM constituaient des candidats incontournables, de par leur conservation à l'échelle des eumétazoaires (Wexler et al., 2014) dans des mécanismes de déterminisme et de différenciation sexuelle (Picard et al., 2015). Chez les insectes, le gène à domaine DM *Dsx* est associé à la présence des gènes *Tra* et *Tra-2*, eux-aussi impliqués dans le déterminisme du sexe (Verhulst et al., 2010). Quatre gènes à domaine DM ont été identifiées dans le génome d'*A. nasatum* (AnasDMRT99B, AnasDMRT93B, AnasDsx1 et AnasDsx2) et 3 ont été identifiés dans le génome d'*A. vulgare* (AvulDMRT99B, AvulDMRT93B et AvulDsx). De plus, 2 gènes *Tra-2* ont été détectés chez *A. nasatum* (AnasTra2A, AnasTra2B) et *A. vulgare* (AvulTra2A, AvulTra2B). L'expression de la totalité des gènes identifiés chez *A. nasatum* a été mesurée au cours de 5 stades de développement (stade 2 indifférencié à stade 6 totalement différencié) ainsi qu'au stade adulte. Ces analyses ont révélé que les gènes AnasDsx1 et AnasDsx2 (dont la séquence codante est identique au niveau nucléotidique) étaient surexprimés chez les individus mâles de stade 4 (premier stade de différenciation sexuelle des gonades). Ces résultats suggèrent donc que l'un de ces gènes au moins est potentiellement impliqué dans le déterminisme ou dans la différenciation sexuelle d'*A. nasatum*. Pour le moment, des analyses similaires n'ont pas pu être effectuées chez *A. vulgare* car des marqueurs W-spécifiques n'étaient pas disponibles. Il s'agit cependant d'une opportunité rare de pouvoir comparer l'expression de gènes potentiellement impliqués dans le déterminisme du sexe chez deux espèces apparentées ayant un système de déterminisme chromosomique différent. De plus, afin d'établir le rôle exact de ces gènes, il est encore nécessaire d'effectuer des tests de validation fonctionnelle (par interférence ARN par exemple). Il paraît aussi déterminant d'analyser l'expression de ces différents gènes au sein d'individus infectés par *Wolbachia*, ce qui pourrait aider à mieux comprendre l'action féminisante de la bactérie sur la cascade de gènes impliquée dans le déterminisme ou la différenciation sexuelle des isopodes terrestres.

Le quatrième chapitre de thèse avait pour objectif d'estimer le nombre et la direction des transitions de systèmes d'hétérogamétie au cours de l'histoire évolutive des isopodes terrestres. Le peu de travaux réalisés jusqu'alors sur la diversité des mécanismes de déterminisme du sexe chez les Oniscidea (Juchault & Rigaud, 1995; Rigaud et al., 1997) laissait supposer une forte variabilité de ces systèmes. Une phylogénie robuste et datée a tout d'abord été réalisée à partir de données transcriptomiques, basée sur l'alignement de 88 gènes orthologues conservés. En parallèle, des expériences de réversion de sexe, permettant de générer des néo-mâles à partir de femelles en cours de différenciation sexuelle par injection de glande androgène (Juchault & Rigaud, 1995), ont permis de caractériser les systèmes d'hétérogamétie chez de nouvelles espèces (Chapitre 4; Becking et al., 2017). Le *mapping* de ces résultats sur la phylogénie a permis d'inférer au moins 3 transitions de systèmes d'hétérogamétie par une méthode d'estimation parcimonie, et jusqu'à 11 transitions par une méthode d'estimation probabiliste (Chapitre 4; Becking et al., 2017). Ces nombres ne sont très probablement qu'une sous-estimation du nombre réel de transitions au sein des Oniscidea, puisque les inférences n'ont été réalisées qu'à partir de 14 systèmes connus. De plus, l'approche par expérience de réversion de sexe ne permet pas de caractériser des transitions de gonosomes conservant le même système d'hétérogamétie. Globalement, le grand nombre de transitions inférées est cohérent avec l'hypothèse que les conflits génétiques entre *Wolbachia* et les isopodes terrestres auraient déclenché des turnovers récurrents de systèmes d'hétérogamétie. Etant donné le rôle important joué par les chromosomes sexuels dans les processus de spéciation (Coyne & Orr, 2004; Presgraves, 2008), nos résultats soulèvent aussi la question de l'influence de *Wolbachia* sur la diversification des Oniscidea (qui comptent plus de 3 700 espèces décrites, Sfenthourakis & Taiti, 2015). L'ensemble des résultats obtenus au cours de cette thèse est synthétisé en Figure 70, qui souligne l'impact de la relation endosymbiotique entre *Wolbachia* et les isopodes terrestres.

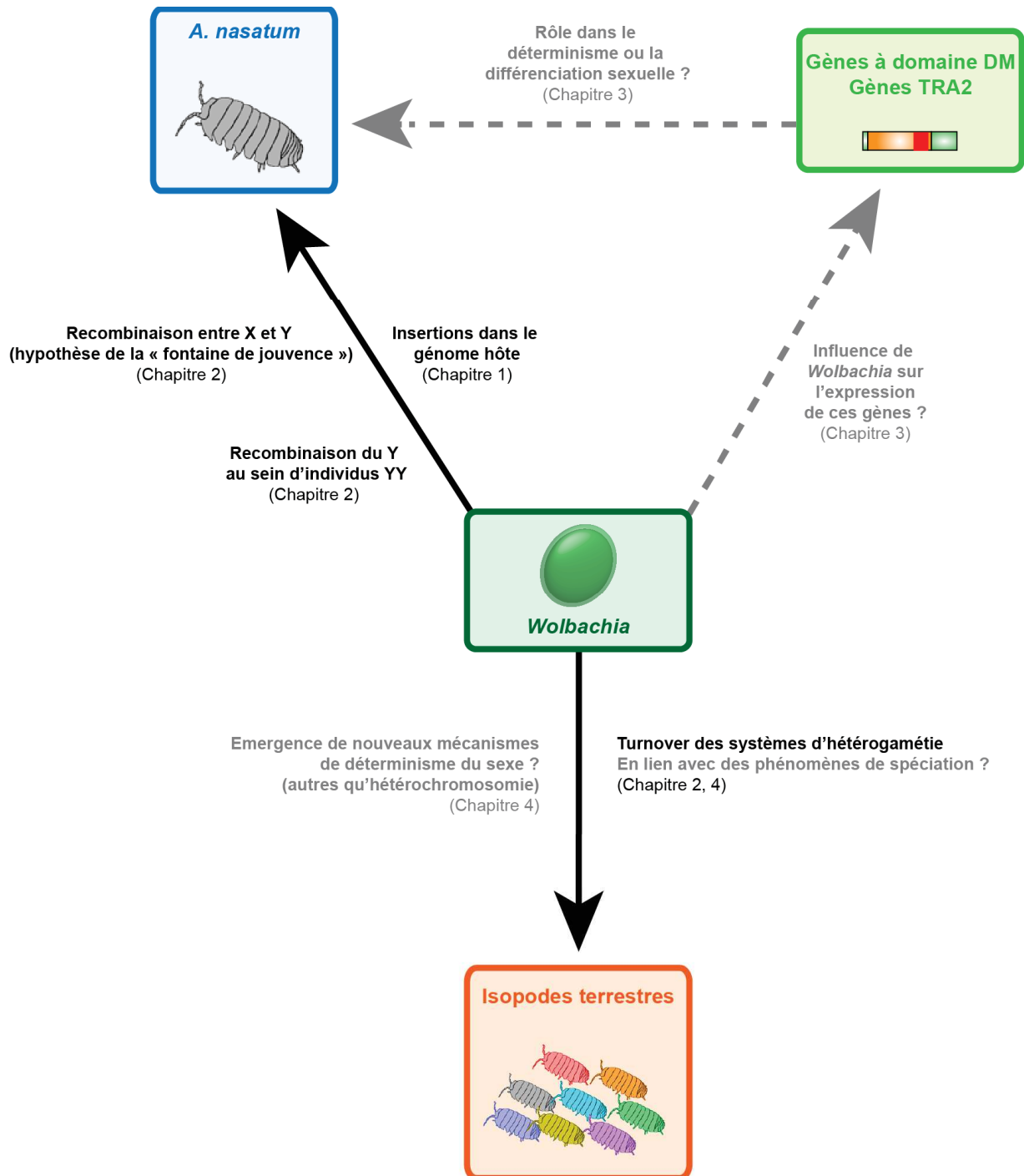


Figure 70 : Schéma récapitulatif des informations issues de cette thèse concernant l'influence potentielle de *Wolbachia* sur l'évolution des isopodes terrestres. Des insertions de *Wolbachia* dans le génome d'*A. nasatum* ont été caractérisées dans le Chapitre 1, soulignant l'impact de la bactérie sur la structure du génome de son hôte. La présence de *Wolbachia* féminisante chez *A. nasatum* permet l'introduction d'individus YY potentiellement recombinant, maintenant l'homomorphie du Y (Chapitre 2). Elle permet aussi l'introduction de femelles phénotypiques XY chez qui le chromosome Y peut potentiellement recombiner (et donc être maintenu à l'état homomorphe) (Chapitre 2). Si les gènes à domaines DM et Tra2 identifiées chez *A. nasatum* et *A. vulgare* sont réellement impliqués dans le déterminisme ou la différenciation sexuelle chez ces espèces, cela soulève la question de l'influence de *Wolbachia* sur l'expression de ces déterminants/différenciants afin de féminiser son hôte (Chapitre 3). Enfin, les conflits génétiques entre *Wolbachia* féminisantes et les déterminants du sexe sont susceptibles de provoquer des turnovers récurrents de chromosomes sexuels, augmentant leur diversité au sein des isopodes terrestres (Chapitre 4) et les maintenant dans un état homomorphe (Chapitre 2). La diversité des systèmes d'hétérogamétie rencontrée soulève aussi la question de l'influence de *Wolbachia* sur la diversification des Oniscidea (Chapitre 4).

Ces travaux illustrent ainsi l'impact de l'endosymbiose sur l'évolution de l'hôte. Cet impact a été évalué sur la structure même des génomes hôtes, mais aussi sur la fonction biologique primordiale que représente le déterminisme du sexe. Les isopodes terrestres constituent donc un modèle de choix pour l'étude des relations endosymbiote/hôte d'une part, mais aussi pour l'étude de l'évolution des mécanismes de déterminisme du sexe et des chromosomes sexuels. De part les résultats produits et les discussions évoquées dans ce manuscrit, les Oniscidea semblent donc parfaitement illustrer les mots de Graham Bell lorsqu'il évoquait "*the queen of problems in evolutionary biology*" que représente l'évolution du sexe (Bell, 1982), et l'étude de ce clade promet de passionnantes perspectives concernant les recherches à venir sur l'évolution du déterminisme du sexe.

Références

- Abeel, T., Van Parys, T., Saeys, Y., Galagan, J., Van de Peer, Y., 2012. GenomeView: a next-generation genome browser. *Nucleic Acids Res.* 40, e12. doi:10.1093/nar/gkr995
- Abrusán, G., Grundmann, N., DeMester, L., Makalowski, W., 2009. TEclass--a tool for automated classification of unknown eukaryotic transposable elements. *Bioinformatics* 25, 1329–1330. doi:10.1093/bioinformatics/btp084
- Ahmed, M.Z., Li, S.-J., Xue, X., Yin, X.-J., Ren, S.-X., Jiggins, F.M., Greeff, J.M., Qiu, B.-L., 2015. The Intracellular Bacterium *Wolbachia* Uses Parasitoid Wasps as Phoretic Vectors for Efficient Horizontal Transmission. *PLOS Pathogens* 11, e1004672. doi:10.1371/journal.ppat.1004672
- Ahmed, S., Cock, J.M., Pessia, E., Luthringer, R., Cormier, A., Robuchon, M., Sterck, L., Peters, A.F., Dittami, S.M., Corre, E., Valero, M., Aury, J.-M., Roze, D., Van de Peer, Y., Bothwell, J., Marais, G.A.B., Coelho, S.M., 2014. A Haploid System of Sex Determination in the Brown *Alga Ectocarpus sp.* *Current Biology* 24, 1945–1957. doi:10.1016/j.cub.2014.07.042
- Amrein, H., Gorman, M., Nöthiger, R., 1988. The sex-determining gene *tra-2* of *Drosophila* encodes a putative RNA binding protein. *Cell* 55, 1025–1035.
- Antkowiak, T., Chase, R., 2003. Sensory innervation of the ovotestis in the snail *Helix aspersa*. *Journal of Experimental Biology* 206, 3913–3921. doi:10.1242/jeb.00625
- Artault, J.C., 1977. Contribution à l'étude des garnitures chromosomiques chez quelques Crustacés Isopodes. (Thèse de 3ème cycle). Poitiers, Poitiers, France.
- Asker, S.E., Jerling, L., 1992. Apomixis in plants CRC Press. Boca Raton, USA.
- Bachtrog, D., 2013. Y-chromosome evolution: emerging insights into processes of Y-chromosome degeneration. *Nat. Rev. Genet.* 14, 113–124. doi:10.1038/nrg3366
- Bachtrog, D., 2003. Adaptation shapes patterns of genome evolution on sexual and asexual chromosomes in *Drosophila*. *Nat. Genet.* 34, 215–219. doi:10.1038/ng1164
- Bachtrog, D., Charlesworth, B., 2002. Reduced adaptation of a non-recombining neo-Y chromosome. *Nature* 416, 323–326. doi:10.1038/416323a
- Bachtrog, D., Mank, J.E., Peichel, C.L., Kirkpatrick, M., Otto, S.P., Ashman, T.-L., Hahn, M.W., Kitano, J., Mayrose, I., Ming, R., Perrin, N., Ross, L., Valenzuela, N., Vamosi, J.C., Tree of Sex Consortium, 2014. Sex determination: why so many ways of doing it? *PLoS Biol.* 12, e1001899. doi:10.1371/journal.pbio.1001899
- Badawi, M., Grève, P., Cordaux, R., 2015. Feminization of the Isopod *Cylisticus convexus* after Transinfection of the wVulC *Wolbachia* Strain of *Armadillidium vulgare*. *PLoS ONE* 10, e0128660. doi:10.1371/journal.pone.0128660
- Baker, B.S., Hoff, G., Kaufman, T.C., Wolfner, M.F., Hazelrigg, T., 1991. The doublesex locus of *Drosophila melanogaster* and its flanking regions: a cytogenetic analysis. *Genetics* 127, 125–138.
- Bao, Z., Eddy, S.R., 2002. Automated De Novo Identification of Repeat Sequence Families in Sequenced Genomes. *Genome Res* 12, 1269–1276. doi:10.1101/gr.88502
- Barnes, H., 1989. Oceanography And Marine Biology. CRC Press.
- Barnett, D.W., Garrison, E.K., Quinlan, A.R., Strömberg, M.P., Marth, G.T., 2011. BamTools: a C++ API and toolkit for analyzing and managing BAM files. *Bioinformatics* 27, 1691–1692. doi:10.1093/bioinformatics/btr174
- Barrangou, R., Doudna, J.A., 2016. Applications of CRISPR technologies in research and beyond. *Nat Biotech* 34, 933–941. doi:10.1038/nbt.3659
- Barriel, V., Darlu, P., Tassy, P., 1993. A propos des conflits entre phylogénies morphologiques et moléculaires: deux exemples empruntés aux mammifères, in: *Annales Des Sciences Naturelles. Zoologie et Biologie Animale*. Elsevier, pp. 157–171.

- Barton, N.H., 1994. The reduction in fixation probability caused by substitutions at linked *loci*. *Genetics Research* 64, 199–208. doi:10.1017/S0016672300032857
- Barzotti, R., Pelliccia, F., Rocchi, A., 2006. DNA methylation, histone H3 methylation, and histone H4 acetylation in the genome of a crustacean. *Genome* 49, 87–90. doi:10.1139/g05-086
- Bayrer, J.R., Zhang, W., Weiss, M.A., 2005. Dimerization of doublesex is mediated by a cryptic ubiquitin-associated domain fold: implications for sex-specific gene regulation. *J. Biol. Chem.* 280, 32989–32996. doi:10.1074/jbc.M507990200
- Becking, T., Giraud, I., Raimond, M., Moumen, B., Chandler, C., Cordaux, R., Gilbert, C., 2017. Diversity and evolution of sex determination systems in terrestrial isopods. *Scientific Reports* 7, 1084. doi:10.1038/s41598-017-01195-4
- Beckmann, J.F., Ronau, J.A., Hochstrasser, M., 2017. A *Wolbachia* deubiquitylating enzyme induces cytoplasmic incompatibility. *Nature Microbiology* 2, nmicrobiol20177. doi:10.1038/nmicrobiol.2017.7
- Bell, G., 1982. *The Masterpiece of Nature :: The Evolution and Genetics of Sexuality*. CUP Archive.
- Bell, G., 1978. The evolution of anisogamy. *Journal of Theoretical Biology* 73, 247–270. doi:10.1016/0022-5193(78)90189-3
- Benaglia, T., Chauveau, D., Hunter, D.R., Young, D.S., 2009. mixtools: An R Package for Analyzing Finite Mixture Models. *Journal of Statistical Software* 32, 1–29.
- Benson, G., 1999. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res.* 27, 573–580.
- Bergero, R., Qiu, S., Charlesworth, D., 2015. Gene loss from a plant sex chromosome system. *Curr. Biol.* 25, 1234–1240. doi:10.1016/j.cub.2015.03.015
- Berlin, K., Koren, S., Chin, C.-S., Drake, J.P., Landolin, J.M., Phillippy, A.M., 2015. Assembling large genomes with single-molecule sequencing and locality-sensitive hashing. *Nat. Biotechnol.* 33, 623–630. doi:10.1038/nbt.3238
- Berset-Brändli, L., Jaquiéry, J., Broquet, T., Ulrich, Y., Perrin, N., 2008. Extreme heterochiasmy and nascent sex chromosomes in European tree frogs. *Proceedings of the Royal Society of London B: Biological Sciences* 275, 1577–1585. doi:10.1098/rspb.2008.0298
- Beukeboom, L.W., Perrin, N., 2014. *The Evolution of Sex Determination*. Oxford University Press.
- Bidon, T., Schreck, N., Hailer, F., Nilsson, M.A., Janke, A., 2015. Genome-Wide Search Identifies 1.9 Mb from the Polar Bear Y Chromosome for Evolutionary Analyses. *Genome Biol Evol* 7, 2010–2022. doi:10.1093/gbe/evv103
- Billiard, S., López-Villavicencio, M., Devier, B., Hood, M.E., Fairhead, C., Giraud, T., 2011. Having sex, yes, but with whom? Inferences from fungi on the evolution of anisogamy and mating types. *Biol Rev Camb Philos Soc* 86, 421–442. doi:10.1111/j.1469-185X.2010.00153.x
- Blaser, O., Grossen, C., Neuenschwander, S., Perrin, N., 2012. Sex-Chromosome Turnovers Induced by Deleterious Mutation Load. *Evolution* 67, 635–645. doi:10.1111/j.1558-5646.2012.01810.x
- Blaser, O., Neuenschwander, S., Perrin, N., 2014. Sex-chromosome turnovers: the hot-potato model. *Am. Nat.* 183, 140–146. doi:10.1086/674026
- Boetzer, M., Pirovano, W., 2012. Toward almost closed genomes with GapFiller. *Genome Biol.* 13, R56. doi:10.1186/gb-2012-13-6-r56
- Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. doi:10.1093/bioinformatics/btu170

- Bordenstein, S.R., O'Hara, F.P., Werren, J.H., 2001. *Wolbachia*-induced incompatibility precedes other hybrid incompatibilities in *Nasonia*. *Nature* 409, 707–710. doi:10.1038/35055543
- Bordenstein, S.R., Reznikoff, W.S., 2005. Mobile DNA in obligate intracellular bacteria. *Nat Rev Micro* 3, 688–699. doi:10.1038/nrmicro1233
- Bouchon, D., Cordaux, R., Grève, P., 2008. Feminizing *Wolbachia* and the evolution of sex determination in isopods, in: *Insect Symbiosis, Volume 3, Contemporary Topics in Entomology*. CRC Press, pp. 273–294.
- Bouchon, D., Rigaud, T., Juchault, P., 1998. Evidence for widespread *Wolbachia* infection in isopod crustaceans: molecular identification and host feminization. *Proceedings of the Royal Society of London B: Biological Sciences* 265, 1081–1090. doi:10.1098/rspb.1998.0402
- Braig, H.R., Zhou, W., Dobson, S.L., O'Neill, S.L., 1998. Cloning and characterization of a gene encoding the major surface protein of the bacterial endosymbiont *Wolbachia pipientis*. *J. Bacteriol.* 180, 2373–2378.
- Brockdorff, N., Turner, B.M., 2015. Dosage Compensation in Mammals. *Cold Spring Harb Perspect Biol* 7, a019406. doi:10.1101/cshperspect.a019406
- Broly, P., Deville, P., Maillet, S., 2013. The origin of terrestrial isopods (Crustacea: Isopoda: Oniscidea). *Evolutionary Ecology* 27, 461–476.
- Bruce, N.L., 1981. Redescription of the isopod (Crustacea) family Phoratopodidae. *Beaufortia* 31, 107–110.
- Brusca, R.C., Wilson, G.D.F., 1991. A phylogenetic analysis of the Isopoda with some classificatory recommendations. *Memoirs of the Queensland Museum* 31, 143–204.
- Buchfink, B., Xie, C., Huson, D.H., 2015. Fast and sensitive protein alignment using DIAMOND. *Nat Meth* 12, 59–60. doi:10.1038/nmeth.3176
- Bull, J.J., 1983. *Evolution of sex determining mechanisms*. Benjamin/Cummings Pub. Co.
- Bull, J.J., 1981. Sex ratio evolution when fitness varies. *Heredity* 46, 9–26.
- Bulnheim, H.-P., 1978. Interaction between genetic, external and parasitic factors in sex determination of the crustacean amphipod *Gammarus duebeni*. *Helgoland Marine Research* 31, 1–33.
- Burtis, K.C., Baker, B.S., 1989. *Drosophila* doublesex gene controls somatic sexual differentiation by producing alternatively spliced mRNAs encoding related sex-specific polypeptides. *Cell* 56, 997–1010. doi:10.1016/0092-8674(89)90633-8
- Butterfield, N.J., 2000. *Bangiomorpha pubescens* n. gen., n. sp.: implications for the evolution of sex, multicellularity, and the Mesoproterozoic/Neoproterozoic radiation of eukaryotes. *Paleobiology* 26, 386–404. doi:10.1666/0094-8373(2000)026<0386:BPNGNS>2.0.CO;2
- Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., Madden, T.L., 2009. BLAST+: architecture and applications. *BMC Bioinformatics* 10, 421. doi:10.1186/1471-2105-10-421
- Carretero-Paulet, L., Fares, M.A., 2012. Evolutionary dynamics and functional specialization of plant paralogs formed by whole and small-scale genome duplications. *Mol. Biol. Evol.* 29, 3541–3551. doi:10.1093/molbev/mss162
- Carvalho, A.B., Clark, A.G., 2013. Efficient identification of Y chromosome sequences in the human and *Drosophila* genomes. *Genome Res.* 23, 1894–1907. doi:10.1101/gr.156034.113
- Carvalho, A.B., Clark, A.G., 1999. Intron size and natural selection. *Nature* 401, 344. doi:10.1038/43827
- Carvalho, A.B., Vibranovski, M.D., Carlson, J.W., Celniker, S.E., Hoskins, R.A., Rubin, G.M., Sutton, G.G., Adams, M.D., Myers, E.W., Clark, A.G., 2003. Y chromosome

- and other heterochromatic sequences of the *Drosophila melanogaster* genome: how far can we go? *Genetica* 117, 227–237.
- Carvalho, G.B., Ja, W.W., Benzer, S., 2009. Non-lethal PCR genotyping of single *Drosophila*. *BioTechniques* 46, 312–314. doi:10.2144/000113088
- Caubet, Hatcher, Mocquard, Rigaud, 2000. Genetic conflict and changes in heterogametic mechanisms of sex determination. *Journal of Evolutionary Biology* 13, 766–777. doi:10.1046/j.1420-9101.2000.00225.x
- Charlesworth, B., 1996. The evolution of chromosomal sex determination and dosage compensation. *Curr. Biol.* 6, 149–162.
- Charlesworth, B., 1991. The evolution of sex chromosomes. *Science* 251, 1030–1033.
- Charlesworth, B., Charlesworth, D., 2000. The degeneration of Y chromosomes. *Philosophical Transactions of the Royal Society B: Biological Sciences* 355, 1563–1572. doi:10.1098/rstb.2000.0717
- Charlesworth, B., Charlesworth, D., 1978. A Model for the Evolution of Dioecy and Gynodioecy. *The American Naturalist* 112, 975–997.
- Charlesworth, B., Wall, J.D., 1999. Inbreeding, heterozygote advantage and the evolution of neo-X and neo-Y sex chromosomes. *Proceedings of the Royal Society of London B: Biological Sciences* 266, 51–56. doi:10.1098/rspb.1999.0603
- Charlesworth, D., 2013. Plant sex chromosome evolution. *J Exp Bot* 64, 405–420. doi:10.1093/jxb/ers322
- Charlesworth, D., 2002. Plant sex determination and sex chromosomes. *Heredity* 88, 94–101. doi:10.1038/sj.hdy.6800016
- Charlesworth, D., Charlesworth, B., Marais, G., 2005. Steps in the evolution of heteromorphic sex chromosomes. *Heredity (Edinb)* 95, 118–128. doi:10.1038/sj.hdy.6800697
- Charlesworth, D., Mank, J.E., 2010. The Birds and the Bees and the Flowers and the Trees: Lessons from Genetic Mapping of Sex Determination in Plants and Animals. *Genetics* 186, 9–31. doi:10.1534/genetics.110.117697
- Charniaux-Cotton, H., 1962. Androgenic gland of crustaceans. *Gen. Comp. Endocrinol. Suppl* 1, 241–247.
- Charniaux-Cotton, H., 1959. Etude comparée du développement post-embryonnaire de l'appareil génital et de la glande androgène chez *Orchestia gammarella* et *Orchestia mediterranea* (Crustacés Amphipodes). Autodifférenciation ovarienne. *Bull. Soc. Zool. Fr* 84, 105–115.
- Charniaux-Cotton, H., Payen, G.G., Ginsburger-Vogel, T., 1992. Arthropoda-Crustacea: sexual differentiation., in: *Reproductive Biology of Invertebrates, Sexual Differentiation and Behaviour*. Adiyodi K. G., Adiyodi R. G., Chichester, NY, pp. 281–323.
- Charnier, M., 1965. Action of temperature on the sex ratio in the *Agama agama* (Agamidae, Lacertilia) embryo. *Comptes Rendus des Seances de la Societe de Biologie et de ses Filiales* 160, 620–622.
- Charnov, E.L., Bull, J., 1977. When is sex environmentally determined? *Nature* 266, 828–830. doi:10.1038/266828a0
- Charnov, E.L., Bull, J.J., Maynard Smith, J., 1976. Why be an hermaphrodite? *Nature* 263, 125–126. doi:10.1038/263125a0
- Chen, N., Bellott, D.W., Page, D.C., Clark, A.G., 2012. Identification of avian W-linked contigs by short-read sequencing. *BMC Genomics* 13, 183. doi:10.1186/1471-2164-13-183
- Chen, P., Xu, S.-L., Zhou, W., Guo, X.-G., Wang, C.-L., Wang, D.-L., Zhao, Y.-L., 2014. Cloning and expression analysis of a transformer gene in *Daphnia pulex* during

- different reproduction stages. *Anim. Reprod. Sci.* 146, 227–237. doi:10.1016/j.anireprosci.2014.03.010
- Chen, S., Zhang, G., Shao, C., Huang, Q., Liu, G., Zhang, P., Song, W., An, N., Chalopin, D., Volff, J.-N., Hong, Y., Li, Q., Sha, Z., Zhou, H., Xie, M., Yu, Q., Liu, Y., Xiang, H., Wang, N., Wu, K., Yang, C., Zhou, Q., Liao, X., Yang, L., Hu, Q., Zhang, J., Meng, L., Jin, L., Tian, Y., Lian, J., Yang, J., Miao, G., Liu, S., Liang, Z., Yan, F., Li, Y., Sun, B., Zhang, H., Zhang, J., Zhu, Y., Du, M., Zhao, Y., Scharl, M., Tang, Q., Wang, J., 2014. Whole-genome sequence of a flatfish provides insights into ZW sex chromosome evolution and adaptation to a benthic lifestyle. *Nat. Genet.* 46, 253–260. doi:10.1038/ng.2890
- Chevalier, F., Herbinière-Gaboreau, J., Charif, D., Mitta, G., Gavory, F., Wincker, P., Grève, P., Braquart-Varnier, C., Bouchon, D., 2012. Feminizing *Wolbachia*: a transcriptomics approach with insights on the immune response genes in *Armadillidium vulgare*. *BMC Microbiology* 12, S1. doi:10.1186/1471-2180-12-S1-S1
- Chibalina, M.V., Filatov, D.A., 2011. Plant Y chromosome degeneration is retarded by haploid purifying selection. *Curr. Biol.* 21, 1475–1479. doi:10.1016/j.cub.2011.07.045
- Chong, T., Iii, J.J.C., Brubacher, J.L., Zarkower, D., Newmark, P.A., 2013. A sex-specific transcription factor controls male identity in a simultaneous hermaphrodite. *Nature Communications* 4, ncomms2811. doi:10.1038/ncomms2811
- Chou, S., Daugherty, M.D., Peterson, S.B., Biboy, J., Yang, Y., Jutras, B.L., Fritz-Laylin, L.K., Ferrin, M.A., Harding, B.N., Jacobs-Wagner, C., 2015. Transferred interbacterial antagonism genes augment eukaryotic innate immune function. *Nature* 518, 98.
- Christodoulou, A., Lederer, C.W., Surrey, T., Vernos, I., Santama, N., 2006. Motor protein KIFC5A interacts with Nubp1 and Nubp2, and is implicated in the regulation of centrosome duplication. *J. Cell. Sci.* 119, 2035–2047. doi:10.1242/jcs.02922
- Coimbra, M.R.M., Kobayashi, K., Koretsugu, S., Hasegawa, O., Ohara, E., Ozaki, A., Sakamoto, T., Naruse, K., Okamoto, N., 2003. A genetic linkage map of the Japanese flounder, *Paralichthys olivaceus*. *Aquaculture* 220, 203–218. doi:10.1016/S0044-8486(02)00353-8
- Colbourne, J.K., Pfrender, M.E., Gilbert, D., Thomas, W.K., Tucker, A., Oakley, T.H., Tokishita, S., Aerts, A., Arnold, G.J., Basu, M.K., Bauer, D.J., Cáceres, C.E., Carmel, L., Casola, C., Choi, J.-H., Detter, J.C., Dong, Q., Dusheyko, S., Eads, B.D., Fröhlich, T., Geiler-Samerotte, K.A., Gerlach, D., Hatcher, P., Jogdeo, S., Krijgsveld, J., Kriventseva, E.V., Kültz, D., Laforsch, C., Lindquist, E., Lopez, J., Manak, J.R., Muller, J., Pangilinan, J., Patwardhan, R.P., Pitluck, S., Pritham, E.J., Rechtsteiner, A., Rho, M., Rogozin, I.B., Sakarya, O., Salamov, A., Schaack, S., Shapiro, H., Shiga, Y., Skalitzy, C., Smith, Z., Souvorov, A., Sung, W., Tang, Z., Tsuchiya, D., Tu, H., Vos, H., Wang, M., Wolf, Y.I., Yamagata, H., Yamada, T., Ye, Y., Shaw, J.R., Andrews, J., Crease, T.J., Tang, H., Lucas, S.M., Robertson, H.M., Bork, P., Koonin, E.V., Zdobnov, E.M., Grigoriev, I.V., Lynch, M., Boore, J.L., 2011. The ecoresponsive genome of *Daphnia pulex*. *Science* 331, 555–561. doi:10.1126/science.1197761
- Colegrave, N., Kaltz, O., Bell, G., 2002. The ecology and genetics of fitness in *Chlamydomonas*. VIII. The dynamics of adaptation to novel environments after a single episode of sex. *Evolution* 56, 14–21.
- Cooper, T.F., 2007. Recombination speeds adaptation by reducing competition between beneficial mutations in populations of *Escherichia coli*. *PLoS Biol.* 5, e225. doi:10.1371/journal.pbio.0050225

- Cordaux, R., Bouchon, D., Grève, P., 2011. The impact of endosymbionts on the evolution of host sex-determination mechanisms. *Trends Genet.* 27, 332–341. doi:10.1016/j.tig.2011.05.002
- Cordaux, R., Gilbert, C., 2017. Evolutionary Significance of *Wolbachia*-to-Animal Horizontal Gene Transfer: Female Sex Determination and the f Element in the Isopod *Armadillidium vulgare*. *Genes (Basel)* 8. doi:10.3390/genes8070186
- Cordaux, R., Michel-Salzat, A., Bouchon, D., 2001. *Wolbachia* infection in crustaceans: novel hosts and potential routes for horizontal transmission. *Journal of Evolutionary Biology* 14, 237–243. doi:10.1046/j.1420-9101.2001.00279.x
- Cordaux, R., Michel-Salzat, A., Frelon-Raimond, M., Rigaud, T., Bouchon, D., 2004. Evidence for a new feminizing *Wolbachia* strain in the isopod *Armadillidium vulgare*: evolutionary implications. *Heredity* 93, 78–84. doi:10.1038/sj.hdy.6800482
- Cordaux, R., Pichon, S., Hatira, H.B.A., Doublet, V., Grève, P., Marcadé, I., Braquart-Varnier, C., Souty-Grosset, C., Charfi-Cheikhrouha, F., Bouchon, D., 2012. Widespread *Wolbachia* infection in terrestrial isopods and other crustaceans. *Zookeys* 123–131. doi:10.3897/zookeys.176.2284
- Cortez, D., Marin, R., Toledo-Flores, D., Froidevaux, L., Liechti, A., Waters, P.D., Grützner, F., Kaessmann, H., 2014. Origins and functional evolution of Y chromosomes across mammals. *Nature* 508, 488–493. doi:10.1038/nature13151
- Coschigano, K.T., Wensink, P.C., 1993. Sex-specific transcriptional regulation by the male and female doublesex proteins of *Drosophila*. *Genes Dev.* 7, 42–54.
- Coyne, J.A., Orr, H.A., 2004. Speciation. Sinauer Associates, Sunderland (Massachusetts).
- Crain, D.A., Guillette, L.J., 1998. Reptiles as models of contaminant-induced endocrine disruption. *Anim. Reprod. Sci.* 53, 77–86.
- Crews, D., Bull, J.J., 2009. Mode and tempo in environmental sex determination in vertebrates. *Semin. Cell Dev. Biol.* 20, 251–255. doi:10.1016/j.semcdb.2009.02.004
- Cuvier, G., 1792. Mémoire sur les cloportes terrestres. *Journal d’Histoire Naturelle* 2, 18–31.
- Danchin, E.G.J., Rosso, M.-N., Vieira, P., Almeida-Engler, J. de, Coutinho, P.M., Henrissat, B., Abad, P., 2010. Multiple lateral gene transfers and duplications have promoted plant parasitism ability in nematodes. *PNAS* 107, 17651–17656. doi:10.1073/pnas.1008486107
- Darby, A.C., Armstrong, S.D., Bah, G.S., Kaur, G., Hughes, M.A., Kay, S.M., Koldkjær, P., Rainbow, L., Radford, A.D., Blaxter, M.L., Tanya, V.N., Trees, A.J., Cordaux, R., Wastling, J.M., Makepeace, B.L., 2012. Analysis of gene expression from the *Wolbachia* genome of a filarial nematode supports both metabolic and defensive roles within the symbiosis. *Genome Res* 22, 2467–2477. doi:10.1101/gr.138420.112
- Darriba, D., Taboada, G.L., Doallo, R., Posada, D., 2011. ProtTest 3: fast selection of best-fit models of protein evolution. *Bioinformatics* 27, 1164–1165. doi:10.1093/bioinformatics/btr088
- Dauwalder, B., Amaya-Manzanares, F., Mattox, W., 1996. A human homologue of the *Drosophila* sex determination factor transformer-2 has conserved splicing regulatory functions. *Proc. Natl. Acad. Sci. U.S.A.* 93, 9004–9009.
- Dedeine, F., Vavre, F., Fleury, F., Loppin, B., Hochberg, M.E., Boulétreau, M., 2001. Removing symbiotic *Wolbachia* bacteria specifically inhibits oogenesis in a parasitic wasp. *PNAS* 98, 6247–6252. doi:10.1073/pnas.101304298
- Denton, J.F., Lugo-Martinez, J., Tucker, A.E., Schrider, D.R., Warren, W.C., Hahn, M.W., 2014. Extensive Error in the Number of Genes Inferred from Draft Genome Assemblies. *PLOS Computational Biology* 10, e1003998. doi:10.1371/journal.pcbi.1003998

- Devlin, R.H., Nagahama, Y., 2002. Sex determination and sex differentiation in fish: an overview of genetic, physiological, and environmental influences. *Aquaculture* 208, 191–364.
- Dittmer, J., Beltran-Bech, S., Lesobre, J., Raimond, M., Johnson, M., Bouchon, D., 2014. Host tissues as microhabitats for *Wolbachia* and quantitative insights into the bacterial community in terrestrial isopods. *Mol Ecol* 23, 2619–2635. doi:10.1111/mec.12760
- Dobson, S.L., Bourtzis, K., Braig, H.R., Jones, B.F., Zhou, W., Rousset, F., O'Neill, S.L., 1999. *Wolbachia* infections are distributed throughout insect somatic and germ line tissues. *Insect Biochem. Mol. Biol.* 29, 153–160.
- Doležel, J., Vrána, J., Šafář, J., Bartoš, J., Kubaláková, M., Šimková, H., 2012. Chromosomes in the flow to simplify genome analysis. *Funct Integr Genomics* 12, 397–416. doi:10.1007/s10142-012-0293-0
- Dufresnes, C., Borzée, A., Horn, A., Stöck, M., Ostini, M., Sermier, R., Wassef, J., Litvinchuck, S.N., Kosch, T.A., Waldman, B., Jang, Y., Brelsford, A., Perrin, N., 2015. Sex-Chromosome Homomorphy in Palearctic Tree Frogs Results from Both Turnovers and X-Y Recombination. *Mol. Biol. Evol.* 32, 2328–2337. doi:10.1093/molbev/msv113
- Dunning Hotopp, J.C., 2011. Horizontal gene transfer between bacteria and animals. *Trends Genet.* 27, 157–163. doi:10.1016/j.tig.2011.01.005
- Dunning Hotopp, J.C., Clark, M.E., Oliveira, D.C.S.G., Foster, J.M., Fischer, P., Muñoz Torres, M.C., Giebel, J.D., Kumar, N., Ishmael, N., Wang, S., Ingram, J., Nene, R.V., Shepard, J., Tomkins, J., Richards, S., Spiro, D.J., Ghedin, E., Slatko, B.E., Tettelin, H., Werren, J.H., 2007. Widespread lateral gene transfer from intracellular bacteria to multicellular eukaryotes. *Science* 317, 1753–1756. doi:10.1126/science.1142490
- Dupeyron, M., Leclercq, S., Cerveau, N., Bouchon, D., Gilbert, C., 2014. Horizontal transfer of transposons between and within crustaceans and insects. *Mobile DNA* 5, 4. doi:10.1186/1759-8753-5-4
- Duron, O., Bouchon, D., Boutin, S., Bellamy, L., Zhou, L., Engelstädter, J., Hurst, G.D., 2008. The diversity of reproductive parasites among arthropods: *Wolbachia* do not walk alone. *BMC Biol.* 6, 27. doi:10.1186/1741-7007-6-27
- Edwards, C.A., Bohlen, P.J., 1996. *Biology and Ecology of Earthworms*. Springer Science & Business Media.
- Ellegren, H., Hultin-Rosenberg, L., Brunström, B., Dencker, L., Kultima, K., Scholz, B., 2007. Faced with inequality: chicken do not have a general dosage compensation of sex-linked genes. *BMC Biology* 5, 40. doi:10.1186/1741-7007-5-40
- Elnagdy, S., Majerus, M. e. n., Handley, L.-J.L., 2011. The value of an egg: resource reallocation in ladybirds (Coleoptera: Coccinellidae) infected with male-killing bacteria. *Journal of Evolutionary Biology* 24, 2164–2172. doi:10.1111/j.1420-9101.2011.02346.x
- Engelstädter, J., Telschow, A., 2009. Cytoplasmic incompatibility and host population structure. *Heredity* 103, 196–207. doi:10.1038/hdy.2009.53
- English, A.C., Richards, S., Han, Y., Wang, M., Vee, V., Qu, J., Qin, X., Muzny, D.M., Reid, J.G., Worley, K.C., Gibbs, R.A., 2012. Mind the Gap: Upgrading Genomes with Pacific Biosciences RS Long-Read Sequencing Technology. *PLOS ONE* 7, e47768. doi:10.1371/journal.pone.0047768
- Eppley, S.M., Jesson, L.K., 2008. Moving to mate: the evolution of separate and combined sexes in multicellular organisms. *J. Evol. Biol.* 21, 727–736. doi:10.1111/j.1420-9101.2008.01524.x
- Erdman, S.E., Burtis, K.C., 1993. The *Drosophila* doublesex proteins share a novel zinc finger related DNA binding domain. *EMBO J.* 12, 527–535.

