Génomique évolutive chez les champignons Microbotryum : adaptation et chromosomes de types sexuels

Hélène Badouin

NNT : 2016SACLS011

THESE DE DOCTORAT
DE L'UNIVERSITE PARIS-SACLAY,
préparée à l'Université Paris-Sud

ÉCOLE DOCTORALE N° 567
Sciences du Végétal : du Gène à l'Ecosystème

Biologie

Par

Mademoiselle Hélène Badouin

Génomique évolutive chez les champignons *Microbotryum* : adaptation à l'hôte et chromosomes de types sexuels

Thèse présentée et soutenue à Orsay, le 9 février 2016 :

Composition du Jury :

M. Pierre Capy	Professeur, Université Paris Sud	Président du Jury
Mme Susana Coelho	CR CNRS, Station biologique de Roscoff	Rapportrice
M. Sylvain Glémin	CR CNRS, ISEM	Rapporteur
M. Gabriel Marais	DR CNRS, LBBE	Examinateur
M. Stéphane De Mita	CR INRA, INRA Nancy	Examinateur
Mme Tatiana Giraud	DR CNRS, ESE	Directrice de thèse

Remerciements

"La grandeur d'un métier est peut-être, avant tout, d'unir les hommes : il n'est qu'un luxe véritable, et c'est celui des relations humaines."

– Antoine de Saint-Exupéry, *Terre des hommes*

Je souhaiterais tout d'abord remercier chaleureusement les membres de mon jury, Susana Coelho et Sylvain Glémin pour avoir accepté d'être rapporteurs de ma thèse, Gabriel Marais, Stéphane de Mita et Pierre Capy pour avoir accepté d'être examinateurs, et bien sûr Tatiana Giraud, ma directrice de thèse.

Cette thèse a été une véritable aventure, pendant laquelle j'ai eu la chance de rencontrer des personnes passionnées, de grandir et m'enrichir à leur contact.

Je remercie Tatiana Giraud d'avoir encadré ma thèse. Tatiana m'a laissé une grande autonomie dans mon travail tout en étant très à l'écoute, et elle m'a donné la possibilité de me former dans les domaines qui m'intéressaient et de travailler sur des sujets passionants. J'ai beaucoup appris en la voyant diriger des projets de recherche à la fois avec souplesse et tenacité. Tatiana, j'attends avec impatience de voir où tu méneras l'aventure *Microbotryum*.

Merci à Carole Smadja, dont la patience, la pédagogie et la passion communicative pour l'évolution et la science en général m'ont fortement inspirée pendant mon stage de M2, et encouragée à continuer en thèse. Je la remercie également, ainsi que Mathieu Joron et Xavier Vekemans, pour les très bons conseils et les encouragements qu'ils m'ont prodigués lors de mon comité de thèse.

Je remercie Jérôme Gouzy de m'avoir accueillie dans son équipe au LIPM, de m'avoir aidée à acquérir des principes de travail solides en bio-informatique, et de m'avoir donné de nombreuses pistes de progressions. Par-dessus tout, merci d'avoir cru en moi.

Je souhaite aussi remercier les membres de l'équipe : Alice et Emilie pour leurs encouragements et nos petites réunions bio-info, Antoine pour ses conseils avisés, Alodie qui nous permet de maniper dans les meilleures conditions possibles, et aussi Taia, Stéphanie, Jeanne, Sylvie, Eric et les autres pour le temps passé ensemble. Alice, bon courage pour la fin, tu vas y arriver ! Emilie, je suis impressionnée de voir tout ce que tu as accompli depuis que tu es arrivée au labo, et je suis sûre que ça va bien marcher pour toi par la suite. Sara, thank you for your support in the last months of my PhD. Merci à toutes les personnes de l'ESE et du bâtiment 360 que j'ai côtoyées pendant ces trois ans et demi, notamment Lucie, Elodie, Safia, Aurélien, Shanerin, Charlotte, Kader, Julie et tous les autres qui sont trop nombreux pour être cités. Une mention particulière pour Martine, Jacqui et Emmanuelle pour m'avoir facilité la vie au maximum dans les démarches administratives inévitables de la vie d'un doctorant. Merci aussi à Valentin pour m'avoir montré qu'un encadrant peut autant apprendre de son étudiant que l'inverse. Bon courage pour ta thèse.

Merci à toutes les membres de la plateforme de bio-informatique du LIPM, Jérôme, les deux Ludovic, Erica, Sophie, Sébastien, Emmanuel et les autres pour leur accueil chaleureux et les discussions geek.

Merci aussi à la communauté des phytopathologistes français et à ses animateurs, qui entretiennent un réseau d'échange informel que j'ai particulièrement apprécié, notamment via le groupe champignon du REID et les JJC à Aussois.

Si je me suis décidée pour l'évolution, c'est d'abord grâce aux enseignants-chercheurs de l'ENS de Lyon dont les cours passionants ont suscité mon intérêt pour cette discipline, en particulier Ludovic Orlando, Marie Semon et Jean-Nicolas Volff. Merci également pour l'excellente école de recherche de génomique comparative organisée par Céline Brochier-Armanet et Jean-Nicolas Volff qui a été pour moi une expérience aussi enthousiasmante que formatrice. L'année passée à préparer l'agrégation m'a permis de découvrir l'écologie. Je remercie les responsables de cette formation à l'ENS de Lyon qui ont à coeur de permettre aux étudiants d'acquérir une culture générale la plus large possible en biologie. Je remercie également les enseignants et enseignants-chercheurs de l'Université Paris-Sud qui m'ont permis de passer de l'autre côté de la barrière et de m'initier à mon tour à l'enseignement, notamment Florence Mougel, Florence Hulot et Hélène Vincent-Schneider.

Au-delà des individus avec qui j'ai interagi, j'ai eu le sentiment d'évoluer au sein d'une communauté qui, malgré les obstacles, tend vers un but commun, celui de faire avancer la science. C'est une des raisons pour laquelle j'ai envie de continuer à évoluer dans cet monde-là.

Je souhaiterais aussi particulièrement remercier ma famille. J'ai beaucoup partagé avec eux pendant cette thèse mes doutes et mes inquiétudes, et j'espère que ma présentation leur fera toucher du doigt ma passion et le bonheur que m'a apporté ce travail. Merci à mon père pour m'avoir soutenue quoiqu'il arrive, et pour m'avoir élevée en me transmettant l'envie d'apprendre, mais sans m'indiquer de direction pré-déterminée afin que je puisse trouver ma propre voie. Merci à ma mère de m'avoir transmis l'amour de la nature, le goût de la lecture et des petites choses du quotidien, et de m'avoir montré que l'envie de vivre était plus forte que tout. Merci à ma sœur Audrey, qui a toujours été là pour moi, et avec qui j'espère partager beaucoup de belles choses à l'avenir. Merci à mon frère Laurent, j'espère que nous aurons l'occasion de nous voir plus souvent. Merci à mon oncle et ma tante Marc et Evelyne pour leur accueil toujours chaleureux et leur soutien, à ma cousine Nathalie, ainsi qu'à tout le reste de ma famille.

Merci enfin aux RATON-LAVEUR et assimilés, en particulier Mathilde, Jonas, Elodie, Pierre, Laetitia, Mickaël, Ophélia, Benoît, Philippe, Benjamin, Irène, Tahina, Antoine, Marthe, Mikaël, Nathanaël, Quentin, Razvan, ainsi que tous les sporzeurs. Ces dernières années nous ont vu nous disperser un peu partout en France et dans le monde, mais les liens perdurent et j'espère que nous parviendrons encore à nous réunir de temps en temps. Merci à mes colocataires de l'Algoloc, Laure, Alex, Amaury, Mikaël et Pim, sans qui mes années en Ile de France n'auraient pas été aussi sympathiques. Enfin, Marthe, pour ton premier anniversaire de soutenance, je t'offre une place toute spéciale dans mes remerciements.

Table des matières

1

Introduction générale

1.1 L'adaptation chez les champignons

Un des buts majeurs de la biologie évolutive est de comprendre comment les êtres vivants s'adaptent à leur environnement. Comprendre l'adaptation est important dans le monde actuel pour prévoir dans quelle mesure les êtres vivants pourront s'adapter au changement climatique et à l'anthropisation croissante de l'environnement. Il est aussi important de comprendre dans quelle mesure et dans quelles conditions la divergence adaptative peut conduire à la formation de nouvelles espèces.

Depuis Darwin et la synthèse néo-darwinienne, on sait que les populations s'adaptent sous l'effet de la sélection naturelle. Il existe au sein des populations une variabilité phénotypique qui confère à différents individus une plus ou moins forte aptitude à la survie et à la reproduction. Si cette variabilité phénotypique repose sur une variabilité génétique, en d'autres termes, si elle est héritable, la fréquence des phénotypes les mieux adaptés augmente dans les générations suivantes. Autrement dit, les allèles responsables des phénotypes adaptatifs ont plus de chances d'être transmis à la descendance.

En pratique, il s'est avéré difficile d'identifier les gènes soumis à la sélection naturelle dans les populations. Il y a plusieurs raisons à cela. Il est souvent difficile de déterminer quels caractères sont responsables de l'adaptation à un environnement donné. Les phénotypes observables ne sont pas toujours adaptatifs, et les caractères responsables de l'adaptation peuvent être difficilement mesurables. De plus, la sélection est variable dans le temps et dans l'espace, et un

caractère sous sélection dans des conditions de laboratoire ne l'est pas forcément en conditions naturelles. Même quand des caractères soumis à la sélection naturelle sont identifiés, il n'est pas forcément simple d'identifier les gènes responsables de ce caractère. L'étude du polymorphisme au sein des populations et la comparaison de gènes entre espèces proches adaptées à des environnements différents peuvent néanmoins permettre d'identifier des gènes candidats dans l'adaptation, particulièrement dans les espèces bien étudiées d'un point de vue écologique.

Les scientifiques ne cherchent pas seulement à identifier les gènes responsables de l'adaptation, mais aussi à comprendre de manière plus générale les processus de l'adaptation, par exemple si l'adaptation implique quelques gènes à effets majeurs ou beaucoup de gènes à faibles effets, quelle est l'importance des interactions épistatiques et des effets additifs, ou quel est l'impact de l'existence de régions non-recombinantes sur l'adaptation. En effet, les régions non-recombinantes peuvent abriter des ensembles d'allèles co-adaptés qui sont ainsi protégés de la recombinaison qui casserait les combinaisons d'allèles favorables [1]. Inversement, la suppression de recombinaison peut empêcher une adaptation optimale. En effet, la recombinaison permet l'apparition de nouvelles combinaisons d'allèles, éventuellement bénéfiques, et favorise la purge des mutations délétères [2, 3]. Même les organismes à reproduction sexuée possèdent cependant souvent de larges régions non-recombinantes, souvent associées à la détermination du sexe génétique ou du type sexuel ; à cause de cette suppression de recombinaison, ces régions peuvent présenter des signes de dégénérescence, comme une accumulation de mutations délétères ou des pertes de gènes [4, 5, 6]. Etudier l'évolution des régions non-recombinantes est évidemment important pour comprendre l'évolution du déterminisme du sexe, mais c'est aussi crucial pour comprendre comment l'existence de ces régions impacte l'adaptation, en particulier à cause de la suppression de recombinaison qui devrait freiner l'adaptation. Pour comprendre l'évolution des régions non-recombinantes, et le lien avec les processus d'adaptation, il est crucial d'étudier une large gamme d'organismes et de modes de vie. Ceci est rendu de plus en plus facile par le développement de la génomique.

Les champignons constituent des modèles intéressants pour étudier l'adaptation car ils occupent une grande diversité de niches écologiques, et possèdent des cycles de vie et des systèmes de reproduction très variés. Ils forment le groupe frère des métazoaires au sein des opisthochontes. Certains champignons, comme la levure *Saccharomyces cerevisiae* ou *Neurospora crassa*, sont des espèces modèles en génétique. Les champignons possèdent en effet des avantages expérimentaux, comme des temps de génération courts et des phénotypes simples à mesurer. L'accès

direct à la phase haploïde, sous forme de produits de méiose alignés, chez certaines espèces comme *Neurospora crassa*, est particulièrement pratique pour étudier la génétique via la ségrégation de marqueurs. Le développement de la biologie moléculaire a révélé l'existence de nombreuses espèces cryptiques adaptées à des environnements différents, qui offrent une excellente opportunité d'étudier les processus de divergence adaptative et de spéciation. De plus, certains champignons possèdent des chromosomes de types sexuels qui incluent de larges régions non-recombinantes. Alors que les chromosomes sexuels des plantes et des animaux font l'objet de nombreuses études, les chromosomes de types sexuels des champignons n'ont été que peu étudiés jusqu'à présent. Ils offrent pourtant une très bonne opportunité d'élargir notre connaissance des régions non-recombinantes et de tester la généralité des théories sur l'évolution de ces régions.

Cette thèse s'intéresse donc à l'adaptation, à la dégénérescence, et à l'évolution des chromosomes portant les types sexuels chez les champignons du complexe d'espèces *Microbotryum violaceum*.

1.2 *Microbotryum violaceum*, un champignon modèle en écologie et évolution

1.2.1 Cycle de vie de *Microbotryum violaceum*

Microbotryum violaceum est un complexe d'espèces de champignons phytopathogènes biotrophes qui causent la maladie du charbon des anthères sur plusieurs centaines d'espèces de plantes de la famille des Caryophyllacées [7, 8]. La maladie du charbon des anthères est une maladie castratrice qui se traduit par le remplacement du pollen par des spores fongiques sur les anthères, conférant un aspect charbonneux (poudreux et sombre) aux fleurs infectées. Cette maladie modifie le développement floral : chez les espèces hermaphrodites, l'ovaire est avorté et les anthères portent les spores fongiques ; chez les espèces dioïques, l'ovaire des fleurs femelles est avorté, et des anthères se développent, portant également des spores [9] (figure 1.1).

Le cycle de vie de *M. violaceum* est haplo-dicaryotique et se déroule le plus souvent sur deux années (figure 1.2). Il n'infecte d'ailleurs que des plantes pérennes [8]. La première année, les téliospores diploïdes formées dans une plante malade sont dispersées par les pollinisateurs. La

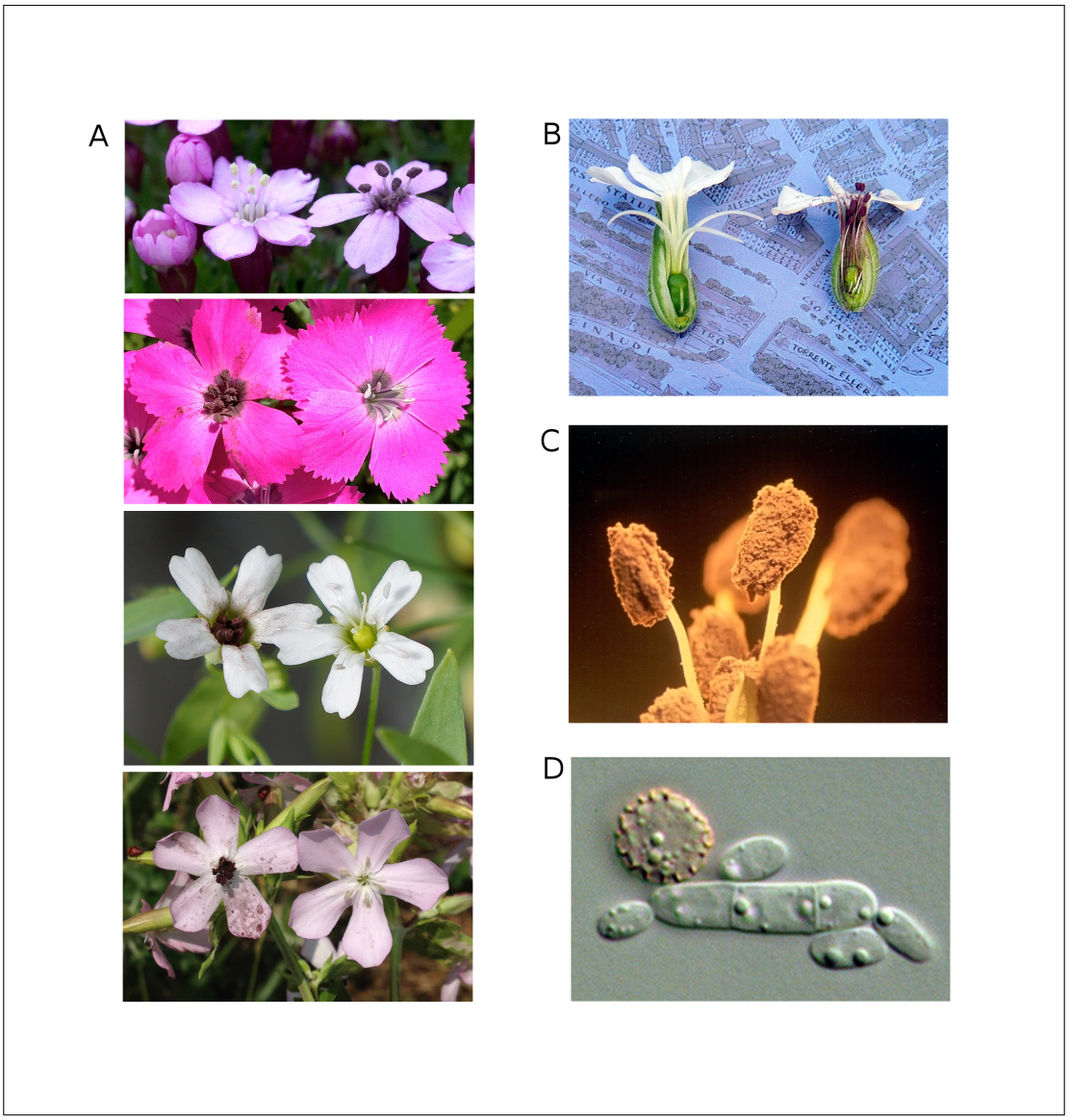


FIGURE 1.1 – Illustrations montrant A) des fleurs de Silènes saines et infectées par *Microbotryum violaceum*, avec de haut en bas *Silene acaulis*, *Dianthus pavonius*, *Atocion ruspestris* et *Saponaria officinalis*. B) Une coupe transversales de fleurs saine et infectée de *Silene latifolia* montrant l'ovaire avorté et les anthères portant des spores fongiques dans la fleur infectée. C) Un gros plan sur des anthères d'une fleur infectée. D) Un produit de germination d'une téliospore montrant une tétrade de spores alignées et quelques cellules issues de bourgeonnement. Crédit : Michael Hood, avec autorisation.

méiose a lieu à la surface des plantes sur lesquelles les téliospores sont déposées, ce qui produit une tétrade de sporidies alignées et ordonnées (figures 1.1 et 2.8). *Microbotryum violaceum* est un champignon hétérothallique, ce qui signifie que la conjugaison ne peut avoir lieu qu'entre spores de types sexuels différents. Il possède uniquement deux types sexuels, nommés a_1 et a_2. Au sein d'une tétrade, on trouve toujours deux sporidies de type a_1 d'un côté et deux sporidies de type a_2 de l'autre (figure 2.8), ce qui montre que le type sexuel ségrége en 1ère division de méiose.

La méiose peut être suivie d'une étape de multiplication des sporidies haploïdes par bourgeonnement. Cette phase est souvent brève, voire même inexistante car la conjugaison (syngamie) a le plus souvent lieu entre sporidies d'une même tétrade, donnant lieu à de forts taux d'autofécondation [10, 11]. Suite à la syngamie, un hyphe dicaryotique infectieux se forme, qui pénètre à l'intérieur de la plante à l'aide d'un appressorium [12]. Le champignon passe l'hiver dans la plante [12], et, au printemps suivant, il colonise les nouvelles hampes florales et produit de nouvelles téliospores, où la caryogamie a lieu.

1.2.2 *Microbotryum violaceum*, un modèle d'étude en génétique, évolution et écologie

Microbotryum violaceum est aussi un modèle en génétique, notamment pour l'étude des systèmes de reproduction [11]. C'est un des premiers champignons pour lequel on a décrit un système de reproduction hétérothallique [14]. Day and Garber ont étudié la ségrégation de nombreux mutants présentant des phénotypes auxotrophes, morphologiques ou de pigmentation [15] et établi des cartes de liaison génétique des mutations responsables. Ils ont aussi inféré l'implication d'éléments génétiques mobiles dans le cas de phénotypes de pigmentation instables [15]. *Microbotryum violaceum* est le premier champignon pour lequel on a démontré l'existence d'un dimorphisme entre les chromosomes qui portent les gènes déterminant la compatibilité sexuelle [16] - appelés chromosomes de types sexuels - de manière analogue à ce qui est observé chez les chromosomes sexuels [4, 17]. Au début de ma thèse, il existait une controverse sur la taille de la région non recombinante mais aussi sur sa nature plus ou moins dégénérée (cf. 2.1.3 p. 32). En bref, les travaux de Hood et Antonovics [10, 16, 17] suggéraient que ses chromosomes de types sexuels étaient non-recombinants sur la plus grande partie de leur longueur, avec la présence de mutations délétères à l'état haploïde liées au type sexuel, une faible densité en gènes

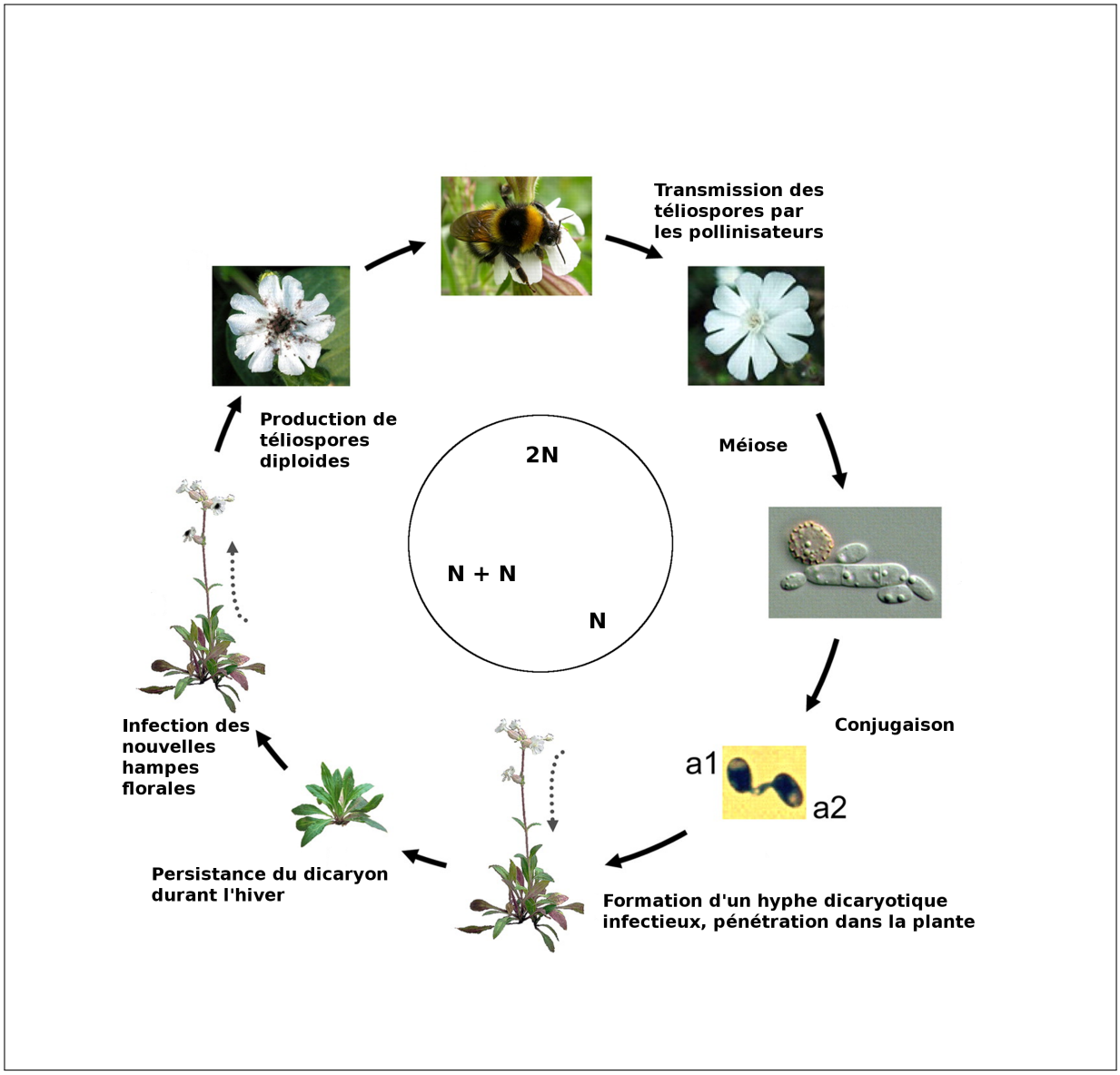


FIGURE 1.2 – Cycle de vie de *Microbotryum violaceum*. Adapté de [13].

et une accumulation d'éléments transposables. Au contraire, une étude de Votintseva et Filatov [18] affirmait que les chromosomes de types sexuels étaient non-recombinants sur seulement un quart de leur longueur, avec une faible différenciation entre types sexuels. Le premier groupe défendait donc l'hypothèse d'un processus de suppression de recombinaison ancien et avancé, le deuxième groupe celle d'une petite région non-recombinante d'origine récente.

Dans le domaine de la spéciation, *M. violaceum* est un cas bien étudié de complexe d'espèces jumelles chez les champignons. Les différentes espèces présentent une forte spécificité d'hôte [19, 20, 21] (figure 1.3), et une forte différenciation génétique [19, 22, 23] qui suggère que les flux génétiques sont limités dans la nature (cf. chapitre 4 p. 111). Les événements de spéciation ont eu lieu par sauts d'hôtes et non pas par cospéciations [24, 22] (figure 4.1). A l'intérieur du complexe d'espèces *M. violaceum*, les espèces ont été délimitées par le critère de congruence entre les différentes généalogies de gènes [20, 22]. Certaines espèces ont été formellement décrites et ont reçu des noms latins [25, 26], mais pas toutes. Dans la suite de ce manuscrit, *Microbotryum* désigne le genre d'espèces phytopathogènes auquel appartient *M. violaceum*, et *Microbotryum violaceum* désigne le complexe d'espèces qui cause le charbon des anthères chez les Caryophyllacées. Les deux espèces du complexe *M. violaceum* auxquelles je fais le plus référence sont *Microbotryum lychnidis-dioicae*, qui parasite le compagnon blanc *Silene latifolia*, et *Microbotryum silenes-dioicae*, qui parasite le compagnon rouge, *Silene dioica*.

Microbotryum violaceum est également un modèle pour l'étude des interactions hôte-pathogène [27, 28, 29, 30]. Dans le domaine de l'étude des maladies, *M. violaceum* est étudié pour son impact sur la distribution et la dynamique des populations de plantes hôtes [31, 32, 8], les dynamiques de méta-populations [33], les infections multiples et leur impact sur la virulence [13, 34]. Le système présente en effet l'intérêt de ne pas être directement impacté par l'homme, les plantes parasitées n'ayant pas d'intérêt économique, ce qui permet donc de pouvoir étudier l'épidémiologie d'une maladie naturelle et de pouvoir faire l'objet de manipulations sans risque pour l'alimentation ou la santé humaine.

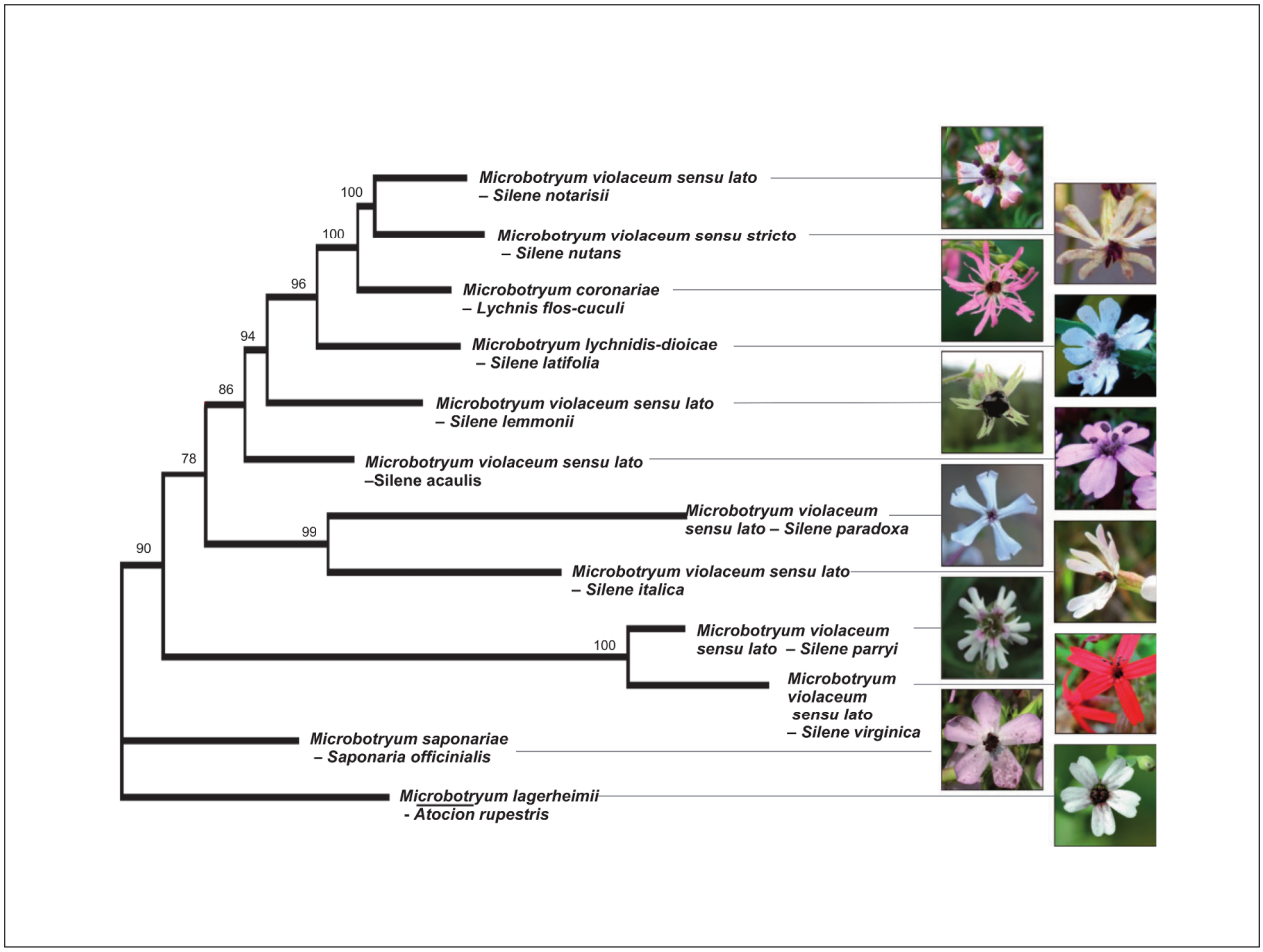


FIGURE 1.3 – Arbre phylogénétique de 12 espèces de *Microbotryum violaceum*. Le nom des espèces hôtes est indiqué ; adapté de [35].

1.3 Motivations de la thèse et plan du manuscrit

Ces dernières années, le développement des techniques de séquençage à haut débit a permis d'obtenir des données de génomique pour un nombre croissant d'organismes, permettant d'étudier de nouvelles questions en écologie et évolution [36]. Chez *M. violaceum*, la génération de ces données a déjà permis des progrès dans notre compréhension de l'évolution des chromosomes de types sexuels et des bases génétiques de l'adaptation du pathogène à son hôte. En effet, la génération de banques d'EST (expressed sequenced tags) a permis d'identifier les séquences des allèles codant pour les récepteurs à phéromones nécessaires à la syngamie [37] (figure 2.5), et de montrer que ceux-ci présentent un polymorphisme trans-spécifique très ancien, estimé à 370 millions d'années [38]. La génération d'une carte optique, c'est-à-dire une carte de restriction à l'échelle du chromosome, pour les chromosomes de types sexuels, a montré que la région non-recombinante couvrait 90% des chromosomes de types sexuels chez *M. lychnidis-dioicae* et semblait réarrangée ou fortement différenciée entre les deux types sexuels [39]. Des banques d'EST réalisées chez plusieurs espèces de *M. violaceum* ont aussi permis d'identifier des gènes sous sélection diversifiante entre différentes espèces de *M. violaceum*, et donc potentiellement impliqués dans l'adaptation à différents hôtes [40, 41] (cf. chapitre 4 p. 111).

Malgré ces avancées, nous ne disposions toujours pas au début de cette thèse d'un assemblage complet des chromosomes de types sexuels de *M. lychnidis-dioicae*, qui aurait permis de clore la controverse sur la taille de la région non-recombinante. D'autre part, les études ayant trait à l'adaptation à l'hôte s'appuyaient sur des données génomiques uniquement codantes et partielles, et n'utilisaient pas de données de polymorphisme (banques d'EST d'espèces différentes) [40, 41].

Dans le cadre de cette thèse, nous avons donc généré de nouvelles données génomiques dans le complexe *M. violaceum* afin de répondre à des questions ayant trait à l'étude des chromosomes de types sexuels et à la génomique de l'adaptation. Nous avons aussi étudié la dégénérescence dans le contexte de l'exposition aux radiations ionisantes dans la région de Tchernobyl. Ce manuscrit est donc structuré en trois parties. La première partie comporte une introduction sur les chromosomes de types sexuels, une analyse de la dégénérescence des chromosomes de types sexuels dans plusieurs espèces du complexe *M. violaceum* et une étude de l'assemblage complet de ces chromosomes chez *M. lychnidis-dioicae*. La deuxième partie comprend une étude de populations de *M. lychnidis-dioicae* de la région de Tchernobyl pour tester si les radiations

ionisantes provoquent une dégénérescence au niveau génomique et une baisse de valeur sélective des individus. Enfin, la troisième partie comprend un article de synthèse sur la génomique de l'adaptation chez les champignons, une étude du génome de *M. lychnidis-dioicae* qui se focalise sur la recherche de gènes impliqués dans les interactions avec l'hôte et le mode de vie biotrophe, et enfin une étude de génomique des populations qui recherche des traces de sélection au sein de deux espèces jumelles du complexe *M. violaceum*, *M. lychnidis-dioicae* et *M. silenes-dioicae*.

2

Etude des chromosomes de types sexuels chez *Microbotryum violaceum*

2.1 Introduction

Dans les génomes eucaryotes sexués, il existe certaines régions spécialisées qui ont cessé de recombiner. Les régions non-recombinantes les mieux étudiées se trouvent au sein des chromosomes sexuels d'animaux ou de plantes comme les chromosomes X/Y ou Z/W [4] et de certaines algues et bryophytes (type U/V) [42]. On trouve aussi des régions non-recombinantes autour des gènes de types sexuels chez les champignons [16, 43, 44]. D'autres exemples incluent les régions d'auto-incompatibilité chez les plantes [45], les régions portant les gènes de toxine - antidote chez certains champignons [46], mais aussi un "chromosome social" qui contrôle l'organisation sociale chez la fourmi *Solenopsis invicta* [47], ou encore un "supergène" qui contrôle les motifs des ailes chez les papillons *Heliconus* [48].