- Ezaz, T., Quinn, A.E., Sarre, S.D., O’Meally, D., Georges, A., Graves, J.A.M., 2009. Molecular marker suggests rapid changes of sex-determining mechanisms in Australian dragon lizards. *Chromosome Res.* 17, 91–98. doi:10.1007/s10577-008-9019-5
- Ezaz, T., Stiglec, R., Veyrunes, F., Graves, J.A.M., 2006. Relationships between Vertebrate ZW and XY Sex Chromosome Systems. *Current Biology* 16, R736–R743. doi:10.1016/j.cub.2006.08.021
- Fautin, D.G., Allen, G.R., 1992. Field guide to anemonefishes and their host sea anemones /. Western Australian Museum, Perth (Australia).
- Ferdy, J.-B., Liu, N., Sicard, M., 2016. Transmission modes and the evolution of feminizing symbionts. *J. Evol. Biol.* 29, 2395–2409. doi:10.1111/jeb.12963
- Ferguson, M.W.J., Joanen, T., 1982. Temperature of egg incubation determines sex in *Alligator mississippiensis*. *Nature* 296, 850–853. doi:10.1038/296850a0
- Ferris, P., Olson, B.J.S.C., De Hoff, P.L., Douglass, S., Casero, D., Prochnik, S., Geng, S., Rai, R., Grimwood, J., Schmutz, J., Nishii, I., Hamaji, T., Nozaki, H., Pellegrini, M., Umen, J.G., 2010. Evolution of an expanded sex-determining locus in *Volvox*. *Science* 328, 351–354. doi:10.1126/science.1186222
- Feschotte, C., Gilbert, C., 2012. Endogenous viruses: insights into viral evolution and impact on host biology. *Nat Rev Genet* 13, 283–296. doi:10.1038/nrg3199
- Fisher, R.A., 1930. The Genetical Theory of Natural Selection: A Complete Variorum Edition. OUP Oxford.
- Folmer, O., Black, M., Hoeh, W., Lutz, R., Vrijenhoek, R., 1994. DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Mol. Marine Biol. Biotechnol.* 3, 294–299.
- Foster, J., Ganatra, M., Kamal, I., Ware, J., Makarova, K., Ivanova, N., Bhattacharyya, A., Kapatral, V., Kumar, S., Posfai, J., Vincze, T., Ingram, J., Moran, L., Lapidus, A., Omelchenko, M., Kyrpides, N., Ghedin, E., Wang, S., Goltsman, E., Joukov, V., Ostrovskaya, O., Tsukerman, K., Mazur, M., Comb, D., Koonin, E., Slatko, B., 2005. The *Wolbachia* Genome of *Brugia malayi*: Endosymbiont Evolution within a Human Pathogenic Nematode. *PLOS Biology* 3, e121. doi:10.1371/journal.pbio.0030121
- Fujii, T., Shimada, T., 2007. Sex determination in the silkworm, *Bombyx mori*: a female determinant on the W chromosome and the sex-determining gene cascade. *Semin. Cell Dev. Biol.* 18, 379–388. doi:10.1016/j.semcdb.2007.02.008
- Fukui, T., Kawamoto, M., Shoji, K., Kiuchi, T., Sugano, S., Shimada, T., Suzuki, Y., Katsuma, S., 2015. The Endosymbiotic Bacterium *Wolbachia* Selectively Kills Male Hosts by Targeting the Masculinizing Gene. *PLOS Pathogens* 11, e1005048. doi:10.1371/journal.ppat.1005048
- Garcia-Vallvé, S., Romeu, A., Palau, J., 2000. Horizontal gene transfer in bacterial and archaeal complete genomes. *Genome Res.* 10, 1719–1725.
- Gautier, M., 2014. Using genotyping data to assign markers to their chromosome type and to infer the sex of individuals: a Bayesian model-based classifier. *Molecular ecology resources* 14, 1141–1159.
- Geddes, P., Thompson, J.A., 1890. Geddes and Thompson on the Evolution of Sex. *The American Naturalist* 24, 754–760.
- Gempe, T., Beye, M., 2011. Function and evolution of sex determination mechanisms, genes and pathways in insects. *Bioessays* 33, 52–60. doi:10.1002/bies.201000043
- Genty, L.-M., Bouchon, D., Raimond, M., Bertaux, J., 2014. *Wolbachia* infect ovaries in the course of their maturation: last minute passengers and priority travellers? *PLoS ONE* 9, e94577. doi:10.1371/journal.pone.0094577

- Geuverink, E., Beukeboom, L.W., 2014. Phylogenetic distribution and evolutionary dynamics of the sex determination genes *doublesex* and *transformer* in insects. *Sex Dev* 8, 38–49. doi:10.1159/000357056
- Ghenu, A.-H., Bolker, B.M., Melnick, D.J., Evans, B.J., 2016. Multicopy gene family evolution on primate Y chromosomes. *BMC Genomics* 17. doi:10.1186/s12864-015-2187-8
- Ghiselin, M.T., 1969. The Evolution of Hermaphroditism Among Animals. *The Quarterly Review of Biology* 44, 189–208. doi:10.1086/406066
- Glawe, G.A., de Jong, T.J., 2009. Complex sex determination in the stinging nettle *Urtica dioica*. *Evol Ecol* 23, 635. doi:10.1007/s10682-008-9261-5
- Glawe, G.A., de Jong, T.J., 2007. Inheritance of progeny sex ratio in *Urtica dioica*. *J. Evol. Biol.* 20, 133–140. doi:10.1111/j.1420-9101.2006.01215.x
- Glennier, H., Thomsen, P.F., Hebsgaard, M.B., Sørensen, M.V., Willerslev, E., 2006. The Origin of Insects. *Science* 314, 1883–1884. doi:10.1126/science.1129844
- Goddard, M.R., Godfray, H.C.J., Burt, A., 2005. Sex increases the efficacy of natural selection in experimental yeast populations. *Nature* 434, 636–640. doi:10.1038/nature03405
- Goffeau, A., Barrell, B.G., Bussey, H., Davis, R.W., Dujon, B., Feldmann, H., Galibert, F., Hoheisel, J.D., Jacq, C., Johnston, M., Louis, E.J., Mewes, H.W., Murakami, Y., Philippsen, P., Tettelin, H., Oliver, S.G., 1996. Life with 6000 genes. *Science* 274, 546, 563–567.
- Goodwin, S., McPherson, J.D., McCombie, W.R., 2016. Coming of age: ten years of next-generation sequencing technologies. *Nat. Rev. Genet.* 17, 333–351. doi:10.1038/nrg.2016.49
- Gordon, H., 1979. Ancient ideas about sex differentiation, in: *Genetic Mechanisms of Sexual Development*. Vallet, H.L. & Porter, I.H, New York (USA), pp. 1–32.
- Gottlieb, Y., Zchori-Fein, E., Werren, J.H., Karr, T.L., 2002. Diploidy restoration in *Wolbachia*-infected *Muscidifurax uniraptor* (Hymenoptera: Pteromalidae). *Journal of Invertebrate Pathology* 81, 166–174. doi:10.1016/S0022-2011(02)00149-0
- Graves, J.A.M., 2008. Weird animal genomes and the evolution of vertebrate sex and sex chromosomes. *Annu. Rev. Genet.* 42, 565–586. doi:10.1146/annurev.genet.42.110807.091714
- Graves, J.A.M., 2006. Sex Chromosome Specialization and Degeneration in Mammals. *Cell* 124, 901–914. doi:10.1016/j.cell.2006.02.024
- Graves, M.J.A., Shetty, S., 2001. Sex from W to Z: evolution of vertebrate sex chromosomes and sex determining genes. *J. Exp. Zool.* 290, 449–462.
- Green, D.M., 1988. Cytogenetics of the endemic New Zealand frog, *Leiopelma hochstetteri*: extraordinary supernumerary chromosome variation and a unique sex-chromosome system. *Chromosoma* 97, 55–70.
- Grève, P., Braquart-Varnier, C., Strub, J.-M., Félix, C., Van Dorsselaer, A., Martin, G., 2004. The glycosylated androgenic hormone of the terrestrial isopod *Porcellio scaber* (Crustacea). *Gen. Comp. Endocrinol.* 136, 389–397. doi:10.1016/j.ygcn.2004.01.015
- Grimaldi, D., Engel, M.S., 2005. *Evolution of the Insects*. Cambridge University Press.
- Gupta, V., Vasanthakrishnan, R.B., Siva-Jothy, J., Monteith, K.M., Brown, S.P., Vale, P.F., 2017. The route of infection determines *Wolbachia* antibacterial protection in *Drosophila*. *Proc. R. Soc. B* 284, 20170809. doi:10.1098/rspb.2017.0809
- Haag, C.R., McTaggart, S.J., Didier, A., Little, T.J., Charlesworth, D., 2009. Nucleotide Polymorphism and Within-Gene Recombination in *Daphnia magna* and *D. pulex*, Two Cyclical Parthenogens. *Genetics* 182, 313–323. doi:10.1534/genetics.109.101147

- Haag, E.S., Doty, A.V., 2005. Sex Determination across Evolution: Connecting the Dots. PLOS Biology 3, e21. doi:10.1371/journal.pbio.0030021
- Hall, A.B., Papathanos, P.-A., Sharma, A., Cheng, C., Akbari, O.S., Assour, L., Bergman, N.H., Cagnetti, A., Crisanti, A., Dottorini, T., Fiorentini, E., Galizi, R., Hnath, J., Jiang, X., Koren, S., Nolan, T., Radune, D., Sharakhova, M.V., Steele, A., Timoshevskiy, V.A., Windbichler, N., Zhang, S., Hahn, M.W., Phillippy, A.M., Emrich, S.J., Sharakhov, I.V., Tu, Z.J., Besansky, N.J., 2016. Radical remodeling of the Y chromosome in a recent radiation of malaria mosquitoes. Proc. Natl. Acad. Sci. U.S.A. 113, E2114-2123. doi:10.1073/pnas.1525164113
- Hall, A.B., Qi, Y., Timoshevskiy, V., Sharakhova, M.V., Sharakhov, I.V., Tu, Z., 2013. Six novel Y chromosome genes in *Anopheles* mosquitoes discovered by independently sequencing males and females. BMC Genomics 14, 273. doi:10.1186/1471-2164-14-273
- Hall, A.B., Timoshevskiy, V.A., Sharakhova, M.V., Jiang, X., Basu, S., Anderson, M.A.E., Hu, W., Sharakhov, I.V., Adelman, Z.N., Tu, Z., 2014. Insights into the preservation of the homomorphic sex-determining chromosome of *Aedes aegypti* from the discovery of a male-biased gene tightly linked to the M-locus. Genome Biol Evol 6, 179–191. doi:10.1093/gbe/evu002
- Hallast, P., Delser, P.M., Batini, C., Zadik, D., Rocchi, M., Schempp, W., Tyler-Smith, C., Jobling, M.A., 2016. Great ape Y Chromosome and mitochondrial DNA phylogenies reflect subspecies structure and patterns of mating and dispersal. Genome Res. 26, 427–439. doi:10.1101/gr.198754.115
- Heath, B.D., Butcher, R.D.J., Whitfield, W.G.F., Hubbard, S.F., 1999. Horizontal transfer of *Wolbachia* between phylogenetically distant insect species by a naturally occurring mechanism. Current Biology 9, 313–316. doi:10.1016/S0960-9822(99)80139-0
- Hedges, L.M., Brownlie, J.C., O'Neill, S.L., Johnson, K.N., 2008. *Wolbachia* and Virus Protection in Insects. Science 322, 702–702. doi:10.1126/science.1162418
- Hediger, M., Henggeler, C., Meier, N., Perez, R., Saccone, G., Bopp, D., 2010. Molecular characterization of the key switch F provides a basis for understanding the rapid divergence of the sex-determining pathway in the housefly. Genetics 184, 155–170. doi:10.1534/genetics.109.109249
- Hedley, M.L., Maniatis, T., 1991. Sex-specific splicing and polyadenylation of dsx pre-mRNA requires a sequence that binds specifically to tra-2 protein in vitro. Cell 65, 579–586.
- Hertel, K.J., Lynch, K.W., Hsiao, E.C., Liu, E.H., Maniatis, T., 1996. Structural and functional conservation of the *Drosophila* doublesex splicing enhancer repeat elements. RNA 2, 969–981.
- Hertig, M., 1936. The *Rickettsia*, *Wolbachia pipientis* (gen. et sp.n.) and Associated Inclusions of the Mosquito, *Culex pipiens*. Parasitology 28, 453–486. doi:10.1017/S0031182000022666
- Hertig, M., Wolbach, S.B., 1924. Studies on *Rickettsia*-Like Micro-Organisms in Insects. J Med Res 44, 329–374.7.
- Hessler, R.R., Elofsson, R., Hessler, A.Y., 1995. Reproductive System of *Hutchinsoniella macracantha* (Cephalocarida). Journal of Crustacean Biology 15, 493–522. doi:10.2307/1548771
- Hildreth, P.E., 1965. Doublesex, a Recessive Gene That Transforms Both Males and Females of *Drosophila* into Intersexes. Genetics 51, 659–678.
- Hilgenboecker, K., Hammerstein, P., Schlattmann, P., Telschow, A., Werren, J.H., 2008. How many species are infected with *Wolbachia*?--A statistical analysis of current data. FEMS Microbiol. Lett. 281, 215–220. doi:10.1111/j.1574-6968.2008.01110.x

- Hiroki, M., Kato, Y., Kamito, T., Miura, K., 2002. Feminization of genetic males by a symbiotic bacterium in a butterfly, *Eurema hecabe* (Lepidoptera: Pieridae). *Naturwissenschaften* 89, 167–170.
- Hoegg, S., Brinkmann, H., Taylor, J.S., Meyer, A., 2004. Phylogenetic timing of the fish-specific genome duplication correlates with the diversification of teleost fish. *J. Mol. Evol.* 59, 190–203. doi:10.1007/s00239-004-2613-z
- Holt, C., Yandell, M., 2011. MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. *BMC Bioinformatics* 12, 491. doi:10.1186/1471-2105-12-491
- Holub, E.B., 2001. The arms race is ancient history in *Arabidopsis*, the wildflower. *Nat. Rev. Genet.* 2, 516–527. doi:10.1038/35080508
- Hornung, E., 2011. Evolutionary adaptation of oniscidean isopods to terrestrial life: Structure, physiology and behavior. *Terrestrial Arthropod Reviews* 4, 95–130. doi:10.1163/187498311X576262
- Hoshijima, K., Inoue, K., Higuchi, I., Sakamoto, H., Shimura, Y., 1991. Control of doublesex alternative splicing by transformer and transformer-2 in *Drosophila*. *Science* 252, 833–836.
- Hoskins, R.A., Carlson, J.W., Kennedy, C., Acevedo, D., Evans-Holm, M., Frise, E., Wan, K.H., Park, S., Mendez-Lago, M., Rossi, F., Villasante, A., Dimitri, P., Karpen, G.H., Celniker, S.E., 2007. Sequence finishing and mapping of *Drosophila melanogaster* heterochromatin. *Science* 316, 1625–1628. doi:10.1126/science.1139816
- Hosokawa, T., Koga, R., Kikuchi, Y., Meng, X.-Y., Fukatsu, T., 2010. Wolbachia as a bacteriocyte-associated nutritional mutualist. *PNAS* 107, 769–774. doi:10.1073/pnas.0911476107
- Huigens, M., Stouthamer, R., 2003. Parthenogenesis Associated With *Wolbachia*, in: *Insect Symbiosis, Contemporary Topics in Entomology*. CRC Press, pp. 247–266. doi:10.1201/9780203009918.ch15
- Huigens, M.E., Almeida, R.P. de, Boons, P. a. H., Luck, R.F., Stouthamer, R., 2004. Natural interspecific and intraspecific horizontal transfer of parthenogenesis-inducing *Wolbachia* in *Trichogramma* wasps. *Proceedings of the Royal Society of London B: Biological Sciences* 271, 509–515. doi:10.1098/rspb.2003.2640
- Hurst, G., Jiggins, F., Majerus, M., 2003. Inherited Microorganisms That Selectively Kill Male Hosts, in: *Insect Symbiosis, Contemporary Topics in Entomology*. CRC Press, pp. 177–198. doi:10.1201/9780203009918.ch12
- Hurst, G.D., Majerus, M.E., 1993. Why do maternally inherited microorganisms kill males? *Heredity* 71, 81–95.
- Huson, D.H., Mitra, S., Ruscheweyh, H.-J., Weber, N., Schuster, S.C., 2011. Integrative analysis of environmental sequences using MEGAN4. *Genome Res.* 21, 1552–1560. doi:10.1101/gr.120618.111
- Isberg, S.R., Johnston, S.M., Chen, Y., Moran, C., 2006. First evidence of higher female recombination in a species with temperature-dependent sex determination: the saltwater crocodile. *Journal of heredity* 97, 599–602.
- Itoh, Y., Melamed, E., Yang, X., Kampf, K., Wang, S., Yehya, N., Van Nas, A., Replogle, K., Band, M.R., Clayton, D.F., Schadt, E.E., Lusi, A.J., Arnold, A.P., 2007. Dosage compensation is less effective in birds than in mammals. *J. Biol.* 6, 2. doi:10.1186/jbiol53
- Jablonka, E., Lamb, M.J., 1988. Meiotic pairing constraints and the activity of sex chromosomes. *J. Theor. Biol.* 133, 23–36.
- Jaenike, J., 2012. Population genetics of beneficial heritable symbionts. *Trends in Ecology & Evolution* 27, 226–232. doi:10.1016/j.tree.2011.10.005

- Jaenike, J., 2001. Sex Chromosome Meiotic Drive. *Annual Review of Ecology and Systematics* 32, 25–49. doi:10.1146/annurev.ecolsys.32.081501.113958
- Jaenike, J., Dyer, K.A., Cornish, C., Minhas, M.S., 2006. Asymmetrical Reinforcement and *Wolbachia* Infection in *Drosophila*. *PLOS Biology* 4, e325. doi:10.1371/journal.pbio.0040325
- Janssens, F.A., 1909. La théorie de la chiasmotypie: Nouvelle interpretation des cinèses de maturation. Van In.
- Jarne, P., Auld, J.R., 2006. Animals mix it up too: the distribution of self-fertilization among hermaphroditic animals. *Evolution* 60, 1816–1824. doi:10.1554/06-246.1
- Jiang, P.-P., Hartl, D.L., Lemos, B., 2010. Y Not a Dead End: Epistatic Interactions Between Y-Linked Regulatory Polymorphisms and Genetic Background Affect Global Gene Expression in *Drosophila melanogaster*. *Genetics* 186, 109–118. doi:10.1534/genetics.110.118109
- Johnson, N.S., Swink, W.D., Brenden, T.O., 2017. Field study suggests that sex determination in sea lamprey is directly influenced by larval growth rate. *Proc. Biol. Sci.* 284. doi:10.1098/rspb.2017.0262
- Juchault, P., 1999. Hermaphroditism and gonochorism. a new hypothesis on the evolution of sexuality in crustacea. *Comptes Rendus de l'Académie des Sciences - Series III - Sciences de la Vie* 322, 423–427. doi:10.1016/S0764-4469(99)80078-X
- Juchault, P., 1966. Contribution à l'étude de la différenciation sexuelle mâle chez les Crustacés Isopodes (Thèse). University of Poitiers, Poitiers, France.
- Juchault, P., Legrand, J.J., 1989. Sex determination and monogeny in terrestrial isopods *Armadillidium vulgare* (Latreille, 1804) and *Armadillidium nasatum* Budde-Lund, 1885. *Monografia. Monitore zoologico italiano* 4, 359–375.
- Juchault, P., Legrand, J.J., 1985. Contribution à l'étude du mécanisme de l'état réfractaire à l'hormone androgène chez les *Armadillidium vulgare* Latr, (crustacé, isopode, oniscoïde) hébergeant une bactérie féminisante. *General and Comparative Endocrinology* 60, 463–467. doi:10.1016/0016-6480(85)90082-6
- Juchault, P., Legrand, J.-J., 1979. Analyse génétique et physiologique de la détermination du sexe dans une population du Crustacé Isopode Oniscoïde *Armadillidium nasatum*. *Arch. Zool. exp. gén* 120, 25–43.
- Juchault, P., Legrand, J.-J., 1972. Croisements de néo-mâles expérimentaux chez *Armadillidium vulgare* Latr. (Crustacé Isopode Oniscoïde). Mise en évidence d'une hétérogamétie femelle. *C. R. Acad. Sc. Paris* 1387–1389, 274–276.
- Juchault, P., Legrand, J.-J., 1964. Démonstration de l'homogamétie femelle par croisement de deux femelles génétiques, chez les Oniscoïdes *Porcellio dilatatus* et *Helleria brevicornis*. *Comptes Rendus de l'Académie des Sciences de Paris* 258, 2685–2686.
- Juchault, P., Legrand, J.-J., Mocquard, J.-P., 1980. Contribution à l'étude qualitative et quantitative des facteurs contrôlant le sexe dans les populations du Crustacé Isopode terrestre *Armadillidium vulgare* Latr. I. La population de Niort (Deux-Sèvres). *Arch. Zool. Exp. Gén* 121, 3–27.
- Juchault, P., Mocquard, J.P., 1993. Transfer of a parasitic sex factor to the nuclear genome of the host: A hypothesis on the evolution of sex-determining mechanisms in the terrestrial Isopod *Armadillidium vulgare* Latr. *Journal of Evolutionary Biology* 6, 511–528. doi:10.1046/j.1420-9101.1993.6040511.x
- Juchault, P., Rigaud, T., 1995. Evidence for female heterogamety in two terrestrial crustaceans and the problem of sex chromosome evolution in isopods. *Heredity* 75, 466–471.

- Juchault, P., Rigaud, T., Mocquard, J.P., 1993. Evolution of sex determination and sex ratio variability in wild populations of *Armadillidium vulgare* (Latr.) (Crustacea, Isopoda): a case study in conflict resolution. *Acta Oecologica* 14, 547–562.
- Juchault, P., Rigaud, T., Mocquard, J.-P., 1992. Evolution of sex-determining mechanisms in a wild population of *Armadillidium vulgare* Latr. (Crustacea, Isopoda): competition between two feminizing parasitic sex factors. *Heredity* 69, 382–390.
- Jurka, J., Kapitonov, V.V., Pavlicek, A., Klonowski, P., Kohany, O., Walichiewicz, J., 2005. Repbase Update, a database of eukaryotic repetitive elements. *Cytogenet. Genome Res.* 110, 462–467. doi:10.1159/000084979
- Kageyama, D., Narita, S., Watanabe, M., 2012. Insect Sex Determination Manipulated by Their Endosymbionts: Incidences, Mechanisms and Implications. *Insects* 3, 161–199. doi:10.3390/insects3010161
- Kamiya, T., Kai, W., Tasumi, S., Oka, A., Matsunaga, T., Mizuno, N., Fujita, M., Suetake, H., Suzuki, S., Hosoya, S., Tohari, S., Brenner, S., Miyadai, T., Venkatesh, B., Suzuki, Y., Kikuchi, K., 2012. A Trans-Species Missense SNP in Amhr2 Is Associated with Sex Determination in the Tiger Pufferfish, *Takifugu rubripes* (Fugu). *PLOS Genetics* 8, e1002798. doi:10.1371/journal.pgen.1002798
- Kao, D., Lai, A.G., Stamataki, E., Rosic, S., Konstantinides, N., Jarvis, E., Di Donfrancesco, A., Pouchkina-Stancheva, N., Semon, M., Grillo, M., Bruce, H., Kumar, S., Siwanowicz, I., Le, A., Lemire, A., Eisen, M.B., Extavour, C., Browne, W.E., Wolff, C., Averof, M., Patel, N.H., Sarkies, P., Pavlopoulos, A., Aboobaker, A., 2016. The genome of the crustacean *Parhyale hawaiiensis*, a model for animal development, regeneration, immunity and lignocellulose digestion. *Elife* 5. doi:10.7554/eLife.20062
- Katakura, Y., 1984. Sex differentiation and androgenic gland hormone in the terrestrial isopod *Armadillidium vulgare*, in: *Symposia of the Zoological Society of London. Presented at the Biology (The) of terrestrial isopods. Symposium*, Cambridge University Press, pp. 127–142.
- Kato, Y., Kobayashi, K., Oda, S., Colbourn, J.K., Tatarazako, N., Watanabe, H., Iguchi, T., 2008. Molecular cloning and sexually dimorphic expression of DM-domain genes in *Daphnia magna*. *Genomics* 91, 94–101. doi:10.1016/j.ygeno.2007.09.002
- Kato, Y., Kobayashi, K., Oda, S., Tatarazako, N., Watanabe, H., Iguchi, T., 2010. Sequence divergence and expression of a transformer gene in the branchiopod crustacean, *Daphnia magna*. *Genomics* 95, 160–165. doi:10.1016/j.ygeno.2009.12.005
- Kato, Y., Kobayashi, K., Watanabe, H., Iguchi, T., 2011. Environmental sex determination in the branchiopod crustacean *Daphnia magna*: deep conservation of a Doublesex gene in the sex-determining pathway. *PLoS Genet.* 7, e1001345. doi:10.1371/journal.pgen.1001345
- 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 Res* 30, 3059–3066.
- Katzourakis, A., Gifford, R.J., 2010. Endogenous Viral Elements in Animal Genomes. *PLOS Genetics* 6, e1001191. doi:10.1371/journal.pgen.1001191
- Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Meintjes, P., Drummond, A., 2012. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28, 1647–1649. doi:10.1093/bioinformatics/bts199
- Keeling, P.J., Palmer, J.D., 2008. Horizontal gene transfer in eukaryotic evolution. *Nat. Rev. Genet.* 9, 605–618. doi:10.1038/nrg2386

- Kennedy, R.C., Unger, M.F., Christley, S., Collins, F.H., Madey, G.R., 2011. An automated homology-based approach for identifying transposable elements. BMC Bioinformatics 12, 130. doi:10.1186/1471-2105-12-130
- Kenny, N.J., Sin, Y.W., Shen, X., Zhe, Q., Wang, W., Chan, T.F., Tobe, S.S., Shimeld, S.M., Chu, K.H., Hui, J.H.L., 2014. Genomic Sequence and Experimental Tractability of a New Decapod Shrimp Model, *Neocaridina denticulata*. Mar Drugs 12, 1419–1437. doi:10.3390/md12031419
- Kent, W.J., 2002. BLAT—The BLAST-Like Alignment Tool. Genome Res. 12, 656–664. doi:10.1101/gr.229202
- Kikuchi, K., Kai, W., Hosokawa, A., Mizuno, N., Suetake, H., Asahina, K., Suzuki, Y., 2007. The Sex-Determining Locus in the Tiger Pufferfish, *Takifugu rubripes*. Genetics 175, 2039–2042. doi:10.1534/genetics.106.069278
- Kim, S., Kettlewell, J.R., Anderson, R.C., Bardwell, V.J., Zarkower, D., 2003. Sexually dimorphic expression of multiple doublesex-related genes in the embryonic mouse gonad. Gene Expression Patterns 3, 77–82. doi:10.1016/S1567-133X(02)00071-6
- Kim, S., Yoo, M.-J., Albert, V.A., Farris, J.S., Soltis, P.S., Soltis, D.E., 2004. Phylogeny and diversification of B-function MADS-box genes in angiosperms: evolutionary and functional implications of a 260-million-year-old duplication. Am. J. Bot. 91, 2102–2118. doi:10.3732/ajb.91.12.2102
- Kimura, M., 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J Mol Evol 16, 111–120. doi:10.1007/BF01731581
- Kitano, J., Ross, J.A., Mori, S., Kume, M., Jones, F.C., Chan, Y.F., Absher, D.M., Grimwood, J., Schmutz, J., Myers, R.M., Kingsley, D.M., Peichel, C.L., 2009. A role for a neo-sex chromosome in stickleback speciation. Nature 461, 1079–1083. doi:10.1038/nature08441
- Kocher, T.D., Thomas, W.K., Meyer, A., Edwards, S.V., Pääbo, S., Villablanca, F.X., Wilson, A.C., 1989. Dynamics of mitochondrial DNA evolution in animals: amplification and sequencing with conserved primers. Proc Natl Acad Sci U S A 86, 6196–6200.
- Koerich, L.B., Dupim, E.G., Faria, L.L., Dias, F.A., Dias, A.F., Trindade, G.S., Mesquita, R.D., Carvalho, A.B., 2016. First report of Y-linked genes in the kissing bug *Rhodnius prolixus*. BMC Genomics 17, 100. doi:10.1186/s12864-016-2425-8
- Kondo, M., Nagao, E., Mitani, H., Shima, A., 2001. Differences in recombination frequencies during female and male meioses of the sex chromosomes of the medaka, *Oryzias latipes*. Genetics Research 78, 23–30.
- Kondrashov, A.S., 1988. Deleterious mutations and the evolution of sexual reproduction. Nature 336, 435–440. doi:10.1038/336435a0
- Kong, L., Lv, W., Huang, Y., Liu, Z., Yang, Y., Zhao, Y., 2015. Cloning, expression and localization of the *Daphnia carinata* transformer gene DcarTra during different reproductive stages. Gene 566, 248–256. doi:10.1016/j.gene.2015.04.057
- Koopman, P., Gubbay, J., Vivian, N., Goodfellow, P., Lovell-Badge, R., 1991. Male development of chromosomally female mice transgenic for Sry. Nature 351, 117–121. doi:10.1038/351117a0
- Korf, I., 2004. Gene finding in novel genomes. BMC Bioinformatics 5, 59. doi:10.1186/1471-2105-5-59
- Korpelainen, H., 1990. Sex Ratios and Conditions Required for Environmental Sex Determination in Animals. Biological Reviews 65, 147–184. doi:10.1111/j.1469-185X.1990.tb01187.x

- Kremer, N., Voronin, D., Charif, D., Mavingui, P., Mollereau, B., Vavre, F., 2009. *Wolbachia* Interferes with Ferritin Expression and Iron Metabolism in Insects. *PLOS Pathogens* 5, e1000630. doi:10.1371/journal.ppat.1000630
- Kumar, S., Jones, M., Koutsovoulos, G., Clarke, M., Blaxter, M., 2013. Blobology: exploring raw genome data for contaminants, symbionts and parasites using taxon-annotated GC-coverage plots. *Front. Genet.* 4. doi:10.3389/fgene.2013.00237
- Kunte, K., Zhang, W., Tenger-Trolander, A., Palmer, D.H., Martin, A., Reed, R.D., Mullen, S.P., Kronforst, M.R., 2014. doublesex is a mimicry supergene. *Nature* 507, 229–232. doi:10.1038/nature13112
- Kussakin, O.G., 1979. Marine and brackish isopods (Isopoda) of cold and temperate waters of the northern hemisphere, vol 1. Suborder Flabellifera. *Opredeliteli po faune SSSR, izdavaemye Zoologicheskim institutom Akademii nauk SSSR* 122, 1–470.
- Langmead, B., Salzberg, S.L., 2012. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9, 357–359. doi:10.1038/nmeth.1923
- Lassy, C.W., Karr, T.L., 1996. Cytological analysis of fertilization and early embryonic development in incompatible crosses of *Drosophila simulans*. *Mechanisms of Development* 57, 47–58. doi:10.1016/0925-4773(96)00527-8
- Le Clec'h, W., Chevalier, F.D., Genty, L., Bertaux, J., Bouchon, D., Sicard, M., 2013a. Cannibalism and Predation as Paths for Horizontal Passage of *Wolbachia* between Terrestrial Isopods. *PLOS ONE* 8, e60232. doi:10.1371/journal.pone.0060232
- Le Clec'h, W., Raimond, M., Guillot, S., Bouchon, D., Sicard, M., 2013b. Horizontal transfers of feminizing versus non-feminizing *Wolbachia* strains: from harmless passengers to pathogens. *Environ Microbiol* 15, 2922–2936. doi:10.1111/1462-2920.12172
- Leclercq, S., Thézé, J., Chebbi, M.A., Giraud, I., Moumen, B., Ernenwein, L., Grève, P., Gilbert, C., Cordaux, R., 2016. Birth of a W sex chromosome by horizontal transfer of *Wolbachia* bacterial symbiont genome. *PNAS* 15036–15041. doi:10.1073/pnas.1608979113
- Leelatanawit, R., Sittikankeaw, K., Yocawibun, P., Klinbunga, S., Roytrakul, S., Aoki, T., Hirono, I., Menasveta, P., 2009. Identification, characterization and expression of sex-related genes in testes of the giant tiger shrimp *Penaeus monodon*. *Comp. Biochem. Physiol., Part A Mol. Integr. Physiol.* 152, 66–76. doi:10.1016/j.cbpa.2008.09.004
- Legrand, J.-J., 1941. Cas d'intersexualité chez l'Isopode terrestre *Armadillidium vulgare* (Latreille). *C. r. hebd. séances Acad. sci.* 213, 808–810.
- Legrand, J.-J., Juchault, P., 1984. Nouvelles données sur le déterminisme génétique et épigénétique de la monogénie chez le crustacé isopode terrestre *Armadillidium vulgare* Latr. *Génétique sélection évolution* 16, 57–84.
- Legrand, J.J., Juchault, P., Artault, J.C., Mocquard, J.P., Picaud, J.L., 1974. Le statut systématique de la «forme» *petiti* Vandel de *Porcellio dilatatus* Brandt, récoltée à l'île Saint-Honorat (Alpes-Maritimes). *Critères morphologiques, génétiques et physiologiques.* *Bull. Soc. Zool. Fr* 99, 461–471.
- Legrand, J.-J., Legrand-Hamelin, E., Juchault, P., 1987. Sex determination in Crustacea. *Biological Reviews* 62, 439–447.
- Lemos, B., Branco, A.T., Hartl, D.L., 2010. Epigenetic effects of polymorphic Y chromosomes modulate chromatin components, immune response, and sexual conflict. *PNAS* 107, 15826–15831. doi:10.1073/pnas.1010383107
- Lenormand, T., 2003. The evolution of sex dimorphism in recombination. *Genetics* 163, 811–822.
- Lenormand, T., Dutheil, J., 2005. Recombination difference between sexes: a role for haploid selection. *PLoS Biol.* 3, e63. doi:10.1371/journal.pbio.0030063

- LePage, D., Bordenstein, S.R., 2013. *Wolbachia*: Can we save lives with a great pandemic? Trends Parasitol. 29, 385–393. doi:10.1016/j.pt.2013.06.003
- LePage, D.P., Metcalf, J.A., Bordenstein, S.R., On, J., Perlmutter, J.I., Shropshire, J.D., Layton, E.M., Funkhouser-Jones, L.J., Beckmann, J.F., Bordenstein, S.R., 2017. Prophage WO genes recapitulate and enhance *Wolbachia*-induced cytoplasmic incompatibility. Nature 543, 243–247.
- Lerat, E., 2010. Identifying repeats and transposable elements in sequenced genomes: how to find your way through the dense forest of programs. Heredity 104, 520–533. doi:10.1038/hdy.2009.165
- Lespinet, O., Wolf, Y.I., Koonin, E.V., Aravind, L., 2002. The role of lineage-specific gene family expansion in the evolution of eukaryotes. Genome Res. 12, 1048–1059. doi:10.1101/gr.174302
- Leutert, V.R., 1974. Zur Geschlechtsbestimmung und Gametogenese von *Bonellia viridis* Rolando. Development 32, 169–193.
- Li, L., Stoeckert, C.J., Roos, D.S., 2003. OrthoMCL: identification of ortholog groups for eukaryotic genomes. Genome Res. 13, 2178–2189. doi:10.1101/gr.1224503
- Li, Q., Zhou, X., Guo, Y., Shang, X., Chen, H., Lu, H., Cheng, H., Zhou, R., 2008. Nuclear localization, DNA binding and restricted expression in neural and germ cells of zebrafish Dmrt3. Biology of the Cell 100, 453–463.
- Li, S., Li, F., Wen, R., Xiang, J., 2012. Identification and characterization of the sex-determiner transformer-2 homologue in Chinese shrimp, *Fenneropenaeus chinensis*. Sex Dev 6, 267–278. doi:10.1159/000341377
- Liew, W.C., Bartfai, R., Lim, Z., Sreenivasan, R., Siegfried, K.R., Orban, L., 2012. Polygenic Sex Determination System in Zebrafish. PLOS ONE 7, e34397. doi:10.1371/journal.pone.0034397
- Livak, K.J., Schmittgen, T.D., 2001. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods 25, 402–408. doi:10.1006/meth.2001.1262
- Lozano-Fernandez, J., Carton, R., Tanner, A.R., Puttick, M.N., Blaxter, M., Vinther, J., Olesen, J., Giribet, G., Edgecombe, G.D., Pisani, D., 2016. A molecular palaeobiological exploration of arthropod terrestrialization. Phil. Trans. R. Soc. B 371, 20150133. doi:10.1098/rstb.2015.0133
- Lucchesi, J.C., Kuroda, M.I., 2015. Dosage Compensation in *Drosophila*. Cold Spring Harb Perspect Biol 7, a019398. doi:10.1101/cshperspect.a019398
- Luo, D., Liu, Y., Hui, M., Song, C., Liu, H., Cui, Z., 2017. Molecular characterization and expression profiles of four transformer-2 isoforms in the Chinese mitten crab *Eriocheir sinensis*. Chin. J. Ocean. Limnol. 35, 782–791. doi:10.1007/s00343-017-6071-z
- Luo, R., Liu, B., Xie, Y., Li, Z., Huang, W., Yuan, J., He, G., Chen, Y., Pan, Q., Liu, Y., Tang, J., Wu, G., Zhang, H., Shi, Y., Liu, Y., Yu, C., Wang, B., Lu, Y., Han, C., Cheung, D.W., Yiu, S.-M., Peng, S., Xiaoqian, Z., Liu, G., Liao, X., Li, Y., Yang, H., Wang, J., Lam, T.-W., Wang, J., 2012. SOAPdenovo2: an empirically improved memory-efficient short-read *de novo* assembler. Gigascience 1, 18. doi:10.1186/2047-217X-1-18
- Lynch, M., 2002. Intron evolution as a population-genetic process. PNAS 99, 6118–6123. doi:10.1073/pnas.092595699
- Magadum, S., Banerjee, U., Murugan, P., Gangapur, D., Ravikesavan, R., 2013. Gene duplication as a major force in evolution. J. Genet. 92, 155–161.
- Makepeace, B.L., Gill, A.C., 2016. *Wolbachia*, in: Rickettsiales. Springer, pp. 465–512.