Une des caractéristiques intrigantes de ces régions, c'est que la suppression de recombinaison s'étend parfois bien au-delà des seuls gènes dont il paraît nécessaire d'empêcher la recombinaison. C'est chez les chromosomes sexuels que la suppression de recombinaison et ses causes évolutives ont été le mieux étudiées.

2.1.1 Les chromosomes sexuels

Chez les espèces où le déterminisme du sexe est génétique, les gènes-clés qui contrôlent le déterminisme du sexe génétique sont souvent portés par une paire de chromosomes qui présentent une forte différenciation cytologique, comme les chromosomes X et Y des mammifères. Ces chromosomes ont cessé de recombiner sur une partie de leur longueur, au point qu'il ne reste parfois que de courtes régions recombinantes à chaque extrémité, appelées régions pseudo-autosomales. C'est le cas par exemple des chromosomes X et Y d'*Homo sapiens*, chez lesquels l'âge de la suppression de recombinaison est estimé à 166 millions d'années [49]. Les régions non-recombinantes peuvent être très différenciées et fortement réarrangées, au point que leur homologie a été sujette à discussion. Elles présentent d'autres caractéristiques particulières, comme une plus faible densité en gènes que les autosomes et une accumulation d'éléments transposables, ainsi qu'une accumulation de mutations délétères et des pertes de gènes [50]. La dégénérescence observée est le résultat d'une baisse d'efficacité de la sélection naturelle en raison de la combinaison de plusieurs facteurs, notamment la baisse d'effectif efficace, l'état hétérozygote permanent et la suppression de recombinaison [5, 51, 6]. De plus, chez les systèmes de type X/Y ou Z/W où il existe un sexe homogamétique (X/X pour les femelles par exemple) et un sexe hétérogamétique (X/Y pour les mâles), la dégénérescence est plus rapide sur le chromosome du sexe hétérogamétique, car d'une part celui-ci a un effectif efficace plus faible que le chromosome de sexe homogamétique, et d'autre part les mutations délétères qu'il porte sont abritées car toujours à l'état hétérozygote (figure 2.1).

Chez les chromosomes sexuels, l'absence de recombinaison concerne des régions bien plus larges que la région de déterminisme du sexe et entraîne une dégénérescence importante. Pourtant, la suppression de recombinaison peut être très ancienne [49]. Comment expliquer l'existence de régions non-recombinantes et leur maintien sur des temps aussi longs ?

La théorie la plus largement acceptée, dites des gènes à effets sexuels antagonistes, mettrait en jeu le recrutement successif de gènes à effets antagonistes chez les mâles et les femelles. Les chromosomes sexuels proviendraient d'une paire ancestrale d'autosomes portant le ou les gènes-clés de déterminisme du sexe [4] (figure 2.2). La recombinaison cesserait d'abord entre les allèles des gènes de déterminisme du sexe afin d'empêcher une recombinaison intra-génique qui serait délétère. L'élargissement progressif de la région sans recombinaison aurait pour origine une pression de sélection pour lier successivement des gènes à effets antagonistes entre les sexes

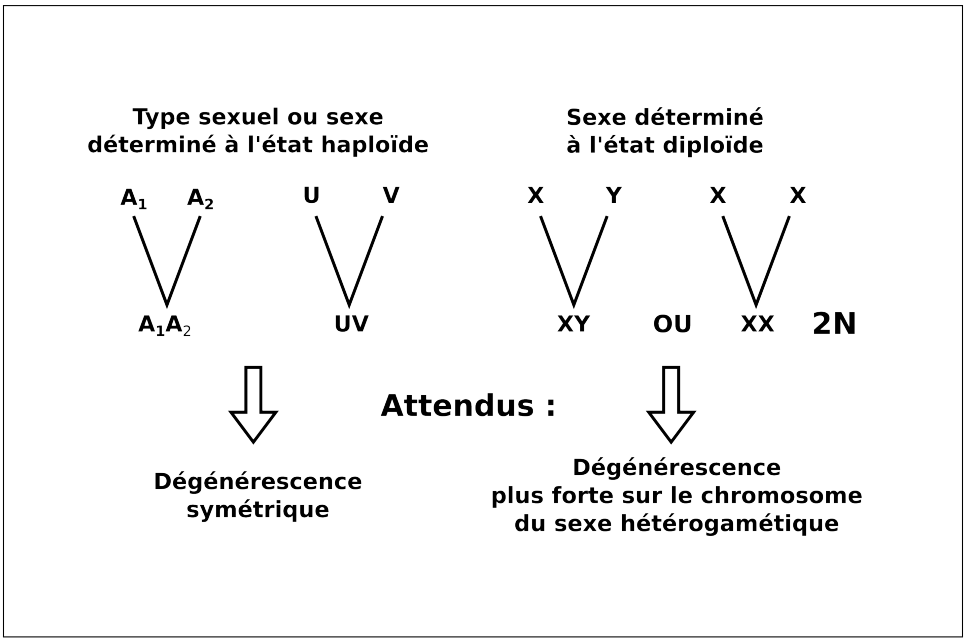


FIGURE 2.1 – Dégénérescence dans des allosomes d'organismes où le sexe/type sexuel est défini à l'état haploïde, par rapport aux organismes qui possèdent un sexe homogamétique et un sexe hétérogamétique.

aux gènes de déterminisme du sexe [6]. La notion d'épisodes successifs de suppression de recombinaison vient de l'observation qu'il existe une corrélation entre les niveaux de divergence entre les deux chromosomes sexuels et la distance aux gènes de déterminisme du sexe, la divergence étant une mesure de la date depuis la suppression de recombinaison (figure 2.3). D'un point de vue mécanistique, l'extension de la suppression de recombinaison pourrait être due à des réarrangements, permis ou facilités par l'accumulation d'éléments transposables.

La théorie des gènes antagonistes ne peut néanmoins pas expliquer la suppression de recombinaison dans les chromosomes de types sexuels des champignons, où la reproduction est essentiellement isogame [52], avec donc très peu de différences phénotypiques entre cellules de types sexuels différents. Il est difficile d'imaginer beaucoup de gènes qui auraient des effets antagonistes dans les différents types sexuels. Pourtant, certains chromosomes de types sexuels des champignons présentent des caractéristiques similaires à celles des chromosomes sexuels, comme une grande région de suppression de recombinaison, une plus faible densité en gènes, une accumulation d'éléments transposables et des réarrangements [53]. Certains auteurs avancent d'ailleurs que la théorie des gènes à effets antagonistes n'expliquerait pas tous les cas de chromosomes sexuels non-recombinants, car on a peu d'exemples de gènes liés aux gènes de déterminisme du sexe ayant des effets antagonistes établis entre les sexes, et que des théories plus générales suffisent à expliquer les données observées [54]. Parmi les hypothèses alternatives, on

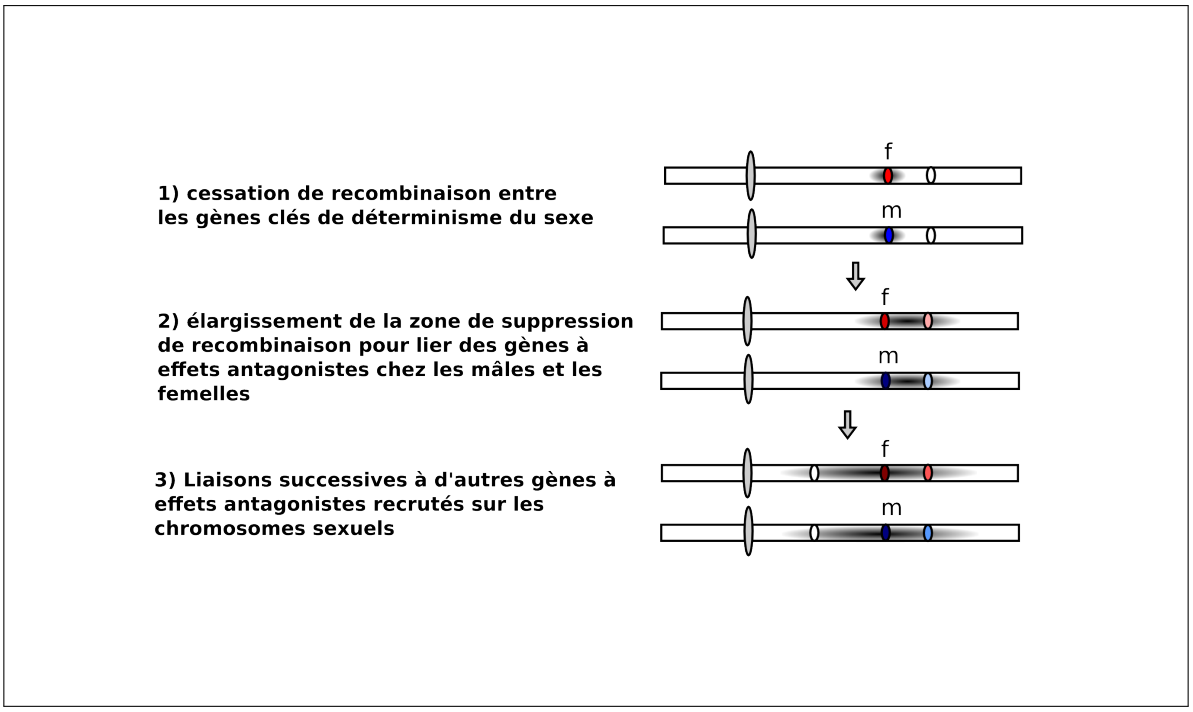


FIGURE 2.2 – La théorie des gènes à effets antagonistes

peut citer les suivantes :

- La suppression de recombinaison pourrait être sélectionnée pour éviter l'homozygotie de mutations délétères récessives, en particulier chez les espèces auto-fécondantes.

- La suppression de recombinaison pourrait permettre de lier des gènes ayant des interactions épistatiques positives, entraînant la formation de "supergènes". Cette théorie est en fait un cas plus général de la théorie des gènes à effets antagonistes.

- Les mutations ou réarrangements qui provoquent la suppression de recombinaison pourraient être fixés par dérive génétique.

Ces phénomènes pourraient se produire également sur des autosomes. Cependant, la différence essentielle entre les régions non-recombinantes autosomales et celles liées aux gènes de déterminisme du sexe serait leur durée de vie : les régions de déterminisme du sexe étant sous sélection équilibrante, les allèles sont maintenus sur des durées plus longues, et donc plus facilement observables.

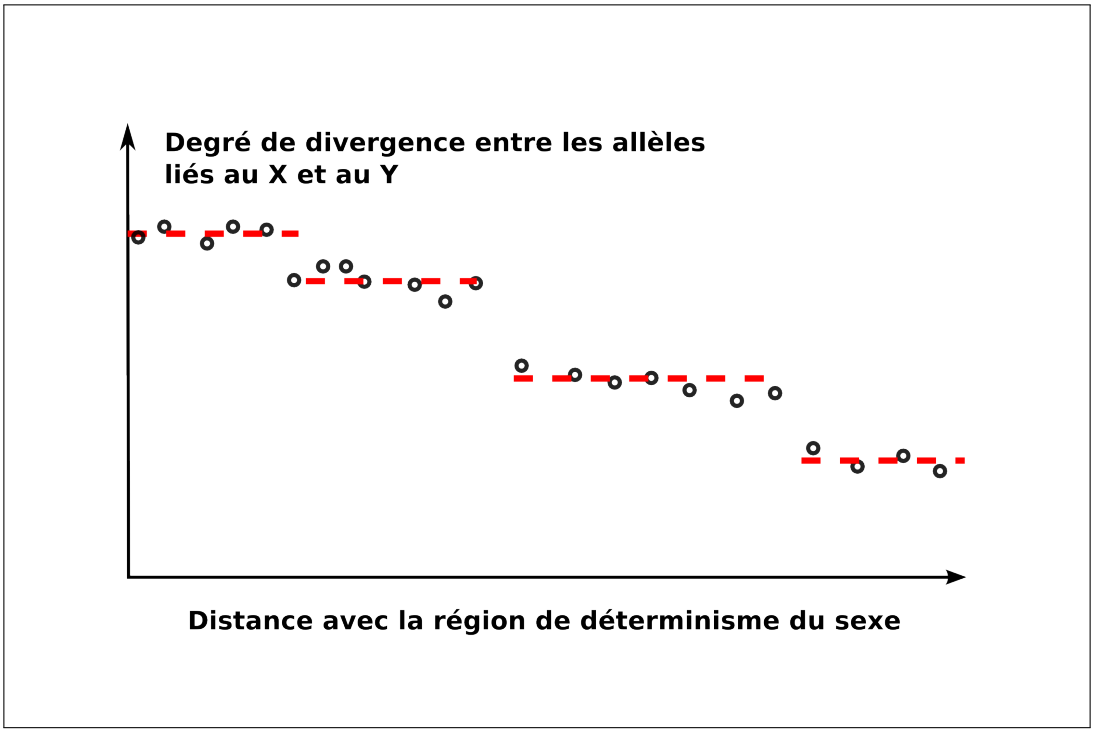


FIGURE 2.3 – Notion de strates de suppression de recombinaison. Dans cet exemple, on observe une corrélation négative entre la distance à la région de déterminisme du sexe et le niveau de divergence entre les allèles des gènes situés dans la région non-recombinante, avec plusieurs discontinuités le long du chromosome. La divergence entre allèles est une mesure de l'âge de la suppression de recombinaison. Ici, on peut inférer l'existence de quatre épisodes successifs de suppression de recombinaison. En pratique, la région de déterminisme du sexe n'est pas toujours connue et on mesure plutôt la divergence en fonction de la distance avec les régions pseudo-autosomales au niveau du chromosome du sexe homogamétique (X ou Z), qui conserve mieux la structure ancestrale.

2.1.2 Les chromosomes de types sexuels des champignons

Exemples et origine évolutive

Chez les champignons, les loci qui contrôlent le type sexuel, appelés loci MAT (pour "mating-type", type sexuel en anglais), peuvent être situés dans des régions qui ne recombinent pas, dont la taille varie de quelques milliers de paires de bases [55] à plusieurs Mpb [43, 44, 16, 39], et qui peuvent englober une majorité du chromosome [56]. Des chromosomes de types sexuels contenant de larges régions non-recombinantes ont été mis en évidence chez *M. lychnidis-dioicae* [16, 18, 39], *Neurospora tetrasperma* [56, 57], *Cryptococcus neoformans* [58, 59], *Ustilago hordei* [60] et *Podospora anserina* [61], entre autres. Les chromosomes de types sexuels présentent certaines caractéristiques similaires aux chromosomes sexuels, comme par exemple un dimorphisme [16] (figure 2.9), des réarrangements entre types sexuels [43], l'accumulation d'éléments transposables [43, 44], et des traces de dégénérescence sous forme d'accumulation de mutations non-synonymes [44].

Quelles forces pourraient être impliquées dans l'initiation et l'extension de la suppression de recombinaison dans les chromosomes de types sexuels des champignons ? On peut citer les suivantes, non mutuellement exclusives, sur lesquels je donnerai plus de détails dans la suite de cette introduction :

- Une pression de sélection pour lier le locus ou les loci MAT et le centromère
- Une pression de sélection pour lier les deux loci MAT déterminant le type sexuel chez certains basidiomycètes
- Une pression de sélection pour abriter des mutations récessives délétères en les liant au type sexuel
- Une pression de sélection pour lier au type sexuel des gènes impliqués dans le processus de syngamie, si les deux types sexuels jouent un rôle asymétrique

Liaison du locus MAT au centromère. Chez les champignons, la syngamie peut avoir lieu suivant les espèces entre spores provenant d'individus différents (allo-fécondation), entre spores provenant de produits de méiose d'une même individu mais issues de tétrades différentes (auto-fécondation classique), ou entre produits de méiose d'une même tétrade (au-fécondation

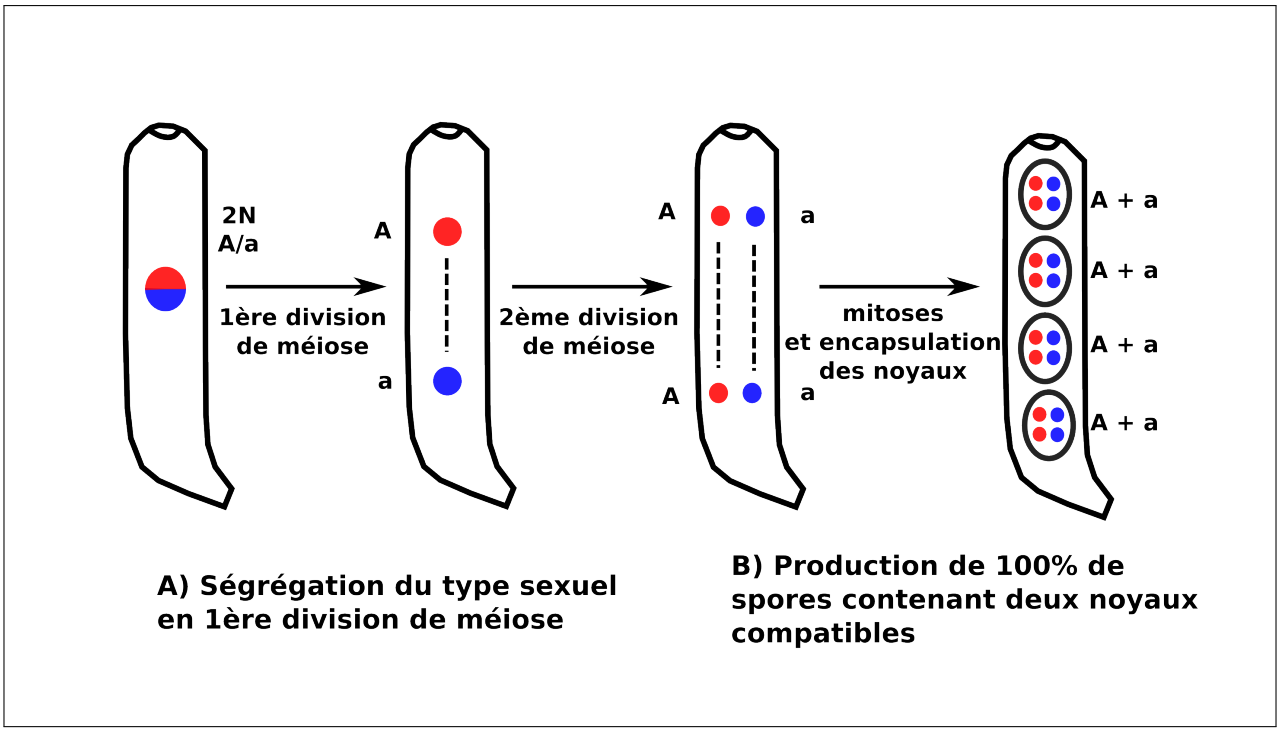


FIGURE 2.4 – Formation d'ascospores chez *Neurospora tetrasperma*. A) Le type sexuel ségrége en 1ère division de méiose. B) Cela permet la production de 100% de spores capables d'auto-fécondation, car contenant deux noyaux de types sexuels opposés. A et a représentent les allèles du locus de type sexuel.

intra-tétrade, ou automixie), voire même entre cellules issues d'un même génotype par multiplication clonale (fécondation intra-clone) [52]. Chez les champignons hétérothalliques, par définition, uniquement des cellules portant des allèles différents au locus ou aux loci MAT peuvent se croiser. Chez *Neurospora tetrasperma*, un champignon ascomycète hétérothallique qui pratique l'automixie, le locus MAT est lié au centromère. Cela permet d'avoir une ségrégation du type sexuel en 1ère division de méiose, et, via des mouvements de noyaux dans les ascospores, de produire 100% de spores contenant deux noyaux de types sexuels opposés et donc auto-compatibles (figure 2.4). Une pression de sélection pour lier le locus de type sexuel au centromère pourrait donc être à l'origine de la suppression de recombinaison chez ce champignon.

Fusion des deux loci MAT chez les basidiomycètes bipolaires. Alors que chez la plupart des champignons le type sexuel est contrôlé par un seul locus MAT, deux loci sont impliqués chez les basidiomycètes. Ce caractère est appelé tétrapolarité, car quatre types sexuels différents sont produits par une même méiose (en opposition avec les cas bipolaire où deux types sexuels sont produits par méiose). Les deux loci impliqués contrôlent respectivement la compatibilité

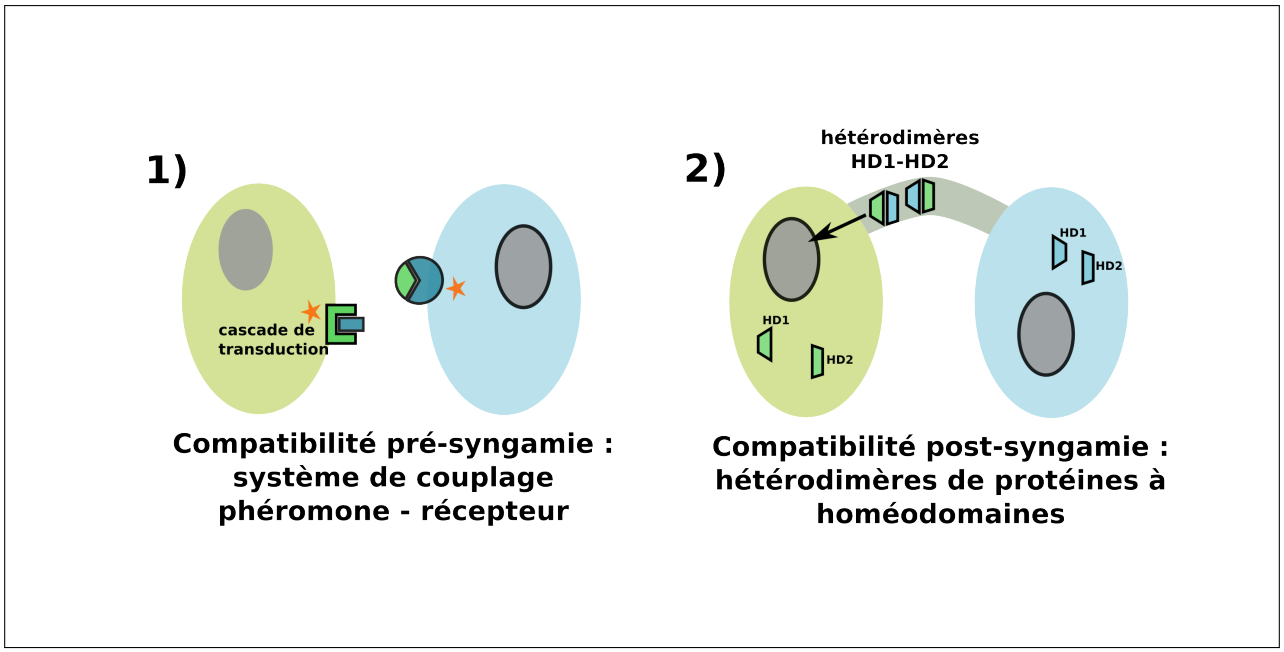


FIGURE 2.5 – Contrôle à deux loci du type sexuel chez les basidiomycètes. 1) La syngamie ne peut avoir lieu que s'il y a liaison entre le récepteur à phéromones et les phéromones compatibles. Cette liaison active une cascade de transcription qui provoque la croissance du tube de conjugaison et la transcription de gènes à homéodomaine. 2) Pour que la croissance post-syngamie puisse s'amorcer, il faut la formation d'hétérodimères de facteurs de transcription à homéodomaines de type sexuel opposés.

pré- et post-syngamie. Ils comprennent respectivement des gènes codant pour un récepteur à phéromone et des précurseurs de phéromones (PR), et pour des facteurs de transcription à homéodomaines (HD) (figure 2.5). Certains basidiomycètes sont néanmoins bipolaires, ce qui est interprété comme un retour à la bipolarité à partir d'un ancêtre tétrapolaire. Chez l'espèce bipolaire *Ustilago hordei*, par exemple, les loci HD et PR sont situés aux extrémités d'une région non-recombinante de 500 kb (figure 2.7), alors qu'ils ne sont pas liés chez l'espèce voisine tétrapolaire *Ustilago maydis*. La fusion des deux loci serait avantageuse en cas d'autofécondation (figure 2.6). Les transitions entre tétrapolarité et bipolarité pourraient donc jouer un rôle important dans l'initiation ou l'expansion de la suppression de recombinaison.

Liaison au locus MAT de mutations délétères récessives. La suppression de recombinaison pourrait résulter d'une sélection pour abriter des mutations délétères récessives partiellement liées au type sexuel car proche de la région non-recombinante [54]. Cette force pourrait être particulièrement importante chez les espèces qui pratiquent l'automixie [62]. Ce processus de liaisons successives de mutations délétères récessives aux loci de type sexuel [62] pourrait créer des strates évolutives.

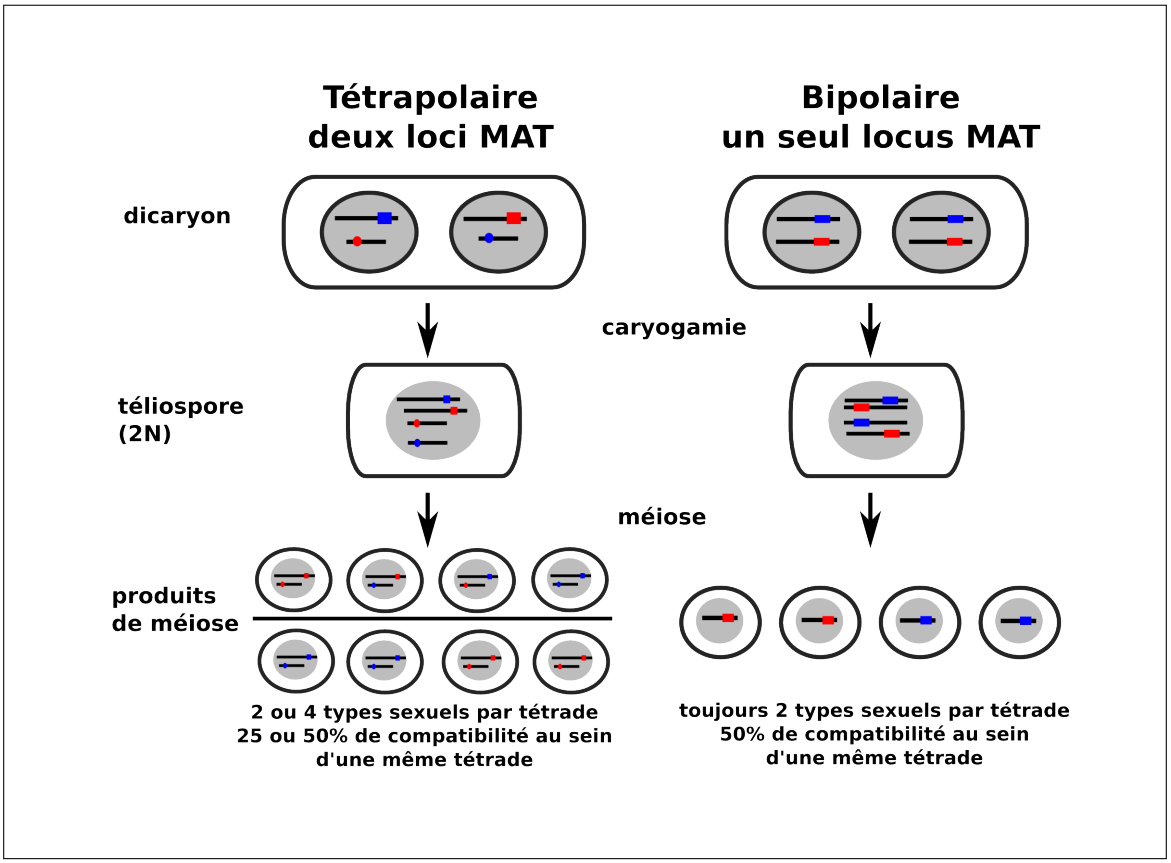


FIGURE 2.6 – Compatibilité entre spores d'une même tétrade dans le cas de système de reproduction bipolaires ou tétrapolaires. Dans le cas d'un système bipolaire, il y a toujours 50% de compatibilité entre spores d'une même tétrade.

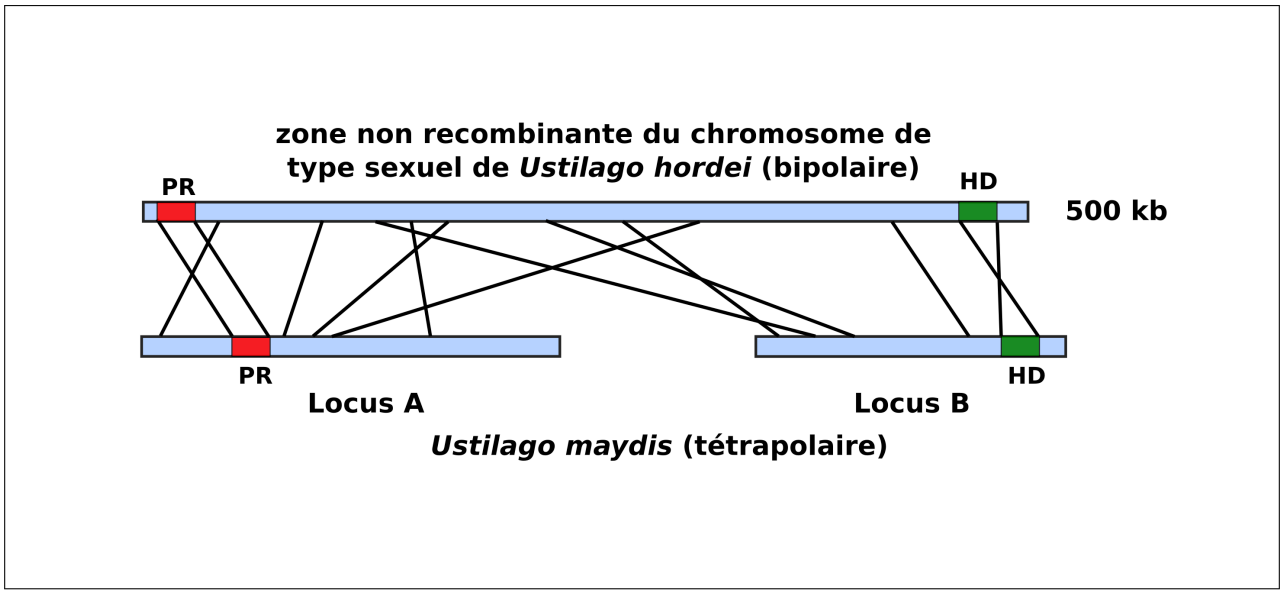


FIGURE 2.7 – Architecture de la zone non-recombinante et origine de la bipolarité chez *Ustilago hordei*. Les loci de phéromones/récepteur de phéromones et de facteurs de transcription à homéodomaine (PR et HD) sont situés à chaque extrémité de la zone non-recombinante, qui présente des réarrangements par rapport à l'espèce tétrapolaire *Ustilago maydis*. La suppression de recombinaison aurait pour origine une pression de sélection pour lier les deux loci de type sexuel, ce qui est avantageux pour une espèce auto-fécondante.

Asymétrie du processus de syngamie. Même s'il existe peu de différences phénotypiques entre types sexuels chez les champignons, les sporidies de types sexuels différents peuvent jouer un rôle asymétrique dans le processus de syngamie, par exemple pour la formation du tube de conjugaison ou la transmission de mitochondries [63, 64, 65]. Il pourrait donc y avoir une pression de sélection pour lier les gènes impliqués dans le processus de syngamie aux gènes de déterminisme du type sexuel.

Présence de strates de suppression de recombinaison

Afin de mieux comprendre les similitudes et différences entre chromosomes sexuels et chromosomes de types sexuels des champignons, et de reconstituer l'histoire de la suppression de recombinaison chez ces derniers, certains auteurs ont testé l'existence de strates de suppression de recombinaison. Chez *Neurospora tetrasperma*, on observe deux niveaux différents de divergence dans la région non-recombinante, de part et d'autre du centromère, qui ont été interprétés comme deux strates de suppression de recombinaison [56]. Chez *Cryptococcus neoformans*, le niveau de divergence est hétérogène au sein de la région non-recombinante, ce qui a aussi été interprété comme la signature d'anciennes strates réarrangées [66].

Cependant, ces niveaux hétérogènes de divergence entre allèles liés aux deux types sexuels pourraient aussi résulter de conversion génique. La conversion génique peut être définie comme le remplacement d'un segment d'ADN par un autre, entre une paire d'allèles ou de manière ectopique dans le cas de séquences répétées [67]. La conversion génique serait fréquente sur le chromosome Y d'*Homo sapiens* [68], et pourrait permettre de ralentir l'accumulation de mutations dans les génomes ou régions génomiques non-recombinants, car elle est souvent biaisée de AT vers GC au contraire de la mutation qui est plutôt biaisée de GC vers AT [69].

Il faut en effet noter que plusieurs difficultés existent pour tester l'existence de strates sur les chromosomes de types sexuels et plus généralement pour reconstituer l'histoire de la suppression de recombinaison chez les champignons. D'une part, les deux chromosomes sont hétérogamétiques et non-recombinants. Or chez les chromosomes sexuels, les strates sont observées en alignant les gènes selon l'ordre du chromosome qui recombine chez le sexe homogamétique et qui conserve donc plus longtemps la structure de l'autosome ancestral [70]. Lorsqu'on observe une hétérogénéité du niveau de divergence, mais sans corrélation entre la distance aux gènes MAT et le niveau de divergence, il est donc difficile de faire la différence entre des événements de conversion génique et des strates réarrangées. De plus, le manque de points de calibration fossile rend plus imprécis que dans d'autres modèles l'utilisation d'un modèle d'horloge moléculaire pour dater les événements de suppression de recombinaison [38].