- Malmberg, R.L., 1977. The evolution of epistasis and the advantage of recombination in populations of bacteriophage T4. *Genetics* 86, 607–621.
- Mank, J.E., 2009. The W, X, Y and Z of sex-chromosome dosage compensation. *Trends Genet.* 25, 226–233. doi:10.1016/j.tig.2009.03.005
- Mank, J.E., Ellegren, H., 2007. Parallel divergence and degradation of the avian W sex chromosome. *Trends Ecol. Evol. (Amst.)* 22, 389–391. doi:10.1016/j.tree.2007.05.003
- Mank, J.E., Hosken, D.J., Wedell, N., 2014. Conflict on the Sex Chromosomes: Cause, Effect, and Complexity. *Cold Spring Harb Perspect Biol* 6, a017715. doi:10.1101/cshperspect.a017715
- Mank, J.E., Promislow, D.E.L., Avise, J.C., 2006. Evolution of alternative sex-determining mechanisms in teleost fishes. *Biological Journal of the Linnean Society* 87, 83–93. doi:10.1111/j.1095-8312.2006.00558.x
- Marcadé, I., Souty-Grosset, C., Bouchon, D., Rigaud, T., Raimond, R., 1999. Mitochondrial DNA variability and *wolbachia* infection in two sibling woodlice species. *Heredity (Edinb)* 83 (Pt 1), 71–78.
- Marçais, G., Kingsford, C., 2011. A fast, lock-free approach for efficient parallel counting of occurrences of *k-mers*. *Bioinformatics* 27, 764–770. doi:10.1093/bioinformatics/btr011
- Marín, I., Baker, B.S., 1998. The evolutionary dynamics of sex determination. *Science* 281, 1990–1994.
- Martin, G., Juchault, P., Legrand, J.J., 1973. Mise en évidence d'un micro-organisme intracytoplasmique symbiote de l'oniscoïde *Armadillidium vulgare* Latr. dont la présence accompagne l'intersexualité ou la féminisation totale des mâles génétiques de la lignée thelygène. *C R Acad Sci Paris* 276, 2213–2216.
- Martin, G., Sorokine, O., Moniatte, M., Bulet, P., Hetru, C., Van Dorsselaer, A., 1999. The structure of a glycosylated protein hormone responsible for sex determination in the isopod, *Armadillidium vulgare*. *Eur. J. Biochem.* 262, 727–736.
- Martin, W., Kowallik, K.V., 1999. Annotated English translation of Mereschowsky's 1905 paper 'Über Natur und Ursprung der Chromatophoren im Pflanzenreiche.' *European Journal of Phycology* 34, 287–295.
- Mathers, T.C., Chen, Y., Kaithakottil, G., Legeai, F., Mugford, S.T., Baa-Puyoulet, P., Bretaudeau, A., Clavijo, B., Colella, S., Collin, O., Dalmay, T., Derrien, T., Feng, H., Gabaldón, T., Jordan, A., Julca, I., Kettles, G.J., Kowitzwanich, K., Lavenier, D., Lenzi, P., Lopez-Gomollon, S., Loska, D., Mapleson, D., Maumus, F., Moxon, S., Price, D.R.G., Sugio, A., van Munster, M., Uzest, M., Waite, D., Jander, G., Tagu, D., Wilson, A.C.C., van Oosterhout, C., Swarbreck, D., Hogenhout, S.A., 2017. Rapid transcriptional plasticity of duplicated gene clusters enables a clonally reproducing aphid to colonise diverse plant species. *Genome Biol* 18. doi:10.1186/s13059-016-1145-3
- Matson, C.K., Zarkower, D., 2012. Sex and the singular DM domain: insights into sexual regulation, evolution and plasticity. *Nat. Rev. Genet.* 13, 163–174. doi:10.1038/nrg3161
- Matsubara, K., Tarui, H., Toriba, M., Yamada, K., Nishida-Umehara, C., Agata, K., Matsuda, Y., 2006. Evidence for different origin of sex chromosomes in snakes, birds, and mammals and step-wise differentiation of snake sex chromosomes. *PNAS* 103, 18190–18195. doi:10.1073/pnas.0605274103
- Matsuda, H., Abrams, P.A., 1999. Why are equally sized gametes so rare? The instability of isogamy and the cost of anisogamy. *Evolutionary Ecology Research* 1, 769–784.
- Matsuda, M., Shinomiya, A., Kinoshita, M., Suzuki, A., Kobayashi, T., Paul-Prasanth, B., Lau, E., Hamaguchi, S., Sakaizumi, M., Nagahama, Y., 2007. DMY gene induces

- male development in genetically female (XX) medaka fish. *Proc Natl Acad Sci U S A* 104, 3865–3870. doi:10.1073/pnas.0611707104
- Mattern, D., Schlegel, M., 2001. Molecular evolution of the small subunit ribosomal DNA in woodlice (Crustacea, Isopoda, Oniscidea) and implications for Oniscidean phylogeny. *Mol. Phylogenet. Evol.* 18, 54–65. doi:10.1006/mpev.2000.0861
- Maynard Smith, J., 1978. *The Evolution of Sex*. CUP Archive.
- McClelland, G.A.H., 1962. Sex-linkage in *Aedes aegypti*. *Trans roy Soc trop Med Hyg* 56.
- McClung, C.E., 1902. The Accessory Chromosome: Sex Determinant? *Biological Bulletin* 3, 43–84. doi:10.2307/1535527
- Metcalf, J.A., Funkhouser-Jones, L.J., Brileya, K., Reysenbach, A.-L., Bordenstein, S.R., 2014. Antibacterial gene transfer across the tree of life. *Elife* 3, e04266.
- Metegnier, G., Becking, T., Chebbi, M.A., Giraud, I., Moumen, B., Schaack, S., Cordaux, R., Gilbert, C., 2015. Comparative paleovirological analysis of crustaceans identifies multiple widespread viral groups. *Mob DNA* 6, 16. doi:10.1186/s13100-015-0047-3
- Meyer, B.J., 2000. Sex in the wormcounting and compensating X-chromosome dose. *Trends Genet.* 16, 247–253.
- Michel-Salzat, A., 2000. Evolution des oniscidea (crustacea, isopoda): analyse de gènes mitochondriaux et implications phylogénétiques. Paris 6, Paris.
- Michel-Salzat, A., Bouchon, D., 2000. Phylogenetic analysis of mitochondrial LSU rRNA in oniscids. *C. R. Acad. Sci. III, Sci. Vie* 323, 827–837.
- Milinkovitch, M.C., 1995. Molecular phylogeny of cetaceans prompts revision of morphological transformations. *Trends in ecology & evolution* 10, 328–334.
- Miller, W.J., Ehrman, L., Schneider, D., 2010. Infectious Speciation Revisited: Impact of Symbiont-Depletion on Female Fitness and Mating Behavior of *Drosophila paulistorum*. *PLOS Pathogens* 6, e1001214. doi:10.1371/journal.ppat.1001214
- Mittal, O.P., Pahwa, S., 1981. Germ cell chromosomes in two species of terrestrial isopods from India. *Experientia* 37, 343–345. doi:10.1007/BF01959853
- Mittal, O.P., Pahwa, S., 1980. Cytogenetical studies on two terrestrial isopods (Crustacea), *Philoscia lodnensis Ramakrishna* and *Porcellio rathkei Brandt*. *Cytobios* 29, 17–24.
- Moore, E.C., Roberts, R.B., 2013. Polygenic sex determination. *Current Biology* 23, R510–R512. doi:10.1016/j.cub.2013.04.004
- Morgan, T.H., 1914. No crossing over in the male of *Drosophila* of genes in the second and third pairs of chromosomes. *The Biological Bulletin* 26, 195–204.
- Müller, H.J., 1964. The relation of recombination to mutational advance. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis* 1, 2–9. doi:10.1016/0027-5107(64)90047-8
- Muyle, A., Käfer, J., Zemp, N., Mousset, S., Picard, F., Marais, G.A., 2016. SEX-DETECTOR: A Probabilistic Approach to Study Sex Chromosomes in Non-Model Organisms. *Genome Biol Evol* 8, 2530–2543. doi:10.1093/gbe/evw172
- Muyle, A., Zemp, N., Deschamps, C., Mousset, S., Widmer, A., Marais, G.A.B., 2012. Rapid de novo evolution of X chromosome dosage compensation in *Silene latifolia*, a plant with young sex chromosomes. *PLoS Biol.* 10, e1001308. doi:10.1371/journal.pbio.1001308
- Nardon, P., Grenier, A.M., 1989. Endocytobiosis in Coleoptera: biological, biochemical, and genetic aspects. *Insect endocytobiosis: morphology, physiology, genetics, evolution*. CRC Press, Boca Raton, Florida 175–216.
- Narita, S., Kageyama, D., Nomura, M., Fukatsu, T., 2007. Unexpected Mechanism of Symbiont-Induced Reversal of Insect Sex: Feminizing *Wolbachia* Continuously Acts on the Butterfly *Eurema hecabe* during Larval Development. *Appl. Environ. Microbiol.* 73, 4332–4341. doi:10.1128/AEM.00145-07

- Natri, H.M., Shikano, T., Merilä, J., 2013. Progressive recombination suppression and differentiation in recently evolved neo-sex chromosomes. *Mol. Biol. Evol.* 30, 1131–1144. doi:10.1093/molbev/mst035
- Navarro-Martín, L., Viñas, J., Ribas, L., Díaz, N., Gutiérrez, A., Croce, L.D., Piferrer, F., 2011. DNA Methylation of the Gonadal Aromatase (cyp19a) Promoter Is Involved in Temperature-Dependent Sex Ratio Shifts in the European Sea Bass. *PLOS Genetics* 7, e1002447. doi:10.1371/journal.pgen.1002447
- Negri, I., Franchini, A., Gonella, E., Daffonchio, D., Mazzoglio, P.J., Mandrioli, M., Alma, A., 2009. Unravelling the *Wolbachia* evolutionary role: the reprogramming of the host genomic imprinting. *Proceedings of the Royal Society of London B: Biological Sciences* 276, 2485–2491. doi:10.1098/rspb.2009.0324
- Negri, I., Pellicchia, M., Mazzoglio, P.J., Patetta, A., Alma, A., 2006. Feminizing *Wolbachia* in *Zyginidia pullula* (Insecta, Hemiptera), a leafhopper with an XX/X0 sex-determination system. *Proceedings of the Royal Society of London B: Biological Sciences* 273, 2409–2416. doi:10.1098/rspb.2006.3592
- Newton, M.E., Southern, D.I., Wood, R.J., 1974. X and Y chromosomes of *Aedes aegypti* (L.) distinguished by Giemsa C-banding. *Chromosoma* 49, 41–49.
- Niimura, Y., 2009. On the origin and evolution of vertebrate olfactory receptor genes: comparative genome analysis among 23 chordate species. *Genome Biology and Evolution* 1, 34–44.
- Niimura, Y., Nei, M., 2005. Evolutionary dynamics of olfactory receptor genes in fishes and tetrapods. *Proceedings of the National Academy of Sciences* 102, 6039–6044.
- Nikoh, N., Hosokawa, T., Moriyama, M., Oshima, K., Hattori, M., Fukatsu, T., 2014. Evolutionary origin of insect-*Wolbachia* nutritional mutualism. *PNAS* 111, 10257–10262. doi:10.1073/pnas.1409284111
- Nishioka, M., Miura, I., Saitoh, K., 1993. Sex Chromosomes of *Rana rugosa* with Special Reference to Local Differences in Sex-determining Mechanism. *Scientific report of the Laboratory for Amphibian Biology* 12, 55–81. doi:info:doi/10.15027/14532
- Nunney, L., 1989. The maintenance of sex by group selection. *Evolution* 43, 245–257. doi:10.1111/j.1558-5646.1989.tb04225.x
- Nur, U., 1980. Evolution of unusual chromosome systems in scale insects (Coccoidea: Homoptera), in: *Insect Cytogenetics*. R. L. Blackman, G. M. Hewitt, and M. Ashburner, Oxford, pp. 97–117.
- Oetjens, M.T., Shen, F., Emery, S.B., Zou, Z., Kidd, J.M., 2016. Y-Chromosome Structural Diversity in the Bonobo and Chimpanzee Lineages. *Genome Biol Evol* 8, 2231–2240. doi:10.1093/gbe/evw150
- Ohno, S., 1967. *Sex Chromosomes and Sex-linked Genes*. Springer-Verlag.
- Okuno, A., Hasegawa, Y., Ohira, T., Katakura, Y., Nagasawa, H., 1999. Characterization and cDNA cloning of androgenic gland hormone of the terrestrial isopod *Armadillidium vulgare*. *Biochem. Biophys. Res. Commun.* 264, 419–423. doi:10.1006/bbrc.1999.1522
- Ospina-Álvarez, N., Piferrer, F., 2008. Temperature-Dependent Sex Determination in Fish Revisited: Prevalence, a Single Sex Ratio Response Pattern, and Possible Effects of Climate Change. *PLOS ONE* 3, e2837. doi:10.1371/journal.pone.0002837
- Otto, S.P., 2009. The evolutionary enigma of sex. *Am. Nat.* 174 Suppl 1, S1–S14. doi:10.1086/599084
- Otto, S.P., Whitton, J., 2000. Polyploid incidence and evolution. *Annu. Rev. Genet.* 34, 401–437. doi:10.1146/annurev.genet.34.1.401
- Pandian, T.J., 2012. *Genetic Sex Differentiation in Fish*. CRC Press.

- Pane, A., Salvemini, M., Delli Bovi, P., Polito, C., Saccone, G., 2002. The transformer gene in *Ceratitis capitata* provides a genetic basis for selecting and remembering the sexual fate. *Development* 129, 3715–3725.
- Pannebakker, B.A., Loppin, B., Elemans, C.P.H., Humblot, L., Vavre, F., 2007. Parasitic inhibition of cell death facilitates symbiosis. *PNAS* 104, 213–215. doi:10.1073/pnas.0607845104
- Parker, G.A., Baker, R.R., Smith, V.G., 1972. The origin and evolution of gamete dimorphism and the male-female phenomenon. *J. Theor. Biol.* 36, 529–553.
- Parker, G.A., Lehtonen, J., 2014. Gamete evolution and sperm numbers: sperm competition versus sperm limitation. *Proc. Biol. Sci.* 281, 20140836. doi:10.1098/rspb.2014.0836
- Peccoud, J., Loiseau, V., Cordaux, R., Gilbert, C., 2017. Massive horizontal transfer of transposable elements in insects. *Proc. Natl. Acad. Sci. U.S.A.* 114, 4721–4726. doi:10.1073/pnas.1621178114
- Peck, J.R., 1994. A ruby in the rubbish: beneficial mutations, deleterious mutations and the evolution of sex. *Genetics* 137, 597–606.
- Perrin, N., 2009. Sex reversal: a fountain of youth for sex chromosomes? *Evolution* 63, 3043–3049. doi:10.1111/j.1558-5646.2009.00837.x
- Picard, M.A.-L., Cosseau, C., Mouahid, G., Duval, D., Grunau, C., Toulza, È., Allienne, J.-F., Boissier, J., 2015. The roles of Dmrt (Double sex/Male-abnormal-3 Related Transcription factor) genes in sex determination and differentiation mechanisms: Ubiquity and diversity across the animal kingdom. *C. R. Biol.* 338, 451–462. doi:10.1016/j.crv.2015.04.010
- Pieau, C., 1971. Sur la proportion sexuelle chez les embryons de deux Chéloniens (*Testudo graeca* L. et *Emys orbicularis* L.) issus d’oeufs incubés artificiellement. *CR Acad. Sci. Paris* 272, 3071–3074.
- Poore, G.C.B., Bruce, N.L., 2012. Global Diversity of Marine Isopods (Except Asellota and Crustacean Symbionts). *PLoS One* 7. doi:10.1371/journal.pone.0043529
- Prak, E.T., Kazazian, H.H., 2000. Mobile elements and the human genome. *Nat. Rev. Genet.* 1, 134–144. doi:10.1038/35038572
- Presgraves, D.C., 2008. Sex chromosomes and speciation in *Drosophila*. *Trends Genet.* 24, 336–343. doi:10.1016/j.tig.2008.04.007
- Price, A.L., Jones, N.C., Pevzner, P.A., 2005. *De novo* identification of repeat families in large genomes. *Bioinformatics* 21 Suppl 1, i351–358. doi:10.1093/bioinformatics/bti1018
- Quevillon, E., Silventoinen, V., Pillai, S., Harte, N., Mulder, N., Apweiler, R., Lopez, R., 2005. InterProScan: protein domains identifier. *Nucleic Acids Res* 33, W116–W120. doi:10.1093/nar/gki442
- R Core Team, 2014. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
- Raymond, C.S., Shamu, C.E., Shen, M.M., Seifert, K.J., Hirsch, B., Hodgkin, J., Zarkower, D., 1998. Evidence for evolutionary conservation of sex-determining genes. *Nature* 391, 691–695. doi:10.1038/35618
- Reichwald, K., Petzold, A., Koch, P., Downie, B.R., Hartmann, N., Pietsch, S., Baumgart, M., Chalopin, D., Felder, M., Bens, M., Sahm, A., Szafranski, K., Taudien, S., Groth, M., Arisi, I., Weise, A., Bhatt, S.S., Sharma, V., Kraus, J.M., Schmid, F., Priebe, S., Liehr, T., Görlach, M., Than, M.E., Hiller, M., Kestler, H.A., Volff, J.-N., Scharl, M., Cellerino, A., Englert, C., Platzer, M., 2015. Insights into Sex Chromosome Evolution and Aging from the Genome of a Short-Lived Fish. *Cell* 163, 1527–1538. doi:10.1016/j.cell.2015.10.071

- Renner, S.S. (University of M., Ricklefs, R.E., 1995. Dioecy and its correlates in the flowering plants. American journal of botany (USA).
- Rhoads, A., Au, K.F., 2015. PacBio Sequencing and Its Applications. Genomics, Proteomics & Bioinformatics, SI: Metagenomics of Marine Environments 13, 278–289. doi:10.1016/j.gpb.2015.08.002
- Ridout, K.E., Veltsos, P., Muyle, A., Emery, O., Rastas, P., Marais, G.A., Filatov, D.A., Pannell, J.R., 2017. Hallmarks of early sex-chromosome evolution in the dioecious plant *Mercurialis annua* revealed by de novo genome assembly, genetic mapping and transcriptome analysis. bioRxiv 106120. doi:10.1101/106120
- Rigaud, null, Moreau, null, Juchault, null, 1999. *Wolbachia* infection in the terrestrial isopod oniscus asellus: sex ratio distortion and effect on fecundity. Heredity (Edinb) 83, 469–475.
- Rigaud, T., 1997. Inherited microorganisms and sex determination of arthropod hosts, in: Influential Passengers: Inherited Microorganisms and Arthropod Reproduction. O'Neill S.L., Hoffmann A.A., Werren J.H., Oxford, pp. 81–101.
- Rigaud, T., Juchault, P., 1993. Conflict between feminizing sex ratio distorters and an autosomal masculinizing gene in the terrestrial isopod *Armadillidium vulgare* Latr. Genetics 133, 247–252.
- Rigaud, T., Juchault, P., 1992. Genetic control of the vertical transmission of a cytoplasmic sex factor in *Armadillidium vulgare* Latr.(Crustacea, Oniscidea). Heredity 68, 47–52.
- Rigaud, T., Juchault, P., Mocquard, J.-P., 1997. The evolution of sex determination in isopod crustaceans. Bioessays 19, 409–416. doi:10.1002/bies.950190508
- Rigaud, T., Mocquard, J.P., Juchault, P., 1992. The spread of parasitic sex factors in populations of *Armadillidium vulgare* Latr (Crustacea, Oniscidea): effects on sex ratio. Genetics Selection Evolution 24, 3.
- Rigaud, T., Pennings, P.S., Juchault, P., 2001. *Wolbachia* bacteria effects after experimental interspecific transfers in terrestrial isopods. J. Invertebr. Pathol. 77, 251–257. doi:10.1006/jipa.2001.5026
- Riparbelli, M.G., Giordano, R., Ueyama, M., Callaini, G., 2012. *Wolbachia*-Mediated Male Killing Is Associated with Defective Chromatin Remodeling. PLOS ONE 7, e30045. doi:10.1371/journal.pone.0030045
- Roberts, R.J., Carneiro, M.O., Schatz, M.C., 2013. The advantages of SMRT sequencing. Genome Biol. 14, 405. doi:10.1186/gb-2013-14-6-405
- Robinson, K.M., Sieber, K.B., Dunning Hotopp, J.C., 2013. A review of bacteria-animal lateral gene transfer may inform our understanding of diseases like cancer. PLoS Genet. 9, e1003877. doi:10.1371/journal.pgen.1003877
- Romiguier, J., Gayral, P., Ballenghien, M., Bernard, A., Cahais, V., Chenuil, A., Chiari, Y., Derrat, R., Duret, L., Faivre, N., Loire, E., Lourenco, J.M., Nabholz, B., Roux, C., Tsagkogeorga, G., Weber, A. a.-T., Weinert, L.A., Belkhir, K., Bierne, N., Glémin, S., Galtier, N., 2014. Comparative population genomics in animals uncovers the determinants of genetic diversity. Nature 515, 261–263. doi:10.1038/nature13685
- Ronshaugen, M., McGinnis, N., McGinnis, W., 2002. Hox protein mutation and macroevolution of the insect body plan. Nature 415, 914–917. doi:10.1038/nature716
- Roulin, A.C., Routtu, J., Hall, M.D., Janicke, T., Colson, I., Haag, C.R., Ebert, D., 2013. Local adaptation of sex induction in a facultative sexual crustacean: insights from QTL mapping and natural populations of *Daphnia magna*. Mol. Ecol. 22, 3567–3579. doi:10.1111/mec.12308
- Rousset, F., Bouchon, D., Pintureau, B., Juchault, P., Solignac, M., 1992. *Wolbachia* Endosymbionts Responsible for Various Alterations of Sexuality in Arthropods.

- Proceedings of the Royal Society of London B: Biological Sciences 250, 91–98. doi:10.1098/rspb.1992.0135
- Ryner, L.C., Baker, B.S., 1991. Regulation of doublesex pre-mRNA processing occurs by 3'-splice site activation. *Genes Dev.* 5, 2071–2085.
- Salmela, L., Rivals, E., 2014. LoRDEC: accurate and efficient long read error correction. *Bioinformatics* 30, 3506–3514. doi:10.1093/bioinformatics/btu538
- Salzberg, S.L., Phillippy, A.M., Zimin, A., Puiu, D., Magoc, T., Koren, S., Treangen, T.J., Schatz, M.C., Delcher, A.L., Roberts, M., Marçais, G., Pop, M., Yorke, J.A., 2012. GAGE: a critical evaluation of genome assemblies and assembly algorithms. *Genome Res* 22. doi:10.1101/gr.131383.111
- Sánchez, L., 2008. Sex-determining mechanisms in insects. *Int. J. Dev. Biol.* 52, 837–856. doi:10.1387/ijdb.072396ls
- Sarre, S.D., Ezaz, T., Georges, A., 2011. Transitions between sex-determining systems in reptiles and amphibians. *Annu Rev Genomics Hum Genet* 12, 391–406. doi:10.1146/annurev-genom-082410-101518
- Schaack, S., Gilbert, C., Feschotte, C., 2010. Promiscuous DNA: horizontal transfer of transposable elements and why it matters for eukaryotic evolution. *Trends Ecol. Evol. (Amst.)* 25, 537–546. doi:10.1016/j.tree.2010.06.001
- Schäfer, M., Nayernia, K., Engel, W., Schäfer, U., 1995. Translational control in spermatogenesis. *Dev. Biol.* 172, 344–352. doi:10.1006/dbio.1995.8049
- Schmalzfuss, H., 2003. World catalog of terrestrial isopods (Isopoda: Oniscidea). Staatliches Museum für Naturkunde.
- Schmalzfuss, H., 1989. Phylogenetics in Oniscidea. *Monografia. Monitore zoologico italiano* 4, 3–27.
- Schmidt, C., 2008. Phylogeny of the terrestrial Isopoda (Oniscidea): a review. *Arthropod Systematics & Phylogeny* 66, 191–226.
- Schmieder, R., Edwards, R., 2011. Quality control and preprocessing of metagenomic datasets. *Bioinformatics* 27, 863–864. doi:10.1093/bioinformatics/btr026
- Selden, P.A., Edwards, D., 1989. Colonisation of the land, in: Allen, K., Briggs, D.E.G. (Eds.), *Evolution and the Fossil Record*. Belhaven Press (Pinter Publishers Ltd), London, pp. 122–152.
- Sellars, M.J., Trewin, C., McWilliam, S.M., Graves, R.S.E., Hertzler, P.L., 2015. Transcriptome profiles of *Penaeus (Marsupenaeus) japonicus* animal and vegetal half-embryos: identification of sex determination, germ line, mesoderm, and other developmental genes. *Mar. Biotechnol.* 17, 252–265. doi:10.1007/s10126-015-9613-4
- Serbus, L.R., Casper-Lindley, C., Landmann, F., Sullivan, W., 2008. The genetics and cell biology of *Wolbachia*-host interactions. *Annu. Rev. Genet.* 42, 683–707. doi:10.1146/annurev.genet.41.110306.130354
- Sfenthourakis, S., Taiti, S., 2015. Patterns of taxonomic diversity among terrestrial isopods. *Zookeys* 13–25. doi:10.3897/zookeys.515.9332
- Shetty, S., Griffin, D.K., Graves, J.A., 1999. Comparative painting reveals strong chromosome homology over 80 million years of bird evolution. *Chromosome Res.* 7, 289–295.
- Shibata, K., Takase, M., Nakamura, M., 2002. The *Dmrt1* expression in sex-reversed gonads of amphibians. *Gen. Comp. Endocrinol.* 127, 232–241.
- Shoemaker, D.D., Machado, C.A., Molbo, D., Werren, J.H., Windsor, D.M., Herre, E.A., 2002. The distribution of *Wolbachia* in fig wasps: correlations with host phylogeny, ecology and population structure. *Proc. Biol. Sci.* 269, 2257–2267. doi:10.1098/rspb.2002.2100

- Sicard, M., Dittmer, J., Grève, P., Bouchon, D., Braquart-Varnier, C., 2014. A host as an ecosystem: *Wolbachia* coping with environmental constraints. *Environ Microbiol* 16, 3583–3607. doi:10.1111/1462-2920.12573
- Simão, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E.V., Zdobnov, E.M., 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31, 3210–3212. doi:10.1093/bioinformatics/btv351
- Skaletsky, H., Kuroda-Kawaguchi, T., Minx, P.J., Cordum, H.S., Hillier, L., Brown, L.G., Repping, S., Pyntikova, T., Ali, J., Bieri, T., Chinwalla, A., Delehaunty, A., Delehaunty, K., Du, H., Fewell, G., Fulton, L., Fulton, R., Graves, T., Hou, S.-F., Latrielle, P., Leonard, S., Mardis, E., Maupin, R., McPherson, J., Miner, T., Nash, W., Nguyen, C., Ozersky, P., Pepin, K., Rock, S., Rohlfing, T., Scott, K., Schultz, B., Strong, C., Tin-Wollam, A., Yang, S.-P., Waterston, R.H., Wilson, R.K., Rozen, S., Page, D.C., 2003. The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. *Nature* 423, 825–837. doi:10.1038/nature01722
- Slater, G.S.C., Birney, E., 2005. Automated generation of heuristics for biological sequence comparison. *BMC Bioinformatics* 6, 31. doi:10.1186/1471-2105-6-31
- Smeds, L., Warmuth, V., Bolivar, P., Uebbing, S., Burri, R., Suh, A., Nater, A., Bureš, S., Garamszegi, L.Z., Hogner, S., Moreno, J., Qvarnström, A., Ružić, M., Sæther, S.-A., Sætre, G.-P., Török, J., Ellegren, H., 2015. Evolutionary analysis of the female-specific avian W chromosome. *Nat Commun* 6, 7330. doi:10.1038/ncomms8330
- Smith, C.A., Roeszler, K.N., Ohnesorg, T., Cummins, D.M., Farlie, P.G., Doran, T.J., Sinclair, A.H., 2009. The avian Z-linked gene DMRT1 is required for male sex determination in the chicken. *Nature* 461, 267–271. doi:10.1038/nature08298
- Song, L., Bian, C., Luo, Y., Wang, L., You, X., Li, J., Qiu, Y., Ma, X., Zhu, Z., Ma, L., Wang, Z., Lei, Y., Qiang, J., Li, H., Yu, J., Wong, A., Xu, J., Shi, Q., Xu, P., 2016. Draft genome of the Chinese mitten crab, *Eriocheir sinensis*. *Gigascience* 5, 5. doi:10.1186/s13742-016-0112-y
- Stamatakis, A., 2006. RAxML-VI-HP: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22, 2688–2690. doi:10.1093/bioinformatics/btl446
- Steinemann, M., Steinemann, S., 1992. Degenerating Y chromosome of *Drosophila miranda*: a trap for retrotransposons. *PNAS* 89, 7591–7595. doi:10.1073/pnas.89.16.7591
- Steinemann, S., Steinemann, M., 2005. Y chromosomes: born to be destroyed. *Bioessays* 27, 1076–1083. doi:10.1002/bies.20288
- Stöck, M., Horn, A., Grossen, C., Lindtke, D., Sermier, R., Betto-Colliard, C., Dufresnes, C., Bonjour, E., Dumas, Z., Luquet, E., Maddalena, T., Sousa, H.C., Martinez-Solano, I., Perrin, N., 2011. Ever-Young Sex Chromosomes in European Tree Frogs. *PLOS Biology* 9, e1001062. doi:10.1371/journal.pbio.1001062
- Stöck, M., Savary, R., Betto-Colliard, C., Biollay, S., Jourdan-Pineau, H., Perrin, N., 2013. Low rates of X-Y recombination, not turnovers, account for homomorphic sex chromosomes in several diploid species of Palearctic green toads (*Bufo viridis* subgroup). *J. Evol. Biol.* 26, 674–682. doi:10.1111/jeb.12086
- Stouthamer, R., Luck, R.F., Hamilton, W.D., 1990. Antibiotics cause parthenogenetic *Trichogramma* (Hymenoptera/Trichogrammatidae) to revert to sex. *Proc Natl Acad Sci U S A* 87, 2424–2427.
- Suzuki, S., 1999. Androgenic gland hormone is a sex-reversing factor but cannot be a sex-determining factor in the female crustacean isopods *Armadillidium vulgare*. *Gen. Comp. Endocrinol.* 115, 370–378. doi:10.1006/gcen.1999.7324
- Suzuki, S., Yamasaki, K., 1995a. Morphological studies on sexual differentiation in *Armadillidium vulgare* (Isopoda: Armadillidae): androgenic gland and male sexual

- characters [WWW Document]. Crustacean Research. December. URL <https://eurekamag.com/research/021/372/021372660.php> (accessed 10.18.16).
- Suzuki, S., Yamasaki, K., 1995b. Morphological studies on sexual differentiation in *Armadillidium vulgare* (Isopoda: Armadillidae): androgenic gland and male sexual characters. Crustacean Research 24, 93–103. doi:10.18353/crustacea.24.0_93
- Tabacaru, I., Danielopol, D.L., 1999. Contribution a la connaissance de la phylogénie des Isopoda (Crustacea). Vie et milieu 49, 163–176.
- Tabacaru, I., Danielopol, D.L., 1996. Phylogénèse et convergence chez les isopodes terrestres. Vie et milieu 46, 171–181.
- Talavera, G., Castresana, J., 2007. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. Syst. Biol. 56, 564–577. doi:10.1080/10635150701472164
- Tamura, K., Stecher, G., Peterson, D., Filipski, A., Kumar, S., 2013. MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. Mol Biol Evol 30, 2725–2729. doi:10.1093/molbev/mst197
- Tanaka, K., Takehana, Y., Naruse, K., Hamaguchi, S., Sakaizumi, M., 2007. Evidence for different origins of sex chromosomes in closely related *Oryzias* fishes: substitution of the master sex-determining gene. Genetics 177, 2075–2081. doi:10.1534/genetics.107.075598
- Tanaka, Y., 1914. A study of Mendelian factors in the Silkworm, *Bombyx mori*. Molecular and General Genetics MGG 12, 161–161.
- Tang, H., 2007. Genome assembly, rearrangement, and repeats. Chem. Rev. 107, 3391–3406. doi:10.1021/cr0683008
- Tarailo-Graovac, M., Chen, N., 2009. Using RepeatMasker to identify repetitive elements in genomic sequences. Curr Protoc Bioinformatics Chapter 4, Unit 4.10. doi:10.1002/0471250953.bi0410s25
- Taylor, D.R., 1990. Evolutionary consequences of cytoplasmic sex ratio distorters. Evol Ecol 4, 235–248. doi:10.1007/BF02214332
- Taylor, J.S., Van de Peer, Y., Meyer, A., 2001. Genome duplication, divergent resolution and speciation. Trends Genet. 17, 299–301.
- Taylor, M.J., Bandi, C., Hoerauf, A., 2005. Wolbachia bacterial endosymbionts of filarial nematodes. Adv. Parasitol. 60, 245–284. doi:10.1016/S0065-308X(05)60004-8
- Teixeira, L., Ferreira, Á., Ashburner, M., 2008. The Bacterial Symbiont Wolbachia Induces Resistance to RNA Viral Infections in *Drosophila melanogaster*. PLOS Biology 6, e1000002. doi:10.1371/journal.pbio.1000002
- Telschow, A., Hammerstein, P., Werren, J.H., 2005. The effect of Wolbachia versus genetic incompatibilities on reinforcement and speciation. Evolution 59, 1607–1619.
- Thézé, J., Leclercq, S., Moumen, B., Cordaux, R., Gilbert, C., 2014. Remarkable diversity of endogenous viruses in a crustacean genome. Genome Biol Evol 6, 2129–2140. doi:10.1093/gbe/evu163
- Tomaszkiewicz, M., Medvedev, P., Makova, K.D., 2017. Y and W Chromosome Assemblies: Approaches and Discoveries. Trends in Genetics 33, 266–282. doi:10.1016/j.tig.2017.01.008
- Tomaszkiewicz, M., Rangavittal, S., Cechova, M., Sanchez, R.C., Fescemyer, H.W., Harris, R., Ye, D., O'Brien, P.C.M., Chikhi, R., Ryder, O.A., Ferguson-Smith, M.A., Medvedev, P., Makova, K.D., 2016. A time- and cost-effective strategy to sequence mammalian Y Chromosomes: an application to the de novo assembly of gorilla Y. Genome Res. 26, 530–540. doi:10.1101/gr.199448.115

- Tram, U., Fredrick, K., Werren, J.H., Sullivan, W., 2006. Paternal chromosome segregation during the first mitotic division determines *Wolbachia*-induced cytoplasmic incompatibility phenotype. *J. Cell. Sci.* 119, 3655–3663. doi:10.1242/jcs.03095
- Traut, W., 2010. New Y chromosomes and early stages of sex chromosome differentiation: sex determination in *Megaselia*. *J. Genet.* 89, 307–313.
- Treangen, T.J., Salzberg, S.L., 2011. Repetitive DNA and next-generation sequencing: computational challenges and solutions. *Nat. Rev. Genet.* 13, 36–46. doi:10.1038/nrg3117
- Trivers, R.L., Willard, D.E., 1973. Natural selection of parental ability to vary the sex ratio of offspring. *Science* 179, 90–92.
- Tuf, I.H., Drábková, L., Šipoš, J., 2015. Personality affects defensive behaviour of *Porcellio scaber* (Isopoda, Oniscidea). *Zookeys* 159–171. doi:10.3897/zookeys.515.9429
- Turner, J.M.A., 2007. Meiotic sex chromosome inactivation. *Development* 134, 1823–1831. doi:10.1242/dev.000018
- Tuttle, E.M., Bergland, A.O., Korody, M.L., Brewer, M.S., Newhouse, D.J., Minx, P., Stager, M., Betuel, A., Cheviron, Z.A., Warren, W.C., Gonser, R.A., Balakrishnan, C.N., 2016. Divergence and Functional Degradation of a Sex Chromosome-like Supergene. *Current Biology* 26, 344–350. doi:10.1016/j.cub.2015.11.069
- Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B.C., Remm, M., Rozen, S.G., 2012. Primer3—new capabilities and interfaces. *Nucleic Acids Res* 40, e115. doi:10.1093/nar/gks596
- Valenzuela, N., Lance, V., 2004. Temperature-dependent sex determination in vertebrates. Smithsonian Books, California (USA).
- Vallender, E.J., Lahn, B.T., 2004. How mammalian sex chromosomes acquired their peculiar gene content. *Bioessays* 26, 159–169. doi:10.1002/bies.10393
- van Doorn, G.S., Kirkpatrick, M., 2010. Transitions between male and female heterogamety caused by sex-antagonistic selection. *Genetics* 186, 629–645. doi:10.1534/genetics.110.118596
- van Doorn, G.S., Kirkpatrick, M., 2007. Turnover of sex chromosomes induced by sexual conflict. *Nature* 449, 909–912. doi:10.1038/nature06178
- Vandel, A., 1962. Faune de France. Isopodes Terrestres (Deuxième Partie), Lechevalier. ed. Lechevalier, Paris.
- Vandel, A., 1960. Faune de France. Isopodes Terrestres (Première Partie), Lechevalier. ed. Lechevalier, Paris.
- Vandel, A., 1943. Essai sur l'origine, L'évolution et la classification des Oniscoidea (Isopodes terrestres). Laboratoire d'évolution des êtres organisés, Paris, France.
- Vandesompele, J., De Preter, K., Pattyn, F., Poppe, B., Van Roy, N., De Paepe, A., Speleman, F., 2002. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol.* 3, research0034.1–research0034.11.
- Veitia, R.A., 2005. Gene dosage balance: deletions, duplications and dominance. *Trends in Genetics* 21, 33–35. doi:10.1016/j.tig.2004.11.002
- Verhulst, E.C., van de Zande, L., Beukeboom, L.W., 2010. Insect sex determination: it all evolves around transformer. *Curr. Opin. Genet. Dev.* 20, 376–383. doi:10.1016/j.gde.2010.05.001
- Veyrunes, F., Waters, P.D., Miethke, P., Rens, W., McMillan, D., Alsop, A.E., Grützner, F., Deakin, J.E., Whittington, C.M., Schatzkamer, K., Kremitzki, C.L., Graves, T., Ferguson-Smith, M.A., Warren, W., Marshall Graves, J.A., 2008. Bird-like sex chromosomes of platypus imply recent origin of mammal sex chromosomes. *Genome Res.* 18, 965–973. doi:10.1101/gr.7101908

- Vicoso, B., Bachtrog, D., 2015. Numerous Transitions of Sex Chromosomes in Diptera. *PLOS Biol* 13, e1002078. doi:10.1371/journal.pbio.1002078
- Vicoso, B., Bachtrog, D., 2013. Reversal of an ancient sex chromosome to an autosome in *Drosophila*. *Nature* 499, 332–335. doi:10.1038/nature12235
- Vicoso, B., Bachtrog, D., 2009. Progress and prospects toward our understanding of the evolution of dosage compensation. *Chromosome Res* 17, 585. doi:10.1007/s10577-009-9053-y
- Vicoso, B., Emerson, J.J., Zektser, Y., Mahajan, S., Bachtrog, D., 2013a. Comparative Sex Chromosome Genomics in Snakes: Differentiation, Evolutionary Strata, and Lack of Global Dosage Compensation. *PLOS Biol* 11, e1001643. doi:10.1371/journal.pbio.1001643
- Vicoso, B., Kaiser, V.B., Bachtrog, D., 2013b. Sex-biased gene expression at homomorphic sex chromosomes in emus and its implication for sex chromosome evolution. *Proc Natl Acad Sci U S A* 110, 6453–6458. doi:10.1073/pnas.1217027110
- Vieira, F.G., Rozas, J., 2011. Comparative genomics of the odorant-binding and chemosensory protein gene families across the Arthropoda: origin and evolutionary history of the chemosensory system. *Genome Biology and Evolution* 3, 476–490.
- Volff, J.-N., Zarkower, D., Bardwell, V.J., Scharf, M., 2003. Evolutionary dynamics of the DM domain gene family in metazoans. *J. Mol. Evol.* 57 Suppl 1, S241–249. doi:10.1007/s00239-003-0033-0
- Voronin, D.A., Dudkina, N.V., Kiseleva, E.V., 2004. A new form of symbiotic bacteria *Wolbachia* found in the endoplasmic reticulum of early embryos of *Drosophila melanogaster*. *Dokl. Biol. Sci.* 396, 227–229.
- Vrijenhoek, R.C., 1998. Animal Clones and Diversity. *BioScience* 48, 617–628. doi:10.2307/1313421
- Wägele, J.-W., 1996. Identification of apomorphies and the role of groundpatterns in molecular systematics. *Journal of Zoological Systematics and Evolutionary Research* 34, 31–39. doi:10.1111/j.1439-0469.1996.tb00807.x
- Wägele, J.-W., 1989. Evolution und phylogenetisches System der Isopoda.
- Walker, B.J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., Cuomo, C.A., Zeng, Q., Wortman, J., Young, S.K., Earl, A.M., 2014. Pilon: An Integrated Tool for Comprehensive Microbial Variant Detection and Genome Assembly Improvement. *PLOS ONE* 9, e112963. doi:10.1371/journal.pone.0112963
- Wallace, H., Wallace, B.M.N., Badawy, G.M.I., 1997. Lampbrush chromosomes and chiasmata of sex-reversed crested newts. *Chromosoma* 106, 526–533.
- Wallau, G.L., Ortiz, M.F., Loreto, E.L.S., 2012. Horizontal transposon transfer in eukarya: detection, bias, and perspectives. *Genome Biol Evol* 4, 689–699. doi:10.1093/gbe/evs055
- Walthour, C.S., Schaeffer, S.W., 1994. Molecular Population Genetics of Sex Determination Genes: The Transformer Gene of *Drosophila Melanogaster*. *Genetics* 136, 1367–1372.
- Wang, J., Na, J.-K., Yu, Q., Gschwend, A.R., Han, J., Zeng, F., Aryal, R., VanBuren, R., Murray, J.E., Zhang, W., Navajas-Pérez, R., Feltus, F.A., Lemke, C., Tong, E.J., Chen, C., Wai, C.M., Singh, R., Wang, M.-L., Min, X.J., Alam, M., Charlesworth, D., Moore, P.H., Jiang, J., Paterson, A.H., Ming, R., 2012. Sequencing papaya X and Yh chromosomes reveals molecular basis of incipient sex chromosome evolution. *Proc. Natl. Acad. Sci. U.S.A.* 109, 13710–13715. doi:10.1073/pnas.1207833109
- Warbrick, E.V., Barker, G.C., Rees, H.H., Howells, R.E., 1993. The effect of invertebrate hormones and potential hormone inhibitors on the third larval moult of the filarial nematode, *Dirofilaria immitis*, *in vitro*. *Parasitology* 107 (Pt 4), 459–463.

- Warburg, M.R., 1993. Evolutionary biology of land isopods. Springer-Verlag, Berlin; New York.
- Watanabe, T.K., 1975. A new sex-transforming gene on the second chromosome of *Drosophila melanogaster*. The Japanese journal of genetics 50, 269–271.
- Weeks, A.R., Marec, F., Breeuwer, J.A., 2001. A mite species that consists entirely of haploid females. Science 292, 2479–2482. doi:10.1126/science.1060411
- Wei, W., Fitzgerald, T.W., Fitzgerald, T., Ayub, Q., Massaia, A., Smith, B.H., Smith, B.B., Dominiczak, A.F., Dominiczak, A.A., Morris, A.D., Morris, A.A., Porteous, D.J., Porteous, D.D., Hurles, M.E., Tyler-Smith, C., Xue, Y., 2015. Copy number variation in the human Y chromosome in the UK population. Hum. Genet. 134, 789–800. doi:10.1007/s00439-015-1562-5
- Weismann, A., 1889. The significance of sexual reproduction in the theory of natural selection, in: Essays upon Heredity and Kindred Biological Problems, Vol I. Poulton E.B., Schönland S., Shipley A. E., Oxford, pp. 255–332.
- Werren, J.H., 1987. The coevolution of autosomal and cytoplasmic sex ratio factors. Journal of Theoretical Biology 124, 317–334. doi:10.1016/S0022-5193(87)80119-4
- Werren, J.H., Baldo, L., Clark, M.E., 2008. *Wolbachia*: master manipulators of invertebrate biology. Nat Rev Micro 6, 741–751. doi:10.1038/nrmicro1969
- Werren, J.H., Beukeboom, L.W., 1998. Sex Determination, Sex Ratios, and Genetic Conflict. Annual Review of Ecology and Systematics 29, 233–261. doi:10.1146/annurev.ecolsys.29.1.233
- Werren, J.H., Zhang, W., Guo, L.R., 1995. Evolution and phylogeny of *Wolbachia*: reproductive parasites of arthropods. Proc. Biol. Sci. 261, 55–63. doi:10.1098/rspb.1995.0117
- Wessler, S.R., 2006. Transposable elements and the evolution of eukaryotic genomes. PNAS 103, 17600–17601. doi:10.1073/pnas.0607612103
- Wexler, J.R., Plachetzki, D.C., Kopp, A., 2014. Pan-metazoan phylogeny of the DMRT gene family: a framework for functional studies. Dev. Genes Evol. 224, 175–181. doi:10.1007/s00427-014-0473-0
- Whitton, J., Sears, C.J., Baack, E.J., Otto, S.P., 2008. The Dynamic Nature of Apomixis in the Angiosperms. International Journal of Plant Sciences 169, 169–182. doi:10.1086/523369
- Wibbels, T., Crews, D., 1994. Putative aromatase inhibitor induces male sex determination in a female unisexual lizard and in a turtle with temperature-dependent sex determination. J. Endocrinol. 141, 295–299.
- Wicker, T., Sabot, F., Hua-Van, A., Bennetzen, J.L., Capy, P., Chalhoub, B., Flavell, A., Leroy, P., Morgante, M., Panaud, O., Paux, E., SanMiguel, P., Schulman, A.H., 2007. A unified classification system for eukaryotic transposable elements. Nat Rev Genet 8, 973–982. doi:10.1038/nrg2165
- Wilkins, A.S., 1995. Moving up the hierarchy: a hypothesis on the evolution of a genetic sex determination pathway. Bioessays 17, 71–77.
- Williams, T.M., Carroll, S.B., 2009. Genetic and molecular insights into the development and evolution of sexual dimorphism. Nat. Rev. Genet. 10, 797–804. doi:10.1038/nrg2687
- Wilson, E.B., 1905. The Chromosomes in Relation to the Determination of Sex in Insects. Science 22, 500–502. doi:10.1126/science.22.564.500
- Winkler, C., Hornung, U., Kondo, M., Neuner, C., Duschl, J., Shima, A., Scharl, M., 2004. Developmentally regulated and non-sex-specific expression of autosomal dmrt genes in embryos of the Medaka fish (*Oryzias latipes*). Mech. Dev. 121, 997–1005. doi:10.1016/j.mod.2004.03.018
- Wootton, R.J., Smith, C., 2014. Reproductive Biology of Teleost Fishes. John Wiley & Sons.

- Wu, B., Novelli, J., Jiang, D., Dailey, H.A., Landmann, F., Ford, L., Taylor, M.J., Carlow, C.K.S., Kumar, S., Foster, J.M., Slatko, B.E., 2013. Interdomain lateral gene transfer of an essential ferrochelatase gene in human parasitic nematodes. *Proc Natl Acad Sci U S A* 110, 7748–7753. doi:10.1073/pnas.1304049110
- Wybouw, N., Pauchet, Y., Heckel, D.G., Van Leeuwen, T., 2016. Horizontal gene transfer contributes to the evolution of arthropod herbivory. *Genome biology and evolution* 8, 1785–1801.
- Xi, Z., Khoo, C.C.H., Dobson, S.L., 2005. Wolbachia Establishment and Invasion in an *Aedes aegypti* Laboratory Population. *Science* 310, 326–328. doi:10.1126/science.1117607
- Xue, W., Li, J.-T., Zhu, Y.-P., Hou, G.-Y., Kong, X.-F., Kuang, Y.-Y., Sun, X.-W., 2013. L_RNA_scaffolder: scaffolding genomes with transcripts. *BMC Genomics* 14, 604. doi:10.1186/1471-2164-14-604
- Yager, J., 1989. The male reproductive system, sperm, and spermatophores of the primitive, hermaphroditic, remipede crustacean *Speleonectes benjamini*. *Invertebrate Reproduction & Development* 15, 75–81. doi:10.1080/07924259.1989.9672024
- Ye, C., Hill, C.M., Wu, S., Ruan, J., Ma, Z.S., 2016. DBG2OLC: Efficient Assembly of Large Genomes Using Long Erroneous Reads of the Third Generation Sequencing Technologies. *Sci Rep* 6, 31900. doi:10.1038/srep31900
- Ye, C., Ma, Z. (Sam), 2016. Sparc: a sparsity-based consensus algorithm for long erroneous sequencing reads. *PeerJ* 4. doi:10.7717/peerj.2016
- Ye, C., Ma, Z.S., Cannon, C.H., Pop, M., Yu, D.W., 2012. Exploiting sparseness in de novo genome assembly. *BMC Bioinformatics* 13, S1. doi:10.1186/1471-2105-13-S6-S1
- Ye, J., Coulouris, G., Zaretskaya, I., Cutcutache, I., Rozen, S., Madden, T.L., 2012. Primer-BLAST: A tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics* 13, 134. doi:10.1186/1471-2105-13-134
- Yen, J.H., Barr, A.R., 1973. The etiological agent of cytoplasmic incompatibility in *Culex pipiens*. *Journal of Invertebrate Pathology* 22, 242–250. doi:10.1016/0022-2011(73)90141-9
- Yen, J.H., Barr, A.R., 1971. New hypothesis of the cause of cytoplasmic incompatibility in *Culex pipiens* L. *Nature* 232, 657–658.
- Yi, W., Zarkower, D., 1999. Similarity of DNA binding and transcriptional regulation by *Caenorhabditis elegans* MAB-3 and *Drosophila melanogaster* DSX suggests conservation of sex determining mechanisms. *Development* 126, 873–881.
- Yoshida, K., Kitano, J., 2012. The Contribution of Female Meiotic Drive to the Evolution of Neo-Sex Chromosomes. *Evolution* 66, 3198–3208. doi:10.1111/j.1558-5646.2012.01681.x
- Yoshida, K., Makino, T., Kitano, J., 2017. Accumulation of Deleterious Mutations on the Neo-Y Chromosome of Japan Sea Stickleback (*Gasterosteus nipponicus*). *J. Hered.* 108, 63–68. doi:10.1093/jhered/esw054
- Yoshimoto, S., Okada, E., Umemoto, H., Tamura, K., Uno, Y., Nishida-Umehara, C., Matsuda, Y., Takamatsu, N., Shiba, T., Ito, M., 2008. A W-linked DM-domain gene, DM-W, participates in primary ovary development in *Xenopus laevis*. *PNAS* 105, 2469–2474. doi:10.1073/pnas.0712244105
- Yusa, Y., 2007. Nuclear Sex-Determining Genes Cause Large Sex-Ratio Variation in the Apple Snail *Pomacea canaliculata*. *Genetics* 175, 179–184. doi:10.1534/genetics.106.060400
- Yusa, Y., 2005. Genetics of sex-ratio variation inferred from parent–offspring regressions and sib correlations in the apple snail *Pomacea canaliculata*. *Heredity* 96, 100–105. doi:10.1038/sj.hdy.6800758

- Yusa, Y., Yoshikawa, M., Kitaura, J., Kawane, M., Ozaki, Y., Yamato, S., Høeg, J.T., 2012. Adaptive evolution of sexual systems in pedunculate barnacles. *Proceedings of the Royal Society of London B: Biological Sciences* 279, 959–966. doi:10.1098/rspb.2011.1554
- Zabalou, S., Apostolaki, A., Pattas, S., Veneti, Z., Paraskevopoulos, C., Livadaras, I., Markakis, G., Brissac, T., Merçot, H., Bourtzis, K., 2008. Multiple rescue factors within a *Wolbachia* strain. *Genetics* 178, 2145–2160. doi:10.1534/genetics.107.086488
- Zarkower, D., 2013. DMRT Genes in Vertebrate Gametogenesis. *Current Topics in Developmental Biology, Gametogenesis* 102, 327–356. doi:10.1016/B978-0-12-416024-8.00012-X
- Zha, X., Xia, Q., Duan, J., Wang, C., He, N., Xiang, Z., 2009. Dosage analysis of Z chromosome genes using microarray in silkworm, *Bombyx mori*. *Insect Biochemistry and Molecular Biology* 39, 315–321. doi:10.1016/j.ibmb.2008.12.003
- Zhang, Z.-Q., 2013. Animal biodiversity: An outline of higher-level classification and survey of taxonomic richness (Addenda 2013). *Zootaxa* 3703, 1–82.
- Zhou, Q., Zhang, J., Bachtrog, D., An, N., Huang, Q., Jarvis, E.D., Gilbert, M.T.P., Zhang, G., 2014. Complex evolutionary trajectories of sex chromosomes across bird taxa. *Science* 346, 1246338. doi:10.1126/science.1246338
- Zhou, R.-N., Hu, Z.-M., 2007. The Development of Chromosome Microdissection and Microcloning Technique and its Applications in Genomic Research. *Curr Genomics* 8, 67–72.
- Zhu, L., Wilken, J., Phillips, N.B., Narendra, U., Chan, G., Stratton, S.M., Kent, S.B., Weiss, M.A., 2000. Sexual dimorphism in diverse metazoans is regulated by a novel class of intertwined zinc fingers. *Genes Dev.* 14, 1750–1764.
- Zug, R., Hammerstein, P., 2012. Still a host of hosts for *Wolbachia*: analysis of recent data suggests that 40% of terrestrial arthropod species are infected. *PLoS ONE* 7, e38544. doi:10.1371/journal.pone.0038544

Annexe I : Numéros d'accension des génomes de *Wolbachia* utilisés pour la recherche d'insertions dans le génome d'*A. nasatum*.

Numéro d'accension (GenBank)	Descriptif (NCBI)
GCA_000008025.1	<i>Wolbachia</i> endosymbiont of <i>Drosophila melanogaster</i> (a-proteobacteria)
GCA_000008385.1	<i>Wolbachia</i> endosymbiont strain TRS of <i>Brugia malayi</i> (a-proteobacteria)
GCA_000022285.1	<i>Wolbachia</i> sp. wRi (a-proteobacteria)
GCA_000073005.1	<i>Wolbachia</i> endosymbiont of <i>Culex quinquefasciatus</i> Pel (a-proteobacteria)
GCA_000153585.1	<i>Wolbachia</i> endosymbiont of <i>Drosophila willistoni</i> TSC#14030-0811.24 (a-proteobacteria)
GCA_000167475.1	<i>Wolbachia</i> endosymbiont of <i>Drosophila ananassae</i> (a-proteobacteria)
GCA_000174095.1	<i>Wolbachia</i> endosymbiont of <i>Muscidifurax uniraptor</i> (a-proteobacteria)
GCA_000204545.1	<i>Wolbachia</i> endosymbiont wVitB of <i>Nasonia vitripennis</i> (a-proteobacteria)
GCA_000208785.1	<i>Wolbachia</i> endosymbiont of <i>Culex pipiens molestus</i> (a-proteobacteria)
GCA_000242415.3	<i>Wolbachia pipientis</i> wAlbB (a-proteobacteria)
GCA_000306885.1	<i>Wolbachia</i> endosymbiont of <i>Onchocerca ochengi</i> (a-proteobacteria)
GCA_000331595.1	<i>Wolbachia</i> endosymbiont of <i>Diaphorina citri</i> (a-proteobacteria)
GCA_000333775.1	<i>Wolbachia pipientis</i> wBol1-b (a-proteobacteria)
GCA_000333795.2	<i>Wolbachia</i> endosymbiont str. valsugana of <i>Drosophila suzukii</i> (a-proteobacteria)
GCA_000338375.1	<i>Wolbachia</i> endosymbiont of <i>Onchocerca volvulus</i> (a-proteobacteria)
GCA_000338395.1	<i>Wolbachia</i> endosymbiont of <i>Wuchereria bancrofti</i> (a-proteobacteria)
GCA_000376585.1	<i>Wolbachia</i> endosymbiont of <i>Drosophila simulans</i> wNo (a-proteobacteria)
GCA_000530755.1	<i>Wolbachia</i> endosymbiont of <i>Onchocerca volvulus</i> str. Cameroon (a-proteobacteria)
GCA_000689175.1	<i>Wolbachia</i> endosymbiont of <i>Glossina morsitans morsitans</i> (a-proteobacteria)
GCA_000723225.1	<i>Wolbachia</i> endosymbiont wPip_Mol of <i>Culex molestus</i> (a-proteobacteria)
GCA_000742435.1	<i>Wolbachia pipientis</i> (a-proteobacteria)
GCA_000829315.1	<i>Wolbachia</i> endosymbiont of <i>Cimex lectularius</i> (a-proteobacteria)
GCA_000832695.1	<i>Wolbachia</i> endosymbiont of <i>Dactylopius coccus</i> (a-proteobacteria)
GCA_001266585.1	<i>Wolbachia</i> endosymbiont of <i>Operophtera brumata</i> (a-proteobacteria)
GCA_001439985.1	<i>Wolbachia</i> endosymbiont of <i>Trichogramma pretiosum</i> (a-proteobacteria)
GCA_001637495.1	<i>Wolbachia</i> endosymbiont of <i>Laodelphax striatella</i> (a-proteobacteria)
GCA_001648015.1	<i>Wolbachia</i> endosymbiont of <i>Dactylopius coccus</i> (a-proteobacteria)
GCA_001648025.1	<i>Wolbachia</i> endosymbiont of <i>Dactylopius coccus</i> (a-proteobacteria)
GCA_001675695.1	<i>Wolbachia</i> endosymbiont of <i>Nomada flava</i> (a-proteobacteria)
GCA_001675715.1	<i>Wolbachia</i> endosymbiont of <i>Nomada leucophthalma</i> (a-proteobacteria)
GCA_001675775.1	<i>Wolbachia</i> endosymbiont of <i>Nomada panzeri</i> (a-proteobacteria)
GCA_001675785.1	<i>Wolbachia</i> endosymbiont of <i>Nomada ferruginata</i> (a-proteobacteria)
GCA_001752665.1	<i>Wolbachia pipientis</i> (a-proteobacteria)
GCA_001758565.1	<i>Wolbachia</i> endosymbiont of <i>Drosophila incompta</i> (a-proteobacteria)
GCA_001758585.1	<i>Wolbachia</i> endosymbiont of <i>Drosophila incompta</i> (a-proteobacteria)

Annexe II : Matériels et Méthodes Général

Extraction ADN : méthode d'extraction par résine de Chelex100 (Biorad)

Cette extraction est réalisée à partir d'une patte prélevée d'un individu, qui a ensuite été broyée à l'aide de pinces fines et placée dans 120 μL d'eau bi-distillée filtrée à 0.2 μ . Afin de faciliter la lyse des tissus, 10 μL de protéinase K (concentrée à $20\text{mg}.\mu\text{L}^{-1}$) ont ensuite été ajoutés. Un volume de 20 μL de résine Chelex100 (Biorad) à 5% est ensuite ajouté. La préparation a ensuite été homogénéisée avec un *MixMate* (Eppendorf) durant 1 minute à 2 000 rpm. L'extrait a ensuite été incubé durant 3h à 56°C , et a été homogénéisé de la même façon que précédemment 2 fois au cours des 3h d'incubation. Une fois l'incubation terminée, l'échantillon a ensuite été homogénéisé une dernière fois avec un *MixMate* (Eppendorf) pendant 1 minutes à 2 000 rpm. La préparation a ensuite été incubée 10 minutes à 95°C puis a été centrifugée brièvement afin de répartir les débris tissulaires au fond du tube.

Protocole de PCR

Toutes les réaction de PCR ont été réalisées avec l'enzyme GoTaq polymérase (Promega). Le mélange réactionnel était composé de 3 μL de tampon 5X, 0.3 μL de dNTP (la concentration de chaque nucléotide était de 2.15 mM), de 0.15 μL de GoTaq (concentrée à 5 $\text{U}.\mu\text{L}^{-1}$) et de 0.42 μL de chaque amorce (10mM). Chaque PCR débutait par une phase de dénaturation de l'ADN de 3 minutes à 94°C , suivie d'une série de 3 étapes répétées 35 fois : (i) une étape de dénaturation de l'ADN pendant 30 secondes à 94°C ; (ii) une étape d'hybridation des amorces pendant 30 secondes à la température de fusion des amorces et (iii) une étape d'élongation à 72°C pendant 1 minute. Enfin, chaque PCR s'est achevée après une phase d'élongation finale de 10 minutes à 72°C .

Visualisation des produits de PCR

Les produits de PCR ont ensuite été séparés sur un gel d'agarose à 1.5% suite à une étape de migration durant 20 à 50 minutes (selon la taille du gel) à 100 V. Les amplifications ont été visualisées sous une lumière UV après un bain d'au moins 15 minutes dans une solution aqueuse de bromure d'éthidium ($0.2 \mu\text{g.mL}^{-1}$).

Transcription inverse (RT)

La phase de transcription inverse a été réalisée à partir du kit SuperScript SSIII (Invitrogen) en suivant les indications fournies. Toutes les réactions de RT ont été réalisées à partir de 250 ng d'ARN, dont la concentration a été mesurée au *Qubit 2.0* (Invitrogen) au préalable. Afin de s'assurer de l'absence d'ADN contaminant, des témoins négatifs ont été réalisés pour chaque réaction et consistaient à remplacer l'enzyme SSIII par de l'eau bidistillée. Les réactions de PCR ont ensuite été réalisées selon la procédure décrite précédemment.

qRT-PCR

La quantification de chaque échantillon a été réalisée relativement par rapport à deux gènes de références dont le niveau est stable en fonction des conditions expérimentales (cf Matériels et Méthodes du chapitre 2). Les réactions ont été réalisées à partir d'un mélange composé de 5 μL de Fast SYBR GREEN Master Mix 2X (Roche), 0.5 μL de chaque amorce (concentrées à 10 mM), de 2,5 μL d'ADN complémentaire synthétisé lors de la phase de transcription inverse et de 1.5 μL d'eau bidistillée. Tout d'abord, l'ADNc a été dénaturé pendant 10 minutes à 95°C. Ensuite la réaction a suivi une série de 3 étapes répétées 45 fois : (i) une étape de dénaturation de l'ADN pendant 10 secondes à 95°C; (ii) une étape d'hybridation des amorces pendant 10 secondes à 60°C et (iii) une étape d'élongation à 72°C pendant 20 secondes. Une courbe de fusion a été réalisée à la fin des cycles en augmentant la température de $0.11^\circ\text{C.s}^{-1}$ jusqu'à atteindre le seuil de 97°C (mesure de la fluorescence tous les 5°C). Cette courbe a été réalisée afin de s'assurer de la spécificité de l'amplification.

Annexe III : Séquences et caractéristiques des amorces testées pour la recherche de séquences candidates Y-spécifiques.

Identifiant de séquence	Séquence des amorces (5'>3')		Taille du produit de PCR (nucléotide)	Température de fusion
	Forward	Reverse		
scf7180001667967	TTTCACCACTTGTCCAACGA	AGAACCCAGCTGATAGAAGGT	322	54°C
scf7180001680167	AATTATTTTGGGCTCTCAAGACA	CCTGGAGAAAAGTGGATGACG	197	54°C
scf7180001690334	CTCAGAAAGTCAATCAACCAT	ATGAGGAAACGATAAATAAATAG	296	48°C
scf7180001768809	CAGCGAATTTTTCAGTTACATT	TGGACGTTTCTTATGCAGGA	404	50°C
scf7180001821350	AACGGTGTTCCTATCACTC	GTAGCGTTTCTTCGCCACTA	242	54°C
scf7180001817225	ATATCTATCATCAACGATCAAA	AACAAAGCTGAAGTATAGAGA	205	48°C
scf7180001796744	GGGAGACTAAAATTTACTGTTG	TTCTTGCTCTATTGTATAATTGAA	413	48°C
scf7180001793998	ACCTGAGATATTGTCATAAATAAA	AAGTCTATTATTCTTTACAGGT	274	48°C
scf7180001777099	TCGTCGTAAAGAAATTTGCACC	TGTTACAGTGTAGTCAGAAAG	266	54°C
scf7180001724780	TGTGCTTTTGAAACTATTTTATT	TAAAACTCATCCCATAAAGC	439	48°C
scf7180001707538	TGCAGGCTATGGGAAAATGA	ATCTCACTATAGTCCGGCTT	317	54°C
scf7180001654076	CTTGACACAAATTTCAAAGTA	AAAATTTTGTGCAAAAATC	438	45°C
scf7180001687914	CCCATTATTTGAAACTACAGTC	GAAATCATAGGTTATTTAATTTGG	101	48°C
scf7180001705172_1	TAGTAGTAGGGGTTCGAGGG	TGGGAACGTATGATTCAAGGA	277	53°C
scf7180001705172_2	GCTCTCTCGTAATGTGTCC	TAAACCTAGGCATAGCGGA	224	53°C
scf7180001668770	AACAAATAGATGAAGTAGCCT	CTGTCTGTATGCTTTAACA	229	48°C
scf7180001797263	ATTGAAAGTGAAGTTCGGGG	CCTGCTTCTATTAGACCTCG	322	52°C
scf7180001792357	GCTCTGTGATTTCTAAAGAATC	TTTAAATCATTTCTGCAACTTT	412	48°C
scf7180001797263	TAATAACTTAAACATGAAATTCG	TCATTACATTAGAGGTCACGTA	347	48°C
scf7180001694435	AAATCTTTTCTCCTTCAGCAC	CAGTGACGTCATTTTCGC	226	52°C
scf7180001729141	CGTTAAAGGTACGATTG	TCTTATCTTAGTATGTCTTC	253	43°C
scf7180001666372	TACGGGTAAAGTCAGTCAGAT	TCCGTATCCACAACAAACAT	285	52°C
scf7180001725282	AAATGAATACCACCGTTGA	ATATGCGTATCTGTGGCAAT	357	52°C
scf7180001736667	CCAAGAAATGTCAATAACGGC	GGGTGCAGAGAACTAAGAAA	199	52°C
scf7180001770442	TGACCCTTTCTAACTACCA	TATCTAAAGTCAGAGGGCGA	256	52°C
scf7180001778226	CCAAAACCTTTTACGTCGG	TTAGTCGTTCTCAGTTTCC	295	52°C
scf7180001827924	AACCACTTATCTAGCAACCC	TAGGTATGCAGTCACATTGC	274	52°C
scf7180001715683	TAAGCTCTTAACCACTGC	TGCATAAAGAGAAGGGTTCG	327	52°C
scf7180001832287	TGCCAAATAAGCTCTTCTCA	AACCCATTATGGAGAATGCC	231	52°C
scf7180001835244	CCTAAGCCTAAATCTGTATTCTT	AATGGTGGGCGAGTTTATCA	132	52°C
scf7180001715110	CGACAAGCTCTAGTCAATCT	GTGATATAGGTTCCAACCCC	169	52°C
scf7180001765933	TTTAGTCGCAACATGTCAGA	TGGAATCGTCGTTTGTGTTT	206	52°C
scf7180001766586	TGAAAGATCGCGTTAGTAGG	CTGTGGTACTCAAAGCCTTA	180	52°C
scf7180001821584	AAGGTCGCAATAAATGAGGA	AGGACAACCTGATTAGGTCTG	183	52°C
scf7180001704461	GAAGGATTAGGGTGTCAAGG	CTGTTAGTGGGACTAGCATC	218	52°C
scf7180001728456	GAGAAGCGCAATTTCAAAGGT	GAAAGTTAGAGTCCCTCGT	235	53°C
scf7180001735089	GTAACCTCGGACTACACACAA	GCTTTCCCGTTTACTGTTAC	390	51°C
scf7180001763822	AATGGTCTCCCTTCATTCC	TAAAGGGTCAAGTGAATCGG	294	52°C
scf7180001703882	AATCAACCATCTTGCCATCT	CATTACGCTCCTCTATACGG	200	52°C
scf7180001757839	ACGAAGACAATACTGAGTGG	GAAGGTTTGTACTGTCGGTT	150	52°C
scf7180001773145	TATACATCGAAATGTATGC	TATGTAGACGCAATATACATG	247	45°C
scf7180001550276	TTTCCAAACTTCGAGGTAG	CTTTTAGTTGGTTCGGTTCG	260	56°C
scf7180001720150	GTACCAACGGATGCTTGTAT	TGTGAGATTTTGACACCCAA	316	52°C
scf7180001728284	CTATCTCCTTTGAGGATGGC	TGTCCCAAAATCTGTGTTGA	188	52°C
scf7180001829015	TGACAGTTGATTCCCTCATC	TGTATTTAAGGGCTAGGAGC	160	51°C
scf7180001583976	GATACTACTCCCTCGAGCTT	GTCCAGTTTAAACCTTGGGT	189	52°C
scf7180001608539	TAGTACAACCTACCAGGGCAT	AAGGTTCTATAAACCCGAGC	227	52°C
scf7180001678246	TGAAGGTTGATAGTGTGTGG	TGCTGAGAGACATTTTCCAT	287	51°C
scf7180001685477	ATCGATTGAGTGAGGGGTTA	TTTAACCCACCATTAACCCAC	135	54°C
scf7180001687135	TATTGTAAGGGGTGGGGTT	GACTCCATAACAAGCTCGAA	172	54°C
scf7180001712093	GACAGAAAATCGTGTGTGG	CACCCATAAACCTTGCAGATA	131	54°C
scf7180001716919	TCCATCAACAGTGAAGTGC	GTGGACGAATAAAGTTCCCT	144	54°C
scf7180001719166	TCAGATCATTTCTCGTTCG	GTAGAAGCTGGCAATAAGGA	265	52°C
scf7180001726347	GTTCTAGATCGACATTCCGT	AGAGGAGGTTTCTACGAAGT	302	51°C
scf7180001735089	GCAGGGAAGTATTACTGCTA	TGTGATGGATCTCATCTGAC	163	51°C
scf7180001831990	TACATTGTCTAACGGCTACG	AGAGATTAAGCTCACGTTCC	368	51°C
scf7180001737975_1	ATGAGTCTGTTCGGTTTCTG	AGTTGACGGTTTACCTTGG	175	48°C
scf7180001737975_2	CTCCGATTTTGAGCTTATGGA	TTTGACTCTTCCCAACAGTTTA	262	51°C
scf7180001737975_3	AGGAAAGTGGAGATTCTTAC	TGTGCCGTAGATGTTATTGG	120	51°C
scf7180001836677	TTGTCAATTGCATTTCTGTC	AGGGCTGAGTTCTGTTATT	219	51°C
scf7180001675306	GATGACCAAAATACTCACTCG	ATACCTCTCCGTCTAAATGC	114	51°C
scf7180001718297_1	TACAGTAGGCTCAGATTGGA	AGTTACGGAAATGGAGTCAC	290	52°C
scf7180001718297_2	TCTTTAAGTCACAAACGGGT	GAGTAGGTGGAATGCGTTTA	203	51°C

Annexe IV : Numéros d'accession des séquences utilisées comme base de données pour la recherche de gènes impliqués dans le déterminisme du sexe.