2.1.3 Les chromosomes de types sexuels de *Microbotryum violaceum*

Découverte et caractéristiques

Microbotryum violaceum est un champignon basidiomycète, hétérothallique et bipolaire [14]. Il possède deux types sexuels, nommés a_1 et a_2. Les types sexuels ségrégent en 1ère division de méiose (figure 2.8). C'est le premier champignon pour lequel il a été montré un dimorphisme entre les chromosomes de types sexuels [16]. Les chromosomes des champignons sont trop petits pour permettre la réalisation de caryotypes classiques par microscopie optique, mais il est possible de les faire migrer par électrophorèse en champ pulsé pour les séparer selon leur taille. La comparaison des caryotypes de sporidies de type sexuel a_1 et a_2 de *Microbotryum lychnidis-dioicae*, l'espèce qui parasite *Silene latifolia*, a montré que les loci de type sexuel étaient portés par des chromosomes différenciés, qui présentaient un polymorphisme de taille de 600 kb [16] (figure 2.9). *Microbotryum lychnidis-dioicae* ayant un très fort taux d'automixie [11], on s'attend à un faible niveau d'hétérozygotie. Pour détecter la présence de zones non-recombinantes, Hood et Antonovics [71] ont étudié des marqueurs AFLP (amplified fragment length polymorphism) dans les tétrades. Ils ont trouvé que tous les marqueurs AFLP hétérozygotes de *Microbotryum lychnidis-dioicae* étaient liés au type sexuel. Ces marqueurs hétérozygotes représentaient 11% des marqueurs AFLP étudiés, alors que les chromosomes de types sexuels représentaient 13% de la longueur totale du génome. Ils en ont conclu que la région non-recombinante devait englober une très large majorité des chromosomes de types sexuels. Des caryotypes d'autres espèces du complexe *M. violaceum* ont également montré la présence de chromosomes de types sexuels dimorphes, mais la taille de ces chromosomes et l'importance du dimorphisme varient entre les espèces [71].

Il est possible d'isoler un chromosome particulier sur un caryotype par gel d'électrophorèse. Pour étudier les caractéristiques des chromosomes de types sexuels chez *M. lychnidis-dioicae* et les comparer à celle des chromosomes sexuels, Hood et Antonovics ont isolé chacun des chromosomes de types sexuels, et réalisé des banques d'ADNc pour chacun d'eux, ainsi que pour les autosomes [17]. Ils ont ensuite séquencé aléatoirement un certain nombre de fragments de chaque banque et caractérisé leur contenu en gènes et en éléments répétés. Les fragments issus des chromosomes de types sexuels étaient deux fois moins riches en gènes et quatre fois

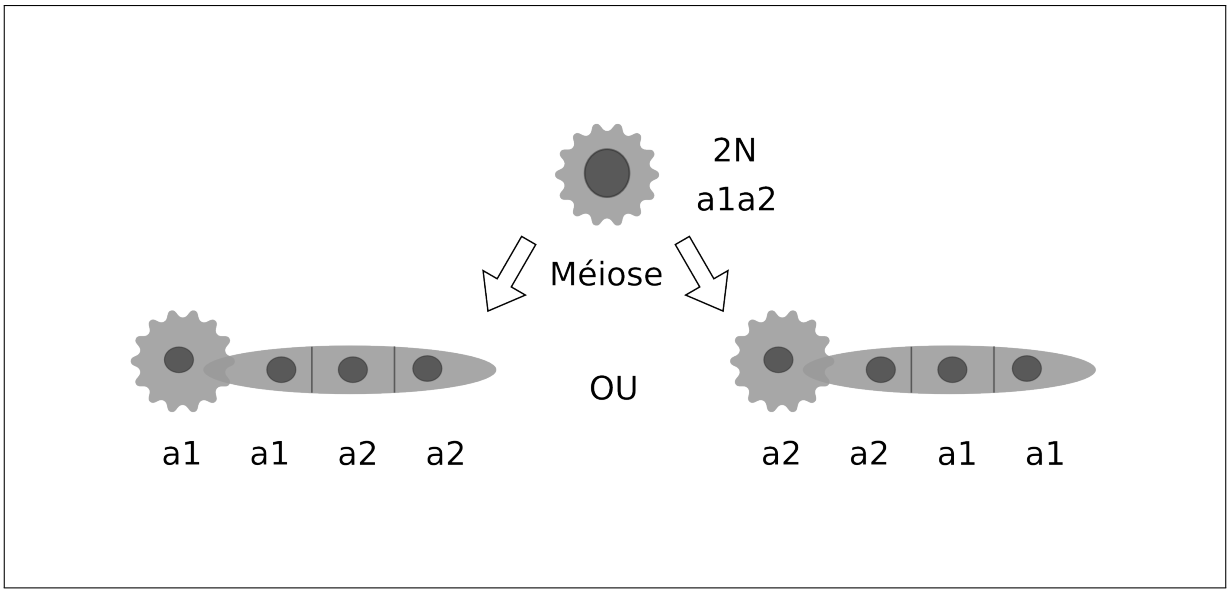


FIGURE 2.8 – Tétrades de spores alignées chez *Microbotryum violaceum*. Le champignon produit des téliospores diploïdes, dont la méiose produit des tétrades alignées. Le type sexuel ségrége en 1ère division de méiose, ce qui montre qu'il est lié à un centromère.

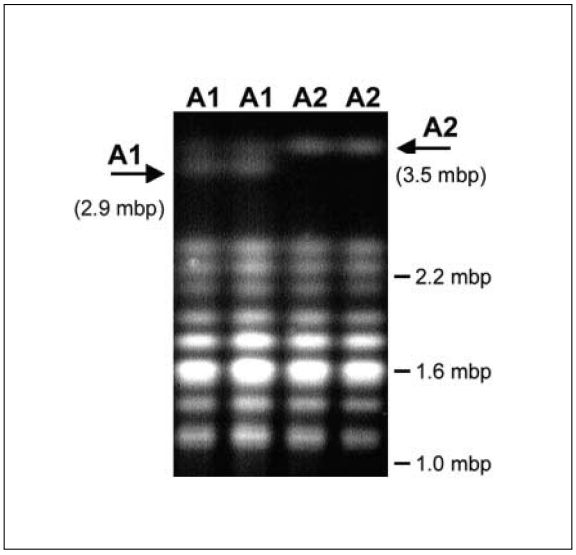


FIGURE 2.9 – Caryotype par électrophorèse sur gel en champ pulsé pour une tétrade de spores de *Microbotryum lychnidis-dioicae*, l'espèce qui parasite *Silene latifolia*. Les flèches indiquent les chromosomes de types sexuels.

plus riches en éléments transposables que les fragments autosomaux. Ces caractéristiques sont proches de celles des chromosomes sexuels de plantes et de mammifères, ce qui renforçait l'idée qu'une large proportion des chromosomes de types sexuels était non-recombinante.

Il existe de plus dans les populations naturelles des individus diploïdes qui présentent des biais de types sexuels quand ils sont cultivés sur milieu nutritif : la méiose se déroule comme habituellement sur milieu riche, mais seules des sporidies de type a_1 ou a_2 arrivent à pousser, alors que l'individu diploïde est nécessairement hétérozygote pour le type sexuel ; Hood et Antonovics ont étudié ces biais dans deux espèces du complexe *M. violaceum*, et montré qu'ils correspondaient à la présence de multiples allèles haplo-létaux liés au type sexuel [10], c'est-à-dire des allèles létaux à l'état haploïde, qui empêchent les sporidies d'un des deux types sexuels de pousser. Ces mutations seraient abritées dans la nature car la phase haploïde est très brève ou inexistante chez *M. violaceum* et ces allèles seraient toujours à l'état hétérozygote. Ces allèles haplo-létaux liés au type sexuel semblent donc être des signes de dégénérescence sur les chromosomes de types sexuels de *M. violaceum*.

Pour estimer l'âge de la suppression de recombinaison, on peut étudier la coalescence des gènes situés dans les régions non-recombinantes. En effet, ces régions sont soumises à une sélection équilibrante qui peut maintenir le polymorphisme sur des temps très longs, dépassant les temps de spéciation [38]. On s'attend donc à ce que les gènes situés dans les régions non-recombinantes présentent un polymorphisme trans-spécifique, c'est-à-dire que, dans la généalogie du gène, les allèles des différentes espèces se groupent par type sexuel plutôt que par espèce. Différentes études se sont donc intéressées aux généalogies de gènes liés au type sexuel. Une généalogie des récepteurs à phéromones chez *M. violaceum* et d'autres champignons [38] a montré que ce gène affiche un polymorphisme trans-spécifique très ancien, estimé à 370 millions d'années. D'autres gènes situés sur les chromosomes de types sexuels de *M. lychnidis-dioicae* ont des généalogies plus complexes au sein du complexe *M. violaceum* [72], avec un polymorphisme trans-spécifique au sein de certains clades mais pas à l'échelle du genre. Ces observations pourraient être expliquées par le fait que ces gènes ne sont pas liés au type sexuel dans toutes les espèces de *M. violaceum*, ou par des événements de conversion génique.

Cette première série d'étude suggérait donc fortement l'existence d'une région non-recombinante très étendue, au moins en partie ancienne, et présentant des signatures de dégénérescence.

En 2009, une étude a tenté d'évaluer la taille de la région non-recombinante chez *M. lychnidis-dioicae* et de tester l'hypothèse de strates évolutives. Pour cela, Votintseva et al ont utilisé les fragments d'ADNc de l'étude de Hood et Antonovics [17], testé leur liaison au type sexuel par ségrégation et mesuré la divergence des allèles entre les deux types sexuels [18]. Ils ont estimé que seuls 25% des marqueurs situés sur les chromosomes de types sexuels étaient liés au type sexuel, que les marqueurs liés au type sexuel étaient faiblement différenciés, et ils ont proposé l'existence d'au moins 3 strates distinctes de suppression de recombinaison. Cette étude affichait cependant deux faiblesses méthodologiques qui pourraient expliquer une telle différence avec les conclusions des travaux précédents. D'une part, les chromosomes de types sexuels isolés sur gel peuvent être dans une certaine mesure contaminés par des fragments autosomaux. Dans l'étude de Hood et Antonovics, cela n'aurait eu pour conséquence que de sous-estimer les différences entre autosomes et chromosomes de types sexuels et restait donc conservatif. Mais dans l'étude de Votintseva et collègues [18], cela a pu entraîner une sur-estimation du pourcentage de marqueurs homozygotes situés sur les chromosomes sexuels. De plus, 20% des marqueurs utilisés par Votintseva et collègues avaient une correspondance avec des éléments transposables et ne représentaient donc pas des marqueurs fiables [39].

En 2013, Hood et collègues [39] ont réalisé des cartes optiques des chromosomes de types sexuels a_1 et a_2. Les cartes optiques sont des cartes de restriction ordonnées à l'échelle des chromosomes [73], obtenues par digestion de larges fragments d'ADN linéarisés et immobilisés sur gel, visualisation des fragments par microscopie optique et construction de cartes consensus. L'ancrage de séquences d'ADN aux cartes optiques permet souvent d'améliorer l'assemblage des génomes [74]. La comparaison des cartes optiques des chromosomes de types sexuels a_1 et a_2 a montré que 90% des chromosomes de types sexuels étaient divergents, seules les extrémités des cartes optiques pouvant être alignées entre les deux chromosomes.

Quelles pourraient être les causes de la suppression de recombinaison entre les chromosomes de types sexuels de *M. lychnidis-dioicae* et d'autres espèces du complexe *M. violaceum* ? Les gènes codant pour les facteurs de transcription à homéodomaines (HD) ont été identifiés par similitude avec le locus MAT d'un groupe externe [75], et sont situés dans la région non-recombinante des chromosomes de types sexuels chez *M. lychnidis-dioicae*, tout comme le gène codant le récepteur à phéromones. La suppression de recombinaison aurait donc pu être initiée en réponse à une pression de sélection pour lier ces deux loci, ce qui est avantageux en cas d'auto-fécondation [52, 76]. La suppression de recombinaison a aussi pu être initiée ou étendue pour lier les loci

MAT au centromère, ce qui est avantageux en cas d'automixie qui est le mode de reproduction prépondérant chez plusieurs espèces de *M. violaceum* [52, 76]. On peut aussi imaginer une pression de sélection pour lier au type sexuel des mutations récessives délétères, et/ou des gènes avec des relations antagonistes dans les deux types sexuels (par exemple les gènes impliqués dans le processus de fusion des gamètes, ou dans la transmission des mitochondries). L'importance relative de ces différentes forces et l'ordre des événements sont pour l'instant inconnus (fusion des loci MAT puis liaison au centromère, ou l'inverse).

2.1.4 Questions et objectifs

Au début de cette thèse, les questions que nous nous posions étaient donc les suivantes :

- Quels sont la taille de la région de suppression de recombinaison et le niveau de divergence entre les chromosomes de types sexuels a_1 et a_2 ?

- Y a-t-il eu dégénérescence dans la région non-recombinante, et notamment accumulation de mutations délétères au sein des chromosomes de types sexuels ? Y a-t-il même eu des pertes des gènes ?

- Quels sont l'âge et l'histoire de la suppression de recombinaison ? Notamment, y a-t-il eu des épisodes successifs de suppression de la recombinaison, créant des "strates" de divergence, ou celle-ci a-t-elle eu lieu en un seul événement, par exemple pour lier les deux loci de type sexuel et le centromère ?

Pour répondre à ces questions, mes efforts se sont concentrés sur l'identification des régions non-recombinantes à l'aide de la comparaison des génomes haploïdes de types sexuels a_1 et a_2, l'obtention d'un assemblage complet des chromosomes de types sexuels de *M. lychnidis-dioicae*, ainsi que la comparaison des chromosomes de types sexuels de différentes espèces du complexe *M. violaceum*.

2.2 Article 1 : Degeneration of the nonrecombining regions in the mating-type chromosomes of the anther-smut fungi

La suppression de recombinaison entraîne une baisse de l'efficacité de la sélection naturelle, en raison notamment de la baisse de la taille de population efficace et du cliquet de Müller - l'accumulation irréversible de mutations délétères dans les régions sans recombinaison - ou la sélection liée (auto-stop de mutations délétères avec des mutations favorables, ou au contraire éliminations de mutations favorables liées à des mutations délétères [5, 51, 6]). La baisse d'efficacité de la sélection naturelle peut entraîner l'accumulation de mutations délétères pouvant aller jusqu'à des pertes de gènes, comme sur le chromosome Y d'*Homo sapiens*. La baisse d'efficacité de la sélection naturelle peut aussi mener à une accumulation d'éléments transposables qui favorisent les réarrangements et peuvent aussi être des mécanismes proximaux des pertes de gènes. Alors que la dégénérescence chez les chromosomes sexuels a été bien caractérisée [5, 51, 6], très peu d'études ont été menées sur les chromosomes de type sexuel des champignons (mais voir [44, 57]). D'une part, les chromosomes de types sexuels ont été découverts plus récemment que les chromosomes sexuels, et d'autre part certains auteurs pensaient que les chromosomes de types sexuels représentaient un stade peu avancé de suppression de recombinaison et étaient donc peu susceptibles d'avoir dégénéré [77, 6]. Cependant, certains chromosomes de types sexuels possèdent des zones non-recombinantes étendues et réarrangées [16, 39, 56, 78], et chez *Neurospora tetrasperma* par exemple [44], une accumulation de mutations délétères a été observée sous la forme de substitutions non-synonymes.

Dans l'étude suivante, nos objectifs étaient d'identifier les séquences des régions non-recombinantes des chromosomes de types sexuels a_1 et a_2, et de caractériser le degré de dégénérescence de ces régions dans différentes espèces du complexe *M. violaceum*. Nous nous attentions à trouver de la dégénérescence [71], et de manière non-asymétrique entre les types sexuels car il n'existe pas d'asymétrie d'hétérozygotie entre ceux-ci [79]. Nous disposions déjà du génome d'une souche haploïde de type a_1 de *M. lychnidis-dioicae*, mais les régions appartenant au chromosome de types sexuels n'y étaient pas identifiées à cause du caractère incomplet de l'assemblage. Nous avons isolé sur gel puis séquencé séparément les chromosomes de types sexuels a_1 et a_2 de la même souche diploïde, ce qui nous a permis d'identifier les régions appartenant aux chromosomes de

types sexuels. Un ancrage des scaffolds des régions pseudo-autosomales sur la carte optique des chromosomes de types sexuels (Hood 2013) a permis d'identifier les régions non-recombinantes et les régions pseudo-autosomales. Nous avons également séquencé les génomes d'une douzaine d'autres espèces du complexe *M. violaceum* à plus faible couverture, et aligné ces séquences au génome de référence. Nous avons ainsi recherché des signes de strates évolutives et de dégénérescence dans les régions non-recombinantes des chromosomes de types sexuels sous la forme d'accumulation de mutations non-synonymes, d'expressions différentes entre types sexuels, et d'accumulation d'éléments transposables.

Ma contribution a été la suivante : j'ai calculé la divergence entre les allèles des gènes situés sur les chromosomes de types sexuels a_1 et a_2, et analysé l'hétérogénéité de la divergence au sein des contigs des chromosomes de types sexuels pour tester la présence de strates de suppression de recombinaison. J'ai aussi analysé l'expression des gènes différentiellement exprimés entre cultures haploïdes de types sexuels a_1 et a_2. Les figures supplémentaires sont en annexe A.1 (page 235).

Degeneration of the Nonrecombining Regions in the Mating-Type Chromosomes of the Anther-Smut Fungi

Eric Fontanillas,[1,2] Michael E. Hood,[3] Hélène Badouin,[1,2] Elsa Petit,[1,2,3] Valérie Barbe,[4] Jérôme Gouzy,[5,6] Damien M. de Vienne,[7,8,9,10] Gabriela Aguileta,[9,10] Julie Poulain,[11] Patrick Wincker,[4,11] Zehua Chen,[12] Su San Toh,[13] Christina A. Cuomo,[12] Michael H. Perlin,[13] Pierre Gladieux,[1,2] and Tatiana Giraud*[,1,2]

[1]Ecologie, Systématique et Evolution, Bâtiment 360, Université Paris-Sud, Orsay, France

[2]CNRS, Orsay, France

[3]Department of Biology, Amherst College

[4]Commissariat à l'Energie Atomique (CEA), Institut de Génomique (IG), Genoscope, Evry, France

[5]INRA, Laboratoire des Interactions Plantes-Microorganismes (LIPM), UMR441, Castanet-Tolosan, France

[6]CNRS, Laboratoire des Interactions Plantes-Microorganismes (LIPM), UMR2594, Castanet-Tolosan, France

[7]Laboratoire de Biométrie et Biologie Evolutive, Centre National de la Recherche Scientifique, Unité Mixte de Recherche 5558, Université Lyon 1, Villeurbanne, France

[8]Université de Lyon, Lyon, France

[9]Bioinformatics and Genomics Programme, Centre for Genomic Regulation (CRG), Dr. Aiguader 88, Barcelona, Spain

[10]Universitat Pompeu Fabra (UPF), Barcelona, Spain

[11]CNRS UMR 8030, Evry, France

[12]Broad Institute of MIT and Harvard, Cambridge, MA

[13]Department of Biology, Program on Disease Evolution, University of Louisville

***Corresponding author:** E-mail: tatiana.giraud@u-psud.fr.

Associate editor: Jianzhi Zhang

Abstract

Dimorphic mating-type chromosomes in fungi are excellent models for understanding the genomic consequences of recombination suppression. Their suppressed recombination and reduced effective population size are expected to limit the efficacy of natural selection, leading to genomic degeneration. Our aim was to identify the sequences of the mating-type chromosomes (a_1 and a_2) of the anther-smut fungi and to investigate degeneration in their nonrecombining regions. We used the haploid a_1 *Microbotryum lychnidis-dioicae* reference genome sequence. The a_1 and a_2 mating-type chromosomes were both isolated electrophoretically and sequenced. Integration with restriction-digest optical maps identified regions of recombination and nonrecombination in the mating-type chromosomes. Genome sequence data were also obtained for 12 other *Microbotryum* species. We found strong evidence of degeneration across the genus in the nonrecombining regions of the mating-type chromosomes, with significantly higher rates of nonsynonymous substitution (*dN*/*dS*) than in nonmating-type chromosomes or in recombining regions of the mating-type chromosomes. The nonrecombining regions of the mating-type chromosomes also showed high transposable element content, weak gene expression, and gene losses. The levels of degeneration did not differ between the a_1 and a_2 mating-type chromosomes, consistent with the lack of homogametic/heterogametic asymmetry between them, and contrasting with X/Y or Z/W sex chromosomes.

Key words: **Y chromosome, *Silene latifolia*, *Microbotryum violaceum*, PAR, evolutionary strata, autosomes, allosomes, genetic map.**

Article

Introduction

Loci determining which genotypes can mate sometimes reside within genomic regions that display exceptional characteristics, with extensive suppression of homologous recombination and cytological differentiation between members of the diploid chromosome pair (e.g., in sex chromosomes; Charlesworth 1991; Marais et al. 2008; Bergero and Charlesworth 2009; Bachtrog et al. 2011). Despite being fundamental to sexual reproduction, the suppression of recombination renders these genomic regions particularly prone to degeneration (Bachtrog 2005). Degenerative changes in the coding sequences (CDS) of sex chromosomes include higher nonsynonymous substitution rates and/or lower expression levels than in recombining regions (e.g., Filatov and Charlesworth 2002; Bartolome and Charlesworth 2006). Degeneration may even extend to the point of gene loss, illustrated for example by the small size and limited gene content of the human Y chromosome (Repping

2006). Nonrecombining regions also tend to accumulate transposable elements (TEs) whose insertions can cause deleterious mutations (e.g., Steinemann M and Steinemann S 1992; Erlandsson et al. 2000; Bachtrog 2003; Marais et al. 2008).

In plants and animals in which one sex is heterogametic (e.g., XY males in systems with XX females), degeneration of the sex chromosomes is asymmetric (Bachtrog 2013). Deleterious recessive mutations on the Y are sheltered, whereas mutations on the X can be exposed to selection when in the homogametic sex. The efficacy of selection on sex chromosomes is further limited because of their reduced effective population size, exacerbated by the process of hitchhiking by deleterious mutations, for example, linked to positively selected Y-linked mutations (Bergero and Charlesworth 2009; Bachtrog 2013). As with the sheltering of mutations, these population-level effects influence the Y chromosome more severely than the X.

Some fungal mating-type chromosomes can also display recombination suppression and size dimorphism analogous to sex chromosomes (Fraser et al. 2004; Menkis et al. 2008; Hood et al. 2013; Grognet et al. 2014). Mating-type chromosomes carry the genes regulating mating compatibility (e.g., through premating pheromones and receptors and postmating homeodomain proteins) but not those determining male/female functions (Billiard et al. 2011). Mating occurs in the haploid stage in fungi, and consequently their mating-type chromosomes are always in a heterogametic condition in diploids. Mating-type chromosomes are therefore expected, like sex chromosomes, to display phenomena of sheltering of deleterious mutations and smaller effective population size than nonmating-type (non-MAT) chromosomes, which may lead to degeneration. However, in contrast to sex chromosomes, the sheltering and reduced effective population size effects are expected to be symmetrical in the mating-type chromosome pair (Bull 1978).

The evolutionary consequences of suppressed recombination on fungal mating-type chromosomes have not been extensively studied. Nevertheless, footprints of degeneration have been reported in the form of high nonsynonymous substitution rates and degeneration of codon usage in the fungus *Neurospora tetrasperma* (Ellison et al. 2011; Whittle and Johannesson 2011; Whittle et al. 2011). Also, the accumulation of TEs has been shown in the mating-type chromosomes of *N. tetrasperma* and *Microbotryum lychnidis-dioicae* (Hood et al. 2004; Ellison et al. 2011). Further work is needed to assess whether reduced expression or loss of genes also occurs, and whether degeneration is similar in the two mating-type chromosomes. In the case of the *M. lychnidis-dioicae* mating-type chromosomes, opposite conclusions have been reached about the extent of suppressed recombination and whether evolutionary strata exist. Evolutionary strata are distinct regions of different levels of divergence between sex chromosomes, probably reflecting successive stages in the expansion of the region with suppressed recombination (Lahn and Page 1999; Bergero et al. 2007). Some results suggest a small region with distinct strata in *M. lychnidis-dioicae* (Votintseva and Filatov 2009), whereas others suggest that much of the chromosome pair is nonrecombining (Hood et al. 2013).

Microbotryum lychnidis-dioicae is a fungus causing anther-smut disease on *Silene latifolia*. It was the first fungus in which dimorphic fungal mating-type chromosomes (a_1, ~3.3 Mb, and a_2, ~4.0 Mb) were described (Hood 2002; Hood et al. 2013). The suppression of recombination between mating-type chromosomes is thought to be favored under a highly selfing mating system as occurs in this fungus (supplementary fig. S1, Supplementary Material online) (Antonovics and Abrams 2004; Hood and Antonovics 2004; Johnson et al. 2005; Giraud et al. 2008). Suppression of recombination linking the centromere, premating pheromones/receptors, and postmating homeodomain proteins renders a higher rate of gamete compatibility under intratetrad selfing (Hood and Antonovics 2000; Nieuwenhuis et al. 2013). Segregation analyses of progeny (Hood and Antonovics 2004) and the lack of colinearity found by analysis of optical maps, that is, ordered, chromosome-wide restriction maps (Hood et al. 2013), indicate that nonrecombining regions (NRRs) make up most of the *M. lychnidis-dioicae* mating-type chromosomes (up to 90%). In contrast, Votintseva and Filatov (2009) estimated that only 25% of the markers from the mating-type chromosomes segregated with the mating type. The NRRs of the mating-type chromosomes in *M. lychnidis-dioicae* are flanked at either end by recombining regions (MATRRs), as shown by segregation analyses of progeny (Votintseva and Filatov 2009) and by colinearity in the optical maps (Hood et al. 2013). Because these distal regions retain the ability to recombine, the transmission and population genetics of these chromosomes are expected to be similar to non-MAT chromosomes (Ellis et al. 1990).

Very ancient trans-specific polymorphism has been observed at the mating-pheromone receptor gene (370 My old; Devier et al. 2009) and, to a lesser extent, at other genes, not involved in mating but linked to the mating-type genes (Abbate and Hood 2010a; Petit et al. 2012). Such ancient trans-specific polymorphism indicates that cessation of recombination began before divergence of the *Microbotryum* anther-smut fungi. Numerous *Microbotryum* species have dimorphic mating-type chromosomes (Perlin 1996; Hood and Antonovics 2004), thus providing a unique opportunity to study the dynamics, roles, and patterns of degeneration in the NRRs in fungal mating-type chromosomes. However, the sequence of the a_1 genome of the reference strain of *M. lychnidis-dioicae*, available at the Broad Institute, was insufficiently assembled to identify the a_1 mating-type chromosome. Furthermore, no reference a_2 genome was available. The aims of this work were therefore: 1) To identify the sequences of the a_1 and a_2 mating-type chromosomes and of their NRRs in *M. lychnidis-dioicae*, and 2) to investigate TE accumulation and degeneration in the NRRs of the mating-type chromosomes in the *Microbotryum* genus, including increased nonsynonymous substitution rates (d*N*/d*S*), reduced expression, and gene loss.

Results

Identification of Mating-Type Chromosomes, MATRRs, and NRRs in *M. lychnidis-dioicae*

We used the reference specimen whose a_1 haploid genome was sequenced by the Broad Institute. Both a_1 and a_2

mating-type chromosomes from this reference specimen were isolated on pulsed-field electrophoretic gels and sequenced at very high coverage (1,175×). Mapping the a_1 chromosome-specific DNA against the a_1 reference assembly available at the Broad Institute allowed the identification of scaffolds corresponding to the a_1 mating-type chromosome. We conducted de novo assembly of the sequences of the karyotype-isolated a_2 mating-type chromosome.

The Broad Institute Supercontigs 37 and 43 could be anchored on previously established optical maps of the *M. lychnidis-dioicae* mating-type chromosomes (Hood et al. 2013). Each of these two scaffolds anchored on one of the two distal regions of structural colinearity between the a_1 and a_2 mating-type chromosomes (Hood et al. 2013). This allowed the sequences of the mating-type chromosome recombining regions (MATRRs) to be identified. As expected for recombining DNA in a selfing species, the sequences assigned to MATRRs showed less than 0.1% nucleotide divergence between the a_1 and a_2 chromosomes sequenced.

Because there were numerous repetitive sequences, the remaining scaffolds assigned to the mating-type chromosome NRRs could not be satisfactorily anchored on the optical maps. Nevertheless, their assignment to the mating-type chromosomes appeared reliable because, after filtering out repetitive elements: 1) These scaffolds were highly represented in the reads from the karyotype-isolated mating-type chromosomes, and 2) significant BLASTn hits were found between the unique genes from the a_1 and a_2 NRR scaffolds, with substantial divergence between most alleles, consistent with expectations for NRRs. Some unique genes within scaffolds assigned to NRRs appeared highly similar between a_1 and a_2. However, these genes were embedded within scaffolds that carried other genes with more divergence between the two mating types (supplementary fig. S2, Supplementary Material online). Their assignment to NRRs therefore appears to be sound. Existing genetic maps were used for further validation of our scaffold assignments to NRRs and MATRRs. All 11 loci previously shown to be completely linked to the mating-type in segregation analyses of a single progeny (Abbate and Hood 2010b; Petit et al. 2012) were found on the scaffolds we assigned to NRRs. The sequences we assigned to NRRs included 11 additional loci shown to be linked to the mating type in another segregation analysis (Votintseva and Filatov 2009). One of the NRR loci (accession BZ782397 in Votintseva and Filatov 2009) corresponded to one of our NRR fragments with zero a_1–a_2 sequence divergence. Thus, even nondiverged regions may belong to the NRR. Five mating-type chromosome loci previously shown to be unlinked to mating type and with zero sequence divergence between a_1 and a_2 (Votintseva and Filatov 2009) were found in scaffolds we assigned to MATRR contigs (accessions BZ782138, BZ782443, BZ782547, BZ782204, and BZ782044).

In the a_1 *M. lychnidis-dioicae* reference genome, 6,891 genes were predicted in non-MAT chromosomes, 349 genes in the NRRs, and 99 genes in the MATRRs. The NRRs thus carry 78% of the genes of the mating-type chromosome. Genes typically determining mating types in basidiomycetes were found in the NRRs: 1) The genes encoding the mating pheromones and receptors, which determine conjugation compatibility, and 2) the genes encoding mating homeodomains regulating growth of the dikaryon after conjugation. Other than genes previously identified as belonging to the NRRs and involved in mating (Petit et al. 2012), no other genes in the NRRs had any putative mating function.

Substantial Divergence between NRRs and Investigation of the Possibility of Evolutionary Strata in *M. lychnidis-dioicae*

Using our de novo assembly of the a_2 mating-type chromosome, we identified the reciprocal best hits between predicted NRR genes on the a_1 and a_2 sequences. We estimated the divergence between these alleles at synonymous coding positions, using alignments of concatenated exons per gene longer than 1,000 bp to avoid stochastic effects due to short sequences (fig. 1). Synonymous divergence was substantial, indicating that the suppression of recombination between the mating-type chromosomes of *M. lychnidis-dioicae* was very ancient, with some extreme values and divergence varying from 0 to 3.4, and a mean of 0.06. There were no marked gaps in the distribution of divergence values (fig. 1), as would be expected if different evolutionary strata had ceased recombining at different times. Because the assembly was incomplete, however, we could not plot the divergence levels against physical distance, which would be the most powerful approach to test for the existence of evolutionary strata. Nevertheless, the large scaffolds included genes displaying high variation in levels of a_1–a_2 divergence (supplementary fig. S2, Supplementary Material online), indicating that heterogeneity in divergence occurs at small scales of physical distance.

We then assessed when recombination suppression between the mating-type chromosomes of *M. lychnidis-dioicae* emerged relative to speciation events. We built trees of the predicted genes assigned to NRRs. Trees were built from alignments of putative orthologous groups (POGs) from the whole haploid genome sequences of either mating type (obtained by 454 technology) from 11 other *Microbotryum* species, as well as the a_1 and a_2 alleles of *M. lychnidis-dioicae*. In principle, recent recombination events would place the a_1 and a_2 alleles of *M. lychnidis-dioicae* close together relative to alleles of other species, whereas suppression of recombination established before speciation events will place *M. lychnidis-dioicae* a_2 alleles closer to a_2 alleles of other species than to *M. lychnidis-dioicae* a_1 alleles (Devier et al. 2009; Petit et al. 2012). Among the 349 predicted NRR genes, only 45 reliable alignments could be obtained that included the alleles of both a_1 and a_2 of *M. lychnidis-dioicae* and of at least five other species. In 19 of the 45 trees built with confidently aligned NRR genes, the a_2 alleles of *M. lychnidis-dioicae* clustered with a_2 alleles of other species: *Microbotryum* from *S. notarsii*, from *L. flos-cuculi*, *S. nutans* and/or S. *lemmoni* (fig. 2). Such trans-specific polymorphism indicates that recombination was suppressed before the corresponding speciation events. The other 26 trees showed clustering of *M. lychnidis-dioicae* a_1

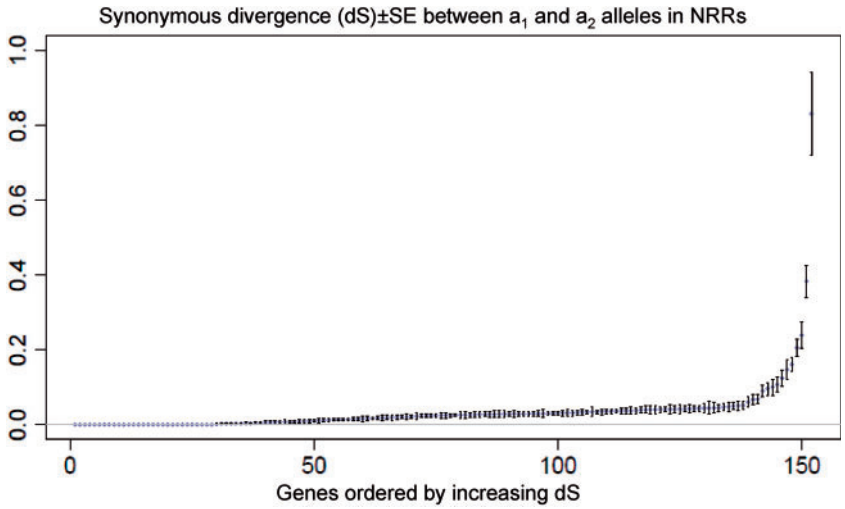


Fig. 1. Synonymous divergence between a_1 and a_2 alleles in predicted genes in the NRRs of the mating-type chromosomes of *Microbotryum lychnidis-dioicae*, ranked by increasing divergence. Only trimmed alignments of concatenated exons longer than 1,000 bp were retained. Synonymous divergence was estimated as the numbers of synonymous substitutions (dS), divided by the number of synonymous sites in the sequence; standard errors are shown. A single dS value at 3.4 is not shown for optimal visualization of the other values. The divergence between the pheromone receptor alleles is not shown either as they could not be aligned in nucleotides.

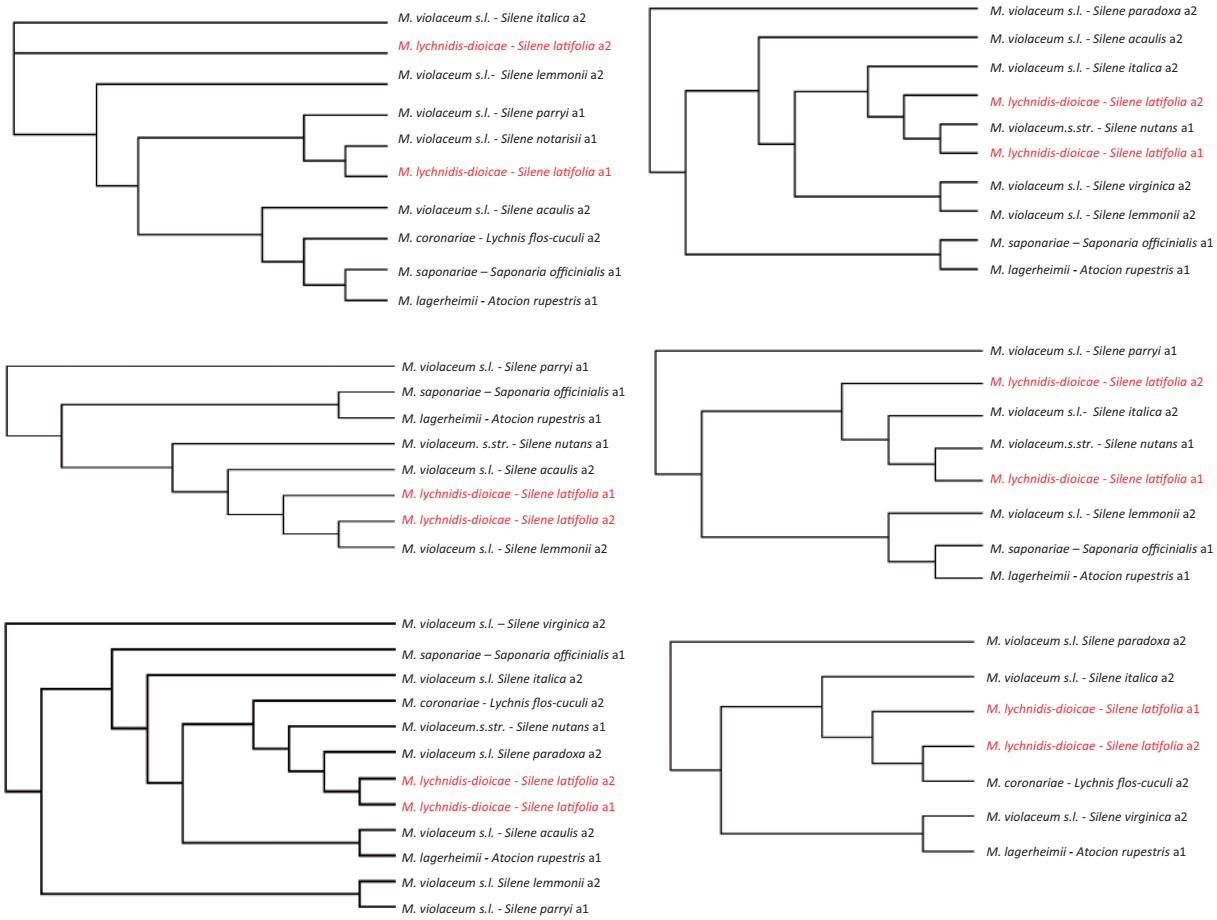


Fig. 2. Examples of gene trees in NRRs of the mating-type chromosomes. Examples are shown of different placements of the a_1 and a_2 alleles of *Microbotryum lychnidis-dioicae genes* (in gray), relative to orthologs in other *Microbotryum* species (in black). The name of the host plant is indicated next to the *Microbotryum* species name.

and a_2 alleles in a specific clade, with high similarity, suggesting that there had been recent gene conversion (fig. 2).

Gene conversion footprints were investigated for predicted genes in NRRs. We used the 196 POG alignments from various *Microbotryum* species and assigned to NRR (see below). A total of 11 gene conversion events were detected within these NRR genes, corresponding to 1.9×10^{-4} events per base pair, affecting 3.6% of the genes.

Identification of POGs and Species Tree Reconstruction

We mapped the whole genome sequences of the 11 other *Microbotryum* species and the outgroup (haploid genomes of either mating type; fig. 3) onto the *M. lychnidis-dioicae* reference genome. We thereby identified 5,726 POGs of loci in which at least 6 of the 12 species were represented and that were longer than 150 bp. These sequences were filtered to remove intron positions and surrounding regions that did not align reliably. Based on their positions in the *M. lychnidis-dioicae* reference genome, 5,453 POGs were assigned to non-MAT chromosomes, 196 to NRRs, and 69 to the MATRRs.

We built a phylogeny of the *Microbotryum* species based upon the set of 5,453 POGs from non-MAT chromosomes (fig. 3); genes on the mating-type chromosomes were excluded because they may display trans-specific polymorphism that could blur species relationships (Devier et al. 2009; Abbate and Hood 2010b; Petit et al. 2012). The relationships between species were consistent with those reported previously for this fungal genus (Kemler et al. 2006). The phylogeny was rooted using a subset of 46 POGs from the outgroup *M. pustulatum* parasitizing *Polygonum bistorta* (Kemler et al. 2006); the topology of the resulting rooted phylogeny was identical to the unrooted version based on the data set from non-MAT chromosomes, although with smaller bootstrap values. The topology obtained using the full set of 5,726 POGs (including from the mating-type chromosome loci) was also identical and

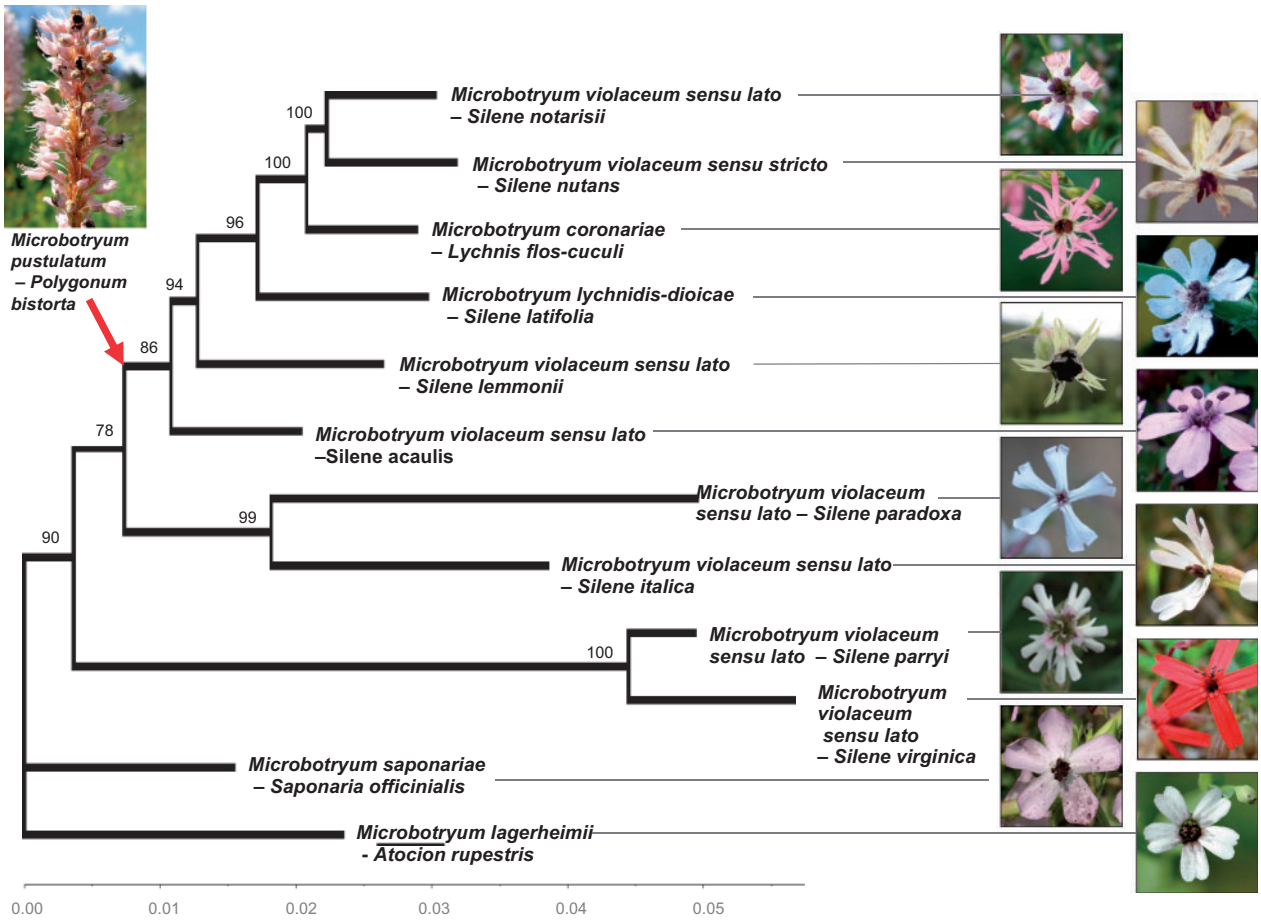


Fig. 3. Unrooted phylogenetic tree of the *Microbotryum* species studied, with a gene partition model under the maximum-likelihood framework. The topologies obtained with the different data sets were all identical: The full set of 5,726 putative orthologous genes (POGs), the 46 POGs including the outgroup, the 5,453 non-MAT chromosome POGs, and the set of 288 non-MAT chromosome POGs with high individual bootstraps. Bootstraps obtained using the full set of 5,726 POGs are indicated at the nodes. The root obtained with the 46 concatenated POGs including the outgroup species, *M. pustulatum* parasitizing *Polygonum bistorta*, is indicated with a red arrow. The scale indicates the total number of substitutions per base accumulated in each lineage. Fungal species names, and host species names, are indicated. *Microbotryum violaceum* sensu lato applies to all host-specific lineages for which no Latin name has been given, whereas *Microbotryum violaceum* sensu stricto indicates the host-specific lineage to which this Latin name was originally given. Pictures of diseased host plants are shown.

with smaller bootstrap values, due to trans-specific polymorphisms of genes in the NRRs.

Bootstrap support of 100% for nodes in phylogenomic trees may nevertheless mask disagreement between trees for individual genes. The phylogeny was therefore reconstructed using only the 288 POGs from non-MAT chromosomes displaying strong bootstrap values in their individual gene trees as previously recommended (Salichos and Rokas 2013). This yielded the same topology as the rooted tree, with 100% bootstrap support for all nodes (fig. 3).

Increased Nonsynonymous Substitution Rates in NRRs

Footprints of less effective selection were identified as the accumulation of nonsynonymous substitutions (d*N*, assumed to be deleterious) at a faster rate than of synonymous substitutions (dS) in the 12-species POG data set. The comparison of median ω (=d*N*/dS) values for terminal branches (fig. 4*A*) indicated that rates of nonsynonymous substitution in NRRs were higher than those in non-MAT chromosomes in all species except in the branch of the *Microbotryum* species parasitizing *S. lemmonii*. The differences in ω values between NRRs and non-MAT chromosomes were significant in four *Microbotryum* species, those parasitizing *S. nutans*, *S. paradoxa*, *S. parryi*, and *S. virginica*. In contrast, ω values did not differ between the MATRRs on the mating-type chromosomes and those on the non-MAT chromosomes; ω values were, however, lower in the MATRR than in non-MAT chromosomes in 8 of the 12 species (fig. 4*A*). Sequencing errors are unlikely to be responsible for the differences in the estimated substitution rates between compartments (NRR, MATRR, or non-MAT chromosomes), as the mean read qualities for the different compartments were similar.

The NRRs of both the a_1 and a_2 mating-type chromosomes had higher rates of nonsynonymous substitutions than in non-MAT chromosomes across the genus as a whole (fig. 4*C*). We then compared degeneration between the a_1

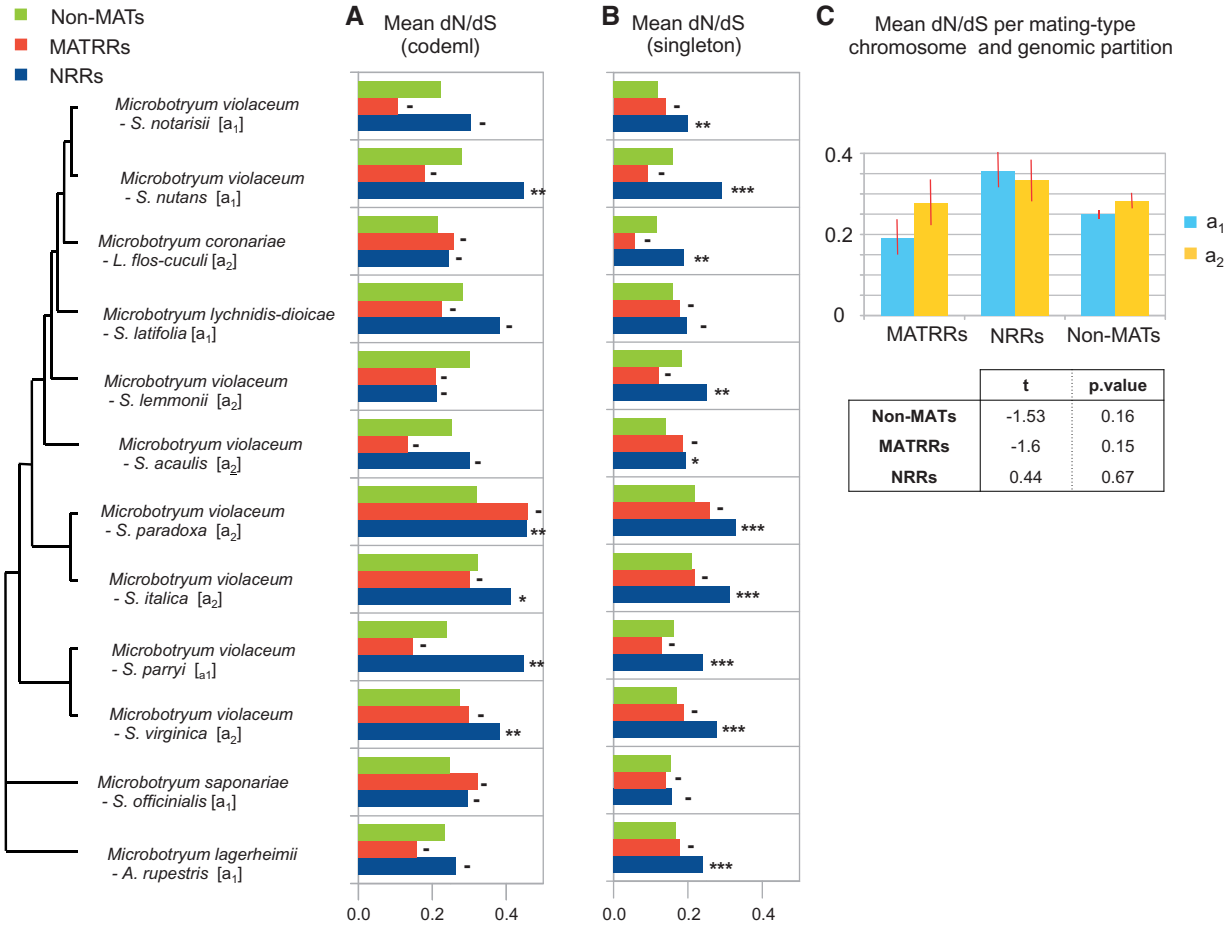


Fig. 4. Degeneration estimated by comparing mean nonsynonymous with synonymous substitutions (ω = d*N*/dS) across CDS in each of non-MAT chromosomes, nonrecombining regions of the mating-type chromosomes (NRRs), and recombining regions of the mating-type chromosomes (MATRRs) in 12 haploid *Microbotryum* genomes (either of a_1 or a_2 mating type): (*A*) With the full frequency spectrum of substitutions used to compute d*N*/dS through a free-ratio branch model (Codeml, PAML), or (*B*) with only singleton substitutions taken into account to focus on recent events. The significance of the differences between NRR and non-MAT chromosomes and between MATRRs and non-MAT chromosomes was assessed by bootstrapping. Significance levels of 1% (***), 5% (**) and 10% (*) and nonsignificance (-) are reported on barplots. (*C*) Mean ω (d*N*/dS) per mating-type chromosome (a_1 and a_2), for NRRs MATRRs and non-MAT chromosomes. Red bars indicate standard deviations. Associated two-sample *t*-tests are presented in the table: No significant difference in mean d*N*/dS was detected between a_1 and a_2.

and a_2 NRR within a single species, *M. lychnidis-dioicae*. We computed the ω values for the 45 NRR POGs with reliable alignments between the a_1 and a_2 alleles of *M. lychnidis-dioicae*. The difference between a_1 and a_2 was significant for only one of these 45 NRR genes, and thus there was no evidence for asymmetry of degeneration between a_1 and a_2 mating-type chromosomes.

To study recent, phylogenetically independent substitutions, we also analyzed ω values for mutations found in only one of the species (singletons): ω ratios were again significantly higher in NRRs than in non-MAT chromosomes (fig. 4*B*) in 10 of the 12 species (for all except *M. lychnidis-dioicae* parasitizing *S. latifolia*, and *M. saponariae* parasitizing *Saponaria officinalis*). MATRR ω values again were not significantly different from those in non-MAT chromosomes (fig. 4*B*).

Distinguishing Decreased Efficacy of Purifying Selection from Positive Selection

The higher nonsynonymous substitution rates in NRRs than in non-MAT chromosomes may reflect decreased efficacy of purifying selection, as would be expected under sheltering and reduced effective population size. Indeed, the results of codon-model tests with and without selection were not consistent with the alternative explanation of positive selection (supplementary table S1, Supplementary Material online). The proportions of loci inferred to evolve under positive selection in the 12-species POG data set appeared to be similar in NRRs (6%, 12 predicted genes), MATRRs and non-MAT chromosomes (4% each). Furthermore, discarding the POGs falling in the "nearly neutral selection" ($0.8 < \omega < 1.2$) and "positive selection" ($\omega > 1.2$) classes led to similarly higher mean ω in NRR than in non-MAT chromosomes and MATRR compartments (supplementary fig. S3*A* and *B*, Supplementary Material online). This suggests that the generally higher ω in NRRs than other compartments is a consequence of a lower efficacy of purifying selection rather than the effects of positive selection on genes.

Higher TE Accumulation in NRRs than Recombining Regions

TE content was significantly higher in the NRRs than in non-MAT chromosomes of all *Microbotryum* species and this was true for both a_1 and a_2 mating-type chromosomes (fig. 5): The TE content was at least four times greater in NRR than in non-MAT chromosomes in all cases, and 26 times more in the case of *M. violaceum* sensu stricto parasitizing *S. nutans*. We examined separately the three most frequent TE families in *M. lychnidis-dioicae*: Copia-like, Gypsy-like, and rolling circle Helitron-like elements (Hood 2005; Hood et al. 2005; Yockteng et al. 2007). The greater TE content in NRR than non-MAT chromosomes applied to all the three families (supplementary fig. S3*D*, Supplementary Material online). MATRRs tended to harbor fewer TEs than the non-MAT chromosomes (fig. 5), but the difference was not significant.

Differential Expression or Gene Loss between the Mating-Type Chromosomes in *M. lychnidis-dioicae*

We then examined gene expression using RNA sequencing (RNA-Seq) in separated a_1 and a_2 haploid cultures of *M. lychnidis-dioicae*, under conducive (low nutrient water agar) and nonconducive (nutrient-rich medium) mating environments. The gene expression profiles (supplementary fig. S4*A*, Supplementary Material online) revealed 177 genes differentially expressed (cutoff $1E^{-10}$) in *M. lychnidis-dioicae* between a_1 and a_2 haploid cultures in either low or high nutrient environments. Among these 177 genes, most of them (93) were differentially expressed under both conditions. Eighty-four of the 177 differentially expressed genes were located in the NRRs and two in the MATRRs. This represents a 49% enrichment of differentially expressed genes in the NRRs, which is highly significant (Pearson $\chi^2 = 52.61$, $P < 0.001$).

Differential expression of NRR genes between a_1 and a_2 cultures can be caused by 1) deleterious mutations impairing proper expression but sheltered in the dikaryotic/diploid stage by permanent heterozygosity in NRRs, or 2) selection favoring linkage to one mating type because of antagonistic effects in cells of the other mating type; this would be similar to genes on sex chromosomes with antagonistic effects in males and females (Charlesworth et al. 2005; Bergero and Charlesworth 2009). Were this second possibility the case, the differentially expressed genes in NRR would be expected to be associated with mating-type functions and thus potentially upregulated during mating. However, the differentially expressed genes in NRRs were not associated with any putative functions involved in mating, except for the mating-type pheromone receptor and homeodomain loci. Furthermore, we did not detect a higher percentage of genes upregulated during mating in NRRs than in non-MAT chromosomes or MATRRs: 2.63% of NRR genes were upregulated during mating, versus 2.02% in MATRRs and 1.71% in non-MAT chromosomes. Moreover, only 10 of the 130 genes upregulated during mating were also differentially expressed between separated a_1 and a_2 cells. Thus, few, if any, of the genes identified as differentially expressed between mating types are involved in mating-type determination, arguing against the idea that numerous genes have antagonistic effects between the two mating types. These observations suggest that most of the genes with weak or no expression in one mating type correspond to genes with deleterious mutations rather than mating-type specific roles.

In our RNA-Seq analysis, fewer genes appear to be expressed specifically in a_2 than a_1 haploid cells. Sixty-eight of the 86 differentially expressed genes mapping to the a_1 mating-type chromosome were significantly more strongly expressed in a_1 cultures and 18 in a_2 cultures (supplementary fig. S3, Supplementary Material online). This was probably because reads were mapped against the gene set from the a_1 reference genome, so genes lost from the a_1 mating-type chromosome (and thus only expressed in a_2 cells) would not have been identified. To recover those genes, RNA-Seq reads from the a_1 and a_2 cell libraries were separately assembled de novo. The resulting transcripts were mapped against both the

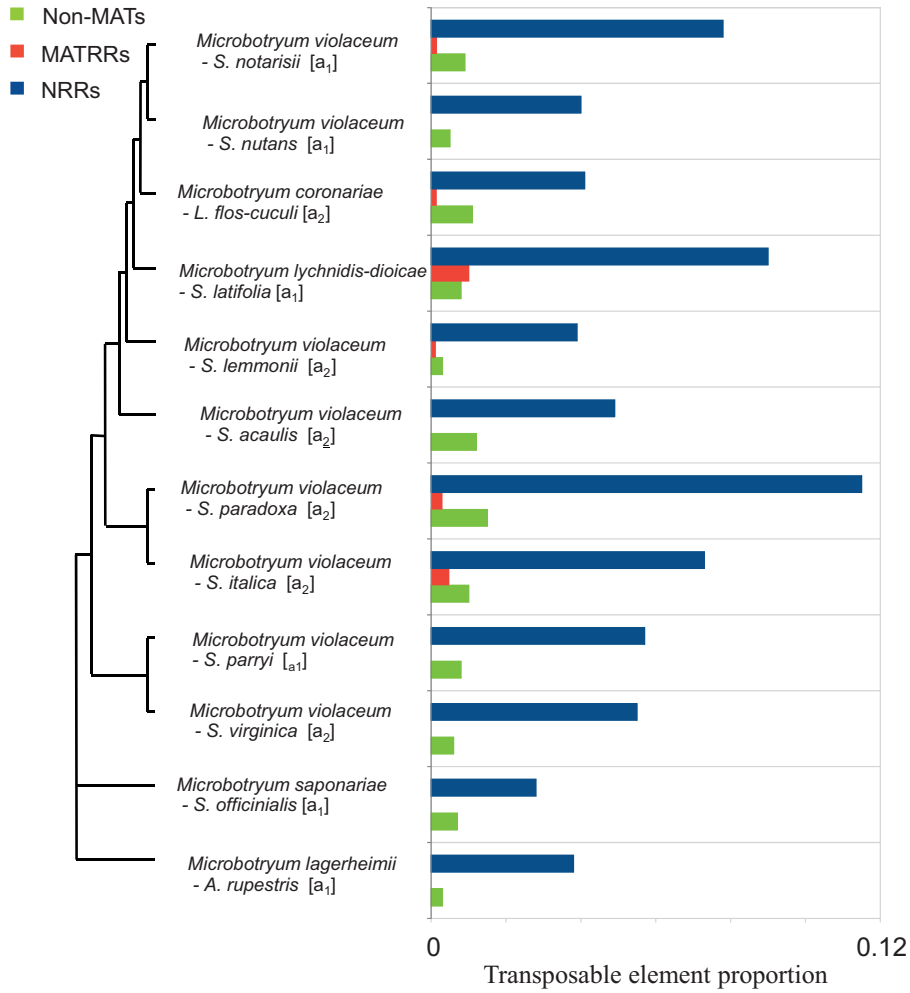


Fig. 5. TE contents in contigs assigned to non-MAT chromosomes, nonrecombining regions of the mating-type chromosomes (NRRs), and recombining regions of the mating-type chromosomes (MATRRs), in 12 haploid *Microbotryum* genomes (either of a_1 or a_2 mating type). The results reported are proportions of sequences corresponding to TEs in 1,000 random 200-bp fragments in each compartment (non-MAT chromosomes, NRRs, and MATRRs).

a_1 reference genome and a de novo assembly of the a_2 mating-type chromosome from the reference specimen. For some genes lacking transcripts in cultures of one mating type, there was evidence of gene loss from the corresponding mating-type chromosome. Sixteen of the 30 genes expressed only in cultures of the a_1 mating type were located in the NRR and not found among a_2 genome sequences. Similarly, 9 of the 21 genes expressed only in cultures of the a_2 mating type were located in the NRR and not found among a_1 genome sequences. The number of gene losses from a_1 and a_2 NRRs were not significantly different ($\chi^2 = 0.2$; $P = 0.35$). We confirmed gene loss from either one of the mating-type chromosomes by polymerase chain reaction (PCR) detection using gene-specific primers in the *M. lychnidis-dioicae* Lamole reference strain (supplementary table S2A, Supplementary Material online). We also screened 18 other *M. lychnidis-dioicae* or *M. silenes-dioicae* strains by PCR using the same primer pairs for the presence of these genes. Overall, the pattern of gene loss observed was same as that in the Lamole reference strain. However, a couple of genes were found in both mating types in several strains, indicating that the gene losses are not in all cases fixed within the species (supplementary table S2B, Supplementary Material online).

Discussion

Identification of DNA Sequences of the Mating-Type Chromosomes in *M. lychnidis-dioicae*, and Its Large NRRs

Sequencing the chromosomes extracted from pulse-field gels allowed the identification of the a_1 and a_2 mating-type chromosomes within the reference genome sequence, as well as their MATRRs and NRRs. Several lines of evidence validated our assignments of scaffolds to NRRs, MATRRs, and non-MAT chromosomes, noting that any failure to properly assign a few sequences would obscure the differences between the mating-type chromosomes and non-MAT chromosomes. The tests for differences between NRRs and recombining regions should therefore be conservative.

Orthologs between the two mating-type chromosomes (a_1 and a_2) were found in the NRRs. This finding supports the view that the dimorphic mating-type chromosomes in fungi

were initially derived from a homologous pair of non-MAT chromosomes, as observed for sex chromosomes and other fungal mating-type chromosomes (Charlesworth 1991; Charlesworth et al. 2005; Menkis et al. 2008). Several hundred genes were identified in the NRRs of the mating-type chromosomes, representing most of the predicted genes of the mating-type chromosomes. This is consistent with suppressed recombination extending over most of the mating-type chromosome pair in *M. lychnidis-dioicae*, as indicated by analysis of optical maps (Hood et al. 2013). Votintseva and Filatov (2009), in contrast, reported that the NRRs covered a much shorter section of the mating-type chromosomes in *M. lychnidis-dioicae*. That result, however, was partly due to the inclusion of loci that were repetitive sequences or were missassigned to MATRRs (Hood et al. 2013).

Divergence between a_1 and a_2 Alleles in the NRRs of the Mating-Type Chromosomes in *M. lychnidis-dioicae*

The high divergence of synonymous substitutions between the a_1 versus a_2 alleles in many of the genes in the NRR indicates that the cessation of recombination is ancient. We found no evidence for allele divergence in discrete stages, which could have been expected were there evolutionary strata in mating-type chromosomes of *M. lychnidis-dioicae*. A high heterogeneity in divergence levels was found, which could result from the evolutionary strata, although the heterogeneity was observed at small physical scale. Physical mapping of loci may be required before a definitive conclusion about the existence of evolutionary strata or other processes is possible.

Dating the suppression of recombination in *M. lychnidis-dioicae* is limited by the inability to use absolute calibration. Indeed, calibration points are rare; exploitable calibrations points require the alignment of numerous genes in fungal species much more distant than the genus level examined here. Using fossil records for calibration, we estimated that the suppression of recombination at the highly conserved mating pheromone receptor gene in *M. lychnidis-dioicae* occurred approximately 370 Ma (Devier et al. 2009). Here, we could only estimate a relative date for the suppression of recombination by investigating the placement of the a_1 and a_2 alleles in the broader gene trees that contained several *Microbotryum* species. Among the subset of NRR genes that could reliably be aligned with other species, their placements often indicated that recombination was suppressed before the speciation of the sister species *M. lychnidis-dioicae* and *M. silenes-dioicae*. This speciation event has been estimated to have been 420,000 years ago (Gladieux et al. 2011). For some predicted genes, however, a_1 and a_2 alleles were very similar, indicating recent cessation of recombination or gene conversion. The heterogeneity of divergence times may be the result of an expansion of the region with suppressed recombination (i.e., evolutionary strata), transposition of genes, and/or the influence of gene conversion. Further studies that provide more complete assembly will aid in disentangling these various possibilities.

Evidence of Degeneration in the Regions of Suppressed Recombination in the Mating-Type Chromosomes of Multiple *Microbotryum* Species

Consistent with the reduced efficacy of selection expected in the nonrecombining mating-type chromosomes, we detected signs of genetic degeneration in the NRRs of *Microbotryum* species, in the form of higher rates of nonsynonymous than synonymous substitutions in NRRs than in MATRRs or non-MAT chromosomes. Note that the nonsynonymous substitutions detected here may represent rare variants within the species, as we did not have polymorphism data to check whether the substitutions had been fixed at the species level. Even in this case, however, the finding of higher numbers of nonsynonymous substitutions in NRRs than in non-MAT chromosomes would still be a strong indication of a lower efficacy of purifying selection, leading to deleterious substitutions being maintained at higher frequencies in the NRRs than in non-MAT chromosomes.

The different patterns of gene expression between a_1 and a_2 haploids also provided strong evidence for degeneration, with some genes showing reduced expression or even having been lost in one mating type. The differential expression was not likely to be due to selection associated with mating-type determination because, other than mating pheromone receptors and homeodomain loci, differentially expressed genes in NRRs did not have any obvious mating functions. Consistent with this, cells of different mating types are not very differentiated and are of similar size, and the haploid stage is very transient and restricted to the tetrad formation (Zakharov 1986; Hood and Antonovics 2000, 2004) (supplementary fig. S1, Supplementary Material online). Only the mating-type genes at the pheromone/receptor and the homeodomain loci are therefore clearly expected to be associated with mating-type determinism, as is typical of basidiomycetes (Feldbrugge et al. 2004). Under "mating" conditions, the genes in the NRRs did not appear more frequently upregulated than genes on non-MAT chromosomes. The high frequency of genes with mating-type-specific expression in the NRRs is thus likely to be associated with the accumulation of deleterious mutations impairing proper expression. Indeed, genes not required for growth during the brief haploid stage may carry loss-of-function mutations at high frequencies, as these would be sheltered by the permanent heterozygosity of these regions.

The substantial degeneration in the NRRs agrees with observations of frequent haplolethal mutations in natural populations of several *Microbotryum* species. Deleterious alleles have been detected that were linked to one mating type (a_1 or a_2 depending on the strain), which prevent sustained growth of cells of one mating type when the haploid stage is artificially extended through in vitro culturing (Oudemans et al. 1998; Hood and Antonovics 2000; Thomas et al. 2003). In the case of the Lamole reference strain, from which the reference genome was obtained, both a_1 and a_2 cells can grow in vitro, so genes lost from the mating-type chromosomes in this genome are not essential for haploid growth.

The TE content of the *Microbotryum* mating-type chromosome NRRs was found in all species to be much higher than that of their MATRRs and non-MAT chromosomes. This is an expected consequence of suppressed recombination, as has been observed in sex chromosomes (Bachtrog 2013); TE insertions may contribute to degeneration by disrupting genes or altering their expression. TE accumulation was more similar between the MATRRs and non-MAT chromosomes, consistent with recombination occurring in MATRR and allowing for more effective selection such that deleterious TE insertions are more likely to be purged. Various families of TEs accumulated differently in NRRs of different *Microbotryum* species, providing further evidence of ongoing and independent mating-type chromosome degeneration during the divergence of these fungal lineages. Analyses of TE transcripts and of mutations also indicate that the activity of TEs in several *Microbotryum* species is continuing (Garber and Ruddat 1995, 1998; Yockteng et al. 2007).

A possible bias in our analyses is that all genomes were mapped against the *M. lychnidis-dioicae* reference assembly. Therefore, genes identified as present in the mating-type chromosomes based on the orthology with *M. lychnidis-dioicae* may be present on non-MAT chromosomes in other *Microbotryum* species. However, our analyses are conservative, as any movements between genomic compartments would obscure differences; our results indicate that sequences assigned to the mating-type chromosome evolved in a distinctive manner across species and that it is unlikely that many of them were misassigned between NRR, non-MAT chromosomes, and MATRR categories.

Overall, there are various lines of evidence for high levels of divergence, degeneration, and TE accumulation on the mating-type chromosomes in numerous species of *Microbotryum*. It is striking that a nonsex chromosome system exhibits such similar phenomena to well-characterized in sex chromosomes, with extensive suppressed recombination, loss of gene content, and accumulation of a high content of repetitive elements. Importantly, similar degeneration was observed on a_1 and a_2 mating-type chromosomes, in contrast to XY sex determining systems, where the Y chromosome degenerates much more than the X (Bachtrog 2013). This is consistent with Bull's prediction that degeneration is not expected to be asymmetrical in species where sex or mating compatibility is determined in the haploid stage and the chromosomes responsible are always heterogametic in the diploid stage (Bull 1978).

Materials and Methods

Reference Genome (*M. lychnidis-dioicae*)

Information on the a_1 haploid genome of the Lamole reference strain of *M. lychnidis-dioicae* parasitizing *S. latifolia* was obtained from the Broad Institute web-server. This included the scaffolds (i.e., supercontigs) corresponding to the nonmitochondrial regions, the associated CDS annotations (gff files), and the annotations obtained with Interproscan, a powerful integrated database and diagnostic tool (Jones et al. 2014).

Resequencing of the a_1 Mating-Type Chromosomes (*M. lychnidis-dioicae*)

The a_1 mating-type chromosome from the same haploid Lamole genotype whose whole genome was sequenced by the Broad Institute was isolated using electrophoretic karyotype analysis as previously described (Hood et al. 2004). The DNA was subjected to multiple displacement amplifications with the REPLI-g Kit (QIAGEN), and then sequenced at a coverage of 1,175 × using 2- and 5-kb insert size, mate-paired libraries for Titanium version 454 technology (www.roche.com). A first assembly using Newbler (the Roche assembler) was performed with approximately 30-fold and approximately 20-fold coverage.