Numéros d'accession - Base de données *Doublesex*

AGB95735	XP_010185565	XP_008474587	XP_007090876	XP_005507218	XP_004428451	AAC38956
EFA07556	XP_010081974	AIE16095	XP_007077581	XP_005507217	XP_004417557	AAR23813
BAN82533	XP_010176727	AIE16094	XP_007077580	XP_005498930	XP_004416091	P19018
BAN82532	XP_010146952	XP_008400653	XP_007077480	XP_005522822	XP_004416088	NP_082772
BAM93342	XP_010143412	XP_008433910	XP_006998467	XP_005522821	XP_004400431	CDQ04911
BAM93339	XP_010044444	XP_008334418	XP_006989884	XP_005521354	XP_004388803	CDQ04608
BAM33613	XP_010068313	XP_008334417	XP_006977780	XP_005521353	XP_004373619	CDP96133
Q8JIR6	XP_009991920	XP_008333719	XP_006977707	XP_005428005	XP_004373589	CDN30080
NP_083654	XP_009991891	XP_008320425	XP_006974063	XP_005428003	XP_004373588	EYC00414
NP_783578	XP_009982327	XP_008291535	XP_006941270	XP_005423049	XP_004372106	AHF81656
NP_001089148	XP_009969033	XP_008289050	XP_003995573	XP_003444528	XP_004372099	CAP22659
NP_063925	XP_009979988	XP_008274623	XP_003995428	XP_003447317	XP_004331636	CAP22218
NP_001005779	XP_010013588	XP_008274622	XP_003990241	XP_005432996	XP_004330868	ERG79893
NP_001007065	XP_009955637	XP_008274619	XP_003990232	ERG85761	XP_004325875	NP_001256883
NP_001033034	XP_009949529	XP_008274618	XP_006908422	ERG82622	XP_004324226	NP_741551
NP_001036871	XP_009938078	JAC85141	XP_006919511	ERE83411	XP_004322303	AFV47371
NP_991191	XP_009936490	XP_008252162	XP_006919506	ERE77350	XP_004315690	AFV47370
BO1IG7	XP_009936489	XP_008271033	XP_006918458	XP_005415255	XP_004312545	AFV47369
NP_001073445	XP_009911906	XP_008271032	XP_006903939	XP_005412484	XP_004276482	AFV47368
NP_056641	XP_00996376	XP_008271029	XP_006871451	XP_005401248	EMP30456	AFV47367
NP_001281161	XP_009900057	XP_008271028	XP_006864558	XP_005398829	XP_004002005	AFV47366
NP_001104725	XP_009900054	XP_008263848	XP_006863797	XP_005388803	XP_003280926	AFV47365
AFQ62105	XP_009882912	XP_008263447	XP_006863796	EPQ18706	XP_002194499	EGT58957
AEX92944	XP_009882617	XP_002708232	XP_006863795	EPQ15248	XP_002194579	EGT54512
NP_001101421	XP_009864039	XP_002708098	XP_006839931	EPQ12836	XP_002158720	EFO95049
XP_008205423	XP_004226593	KDR19787	XP_006813729	NP_001232910	EMC79089	EFO91664
BAO23811	XP_009700820	KDR14494	XP_006900983	XP_005342896	ELW67051	CAC42778
BAO23809	XP_009695797	XP_001603339	XP_006821777	XP_005335976	ELW65439	CAC42783
AGS48315	XP_009639799	XP_008199135	XP_006881430	XP_005326104	ELW63100	Q7KBL8
AGS48309	XP_009639798	XP_008174580	XP_006881400	XP_005353686	ELW63099	Q02427
NP_001033035	XP_009681624	XP_008166985	XP_006881386	XP_005352150	ELV09399	CAB01491
AFQ62108	XP_009675974	XP_005284898	XP_006881385	XP_005244443	XP_004087084	CAB05899
AEX92947	XP_009675322	XP_008160002	XP_006879872	XP_005244441	XP_003259177	CAB01490
ACY95408	XP_009667936	XP_008158466	XP_002734495	XP_005231730	ELR57086	AGR88662
ACN73402	XP_009560254	XP_008143272	XP_002732510	XP_005219381	ELR53307	AGR88661
BAJ25852	XP_009558978	XP_008137724	XP_006746523	XP_002686430	DAA26910	AGR88660
BAJ25850	XP_009554451	XP_008188194	XP_006774884	EPY84582	ELK32318	Q8NF50
ADG37648	XP_009554449	XP_005296971	XP_006768453	EPY83517	ELK26739	O02008
ACJ65505	XP_009498848	XP_001949339	XP_006765373	EPY74925	ELK19630	P11596
ACJ65501	XP_009454536	XP_001950330	XP_006806330	XP_005155070	ELK15685	Q6IUP1
ABF50952	XP_009460512	XP_975675	XP_006797837	XP_005151344	ELK07038	P70218
AAK38833	XP_009489941	XP_008072463	XP_006797793	XP_005151337	ELK06173	CEF71611
AAK38832	XP_009356150	XP_008070856	XP_006794474	XP_005060932	ELK06140	CEF64124
ACY78702	XP_009278229	XP_008069229	AHG06462	XP_005050120	JAA57828	NP_001038717
ACY78701	XP_009278227	XP_008064268	XP_006635098	XP_005088421	XP_004064446	XP_009169810
ACY78700	KGB36073	XP_008050656	XP_006634975	XP_005087858	XP_004047935	XP_009169560
AAZ48939	KGB34794	XP_008122698	XP_006630066	XP_005081289	XP_004047798	NP_001002911
AAZ78362	KGB33162	XP_008107806	XP_006627187	XP_005074348	XP_004025823	AAF48261
AIY68268	XP_009092297	XP_008101683	XP_006627186	XP_005072298	XP_004005353	AAF56919
KHG25876	XP_009092296	XP_008101529	XP_006627185	XP_005096933	XP_004004391	AAF55843
XP_010344492	XP_009087161	AHZ95776	XP_006624379	XP_005096931	BAM62887	CDQ08335
XP_010337086	KFZ58784	XP_003220348	XP_006611457	JAA93927	EKC42823	CDN30083
XP_010221441	KFW02464	XP_003216601	XP_006607559	WP_016643793	EKC22641	KFD62256
XP_010179983	KFV82739	XP_003216535	XP_006570066	XP_005029316	EGW113936	KFD48631
XP_010143385	KFV16031	XP_007990272	XP_006528443	XP_005011697	EGW09718	KFB41255
XP_009957423	KFV15427	XP_007954185	XP_006528442	XP_005011695	EHB18221	WP_022810525
XP_009818793	XP_009078346	XP_007954175	ETN58485	XP_005004778	EHB16424	WP_025041374
XP_009667934	XP_009082875	XP_007954174	AHC08007	XP_003463125	EHA98146	KDB23267
XP_009554450	KFQ94077	XP_007946600	XP_006268809	XP_003478909	EGI67132	EYC33706
XP_009460513	KFQ68670	XP_007944154	XP_006268224	XP_003472315	EGI62789	EYC21849
XP_009285756	KFQ40473	XP_007941145	XP_006261330	XP_003464114	EGS57411	EYB85570
KGB38632	KFQ25870	EAWS8818	XP_006171643	XP_003643035	EFN87725	EYB84210
XP_009206658	KFQ15129	XP_007909461	XP_006171637	XP_004875581	EFN82716	CDJ84121
AIL86779	KFP85300	XP_007898236	XP_006171633	XP_004873128	EFN81038	CDJ85098
XP_006527244	KFP79117	XP_007898235	XP_006167088	XP_004838410	EFN77976	AHF81661
XP_006540420	KFP75534	XP_007890871	XP_006152568	XP_004838268	EFN71326	AHF81660
CDS33635	KFO96444	XP_007890813	XP_006152505	XP_004838266	EFN67821	XP_006623927
CDS39593	KFO35450	XP_007890812	XP_006149434	XP_004908875	EFN63112	ETN82947
AAF49460	KFO31254	EUB63223	XP_006142049	XP_004932028	XP_003803744	ETN81016
XP_008758529	KFO26519	XP_007631618	XP_006218537	XP_004906078	XP_003799503	BAO23812
XP_008758523	KFO26517	XP_007631617	XP_006217254	XP_004930266	XP_003793190	BAO23813
XP_008554572	KFO20944	XP_007629749	XP_006217252	XP_004902092	XP_003783067	CAP36235
XP_008554550	XP_008689150	XP_007627155	XP_006215193	XP_004902091	XP_003782965	CAP29170
ETE67544	XP_006977698	XP_007623043	XP_006208660	XP_004924389	XP_003761708	AGW99161
ETE67477	XP_006528342	XP_003514808	XP_006200637	XP_004877115	EDS41189	EAL28163

ETE67265	XP_006503315	XP_007642746	XP_006200631	XP_002931473	EEC09723	CAP38316
ETE60372	XP_006104352	XP_003515526	XP_006181640	XP_429193	EDL90429	AFK01539
NP_001273240	XP_005874536	XP_003495792	XP_006180217	XP_004827309	XP_003742524	WP_018185150
NP_001243149	XP_004637231	XP_007664048	AHA85564	XP_004821839	XP_786938	XP_004529238
NP_001093726	NP_001178690	XP_007653748	AGN49325	XP_004803962	BAM20792	NP_001084923
NP_001265833	XP_008953193	XP_007653746	XP_006133574	XP_004803170	BAM20278	BAM93350
NP_001254571	XP_008975314	XP_007653745	XP_006104351	XP_004803169	EFX89054	NP_001041089
NP_001133069	XP_002763045	XP_001506523	XP_006129005	XP_004785000	EDL41632	NP_871909
NP_001117741	XP_008966173	XP_001513790	XP_006126511	XP_004784697	AFJ20714	NP_495138
NP_001089725	XP_008960872	XP_004086504	XP_006099394	XP_004777329	XP_003705132	NP_510466
NP_001033037	XP_008999505	XP_004085543	XP_006087451	XP_004694204	ADN23487	NP_500305
EKC24466	XP_008994593	XP_004079711	XP_006087440	XP_004685930	AFA46804	EFO22858
ADI86270	XP_008943347	XP_007534265	XP_005186857	XP_004679550	AFA46803	EFO17042
ADI86269	XP_003812568	XP_007533177	XP_005184871	XP_004678078	AFA46802	EFO15216
EDS41190	XP_002742974	XP_007532974	XP_006025859	XP_004678055	AFA46801	AFJ15562
AEH68214	XP_008923483	XP_007530835	XP_006022663	XP_004677681	GAA52835	AFJ15561
NP_003700217	XP_008919961	XP_007530834	XP_006013577	XP_004714606	GAA51980	EJW87147
GAA34150	AIK28932	XP_007528525	XP_006004087	XP_004714605	CCD83081	EJW80816
GAA52246	CDS33776	XP_007506398	XP_006001262	XP_004714604	CCD82280	EJW77112
EHJ78136	CDS15764	XP_007480338	XP_006001259	XP_004710412	AAR89620	BAM33612
CCE46022	XP_008853819	XP_001374295	XP_005999966	XP_004638630	XP_003489387	BAM33611
ACY95410	XP_008844673	XP_001365681	XP_005978027	XP_004706683	AEN92271	BAM33610
ACY95407	XP_008839920	XP_001365618	XP_005976419	XP_004699710	BAK62678	EGT55583
ACY95406	XP_008839916	XP_001374014	XP_005966259	XP_004623190	AEK27010	EGT53972
AAN77230	XP_008836796	XP_001362692	JAB78361	XP_004671412	ADM07317	EGT47208
ACY78703	XP_008835512	EZA59919	XP_005911591	XP_004670369	ACD62474	EFX74782
JAG33399	XP_008830651	EZA58403	XP_005894812	XP_004660962	XP_002931394	EFP11031
JAG01169	XP_008827159	EZA53129	XP_005922062	XP_004660961	AAR89621	EFP10936
XP_010598309	NP_001020459	EZA49556	XP_005921989	XP_004658557	XP_002800204	EFP10925
XP_003411121	XP_008771551	XP_007471244	XP_005921988	XP_004653048	ADH94552	EFO88111
XP_003420509	XP_008765122	XP_007466821	XP_005886271	XP_004620764	AAO41737	EFO87235
XP_003420508	XP_008682363	XP_007442857	XP_005878744	XP_004607395	AAN61065	EDW96315
XP_003407697	XP_008709704	XP_007429329	XP_005874538	XP_004607380	AAP80401	EDW84778
XP_010608723	XP_008697616	XP_007423833	XP_005811528	XP_004600422	AAP80398	EDW59710
XP_010638444	XP_008693429	XP_007423832	XP_005811522	XP_004600388	ACN54528	EDW15889
XP_010636197	XP_008689241	XP_007423825	XP_005806418	XP_004600387	ACT83604	EDV94806
XP_009325572	XP_008689240	XP_007422318	XP_005678544	XP_004598114	CAQ52797	EDV43362
XP_010621579	XP_008636867	AHI47025	XP_005730702	XP_004592852	CAQ52796	EDV22336
XP_010581169	CDW54559	AHI47024	XP_005657892	XP_004591934	BAF44086	EDV19257
XP_010581168	AIK62892	XP_007257776	XP_005655960	XP_004591890	BAF44085	XP_003489222
XP_010581167	XP_008592272	XP_007257775	XP_005673457	XP_004591889	BAF44084	XP_003436697
XP_010407737	XP_008588993	XP_007257774	XP_005660196	XP_004588846	BAH05021	XP_001688199
XP_010391395	XP_008580957	XP_007257773	XP_003128031	XP_004535662	ACD74917	BAJ78309
XP_003945149	XP_008580604	XP_007235585	XP_003121928	XP_004521323	ACD74916	BAJ49799
XP_010344491	XP_008575480	XP_007231728	XP_005641556	XP_004484355	ACD62372	BAE24347
XP_010349250	XP_008573429	XP_007188097	XP_005629470	XP_004468901	ACG69837	ABM54574
XP_010349176	XP_008552718	XP_007182367	XP_005620332	XP_004460133	ACG69836	AAL18253
XP_010362718	XP_008552294	XP_007182288	XP_005616564	XP_004459455	ACG69835	AAL18252
XP_010306348	XP_008560071	XP_007164819	XP_005614335	XP_004459445	BAC65667	AAL02164
XP_010309156	XP_008506088	XP_007164816	XP_005607472	XP_004456130	ABB83370	AAL02161
XP_010292857	XP_008537704	XP_006061815	XP_005605197	AGH33736	ACS36669	XP_001901119
XP_010288429	XP_008510968	XP_006044531	XP_005605094	AGH69790	XP_001900187	AAL02163
KFP43239	XP_008508692	XP_007065505	XP_538686	EOA94527	ABQ44209	AAF72594
XP_010224483	XP_008501756	XP_007062267	XP_851495	AGJ00070	ABC48237	XP_002638637
XP_010221452	XP_008501702	XP_007062258	XP_005589480	AGJ00068	ABB72995	ACB59349
XP_010197371	XP_008500754	XP_007129268	XP_005589478	AGJ00067	AAY64468	CAC37947
XP_010194841	XP_008483100	XP_007128230	XP_005581667	XP_004443941	Q9PTQ7	ABK76309
XP_010121435	XP_008477210	XP_007108541	XP_005543380	XP_004440732	P85119	ABK88911
XP_010186953	XP_008474591	XP_007102230	XP_005511730	XP_004438631	NP_503176	CAC42777
CAG29242	CAD98335	BAE06379	FAA00223			

Numéros d'accession - Base de données *DMRT*

AEK27010	P57690	AGJ00067	NP_082772	NP_001014244	ACI43914	Q8BG36
ACI62766	NP_001093460	NP_001256883	NP_063925	NP_001101421	ACI43913	Q8CGW9
AAY64468	NP_00101274025	NP_001041089	AAF24509	CDQ04911	ACI43911	Q8CFG4
CAD44608	NP_999879	NP_503176	Q9QZ59	CDQ04608	ADK55063	Q96MA1
CAD44607	NP_001107909	NP_871909	Q9PTQ7	CDP96133	ADH15934	Q5VZB9
AAP84972	NP_001107930	NP_741551	D4A218	CDP96744	AAL83920	NP_956236
AFV47371	NP_001073445	NP_500305	COLZJ1	CDP98217	AAL83919	AGJ00070
AFV47370	NP_001004008	BAM37631	BOIIG7	CDQ08335	AAF79931	AGJ00069
AFV47369	NP_571027	BAM37632	Q9Y5R5	AEN11427	ACS36669	AGJ00068
AFV47368	XP_006723383	EFX89054	Q96SC8	AHI47025	BAF44086	AAY26896
AFV47367	XP_006528347	EFX79095	A4QN97	AHI47024	BAF44085	AAX08123
AFV47366	XP_006528342	AFA46804	P85119	ACV77084	BAF44084	BAE16955
AFV47365	XP_006528297	AFA46803	P83758	NP_001266410	ACB59349	BAE16953
Q2I327	XP_006503315	AFA46802	Q90WM5	AGZ63420	ABU92554	BAE16952
Q4QR87	XP_005259203	AFA46801	Q9Y5R6	NP_001179569	ABM53669	BAB63259
Q5HYR2	NP_001178690	ACI43917	Q8JIR6	NP_001232910	ABD77574	ABD62706
NP_001007065	NP_083654	ACI43916	Q80IF8	NP_001072447	ABD62707	XP_005219381
AAÿ26897	NP_037133	ACI43915	Q71MM5			

Numéros d'accession - Base de données *Transformer/Transformer-2*

AAF49441	XP_008884979	XP_007737639	XP_006352512	WP_006865573	EIE87805	XP_007877230
AAF67591	AIK25411	XP_007741073	ETE68519	WP_006409750	EFX88901	XP_007862612
AA53589	CEA07950	XP_007732152	ACY23449	WP_005192114	EFX69428	XP_007856967
AAO38911	CDS29959	AHY79612	XP_006272472	CCD55315	EFX66653	XP_007850789
ABW04174	CDS21675	XP_007724819	XP_006272471	XP_004615620	EFX65303	XP_007847123
AGM48362	XP_008843517	XP_007724461	XP_006259509	XP_004608986	EFO99465	XP_007843149
NP_001128299	XP_008839535	XP_007704551	XP_006258005	XP_004591293	EFO98598	XP_007767617
AGE31794	CDI56510	EYU40880	ESZ97062	XP_004499153	EFO88717	XP_007305599
ACS34687	CDM30702	XP_003500170	ESZ93571	XP_004491420	EFJ29491	XP_004073891
BAI66432	ETS80459	XP_007653736	ESZ91658	XP_004562584	EFA10022	XP_007485990
ACD39776	AFZ80294	XP_007658522	XP_006151326	XP_004562580	EEE55887	EST06099
ACY40710	EKKX73265	CDO88714	XP_006211868	XP_004562575	EEC50424	EST05979
ACY40709	KFH71949	XP_004086588	XP_006175504	XP_004556218	EDW90656	XP_007407078
AAF55934	KFH65370	XP_004080039	ESU42152	XP_004521799	EDW69741	XP_007392122
Q24761	KFH47644	XP_004079780	ESU39179	WP_003645590	EDW56932	XP_007381759
Q23949	NP_001169549	XP_004078986	EST24453	WP_001944160	EDW47683	XP_007337845
Q23935	ACL83207	XP_007532521	XP_006119565	WP_008716624	EDW32046	XP_007364416
NP_001119708	KFH16620	XP_007531001	XP_006119564	CCE29090	EDW30353	XP_007327850
NP_476763	KFH05054	XP_007543408	XP_005191965	NP_001084800	EDW17317	XP_007315263
AAM88673	KFG83504	XP_007574517	XP_006025722	ENI01203	EDV95829	EYB82470
CAG29241	KFG83294	XP_007574155	XP_006022615	ENH80563	EDV93180	XP_007246713
AIK25400	NP_001150405	XP_007574149	XP_006017419	ENH79054	EDV56202	XP_007240738
AGI50962	XP_008780214	XP_007568428	XP_006015366	ENH78418	EDV55444	XP_007233832
AGW27098	XP_008816144	XP_007567566	XP_006012878	EMT20948	EDV27760	XP_007268255
AGW27097	XP_008814097	XP_007555814	XP_006011223	EMT03161	EDV26870	EXX69246
NP_001004819	XP_008789198	XP_007551384	XP_006007322	EMS63882	EDV22078	XP_007003211
ACS34693	XP_008782424	XP_007596620	XP_006004482	EMS63073	EDK39010	XP_006959286
ACS34688	CDQ04436	EZP29534	XP_006001530	EMS58637	EDQ71178	XP_006887086
P34708	CDQ01102	XP_007494393	XP_006001529	EMS51747	EDO48634	XP_006814074
Q17307	XP_008728416	XP_007482376	XP_005997793	EMR90667	EDN04625	XP_006882634
Q17308	XP_008721811	XP_006399577	XP_005994105	EMR90122	EAZ16547	CDJ81337
P34709	XP_008704650	XP_001362215	XP_005982188	EMR81711	CCD25602	EUC57683
CDS33498	XP_008702439	EZG78631	XP_005972202	EMR81248	AFI99937	XP_006676633
CDS39557	XP_008700964	EZG58343	AEB06874	XP_004390908	XP_003704798	XP_006636782
CDS41593	XP_008652333	EZG49312	XP_005951084	XP_004386526	CCF34158	XP_006636269
AFH07473	XP_008647455	EZG08523	XP_005945985	XP_004385946	CCF47647	ETN60556
AHW45716	XP_008680345	EZA58735	XP_005942595	XP_004261723	XP_003694586	ETE68860
AGH39950	XP_008670369	EZA54276	XP_005935911	XP_004321843	CBN82191	EGU11104
CAD25383	KFD66446	XP_007448570	XP_005912225	XP_004302118	CCA66394	XP_006160845
AET31470	KFD53349	XP_007448569	XP_005884713	XP_004287912	GAA52209	XP_006138419
AAM88675	XP_008633393	XP_007462031	XP_005880332	XP_004286001	EAN33238	CAP32108
AAM88674	XP_007777325	XP_007438784	XP_005869698	XP_004274474	AER37951	XP_005997166
ACS34692	KFA79026	XP_007430795	DAA40710	EMP38023	CCD74754	XP_005994333
AAPO6839	KFA69881	XP_007430738	ERT63860	EMP33303	CCE44605	XP_005918332
AAW34233	XP_008601966	CDO24833	XP_005823619	EMP26535	CCE42551	AGW99165
AAZ14853	XP_008598736	XP_007411678	XP_005800482	XP_004223843	XP_312950	CCF54532
WP_017453287	KEZ40368	XP_007410940	WP_022548272	EMF14776	XP_308255	XP_005679120
Q22036	XP_008546245	XP_007386899	XP_005754899	EMF13425	ACI49173	XP_005675227
XP_009264606	WP_031146098	EYD76148	XP_005748444	XP_004208367	EGX46773	ERL90597
ADM11622	WP_031129256	XP_007347162	XP_005747591	XP_004208028	CBQ72246	XP_005421840
AIK25407	WP_031093453	XP_007322521	XP_005730470	XP_002189088	CCD21763	ERG83502
CDS28764	XP_008521478	XP_007299753	XP_005728462	XP_002193464	CCA18460	ERE66072
CDS33499	WP_031083001	EYB94880	WP_022442987	XP_002193613	XP_003390187	EPQ02843
KEY75498	WP_031065743	XP_007259040	WP_022378222	XP_002154565	BAJ97762	CCW00012
EUB58298	WP_031011931	XP_007256393	WP_022180188	XP_004162101	BAJ96674	EPT03124
JAC29154	WP_030976832	XP_007245807	WP_021903444	XP_004156139	CCA30123	XP_005050939
EWY39439	WP_030904363	XP_007237738	ERL93400	CCH59523	XP_003210803	XP_005041191
XP_002314126	WP_030801670	XP_007236589	XP_005708204	ELW66550	XP_003207015	XP_005087178
WP_022981486	WP_030667771	XP_007234698	WP_021650409	ELW66202	XP_003205348	XP_005071760
CCX13347	WP_030615064	XP_007234696	XP_001500275	ELW65135	XP_003205269	XP_005089800
ERG80442	XP_008497739	XP_007292262	XP_001490939	ELV10201	ADI16916	GAC96919
WP_019572761	XP_008490837	XP_007291655	XP_005576322	ELU14016	CAK80294	XP_004994829
WP_018423186	KEQ85253	XP_007291433	WP_021030852	ELU00184	CAK57540	EPB85118
WP_017773461	KEQ83635	XP_007284083	XP_005508358	EJD76219	CBX45935	XP_004626622
WP_016919579	KEQ74438	XP_007270475	XP_005530051	EAT47304	XP_003022428	BAN20641
JAA89154	XP_007046264	EXX60199	XP_005430452	EAT46229	XP_002927049	XP_004555494
ENH62449	XP_007043371	EXX58902	XP_005423911	EJB11413	XP_002918652	XP_004547942
EMT63448	XP_008480358	EXV06503	XP_005421741	XP_003273107	XP_002793720	XP_004540367
JAA59270	XP_008475770	EXU94892	XP_005448987	ELR57527	EDL46908	XP_004461410
CCH41310	XP_008474851	XP_007179288	XP_005494223	ELQ60929	ABY51976	GAA97500
ADY44475	XP_008469137	XP_006966353	XP_005439402	XP_004059139	AAW27686	EMS19823
ADY40996	XP_008467824	XP_006073779	XP_005433095	XP_004043267	AAD51821	CCO26540
EED47961	CDP01712	XP_006041857	AGV05019	XP_004004588	AAK32904	XP_004360968
EER05937	CDP83580	XP_007141988	WP_019024054	XP_004003989	AAD32882	EMP29840
AET31468	XP_008463539	XP_007132039	WP_004948172	CCO75453	CAX73752	XP_002158647
AAD29784	XP_008400307	XP_007062021	WP_004848769	CCC08192	CAX72974	EMD40972
KEH29571	XP_008425772	GAD93323	ERE80288	CCC12515	ACO12866	ELT87496
ERE80299	XP_008406132	XP_007108217	EPS30139	CCC13118	XP_001019099	ACH81147
CCF75983	XP_008406131	XP_007105845	EPQ16396	XP_003972322	XP_002261613	EFO27260
CBY13204	XP_008403582	XP_007100916	EPQ03871	XP_003971411	ACO09209	EFO22008
EGC32884	KEI39923	EXM13212	WP_004985793	XP_003969911	ABK24504	EJD75619
AFU60542	KEI37990	GAC75786	EN001966	XP_003968221	BAH14190	EAT42019
Q8IN81	XP_008360079	GAC72597	YP_005266615	AFP24338	BAG93673	EAT38922
YP_004020013	XP_008385863	GAC72233	EAL30002	AFK10632	XP_725465	ELK00780
YP_002378176	XP_008332762	XP_007085844	EAL31605	EKD04357	XP_001613442	2KXN

YP_001521374	XP_008314452	XP_007007556	NP_001006379	EKC33746	ABR21771	2CQC
YP_192259	WP_029469977	XP_006958181	YP_006494332	EJY86355	XP_001244996	EKD03814
NP_296164	XP_008303790	EXB77634	XP_005341678	XP_003320855	ABL68115	EKC35695
XP_002275480	XP_008300432	EXA31095	XP_005360443	YP_003517744	XP_001556797	XP_003324121
XP_002271267	XP_008293400	XP_003988634	EQL01584	EJT41479	XP_001549058	AFQ62109
XP_002267339	XP_008291633	NP_001241952	EQB59502	XP_003877602	BAD87730	EJW76920
WP_025641044	XP_008290739	EAR91290	EQB48519	EGW12177	CAN80991	EJU06491
Q9Z340	KEH44244	EWS73617	XP_005240351	EGW12020	CAN83201	EJT46854
XP_004250062	KEH26020	EWS61856	EPY87160	EGW08609	CAN82029	EHB13067
XP_004242571	WP_029336405	CDH10474	EPY84665	EGW05291	CAC28640	BAF23620
XP_004237943	AET00576	XP_006918467	EPY18519	EHB17980	XP_761572	BAF11564
XP_003562648	KEF55406	XP_006861776	EPS97009	BAF07111	XP_761262	EGI70155
AGH60516	CDR97904	XP_006823737	EPS71895	EGI70471	ABD72238	EFN80772
CAA93454	XP_008250628	XP_006822941	EPS43940	EGI64188	CAE05326	CBY33705
AIDHK2	XP_008269626	XP_006889443	EPS42358	EFZ17147	CAC48394	CBY18192
QOCKU4	XP_008257581	XP_006888238	CCT70071	EFZ10065		EFH68654
Q54Y98	WP_028748950	XP_002739620	XP_005059873	EFQ92573	BAD24702	EGC37204
Q9W1R5	WP_026655384	XP_002739463	XP_005055958	EFN85452	BAD24699	EFJ02026
XP_002123239	WP_026672510	XP_002736762	XP_005049932	EFN69463	CAF94089	ADI86271
XP_002131898	WP_026555081	XP_006809009	XP_005042609	EFN69457	KHJ43737	EGT36193
XP_002130665	KDR21936	XP_006800350	XP_005042236	EJN40473	CEJ03402	EEN62196
CDX77479	KDR08758	XP_006799316	XP_005041577	XP_003862311	P19339	EED81127
CDX84298	KDQ17365	XP_006797673	XP_005076088	XP_003860192	CAB07322	EDR04908
CDY29619	XP_008243207	EWG10571	XP_005104432	XP_003858733	P93395	EDQ90533
XP_007931330	XP_008230126	EUT83569	XP_005100365	EJK56446	P49344	EDS43828
XP_007929324	XP_008218941	CDJ91704	XP_003074417	EGD93389	P49343	ACD13597
AEE84818	XP_008205663	AGF63234	WP_013868981	CBY07372	GAK64965	EES13772
AEE36468	XP_008211867	AFU19669	WP_020571925	CBY21964	XP_009860260	EEC13476
AEE28565	KDP43732	AFT84390	WP_020096859	CBY21310	CDR35515	EEB19170
Q5UP83	KDP39336	AGP62864	WP_020065470	CBY43665	XP_009543934	XP_003747013
Q291E4	KDP39161	EUA72278	WP_019182257	EFQ25894	AAW40908	XP_796831
P56698	KDP27299	ADE15838	WP_019058448	EFH42366	AAW40816	EIE88854
AGO10830	KDP23770	ACY98679	WP_018846632	CAZ79898	CEAF63868	EIE82806
AFR96591	XP_008198128	ABV38455	WP_018575134	XP_003797953	AFR92532	EFX90042
NP_956710	XP_008199782	ABC81246	WP_018511480	XP_003793682	ADV19564	EFP07598
NP_957491	XP_008197930	NP_989624	WP_016568230	EFJ01569	CDQ74157	EFP04657
XP_687895	CDQ69710	CCO54678	XP_005031146	EEQ330342	KGB75667	EFP02409
XP_006451127	CDQ70211	ETW48586	XP_005020703	EEC72030	KGB74473	EFA10734
XP_006297086	XP_008162541	ETW15352	XP_004962180	XP_003774430	KGB33272	EDW74745
XP_006296861	XP_008144264	XP_006667621	XP_005019762	EGC48069	XP_009268354	EDW60892
XP_006290570	XP_008188584	XP_006667006	XP_005009108	EGT58941	XP_009169687	EDW47879
KFZ67759	XP_008188583	XP_006661113	EFW46680	EGT56850	XP_009064723	EDW32060
KFZ17715	XP_008187602	XP_006658528	XP_003476129	EGT36714	KFO33187	EDW10078
KFZ08816	XP_008180168	XP_006646608	XP_003463067	EGT36467	XP_009030323	EDW00854
KFY98736	XP_008178521	XP_006640421	WP_016535526	CB123686	XP_009030322	EDW56014
KFY66164	XP_003241090	XP_006636623	XP_418290	EED84731	XP_009015905	EDV37594
KFY64107	XP_008178712	XP_006635319	XP_004939964	EED80694	KFM80326	EDV23184
KFY58656	XP_005300478	XP_006635023	XP_004853658	EEA21237	NP_001149397	EDO44802
KFY46525	XP_001946255	XP_006634485	XP_004907415	EDR06956	KFG25163	EDK98603
KFY10717	XP_003240356	XP_006627627	XP_004918041	EDO18557	AAZ10327	XP_003700631
CDP28697	XP_001949110	XP_006627015	XP_002935224	EY62225	CDS11072	AAH44695
KFW81056	KDN81168	XP_006457496	XP_004913737	EEU46061	CDP90807	CCA75325
KFW78767	CDM78478	ETS59693	XP_004913733	EEU44096	KFD71267	XP_308949
KFV98531	KDN74351	XP_006614741	XP_002936474	EEU44064	KFD54234	CBQ70737
CDW89961	KDN68066	XP_006610868	EFM49019	EEF33239	KFD47880	XP_003383668
CDW91213	WP_024468518	XP_006608480	EFK55494	EDS45104	KFB51562	XP_003376018
KFV58306	XP_003223661	XP_006564394	YP_001617913	EDS30059	XP_008548004	XP_003136815
KFQ00659	XP_008111535	XP_006564390	EPB90806	ABG22114	XP_008467503	BAE41904
KFP27583	XP_008109207	XP_006564385	EFS98390	EHA52035	EPZ32016	XP_002915754
XP_009051972	XP_008088564	XP_006562912	WP_009328338	EES19773	XP_008402178	XP_001839345
XP_009048189	XP_008086017	XP_006562907	EEI63664	EES17511	XP_008329504	ADE10071
XP_006531590	XP_003217630	XP_006562552	WP_008461006	EES13022	XP_008326981	AAW27493
XP_006531589	XP_008045529	XP_006571440	WP_008396052	EER98466	XP_008307903	AAZ14854
XP_006538151	XP_008031952	XP_006592826	WP_007338102	EEP81362	KDR74388	AAK96648
XP_005269287	KDE80068	XP_006592825	WP_007270879	EEP76266	KDR12437	AAD19279
XP_005252370	KDE76182	XP_006592400	WP_007259531	EEH56413	KDQ65030	CAX77531
XP_005252368	CCI43254	ETP65061	WP_006898884	EEB15471	KDQ28892	XP_002390474
NP_001178912	CCI10108	XP_003544346	WP_006898674	EAT89855	KDQ08825	BAF81027
XP_008967057	XP_007967627	ETO99583	WP_006565309	EAT77474	XP_008205153	BAG95938
XP_008979214	KDB27203	ETO62292	WP_006509663	EDO03426	XP_008197653	XP_001732525
XP_008941762	KDB21071	XP_002297938	EEG27206	EAY05515	CDQ76268	XP_001900358
AEO63718	XP_007900832	ETO34690	XP_004707171	EAW23557	CDQ81807	XP_777899
AEO63347	XP_007899390	ETO24050	XP_004635693	XP_003745262	XP_008178883	XP_778234
AAN35931	XP_007894449	ETO22837	XP_004652622	XP_003739094	XP_001944825	BAE87345
XP_008929262	XP_007892668	XP_006484451	WP_010691933	XP_003730814	KDN40836	XP_757472
XP_008932199	XP_007890453	XP_006484450	WP_010478557	XP_003724168	XP_008104608	CAA70700
KFM57637	XP_007851575	ACY16387	WP_009623982	XP_001033297	KDE04808	CAG10350
CDU19388	XP_007797685	ETN80007	WP_009605732	CBA09476	XP_007909596	XP_007879936
CDS50615	XP_007794000	ETN70462	WP_009348080	EED56489	XP_007902599	EIE91598
CDS45583	XP_007793835	XP_006414557	WP_007322241	EED53163	XP_007901379	WP_006952201
XP_008916645	XP_007803276	XP_006357912	WP_007241269	EIE91599	EUB61593	XP_006357911
XP_008888553	XP_007766839					

Annexe V : Numéros d'accèsion des gènes à domaine DM utilisés pour les analyses phylogénétiques.