Identifying the Mating-Type Chromosome Scaffolds on the a_1 Reference Genome

To determine which scaffolds of the *M. lychnidis-dioicae* genome belonged to the a_1 mating-type chromosomes and which to non-MAT chromosomes, the assembled contigs from the gel-isolated mating-type chromosomes were mapped against the *M. lychnidis-dioicae* reference genome using NUCmer (http://mummer.sourceforge.net/). Matches were assigned to the mating-type chromosome, and scaffolds without any match were assigned to non-MAT chromosomes. To validate these results, our 454 data (Newbler) were mapped onto the Broad Institute assembly scaffolds to identify contamination from non-MAT chromosomes by their scattered weak read coverage. Finally, all nonmitochondrial Broad Institute assembly scaffolds with coverage greater than 20-fold when mapped with contigs from the gel-isolated mating-type chromosomes were assigned to mating-type chromosomes.

Sequencing and Assembly of the a_2 Mating-Type Chromosome (*M. lychnidis-dioicae*)

The a_2 mating-type chromosome from the Lamole reference strain was isolated and amplified as described above for the a_1 mating-type chromosome. Four libraries were sequenced on Illumina HiSeq 2000, one paired-end library (insert size of 250 bp) and three mate-pair libraries (expected insert sizes 3, 8, 20 kb; observed insert sizes of 2.2, 9, 12.5 kb, respectively). The sequence data for the paired-end library (100 bp sequenced at both ends of fragments) represent a coverage of 1,175 × (given the expected genome size of 25 Mb). The 2.2-, 9- and 12.5-kb libraries include 34, 6.4 and 2.2 millions of base pairs, respectively.

Contigs were generated from the paired-end data, with SOAPdenovo 2.04 (Li et al. 2010) using a kmer of 91. Then, a three-step process was applied to remove artifacts from the assembly: 1) Contigs fully included in longer contigs were removed; 2) contigs containing 15-mers that were not present in raw reads were split, as they are likely to be chimeras; and 3) contigs were trimmed (46 bp ~ 1/2 kmer length) to remove the artificial overlaps created by the large kmer used.

Paired-end and mate-pair reads were mapped on contigs using the glint software (Faraut T and Courcelle E, unpublished data; http://lipm-bioinfo.toulouse.inra.fr/

download/glint; parameters --best-score --step 2 --no-lc-filtering --mmis 5 --lrmin 80). Due to the large numbers of highly conserved repeats, additional stringent criteria were used for further hit selection (length of the hit $\geq$ 90 bp; identity = 100%). LYNX scaffolder (Gouzy J, unpublished data) was used to generate scaffolds. A minimum number of five links was required for linking information from the 250-nt and 2.2-kb libraries to be considered, and at least two links for the 9- and 12.5-kb libraries. Finally, SOAPGapCloser was used to fill in gaps in the assembly. The final assembly was 24,832,594 nucleotides long in 388 scaffolds (N50 of scaffolds longer than 1,000 nt = 299 kb; 4.75% of *N*).

Identification of the MATRRs (*M. lychnidis-dioicae*)

To identify sequence assemblies corresponding to the two MATRRs of *M. lychnidis-dioicae* mating-type chromosomes reported in optical maps (Hood et al. 2013), contigs were aligned with restriction-digest optical maps using the MapSolver software (OpGen) (Latreille et al. 2007). This software predicts restriction digest patterns from DNA sequence data for comparison with the observed optical maps (restriction-digest fragments sizes); alignments are based on a cumulative scoring function that rewards matching cut-site distributions and penalizes mismatch or missing sites. The software's default settings were used to calculate alignments.

Genome Sequencing of Additional *Microbotryum* Species and Mapping against the Reference

The fungal material used in this study was collected as diploid teliospores from natural populations in North America and Europe. Some *Microbotryum* species have not received a specific Latin name yet and are referred to as *M. violaceum* sensu lato with the additional indication of their host plant species (fig. 3). *Microbotryum* species have been distinguished on the basis of a combination of multiple gene genealogies, interfertility assays, and hybrid fitness and fertility assessments (Le Gac, Hood, Fournier, et al. 2007; Le Gac, Hood, and Giraud 2007; de Vienne, Hood, et al. 2009; de Vienne, Refrégier, et al. 2009). Haploid cultures of the 11 *Microbotryum* species were obtained by micromanipulation of the postmeiotic yeast-like sporidial cells, which were grown on potato dextrose agar (Difco). Mating types of the haploid cultures were identified by pairing with cultures of known mating types and examining the conjugation response elicited by the alternate mating pheromone (Day 1979). Also, PCR primers that discriminate between a_1 and a_2 pheromone receptors (Devier et al. 2009) were used to test extracted DNA for mating type with the DNeasy Plant Mini Kit (QIAGEN). The extracted DNA from the *Microbotryum* species was also used to obtain shotgun sequence libraries using the Titanium version 454 technology (performed at the University of Virginia, Genomics Core Facility) with a coverage of 2–2.5 $\times$. Only reliable sequenced positions, with sufficient qualities, were taken into account.

Sequence libraries for each of the 11 *Microbotryum* species were assembled individually by mapping against the *M. lychnidis-dioicae* reference genome using Newbler. Pairwise mapped sequences were realigned within species libraries using Prank (Loytynoja and Goldman 2005, 2008) with the "-codon" option in accordance with recent recommendations based on simulations for the use of codon-based models to detect selection (Fletcher and Yang 2010; Markova-Raina and Petrov 2011). In the pairwise alignments obtained, the positions creating gaps on the reference sequence were discarded to maintain the initial coordinates in the reference.

Sequences matching the same fragments of the reference *M. lychnidis-dioicae* genome were clustered together as putative orthologs using Newbler-generated coordinates of the mapped contigs on the reference genome. Noncoding regions were filtered out from POGs using the CDS annotations of the reference genome provided by the Broad Institute. Loci shorter than 300 bp and those missing from more than six *Microbotryum* species were discarded.

An outgroup species, *M. pustulatum*, parasitizing *P. bistorta* (Kemler et al. 2006), was also sequenced and assembled against the reference genome as described above. Given its distance from the *Microbotryum* species (mean number of substitutions per site between the outgroup and the ingroups is 0.085 ± 0.067), few POGs (46) were available for the phylogenetic studies. Consequently a second, larger data set was generated without the outgroup species, including 5,726 POGs, which was used for all the subsequent analyses.

Species Phylogeny Inference

To avoid potential biases in the phylogenetic signal caused by trans-specific polymorphism, as observed at some mating-type genes (Devier et al. 2009; Abbate and Hood 2010b; Petit et al. 2012), only POGs not belonging to mating-type chromosomes were used for the species tree reconstruction. The 5,453 POGs not belonging to mating-type chromosomes were concatenated and JMODELTEST (Posada 2008) was used to infer the most suitable nucleotide substitution model. Phylogenetic trees were constructed in a maximum-likelihood framework using RAxML 7.8.1 (Stamatakis et al. 2005) with a GTR + gamma substitution model and on gene-partitioned data sets (substitution model parameters were optimized for each partition of genes with JMODELTEST). Other phylogenies based on codon position partitioning (computing estimations separately for the first two codon positions, GC_{12}, vs. the third codon position, GC_3) or on other amino acid substitution models, had identical topologies. A rooted phylogeny was also produced based on the small subset of 46 POGs for which orthologs were found in the *Microbotryum* species and the *P. bistorta* outgroup species (Lutz et al. 2005; Kemler et al. 2006).

The robustness of nodes in our tree was tested using only the POGs not belonging to the mating-type chromosomes yielding individual trees with a high mean bootstrap support (MBS) or a high Tree Certainty (TC) (Salichos and Rokas 2013). For each POG, ModelTest v 2.3.1 (Darriba et al. 2012) was used to find the best model, and PhyML v3 (Guindon and Gascuel 2003) to reconstruct each tree with 100 bootstrap replicates. For each tree, the MBS and the TC (using RAxML 7.8.1) were then computed. MBS and TC values being highly

correlated ($R^2 = 0.8276$, $P < 2.2e\text{-}16$), only MBS was used thereafter. POGs with MBS ≥ 80 were selected: The sequences of the 288 POGs selected were concatenated. RAxML v 7.2.8 was used to reconstruct the final trees using the gtr+gamma substitution model and 100 bootstrap replicates. Note that the tree obtained when selecting the genes based on their TC was the same as that based on MBS.

Placement of NRR Alleles in a_1 and a_2 of *M. lychnidis-dioicae* in the Species Tree and Computation of Their Divergence

BLASTn and tBLASTn searches (NCBI blast 2.2.29 universal macosx) of the *M. lychnidis-dioicae* a_1 alleles for predicted genes in NRRs were performed to search for hits in the a_2 mating-type chromosome of *M. lychnidis-dioicae*. Both methods converged on the same hits. The a_2 protein hits obtained from the tBLASTn search were added to the respective POG protein alignments obtained above and realigned using t-coffee 10.00.r1613 (Notredame et al. 2000) with default settings. The resulting alignments were trimmed with Trimal version 1.2rev59 (Capella-Gutierrez et al. 2009), removing 20% of the gaps introduced when adding the a_2 sequences. We kept only the 45 reliable alignments, and used them to build the trees with RAxML version 7.4.4 (Stamatakis 2006) and the GTRGAMMA model. The placement of the a_1 and a_2 alleles relative to alleles of other species was visually inspected in the predicted trees. The synonymous divergence of NRR alleles between a_1 and a_2 of *M. lychnidis-dioicae* was computed as the estimated number of synonymous substitutions, including inferred multiple substitutions, divided by the number of synonymous sites in the sequence. The synonymous divergence and their associated standard error values were computed using the yn00 program in the PAML package (v. 4.7a) (Yang 2007) (fig. 1).

Detection of Selection Signatures

The efficacy of purifying selection to maintain gene function was assessed by two approaches. First classic models of codon substitution (Yang 2007) were used: Parameters were inferred from the species phylogeny to reconstruct changes in the ratios of nonsynonymous to synonymous substitutions ($\omega = dN/dS$). The rationale is that nonsynonymous, deleterious substitutions are expected to accumulate if purifying selection is weak or relaxed. Ratios of dN/dS were estimated for each POG using codon models with CODEML in the PAML package (Yang 2007). The tree used was the species tree (fig. 3). Species not represented in these POGs were pruned from the species tree using the Ape package in R. We also ran analyses using individual gene trees for NRR POGs instead of species tree, and similar results were obtained (supplementary fig. S3C, Supplementary Material online).

The efficacy of purifying selection was also investigated using an estimation of the global ω per branch (the "free-ratio branch model" under CODEML): Mean and median ω values per branch for each terminal lineage were calculated, and genes with dS values close to 0 that would produce artifactually very large ω values were excluded. Positive selection was tested by comparing neutral and selective codon models. Genes displaying unrealistically high dN or dS values (i.e., "999" in the codeml output) were discarded from the analysis as they probably corresponded to saturated levels of substitutions and would bias ω estimations. We also excluded from the analysis those genes with null dS and dN values. More stringent filtering (discarding sequences with SdS lower than 2, 3, or 4, or filtering using branch length parameters) led to results for ω similar to those presented.

To control for the possibility of positive selection increasing dN/dS, which would not constitute degeneration, positive selection was also tested using a branch-site test based on codon models which accounts for variation of the ω ratio between both sites and lineages (Yang and Nielsen 2002; Zhang et al. 2005). A likelihood ratio test was used to compare a null model without positive selection and an alternative model allowing positive selection (Nielsen and Yang 1998). In cases where the null hypothesis was rejected, and the "selective" model retained, positions under positive selection among foreground branches were determined using a Bayesian empirical Bayes approach (Yang et al. 2005). Similar automated procedures (python + R scripts) were used to run codeml automatically and both extract and filter parameter values as in the free-ratio branch model described above.

The second approach used to assess the efficacy of purifying selection focused on recently derived substitutions; these have the advantage of being free of assumptions on phylogenic relationships and are more likely to have occurred while being in the current genomic compartment; indeed, substitutions may have occurred before the putative transfer of genes between recombining regions and NRRs (Abbate and Hood 2010b; Petit et al. 2012). Recent substitutions were identified as singletons in aligned POGs. From the singleton data set for all alignments of orthologous loci, median dN/dS ratios between different genomic compartments (NRRs, MATRRs, and non-MAT chromosomes) were computed and compared. The nonsynonymous rate was computed as $dN = n/N$, where n is the nonsynonymous singleton count and N is the number of nonsynonymous sites in the alignment. Similarly, the synonymous rate was computed as $dS = s/S$ where s is the count of synonymous singleton substitutions and S the number of synonymous sites in the alignment. This approach, however, does not take into account multiple substitutions at a given site so it may underestimate substitution counts.

Gene Conversion Analyses

Geneconv v1.81 (Sawyer 1989) was used to look for footprints of gene conversion in NRRs directly. To be conservative, only results from estimations of "inner" gene converted fragments are reported. As defined in the Geneconv manual, such fragments represent evidence of possible gene conversion events between ancestors of two sequences in the alignment. Geneconv was executed through a batch script independently for the alignments of putative CDS regions (i.e., POGs) in NRRs.

TE Analyses

TE contents of contigs were estimated and assigned to non-MAT chromosomes, NRRs, and MATRRs in each *Microbotryum* species. Because many contigs were small, especially in the NRRs, the analysis was run on fragments of similar sizes using databases of 200-bp fragments generated in each species and genomic compartment (non-MAT chromosomes, NRR, and MATRR). To compare databases of equivalent sizes, 1,000 reads were randomly resampled without replacement from each database. A set of full-length TEs of the most common types in *Microbotryum* (Hood 2005; Hood et al. 2005; Yockteng et al. 2007) (Copia-like, e.g., Supercontig 147: 1,901–7,063 bp; Gypsy-like, e.g., Supercontig 151: 14,408–18,631 bp; and rolling circle Helitron-like, e.g., Supercontig 124:8,251–12,551 bp) were used as queries to search against the 1,000 read databased using tBLASTx (Basic Local Alignment Search Tool; blast.ncbi.nlm.nih.gov/Blast.cgi) with the significance threshold set at E value $< 10^{-6}$. To assess the significance of the differences in TE counts between genomic compartments, Z tests were used. P values were assigned by comparing the computed Z value to the critical Z value for a two-tailed test in a standard normal distribution table. The null hypothesis was that the difference between the proportions of the two samples considered is zero.

Expression Analysis in the Reference Species (*M. lychnidis-dioicae*)

Expression data were generated under five different conditions using the Lamole reference strain, with two biological replicates per condition. Haploid fungal cells of Lamole a_1 sporidial cells (p1A1Rich) and a_2 sporidial cells (p1A2Rich) were grown separately for 5 days on yeast peptone dextrose media (1% yeast extract, 10% dextrose, 2% peptone, 2% agar; rich growth conditions) at room temperature, and then harvested for RNA extraction. Similarly, haploid cells grown separately on 2% water agar for 2 days were harvested for RNA extraction and named p1A1Water and p1A2Water; these samples allowed comparison between the gene expression when haploid cells of different mating types were subjected to a nutrient-free environment without a mating partner. We also studied a "mated" condition: An equal mixture of a1 and a2 cells was kept at 14 °C on 2% water agar for 2 days, which induces conjugation, the first stage of mating (Day 1979).

RNA was extracted from samples of the cells cultured in these various conditions using the RNeasy Plant Mini Kit (QIAGEN, cat no: 74904) according to the manufacturer's instructions. Ambion's TURBO DNA-free (Applied Biosystems, cat no: AM1907) DNase was used according to the manufacturer's instruction. For quality assessment before Illumina sequencing, 5 μg of DNase-treated RNA was reverse transcribed with SuperScript III First Strand Synthesis System for reverse transcription PCR (Life Technologies, cat no: 18080-051). PCR was performed using TaKaRa Ex Taq Hot-Start DNA Polymerase (Takara, cat. no: RR001B) in a reaction volume of 25 μl. To check for DNA contamination and possible inhibitory substances in the RNA, two sets of housekeeping primers (Eurofins/MWG/Operon) were used. Targeting the *Microbotryum mepA* gene, the forward primer, 5′-CTTT TGCGTAGGAAGAATGC-3′ and the reverse primer, 5′-AGCA CTGAACACCCCAACTT-3′ were used; this combination yielded a 532-bp fragment from cDNA, and a 1,039-bp fragment from genomic DNA. The other primer combination targeted the *Microbotryum* beta-tubulin gene, with forward primer, 5′-CGGACACCGTTGTCGAGCCT-3′, and reverse primer, 5′-TGAGGTCGCCGTGAGTCGGT-3′, yielding a 150-bp fragment from cDNA and a 215-bp fragment from genomic DNA. The PCR program was 30 s at 94 °C, 30 s at 60 °C, and 1 min at 72 °C for 35 cycles. An Agilent BioAnalyzer was also used to determine the quality of the RNA prior to Illumina RNA-Seq.

De novo assemblies of both the a_1 and a_2 RNA-Seq data were generated with Trinity (Grabherr et al. 2011) to identify genes specific to each mating chromosome. Assembled transcripts were compared between a_1 and a_2 using blast to identify transcripts specific for each cell type: a_1 transcripts were mapped to the gene set based on the whole genome assembly; the a_2 assembly is available at the Broad website, at http://www.broadinstitute.org/annotation/genome/Microbotryum_violaceum/assets/Mvio_A2_trinity.fa.gz.

Differential expression scripts in the Trinity pipeline version r2013-02-25 (Grabherr et al. 2011) were run to process RNA-Seq reads from the ten libraries (two biological replicates of each of p1A1Water, p1A2Water, p1A1Rich, p1A2Rich, and Mated). The RNA-Seq reads were aligned by bowtie (Langmead et al. 2009) against protein CDS extracted from the reference genome, adding 100 bases of flanking sequence to approximate UTRs. The read alignment files were then processed with RSEM (Li and Dewey 2011) to quantify transcript abundances. To identify genes differentially expressed between each pair of conditions, edgeR with the trimmed mean of M-value normalization method (TMM) (Robinson et al. 2010; Kadota et al. 2012) was run using an adjusted P value cutoff of 1E-10; TMM normalization is a method for estimating relative RNA production levels from RNA-Seq data by estimating scale factors between samples. The agglomerative hierarchical clustering method implemented in the R package cluster (Maechler et al. 2013) was then used to identify coexpressed gene clusters of those genes with significant differential expression between conditions. Expression data are available at the Broad Institute website, www.broadinstitute.org/annotation/genome/Microbotryum_violaceum.

Some genes displayed mating-type-specific expression profiles, which could be the result of a lack of expression of these genes on one of the mating-type chromosomes or of hemizygosity (there is only a single copy of the gene, on only one of the mating-type chromosomes). To distinguish these possibilities, the presence of the genes expressed only from one mating-type chromosome was investigated using BLASTn by aligning the assembled cDNA against the genomic sequence of the opposite mating type (the published a_1 haploid genome for a_2-specific transcripts, and the a_2 mating type chromosome for a_1-specific ones).

The absence of some genes from one of the mating-type chromosomes was then checked with primers specific for

these genes (supplementary table S2, Supplementary Material online): DNAs extracted from a_2 and a_1 cells from the reference Lamole strain, 20 other *M. lychnidis-dioicae* strains, and one strain from each other species studied here, were tested for amplification with these primers.

Statistical Tests Using Monte Carlo Resampling of the Genomic Background

To test for the significance of the differences in the values of several variables described above (d*N*/d*S*, GC content, preferred/unpreferred codons) between the NRRs, MATRRs, and non-MAT chromosomes, a null genomic distribution of the variable values was constructed through randomization. To do this, 1,000 sets of non-MAT chromosome loci were randomly sampled and the mean and median values were computed for the variables of interest. The size of each of these sets was selected to be equal to the number of loci observed in the region of interest (NRR or MATRR). Significance thresholds of 0.01 or 0.05 were used to score the mating-type chromosome data set (MATRR or NRR) as significantly different from the non-MAT chromosome set.

Supplementary Material

Supplementary tables S1 and S2 and figures S1–S4 are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

The authors thank Pierre Fontanillas, Laurent Duret, Gabriel Marais, and Antoine Branca for valuable discussions. This work was supported by the Agence Nationale de la Recherche (ANR-09-BLAN-064 to T.G.), the European Research Council (starting grant GenomeFun 309403 to T.G.), and the National Science Foundation (DEB 0747222 to M.E.H., 0947963 to M.H.P. and C.A.C.). The authors thank the IT service at the Station Biologique Marine de Roscoff and the Centre National de Séquencage (Génoscope) for computing cluster facilities. They are grateful to Stéphanie Le Prieur, Alodie Snirc, and Odile Jonot for help with molecular biology experiments. They also thank Janis Antonovics, Mel Harte, and Steve Matson for some of the pictures in figure 2. The accession numbers are PRJEB6323 for sequences of the a_2 mating type chromosome of *M. lychnidis-dioicae* and PRJEB6548 for the genomes of the other species. T.G. and M.E.H. designed the research project; J.P., P.W., E.P., M.E.H., Z.C., S.T., C.A.C., and M.H.P. obtained the genomic sequences and transcriptome data; E.F., V.B., Z.C., C.A.C., H.B., J.G., and M.P. performed assembly and mapping; E.F., H.B., E.P., J.G., V.B., C.A.C., M.P., D.M.d.V., H.B., P.G., and G.A. performed analyses; E.F., T.G., and M.E.H. wrote the paper; all authors approved the final manuscript.

References

Abbate JL, Hood ME. 2010. Dynamic linkage relationships to the mating-type locus in automictic fungi of the genus *Microbotryum*. *J Evol Biol.* 23:1800–1805.

Antonovics J, Abrams J. 2004. Intratetrad mating and the evolution of linkage relationships. *Evolution* 58:702–709.

Bachtrog D. 2003. Accumulation of Spock and Worf, two novel non-LTR retrotransposons, on the neo-Y chromosome of *Drosophila miranda*. *Mol Biol Evol.* 20:173–181.

Bachtrog D. 2005. Sex chromosome evolution: molecular aspects of Y chromosome degeneration in *Drosophila*. *Genome Res.* 15: 1393–1401.

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

Bachtrog D, Kirkpatrick M, Mank JE, McDaniel SF, Pires JC, Rice WR, Valenzuela N. 2011. Are all sex chromosomes created equal? *Trends Genet.* 27:350–357.

Bartolome C, Charlesworth B. 2006. Evolution of amino-acid sequences and codon usage on the *Drosophila miranda* neo-sex chromosomes. *Genetics* 174:2033–2044.

Bergero R, Charlesworth D. 2009. The evolution of restricted recombination in sex chromosomes. *Trends Ecol Evol.* 24:94–102.

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

Billiard S, Lopez-Villavicencio M, Devier B, Hood M, Fairhead C, Giraud T. 2011. Having sex, yes, but with whom? Inferences from fungi on the evolution of anisogamy and mating types. *Biol Rev.* 86:421–442.

Bull JJ. 1978. Sex chromosomes in haploid dioecy—unique contrast to Mullers theory for diploid dioecy. *Am Nat.* 112:245–250.

Capella-Gutierrez S, Silla-Martinez JM, Gabaldon T. 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25:1972–1973.

Charlesworth B. 1991. The evolution of sex chromosomes. *Science* 251: 1030–1033.

Charlesworth D, Charlesworth B, Marais G. 2005. Steps in the evolution of heteromorphic sex chromosomes. *Heredity* 95:118–128.

Darriba D, Taboada G, Doallo R, Posada D. 2012. jModelTest 2: more models, new heuristics and parallel computing. *Nat Methods.* 9:772.

Day DW. 1979. Mating type and morphognesis in *Ustilago violacea*. *Bot Gaz.* 140:94–101.

de Vienne D, Hood M, Giraud T. 2009. Phylogenetic determinants of potential host shifts in fungal pathogens. *J Evol Biol.* 22:2532–2541.

de Vienne DM, Refrégier G, Hood M, Guigue A, Devier B, Vercken E, Smadja C, Deseille A, Giraud T. 2009. Hybrid sterility and inviability in the parasitic fungal species complex *Microbotryum*. *J Evol Biol.* 22: 683–698.

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

Ellis N, Taylor A, Bengtsson BO, Kidd J, Rogers J, Goodfellow P. 1990. Population-structure of the human pseudoautosomal boundary. *Nature* 344:663–665.

Ellison CE, Stajich JE, Jacobson DJ, Natvig DO, Lapidus A, Foster B, Aerts A, Riley R, Lindquist EA, Grigoriev IV, et al. 2011. Massive changes in genome architecture accompany the transition to self-fertility in the filamentous fungus *Neurospora tetrasperma*. *Genetics* 189:55–69.

Erlandsson R, Wilson JF, Paabo S. 2000. Sex chromosomal transposable element accumulation and male-driven substitutional evolution in humans. *Mol Biol Evol.* 17:804–812.

Feldbrugge M, Kamper J, Steinberg G, Kahmann R. 2004. Regulation of mating and pathogenic development in *Ustilago maydis*. *Curr Opin Microbiol.* 7:666–672.

Filatov DA, Charlesworth D. 2002. Substitution rates in the X- and Y-linked genes of the plants, *Silene latifolia* and *S. dioica*. *Mol Biol Evol.* 19:898–907.

Fletcher W, Yang Z. 2010. The effect of insertions, deletions, and alignment errors on the branch-site test of positive selection. *Mol Biol Evol.* 27:2257–2267.

Fraser JA, Diezmann S, Subaran RL, Allen A, Lengeler KB, Dietrich FS, Heitman J. 2004. Convergent evolution of chromosomal sex-determining regions in the animal and fungal kingdoms. *PLoS Biol.* 2: 2243–2255.

Garber ED, Ruddat M. 1995. Genetics of *Ustilago violacea*. XXXII. Genetic evidence for transposable elements. *Theor Appl Genet.* 89:838–846.

Garber ED, Ruddat M. 1998. Genetics of *Ustilago violacea*. XXXIV. Genetic evidence for a transposable element functioning during mitosis and txo transposable elements functioning during meiosis. *Int J Plant Sci.* 159:1018–1022.

Giraud T, Yockteng R, Lopez-Villavicencio M, Refregier G, Hood ME. 2008. The mating system of the anther smut fungus, *Microbotryum violaceum*: selfing under heterothallism. *Eukaryot Cell.* 7:765–775.

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

Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng QD, et al. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol.* 29:644–652.

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

Guindon S, Gascuel O. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol.* 52: 696–704.

Hood ME. 2002. Dimorphic mating-type chromosomes in the anther-smut fungus. *Genetics* 160:457–461.

Hood ME. 2005. Repetitive DNA in the automictic fungus *Microbotryum violaceum*. *Genetica* 124:1–10.

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

Hood ME, Antonovics J, Koskella B. 2004. Shared forces of sex chromosome evolution in haploids-mating and diploids-mating organisms: *Microbotryum violaceum* and other model organisms. *Genetics* 168: 141–146.

Hood ME, Antonovics JA. 2004. Mating within the meiotic tetrad and the maintenance of genomic heterozygosity. *Genetics* 166: 1751–1759.

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

Hood ME, Petit E, Giraud T. 2013. Extensive divergence between mating-type chromosomes of the anther-smut fungus. *Genetics* 193: 309–315.

Johnson LJ, Antonovics J, Hood ME. 2005. The evolution of intratetrad mating rates. *Evolution* 59:2525–2532.

Jones P, Binns D, Chang H, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G, et al. 2014. InterProScan 5: genome-scale protein function classification. *Bioinformatics* 30: 1236–1240.

Kadota K, Nishiyama T, Shimizu K. 2012. A normalization strategy for comparing tag count data. *Algorithms Mol Biol.* 7:5.

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

Lahn BT, Page DC. 1999. Four evolutionary strata on the human X chromosome. *Science* 286:964–967.

Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* 10:R25.

Latreille P, Norton S, Goldman BS, Henkhaus J, Miller N, Barbazuk B, Bode HB, Darby C, Du ZJ, Forst S, et al. 2007. Optical mapping as a routine tool for bacterial genome sequence finishing. *BMC Genomics* 8:6.

Le Gac M, Hood ME, Fournier E, Giraud T. 2007. Phylogenetic evidence of host-specific cryptic species in the anther smut fungus. *Evolution* 61:15–26.

Le Gac M, Hood ME, Giraud T. 2007. Evolution of reproductive isolation within a parasitic fungal complex. *Evolution* 61:1781–1787.

Li B, Dewey CN. 2011. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* 12:323.

Li R, Zhu H, Ruan J, Qian W, Fang X, Shi Z, Li Y, Li S, Shan G, Kristiansen K, et al. 2010. De novo assembly of human genomes with massively parallel short read sequencing. *Genome Res.* 20:265–272.

Loytynoja A, Goldman N. 2005. An algorithm for progressive multiple alignment of sequences with insertions. *Proc Natl Acad Sci U S A.* 102:10557–10562.

Loytynoja A, Goldman N. 2008. Phylogeny-aware gap placement prevents errors in sequence alignment and evolutionary analysis. *Science* 320:1632–1635.

Lutz M, Goker M, Piatek M, Kemler M, Begerow D, Oberwinkler F. 2005. Anther smuts of Caryophyllaceae: molecular characters indicate host-dependent species delimitation. *Mycol Prog.* 4: 225–238.

Maechler M, Rousseeuw P, Struyf A, Hubert M, Hornik K. 2013. cluster: cluster analysis basics and extensions. R package version 1.14.4.

Marais GAB, Nicolas M, Bergero R, Chambrier P, Kejnovsky E, Moneger F, Hobza R, Widmer A, Charlesworth D. 2008. Evidence for degeneration of the Y chromosome in the dioecious plant *Silene latifolia*. *Curr Biol.* 18:545–549.

Markova-Raina P, Petrov D. 2011. High sensitivity to aligner and high rate of false positives in the estimates of positive selection in the 12 *Drosophila* genomes. *Genome Res.* 21:863–874.

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

Nielsen R, Yang Z. 1998. Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. *Genetics* 148:929–936.

Nieuwenhuis BPS, Billiard S, Vuilleumier S, Petit E, Hood ME, Giraud T. 2013. Evolution of uni- and bifactorial sexual compatibility systems in fungi. *Heredity* 111:445–455.

Notredame C, Higgins D, Heringa J. 2000. T-Coffee: a novel method for fast and accurate multiple sequence alignment. *J Mol Biol.* 302:205–217.

Oudemans PV, Alexander HM, Antonovics J, Altizer S, Thrall PH, Rose L. 1998. The distribution of mating-type bias in natural populations of the anther-smut *Ustilago violacea* on *Silene alba* in Virginia. *Mycologia* 90:372–381.

Perlin MH. 1996. Pathovars or formae speciales of *Microbotryum violaceum* differ in electrophoretic karyotype. *Int J Plant Sci.* 157:447–452.

Petit E, Giraud T, de Vienne DM, Coelho M, Aguileta G, Amselem J, Kreplak J, Poulain J, Gavory F, Wincker P, et al. 2012. Linkage to the mating-type locus across the genus *Microbotryum*: insights into non-recombining chromosomes. *Evolution* 66:3519–3533.

Posada D. 2008. jModelTest: phylogenetic model averaging. *Mol Biol Evol.* 25:1253–1256.

Repping S. 2006. High mutation rates have driven extensive structural polymorphism among human Y chromosomes. *Nat Genet.* 38: 463–467.

Robinson MD, McCarthy DJ, Smyth GK. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26:139–140.

Salichos L, Rokas A. 2013. Inferring ancient divergences requires genes with strong phylogenetic signals. *Nature* 497:327–331.

Sawyer SA. 1989. Statistical tests for detecting gene conversion. *Mol Biol Evol.* 6:526–538.

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

Stamatakis A, Ludwig T, Meier H. 2005. RAxML-III: a fast program for maximum likelihood-based inference of large phylogenetic trees. *Bioinformatics* 21:456–463.

Steinemann M, Steinemann S. 1992. Degenerating Y chromosome of *Drosophila miranda*: a trap for retrotransposons. *Proc Natl Acad Sci U S A.* 89:7591–7595.

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

Votintseva AA, Filatov DA. 2009. 'Evolutionary Strata' in a small mating type-specific region of the smut fungus *Microbotryum violaceum*. *Genetics* 182:1391–1396.

Whittle CA, Johannesson H. 2011. Evidence of the accumulation of allele-specific non-synonymous substitutions in the young region of recombination suppression within the mating-type chromosomes of *Neurospora tetrasperma*. *Heredity* 107:305–314.

Whittle CA, Sun Y, Johannesson H. 2011. Degeneration in codon usage within the region of suppressed recombination in the mating-type chromosomes of *Neurospora tetrasperma*. *Eukaryot Cell.* 10: 594–603.

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

Yang Z, Nielsen R. 2002. Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. *Mol Biol Evol.* 19:908–917.

Yang Z, Wong WSW, Nielsen R. 2005. Bayes Empirical Bayes inference of amino acid sites under positive selection. *Mol Biol Evol.* 22: 1107–1118.

Yockteng R, Marthey S, Chiapello H, Hood M, Rodolphe F, Devier B, Wincker P, Dossat C, Giraud T. 2007. Expressed sequence tags of the anther smut fungus, *Microbotryum violaceum*, identify mating and pathogenicity genes. *BMC Genomics* 8:272.

Zakharov IA. 1986. Some principles of the gene localization in eukaryotic chromosomes. Formation of the problem and analysis of nonrandom localization of the mating-type loci in some fungi. *Genetika* 22: 2620–2624.

Zhang J, Nielsen R, Yang Z. 2005. Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. *Mol Biol Evol.* 22:2472–2479.

2.3 Article 2 : Chaos of rearrangements in the mating-type chromosomes of the anther-smut fungus *Microbotryum lychnidis-dioicae*.

L'étude précédente a permis de montrer que les chromosomes de types sexuels de *M. lychnidis-dioicae* et de plusieurs autres espèces du complexe *M. violaceum* présentaient des signatures de dégénérescence similaires à celles observées chez les chromosomes sexuels, et que cette dégénérescence n'était pas asymétrique, conformément à ce qu'on attend dans un système où il n'y a pas d'asymétrie d'hétérozygotie entre les type sexuels. Nous avons pu déterminer quelles séquences génomiques appartenaient aux zones non-recombinantes ou recombinantes des chromosomes de types sexuels, aux régions pseudo-autosomales ou aux autosomes.

Cependant, l'assemblage restant très fragmenté (plus de 1000 contigs pour environ 26 Mb), et les scaffolds des régions non-recombinantes n'ayant pas pu être ancrés physiquement sur les chromosomes, nous n'avons pas pu tester l'existence de strates évolutives ni étudier les réarrangements au sein de la région non-recombinante. Whittle et al [80] ont publié en 2015 une étude comparant plusieurs génomes de *M. lychnidis-dioicae*. Ils ont réévalué à la hausse la proportion non-recombinante des chromosomes de types sexuels par rapport à leur étude précédente [18], mais ont estimé que la plus grande partie de la région non-recombinante présentait une faible divergence entre les types sexuels, et proviendrait donc d'une expansion massive et récente de la suppression de recombinaison. S'appuyant sur un assemblage très fragmenté des chromosomes a_1 et a_2 et sur des données de reséquençage d'individus de *M. lychnidis-dioicae*, ils ont inféré la présence d'au moins trois strates évolutives, correspondant à trois épisodes distincts de suppression de recombinaison, selon le scénario suivant :

1. Suppression très ancienne [38] de la recombinaison autour du locus de récepteur à phéromones.

2. Fusion des deux loci MAT, HD et PR (distants de 60 kb d'après leur étude), lors du passage l'état tétrapolaire à l'état bipolaire chez un ancêtre de *M. lychnidis-dioicae*, entraînant la formation d'une 2ème strate de suppression de recombinaison.

3. Expansion massive et récente (au plus quelques millions d'années d'après un modèle d'horloge moléculaire) de la suppression de recombinaison, peut-être en réponse à une

pression de sélection pour lier le centromère et le locus MAT contenant les deux facteurs de types sexuels.

Dans l'étude de Whittle et al [80], l'identification des régions homologues entre chromosomes des types sexuels a_1 et a_2 et le calcul de la divergence entre ces régions ont été effectués uniquement à partir des alignements des séquences d'ADN, sans utiliser l'alignement des séquences protéiques qui aurait pu identifier des régions homologues plus divergentes. Leur méthode est donc biaisée vers l'identification de régions faiblement divergentes, ce qui a pu entraîner une sous-estimation du niveau de divergence entre les chromosomes de types sexuels. De plus, leur assemblage des chromosomes a_1 et a_2 reste très fragmenté, ce qui limite la possibilité d'élaborer des scénarios sur l'histoire de la suppression de recombinaison.

Dans l'article suivant, nous présentons un assemblage complet des chromosomes de types sexuels de l'espèce *M. lychnidis-dioicae*, obtenu grâce aux améliorations récentes de la technologie PacBio et à des cartes optiques générées précédemment [39].

Dans cette étude, j'ai réalisé toute l'analyse des assemblages ; j'ai identifié les régions et gènes homologues entre les chromosomes a_1 et a_2, mesuré le niveau de divergence synonyme entre ces allèles, caractérisé le contenu en gènes des chromosomes de types sexuels et identifié les gènes présents à l'état hémizygote (c'est-à-dire perdus dans un des deux chromosomes de types sexuels). J'ai étudié les réarrangements entre a_1 et a_2. J'ai aussi identifié et caractérisé la structure des centromères, dont certains étaient entièrement séquencés, et identifié des motifs spécifiques des centromères ou des régions sub-télomériques. Les figures supplémentaires sont en annexe A.2 (page 240).

Chaos of Rearrangements in the Mating-Type Chromosomes of the Anther-Smut Fungus *Microbotryum lychnidis-dioicae*

Hélène Badouin,*,†,1 Michael E. Hood,‡,1 Jérôme Gouzy,§, Gabriela Aguileta,*,† Sophie Siguenza,§,** Michael H. Perlin,†† Christina A. Cuomo,‡‡ Cécile Fairhead,†,§§ Antoine Branca,*,† and Tatiana Giraud*,†,2**

*Ecologie, Systématique et Evolution, Univ Paris-Sud, 91405 Orsay, France, †Centre National de la Recherche Scientifique, F-91405 Orsay, France, ‡Department of Biology, Amherst College, Amherst, Massachusetts 02142, §INRA and **Centre National de la Recherche Scientifique, Laboratoire des Interactions Plantes-Microorganismes, 31326 Castanet-Tolosan, France, ††Department of Biology, Program on Disease Evolution, University of Louisville, Louisville, Kentucky 40292, ‡‡Broad Institute of MIT and Harvard, Cambridge, Massachusetts 02142, and §§Institut de Génétique et Microbiologie, Université Paris-Sud, F-91405 Orsay cedex, France

ORCID IDs: 0000-0002-2456-5968 (H.B); 0000-0002-5778-960X (C.A.C).

ABSTRACT Sex chromosomes in plants and animals and fungal mating-type chromosomes often show exceptional genome features, with extensive suppression of homologous recombination and cytological differentiation between members of the diploid chromosome pair. Despite strong interest in the genetics of these chromosomes, their large regions of suppressed recombination often are enriched in transposable elements and therefore can be challenging to assemble. Here we show that the latest improvements of the PacBio sequencing yield assembly of the whole genome of the anther-smut fungus, *Microbotryum lychnidis-dioicae* (the pathogenic fungus causing anther-smut disease of *Silene latifolia*), into finished chromosomes or chromosome arms, even for the repeat-rich mating-type chromosomes and centromeres. Suppressed recombination of the mating-type chromosomes is revealed to span nearly 90% of their lengths, with extreme levels of rearrangements, transposable element accumulation, and differentiation between the two mating types. We observed no correlation between allelic divergence and physical position in the nonrecombining regions of the mating-type chromosomes. This may result from gene conversion or from rearrangements of ancient evolutionary strata, *i.e.*, successive steps of suppressed recombination. Centromeres were found to be composed mainly of copia-like transposable elements and to possess specific minisatellite repeats identical between the different chromosomes. We also identified subtelomeric motifs. In addition, extensive signs of degeneration were detected in the nonrecombining regions in the form of transposable element accumulation and of hundreds of gene losses on each mating-type chromosome. Furthermore, our study highlights the potential of the latest breakthrough PacBio chemistry to resolve complex genome architectures.

KEYWORDS *Microbotryum violaceum*; finished genome assembly; intratetrad mating; MAT; basidiomycete; selfing; bipolarity

SEX chromosomes often show exceptional genome features with extensive suppression of homologous recombination and cytological differentiation between members of the diploid chromosome pair (Bergero and Charlesworth 2009). Determination of male *vs.* female development by such nonrecombining sex chromosomes has evolved independently in diverse lineages, including mammals, birds, fishes insects, vascular plants, bryophytes, and macroalgae (Ohno 1967; Charlesworth 1991; Itoh *et al.* 2006; Yamato *et al.* 2007; Marais *et al.* 2008; Bergero and Charlesworth 2009; Cock *et al.* 2010; Kaiser and Bachtrog 2010; Page *et al.* 2010; Bachtrog *et al.* 2011; McDaniel *et al.* 2013), where incorporation of multiple genes responsible for sexually antagonistic traits is a common explanation for expansion of recombination suppression (Rice 1984; Lahn and Page 1999). Despite the fundamental roles that sex

doi: 10.1534/genetics.115.177709
Manuscript received May 2, 2015; accepted for publication June 2, 2015; published Early Online June 3, 2015.
Available freely online through the author-supported open access option.
Supporting information is available online at www.genetics.org/lookup/suppl/doi:10.1534/genetics.115.177709/-/DC1.
Sequence data from this article have been deposited with EMBL-ENA under accession no. PRJEB7910.
[1]These authors contributed equally to this work.
[2]Corresponding author: Laboratoire Ecologie, Systématique et Evolution, Bâtiment 360, Université de Paris-Sud, 91405 Orsay cedex, France. E-mail: tatiana.giraud@u-psud.fr

chromosomes play, their genic content is particularly prone to degenerative influences of recombination suppression and enforced heterozygosity (Bergero and Charlesworth 2009).

In fungi, mating-type chromosomes contain the genes for molecular mechanisms of mating compatibility (*e.g.*, via pheromones and receptors), and they can display recombination suppression and size dimorphism analogous to sex chromosomes (Bakkeren and Kronstad 1994; Hood 2002; Fraser *et al.* 2004; Fraser and Heitman 2004; Hood *et al.* 2004, 2013; Fraser and Heitman 2005; Menkis *et al.* 2008; Ellison *et al.* 2011; Grognet *et al.* 2014), as well as degeneration (Hood *et al.* 2004; Whittle and Johannesson 2011; Whittle *et al.* 2011; Fontanillas *et al.* 2015). Grouped together, heteromorphic sex chromosomes and mating-type chromosomes can therefore be considered allosome pairs (Montgomery 1911), which are similarly maintained and influenced by their central role in regulating gamete fusion, contrasted by the homomorphic autosomes. Nonrecombining fungal mating-type chromosomes can also show footprints of degeneration (Hood *et al.* 2004; Whittle and Johannesson 2011; Whittle *et al.* 2011; Fontanillas *et al.* 2015). However, important differences exist between sex and mating-type chromosomes: because mating types are expressed as non-self-recognition at the haploid stage, they are found only in the heterogametic condition in diploids (similar to macroalgae and bryophytes), and thus with symmetrical roles, and mating types are not associated with male/female functions in fungi (Billiard *et al.* 2011). These common features and differences help make fungal mating-type chromosomes valuable models for the broader phenomenon of allosome evolution, with some shared and some distinguishing expectations compared to sex chromosomes, *e.g.*, on the asymmetry of allosome degeneration (Bull 1978) and on the role of sexually antagonistic selection in driving suppressed recombination (Bergero and Charlesworth 2009).

Examples of nonrecombining fungal allosomes include the mating-type chromosomes of *Microbotryum lychnidis-dioicae* (Hood 2002) and the fungus-causing anther-smut disease on *Silene latifolia* (Votintseva and Filatov 2009; Hood *et al.* 2013; Fontanillas *et al.* 2015; Perlin *et al.* 2015; Whittle *et al.* 2015). These nonrecombining fungal mating-type chromosomes have been extensively studied, and degeneration has been reported in the form of transposable element (TE) accumulation and high rates of nonsynonymous substitutions (Hood *et al.* 2004; Fontanillas *et al.* 2015). The buildup of TEs in regions of suppressed recombination has, however, prevented genome assembly for these mating-type chromosomes. Therefore, the exact content in genes and transposable elements, the extent of the nonrecombining regions (NRRs), and the existence of evolutionary strata of recombination suppression have remained elusive (Votintseva and Filatov 2009; Hood *et al.* 2013; Fontanillas *et al.* 2015; Perlin *et al.* 2015; Whittle *et al.* 2015).

The spread of recombination suppression outward from key compatibility-determining loci to form evolutionary strata of differentiation is described for plant and animal sex chromosomes (Bergero and Charlesworth 2009) and may also occur on fungal mating-type chromosomes (Fraser *et al.* 2004; Fraser and Heitman 2004, 2005; Menkis *et al.* 2008; Votintseva and Filatov 2009). Suppressed recombination has been shown to link the two mating-type-determining loci (*i.e.*, encoding mating pheromone/receptor and homeodomain proteins, respectively), including basidiomycete fungal pathogens of plants and humans (*Ustilago hordei* and *Cryptococcus neoformans*, respectively) (Bakkeren and Kronstad 1994; Fraser *et al.* 2004). This linkage of two mating-type-determining loci may be advantageous under selfing mating systems, where it increases the compatibility of gamete combinations (Billiard *et al.* 2011; Nieuwenhuis *et al.* 2013). In other fungi, suppressed recombination links mating-type loci to the centromere, causing segregation of mating types in the first meiotic division. This is associated with automictic reproduction (mating within the meiotic tetrad), *e.g.*, in *M. lychnidis-dioicae* and *Neurospora tetrasperma* (Zakharov 1987; Hood and Antonovics 2000; Giraud *et al.* 2008; Menkis *et al.* 2008). Such intratetrad mating may favor successive linkage to mating type for a suite of genes that experience deleterious, recessive mutations, as shown by theoretical models (Antonovics and Abrams 2004; Johnson *et al.* 2005) (Supporting Information, Figure S1). Indeed, deleterious mutation occurring at the margin of the nonrecombining regions may be partially sheltered in a heterozygous state due to less frequent recombination, and extension of the region of suppressed recombination would then be selected for permanent sheltering of these deleterious mutations (Ironside 2010). This could lead to the formation of evolutionary strata, reflected in increasing allelic differentiation at physical positions along the chromosome when coming closer to the mating-type genes. Although in plants and in animals evolutionary strata are generally thought to be due to sexually antagonistic selection, the definitive linkage of partially linked deleterious alleles to the sex-determining region has also been proposed as an alternative explanation (Ironside 2010). Elucidating the existence of evolutionary strata in fungal mating-type chromosomes may thus generally provide insights on the shared evolutionary processes acting on allosomes.

In this research, we aimed to produce a finished assembly of the mating-type chromosomes of the anther-smut fungus *M. lychnidis-dioicae* for elucidating the extent of suppressed recombination, transposable element accumulation, gene content, and rearrangements and whether evolutionary strata are present. For this goal, we took advantage of the recent developments of the PacBio sequencing technology, which allowed full-length physical mapping of the mating-type chromosomes and the resolution of these important genomic properties.

Materials and Methods

DNA extraction and sequencing

DNA was extracted with the Qiagen Kit 10243 (Courtaboeuf, France) following manufacturer instructions and using a Carver hydraulic press (reference 3968, Wabash, IN) for breaking cell

walls. Haploid a_1 and a_2 genomes of the Lamole reference strain of *M. lychnidis-dioicae* were sequenced separately using the Pacific Bioscience (PacBio) method (Institute for Genomic Medicine, University of California, San Diego, La Jolla, CA).

Assembly and annotation

The PacBio P5/C3 sequencing of haploid a_1 and a_2 genomes produced 658,501 and 648,664 reads, respectively (mean read length: 7.925/7.569; coverage: 180×/160×). Independent assemblies of the a_1 and a_2 genomes were generated with the wgs-8.2beta version of the PBcR assembler (Koren *et al.* 2012), with the following parameters: -pbCNS -length 500 -partitions 200 genomeSize = 27000000 -maxCoverage 150. Contigs were aligned with optical maps of the two mating-type chromosomes (Hood *et al.* 2013) with MapSolver software (OpGen). The alignment of the contig corresponding to the a_2 mating-type chromosome with the corresponding optical map was highly significant. The a_1 optical map was used to create an oriented a_1 pseudomolecule composed of two contigs for each chromosome arm, aligned on the optical map, plus a 50-kb unanchored contig in the centromeric region. The two contigs corresponding to the two a_1 chromosome arms produced full-length alignments that matched only with the a_1 mating-type chromosome optical map; alignment scores of the contigs were 46 and 143, exceeding the default MapSolver threshold range for significance of 3 to 6.

Mauve was used to compare the two assemblies (Darling *et al.* 2010) (Figure S6). In this genome comparison, a contig belonging to autosomes in one assembly was found to be a subsequence of a contig in the second assembly (and vice versa). Manual curation was performed to select the largest contigs and remove the few short degenerate contigs, probably resulting from the incorrect editing of reads (Figure S6). Comparison with the reference sequence (GenBank no. NC_020353) revealed that the mitochondrial genome was not correctly solved by the assembler, and this genome sequence was therefore removed from the assembly. The PacBio assembly was compared with the Illumina-based assembly available from the Broad Institute and the Illumina reads generated in a previous study (Perlin *et al.* 2015). The PacBio assembly was edited at positions for which both the genome alignment obtained with Mugsy (Angiuoli and Salzberg 2011) and a standard variant-calling method based on read mapping (samtools, bcftools, and VarScan) (Li *et al.* 2009; Koboldt *et al.* 2012) gave concordant findings for the polymorphism. This conservative approach led to the editing of 1332 SNPs, 456 insertions, and 19,528 deletions. The 11,057 protein-coding gene models were predicted with EuGene (Foissac *et al.* 2008), trained for *Microbotryum*. Available *Microbotryum* EST data (Aguileta *et al.* 2010) and similarities to the fungi subset of the uniprot database (Consortium 2011) were integrated into EuGene for gene prediction. InterPro (Mitchell *et al.* 2014) was used to identify protein domains and families and to infer an automatic functional annotation from protein domain content.

Analysis of genome structure

Repetitive DNA content was analyzed with RepeatMasker (Smit 2013), using REPBASE v19.11 (Jurka 1998). We searched for subtelomeric motifs with Meme (Bailey *et al.* 2006) (Figure S2). Figure 1, Figure 2, and Figure S5 were prepared with Circos (Krzywinski *et al.* 2009). We identified syntenic blocks from the alignment of the two chromosomes with Mauve, using standard parameters (Darling *et al.* 2010). We analyzed gene order after filtering out transposable elements (TEs) to identify larger blocks of synteny. Alleles were assigned between the two mating-type chromosomes by applying orthomcl (Li *et al.* 2003) to the protein data sets for unique a_1 and a_2 orthologs. The signed permutations were used to calculate minimal reversal distances with Grimm (Tesler 2002). The sequences of TE-like copias and minisatellites specific for the centromeres are provided in File S1.

Divergence between a_1 and a_2 alleles

Alleles were identified as indicated above, and MACSE (Ranwez *et al.* 2011) was used to align DNA-coding sequences, respecting phase. Synonymous divergence and its standard error were estimated with the yn00 program of the PAML package (Yang 2007). The correlation between d_S and physical distance was assessed with JMP (SAS Institute). As we have here a single individual genome sequence, some of the substitutions between a_1 and a_2 identified here could theoretically include within-species polymorphism rather than fixed differences between the a_1 and a_2 alleles. However, this should not be a problem for identifying the nonrecombining regions from the pseudo-autosomal regions (PARs). Indeed, given its highly selfing mating system, *M. lychnidis-dioicae* is largely homozygous in regions unlinked to mating type (Giraud *et al.* 2008; Vercken *et al.* 2010), so that high levels of synonymous divergence between two a_1 and a_2 alleles from a given diploid individual reliably indicate suppressed recombination.

Data access

The assembly is available at EMBL-ENA (accession no. PRJEB7910).

Results and Discussion

Finished genome assembly and new insights about centromere and telomere structures

The recent PacBio P5/C3 chemistry produced long reads, averaging 8 kb. The imperfect quality of PacBio reads was fully compensated by >100× sequencing depth, as illustrated by the almost complete assembly of the *M. lychnidis-dioicae* genome and including the mating-type chromosomes (Figure 1 and Figure 2), while several previous efforts and chemistries had been unsuccessful (Votintseva and Filatov 2009; Hood *et al.* 2013; Fontanillas *et al.* 2015). The PacBio genome assembly indeed yielded 22 contigs, including 18 autosomal contigs, 1 contig for the a_2 mating-type

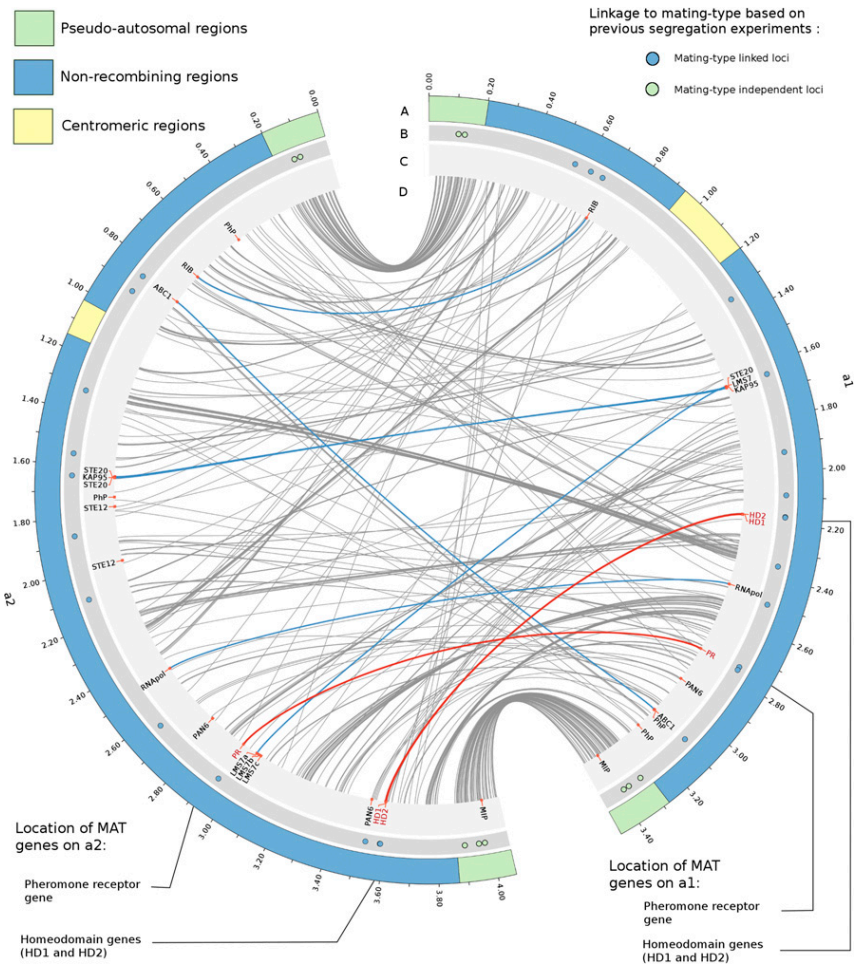


Figure 1 Finished assembly of the mating-type chromosomes of the anther-smut fungus *M. lychnidis-dioicae*. Tracks A–D show the location of different genomic elements. A: Structure of the chromosomes, with the PARs in green, the NRRs in blue, and the centromeres in yellow. B: Location of loci that have been shown to be linked (blue circles) or unlinked (green circles) to mating type by previous segregation analyses (Votintseva and Filatov 2009; Abbate and Hood 2010; Petit *et al.* 2012). C: Location of the genes related to the mating-type function—pheromone receptor (PR) and HD homeodomain genes (in red), of the other genes likely involved in mating (STE12, STE20, and the precursors of pheromones, PhP), and of the genes located around the pheromone receptor gene in the closely related *S. salmonicolor* (Coelho *et al.* 2010) (KAP95, RNAPol, RIB, and ABC1). D: Links between orthologous alleles of a_1 and a_2, in red for HD and PR and in blue for KAP95, RNAPol, RIB, and ABC1.

chromosome, and 3 contigs for the a_1 mating-type chromosome that could be assembled in a scaffold (see below) (Figure 2, Table 1). The total size of the assembly was 33.3 Mb, *i.e.*, almost 7 Mb larger than that of the very same a_1 haploid *M. lychnidis-dioicae* strain sequenced at the Broad Institute using Illumina technology (Perlin *et al.* 2015), as the long reads of the PacBio chemistry allowed assembly by more complete incorporation of repetitive DNA elements and included both mating-type chromosomes. In particular, putative centromeric regions could be assembled and were identified as the most repeat-rich regions, where we found two copia-like TEs and two minisatellite repeats specifically clustering in these regions at one location per contig (tracks E and F in Figure 2). Both copia-like TEs possessed two open reading frames, coding for an integrase and for a reverse transcriptase, but no long terminal repeats, and many copies were truncated. The minisatellites consisted of a 66-bp repeat motif of GGCCCA and a 110-bp repeat motif of CGACGG; the 66-bp repeat was specific for centromeric regions and identical in all iterations. The structures of the finished centromeres are shown in Figure 3. Eleven of the contigs had these specific repeats at one end, suggesting that some autosomes were split at their centromeres in the assembly (Figure 2). On finished chromosomes, the centromeric regions spanned 65–155 kb. Centromeric regions have been elucidated in a single other basidiomycete fungus so far, *Cryptococcus neoformans* var. grubii, in which they were instead enriched in Tcn transposons (Janbon *et al.* 2014).

We found no typical telomeric repeats (TTAGGG) while they were identified at the edge of five scaffolds in the Broad Institute *M. lychnidis-dioicae* genome (Perlin *et al.* 2015). Here, we identified instead a specific 50-bp motif repeated at some, but not all, chromosome ends (Figure 2, Figure S2), which thus likely represent a subtelomeric motif. This motif was in fact found near the TTAGGG telomeric repeats (80–1500 bp apart) in two of the five contigs from the Broad Institute containing TTAGGG repeats (the three other contigs being only 500–850 bp long). This suggests that PacBio technology does not sequence the very edges of telomeres well and that other motifs may be useful for detecting ends of chromosomes.

Previous pulse-field gel electrophoresis estimated 11 autosomes for the haploid stage (Perlin *et al.* 2015), corresponding precisely to the 5 complete autosomes plus the 6 autosomes as two contigs split in their centromeric regions (the S18 contig corresponded only to a telomeric region). The autosomes were thus assembled into full or two-arm contigs (Figure 2). The a_2 mating-type chromosome was

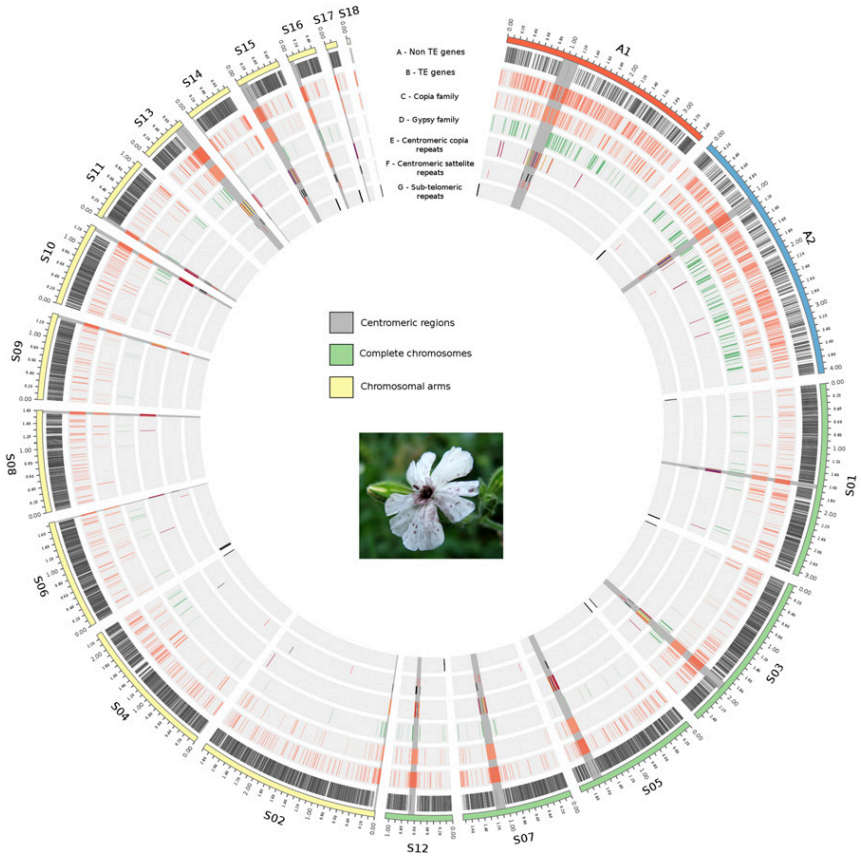


Figure 2 PacBio assembly of the nuclear *M. lychnidis-dioicae* genome. The picture shows the fungal dark spores in the anthers of a *S. latifolia* flower, replacing the pollen. The mating-type chromosomes (blue) and finished autosomes (green) or autosome arms (yellow) are represented. The identified centromeric regions are shaded. Tracks A–G show the location of different genomic elements. A: Genes after filtering out TEs. B: Genes encoding TEs. C: TEs from the copia family. D: TEs from the gypsy family. E: TEs from the copia subfamily particularly abundant at centromeres (Figure 3). F: Satellite repeat of the type particularly abundant at centromeres (Figure 3). G: Subtelomeric motifs (Figure S2).

completely assembled from PacBio sequences, including the submetacentric centromeric region. Available optical maps (Hood *et al.* 2013), *i.e.*, ordered, genome-wide restriction maps obtained from single DNA molecules, allowed orienting the two contigs covering the arms of the a_1 mating-type chromosome and attributing a 50-kb contig composed of repeated DNA to its centromeric region. Collinearity between the optical maps and the contigs confirmed the high quality of the assembly (Figure S3). The combination of PacBio chemistry and optical maps thus revealed the right approach to finally resolve the mating-type chromosome structure and content.

Identification of pseudo-autosomal regions and extensive nonrecombining regions on the mating-type chromosomes

Assemblies of the dimorphic *M. lychnidis-dioicae* mating-type chromosomes (a_1, 3.5 Mbp; a_2, 4.0 Mbp) encompassed 614 and 683 predicted genes, respectively, filtering out TEs. Among these genes, only a few had putative functions known to be involved in mating (Petit *et al.* 2012; Fontanillas *et al.* 2015), *i.e.*, the PR genes encoding the pheromone receptors, the PhP genes encoding the precursors of pheromones, the HD1 and HD2 homeodomain genes, STE20 encoding a protein kinase regulating mating (Smith *et al.* 2004), and STE12 encoding a transcription factor regulating mating and invasive hyphal growth in fungal pathogens (Hoi and Dumas 2010) (Figure 1). Actually, two homologs of STE12 were found, both present only on the a_2 mating-type chromosome, while absent from the a_1 mating-type chromosome; this seems consistent with previous observations that initiation of the conjugation tubes is earlier and to a greater extent from cells of the a_2 mating type in *M. lychnidis-dioicae* (Day 1976; Xu *et al.* 2015). The PhP pheromone genes were recently characterized as encoding proteins with the typical C terminus CAAX motif and producing functional pheromone peptides triggering the mating response in several *Microbotryum* species (Xu *et al.* 2015).

We identified 305 predicted genes shared between the a_1 and a_2 mating-type chromosomes. Synonymous divergence between a_1 and a_2 alleles of these shared genes, when plotted along their physical positions on both mating-type chromosomes, clearly identified the nonrecombining regions; suppressed recombination can lead to divergence between alleles while recombination homogenizes allele sequences. Synonymous divergence values between alleles confirmed the existence of recombining regions at both ends of the mating-type chromosomes, called PARs (Figure 4) (Votintseva and Filatov 2009; Hood *et al.* 2013). The PARs displayed almost no substitutions between a_1 and a_2 alleles (Figure 4) and included loci that have previously been shown to be unlinked to the mating type (Votintseva and Filatov 2009) (Figure 1). PARs on the short arm (pPAR) and long arm (qPAR) of the mating-type chromosomes spanned 0.20 and 0.19 Mb,

Table 1 Assembly statistics of the *M. lychnidis-dioicae* genome

No. of contigs	22
Minimum contig size	44,493
Maximum contig size	4,061,474
N50 (bp)	2,275,168
N50 (no. of contigs)	6
N90 (bp)	1,030,812
N90 (no. of contigs)	15
Mean contig size	1,505,578.59
Median contig size	1,345,833
Total length of the assembly (bp)	33,122,729

The PacBio genome assembly yielded 22 contigs—18 for autosomes, 1 for the a_2 mating-type chromosome, and 3 for the a_1 mating-type chromosome—that could be assembled in a single scaffold.

respectively, and contained 102 genes. PAR boundaries corresponded precisely to the two distal regions of collinearity between a_1 and a_2 mating-type chromosomes on the restriction digest optical maps (Hood *et al.* 2013).

The NRRs were characterized by high divergence between the a_1 and a_2 alleles of the single-copy genes (Figure 4 and Figure S4). Fourteen widely distributed loci known to cosegregate with mating-type (Votintseva and Filatov 2009; Abbate and Hood 2010; Petit *et al.* 2012) were found within these divergent regions, validating the NRRs (Figure 1). The a_1 and a_2 NRRs spanned the majority of the mating-type chromosomes (>88%), covering 3.08 Mb with 506 genes and 3.67 Mb with 580 genes for the a_1 and a_2 mating-type chromosomes, respectively (Figure 1), confirming conclusions based on optical maps and marker segregation (Hood *et al.* 2004, 2013; Fontanillas *et al.* 2015) and contrasting with reports of smaller regions of suppressed recombination (Votintseva and Filatov 2009; Whittle *et al.* 2015).

Extensive rearrangements between mating-type chromosomes and degeneration

Comparison of aligned sequences between the a_1 and a_2 NRRs revealed extensive rearrangement (Figure 1 and Figure S5), probably reflecting and contributing to the current degree of suppressed recombination. We identified 40 syntenic blocks of shared genes in the NRRs, encompassing 2–18 genes. At least 210 inversion events were inferred to account for these rearrangements (Figure 1 and Figure S5). In contrast, all autosomes were collinear (Figure S6). Suppressed recombination on fungal mating-type chromosomes has been shown to be associated with rearrangements in *N. tetrasperma* (Ellison *et al.* 2011) and *C. neoformans* (Fraser *et al.* 2004; Fraser and Heitman 2004, 2005) or to occur despite conserved collinearity in *Podospora anserina* (Grognet *et al.* 2014). The lack of collinearity across several megabases and patterns of major inversions or translocations in *M. lychnidis-dioicae* is, however, to an exceptional degree among fungal mating-type chromosomes.

The two mating-type loci typical of basidiomycete fungi (*i.e.*, the mating pheromone receptor determining premating recognition and the homeodomain proteins determining postmating compatibility) were 0.60 Mb apart on the a_1 and a_2 mating-type chromosomes, while inverted in orientation (Figure 1 and Figure S5). This distance is much greater than the one estimated recently based on Illumina sequencing (56 kb) (Whittle *et al.* 2015), again showing the great advantage of PacBio sequencing for chromosome assembly. These mating-type loci were not near the edges of the NRRs, unlike other basidiomycete fungi with linkage suppression in their mating-type regions, *i.e.*, *U. hordei* (Bakkeren and Kronstad 1994) and *C. neoformans* (Fraser *et al.* 2004). The mating-type loci were instead on the same side of the centromere, on the long arm of the chromosomes. The pheromone precursor loci were linked to the receptor loci in the NRR, but were up to 1.4 Mb away, on opposite sides of the centromere on the a_2 mating-type chromosome and the same side in the a_1 mating-type chromosome. This distance between pheromone receptor and pheromone is also a novel observation among fungi. Despite rearrangements in NRRs, several genes close to the pheromone receptor genes in the closely related fungus *Sporidiobolus salmonicolor* (Microbotryomycetes) (Coelho *et al.* 2010) were also found on the mating-type chromosomes of *M. lychnidis-dioicae* (KAP95, RNAPol, RIB, and ABC1, Figure 1). In contrast, the homologs of the 40 predicted genes in the scaffold containing the HD locus in *S. salmonicolor* with blast hits in *M. lychnidis-dioicae* were all found on autosomes but one, and 34 of them on a single autosome arm (S09). Interestingly, we also found on the S09 contig a sequence with similarity to one of the two homeodomain genes (HD1). The possibility that the HD locus may have duplicated to a location in linkage with the PhP/PR locus warrants further study as a potential transition to mating-type bipolarity within the Microbotryomycetes.

The high degree of rearrangements between the two mating-type chromosomes in *M. lychnidis-dioicae* prevents us from obtaining definitive evidence for the evolutionary cause of the expansive nature of recombination suppression; nevertheless, the linkage of mating-type loci and the centromere within NRRs indicates that first-division segregation of both mating-type loci during meiosis may have been a primary evolutionary factor in recombination suppression, as suggested for other automictic fungi (Zakharov 2005).

As expected for nonrecombining regions, the NRRs had a higher frequency of repetitive elements than the PARs and autosomes (Figure 2). Repetitive elements accounted for 16.0 and 18.7% of the a_1 and a_2 NRRs, respectively, explaining why they have remained unassembled until now. Most of the repetitive elements were long terminal repeat retro-elements of the copia and gypsy families, as found previously (Hood *et al.* 2005; Fontanillas *et al.* 2015). In addition, gene density was lower in the NRRs (25.8% were coding sequences) than in the autosomes (62.7%) or PARs (51.2%).