Organisme	Nom du gène (définition selon NCBI)	Numéro d'accèsion
Branchiopoda		
<i>Daphnia magna</i>	doublesex-Mab related 93B	BAG12872
	doublesex-Mab related 99B	BAG12873
	doublesex-Mab related 11E	BAG12871
<i>Daphnia pulex</i>	DMRT-99B-like protein	EFX84867
	DM DNA-binding protein	EFX74782
	Doublesex and mab-3 related transcription factor 3	EFX89054
Decapoda		
<i>Eriocheir sinensis</i>	Dmrt-like protein	ADH15934
<i>Penaeus monodon</i>	testis-specific DMRT1	ACC94178
Diptera		
<i>Bactrocera oleae</i>	male-specific doublesex	CAD67987
<i>Anastrepha obliqua</i>	doublesexM	AAY25167
<i>Musca domestica</i>	DSXM	AAR23813
<i>Drosophila melanogaster</i>	doublesex, isoform A	NP_731197
	doublesex-Mab related 93B	NP_524428
	doublesex-Mab related 99B	NP_524549
	doublesex-Mab related 11E	NP_511146
<i>Aedes aegypti</i>	male specific 1 protein	ABD96573
	AAEL004696-PA	EAT43900
<i>Anopheles gambiae</i>	AGAP004050-PC	XP_560052
	AGAP000431-PA	XP_310668
	AGAP001388-PA	XP_321748
Lepidoptera		
<i>Bombyx mori</i>	doublesex isoform M	NP_001104815
Hymenoptera		
<i>Apis mellifera</i>	doublesex isoform M	NP_001104725
Coleoptera		
<i>Tribolium castaneum</i>	PREDICTED: doublesex- and mab-3-related transcription factor A2	XP_975675
Anura		
<i>Xenopus laevis</i>	doublesex and mab-3 related transcription factor 5	ABC55871
	Doublesex- and mab-3-related transcription factor 1A	Q3LH63
	DMRT4	AAV66322
Cypriniformes		
<i>Danio rerio</i>	doublesex and mab-3 related transcription factor 3	AAU89440
Beloniformes		
<i>Oryzias latipes</i>	DMY	BAB92012
	DMRT4	BAB63259
	DMRT2	AAL02163
	doublesex and mab-3 related transcription factor 5	BAD00703
Tetraodontiformes		
<i>Takifugu rubripes</i>	doublesex- and mab-3-related transcription factor A2	NP_001033039
	doublesex and mab-3 related transcription factor 4	NP_001033037
	doublesex and mab-3 related transcription factor 2	NP_001033035
	doublesex- and mab-3-related transcription factor 3	NP_001033034
	doublesex- and mab-3-related transcription factor 1	NP_001033038
Primates		
<i>Homo sapiens</i>	doublesex- and mab-3-related transcription factor A2	NP_115486
	doublesex- and mab-3-related transcription factor 2 isoform 1	NP_006548
	doublesex- and mab-3-related transcription factor 3	NP_067063
	doublesex and mab-3 related transcription factor 1 b	AAR89619
	doublesex- and mab-3-related transcription factor A1	NP_071443

Annexe VI : Numéros d'accension des gènes *Transformer-2* utilisés pour les analyses phylogénétiques.

Organisme	Nom du gène (définition selon NCBI)	Numéro d'accension
Diptera		
<i>Anastrepha aff. fraterculus</i>	transformer-2	FN658609
<i>Anastrepha amita</i>	transformer-2	FN658617
<i>Anastrepha bistrigata</i>	transformer-2	FN658616
<i>Anastrepha grandis</i>	transformer-2	FN658612
<i>Anastrepha obliqua</i>	transformer-2	FN658607
<i>Anastrepha serpentina</i>	transformer-2	FN658613
<i>Anastrepha sororcula</i>	transformer-2	FN658614
<i>Anastrepha striata</i>	transformer-2	FN658615
<i>Ceratitis capitata</i>	transformer-2	EU999754
<i>Bactrocera oleae</i>	transformer-2	AJ547623
<i>Musca domestica</i>	transformer-2	AY847518
<i>Lucilia cuprina</i>	transformer-2	FJ461620
<i>Drosophila melanogaster</i>	transformer-2	M23633
<i>Drosophila virilis</i>	transformer-2 variant D	XM_002049663
<i>Drosophila pseudoobscura pseudoobscura</i>	uncharacterized protein, variant A	XM_001360568
Lepidoptera		
<i>Bombyx mori</i>	transformer-2 variant A	NM_001126233
Hymenoptera		
<i>Apis mellifera</i>	transformer-2	NM_001265585
<i>Nasonia vitripennis</i>	transformer-2 protein homolog alpha isoform X3	XP_001601106
Decapoda		
<i>Penaeus monodon</i>	transformer 2	ACD13597
<i>Macrobrachium nipponense</i>	transformer 2	AGI50962
<i>Eriocheir sinensis</i>	transformer-2a	APJ36536
	transformer-2b	APJ36537
	transformer-2c	APJ36538
	transformer-2d	APJ36539
Copepoda		
<i>Lepeophtheirus salmonis</i>	transformer-2 beta	CDW20248
Branchiopoda		
<i>Daphnia magna</i>	transformer-2 protein-like	EFX90042

Annexe VII : Séquences et caractéristiques des amorces générales utilisées au cours de cette étude.

Gène	Nom des amorces	Séquence (5' vers 3')	Taille du produit de PCR (pb)	TM (°C)	Source
Isopodes terrestres					
18S (marqueur nucléaire)	18S-F	GCGTAATAAAAAGACCGATT	620	48	Cette étude
	18S-R	CTTTTGTAACACGAAGCC			
COI (marqueur mitochondrial)	LCO	GGTCAACAAATCATTAAAGATATTGG	700	52	Folmer et al., 1994
	HCO	TAAACTTCAGGGTGACCAAAAAATCA			
<i>Wolbachia</i>					
<i>wsp</i>	81F	TGGTCCAATAAGTGATGAAGAAAC	650	55	Zhou et al., 1998
	691R	AAAAATTAAACGCTACTCCA			
<i>ftsZ</i>	ZF1	GTTGTCGCAAATACCGATGC	1 000	47	Werren et al., 1995
	ZR1	CTTAAGTAAGCTGGTATATC			

Annexe VIII : Matériels supplémentaires issus de Becking et al., 2017 (Chapitre 4).

Diversity and evolution of sex determination systems in terrestrial isopods

Thomas Becking¹, Isabelle Giraud¹, Maryline Raimond¹, Bouziane Moumen¹, Christopher Chandler², Richard Cordaux^{1,#,*} and Clément Gilbert^{1,#,*}

¹Université de Poitiers, UMR CNRS 7267 Ecologie et Biologie des Interactions, Equipe Ecologie Evolution Symbiose, TSA 51106, 86073 Poitiers Cedex 9, France

²Department of Biological Sciences, SUNY Oswego, Oswego, New York 13126

These authors contributed equally to this work

* Corresponding authors: RC (richard.cordaux@univ-poitiers.fr) and CG (clement.gilbert@univ-poitiers.fr)

Supplementary Figure legends

Supplementary Figure S1. Cladogram showing the relationships among the 26 species used in this study. The selected model is GTR + I + G for each combination of markers. The values written next to each node correspond to the Bayesian posterior probabilities from BEAST analysis (without a constrained topology) and to the bootstrap score from the Maximum Likelihood analysis (200 bootstrap replicates) using all 88 markers (alignment length : 69,570 bp) / nuclear markers (81 markers, alignment length : 63,723 bp) / mitochondrial markers (7 markers, alignment length : 5,847 bp) / PCR markers (10 markers, alignment length : 10,773 bp), following this order.

Supplementary Figure S2 : Chronogram of terrestrial isopod crustaceans, obtained using BEAST and the tree shown in Supplementary Figure S1 as a backbone. Branch length is scaled to time, with nodal ages corresponding to median posterior estimates. Grey bars represent the Highest Posterior Density (HPD) intervals at 95%. Orange bars denote the calibration points.

Supplementary Figure S3. Cladogram indicating the silent substitution rate (dS) of each branch of isopods for mitochondrial sequences (red), nuclear sequences (blue) and the ratio between mitochondrial dS and nuclear dS (black). The per-site rate of amino acid replacements/ silent substitution rate ratio (dN/dS) of mitochondrial and nuclear dataset are estimated respectively at 0.2785 and 0.0674.

Supplementary Table legends

Supplementary Table S1. Information on transcriptomic data used.

Supplementary Table S2. Detailed information on the sex reversal experiment success.

Supplementary Table S3. Statistical analyses of sex ratio from the broods obtained in crosses between females and neomales. Red cells indicate females infected by *Wolbachia* (excluded results). Orange cells indicate that the progeny sex-ratio is not significantly different from 50/50 (excluded results). Green cells indicate that the progeny sex-ratio is significantly different from 50/50 (included results).

Supplementary Table S4. Summary of the statistics used to determine the heterogametic systems.

Supplementary Table S5. Ancestral states reconstruction of the estimated number of transitions between sex determination systems. The numbers written for the maximum likelihood analysis correspond to the mean number of transitions estimated with 1,000 stochastic mappings ($\pm$ standard deviation).

Supplementary Table S6. Metrics of the 19 isopod transcriptomes assembled in this study.

Supplementary Table S7. Detailed annotation of the 88 markers used in this study (5 first best BLASTX versus NCBI non-redundant database hit are shown).

Supplementary Table S8. Sequences, PCR product size, and melting temperature of the primers designed to complete the phylogeny.

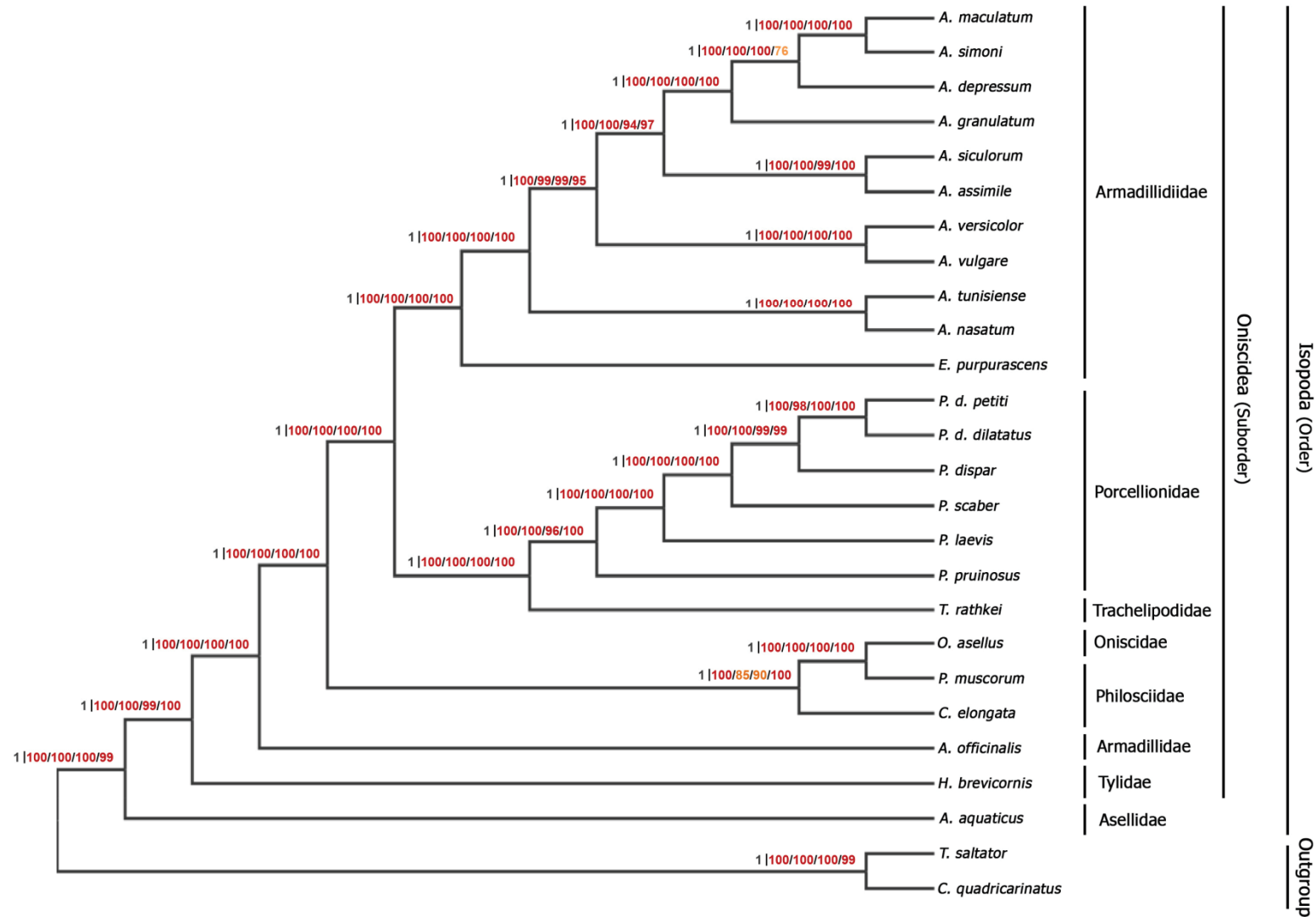
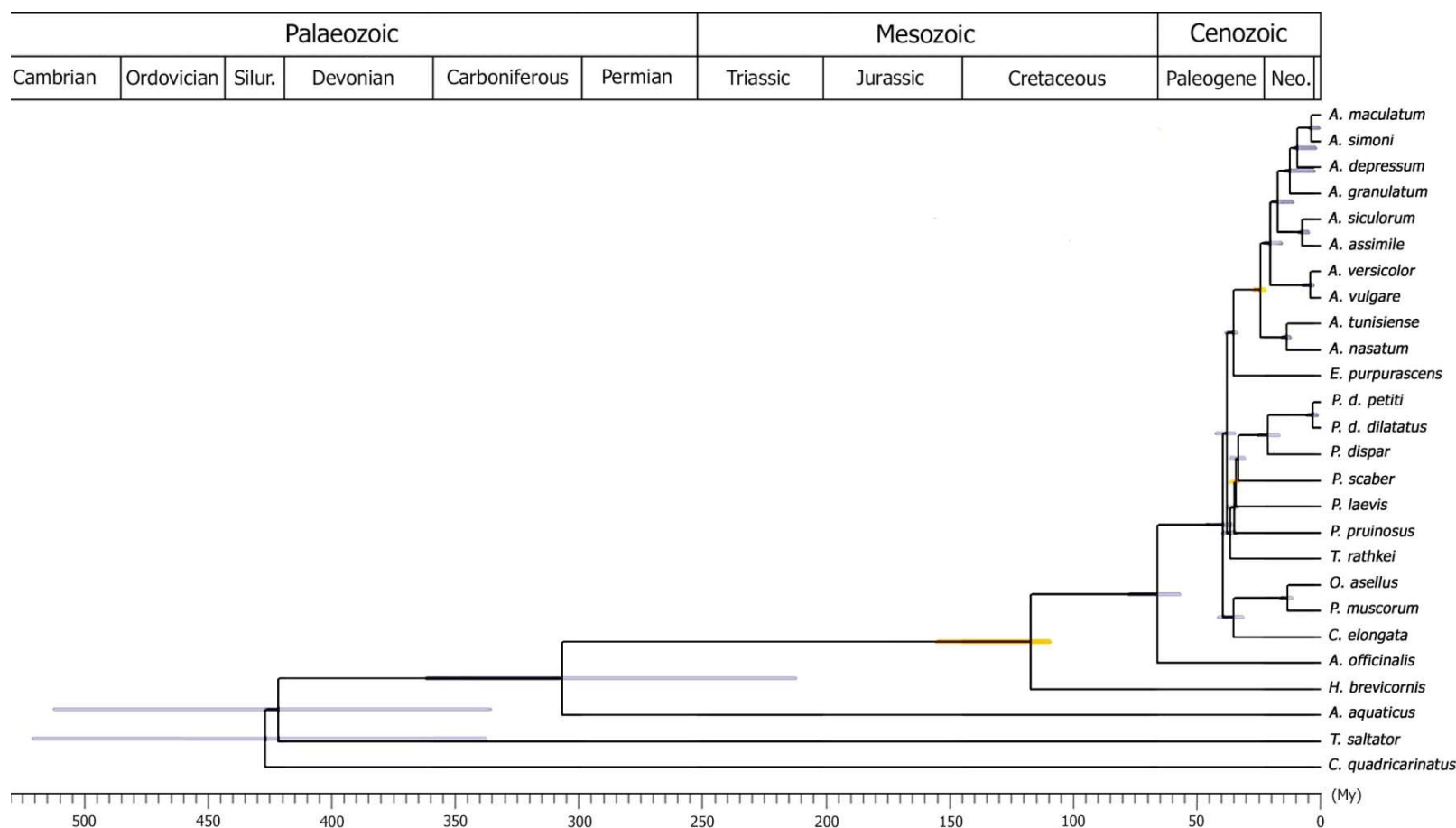


Figure S1. Cladogram showing the relationships among the 26 species used in this study. The selected model is GTR + I + G for each combination of markers. The values written next to each node correspond to the Bayesian posterior probabilities from BEAST analysis (without a constrained topology) and to the bootstrap score from the Maximum Likelihood analysis (200 bootstrap replicates) using all 88 markers (alignment length : 69,570 bp) / nuclear markers (81 markers, alignment length : 63,723 bp) / mitochondrial markers (7 markers, alignment length : 5,847 bp) / PCR markers (10 markers, alignment length : 10,773 bp), following this order.



Supplementary Figure S2 : Chronogram of terrestrial isopod crustaceans, obtained using BEAST and the tree shown in Supplementary Figure S1 as a backbone. Branch length is scaled to time, with nodal ages corresponding to median posterior estimates. Grey bars represent the Highest Posterior Density (HPD) intervals at 95%. Orange bars denote the calibration points.

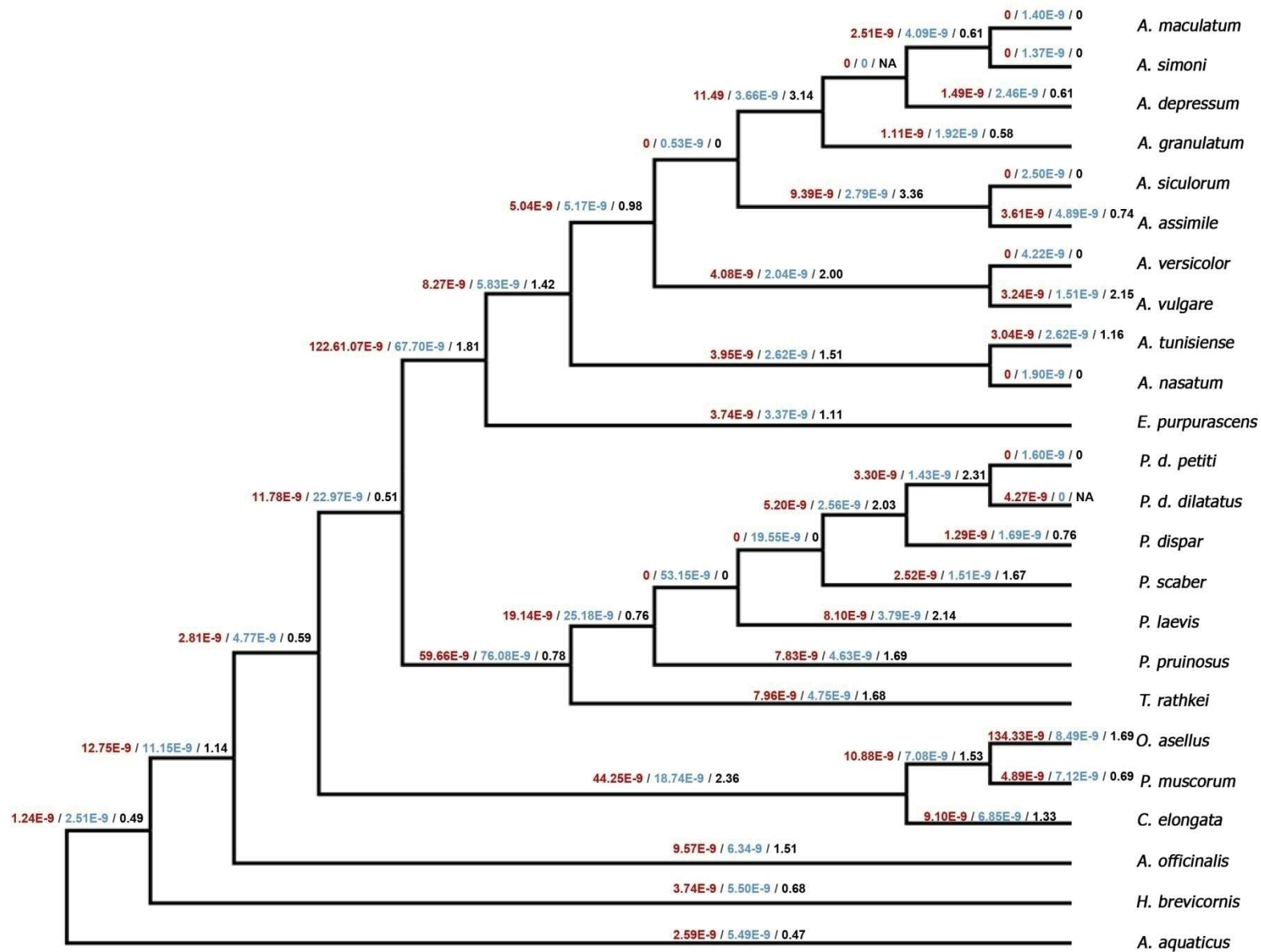


Figure S3. Cladogram indicating the silent substitution rate (dS) of each branch of isopods for mitochondrial sequences (red), nuclear sequences (blue) and the ratio between mitochondrial dS and nuclear dS (black). The per-site rate of amino acid replacements/ silent substitution rate ratio (dN/dS) of mitochondrial and nuclear dataset are estimated respectively at 0.2785 and 0.0674.

Supplementary Table S1. Information on transcriptomic data used.

Genus	Species	Origin (country)	Gender	Number of individuals used	Total number of raw reads	Accession number (source)
<i>Armadillidium</i>	<i>maculatum</i>	<i>Sainte Marguerite (France)</i>	♀	1	12,766,990	SRX2600479 (This study)
<i>Armadillidium</i>	<i>simoni</i>	<i>Villeneuve-loubet (France)</i>	♀	1	17,933,434	SRX2600482 (This study)
<i>Armadillidium</i>	<i>depressum</i>	<i>Saint Guilherm le Désert (France)</i>	♀	1	17,940,202	SRX2600477 (This study)
<i>Armadillidium</i>	<i>granulatum</i>	<i>Crete (Greece)</i>	♀	1	13,084,538	SRX2600478 (This study)
<i>Armadillidium</i>	<i>assimile</i>	<i>Saint-Maixent (France)</i>	♂	1	16,563,498	SRX2600476 (This study)
<i>Armadillidium</i>	<i>siculorum</i>	<i>Valle dei Templi (Sicily, Italy)</i>	♂	1	13,304,442	SRX26004781 (This study)
<i>Armadillidium</i>	<i>versicolor</i>	<i>Saint-Veit (Carynthia, Austria)</i>	♀	1	17,200,162	SRX2600484 (This study)
<i>Armadillidium</i>	<i>vulgare</i>	?	?	?	33,692,087	SRX564995 (Romiguier <i>et al.</i> , 2014)
<i>Armadillidium</i>	<i>nasatum</i>	?	?	?	15,552,480	SRX564993 and SRX564994 (Romiguier <i>et al.</i> , 2014)
<i>Armadillidium</i>	<i>tunisiense</i>	<i>Khmiss Monastir (Tunisia)</i>	♀	1	24,948,176	SRX2600483 (This study)
<i>Eluma</i>	<i>purpurascens (caelatum)</i>	<i>Chizé (France)</i>	♀	1	14,402,732	SRX2600486 (This study)
<i>Porcellionides</i>	<i>pruinosis</i>	<i>Guadalquivir (Spain)</i>	♂	1	15,362,180	SRX2600492 (This study)
<i>Porcellio</i>	<i>laevis</i>	<i>Fréjus (France)</i>	♀	1	13,244,040	SRX2600491 (This study)
<i>Porcellio</i>	<i>scaber</i>	<i>Bois de la Borne (France)</i>	♂	1	16,303,962	SRX2600493 (This study)
<i>Porcellio</i>	<i>dispar</i>	<i>Santa Maria del Sol (Brasil)</i>	♀	1	15,455,100	SRX2600490 (This study)
<i>Trachelipus</i>	<i>rathkei</i>	<i>New York State (USA)</i>	1 ♀ and 1 ♂	2	237,910,854	SRR5198727 and SRR5198726 (This study)
<i>Chaetophiloscia</i>	<i>elongata</i>	<i>Turée (France)</i>	♂	2	16,269,144	SRX2600485 (This study)
<i>Philoscia</i>	<i>muscorum</i>	<i>Poitiers (France)</i>	♀	1	16,763,368	SRX2600489 (This study)
<i>Oniscus</i>	<i>asellus</i>	<i>Angers (France)</i>	♀	1	16,889,482	SRX2600488 (This study)
<i>Armadillo</i>	<i>officinalis</i>	<i>Crete (Greece)</i>	♀	1	15,903,154	SRX2600480 (This study)
<i>Helleria</i>	<i>brevicornis</i>	<i>Fréjus (France)</i>	♀	1	17,082,882	SRX2600486 (This study)
<i>Asellus</i>	<i>aquaticus</i>	<i>Planina cave (Slovenia)</i>	?	?	637,023	SRX1097282, SRX1097283, SRX1097286 and SRX1097290 (Speiser <i>et al.</i> 2014)
<i>Talitrus</i>	<i>saltator</i>	<i>Ynyslas beach (Wales, United Kingdom)</i>	♀ and ♂	6	49,459,410	SRX1896418 and SRX1896419 (O'Grady <i>et al.</i> 2016)
<i>Cherax</i>	<i>quadricarinatus</i>	<i>Darwin (Australia)</i>	?	1	32,344,260	ERX363978 (Tan <i>et al.</i> 2016)

Supplementary Table S2. Detailed information on sex reversal experiment success.

Species	number of injected females	number of injected individuals developing as females		number of dead individuals		number of putative neomales		number of attempted crosses	number of progenies
		n	%	n	%	n	%		
<i>Armadillidium maculatum</i>	12	2	16.67	4	33.33	6	50.00	18	8
<i>Armadillidium depressum</i>	7	1	14.29	2	28.57	4	57.14	12	7
<i>Armadillidium granulatum</i>	10	2	20.00	5	50.00	3	30.00	9	3
<i>Armadillidium assimile</i>	8	0	0.00	2	25.00	6	75.00	12	6
<i>Armadillidium siculorum</i>	4	2	50.00	0	0.00	2	50.00	4	1
<i>Armadillidium versicolor</i>	8	3	37.50	1	12.50	4	50.00	16	2
<i>Armadillidium vulgare</i>	10	2	20.00	4	40.00	4	40.00	8	1
<i>Armadillidium nasatum</i>	10	2	20.00	4	40.00	4	40.00	12	8
<i>Armadillidium tunisiense</i>	6	4	66.67	1	16.67	1	16.67	2	0
<i>Porcellionides pruinosus</i>	7	1	14.29	4	57.14	2	28.57	4	0
<i>Porcellio scaber</i>	9	3	33.33	4	44.44	2	22.22	4	2
<i>Porcellio dispar</i>	11	2	18.18	6	54.55	3	27.27	6	2
<i>Chaetophiloscia elongata</i>	7	0	0.00	7	100.00	0	0.00	0	0
<i>Philoscia muscorum</i>	8	0	0.00	8	100.00	0	0.00	0	0
<i>Armadillo officinalis</i>	18	1	5.56	8	44.44	9	50.00	18	3
TOTAL :	135	25		60		50		125	43

Supplementary Table S3. Statistical analyses of sex ratio from the broods obtained in crosses between females and neomales. Red cells indicate females infected by *Wolbachia* (excluded results). Orange cells indicate that the progeny sex-ratio is not significantly different from 50/50 (excluded results). Green cells indicate that the progeny sex-ratio is significantly different from 50/50 (included results).

Species		Male identification number	Progenies					Wolbachia ?	χ^2 Tests								Reference
			number of males		number of females		Total		Sex-ratio $\neq$ 50/50		Sex-ratio $\neq$ 66/33		Sex-ratio $\neq$ 75/25		Sex-ratio $\neq$ 100/0		
			n	%	n	%			X- squared	p-values	X- squared	p-values	X- squared	p- values	X- squared	p- values	
<i>Armadillidium</i>	<i>maculatum</i>	1	12	20.00	48	80.00	60	-	21.6	3.36E-06	4.8	0.02846	0.8	0.3711	∞	0	This study
<i>Armadillidium</i>	<i>maculatum</i>	2	10	55.56	8	44.44	18	-	0.2222	0.6374							This study
<i>Armadillidium</i>	<i>maculatum</i>	3	16	40.00	24	60.00	40	-	1.6	0.2059							This study
<i>Armadillidium</i>	<i>maculatum</i>	4	8	28.57	20	71.43	28	-	5.1429	0.02334	0.2857	0.593	0.1905	0.6625	∞	0	This study
<i>Armadillidium</i>	<i>maculatum</i>	4	9	21.43	33	78.57	42	-	13.7143	0.0002128	2.6786	0.1017	0.2857	0.593	∞	0	This study
<i>Armadillidium</i>	<i>maculatum</i>	4	7	20.00	28	80.00	35	-	12.6	3.86E-04	2.8	0.09426	0.4667	0.4945	∞	0	This study
<i>Armadillidium</i>	<i>maculatum</i>	4	5	31.25	11	68.75	16	-	2.25	0.1336							This study
<i>Armadillidium</i>	<i>maculatum</i>	4	4	22.22	14	77.78	18	-	5.5556	0.01842	1	0.3173	0.0741	0.7855	∞	0	This study
<i>Armadillidium</i>	<i>depressum</i>	1	11	20.75	42	79.25	53	-	18.1321	2.06E-05	3.7736	0.05207	0.5094	0.4754	∞	0	This study
<i>Armadillidium</i>	<i>depressum</i>	2	7	17.95	32	82.05	39	-	16.0256	6.25E-05	4.1538	0.04154	1.0342	0.3092	∞	0	This study
<i>Armadillidium</i>	<i>depressum</i>	2	3	21.43	11	78.57	14	+									This study
<i>Armadillidium</i>	<i>depressum</i>	3	13	22.03	46	77.97	59	+									This study
<i>Armadillidium</i>	<i>depressum</i>	3	7	16.67	35	83.33	42	-	18.6667	1.56E-05	5.25	0.02195	1.5556	0.2123	∞	0	This study
<i>Armadillidium</i>	<i>depressum</i>	4	7	31.82	15	68.18	22	-	2.9091	8.81E-02							This study
<i>Armadillidium</i>	<i>depressum</i>	4	7	41.18	10	58.82	17	+									This study
<i>Armadillidium</i>	<i>granulatum</i>	1	12	27.91	31	72.09	43	-	8.3953	0.003762	0.5698	0.4504	0.1938	0.6598	∞	0	This study
<i>Armadillidium</i>	<i>granulatum</i>	1	9	19.57	37	80.43	46	-	17.0435	3.65E-05	3.9239	0.0476	0.7246	0.3946	∞	0	This study
<i>Armadillidium</i>	<i>granulatum</i>	1	8	24.24	25	75.76	33	-	8.7576	0.003083	1.2273	0.2679	0.0101	0.9199	∞	0	This study

<i>Armadillidium</i>	<i>assimile</i>	1	25	100.0 0	0	0.00	25	-	25	5.73E-07	50	1.54E-12	75	< 2.2e-16	∞	0	This study
<i>Armadillidium</i>	<i>assimile</i>	1	6	100.0 0	0	0.00	6	-	6	0.01431	12	0.000532	18	2.21E-05	∞	0	This study
<i>Armadillidium</i>	<i>assimile</i>	2	20	100.0 0	0	0.00	20	-	20	7.74E-06	40	2.54E-10	60	9.49E-15	∞	0	This study
<i>Armadillidium</i>	<i>assimile</i>	3	60	100.0 0	0	0.00	60	-	60	9.49E-15	120	< 2.2e-16	180	< 2.2e-16	∞	0	This study
<i>Armadillidium</i>	<i>assimile</i>	3	14	100.0 0	0	0.00	14	-	14	0.0001828	28	1.21E-07	42	9.13E-11	∞	0	This study
<i>Armadillidium</i>	<i>assimile</i>	4	39	100.0 0	0	0.00	39	-	39	4.24E-10	78	< 2.2e-16	117	< 2.2e-16	∞	0	This study
<i>Armadillidium</i>	<i>siculorum</i>	1	12	40.00	18	60.00	30	-	1.2	0.2733							This study
<i>Armadillidium</i>	<i>versicolor</i>	1	7	58.33	5	41.67	12	-	0.3333	0.5637							This study
<i>Armadillidium</i>	<i>versicolor</i>	2	16	57.14	12	42.86	28	-	0.5714	0.4497							This study
<i>Armadillidium</i>	<i>vulgare</i>	1	3	25.00	9	75.00	12	-	3	0.08326							This study
<i>Armadillidium</i>	<i>vulgare</i>	1	64	28.57	160	71.43	224	NA	41.1429	1.42E-10	2.2857	0.1306	1.5238	0.217	∞	0	Juchault and Legrand, 1972
<i>Armadillidium</i>	<i>vulgare</i>	2	16	22.54	55	77.46	71	NA	21.4225	3.68E-06	3.7254	0.05359	0.23	0.6315	∞	0	Juchault and Legrand, 1972
<i>Armadillidium</i>	<i>vulgare</i>	3	39	23.35	128	76.65	167	NA	47.4311	5.70E-12	7.485	0.006221	0.2415	0.6231	∞	0	Juchault and Legrand, 1972
<i>Armadillidium</i>	<i>vulgare</i>	4	57	27.01	154	72.99	211	NA	44.5924	2.43E-11	3.7915	0.05151	0.4566	0.4992	∞	0	Juchault and Legrand, 1972
<i>Armadillidium</i>	<i>nasatum</i>	1	0	0.00	12	100.0 0	12	-	12	0.000532	6	0.01431	4	0.0455	NA	NA	This study
<i>Armadillidium</i>	<i>nasatum</i>	1	0	0.00	11	100.0 0	11	-	11	0.0009111	5.5	0.01902	3.6667	0.05551	NA	NA	This study
<i>Armadillidium</i>	<i>nasatum</i>	2	0	0.00	29	100.0 0	29	-	29	7.24E-08	14.5	0.0001402	9.6667	0.001876	NA	NA	This study
<i>Armadillidium</i>	<i>nasatum</i>	2	0	0.00	8	100.0 0	8	-	8	0.004678	4	0.0455	2.6667	0.1025	NA	NA	This study
<i>Armadillidium</i>	<i>nasatum</i>	3	7	58.33	5	41.67	12	-	0.3333	0.5637							This study
<i>Armadillidium</i>	<i>nasatum</i>	3	14	63.64	8	36.36	22	-	1.6364	0.2008							This study
<i>Armadillidium</i>	<i>nasatum</i>	4	0	0.00	29	100.0 0	29	-	29	7.24E-08	14.5	1.40E-04	9.6667	0.001876	NA	NA	This study
<i>Armadillidium</i>	<i>nasatum</i>	4	0	0.00	20	100.0 0	20	-	20	7.74E-06	10	0.001565	6.6667	0.009823	NA	NA	This study
<i>Eluma</i>	<i>purpurascens (caelatum)</i>	1	7	21.88	25	78.13	32	-	10.125	0.001463	1.8906	0.1691	0.1667	0.6831	∞	0	Juchault and Rigaud 1995
<i>Eluma</i>	<i>purpurascens (caelatum)</i>	2	2	16.67	10	83.33	12	-	5.3333	0.02092	1.5	0.2207	0.4444	0.505	∞	0	Juchault and Rigaud 1995
<i>Eluma</i>	<i>purpurascens (caelatum)</i>	3	7	28.00	18	72.00	25	-	4.84	0.02781	0.32	0.5716	0.12	0.729	∞	0	Juchault and Rigaud 1995

<i>Eluma purpurascens (caelatum)</i>	4	11	22.92	37	77.08	48	-	14.0833	0.0001749	2.3438	0.1258	0.1111	0.7389	∞	0	Juchault and Rigaud 1995
<i>Porcellio scaber</i>	1	9	30.00	21	70.00	30	-	4.8	0.02846	0.15	0.6985	0.4	0.5271	∞	0	This study
<i>Porcellio scaber</i>	2	23	44.23	29	55.77	52	-	0.6923	0.4054							This study
<i>Porcellio dispar</i>	1	20	64.52	11	35.48	31	+									This study
<i>Porcellio dispar</i>	1	7	38.89	11	61.11	18	+									This study
<i>Porcellio dilatatus petiti</i>	NA	185 (24)	23.30	609 (24)	76.70	794	NA	226.4181	< 2.2e-16	35.9704	2.00E-09	1.2242	0.2685	∞	0	Legrand <i>et al.</i> , 1974
<i>Porcellio dilatatus dilatatus</i>	NA	0 (15)	0.00	207 (15)	100.00	207 (15)	NA	207	< 2.2e-16	103.5	< 2.2e-16	69	< 2.2e-16	NA	NA	Juchault and Legrand, 1964
<i>Oniscus asellus</i>	1	5	21.74	18	78.26	23	-	7.3478	0.006714	1.3913	0.2382	0.1304	0.718	∞	0	Juchault and Rigaud 1995
<i>Oniscus asellus</i>	2	3	16.67	15	83.33	18	-	8	0.004678	2.25	0.1336	0.6667	0.4142	∞	0	Juchault and Rigaud 1995
<i>Oniscus asellus</i>	3	6	23.08	20	76.92	26	-	7.5385	0.00604	1.2308	0.2673	0.0513	0.8208	∞	0	Juchault and Rigaud 1995
<i>Oniscus asellus</i>	4	5	18.52	22	81.48	27	-	10.7037	0.001069	2.6667	0.1025	0.6049	0.4367	∞	0	Juchault and Rigaud 1995
<i>Oniscus asellus</i>	5	2	13.33	13	86.67	15	-	8.0667	0.004509	2.7	0.1003	1.0889	0.2967	∞	0	Juchault and Rigaud 1995
<i>Oniscus asellus</i>	6	5	20.83	19	79.17	24	-	8.1667	0.004267	1.6875	0.1939	0.2222	0.6374	∞	0	Juchault and Rigaud 1995
<i>Oniscus asellus</i>	7	7	25.00	21	75.00	28	-	9.8462	0.001702	2.3269	0.1272	0.4615	0.4969	∞	0	Juchault and Rigaud 1995
<i>Armadillo officinalis</i>	1	0	0.00	17	100.00	17	-	17	3.74E-05	8.5	0.003551	5.6667	0.01729	NA	NA	This study
<i>Armadillo officinalis</i>	2	0	0.00	18	100.00	18	-	18	2.21E-05	9	0.0027	6	0.01431	NA	NA	This study
<i>Armadillo officinalis</i>	3	0	0.00	12	100.00	12	-	12	0.000532	6	0.01431	4	0.0455	NA	NA	This study
<i>Helleria brevicornis</i>	NA	0 (2)	0.00	32 (2)	100.00	32 (2)	NA	32	1.54E-08	16	6.33E-05	10.6667	0.001091	NA	NA	Juchault and Legrand, 1964

Supplementary Table S4. Summary of the statistics used to determine heterogametic systems.