The almost finished assembly of the genome allowed detection of 245 and 289 genes in a hemizygous state on the mating-type chromosomes a_1 and a_2, respectively, suggesting that functional copies have been lost from a single one of the mating-type chromosomes or have degenerated to

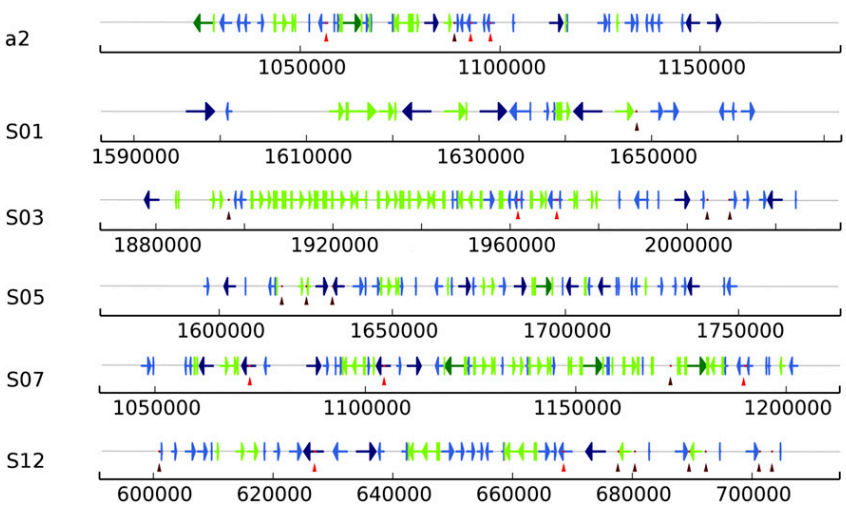


Figure 3 Structure of the centromeres. Localization of copia-like transposable elements and specific mini-satellite repeats in the finished centromeric regions (scaffolds names are indicated as in Figure 1). Full-length copies of two types of copia-like transposable elements are shown in blue and green. Lighter shades indicate incomplete copies of these two types. Arrows indicate the orientation of copies. Dots and vertical arrows indicate repeats of the 66-bp motif minisatellite (GGCCCA)n and of the 110-bp motif minisatellite (CGACGG)n in brown and red, respectively.

a great extent. None of these hemizygous genes had putative functions known to be involved in mating except the STE12 homologs on the a_2 mating-type chromosome. This degree of gene losses on mating-type chromosomes appears unprecedented among fungi. Together with the accumulation of TEs in NRRs, low gene density and hemizygous loss of genes constitute signs of advanced mating-type chromosome degeneration, as expected for long-term, nonrecombining regions (Bergero and Charlesworth 2009). Degeneration has also been reported for the mating-type chromosome NRRs of *M. lychnidis-dioicae* in terms of higher rates of nonsynonymous substitutions than for recombining regions, and these degenerative characteristics were shown to occur in homologous sequences from several other *Microbotryum* species (Fontanillas *et al.* 2015).

The a_2 mating-type chromosome was 590 kb larger than the a_1 mating-type chromosome, where TEs accounted for 26.4% (154 kb) and non-TE genes for 6.8% (40 kb) of the difference. In particular, gypsy copies were more abundant on a_2 than a_1 mating-type chromosome (129 *vs.* 93 copies, Figure 2). However, intergenic, nonrepetitive DNA accounted for the majority of the difference in size (59%, or 0.347 kb). This finding supports previous conclusions that TE accumulation and degeneration are not asymmetrical between the mating-type chromosomes (Fontanillas *et al.* 2015), as expected from their symmetrical role in mating (Bull 1978).

Ancient recombination suppression without clear evolutionary strata despite heterogeneity in a_1–a_2 divergence

Divergence between a_1 and a_2 alleles of predicted genes in the NRRs was substantial (mean synonymous substitution rates: $d_S = 0.064$, Figure 4), indicating relatively ancient recombination suppression. However, this divergence was lower than for the pheromone receptor gene, which was previously shown to constitute the oldest trans-specific polymorphism in any organism so far (Devier *et al.* 2009).

In the evaluation of evolutionary strata within NRRs, we quantified synonymous divergence (d_S) along the mating-type chromosomes, and there was no correlation with physical distance to the PAR boundary or to the mating-type loci on either the a_1 or a_2 chromosomes (Figure 3). In particular, the correlations between d_S and gene order from PAR were nonsignificant (on a_1, $r = 0.06$, $P = 0.36$; on a_2, $r = 0.02$, $P = 0.77$), as were correlations between d_S and gene order from the nearest mating-type gene (on a_1 $r = -0.04$, $P = 0.58$; on a_2 $r = 0.02$, $P = 0.77$). There was a particularly sharp boundary between the pPARs, with zero synonymous divergence between a_1 and a_2 alleles, and the adjacent NRRs, with extensive divergence (mean $d_S = 0.064$). Thus, there was no evidence of evolutionary strata.

However, given that there is no homogametic diploid in heterothallic fungi (*i.e.*, no a_1a_1 or a_2a_2 genotypes), which tends to retain chromosome structure in homogametic sex chromosomes of plants and animals (X or Z chromosomes), the mating-type chromosomes may more rapidly undergo rearrangements that obscure any history of evolutionary strata, had they existed. Actually, in XY or ZW systems, evolutionary strata can be inferred only by using as reference the less rearranged homogametic sex chromosome (Lahn and Page 1999).

It is possible that we may have detected one recent mini-stratum as a very small region (23 kb) at the edge of the NRR close to the qPAR that was collinear between a_1 and a_2 mating-type chromosomes but with a high rate of synonymous substitutions between mating types. This region was separated from the qPAR by a rearranged region rich with TEs (Figure S7). This pattern is consistent with a recent incorporation of a PAR fragment into the NRR, which would support the scenario of an extension of the NRR, and thus evolutionary strata.

Allele divergence between mating types in NRRs was characterized by heterogeneous values (Figure 4 and Figure S4), which may result from a continuing process of translocations between the chromosomes or the remnants of ancient evolutionary strata that have been scrambled by rearrangements. Alternatively, gene conversion may have played a role by recurrently resetting divergence to zero in some stretches of DNA.

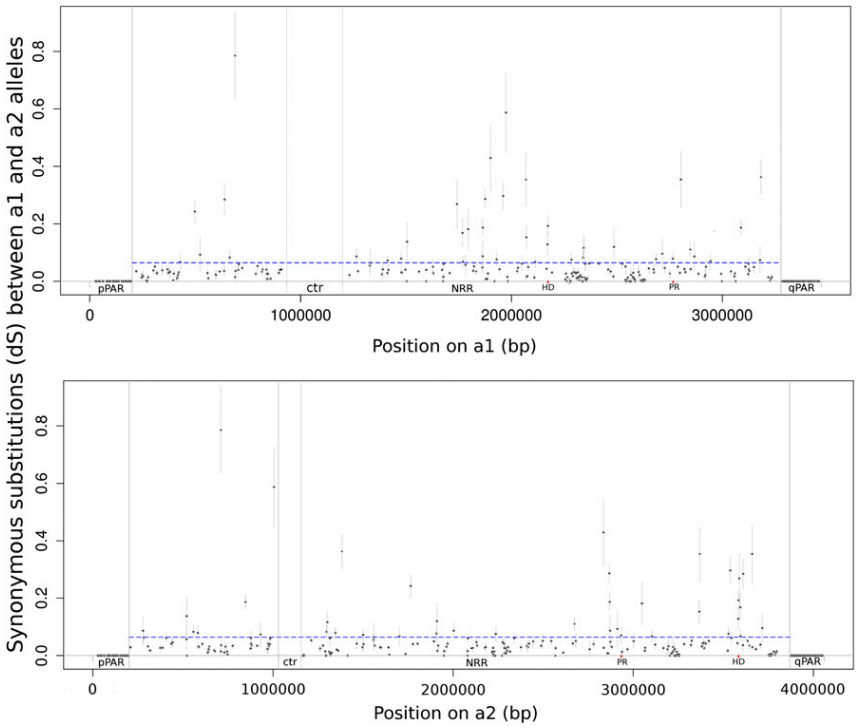


Figure 4 Divergence between a_1 and a_2 mating-type chromosomes in *M. lychnidis-dioicae*. (A) Synonymous divergence $d_S \pm$ SE is plotted against the genomic coordinates of the a_1 (A) and a_2 (B) mating-type chromosomes for all nontransposable element genes shared by the mating-type chromosomes. The boundaries between the PARs and the NRRs are indicated, as well as the locations of the mating-type loci (PR: pheromone receptor gene; HD1 and HD2: homeodomain genes). The mean value of d_S in the NRRs is shown as a blue dotted line. The pheromone receptor could not be plotted as its a_1–a_2 divergence was too extensive to be aligned in nucleotides.

Conclusions

The nearly finished assembly of the *M. lychnidis-dioicae* genome allowed the resolution of controversies over the extent of the nonrecombining regions of the mating-type chromosomes, the degree of divergence between the a_1 and a_2 mating-type chromosomes, the patterns of evolutionary strata, and the mechanisms of evolution to bipolarity (Votintseva and Filatov 2009; Hood *et al.* 2013; Fontanillas *et al.* 2015; Whittle *et al.* 2015). The great extent to which the mating-type chromosomes are consumed by the NRR and the chaos of rearrangements within this region indicate that these dimorphic chromosomes are the result of ancient and impactful evolutionary processes, approaching an ultimate state of genomic entropy, and thus showing striking convergence with some animals (Repping *et al.* 2006; Bachtrog 2013). Our study allows the broadening of evolutionary biology for nonrecombining chromosomes by contributing to evidence of convergence in genomic patterns between fungal mating-type chromosomes and sex chromosomes, as previously highlighted (Hood 2002; Fraser *et al.* 2004; Menkis *et al.* 2008). These distant systems share extensive suppressed recombination, a high level of rearrangements, numerous losses of genes, and accumulation of a high repetitive elements content. The high level of rearrangements and the possibility of gene conversion still make it difficult to reconstitute the full history of the recombination suppression of the mating-type chromosomes. Comparative genomics between closely related *Microbotryum* species and other members of the Microbotryomycetes with different degrees and/or ages of suppression of recombination between their mating-type chromosomes should help in resolving these questions in future studies.

Next generation sequencing is increasingly making it possible to obtain sequence data at low cost for nonmodel organisms. However, the short length of reads remains a major limitation, potentially preventing large-scale assembly, particularly for genomic regions rich in transposable elements (Treangen and Salzberg 2012). The lack of a complete physical assembly precludes investigation of some of the most fundamental evolutionary inferences. PacBio chemistry now makes it possible to obtain a complete and affordable assembly, even for notoriously challenging regions, such as highly repetitive centromeres and nonrecombining chromosomes (Treangen and Salzberg 2012). This breakthrough technology will now facilitate studies of the dynamics and evolutionary role of genome structures with unprecedented resolution and power.

Acknowledgments

We thank Stéphanie Le Prieur, Alodie Snirc, and Gilles Deparis for help with DNA extraction. PacBio sequencing was conducted at the Institute for Genomic Medicine, University of California, San Diego, La Jolla, CA. This work was supported by a European Research Council starting grant, GenomeFun 309403 and ANR-12-ADAP-0009 (Gandalf project) (to T.G.); an Institut Diversité et Evolution du Vivant grant (to T.G. and C.F.); National Science Foundation (NSF) grant DEB-1115765 to (M.E.H.); and NSF award #0947963 (to M.H.P. and C.A.C.). We declare no conflict of interest.

Literature Cited

Abbate, J. L., and M. E. Hood, 2010 Dynamic linkage relationships to the mating-type locus in automictic fungi of the genus *Microbotryum*. J. Evol. Biol. 23: 1800–1805.

Aguileta, G., J. Lengelle, S. Marthey, H. Chiapello, F. Rodolphe *et al.*, 2010 Finding candidate genes under positive selection in non-model species: examples of genes involved in host specialization in pathogens. Mol. Ecol. 19: 292–306.

Angiuoli, S., and S. Salzberg, 2011 Mugsy: fast multiple alignment of closely related whole genomes. Bioinformatics 27: 334–342.

Antonovics, J., and J. Y. Abrams, 2004 Intratetrad mating and the evolution of linkage relationships. Evolution 58: 702–709.

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

Bachtrog, D., M. Kirkpatrick, J. E. Mank, S. F. McDaniel, J. C. Pires *et al.*, 2011 Are all sex chromosomes created equal? Trends Genet. 27: 350–357.

Bailey, T. L., N. Williams, C. Misleh, and W. W. Li, 2006 MEME: discovering and analyzing DNA and protein sequence motifs. Nucleic Acids Res. 34: W369–W373.

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

Bergero, R., and D. Charlesworth, 2009 The evolution of restricted recombination in sex chromosomes. Trends Ecol. Evol. 24: 94–102.

Billiard, S., M. Lopez-Villavicencio, B. Devier, M. Hood, C. Fairhead *et al.*, 2011 Having sex, yes, but with whom? Inferences from fungi on the evolution of anisogamy and mating types. Biol. Rev. Camb. Philos. Soc. 86: 421–442.

Bull, J. J., 1978 Sex chromosomes in haploid dioecy: unique contrast to Mullers theory for diploid dioecy. Am. Nat. 112: 245–250.

Charlesworth, B., 1991 The evolution of sex chromosomes. Science 251: 1030–1033.

Cock, J. M., L. Sterck, P. Rouze, D. Scornet, A. E. Allen *et al.*, 2010 The Ectocarpus genome and the independent evolution of multicellularity in brown algae. Nature 465: 617–621.

Coelho, M. A., J. P. Sampaio, and P. Goncalves, 2010 A deviation from the bipolar-tetrapolar mating paradigm in an early diverged Basidiomycete. PLoS Genet. 6: e1001052.

Consortium, T. U., 2011 Ongoing and future developments at the Universal Protein Resource. Nucleic Acids Res. 39: D214–D219.

Darling, A. E., B. Mau, and N. T. Perna, 2010 ProgressiveMauve: multiple genome alignment with gene gain, loss, and rearrangement. PLoS ONE 5: e11147.

Day, A., 1976 Communication through fimbriae during conjugation in a fungus. Nature 262: 583–584.

Devier, B., G. Aguileta, M. Hood, and T. Giraud, 2009 Ancient trans-specific polymorphism at pheromone receptor genes in basidiomycetes. Genetics 181: 209–223.

Ellison, C. E., J. E. Stajich, D. J. Jacobson, D. O. Natvig, A. Lapidus *et al.*, 2011 Massive changes in genome architecture accompany the transition to self-fertility in the filamentous fungus *Neurospora tetrasperma*. Genetics 189: 55–69.

Foissac, S., J. Gouzy, S. Rombauts, C. Mathé, J. Amselem *et al.*, 2008 Genome annotation in plants and fungi: EuGène as a model platform. Current Bioinformatics 3: 87–97.

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

Fraser, J. A., and J. Heitman, 2004 Evolution of fungal sex chromosomes. Mol. Microbiol. 51: 299–306.

Fraser, J., and J. Heitman, 2005 Chromosomal sex-determining regions in animals, plants and fungi. Curr. Opin. Genet. Dev. 15: 645–651.

Fraser, J. A., S. Diezmann, R. L. Subaran, A. Allen, K. B. Lengeler *et al.*, 2004 Convergent evolution of chromosomal sex-determining regions in the animal and fungal kingdoms. PLoS Biol. 2: 2243–2255.

Giraud, T., R. Yockteng, M. Lopez-Villavicencio, G. Refrégier, and M. E. Hood, 2008 The mating system of the anther smut fungus, *Microbotryum violaceum*: selfing under heterothallism. Eukaryot. Cell 7: 765–775.

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

Hoi, J. W. S., and B. Dumas, 2010 Ste12 and Ste12-like proteins, fungal transcription factors regulating development and pathogenicity. Eukaryot. Cell 9: 480–485.

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

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

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

Hood, M. E., M. Katawezik, and T. Giraud, 2005 Repeat-induced point mutation and the population structure of transposable elements in *Microbotryum violaceum*. Genetics 170: 1081–1089.

Hood, M. E., E. Petit, and T. Giraud, 2013 Extensive divergence between mating-type chromosomes of the anther-smut fungus. Genetics 193: 309–315.

Ironside, J. E., 2010 No amicable divorce? Challenging the notion that sexual antagonism drives sex chromosome evolution. BioEssays 32: 718–726.

Itoh, Y., K. Kampf, and A. P. Arnold, 2006 Comparison of the chicken and zebra finch Z chromosomes shows evolutionary rearrangements. Chromosome Res. 14: 805–815.

Janbon, G., K. L. Ormerod, D. Paulet, E. J. Byrnes, III, V. Yadav *et al.*, 2014 Analysis of the genome and transcriptome of Cryptococcus neoformans var. grubii reveals complex RNA expression and microevolution leading to virulence attenuation. PLoS Genet. 10: e1004261.

Johnson, L. J., J. Antonovics, and M. E. Hood, 2005 The evolution of intratetrad mating rates. Evolution 59: 2525–2532.

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

Kaiser, V. B., and D. Bachtrog, 2010 Evolution of sex chromosomes in insects. Annu. Rev. Genet. 44: 91–112.

Koboldt, D., Q. Zhang, D. Larson, D. Shen, M. McLellan *et al.*, 2012 VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. Genome Res. 22: 568–576.

Koren, S., M. Schatz, B. Walenz, J. Martin, J. Howard *et al.*, 2012 Hybrid error correction and de novo assembly of single-molecule sequencing reads. Nat. Biotechnol. 30: 693–700.

Krzywinski, M. I., J. E. Schein, I. Birol, J. Connors, R. Gascoyne *et al.*, 2009 Circos: an information aesthetic for comparative genomics. Genome Res. 19: 1639–1645.

Lahn, B. T., and D. C. Page, 1999 Four evolutionary strata on the human X chromosome. Science 286: 964–967.

Li, H., B. Handsaker, A. Wysoker, and T. Fennell, J. Ruan, *et al.*, 2009 The Sequence alignment/map (SAM) format and SAMtools. Bioinformatics 25: 2078–2079.

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

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

McDaniel, S. F., K. M. Neubig, A. C. Payton, R. S. Quatrano, and D. J. Cove, 2013 Recent capture on the UV sex chromosomes of the moss Ceratodon purpureus. Evolution 67: 2811–2822.

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

Mitchell, A., H. Chang, L. Daugherty, M. Fraser, S. Hunter *et al.*, 2014 The InterPro protein families database: the classification resource after 15 years. Nucleic Acids Res. 43: D213–D221.

Montgomery, T., 1911 Are particular chromosomes sex determinants? Biol. Bull. 19: 1–17.

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

Ohno, S., 1967 *Sex Chromosomes and Sex-Linked Genes*. Springer-Verlag, Berlin.

Page, D. C., J. F. Hughes, D. W. Bellott, J. L. Mueller, M. E. Gill *et al.*, 2010 Reconstructing sex chromosome evolution. Genome Biol. 11: I21.

Perlin, M., J. Amselem, E. Fontanillas, S. Toh, Z. Chen *et al.*, 2015 Sex and parasites: genomic and transcriptomic analysis of *Microbotryum lychnidis-dioicae*, the biotrophic and plant-castrating anther smut fungus. BMC Genomics 16: 461.

Petit, E., T. Giraud, D. M. de Vienne, M. Coelho, G. Aguileta *et al.*, 2012 Linkage to the mating-type locus across the genus Microbotryum: insights into non-recombining chromosomes. Evolution 66: 3519–3533.

Ranwez, V., S. Harispe, F. Delsuc, and E. Douzery, 2011 MACSE: multiple alignment of coding sequences accounting for frameshifts and stop codons. PLoS ONE 6: e22594.

Repping, S., S. K. M. van Daalen, L. G. Brown, C. M. Korver, J. Lange *et al.*, 2006 High mutation rates have driven extensive structural polymorphism among human Y chromosomes. Nat. Genet. 38: 463–467.

Rice, W. R., 1984 Sex chromosomes and the evolution of sexual dimorphism. Evolution 38: 735–742.

Smit, A. F. A, R. Hubley, and P. Green 2013 RepeatMasker Open-4.0. Available at: http://www.repeatmasker.org.

Smith, D. G., M. D. Garcia-Pedrajas, W. Hong, Z. Y. Yu, S. E. Gold *et al.*, 2004 An ste20 homologue in *Ustilago maydis* plays a role in mating and pathogenicity. Eukaryot. Cell 3: 180–189.

Tesler, G., 2002 GRIMM: genome rearrangements web server. Bioinformatics 18: 492–493.

Treangen, T. J., and S. L. Salzberg, 2012 Repetitive DNA and next-generation sequencing: computational challenges and solutions. Nat. Rev. Genet. 13: 36–46.

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

Votintseva, A. A., and D. A. Filatov, 2009 Evolutionary strata in a small mating-type-specific region of the smut fungus *Microbotryum violaceum*. Genetics 182: 1391–1396.

Whittle, C. A., and H. Johannesson, 2011 Evidence of the accumulation of allele-specific non-synonymous substitutions in the young region of recombination suppression within the mating-type chromosomes of *Neurospora tetrasperma*. Heredity 107: 305–314.

Whittle, C. A., Y. Sun, and H. Johannesson, 2011 Degeneration in codon usage within the region of suppressed recombination in the mating-type chromosomes of *Neurospora tetrasperma*. Eukaryot. Cell 10: 594–603.

Whittle, C. A., A. Votintseva, K. Ridout, and D. A. Filatov, 2015 Recent and massive expansion of the mating-type specific region in the smut fungus *Microbotryum*. Genetics 199: 809–816.

Xu, L., E. Petit, and M. E. Hood, 2015 Variation in mate-recognition pheromones of the fungal genus *Microbotryum*. Heredity (in press).

Yamato, K. T., K. Ishizaki, M. Fujisawa, S. Okada, S. Nakayama *et al.*, 2007 Gene organization of the liverwort Y chromosome reveals distinct sex chromosome evolution in a haploid system. Proc. Natl. Acad. Sci. USA 104: 6472–6477.

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

Zakharov, I. A., 1987 Some principles of the gene localization in eukaryotic chromosomes. Formation of the problem and analysis of nonrandom localization of the mating-type loci in some fungi. Genetika 22: 2620–2624.

Zakharov, I. A., 2005 Intratetrad mating and its genetic and evolutionary consequences. Russ. J. Genet. 41: 508–519.

Communicating editor: J. Heitman

2.4 Discussion et perspectives

2.4.1 Synthèse des résultats

Les travaux présentés dans cette première partie ont permis d'établir définitivement que les chromosomes de types sexuels de *M. lychnidis-dioicae* sont non-recombinants sur près de 90% de leur longueur. De plus, ils présentent un très fort degré de réarrangements entre types sexuels a_1 et a_2. Des centaines de gènes sont présents à l'état hémizygote et ont donc potentiellement été perdus dans un des types sexuels. La divergence synonyme entre allèles est de 6% en moyenne mais affiche une forte hétérogénéité le long de la région non-recombinante. L'ampleur des réarrangements n'a pas permis de tester si cette hétérogénéité proviendrait de strates de suppression de recombinaison ou d'événements de conversion génique. Deux arguments sont en faveur de l'existence de conversion génique sur les chromosomes de types sexuels de *M. lychnidis-dioicae*. D'une part plusieurs gènes situés dans les régions non-recombinantes ont des généalogies qui montrent une perte dérivée du polymorphisme trans-spécifique dans certaines espèces, les espèces impliquées variant d'un gène à un autre [75]. D'autre part, les copies de rétro-transposons à LTR ("long terminal repeat") de type copia sont plus semblables entre elles sur les chromosomes de types sexuels que sur le reste du génome [81].

Au sein du complexe *M. violaceum*, on observe une dégénérescence sous forme d'une accumulation de mutations délétères et d'éléments transposables, et qui sont spécifiques aux différentes espèces. Nous avons vérifié que l'augmentation du taux de substitutions non-synonymes ne correspondait pas à de la sélection positive en comparant des modèles de sélection relâchée (dN/dS plus fort que le groupe témoin mais inférieur à 1) et de sélection positive (dN/dS > 1). On observe aussi une expression des gènes plus faible au sein de la région non-recombinante chez *M. lychnidis-dioicae*. La dégénérescence observée n'est pas asymétrique entre les deux chromosomes de types sexuels a_1 et a_2, contrairement à ce qui est observé dans les systèmes où il existe une asymétrie avec un sexe homogamétique et un sexe hétérogamétique.

L'étude de Fontanillas et al [35] a montré qu'il existait d'importantes différences d'expression entre a_1 et a_2. Pour mieux caractériser la dégénérescence dans les chromosomes de types sexuels de *M. lychnidis-dioicae* et étudier le lien entre mutations et différences d'expression, nous pourrons ré-analyser les données d'expression déjà générées [82, 35] avec l'assemblage complet des chromosomes de types sexuels, pour détecter les types de mutations (mutations non-sens, dans

les sites d'épissages, les séquences régulatrices) associées aux différences ou pertes d'expression.

2.4.2 Histoire de la suppression de recombinaison, transitions entre état bipolaire et état tétrapolaire

Les chromosomes de types sexuels de *M. lychnidis-dioicae* sont tellement réarrangés qu'il est difficile d'utiliser leur structure pour élaborer un scénario sur l'histoire de la suppression de recombinaison. La suite logique de ce travail est donc d'obtenir l'assemblage des chromosomes de types sexuels chez d'autres espèces du complexe *M. violaceum* et de groupes externes, ce qui devrait permettre de reconstituer l'histoire de la suppression de recombinaison dans ces chromosomes de types sexuels.

On pourra en particulier essayer de comprendre comment se sont produites les transitions entre bipolarité et tétrapolarité. En effet, chez les champignons basidiomycètes, le type sexuel est le plus souvent déterminé par deux loci indépendants (figure 2.5), et les croisements ne peuvent avoir lieu qu'entre cellules haploïdes portant des allèles différents aux deux loci. Cependant chez certains basidiomycètes, il y a eu des retours à l'état bipolaire par fusion des deux loci MAT ou perte de fonction de l'un des deux dans le déterminisme du type sexuel [83]. On pensait que toutes les espèces du complexe *M. violaceum* étaient bipolaires, mais Hood et collègues ont très récemment découvert que deux espèces d'un même clade étaient tétrapolaires, les loci HD et PR étant impliqués dans le déterminisme du type sexuel, situés sur des chromosomes différents, et liés chacun à leur centromère [84]. Le chromosome portant le locus PR présente une taille réduite par rapport aux chromosomes de types sexuels des espèces bipolaires, mais affiche un dimorphisme important entre les types sexuels, alors que le chromosome portant le locus HD est pratiquement monomorphe. La plupart des espèces du complexe *M. violaceum* étant bipolaires, Hood et collègues [84] ont proposé que l'état ancestral était la bipolarité, et qu'il y aurait eu un retour à l'état tétrapolaire au sein d'un clade du complexe *M. violaceum*. Mais leur conclusion repose sur une analyse de reconstruction de l'état ancestral qui dépend très fortement de l'échantillonnage, et on peut aussi imaginer que l'état ancestral soit la tétrapolarité même au sein du complexe *Microbotryum*, ce qui impliquerait deux acquisitions indépendantes de l'état bipolaire au sein de *M. violaceum* (voir phylogénie sur la figure 2.10). Nous pourrons donc comparer les génomes d'espèces bipolaires et tétrapolaires du complexe *M. violaceum* pour

comprendre dans quel sens se sont produits les changements de mode de reproduction et quels types de réarrangements sont impliqués.

Hood et collègues [84] ont aussi étudié les caryotypes d'autres espèces bipolaires du complexe *M. violaceum* : toutes les espèces bipolaires étudiées possèdent des chromosomes de types sexuels qui figurent parmi les chromosomes les plus larges de leurs génomes, et ces chromosomes de types sexuels présentent un dimorphisme important. Cela montrerait que l'existence de larges régions non-recombinantes est un caractère ancien, voire ancestral dans le complexe *M. violaceum*. Comme il n'existe pas d'estimation des temps de spéciation au sein du complexe *M. violaceum* (excepté la séparation entre les espèces très proches *M. lychnidis-dioicae* et *M. silenes-dioicae*, estimée à 400 000 ans [23]), il est difficile de dire à quel point ce résultat entre en contradiction avec les résultats de Whittle et al [80] qui proposent que la grande majorité de la région non-recombinante est d'âge "très récent" (1 ou 2 millions d'années).

Plus généralement, l'assemblage des chromosomes de types sexuels chez d'autres espèces du complexe *M. violaceum* et de groupes externes permettra de comparer le contenu en gènes et en éléments transposables pour comprendre les différences de taille observées entre chromosomes de types sexuels chez différentes espèces du complexe *M. violaceum*.

Pour comprendre l'histoire de la suppression de recombinaison, nous pourrons tester si les régions non-recombinantes sont aussi larges et réarrangées dans d'autres espèces de *M. violaceum* que chez *M. lychnidis-dioicae*, si on trouve des traces de strates de suppression de recombinaison récentes, sous forme de régions colinéaires entre types sexuels mais liées au type sexuel et non-recombinantes, et quels sont les types des réarrangements impliqués dans la formation de ces strates (inversions, translocations, insertions d'éléments transposables). Des analyses de ségrégation avec le type sexuel pourront être effectuées pour évaluer le caractère recombinant de ces régions. Une "mini-strate" potentielle a déjà été détectée chez *M. lychnidis-dioicae* [85], mais nous n'avons pas encore établi si cette région est liée au type sexuel. Il sera également intéressant de tester de degré de réarrangements entre chromosomes d'un même type sexuel mais d'espèces différentes, pour tenter de reconstituer l'histoire de ces réarrangements et à quelle vitesse ils se produisent. Nous pourrons aussi systématiquement rechercher des gènes à l'état hémizygote, et caractériser les mutations impliquées quand la pseudogénisation est suffisamment récente (mutation non-sens, insertion d'éléments transposables). Si des chromosomes de types sexuels avec des suppressions de recombinaison récentes sont identifiés, il sera intéressant

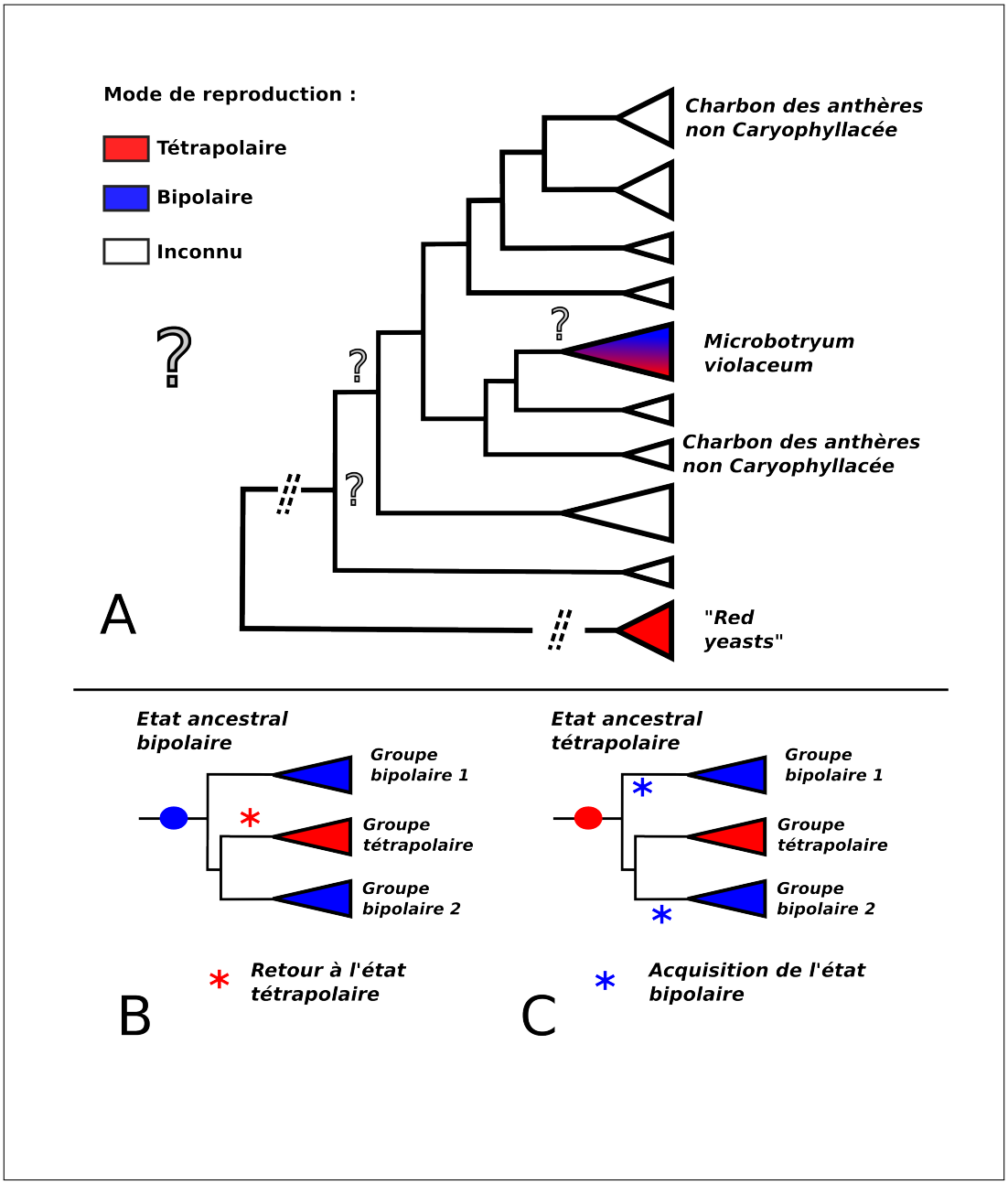


FIGURE 2.10 – Transitions entre bipolarité et tétrapolarité chez Microbotryum. A) Phylogénie simplifiée de *Microbotryum* montrant la position du groupe externe le plus proche, les levures rouges. Les levures rouges sont tétrapolaires, et *M. violaceum* inclut des espèces bipolaires et tétrapolaires. Le statut des autres espèces de *Microbotryum* et l'architecture de leurs chromosomes de types sexuels sont inconnus, et donc le nombre et la position des événements d'acquisition de la bipolarité. B) Répartition des espèces bi- et tétrapolaires dans le complexe *M. violaceum*. D'après le scénario le plus probable, l'ancêtre commun était bipolaire et il y a eu retour à la tétrapolarité au sein d'un clade. C) Le scénario alternatif, celui d'un ancêtre commun tétrapolaire, impliquerait deux acquisitions indépendantes de la bipolarité.

d'observer si une dégénérescence est déjà détectable, pour comprendre à quelle vitesse elle se produit.