Species		Number of males (number of progenies)	Number of females (number of progenies)	Sex ratio (% of females / % of males)	χ^2 tests								Interpretation	Reference
					Sex-ratio $\neq$ 50/50		Sex-ratio $\neq$ 66/33		Sex-ratio $\neq$ 75/25		Sex-ratio $\neq$ 100/0			
					X-squared	p-values	X-squared	p-values	X-squared	p-values	X-squared	p-values		
<i>Armadillidium</i>	<i>maculatum</i>	40 (5)	143 (5)	78.1 / 21.9	57.9727	2.658E-14	10.8443	0.000991	0.9636	0.3263	∞	0	ZW/ZZ with viable WW genotype	This study
<i>Armadillidium</i>	<i>depressum</i>	25 (3)	109 (3)	81.3 / 18.7	52.6567	3.973E-13	12.9888	0.0003134	2.8756	0.08993	∞	0	ZW/ZZ with viable WW genotype	This study
<i>Armadillidium</i>	<i>granulatum</i>	29 (3)	93 (3)	76.2 / 23.8	33.5738	6.861E-09	5.0205	0.02505	0.0984	0.7538	∞	0	ZW/ZZ with viable WW genotype	This study
<i>Armadillidium</i>	<i>assimile</i>	164 (6)	0 (6)	0 / 100	164	< 2.2e-16	328	< 2.2e-16	492	< 2.2e-16	∞	0	?	This study
<i>Armadillidium</i>	<i>vulgare</i>	176 (4)	497 (4)	72.6 / 27.4	153.107	< 2.2e-16	15.6204	7.74E-05	0.476	0.4902	∞	0	ZW/ZZ with viable WW genotype	Juchault and Legrand, 1972
<i>Armadillidium</i>	<i>nasatum</i>	0 (6)	109 (6)	100 / 0	109	< 2.2e-16	54.5	1.554E-13	36.3333	1.663E-09	NA	NA	XY/XX	This study
<i>Eluma</i>	<i>purpurascens (caelatum)</i>	27 (4)	90 (4)	76.9 / 23.1	33.9231	5.73E-09	5.5385	0.0186	0.2308	0.631	∞	0	ZW/ZZ with viable WW genotype	Juchault and Rigaud, 1995
<i>Porcellio</i>	<i>scaber</i>	9 (1)	21 (1)	70 / 30	4.8	0.02846	0.15	0.6985	0.4	0.5271	∞	0	ZW/ZZ	This study
<i>Porcellio</i>	<i>dilatatus petiti</i>	185 (24)	609 (24)	76.7 / 23.3	226.4181	< 2.2e-16	35.9704	2.00E-09	1.2242	0.2685	∞	0	ZW/ZZ with viable WW genotype	Legrand <i>et al</i> , 1974
<i>Porcellio</i>	<i>dilatatus</i>	0 (15)	207 (15)	100 / 0	207	< 2.2e-16	103.5	< 2.2e-16	69	< 2.2e-16	NA	NA	XY/XX	Juchault and Legrand, 1964
<i>Armadillo</i>	<i>officinalis</i>	0 (3)	47 (3)	100 / 0	47	7.10E-12	23.5	1.25E-06	15.6667	7.55E-05	NA	NA	XY/XX	This study
<i>Oniscus</i>	<i>asellus</i>	33 (7)	128 (7)	79.5 / 20.5	56.0559	7.04E-14	11.9379	0.00055	1.7412	0.187	∞	0	ZW/ZZ with viable WW genotype	Juchault and Rigaud, 1995
<i>Helleria</i>	<i>brevicornis</i>	0 (2)	32 (2)	100 / 0	32	1.54E-08	16	6.33E-05	10.6667	0.001091	NA	NA	XY/XX	Juchault and Legrand, 1964

Supplementary Table S5. Ancestral states reconstruction of the estimated number of transitions between sex determination systems. The numbers written for the maximum likelihood analysis correspond to the mean number of transitions estimated with 1,000 stochastic mappings ($\pm$ standard deviation).

Types of discrete states used	Transition directions	Parsimony analysis	Maximum likelihood analysis	
			"ER" model	"SYM" model
<i>A. assimile</i> system treated as independent sex determination mechanism (1)	XY/XX to ZZ/ZW	1	1.66 ($\pm$ 1.31)	3.54 ($\pm$ 1.85)
	XY/XX to undetermined	0	1.04 ($\pm$ 0.98)	0
	ZZ/ZW to XY/XX	2	4.79 ($\pm$ 1.51)	6.86 ($\pm$ 1.83)
	ZZ/ZW to undetermined	1	2.81 ($\pm$ 1.28)	1.31 ($\pm$ 0.54)
	undetermined to XY/XX	0	1.87 ($\pm$ 1.23)	0
	undetermined to ZZ/ZW	0	1.36 ($\pm$ 1.19)	0.13 ($\pm$ 0.44)
	Total	4	13.53 ($\pm$ 3.26)	11.84 ($\pm$ 3.11)
<i>A. assimile</i> system treated as unknown mechanism (2)	XY/XX to ZZ/ZW	1	3.67 ($\pm$ 1.94)	3.71 ($\pm$ 1.91)
	ZZ/ZW to XY/XX	2	7.51 ($\pm$ 1.83)	7.61 ($\pm$ 1.84)
	Total	3	11.18 ($\pm$ 3.08)	11.32 ($\pm$ 3.07)

Supplementary Table S6. Metrics of the 19 isopod transcriptomes assembled in this study.

Species	Number of transcripts	Total size of assembly (nuc)	Mean transcript size (nuc)	Median transcript size (nuc)	N50 transcript length (nuc)
<i>Armadillidium maculatum</i>	49,506	31,897,914	644	380	928
<i>Armadillidium simoni</i>	53,420	37,351,839	699	381	1,112
<i>Armadillidium depressum</i>	57,394	42,641,755	743	381	1,283
<i>Armadillidium granulatum</i>	60,651	39,394,793	650	358	1,011
<i>Armadillidium siculorum</i>	60,934	37,141,038	610	346	884
<i>Armadillidium assimile</i>	56,481	37,002,359	655	378	995
<i>Armadillidium versicolor</i>	54,830	37,553,127	685	379	1,075
<i>Armadillidium vulgare</i>	112,787	50,040,006	444	234	755
<i>Armadillidium tunisiense</i>	77,764	52,869,378	680	343	1,179
<i>Armadillidium nasatum</i>	48,570	38,673,525	796	423	1,347
<i>Eluma purpurascens (caelatum)</i>	69,858	45,004,684	644	343	1,027
<i>Porcellio dispar</i>	65,731	41,771,358	635	352	961
<i>Porcellio scaber</i>	54,400	33,842,101	622	372	881
<i>Porcellio laevis</i>	48,428	29,369,984	606	376	821
<i>Porcellionides pruinosus</i>	67,260	46,231,902	687	369	1,114
<i>Trachelipus rathkei</i>	114,623	73,831,912	644	353	1,021
<i>Oniscus asellus</i>	61,150	39,337,685	643	367	953
<i>Philoscia muscorum</i>	64,025	44,470,959	695	379	1,115
<i>Chaetophiloscia elongata</i>	69,223	43,969,065	635	356	952
<i>Armadillo officinalis</i>	55,357	32,716,309	591	359	794
<i>Helleria brevicornis</i>	37,264	18,531,321	497	328	575
<i>Asellus Aquaticus</i>	21,684	11,582,160	534	464	506
<i>Talitrus saltator</i>	156,706	151,674,147	968	513	1,533
<i>Cherax quadricarinatus</i>	189,772	132,835,005	700	368	1,154

Supplementary Table S7. Detailed annotation of the 88 markers used in this study (5 first best BLASTX versus NCBI non-redundant database hit are shown).

Annotation					
Best Blastp hits		Max score	Total score	Query cover	E value Ident
Ortho 1	Methyl-CpG-binding domain protein 2 [Zootermopsis nevadensis]	372	372	97%	1.00E-126 69%
	methyl-CpG binding transcription regulator, putative [Ixodes scapularis]	349	349	95%	2.00E-117 64%
	PREDICTED: LOW QUALITY PROTEIN: methyl-CpG-binding domain protein 2-like [Limulus polyphemus]	344	344	95%	7.00E-116 65%
	PREDICTED: methyl-CpG-binding domain protein 2-like [Limulus polyphemus]	342	342	95%	8.00E-115 64%
	Methyl-CpG-binding domain protein 2 [Stegodyphus mimosarum]	333	333	94%	3.00E-111 62%
Ortho 2	innexin 1 [Homarus americanus]	562	562	99%	0 68%
	innexin 1 [Cancer borealis]	536	536	94%	0 67%
	innexin 1 [Schistocerca gregaria]	417	417	92%	1.00E-140 53%
	PREDICTED: innexin inx1 [Cimex lectularius]	417	417	92%	1.00E-140 52%
	RecName: Full=Innexin inx1; Short=Innexin-1; AltName: Full=G-Inx1 [Schistocerca americana]	416	416	92%	4.00E-140 53%
Ortho 3	PREDICTED: proteasome subunit alpha type-4 [Fopius arisanus]	424	424	95%	4.00E-147 78%
	PREDICTED: proteasome subunit alpha type-4 [Athalia rosae]	423	423	95%	1.00E-146 78%
	PREDICTED: proteasome subunit alpha type-4 [Cerapachys biroii]	422	422	95%	4.00E-146 78%
	PREDICTED: proteasome subunit alpha type-4 [Linepithema humile]	422	422	95%	4.00E-146 78%
	PREDICTED: proteasome subunit alpha type-4 [Dinoponera quadriceps]	421	421	95%	1.00E-145 78%
Ortho 4	PREDICTED: uncharacterized protein LOC106686239 [Halyomorpha halys]	99.4	99.4	78%	2.00E-20 29%
	conserved hypothetical protein [Riptortus pedestris]	94.4	94.4	72%	9.00E-19 29%
	PREDICTED: transmembrane protein 53 [Strongylocentrotus purpuratus]	93.2	93.2	86%	2.00E-18 26%
	Transmembrane protein 53 [Stegodyphus mimosarum]	90.9	90.9	85%	1.00E-17 28%
	hypothetical protein YQE_00956 [Dendroctonus ponderosae]	90.5	90.5	67%	2.00E-17 30%
Ortho 5	eukaryotic translation initiation factor 5A [Scylla paramamosain]	274	274	99%	5.00E-91 82%
	eukaryotic translation initiation factor 5 alpha [Procambarus clarkii]	270	270	99%	1.00E-89 80%
	eukaryotic translation initiation factor 5A [Penaeus monodon]	268	268	99%	1.00E-88 80%
	eukaryotic translation initiation factor 5A [Litopenaeus vannamei]	267	267	99%	2.00E-88 80%
	PREDICTED: eukaryotic translation initiation factor 5A [Stomoxys calcitrans]	242	242	98%	1.00E-78 71%
Ortho 6	AAEL011293-PA [Aedes aegypti]	144	144	44%	5.00E-38 48%
	PREDICTED: vitamin K epoxide reductase complex subunit 1-like protein 1 [Limulus polyphemus]	143	143	46%	1.00E-37 47%
	vitamin K epoxide reductase complex subunit 1 [Aedes aegypti]	143	143	44%	2.00E-37 48%
	vitamin K epoxide reductase complex subunit 1-like protein 1 [Danio rerio]	139	139	47%	7.00E-36 46%
	vitamin K epoxide reductase complex subunit 1-like protein 1 [Danio rerio]	139	139	47%	7.00E-36 46%
Ortho 7	PREDICTED: dolichol-phosphate mannosyltransferase subunit 1-like [Crassostrea gigas]	400	400	90%	7.00E-138 80%
	PREDICTED: probable dolichol-phosphate mannosyltransferase [Amyeloidis transitella]	399	399	91%	2.00E-137 78%
	PREDICTED: dolichol-phosphate mannosyltransferase subunit 1 [Lingula anatina]	397	397	91%	2.00E-136 77%
	PREDICTED: dolichyl-phosphate mannosyltransferase isoform X1 [Bombyx mori]	395	395	93%	9.00E-136 78%
	putative dolichol-phosphate mannosyltransferase [Danaus plexippus]	394	394	91%	2.00E-135 79%

Ortho 8	hypothetical protein DAPPUDRAFT_220118 [Daphnia pulex]	114	114	97%	4.00E-27	32%
	coiled-coil protein [Artemia franciscana]	95.1	95.1	97%	3.00E-20	28%
	PREDICTED: uncharacterized protein LOC105697737 isoform X1 [Orussus abietinus]	94.4	94.4	96%	9.00E-20	29%
	PREDICTED: uncharacterized protein LOC105697737 isoform X2 [Orussus abietinus]	94	94	96%	1.00E-19	29%
	Coiled-coil-helix-coiled-coil-helix domain-containing protein 3, mitochondrial precursor [Lepeophtheirus salmonis]	89.4	89.4	96%	5.00E-18	27%
Ortho 9	TMEM9 family protein [Coptotermes formosanus]	219	219	91%	4.00E-68	63%
	hypothetical protein L798_04988 [Zootermopsis nevadensis]	217	217	91%	2.00E-67	63%
	PREDICTED: uncharacterized protein CG1161 [Papilio polytes]	217	217	94%	2.00E-67	58%
	PREDICTED: uncharacterized protein CG1161 [Papilio machaon]	215	215	92%	1.00E-66	61%
	PREDICTED: uncharacterized protein CG1161 [Papilio xuthus]	213	213	81%	9.00E-66	67%
Ortho 10	PREDICTED: 28S ribosomal protein S21, mitochondrial [Halyomorpha halys]	115	115	78%	1.00E-30	57%
	AGAP009152-PA [Anopheles gambiae str. PEST]	114	114	75%	4.00E-30	57%
	mitochondrial 28S ribosomal protein S21 [Culex quinquefasciatus]	114	114	75%	8.00E-30	56%
	AAEL001909-PA [Aedes aegypti]	113	113	75%	1.00E-29	56%
	PREDICTED: 28S ribosomal protein S21, mitochondrial [Athalía rosae]	113	113	77%	1.00E-29	58%
Ortho 11	ribosomal protein L8 [Litopenaeus vannamei]	468	468	98%	2.00E-164	86%
	60S ribosomal protein L8 [Lepeophtheirus salmonis]	443	443	96%	1.00E-154	84%
	60S ribosomal protein L8 [Coptotermes formosanus]	442	442	97%	3.00E-154	82%
	PREDICTED: 60S ribosomal protein L8 [Athalía rosae]	441	441	98%	8.00E-154	81%
	60S ribosomal protein L8 [Laodelphax striatella]	440	440	97%	4.00E-153	83%
Ortho 12	PREDICTED: cold shock domain-containing protein CG9705-like [Plutella xylostella]	168	168	89%	2.00E-50	69%
	PREDICTED: cold shock domain-containing protein CG9705 [Bombyx mori]	166	166	89%	1.00E-49	70%
	PREDICTED: cold shock domain-containing protein CG9705 [Amyeloidis transitella]	166	166	89%	2.00E-49	70%
	calcium-regulated heat stable protein 1 [Acyrtosiphon pisum]	164	164	84%	8.00E-49	68%
	cold shock domain-containing protein CG9705 [Papilio xuthus]	161	161	89%	1.00E-47	68%
Ortho 13	conserved hypothetical protein [Pediculus humanus corporis]	223	223	82%	5.00E-67	45%
	Integral membrane protein 2B [Zootermopsis nevadensis]	222	222	95%	3.00E-66	41%
	PREDICTED: integral membrane protein 2B [Megachile rotundata]	204	204	95%	2.00E-59	38%
	PREDICTED: integral membrane protein 2C-like isoform X1 [Apis dorsata]	202	202	95%	1.00E-58	39%
	PREDICTED: integral membrane protein 2A [Tribolium castaneum]	201	201	95%	3.00E-58	42%
Ortho 14	PREDICTED: activator of 90 kDa heat shock protein ATPase homolog 1-like [Limulus polyphemus]	369	369	98%	1.00E-122	50%
	PREDICTED: activator of 90 kDa heat shock protein ATPase homolog 1 [Fopius arisanus]	360	360	98%	2.00E-119	53%
	PREDICTED: activator of 90 kDa heat shock protein ATPase homolog 1 [Halyomorpha halys]	358	358	98%	3.00E-118	51%
	PREDICTED: activator of 90 kDa heat shock protein ATPase homolog 1 isoform X1 [Acromyrmex echinator]	356	356	98%	1.00E-117	52%
	hypothetical protein L798_08367 [Zootermopsis nevadensis]	356	356	98%	2.00E-117	51%
Ortho 15	adenine nucleotide translocase 2 [Litopenaeus vannamei]	559	559	94%	0	89%
	adenine nucleotide translocase [Litopenaeus vannamei]	555	555	95%	0	88%
	ATP/ADP translocase [Marsupenaeus japonicus]	548	548	94%	0	88%
	adenine nucleotide translocase [Penaeus monodon]	548	548	94%	0	88%
	ATP/ADP translocase [Pacifastacus leniusculus]	546	546	95%	0	88%
Ortho 16	V-ATPase A [Cherax destructor]	1112	1112	98%	0	84%
	V-H-ATPase subunit A [Cherax cainii]	1107	1107	98%	0	84%
	v-type proton atpase catalytic subunit a [Lasius niger]	1043	1043	96%	0	80%
	PREDICTED: V-type proton ATPase catalytic subunit A [Megachile rotundata]	1041	1041	96%	0	81%

	PREDICTED: V-type proton ATPase catalytic subunit A [Trichogramma pretiosum]	1040	1040	96%	0	80%
Ortho 17	ribosomal protein S9 [Procambarus clarkii]	337	337	91%	1.00E-114	92%
	hypothetical protein HELRODRAFT_155894 [Helobdella robusta]	322	322	94%	1.00E-108	85%
	ribosomal protein S9 [Riptortus pedestris]	322	322	96%	1.00E-108	81%
	PREDICTED: 40S ribosomal protein S9 [Halyomorpha halys]	320	320	96%	3.00E-108	81%
	S9e ribosomal protein [Meladema coriacea]	319	319	96%	9.00E-108	82%
Ortho 18	PREDICTED: 60S acidic ribosomal protein P2 [Thamnophis sirtalis]	78.2	78.2	95%	1.00E-15	45%
	60S acidic ribosomal protein P2 [Salmo salar]	77	77	95%	3.00E-15	50%
	PREDICTED: LOW QUALITY PROTEIN: ribosomal protein, large, P2 [Pantholops hodgsonii]	76.6	76.6	95%	4.00E-15	50%
	PREDICTED: 60S acidic ribosomal protein P2 isoform X2 [Anolis carolinensis]	76.3	76.3	95%	6.00E-15	49%
	hypothetical protein BN946_scf185008.g40 [Trametes cinnabarina]	75.5	75.5	95%	1.00E-14	45%
Ortho 19	hypothetical protein DAPPUDRAFT_300779 [Daphnia pulex]	624	624	90%	0	65%
	PREDICTED: nucleolar protein 58 isoform X1 [Apis florea]	619	619	87%	0	66%
	PREDICTED: nucleolar protein 58-like [Apis dorsata]	619	619	87%	0	66%
	PREDICTED: nucleolar protein 58 isoform X2 [Apis florea]	619	619	87%	0	66%
	PREDICTED: nucleolar protein 58-like [Apis mellifera]	618	618	87%	0	66%
Ortho 20	hypothetical protein DAPPUDRAFT_302032 [Daphnia pulex]	199	199	87%	1.00E-59	48%
	PREDICTED: ran-specific GTPase-activating protein [Microplitis demolitor]	197	197	65%	2.00E-58	56%
	Ran-specific GTPase-activating protein [Stegodyphus mimosarum]	196	196	81%	5.00E-58	54%
	PREDICTED: LOW QUALITY PROTEIN: ran-specific GTPase-activating protein-like [Bombus terrestris]	196	196	70%	7.00E-58	52%
	PREDICTED: ran-specific GTPase-activating protein-like [Bombus impatiens]	196	196	70%	9.00E-58	52%
Ortho 21	Aldo-keto reductase family 1 member B10 [Zootermopsis nevadensis]	474	474	96%	1.00E-164	67%
	PREDICTED: aldose reductase-like [Stomoxys calcitrans]	451	451	94%	2.00E-155	66%
	PREDICTED: aldose reductase-like isoform X2 [Vollenhovia emeryi]	449	449	97%	6.00E-155	64%
	PREDICTED: aldose reductase-like [Monomorium pharaonis]	449	449	97%	9.00E-155	65%
	prostaglandin F synthase [Penaeus monodon]	448	448	95%	2.00E-154	69%
Ortho 22	PREDICTED: LOW QUALITY PROTEIN: 60S ribosomal protein L18 [Solenopsis invicta]	304	304	98%	5.00E-101	75%
	PREDICTED: 60S ribosomal protein L18 [Wasmannia auropunctata]	304	304	98%	1.00E-100	76%
	PREDICTED: LOW QUALITY PROTEIN: 60S ribosomal protein L18 [Monomorium pharaonis]	302	302	98%	9.00E-100	75%
	PREDICTED: LOW QUALITY PROTEIN: 60S ribosomal protein L18 [Atta cephalotes]	300	300	98%	1.00E-99	75%
	hypothetical protein SINV_04174 [Solenopsis invicta]	296	296	96%	8.00E-99	75%
Ortho 23	muscle-specific protein 20 [Solenopsis invicta]	238	238	96%	3.00E-76	62%
	hypothetical protein SINV_13357 [Solenopsis invicta]	239	239	93%	3.00E-75	64%
	PREDICTED: muscle-specific protein 20 [Bombus terrestris]	234	234	96%	7.00E-75	61%
	PREDICTED: muscle-specific protein 20 [Bombus impatiens]	234	234	96%	8.00E-75	61%
	PREDICTED: muscle-specific protein 20 [Pogonomyrmex barbatus]	233	233	96%	2.00E-74	63%
Ortho 24	translationally controlled tumor protein [Marsupenaeus japonicus]	262	262	97%	4.00E-86	77%
	translationally controlled tumor protein [Fenneropenaeus chinensis]	261	261	97%	1.00E-85	78%
	translationally controlled tumor protein [Fenneropenaeus merguensis]	261	261	97%	1.00E-85	77%
	translationally controlled tumor protein [Litopenaeus vannamei]	260	260	97%	3.00E-85	76%
	translationally controlled tumor protein [Marsupenaeus japonicus]	260	260	97%	4.00E-85	76%
Ortho 25	NADH dehydrogenase [ubiquinone] 1 beta subcomplex subunit 11, mitochondrial [Stegodyphus mimosarum]	112	112	73%	1.00E-27	49%
	mitochondrial NADH:ubiquinone oxidoreductase ESSS subunit, putative [Riptortus pedestris]	111	111	76%	2.00E-27	46%
	PREDICTED: NADH dehydrogenase [ubiquinone] 1 beta subcomplex subunit 11,	111	111	94%	3.00E-27	40%

	mitochondrial [Halyomorpha halys]					
	hypothetical protein CAPTEDRAFT_157472 [Capitella teleta]	111	111	94%	5.00E-27	37%
	PREDICTED: NADH dehydrogenase [ubiquinone] 1 beta subcomplex subunit 11, mitochondrial [Tribolium castaneum]	106	106	84%	1.00E-25	41%
Ortho 26	uncharacterized protein Dmoj_GI22699, isoform B [Drosophila mojavensis]	290	290	97%	1.00E-96	79%
	Cyclophilin-like [Drosophila melanogaster]	289	289	95%	1.00E-96	80%
	GI22699 [Drosophila mojavensis]	288	288	97%	2.00E-96	79%
	PREDICTED: peptidyl-prolyl cis-trans isomerase-like 1 [Ceratitis capitata]	288	288	95%	3.00E-96	81%
	PREDICTED: peptidyl-prolyl cis-trans isomerase-like 1 [Bactrocera oleae]	288	288	95%	3.00E-96	81%
Ortho 27	PREDICTED: p53 and DNA damage-regulated protein 1 [Latimeria chalumnae]	110	110	78%	4.00E-27	46%
	PREDICTED: p53 and DNA damage-regulated protein 1 [Clupea harengus]	101	101	78%	7.00E-24	43%
	p53 and DNA damage-regulated protein 1 [Anoplopoma fimbria]	100	100	76%	1.00E-23	41%
	PREDICTED: p53 and DNA damage-regulated protein 1 isoform X1 [Cynoglossus semilaevis]	100	100	78%	2.00E-23	41%
	PREDICTED: p53 and DNA damage-regulated protein 1 [Larimichthys crocea]	99.4	99.4	78%	6.00E-23	42%
Ortho 28	Regulator of G-protein signaling 3 [Stegodyphus mimosarum]	197	197	76%	2.00E-59	60%
	PREDICTED: regulator of G-protein signaling 2-like [Limulus polyphemus]	194	194	76%	9.00E-59	58%
	PREDICTED: regulator of G-protein signaling 12-like [Limulus polyphemus]	198	198	88%	7.00E-58	55%
	PREDICTED: regulator of G-protein signaling 3-like [Limulus polyphemus]	195	195	74%	1.00E-56	62%
	PREDICTED: regulator of G-protein signaling 2-like [Limulus polyphemus]	188	188	74%	3.00E-56	60%
Ortho 29	core histone H2A/H2B/H3/H4 [Dictyocaulus viviparus]	275	275	98%	4.00E-92	99%
	PREDICTED: LOW QUALITY PROTEIN: histone H3.3 [Felis catus]	275	275	97%	6.00E-92	100%
	Histone H3.3 [Habropoda laboriosa]	274	274	98%	1.00E-91	99%
	histone H3.3 [Homo sapiens]	273	273	97%	1.00E-91	100%
	Chain A, The Human Nucleosome Structure Containing The Histone Variant H3.3	273	273	97%	2.00E-91	100%
Ortho 30	ribonucleotide reductase, beta subunit, putative [Ixodes scapularis]	555	555	96%	0	73%
	PREDICTED: ribonucleoside-diphosphate reductase subunit M2 B-like [Limulus polyphemus]	537	537	98%	0	72%
	Ribonucleoside-diphosphate reductase subunit M2 [Stegodyphus mimosarum]	535	535	100%	0	71%
	ribonucleotide reductase small subunit R2ii [Carassius carassius]	535	535	98%	0	72%
	PREDICTED: ribonucleoside-diphosphate reductase subunit M2 B isoform X2 [Halyomorpha halys]	533	533	100%	0	71%
Ortho 31	cathepsin B (Marsupenaeus japonicus)	433	433	96%	3.00E-148	60%
	cathepsin B (Fenneropenaeus chinensis)	432	432	95%	7.00E-148	60%
	cathepsin B (Litopenaeus vannamei)	432	432	91%	1.00E-147	63%
	cathepsin B (Palaemon carinicauda)	430	430	99%	8.00E-147	59%
	cathepsin B (Penaeus monodon)	429	429	95%	9.00E-147	60%
Ortho 32	PREDICTED: MFS-type transporter SLC18B1-like (Trichogramma pretiosum)	488	488	97%	1.00E-164	54%
	uncharacterized protein LOC100123296 precursor (Nasonia vitripennis)	479	479	89%	2.00E-161	56%
	Chromaffin granule amine transporter, putative (Pediculus humanus corporis)	474	474	85%	3.00E-159	56%
	PREDICTED: MFS-type transporter SLC18B1-like (Solenopsis invicta)	472	472	88%	2.00E-158	55%
	PREDICTED: MFS-type transporter SLC18B1-like (Monomorium pharaonis)	471	471	88%	3.00E-158	55%
Ortho 33	PREDICTED: cytochrome b-c1 complex subunit 2, mitochondrial (Megachile rotundata)	304	304	91%	2.00E-94	42%
	Cytochrome b-c1 complex subunit 2, mitochondrial (Habropoda laboriosa)	299	299	91%	1.00E-92	41%
	cytochrome b-c1 complex subunit mitochondrial (Lasius niger)	293	293	99%	5.00E-90	40%
	PREDICTED: cytochrome b-c1 complex subunit 2, mitochondrial (Acromyrmex echinator)	289	289	99%	1.00E-88	39%
	Ubiquinol-cytochrome C reductase complex (Xenopus laevis)	289	289	92%	2.00E-88	40%
Ortho	PREDICTED: probable dynactin subunit 2 (Linepithema humile)	307	307	98%	9.00E-97	41%

34	putative dynactin subunit 2 (<i>Zootermopsis nevadensis</i>)	302	302	99%	7.00E-95	42%
	unknown (<i>Dendroctonus ponderosae</i>)	300	300	99%	3.00E-94	42%
	PREDICTED: probable dynactin subunit 2 (<i>Camponotus floridanus</i>)	301	301	98%	3.00E-94	40%
	PREDICTED: probable dynactin subunit 2 (<i>Monomorium pharaonis</i>)	297	297	98%	8.00E-93	40%
Ortho 35	PREDICTED: NADH-ubiquinone oxidoreductase 49 kDa subunit (<i>Bactrocera dorsalis</i>)	738	738	96%	0	74%
	PREDICTED: NADH-ubiquinone oxidoreductase 49 kDa subunit (<i>Bactrocera oleae</i>)	737	737	96%	0	74%
	PREDICTED: NADH-ubiquinone oxidoreductase 49 kDa subunit (<i>Bactrocera cucurbitae</i>)	736	736	96%	0	74%
	GH23924 (<i>Drosophila grimshawi</i>)	733	733	90%	0	79%
	PREDICTED: NADH-ubiquinone oxidoreductase 49 kDa subunit (<i>Ceratitis capitata</i>)	733	733	91%	0	78%
Ortho 36	PREDICTED: transmembrane protein 165 isoform X1 (<i>Plutella xylostella</i>)	282	282	75%	1.00E-89	56%
	PREDICTED: transmembrane protein 165 isoform X2 (<i>Plutella xylostella</i>)	281	281	80%	2.00E-89	54%
	PREDICTED: transmembrane protein 165 (<i>Microplitis demolitor</i>)	280	280	75%	8.00E-89	56%
	PREDICTED: transmembrane protein 165-like isoform X4 (<i>Halyomorpha halys</i>)	278	278	73%	1.00E-88	56%
	PREDICTED: transmembrane protein 165 isoform X3 (<i>Megachile rotundata</i>)	278	278	78%	1.00E-88	53%
Ortho 37	Xaa-Pro dipeptidase (<i>Zootermopsis nevadensis</i>)	303	303	70%	5.00E-87	34%
	PREDICTED: FAST kinase domain-containing protein 3-like (<i>Limulus polyphemus</i>)	253	253	97%	6.00E-70	26%
	AAEL008901-PA (<i>Aedes aegypti</i>)	243	243	92%	1.00E-66	27%
	PREDICTED: uncharacterized protein LOC105205340 (<i>Solenopsis invicta</i>)	236	236	91%	7.00E-64	28%
	conserved hypothetical protein (<i>Culex quinquefasciatus</i>)	233	233	92%	7.00E-63	28%
Ortho 38	PREDICTED: uncharacterized protein LOC106174438 isoform X2 (<i>Lingula anatina</i>)	275	275	94%	1.00E-82	56%
	PREDICTED: uncharacterized protein LOC106174438 isoform X5 (<i>Lingula anatina</i>)	273	273	82%	4.00E-82	61%
	PREDICTED: uncharacterized protein LOC106174438 isoform X4 (<i>Lingula anatina</i>)	273	273	82%	4.00E-82	61%
	PREDICTED: uncharacterized protein LOC106174438 isoform X1 (<i>Lingula anatina</i>)	274	274	82%	5.00E-82	61%
	PREDICTED: uncharacterized protein LOC106174438 isoform X3 (<i>Lingula anatina</i>)	273	273	82%	6.00E-82	61%
Ortho 40	ADP-ribosylation factor-like protein 8B-A (<i>Zootermopsis nevadensis</i>)	374	374	98%	1.00E-129	96%
	PREDICTED: ADP-ribosylation factor-like protein 8 (<i>Tribolium castaneum</i>)	372	372	98%	1.00E-128	95%
	PREDICTED: ADP-ribosylation factor-like protein 8B-A (<i>Acromyrmex echinator</i>)	371	371	98%	2.00E-128	96%
	PREDICTED: ADP-ribosylation factor-like protein 8B-A (<i>Trichogramma pretiosum</i>)	370	370	98%	3.00E-128	95%
	PREDICTED: ADP-ribosylation factor-like protein 8B-A (<i>Nasonia vitripennis</i>)	370	370	98%	3.00E-128	95%
Ortho 41	40S ribosomal protein S6 (<i>Zootermopsis nevadensis</i>)	408	408	42%	6.00E-136	79%
	PREDICTED: LOW QUALITY PROTEIN: 40S ribosomal protein S6 (<i>Linepithema humile</i>)	405	405	42%	6.00E-135	78%
	ribosomal protein S6 (<i>Riptortus pedestris</i>)	404	404	42%	3.00E-134	78%
	40s ribosomal protein s6-like protein (<i>Lasius niger</i>)	403	403	42%	6.00E-134	77%
	PREDICTED: 40S ribosomal protein S6 (<i>Camponotus floridanus</i>)	403	403	42%	8.00E-134	77%
Ortho 42	Thioredoxin domain-containing protein 15 (<i>Papilio xuthus</i>)	117	117	50%	2.00E-27	46%
	PREDICTED: thioredoxin domain-containing protein 15 (<i>Papilio xuthus</i>)	117	117	50%	3.00E-27	46%
	Thioredoxin domain-containing protein 15 (<i>Papilio machaon</i>)	117	117	50%	4.00E-27	45%
	PREDICTED: thioredoxin domain-containing protein 15 (<i>Papilio machaon</i>)	117	117	50%	4.00E-27	45%
	thioredoxin domain-containing protein 15 precursor (<i>Papilio xuthus</i>)	116	116	50%	4.00E-27	46%
Ortho 43	PREDICTED: tribbles homolog 2 (<i>Tribolium castaneum</i>)	256	256	79%	8.00E-79	49%
	PREDICTED: tribbles homolog 2 (<i>Cimex lectularius</i>)	250	250	78%	1.00E-76	48%
	PREDICTED: tribbles homolog 2 (<i>Atta cephalotes</i>)	250	250	87%	4.00E-76	47%
	PREDICTED: tribbles homolog 2 (<i>Orussus abietinus</i>)	249	249	88%	5.00E-76	45%
	PREDICTED: tribbles homolog 2 (<i>Plutella xylostella</i>)	249	249	79%	6.00E-76	51%

Ortho 44	troponin C2 (Litopenaeus vannamei)	234	234	98%	2.00E-75	79%
	troponin C isoform 4' (Homarus americanus)	229	229	96%	2.00E-73	77%
	troponin C (Crangon crangon)	224	224	98%	2.00E-71	74%
	troponin C isoform 3 (Homarus americanus)	223	223	98%	2.00E-71	75%
	troponin C1 (Litopenaeus vannamei)	220	220	98%	4.00E-70	74%
Ortho 45	N(G),N(G)-dimethylarginine dimethylaminohydrolase 1 (Stegodyphus mimosarum)	327	327	97%	2.00E-108	56%
	PREDICTED: N(G),N(G)-dimethylarginine dimethylaminohydrolase 1-like (Limulus polyphemus)	324	324	97%	4.00E-107	58%
	PREDICTED: N(G),N(G)-dimethylarginine dimethylaminohydrolase 1-like (Limulus polyphemus)	318	318	97%	6.00E-105	58%
	PREDICTED: LOW QUALITY PROTEIN: N(G),N(G)-dimethylarginine dimethylaminohydrolase 1-like (Limulus polyphemus)	310	310	94%	8.00E-102	56%
	PREDICTED: N(G),N(G)-dimethylarginine dimethylaminohydrolase 1 (Trichogramma pretiosum)	299	299	97%	2.00E-97	56%
Ortho 46	arginine kinase (Litopenaeus vannamei)	687	687	96%	0	90%
	arginine kinase (Fenneropenaeus merguensis)	686	686	96%	0	90%
	arginine kinase (Metapenaeus ensis)	686	686	96%	0	90%
	arginine kinase (Litopenaeus vannamei)	685	685	96%	0	90%
	RecName: Full=Arginine kinase; Short=AK; AltName: Allergen=Pen m 2 (Penaeus monodon)	685	685	96%	0	90%
Ortho 47	PREDICTED: phospholipase A1 member A (Nannospalax galili)	104	104	77%	5.00E-22	33%
	PREDICTED: phospholipase A1 member A isoform X2 (Camelus bactrianus)	100	100	77%	6.00E-21	32%
	PREDICTED: phospholipase A1 member A isoform X2 (Camelus dromedarius)	100	100	77%	7.00E-21	32%
	PREDICTED: phospholipase A1 member A isoform X2 (Camelus ferus)	100	100	77%	1.00E-20	32%
	PREDICTED: phospholipase A1 member A (Dipodomys ordii)	100	100	77%	2.00E-20	32%
Ortho 48	PREDICTED: stromal membrane-associated protein 1 (Amyeloid transitella)	327	327	77%	1.00E-103	49%
	PREDICTED: stromal membrane-associated protein 1 (Solenopsis invicta)	331	331	75%	1.00E-103	51%
	PREDICTED: stromal membrane-associated protein 1 (Papilio polytes)	327	327	64%	4.00E-103	53%
	PREDICTED: stromal membrane-associated protein 1 (Camponotus floridanus)	328	328	75%	4.00E-103	51%
	PREDICTED: stromal membrane-associated protein 1 (Wasmannia auropunctata)	329	329	88%	1.00E-102	45%
Ortho 49	C11orf46 homolog (Caligus clemensi)	93.6	93.6	31%	1.00E-19	45%
	PREDICTED: ARL14 effector protein (Amyeloid transitella)	92.8	92.8	27%	1.00E-19	49%
	Cdc2c (Operophtera brumata)	91.7	91.7	27%	3.00E-19	47%
	PREDICTED: ARL14 effector protein (Papilio polytes)	90.9	90.9	24%	6.00E-19	52%
	PREDICTED: ARL14 effector protein (Papilio machaon)	90.5	90.5	24%	6.00E-19	50%
Ortho 50	lethal 35Di (Culex quinquefasciatus)	160	160	81%	3.00E-46	53%
	lethal 35Di (Anopheles darlingi)	157	157	81%	6.00E-45	52%
	PREDICTED: NADH dehydrogenase (ubiquinone) 1 beta subcomplex subunit 6 (Amyeloid transitella)	154	154	82%	1.00E-43	51%
	PREDICTED: NADH dehydrogenase (ubiquinone) 1 beta subcomplex subunit 6 (Fopius arisanus)	154	154	81%	2.00E-43	50%
	AAEL009066-PA (Aedes aegypti)	152	152	81%	5.00E-43	52%
Ortho 51	PREDICTED: stromal cell-derived factor 2 (Trichogramma pretiosum)	252	252	99%	1.00E-80	53%
	PREDICTED: stromal cell-derived factor 2 (Tribolium castaneum)	251	251	85%	4.00E-80	59%
	Stromal cell-derived factor 2 (Zootermopsis nevadensis)	252	252	84%	3.00E-79	63%
	hypothetical protein DAPPUDRAFT_227363 (Daphnia pulex)	247	247	92%	2.00E-78	56%
	PREDICTED: stromal cell-derived factor 2 (Harpegnathos saltator)	247	247	85%	2.00E-78	56%
Ortho 52	PREDICTED: actin-related protein 2/3 complex subunit 2 (Bactrocera dorsalis)	472	472	97%	1.00E-164	75%
	PREDICTED: actin-related protein 2/3 complex subunit 2 (Bactrocera cucurbitae)	471	471	97%	4.00E-164	75%
	putative actin-related protein 2/3 complex subunit 2 (Zootermopsis nevadensis)	470	470	96%	1.00E-163	75%
	PREDICTED: actin-related protein 2/3 complex subunit 2 (Ceratitis capitata)	470	470	97%	1.00E-163	75%

	PREDICTED: actin-related protein 2/3 complex subunit 2 (<i>Bactrocera oleae</i>)	469	469	97%	2.00E-163	75%
Ortho 53	hypothetical protein AMK59_1816 (<i>Oryctes borbonicus</i>)	296	296	98%	5.00E-97	58%
	Proteasome activator complex subunit 3 (<i>Stegodyphus mimosarum</i>)	296	296	96%	5.00E-97	58%
	PREDICTED: proteasome activator complex subunit 3 isoform X2 (<i>Musca domestica</i>)	295	295	96%	2.00E-96	58%
	LOW QUALITY PROTEIN: uncharacterized protein Dmoj_GI15909 (<i>Drosophila mojavensis</i>)	293	293	98%	6.00E-96	57%
	PREDICTED: proteasome activator complex subunit 3 isoform X1 (<i>Musca domestica</i>)	293	293	97%	1.00E-95	57%
Ortho 54	PREDICTED: uncharacterized protein C19orf60 homolog (<i>Limulus polyphemus</i>)	113	113	75%	5.00E-28	46%
	fed tick salivary protein 6 (<i>Ixodes scapularis</i>)	108	108	79%	3.00E-26	43%
	secreted salivary gland peptide, putative (<i>Ixodes scapularis</i>)	109	109	79%	4.00E-26	43%
	PREDICTED: uncharacterized protein C19orf60 homolog (<i>Lingula anatina</i>)	105	105	76%	8.00E-25	42%
	PREDICTED: uncharacterized protein C19orf60 homolog (<i>Meleagris gallopavo</i>)	105	105	75%	1.00E-24	41%
Ortho 55	PREDICTED: synaptosomal-associated protein 29 (<i>Amyeloidis transitella</i>)	172	172	95%	1.00E-47	40%
	Synaptosomal-associated protein 29 (<i>Zootermopsis nevadensis</i>)	168	168	84%	3.00E-46	42%
	hypothetical protein DAPPUDRAFT_194582 (<i>Daphnia pulex</i>)	165	165	80%	5.00E-45	43%
	PREDICTED: soluble NSF attachment protein 29 (<i>Bombyx mori</i>)	163	163	98%	2.00E-44	38%
	PREDICTED: synaptosomal-associated protein 29 (<i>Papilio machaon</i>)	162	162	98%	3.00E-44	38%
Ortho 56	mitochondrial cytochrome c oxidase subunit Va (<i>Litopenaeus vannamei</i>)	229	229	95%	2.00E-73	68%
	PREDICTED: cytochrome c oxidase subunit 5A, mitochondrial (<i>Tribolium castaneum</i>)	223	223	96%	5.00E-71	67%
	AAEL014944-PA (<i>Aedes aegypti</i>)	218	218	96%	2.00E-69	67%
	cytochrome c oxidase polypeptide (<i>Anopheles darlingi</i>)	216	216	96%	2.00E-68	66%
	Cytochrome c oxidase subunit 5A, mitochondrial (<i>Zootermopsis nevadensis</i>)	216	216	96%	3.00E-68	64%
Ortho 57	Pinx1 (<i>Litopenaeus vannamei</i>)	172	172	29%	2.00E-46	51%
	PREDICTED: PIN2/TERF1-interacting telomerase inhibitor 1 (<i>Fundulus heteroclitus</i>)	176	176	40%	2.00E-45	42%
	Pin2-interacting protein X1 (<i>Esox lucius</i>)	170	170	36%	4.00E-45	42%
	PREDICTED: PIN2/TERF1-interacting telomerase inhibitor 1 (<i>Esox lucius</i>)	172	172	36%	3.00E-44	42%
	PREDICTED: PIN2/TERF1-interacting telomerase inhibitor 1 (<i>Clupea harengus</i>)	169	169	40%	5.00E-43	42%
Ortho 58	Tubulin-specific chaperone cofactor E-like protein (<i>Cerapachys biroi</i>)	365	365	99%	2.00E-118	46%
	Tubulin-specific chaperone cofactor E-like protein (<i>Habropoda laboriosa</i>)	357	357	99%	3.00E-115	45%
	PREDICTED: tubulin-specific chaperone cofactor E-like protein isoform X2 (<i>Cimex lectularius</i>)	347	347	99%	1.00E-111	43%
	PREDICTED: tubulin-specific chaperone cofactor E-like protein isoform X1 (<i>Cimex lectularius</i>)	348	348	99%	2.00E-111	43%
	PREDICTED: tubulin-specific chaperone cofactor E-like protein (<i>Plutella xylostella</i>)	341	341	99%	5.00E-109	42%
Ortho 59	nucleotide excision repair protein (<i>Marsupenaeus japonicus</i>)	423	423	98%	9.00E-143	65%
	UV excision repair protein RAD23-like protein B (<i>Zootermopsis nevadensis</i>)	360	360	98%	4.00E-118	52%
	PREDICTED: UV excision repair protein RAD23 homolog B-like (<i>Limulus polyphemus</i>)	343	343	98%	3.00E-111	52%
	uv excision repair protein rad23, putative (<i>Pediculus humanus corporis</i>)	338	338	98%	8.00E-110	49%
	PREDICTED: UV excision repair protein RAD23 homolog B (<i>Orussus abietinus</i>)	337	337	98%	3.00E-109	49%
Ortho 61	PREDICTED: charged multivesicular body protein 2b (<i>Bombyx mori</i>)	254	254	94%	1.00E-81	60%
	vacuolar assembly/sorting protein DID4, putative (<i>Ixodes scapularis</i>)	254	254	93%	2.00E-81	60%
	PREDICTED: charged multivesicular body protein 2b (<i>Amyeloidis transitella</i>)	253	253	93%	3.00E-81	60%
	PREDICTED: charged multivesicular body protein 2b-B (<i>Papilio xuthus</i>)	251	251	93%	4.00E-80	58%
	PREDICTED: charged multivesicular body protein 2b-B (<i>Plutella xylostella</i>)	249	249	93%	1.00E-79	60%
Ortho 62	PREDICTED: DNA-directed RNA polymerases I and III subunit RPAC1-like (<i>Lingula anatina</i>)	317	317	85%	2.00E-102	52%
	hypothetical protein DAPPUDRAFT_213644 (<i>Daphnia pulex</i>)	310	310	84%	2.00E-99	52%
	DNA-directed RNA polymerases I and III subunit RPAC1 (<i>Zootermopsis nevadensis</i>)	307	307	84%	1.00E-98	51%

	PREDICTED: DNA-directed RNA polymerases I and III subunit RPAC1-like isoform X1 (<i>Limulus polyphemus</i>)	307	307	87%	2.00E-98	52%
	PREDICTED: DNA-directed RNA polymerases I and III subunit RPAC1 (<i>Bombyx mori</i>)	305	305	82%	8.00E-98	54%
Ortho 63	PREDICTED: methylmalonic aciduria and homocystinuria type D homolog, mitochondrial-like (<i>Chrysochloris asiatica</i>)	214	214	87%	3.00E-63	41%
	PREDICTED: methylmalonic aciduria and homocystinuria type D protein, mitochondrial (<i>Erinaceus europaeus</i>)	213	213	88%	4.00E-63	41%
	PREDICTED: methylmalonic aciduria and homocystinuria type D protein, mitochondrial (<i>Loxodonta africana</i>)	213	213	87%	7.00E-63	42%
	PREDICTED: methylmalonic aciduria and homocystinuria type D protein, mitochondrial (<i>Condylura cristata</i>)	212	212	87%	2.00E-62	41%
	PREDICTED: methylmalonic aciduria and homocystinuria type D homolog, mitochondrial-like (<i>Bubalus bubalis</i>)	211	211	87%	3.00E-62	42%
Ortho 64	E3 ubiquitin-protein ligase TM129 (<i>Xenopus laevis</i>)	303	303	93%	4.00E-96	42%
	E3 ubiquitin-protein ligase TM129 (<i>Xenopus (Silurana) tropicalis</i>)	301	301	93%	2.00E-95	42%
	PREDICTED: transmembrane protein 129 (<i>Chrysemys picta bellii</i>)	296	296	93%	2.00E-93	40%
	E3 ubiquitin-protein ligase TM129 isoform X1 (<i>Amazona aestiva</i>)	296	296	93%	2.00E-93	42%
	PREDICTED: E3 ubiquitin-protein ligase TM129 isoform X1 (<i>Falco peregrinus</i>)	295	295	93%	6.00E-93	41%
Ortho 65	unknown (<i>Dendroctonus ponderosae</i>)	377	377	98%	6.00E-124	54%
	PREDICTED: zinc finger CCCH domain-containing protein 15 homolog (<i>Fopius arisanus</i>)	372	372	94%	5.00E-122	55%
	PREDICTED: zinc finger CCCH domain-containing protein 15 homolog (<i>Monomorium pharaonis</i>)	364	364	94%	7.00E-119	56%
	PREDICTED: zinc finger CCCH domain-containing protein 15 homolog (<i>Atta cephalotes</i>)	362	362	88%	5.00E-118	55%
	PREDICTED: zinc finger CCCH domain-containing protein 15 homolog (<i>Acromyrmex echinator</i>)	362	362	88%	7.00E-118	56%
Ortho 66	PREDICTED: coiled-coil domain-containing protein 86-like (<i>Salmo salar</i>)	104	104	76%	2.00E-24	50%
	unnamed protein product (<i>Oncorhynchus mykiss</i>)	106	106	76%	2.00E-24	50%
	Coiled-coil domain-containing protein 86 (<i>Melipona quadrifasciata</i>)	101	101	63%	2.00E-23	51%
	PREDICTED: coiled-coil domain-containing protein 86 (<i>Esox lucius</i>)	103	103	75%	5.00E-23	50%
	PREDICTED: coiled-coil domain-containing protein 86 (<i>Wasmannia auropunctata</i>)	100	100	72%	6.00E-23	46%
Ortho 67	hypothetical protein SINV_14247 (<i>Solenopsis invicta</i>)	157	157	88%	1.00E-45	81%
	TPA: alternative splicing factor ASF/SF2 (<i>Amblyomma variegatum</i>)	157	157	82%	4.00E-45	89%
	Splicing factor, arginine/serine-rich 1 (<i>Harpegnathos saltator</i>)	155	155	87%	6.00E-45	82%
	PREDICTED: serine/arginine-rich splicing factor 1B (<i>Megachile rotundata</i>)	156	156	87%	9.00E-45	82%
	PREDICTED: serine/arginine-rich splicing factor 1A (<i>Orussus abietinus</i>)	156	156	87%	1.00E-44	82%
Ortho 68	PREDICTED: bridging integrator 3-like (<i>Athalia rosae</i>)	88.2	88.2	91%	5.00E-17	26%
	PREDICTED: bridging integrator 3-like (<i>Solenopsis invicta</i>)	83.6	83.6	91%	1.00E-15	25%
	PREDICTED: bridging integrator 3-like (<i>Monomorium pharaonis</i>)	83.2	83.2	91%	2.00E-15	25%
	PREDICTED: bridging integrator 3-like isoform X1 (<i>Cerapachys biroii</i>)	83.2	83.2	91%	2.00E-15	25%
	PREDICTED: bridging integrator 3 homolog (<i>Harpegnathos saltator</i>)	83.2	83.2	91%	2.00E-15	26%
Ortho 69	hypothetical protein CAPTEDRAFT_152195 (<i>Capitella teleta</i>)	231	231	99%	1.00E-72	51%
	hypothetical protein LOTGIDRAFT_222883 (<i>Lottia gigantea</i>)	229	229	99%	9.00E-72	52%
	PREDICTED: Golgi SNAP receptor complex member 2-like (<i>Limulus polyphemus</i>)	216	216	99%	2.00E-66	49%
	PREDICTED: Golgi SNAP receptor complex member 2-like (<i>Biomphalaria glabrata</i>)	212	212	99%	4.00E-65	50%
	PREDICTED: Golgi SNAP receptor complex member 2-like isoform X1 (<i>Saccoglossus kowalevskii</i>)	211	211	99%	6.00E-65	53%
Ortho 70	NEDD8-conjugating enzyme Ubc12 (<i>Zootermopsis nevadensis</i>)	325	325	96%	2.00E-110	85%
	PREDICTED: NEDD8-conjugating enzyme Ubc12 (<i>Fopius arisanus</i>)	322	322	96%	5.00E-109	85%
	NEDD8-conjugating enzyme Ubc12, putative (<i>Pediculus humanus corporis</i>)	321	321	96%	1.00E-108	83%
	PREDICTED: NEDD8-conjugating enzyme Ubc12 (<i>Monomorium pharaonis</i>)	320	320	96%	3.00E-108	85%
	hypothetical protein CAPTEDRAFT_149080 (<i>Capitella teleta</i>)	320	320	96%	3.00E-108	83%
Ortho	OTU domain ubiquitin aldehyde binding protein (<i>Scylla paramamosain</i>)	392	392	84%	8.00E-134	73%

71	OTU domain, ubiquitin aldehyde binding protein, putative (<i>Ixodes scapularis</i>)	276	276	85%	5.00E-88	53%
	PREDICTED: ubiquitin thioesterase OTUB1-like isoform X1 (<i>Limulus polyphemus</i>)	276	276	84%	7.00E-88	53%
	PREDICTED: ubiquitin thioesterase otubain-like (<i>Fopius arisanus</i>)	273	273	88%	4.00E-87	50%
	PREDICTED: ubiquitin thioesterase otubain-like (<i>Microplitis demolitor</i>)	273	273	88%	5.00E-87	49%
Ortho 72	Ribosomal RNA small subunit methyltransferase NEP1 (<i>Habropoda laboriosa</i>)	302	302	89%	3.00E-99	59%
	hypothetical protein LOTGIDRAFT_208870 (<i>Lottia gigantea</i>)	301	301	81%	5.00E-99	68%
	PREDICTED: ribosomal RNA small subunit methyltransferase NEP1 isoform X1 (<i>Diaphorina citri</i>)	300	300	82%	1.00E-98	67%
	PREDICTED: ribosomal RNA small subunit methyltransferase NEP1 (<i>Pogonomyrmex barbatus</i>)	300	300	82%	2.00E-98	65%
	PREDICTED: ribosomal RNA small subunit methyltransferase NEP1 (<i>Apis florea</i>)	299	299	82%	3.00E-98	66%
Ortho 73	PREDICTED: fucose mutarotase-like isoform X1 (<i>Crassostrea gigas</i>)	164	164	89%	4.00E-48	53%
	PREDICTED: fucose mutarotase-like (<i>Limulus polyphemus</i>)	161	161	90%	8.00E-47	53%
	PREDICTED: fucose mutarotase isoform X1 (<i>Oryzias latipes</i>)	157	157	90%	4.00E-45	53%
	PREDICTED: fucose mutarotase (<i>Monodelphis domestica</i>)	156	156	90%	8.00E-45	50%
	PREDICTED: fucose mutarotase (<i>Xiphophorus maculatus</i>)	155	155	90%	2.00E-44	52%
Ortho 74	RING finger protein 141 (<i>Stegodyphus mimosarum</i>)	160	160	86%	5.00E-43	31%
	AGAP000736-PA-like protein (<i>Anopheles sinensis</i>)	159	159	77%	4.00E-42	35%
	zinc finger protein (<i>Oryctes borbonicus</i>)	154	154	74%	4.00E-41	36%
	hypothetical protein Phum_PHUM055710 (<i>Pediculus humanus corporis</i>)	156	156	86%	5.00E-41	32%
	PREDICTED: RING finger protein 141-like isoform X2 (<i>Musca domestica</i>)	154	154	82%	6.00E-41	33%
Ortho 75	conserved hypothetical protein (<i>Pediculus humanus corporis</i>)	332	332	69%	2.00E-109	69%
	Ubiquitin domain-containing protein UBFD1 (<i>Zootermopsis nevadensis</i>)	331	331	74%	3.00E-109	65%
	PREDICTED: ubiquitin domain-containing protein UBFD1-like isoform X2 (<i>Limulus polyphemus</i>)	327	327	70%	7.00E-108	67%
	PREDICTED: ubiquitin domain-containing protein UBFD1-like isoform X1 (<i>Limulus polyphemus</i>)	322	322	70%	8.00E-106	67%
					3.00E-104	
	PREDICTED: ubiquitin domain-containing protein UBFD1-like (<i>Apis mellifera</i>)	319	319	89%	104	56%
Ortho 76	PREDICTED: potassium channel subfamily K member 1-like isoform X1 (<i>Cimex lectularius</i>)	383	383	84%	7.00E-128	56%
	PREDICTED: potassium channel subfamily K member 1-like isoform X2 (<i>Cimex lectularius</i>)	376	376	83%	3.00E-125	58%
	PREDICTED: potassium channel subfamily K member 1-like (<i>Limulus polyphemus</i>)	374	374	86%	2.00E-123	56%
	PREDICTED: potassium channel subfamily K member 6-like isoform X2 (<i>Halyomorpha halys</i>)	357	357	87%	2.00E-117	53%
	hypothetical protein DAPPUDRAFT_218276 (<i>Daphnia pulex</i>)	356	356	84%	9.00E-117	53%
Ortho 77	PREDICTED: 26S proteasome non-ATPase regulatory subunit 14 (<i>Athalia rosae</i>)	575	575	99%	0	88%
	PREDICTED: 26S proteasome non-ATPase regulatory subunit 14 (<i>Fopius arisanus</i>)	574	574	99%	0	88%
	PREDICTED: 26S proteasome non-ATPase regulatory subunit 14 (<i>Trichogramma pretiosum</i>)	571	571	99%	0	88%
	PREDICTED: 26S proteasome non-ATPase regulatory subunit 14 (<i>Acromyrmex echinator</i>)	571	571	99%	0	88%
	PREDICTED: 26S proteasome non-ATPase regulatory subunit 14 (<i>Nasonia vitripennis</i>)	571	571	99%	0	88%
Ortho 78	ribosomal protein L5 (<i>Riptortus pedestris</i>)	453	453	97%	4.00E-157	73%
	PREDICTED: 60S ribosomal protein L5 (<i>Tribolium castaneum</i>)	451	451	96%	3.00E-156	72%
	PREDICTED: 60S ribosomal protein L5-like (<i>Metaseiulus occidentalis</i>)	448	448	97%	3.00E-155	71%
	PREDICTED: 60S ribosomal protein L5 (<i>Ciona intestinalis</i>)	447	447	97%	1.00E-154	71%
	60S ribosomal protein L5 (<i>Lucilia cuprina</i>)	444	444	96%	1.00E-153	71%
Ortho 79	hypothetical protein DAPPUDRAFT_92230 (<i>Daphnia pulex</i>)	165	165	92%	3.00E-50	84%
	PREDICTED: barrier-to-autointegration factor-like (<i>Limulus polyphemus</i>)	162	162	92%	2.00E-49	84%
	conserved hypothetical protein (<i>Pediculus humanus corporis</i>)	160	160	92%	3.00E-48	80%
	PREDICTED: barrier-to-autointegration factor-like (<i>Limulus polyphemus</i>)	159	159	92%	4.00E-48	81%

	PREDICTED: barrier-to-autointegration factor (<i>Halyomorpha halys</i>)	158	158	92%	2.00E-47	82%
Ortho 80	myosin light chain 2 (<i>Procambarus clarkii</i>)	215	215	77%	3.00E-67	73%
	myosin light chain (<i>Artemia franciscana</i>)	199	199	86%	4.00E-61	57%
	myosin light chain 2 (<i>Antheraea pernyi</i>)	196	196	86%	1.00E-59	61%
	Myosin regulatory light chain 2 (<i>Papilio xuthus</i>)	197	197	88%	1.00E-59	60%
	myosin regulatory light chain 2 (<i>Papilio xuthus</i>)	196	196	86%	2.00E-59	60%
Ortho 81	hypothetical protein DAPPUDRAFT_216285 (<i>Daphnia pulex</i>)	323	323	86%	7.00E-105	54%
	PREDICTED: growth hormone-inducible transmembrane protein (<i>Tribolium castaneum</i>)	315	315	92%	2.00E-101	51%
	hypothetical protein LOTGIDRAFT_210320 (<i>Lottia gigantea</i>)	312	312	77%	3.00E-100	56%
	hypothetical protein OCBIM_22018462mg (<i>Octopus bimaculoides</i>)	310	310	77%	2.00E-99	55%
	PREDICTED: growth hormone-inducible transmembrane protein-like (<i>Halyomorpha halys</i>)	306	306	95%	4.00E-98	49%
Ortho 82	hypothetical protein DAPPUDRAFT_189030 (<i>Daphnia pulex</i>)	379	379	94%	2.00E-128	66%
	PREDICTED: eukaryotic translation initiation factor 3 subunit F-like (<i>Lingula anatina</i>)	357	357	94%	5.00E-120	62%
	ukaryotic translation initiation factor 3 subunit, putative (<i>Pediculus humanus corporis</i>)	345	345	95%	3.00E-115	58%
	Eukaryotic translation initiation factor 3 subunit F-1 (<i>Zootermopsis nevadensis</i>)	339	339	95%	1.00E-112	57%
	PREDICTED: eukaryotic translation initiation factor 3 subunit F-like (<i>Biomphalaria glabrata</i>)	335	335	94%	2.00E-111	58%
Ortho 83	PREDICTED: 26S proteasome non-ATPase regulatory subunit 11 (<i>Tribolium castaneum</i>)	574	574	89%	0	71%
	hypothetical protein TcasGA2_TC012621 (<i>Tribolium castaneum</i>)	574	574	89%	0	71%
	uncharacterized protein Dyak_GE12293, isoform B (<i>Drosophila yakuba</i>)	570	570	96%	0	68%
	GE12293 (<i>Drosophila yakuba</i>)	570	570	97%	0	67%
	GG22404 (<i>Drosophila erecta</i>)	570	570	97%	0	67%
Ortho 84	PREDICTED: mRNA-decapping enzyme 1A (<i>Papilio machaon</i>)	155	155	96%	9.00E-39	30%
	hypothetical protein KGM_11333 (<i>Danaus plexippus</i>)	152	152	96%	8.00E-38	30%
	PREDICTED: mRNA-decapping enzyme 1B isoform X1 (<i>Bombyx mori</i>)	150	150	96%	5.00E-37	29%
	PREDICTED: mRNA-decapping enzyme 1B isoform X2 (<i>Bombyx mori</i>)	149	149	96%	1.00E-36	29%
	PREDICTED: mRNA-decapping enzyme-like protein (<i>Cerapachys biroi</i>)	146	146	98%	2.00E-35	29%
Ortho 85	PREDICTED: probable 39S ribosomal protein L45, mitochondrial (<i>Halyomorpha halys</i>)	318	318	86%	7.00E-103	51%
	putative 39S ribosomal protein L45, mitochondrial (<i>Zootermopsis nevadensis</i>)	315	315	81%	2.00E-101	53%
	PREDICTED: probable 39S ribosomal protein L45, mitochondrial (<i>Limulus polyphemus</i>)	313	313	82%	2.00E-100	52%
	conserved hypothetical protein (<i>Ixodes scapularis</i>)	310	310	73%	1.00E-99	53%
	AAEL012025-PA (<i>Aedes aegypti</i>)	303	303	93%	9.00E-97	46%
Ortho 86	RNA polymerase II subunit A C-terminal domain phosphatase SSU72 (<i>Zootermopsis nevadensis</i>)	327	327	96%	7.00E-111	78%
	PREDICTED: RNA polymerase II subunit A C-terminal domain phosphatase SSU72 isoform X1 (<i>Vollenhovia emeryi</i>)	315	315	95%	6.00E-106	76%
	PREDICTED: RNA polymerase II subunit A C-terminal domain phosphatase SSU72 isoform X2 (<i>Orussus abietinus</i>)	315	315	95%	8.00E-106	75%
	PREDICTED: RNA polymerase II subunit A C-terminal domain phosphatase SSU72 isoform X3 (<i>Orussus abietinus</i>)	314	314	95%	1.00E-105	75%
	PREDICTED: RNA polymerase II subunit A C-terminal domain phosphatase SSU72 (<i>Megachile rotundata</i>)	314	314	95%	1.00E-105	75%
Ortho 87	hypothetical protein DAPPUDRAFT_306247 (<i>Daphnia pulex</i>)	124	124	97%	1.00E-26	30%
	PREDICTED: mediator of RNA polymerase II transcription subunit 26 (<i>Athalia rosae</i>)	120	120	97%	2.00E-25	26%
	PREDICTED: mediator of RNA polymerase II transcription subunit 26-like (<i>Ceratosolen solmsi marchali</i>)	112	112	97%	9.00E-23	25%
	PREDICTED: mediator of RNA polymerase II transcription subunit 26-like (<i>Lingula anatina</i>)	111	111	97%	4.00E-22	26%
	PREDICTED: uncharacterized protein LOC106473716 (<i>Limulus polyphemus</i>)	107	107	97%	5.00E-21	26%
Ortho 88	PREDICTED: PCTP-like protein (<i>Cimex lectularius</i>)	282	282	84%	4.00E-91	59%
	PREDICTED: PCTP-like protein (<i>Halyomorpha halys</i>)	281	281	98%	1.00E-90	52%

	PCTP-like protein (<i>Zootermopsis nevadensis</i>)	280	280	87%	4.00E-90	55%
	PREDICTED: PCTP-like protein (<i>Crassostrea gigas</i>)	269	269	87%	4.00E-86	52%
	PREDICTED: PCTP-like protein (<i>Solenopsis invicta</i>)	269	269	82%	6.00E-86	58%
Ortho 89	conserved hypothetical protein (<i>Ixodes scapularis</i>)	315	315	97%	1.00E-106	82%
	hypothetical protein DAPPUDRAFT_303043 (<i>Daphnia pulex</i>)	309	309	97%	4.00E-104	80%
	n-terminal acetyltransferase complex ard1 subunit (<i>Loa loa</i>)	297	297	97%	2.00E-99	80%
	N-terminal acetyltransferase complex ARD1 subunit homolog, putative (<i>Brugia malayi</i>)	296	296	97%	3.00E-99	80%
	N-terminal acetyltransferase complex ARD1 subunit homolog, putative (<i>Brugia malayi</i>)	296	296	97%	3.00E-99	80%
Ortho 90	Fkbp13, isoform B (<i>Drosophila yakuba</i>)	272	272	96%	1.00E-88	62%
	uncharacterized protein Dsimw501_GD25197, isoform C (<i>Drosophila simulans</i>)	272	272	96%	1.00E-88	62%
	GD25197 (<i>Drosophila simulans</i>)	273	273	96%	1.00E-88	62%
	GM15719 (<i>Drosophila sechellia</i>)	272	272	96%	2.00E-88	62%
	Fkbp13 (<i>Drosophila yakuba</i>)	272	272	96%	3.00E-88	62%

Supplementary Table S8. Sequences, PCR product size, and melting temperature of the primers designed to complete the phylogeny.

<i>Locus</i>	Primer sequences		PCR Product size (nuc)	Melting Temperature
	Forward	Reverse		
CYTOB2	AATTGGAAGTGARGAAGCAGCA	TTACRGACTGAACWGATGARGCW	437	58°C
UBI4	GAAACTTCTGGGGATTGAAGCAC	ACCGGYWGGCATTGATTGA	595	57°C
NADH19	GAT TCCWCCTCTGCTYTGGT	GAYTCTCGCTTGACCCAATCM	248	56°C
METH24	TCTTGTTTCGAGCCAAARAYT	TGGGCATGGATAMCCRTTAAG	421	54°C
RIB46	GAACCWTATGTGCCYCCMGA	GKAYACGCCACAACCCRTA	529	56°C
INN2	GCAGCTCTTTYTGGTKGACAG	AAGCGTGAAGCCCTTCTTTG	571	57°C
XAA6	GCTGAAAAAGYRTGGGTGGA	AAKCMTGRTCTGGGATATAAGM	556	52°C
VATP10	GATCCAGCGTCCRCTTCARG	CTCCRCAACCCACGTACACA	536	58°C
GROW42	YTTRGGKCTTGAGCMCTGT	AGGACCTCCCCARCTCAAGA	481	58°C
PROT44	AAGCAGGRATTWTGGSTYTGG	RATGGCATTTCAGTRGTWCG	461	56°C

Impact des bactéries endosymbiotiques féminisantes du genre *Wolbachia* sur l'évolution des chromosomes sexuels d'isopodes terrestres

Les Oniscidea présentent une diversité remarquable de systèmes chromosomiques de déterminisme du sexe (hétérogamétie mâle XX/XY ou hétérogamétie femelle ZW/ZZ), dont l'origine reste encore largement inconnue à ce jour. Il a été proposé que ces différents systèmes puissent être le produit de la coévolution entre les isopodes terrestres et *Wolbachia*, une bactérie endosymbiotique féminisante transmise verticalement par voie ovocytaire. Dans le but de caractériser l'impact de l'endosymbiose à *Wolbachia* sur l'évolution des mécanismes de déterminisme du sexe, nous avons utilisé une combinaison d'approches génomique, transcriptomique et d'expression de gènes. Tout d'abord, le génome de l'espèce *Armadillidium nasatum* (caractérisée par un système XX/XY) a été généré et ensuite annoté structurellement et fonctionnellement. A partir de ce génome, des approches de génomique comparatives ont permis la caractérisation de séquences liées au chromosomes Y, afin de mieux comprendre les processus impliqués dans la dégénérescence des gonosomes. Afin d'identifier des effecteurs liés au déterminisme ou à la différenciation du sexe, une approche par gènes candidats a permis de caractériser des gènes à domaines DM, connus pour être impliqués dans le déterminisme du sexe de nombreuses espèces, et d'en mesurer l'expression au cours du temps. Enfin, une phylogénie des Oniscidea a été réalisée en parallèle d'expériences de réversion de sexe afin d'estimer le nombre et la direction des transitions de systèmes d'hétérogamétie au cours de l'évolution des isopodes terrestres. Ces travaux contribuent à illustrer l'impact de l'endosymbiose sur l'évolution des mécanismes de déterminisme du sexe de l'hôte.

Impact of a feminizing endosymbiotic bacteria (genus *Wolbachia*) on the evolution of terrestrial isopods sex chromosomes

Oniscidea show a remarkable diversity of chromosomal sex determination systems (male heterogamety XX/XY or female heterogamety ZW / ZZ). However, the origin of such diversity is still largely unknown to date. It has been proposed that these different systems may be the product of the coevolution between terrestrial isopods and *Wolbachia*, a feminizing endosymbiotic bacteria transmitted vertically through oocytes. In order to characterize the impact of *Wolbachia* endosymbiosis on the evolution of sex determination mechanisms, we used a combination of genomic, transcriptomic and gene expression approaches. First, the genome of the species *Armadillidium nasatum* (characterized by an XX/XY system) was generated and then structurally and functionally annotated. From this genome, a comparative genomic approach allowed us to characterize sequences Y-linked, in order to better understand the processes involved in the sex chromosome degeneration. In order to identify effectors potentially related to sex determination or differentiation, a candidate gene approach has been used to characterize DM-domain genes, known to be involved in the sex determination pathways of many species, and then to measure their expression over development. Finally, a Oniscidea phylogeny was generated in parallel with sex-reversal experiments in order to characterize the number and the direction of the transitions of heterogenetic systems during the terrestrial isopods evolution. This work emphasize the impact of endosymbiosis on the evolution of host sex determination mechanisms.