Tous ces événements (formation de nouvelles strates de suppression de recombinaison, réarrangements, pertes de gènes) pourront être replacés sur la phylogénie du complexe *M. violaceum* pour reconstituer l'histoire de la suppression de recombinaison (figure 2.12). *Microbotryum violaceum* constitue un système exceptionnel pour ce type d'analyses, avec de multiples espèces possédant des chromosomes de types sexuels non-recombinants, des niveaux de divergence variables, et une phylogénie bien résolue [35]. Cependant, la plus grande inconnue est l'architecture des chromosomes de types sexuels des espèces les plus basales du complexe *M. violaceum*. Si celles-ci ont des chromosomes de types sexuels non-recombinants et très réarrangés comme ceux de *M. lychnidis-dioicae*, nous ne pourrons pas reconstituer les premières étapes de suppression de recombinaison. Dans ce cas, il sera très important d'obtenir les génomes de groupes externes tétrapolaires ou bipolaires mais possédant des chromosomes de types sexuels recombinants. Nous pourrons inclure des espèces plus basales du genre *Microbotryum*, pour lesquelles des difficultés pourraient être rencontrées pour obtenir des cultures *in vitro* et de l'ADN de qualité suffisante pour le séquençage en PacBio. Nous pourrons aussi utiliser comme groupes externes des espèces d'un groupe de levures saprophytes, les 'levures rouges ("red yeast"), qui sont tétrapolaires [86] et constituent le groupe externe le plus proche du genre *Microbotryum* au sein des pucciniomycètes. Cette étude comparative devrait permettre de mieux comprendre l'importance des transitions entre bipolarité et tétrapolarité dans la formation de larges régions non-recombinantes, ainsi que l'importance relative de la liaison au centromère et de la liaison des loci MAT dans ce processus.

Pour étudier la dynamique et l'histoire récente des régions non-recombinantes des chromosomes de types sexuels du complexe *M. violaceum*, on pourra aussi étudier les variations à l'intérieur d'une même espèce. Nous ignorons en effet si les différences observées entre types sexuels sont fixées, à quelle vitesse se produisent les réarrangements entre chromosomes de types sexuels et la dégénérescence. Il sera particulièrement intéressant de chercher des traces de conversion génique au sein d'une espèce, en détectant par exemple des régions avec une divergence synonyme nulle entre allèles a1 et a2 chez certains individus mais élevée chez d'autres. Nous disposons déjà pour cela de données de reséquençage de plusieurs dizaines d'individus au sein de *M. lychnidis-dioicae*. Enfin, il a été proposé chez *Neurospora tetrasperma* [87] que l'introgression de régions recombinantes provenant d'une espèce proche avec des chromosomes de types sexuels

qui recombinent pourrait limiter la dégénérescence des régions non-recombinantes. Il sera donc intéressant de chercher des signaux d'introgression dans les génomes du complexe *M. violaceum*, surtout si certaines espèces ont des chromosomes sexuels recombinants sur une grande partie de leur longueur.

On n'attend pas de différence de dosage d'expression entre gènes de types sexuels différents, comme ce qui est observé chez les chromosomes sexuels, le stade diploïde étant toujours hétérogamétique chez les champignons. Néanmoins, il pourrait y avoir une compensation de dosage suite aux pertes différentielles de gènes entre chromosomes de types sexuels. Cette question n'a jamais été étudiée chez les champignons à ma connaissance. Il sera intéressant d'étudier si les pertes de gènes dans un type sexuel sont compensées par une augmentation du niveau de l'expression de la copie restante par rapport à des espèces où il n'y a pas eu les mêmes pertes de gènes, ou mieux, par rapport à une espèce où les gènes correspondant sont dans une région recombinante.

Nous pourrons également essayer de dater de manière absolue les événements de suppression de recombinaison, pour essayer d'évaluer à quelle vitesse ils se sont produits dans le complexe *M. violaceum*. Nous avions proposé une datation de six marqueurs présentant différents niveaux de divergence synonyme dans une version précédente du manuscrit sur l'assemblage des chromosomes de types sexuels de *M. lychnidis-dioicae*, ce qui donnait des temps de divergence d'environ 500 000 à 900 000 ans (figure 2.11). Il nous avait été reproché d'avoir utilisé un nombre de marqueurs trop faible compte tenu de l'imprécision des modèles d'horloge moléculaire chez les champignons et du risque de conversion génique. Le faible nombre de gènes avec des datations était dû au fait que peu de gènes avaient des orthologues dans les espèces de champignons utilisables pour les calibrations fossiles. Une approche plus robuste pourra être de calibrer les nœuds de la phylogénie du complexe *M. violaceum* et de groupes externes en utilisant des marqueurs autosomaux, et de placer les événements de suppression de recombinaison et de réarrangements sur cette phylogénie calibrée.

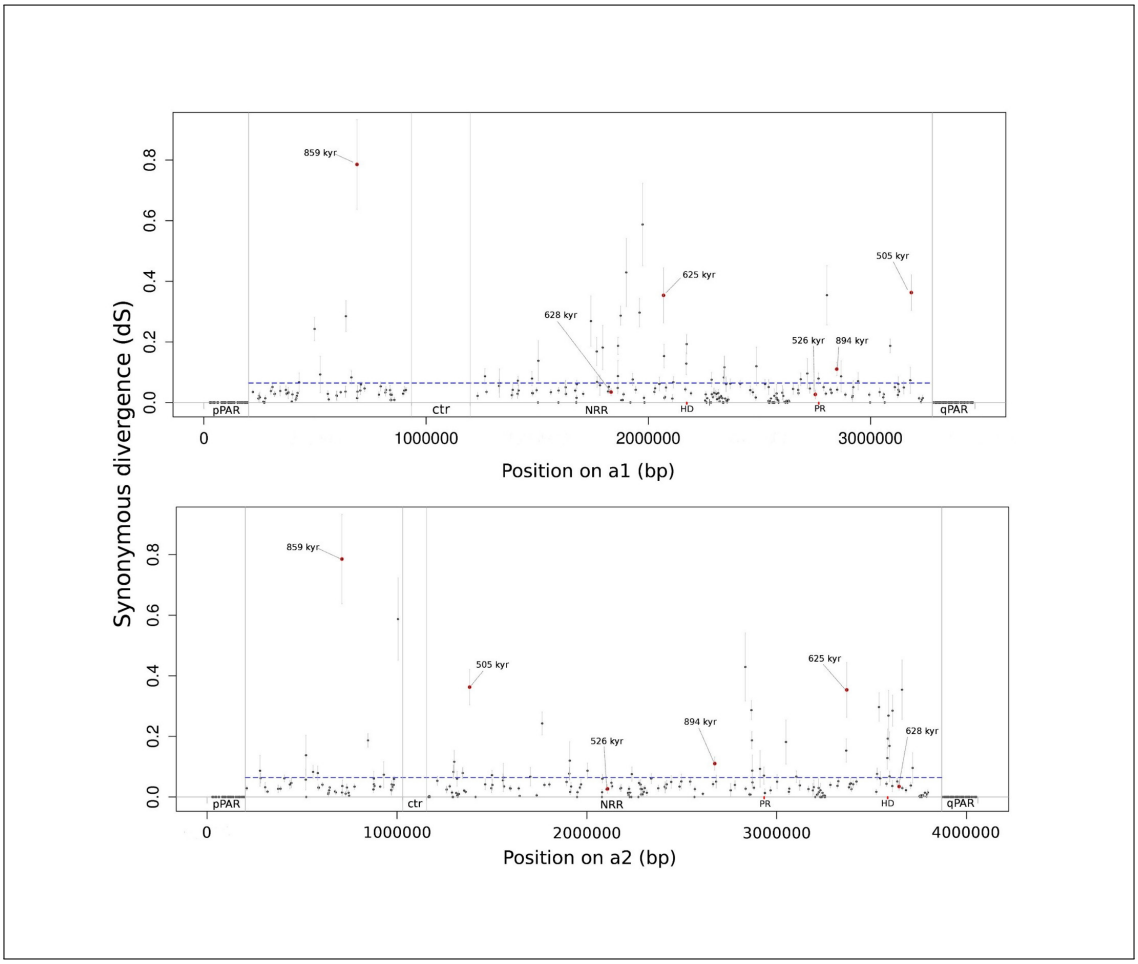


FIGURE 2.11 – Niveau de divergence synonyme entre allèles le long des chromosomes de types sexuels de *Microbotryum lychnidis-dioicae*. Le niveau de divergence synonyme entre allèles est représenté en fonction de la position physique sur a_1 (haut) et a_2 (bas). Le trait horizontal en pointillés correspond à la divergence moyenne, et les traits verticaux aux limites entre régions pseudo-autosomales et région non-recombinante.

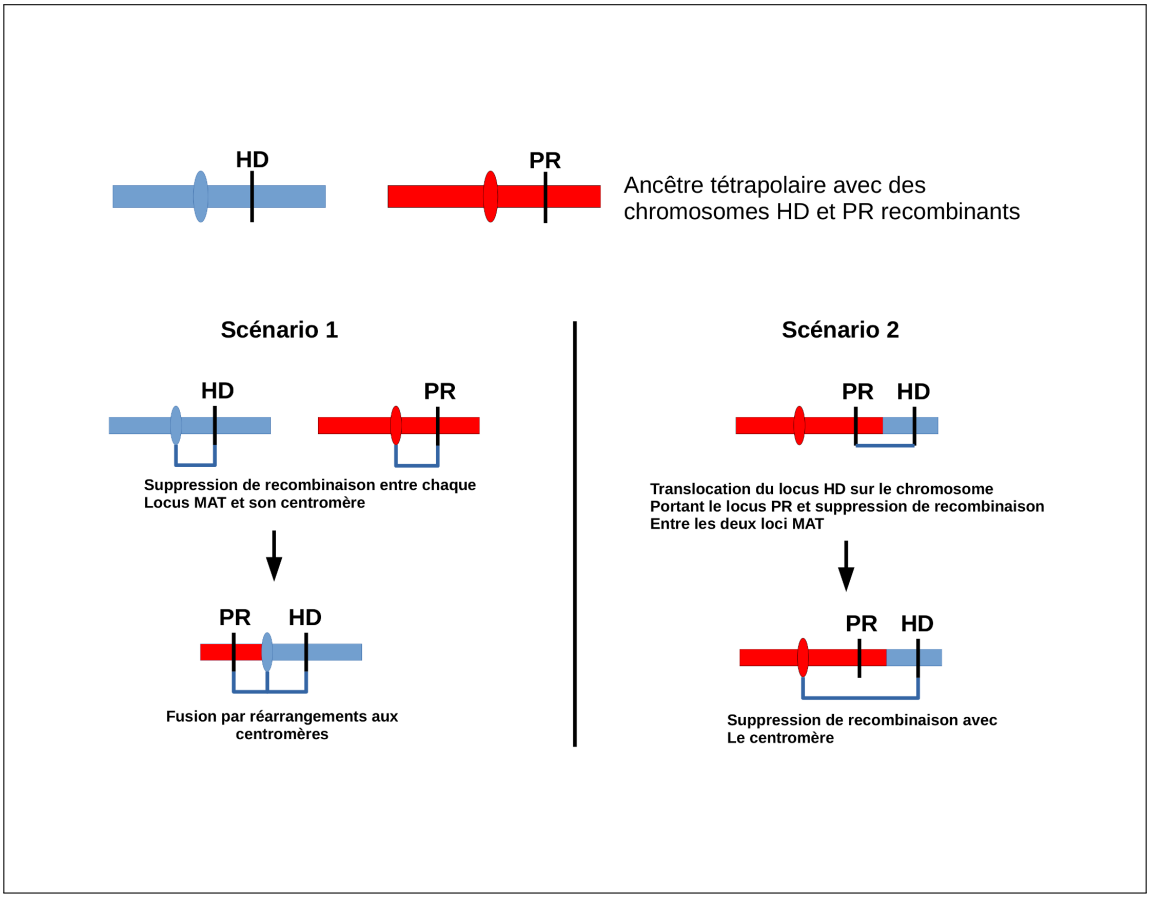


FIGURE 2.12 – Deux scénarios possibles de l'émergence de la bipolarité chez *Microbotryum violaceum*. Les réarrangements éventuels qui peuvent conduire à la suppression de recombinaison ne sont pas représentés.

3

Etude de la dégénérescence dans des populations de *Microbotryum lychnidis-dioicae* de la région de Tchernobyl

3.1 Introduction

La dégénérescence génomique peut être définie comme l'accumulation de mutations délétères dans les génomes, conduisant à une baisse de la valeur sélective à l'échelle des individus ou des populations. Ces mutations délétères peuvent être par exemple des mutations ponctuelles non-sens ou faux-sens ou des insertions d'éléments transposables dans une séquence codante ou un promoteur. Lorsque les mutations ont lieu dans la lignée somatique, c'est-à-dire dans des cellules qui ne participent pas à la reproduction, les conséquences sont limitées à l'individu dans lequel elles se produisent, et contribuent au processus de sénescence et au développement de maladies comme le cancer. Au contraire, lorsque des mutations délétères sont transmises à la descendance, cela peut entraîner une baisse globale de la valeur sélective d'une population.

Il peut y avoir augmentation ou accélération de la dégénérescence si le taux de mutation augmente, par exemple en présence d'agents mutagènes dans l'environnement, ou à cause d'une moindre efficacité des systèmes de réparation de l'ADN. Une autre cause de dégénérescence peut

être une diminution de l'efficacité de la sélection, ce qui peut être lié à une multitude de facteurs (baisse de l'effectif efficace de la population, changement du régime de reproduction, diverses formes d'auto-stop génétique, baisse du taux de recombinaison méiotique) [88, 89, 3].

Dans le chapitre précédent, nous avons étudié la dégénérescence des génomes dans un cadre particulier, celui de régions non-recombinantes situées sur les chromosomes de types sexuels. La dégénérescence de ces régions est principalement la conséquence de la baisse d'efficacité de la sélection associée à la suppression de recombinaison et de l'hétérozygotie obligatoire [5, 6].

La dégénérescence peut aussi se produire dans d'autres contextes, comme celui de l'exposition à des agents mutagènes. La région de Tchernobyl comprend de nombreux sites qui présentent différents niveaux de radiations ionisantes, permettant d'étudier l'effet des radiations sur les populations naturelles en fonction de la dose de radiation. Il n'est pas simple de prédire les conséquences de l'exposition aux radiations sur l'accumulation de mutations délétères dans les génomes. D'un côté, on peut s'attendre à ce que la présence d'agents mutagènes entraîne une augmentation des taux de mutations délétères et donc leur accumulation dans les génomes [90], mais d'un autre côté, l'augmentation du taux de mutations pourrait aussi entraîner une augmentation de la pression de sélection purifiante suffisante pour éviter une telle accumulation [91]. De plus, les espèces varient dans leur niveau de résistance aux mutations. De fait, parmi les espèces étudiées sur le site de Tchernobyl, il existe une grande variabilité dans les conséquences génomiques de la contamination radioactive [92]. Certains champignons ont déjà été étudiés pour leur susceptibilité aux radiations, et certains peuvent être très résistants [93, 94]. Mais peu d'études se sont intéressées aux conséquences des radiations à la fois au niveau génomique et en termes de valeur sélective dans des populations naturelles [93].

Pour étudier l'effet des radiations sur les populations de *M. lychnidis-dioicae*, nous avons analysé plusieurs populations présentes dans la région de Tchernobyl et exposées à des niveaux différents de radiations. Nous avons mesuré l'influence du niveau de radiation sur la prévalence de l'infection par *M. lychnidis-dioicae*, effectué des mesures de valeur sélective des individus, et testé l'existence de dégénérescence, en termes de d'accumulation de mutations non-synonymes dans les génomes (et donc probablement majoritairement délétères).

Dans le manuscrit suivant, j'ai généré les données génomiques, contribué à définir la stratégie d'analyse des génomes et participé aux analyses bio-informatiques. Ce manuscrit va être soumis très prochainement à la revue Molecular Ecology.

3.2 Article : Lower prevalence but similar viability and non-synonymous substitution rates suggest radioresistance and increased purifying selection in a parasitic fungus at Chernobyl

Lower prevalence but similar viability and non-synonymous substitution rates suggest radioresistance and increased purifying selection in a parasitic fungus at Chernobyl

Aguileta[1*] G., Badouin[1*] H., Hood[2] M. E., Møller[1] A.P., Le Prieur[1] S., Snirc[1] A, Siguenza[3,4] S., Mousseau[5] T.A., Shykoff[1] J.A., Cuomo[6] C.A., and Giraud[1] T.

*These authors share first co-authorship

[1] Ecologie Systématique Evolution, CNRS, Univ. Paris-Sud, AgroParisTech, Université Paris-Saclay, 91400 Orsay, France;

[2] Biology Department, Amherst College, Amherst, MA USA;

[3] INRA, Laboratoire des Interactions Plantes-Microorganismes (LIPM), UMR441, Castanet-Tolosan, F-31326, France;

[4] CNRS, Laboratoire des Interactions Plantes-Microorganismes (LIPM), UMR2594, Castanet-Tolosan, F-31326, France;

[5] Department of Biological Sciences, University of South Carolina, Columbia SC 29208, USA;

[6] Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA

Correspondence to TG:
Tel: (+33) 1 69 15 56 88
Fax: (+33) 1 69 15 56 96
E-mail: tatiana.giraud@u-psud.fr

Running headline:
Radio-resistance in a parasitic fungus in Chernobyl

Abstract

Nuclear disasters at Chernobyl and Fukushima provide examples of effects of acute ionizing radiation on mutations. Here, we investigated the accumulation of non-synonymous mutations, the viability and the prevalence of *Microbotryum* fungi causing the anther smut-disease, transmitted by plant pollinators. We collected flowers of *Silene latifolia* infected with *Microbotryum lychnidis-dioicae* in locations ranging by two orders of magnitude in background radiation from 0.08 to 8.35 μSv/h. Disease prevalence decreased significantly with increasing radiation level, most likely due to pollinator disappearance and abnormal behaviour. Viability and fertility, measured as the budding rate of haploid sporidia germinating from diploid teliospores, did not decrease with increasing radiation levels. We sequenced the genomes of twelve samples from Chernobyl and of four Eastern European samples collected from uncontaminated areas in Europe and analyzed alignments of 6,068 predicted genes, corresponding to 1.04×10^7 base pairs. We found no dose-dependent differences in non-synonymous substitution rates, and our analyses suggested that this was not due to a low statistical power. Thus, there was no significant evidence of increased deleterious mutation rates at higher levels of background radiation. We even found lower levels of non-synonymous substitution rates in contaminated areas compared to control regions, suggesting that purifying selection was stronger in contaminated than uncontaminated areas. We briefly discuss the mechanistic basis for radio-resistance in this non-melanized fungus.

Keywords: bumblebees; butterflies; dN/dS; fertility, *Microbotryum violaceum*; radioresistance; red pigment; melanine; spore germination; susceptibility to radiation, synonymous and non-synonymous substitution rates; viability rate; degeneration.

Introduction

Natural levels of radioactivity on Earth vary by more than one thousand-fold, and such spatial heterogeneity suffices to create heterogeneous effects on mutation. Levels of background ionizing radiation (hereafter radiation) have decreased by a factor of ten since life first originated in the pre-Cambrian (Karam & Leslie 2005). Fluctuations in natural levels of radiation due to solar flares, supernovae and gamma ray bursts, and large meteor impacts in Mexico, India, Russia and other sites likely caused the release of vast amounts of radioactive material on Earth (e.g. (Alvarez *et al.* 1980). Such fluctuations may have contributed to current levels of resistance to radiation damage in free-living organisms. Today, typical background radiation dose rates vary from around a minimum of only 0.01 to 0.10 μSv/h, with the natural level in Chernobyl before the nuclear accident being 0.01-0.02 μSv/h (Møller & Mousseau 2013). There are many sites with high levels of radiation in the oceans, with thermal vents being a well-known example (e.g. (Cherry *et al.* 1992; Fiala-Médioni *et al.* 1986; Jollivet *et al.* 1995a, b). Maximum terrestrial levels of radioactivity reach 29.7 μSv/h in Ramsar, Iran (Ghiassi-Nejad *et al.* 2002).

Mutations are changes in genomic sequences of DNA that may occur as a consequence of single or double strand breakage. DNA repair mechanisms readily repair single-strand, but also double-strand breaks of DNA (Lehman 2006; von Sonntag 2010). Radiation was first shown to be a powerful mutagen in classical laboratory experiments almost a century ago (Muller 1954; Nadson & Philippov 1925; National Academy of Sciences – Natural Resources Council. Committee on the Biological Effects of Ionizing Radiation. (BEIR V). 1990; UNSCEAR 1988). However, it is less well known that natural variation in levels of background radiation is a source of significant mutation in nature (*e.g.*, (Forster *et al.* 2002); meta-analysis in (Møller & Mousseau 2013). Such naturally occurring mutations are an important source of novel genetic variation that forms the raw material for evolution (charmHartl 1988), but both germline and somatic mutations may also cause genetic diseases, including cancer.

Fungi show extraordinary abilities to cope with ionizing radiation. For example micro-fungi associated with thermal vents can live under extremely high radiation levels (e.g. (Charmasson *et al.* 2009; Shravage *et al.* 2007). Several micro-fungi from irradiated areas are directly attracted by radionuclides (positive radiotropism), being able to grow on "hot particles" and even degrade them (Zhdanova *et al.* 2004). Furthermore, ionizing radiation may have a positive stimulatory effect on spore germination (Tugay *et al.* 2006). Because ionizing radiation promotes growth of some melanized fungi, while simultaneously transferring electrons to melanin, this suggests that redox properties may even be used to transduce energy for cell metabolism (Dadachova *et al.* 2007; Dadachova & Casadevall 2008). Thus most radio-resistant fungi have been found so far to differ from non-radio-resistant species in their level of accumulation of intracellular melanin (Dadachova *et al.* 2007; Khajo *et al.* 2011 ; Tugay *et al.* 2011).

The nuclear accident at Chernobyl on 26 April 1986 was followed by research providing extensive evidence of somatic and germline mutations in plants (*e.g.*, (Kovalchuk *et al.* 2000) and animals (*e.g.*, (Ellegren *et al.* 1997). Mutation estimates from 45 studies of 30 species revealed a mean effect size of 0.67 measured as Pearson's product-moment correlation coefficient, accounting for 44.3% of the variance (Møller & Mousseau 2015). This effect size is one of the largest ever recorded in biological sciences. Although there was consistency within species in effect size (*i.e.*, significant repeatability), and there was significant heterogeneity among species, there was no evidence of phylogenetic signal, nor were there any clear ecological predictors of effect size (Møller & Mousseau 2015).

The predicted effects of radiation exposure include chromosomal aberrations (Møller & Mousseau 2013) (Wang *et al.* 1990), an increased substitution rate due to damaged DNA (Martincorena & Campbell 2015) and consequently a likely increase in the rate of deleterious mutations (Møller & Mousseau 2013; Premi *et al.* 2009). Several studies report the detection of chromosomal aberrations subsequent to exposure to high radiation levels (Chen & Wei 1991; Cheriyan *et al.* 1999; Hayata *et al.* 2004; Jiang *et al.* 2000; Kochupillai *et al.* 1976; Møller & Mousseau 2013; Wang *et al.* 1990). Computational analysis of genes and genomes of organisms exposed to

excessive radiation typically as a result of a nuclear disaster such as those experienced in Chernobyl and Fukushima reveal signatures of increased amino-acid substitution rates, often with deleterious effects (Møller & Mousseau 2013, 2015). Increased rates of deleterious substitutions can be detected by increased ratios of non-synonymous to synonymous rates of substitutions dN/dS (Anisimova & Liberles 2012; Fontanillas *et al.* 2015). On the other hand, we may expect that selection is stronger in radiated areas (Ellegren *et al.* 1997; Moller 1993, 2002; Moller *et al.* 2013; Moller *et al.* 2012; Moller & Mousseau 2001, 2003; Moller *et al.* 2005a; Moller *et al.* 2005b), which would lead to lower rates of nonsynonymous to synonymous rates of substitutions (i.e., lower dN/dS rates).

Although radionucleotide accumulation has been investigated in fungi, in particular in edible and mycorrhizal fungi (Gwynn *et al.* 2013 ; Mascanzoni 2001 ; Mietelski *et al.* 2010), few studies have investigated the effect of contamination in Chernobyl on the abundance and fitness of fungi from the contaminated environments (Møller & Mousseau 2013; Møller & Mousseau 2014), with a few exceptions (Dadachova & Casadevall 2008; Tugay *et al.* 2006). Certain fungi were found to cope very well with these high radiation levels, and some even thriving in the defunct Chernobyl nuclear reactor (Dadachova & Casadevall 2008; Zhdanova *et al.* 2000). Fungi are good models for studying genomic consequences of radiation because they have small genomes that can easily be fully sequenced (Gladieux *et al.* 2014), and some are easy to grow *in vitro*, allowing viability and fertility measures. In addition, they play important ecological roles as pathogens, mutualists or decomposers.

Here we investigated consequences of the Chernobyl disaster on a plant pathogenic fungus. *Microbotryum lychnidis-dioicae* is a fungus causing anther smut disease on the white campion *Silene latifolia*. The pathogen castrates the plant by producing its spores in anthers in the place of pollen in males and aborting ovaries while inducing spore-bearing anthers in female plants. Spores are transmitted to healthy flowers by pollinators (Roche *et al.* 1995), and insect abundance has been shown to decrease dramatically with radiation level in Chernobyl (Moller & Mousseau 2009). *Microbotryum lychnidis-dioicae* is highly selfing, undergoing mostly intra-tetrad mating (Giraud *et al.* 2005; Hood & Antonovics 2000; Zakharov 2005), and is therefore highly homozygous (Giraud 2004; Vercken *et al.* 2010). *Microbotryum*

lychnidis-dioicae is a model organism in ecology and evolution (Antonovics *et al.* 2002; Bernasconi *et al.* 2009), and a reference genome has been published recently (Badouin *et al.* 2015; Perlin *et al.* 2015). Previous studies based on microsatellite markers and gene sequences revealed a strong population structure in *M. lychnidis-dioicae* at the European scale, with three main clusters corresponding to glacial refugia, in Italy and western and eastern Europe, respectively (Gladieux *et al.* 2011; Vercken *et al.* 2010). Therefore, we focused here on the eastern cluster next to the Chernobyl area to avoid biases due to population structure.

The objectives of this study were to 1) assess anther-smut disease prevalence in *S. latifolia* in relation to pollinator abundance and radiation levels in the field, 2) estimate deleterious effects of radiation in *M. lychnidis-dioicae* samples from Chernobyl, in terms of spore viability and fertility, and of non-synonymous substitution accumulation rates. For this second goal, we estimated one the one hand the rate of haploid spore budding from diploid teliospores, and on the other hand the genome-wide ratio of non-synonymous mutations over synonymous mutations (dN/dS). This latter objective was addressed by sequencing the genomes of twelve strains from radiated areas, and the genomes of four strains of the same genetic cluster (previously called the "eastern cluster" (Gladieux *et al.* 2011; Vercken *et al.* 2010)) from uncontaminated areas. More specifically, we tested whether dN/dS was significantly higher in contaminated regions and whether a significant correlation between dN/dS and radiation level in the field could be detected.

Materials and Methods

Data sampling in Chernobyl

APM recorded pollinator abundance (butterflies and bumblebees) and anther smut disease prevalence in *S. latifolia* in Chernobyl during fieldwork at 16 different sites inside and just outside the Chernobyl exclusion zone during 2010-2015 (Table 1, Fig. 1). APM surveyed more than 30 study sites across the Chernobyl Exclusion Zone and the surroundings. Each study site was checked for presence of caryophyllaceous plants including *Silene latifolia*. None were recorded in 14 out of 30 sites, and these 14 sites all had high levels of background radiation. If *S. latifolia* was present, the

total number of flowers and of infected flowers, relying on the presence of smut on the petals, was counted in the immediate vicinity and the number of butterflies and bumblebees seen during an observation period of 5 minutes was recorded. In large populations of *S. latifolia,* APM then moved 100 m to a new site within the *S. latifolia* population where flower number was noted and butterflies and bumblebees were again recorded for 5 minutes. This process continued until there were no more plants of *S. latifolia* recorded (Møller et al. 2012). The abundance of butterflies and bumblebees was subsequently standardized to numbers per 5 minutes of observation. APM collected flower samples in paper bags to bring to the lab for further analyses. All surveys and collections were only made on days without rain or strong wind.

Spore viability measures

MEH spread teliospores on Potato Dextrose Agar (PDA) plates for 12 strains from the Chernobyl area without prior knowledge of the level of background radiation of the sample (Table 1). On nutritive media, the diploid teliospores undergo meiosis and replicate clonally as haploid sporidia. On the plant, the sporidia undergo mating, producing an infectious hypha. Sporidia budding rate is thus a measure of both viability and fertility of the diploid individual present in a given flower. After 48 hours at 22°C MEH photographed the germinations. In the photographs, he counted the number of sporidia for 100 teliospores per sample, excluding the promycelium's cells (the four cells forming the aligned tetrad). We did not use the reference strains from outside the Chernobyl area for this experiment because they were older.

Strains, DNA extraction and sequencing

The genomes of 18 strains collected on *S. latifolia* were sequenced. For this goal, diploid spores from one anther were spread on petri dishes on PDA medium at 23°C for a few days. A given flower bears diploid spores from a single individual (Lopez-Villavicencio *et al.* 2007). Therefore, the harvested haploid sporidia on PDA represented thousands of meiotic products of a single diploid individual. Haploid harvested cells were stored at -20°C upon use. DNAs were extracted using the Macherey-Nagel NucleoSpin Soil kit #740780.250 following manufacturer's instructions and resuspended in desionised water (100 µl). DNA purity was assessed by measuring ratio of 230/260 and 280/260 nm with a NanoDrop 2000

spectrophotometer (Thermo Scientific), and double-strained DNA concentration was measured with a Qubit 2.0 fluorometer.

Paired-end libraries of 2x100 bp fragments with an insert size of 300 bp were prepared with Illumina TruSeq Nano DNA Library Prep Kits, and sequencing was performed on a HiSeq2000 Illumina sequencer, at a depth of coverage of 100X on average.

We checked that the strains belonged to *M. lychnidis-dioicae* by building a phylogenetic tree using orthologous genes of other *Microbotryum* species; two strains placed close together in the tree with other *Microbotryum* species than *M. lychnidis-dioicae*, thus representing spill-over from other host plant species, were discarded (not shown). The genomes of 16 *M. lychnidis-dioicae* strains were thus retained (Table 1). Of these, the first 12 corresponded to strains sampled in the area of Chernobyl (Figure 1, hereafter referred to as Chernobyl group), while the remaining four were sampled in different locations belonging to the Eastern genetic cluster in Europe in non-radiated areas (hereafter referred to as the reference group).

Sequence analyses

Accession numbers of the genomes 16 *M. lychnidis-dioicae* strains analyzed are included in Table 1. The complete genomes (predicted genes and corresponding proteins) of the 16 *M. lychnidis-dioicae* strains and the reference genome (Fontanillas *et al.* 2015) were used. For each genome, the total collection of predicted gene sequences was used as queries in the orthomcl analysis, which implements a sequential bioinformatics pipeline based on blast searches and clustering methods for the prediction of orthologous relationships. The script orthomcl.pl version 1.4 (Li *et al.* 2003) was used with default settings, allowing us to retrieve the full set of shared orthologs that are present in all analyzed strains as a single copy (1:1 orthologs). We therefore obtained alignments of alleles among strains.

Initially, orthomcl predicted 6,145 single-copy orthologous protein-coding genes shared among the 16 analyzed strains. However, in 77 cases, no matching protein sequence was found, so all subsequent analyses were conducted with the remaining 6068 genes. For each of these, the predicted genes and the corresponding protein sequences were extracted. First, the protein sequences were aligned and those

alignments were subsequently used to guide the predicted gene alignments, taking codons into account, with the pal2nal v.14 software (Suyama *et al.* 2006). Next, the program gestimator from the Libsequence library (Thornton 2003) was implemented to obtain all possible pairwise rate estimates of synonymous (dS), non-synonymous (dN), and the corresponding dN/dS rate ratio; dN/dS ratios equal to 999 were removed from the dataset, as they reflect the absence of synonymous differences (*i.e.*, in those cases dN/dS is estimated to be infinity). Then, the average dN/dS ratio of each of the 12 Chernobyl strains against each of the four eastern reference strains was calculated. Diversity estimates were computed using egglib (De Mita & Siol 2012). We checked using a larger genome dataset (unpublished data) that all the Eastern reference strains belonged to the genetic Eastern European cluster as previously identified based on microsatellite data (Vercken *et al.* 2010) and that no further population subdivision was found within this cluster (not shown).

To examine the capacity for melanin production, genes involved in three melanin pathways in fungi were examined: the DHN-melanin (Wheeler *et al.* 2008), DOPA melanin (Langfelder *et al.* 2003), and L-tyrosine degradation (Keller *et al.* 2011; Schmaler-Ripcke *et al.* 2009). Genes for each pathway in *M. lychnidis-dioicae* were identified by identifying orthologs using Orthomcl as above with the gene set of *Aspergillus niger*. Where orthologs were not identified, the most similar sequences based on blastp were examined; in all cases this did not identify any additional orthologs, *i.e.*, for the multicopper-oxidases (MCOs) involved in DHN-melanin synthesis this identified only ascomycete laccase MCOs.

Statistical analyses

ANOVA, mean comparisons and correlations were performed using JMP (SAS Institute) and power analysis using the Z transformation method (Lachin 1981) at http://www.cct.cuhk.edu.hk/stat/other/correlation.htm. Radiation and abundance data were log-transformed for improving the normality of their distributions.

Results

Prevalence data of the disease

We analyzed the effect of the abundance of butterflies, bumblebees and radiation level on anther-smut disease prevalence in *S. latifolia* (Table 2). Disease prevalence

increased with pollinator abundance, and significantly so with butterfly abundance (Figure 2; analysis weighted by sample size: r = 0.65, df = 19, P = 0.0008), and log butterfly abundance decreased with log radiation levels (Figure 3; analysis weighted by sample size: $F_{(1, 19)}$ = 15.48, r^2 = 0.42, P = 0.0009, slope (SE) = -0.107 (0.027)). Therefore, we included pollinator abundance in the model testing for an effect of radiation levels on disease prevalence. Disease prevalence significantly decreased with radiation level (Table 2, Figure 4), while it was not directly affected by the abundance of either class of pollinator as main factors. The interaction between radiation level and abundance of butterflies was however significantly correlated with disease prevalence (Table 2). Indeed, prevalence increased with butterfly abundance at low radiation levels, while it decreased with butterfly abundance at high pollinator abundance (Figure 5).

Spore viability measures

The numbers of haploid sporidia germinating from diploid teliospores following meiosis and after 48 hours were counted for 100 teliospores for each of 12 strains from the Chernobyl area (Table 1). There was no significant correlation between spore germination rates and radiation level (r = 0.19, P = 0.55, on log transformed data; N = 12; Figure 6). With this correlation, a sample size of more than 105 strains would be required to detect a significant weak correlation.

Genome sequences

A total of 6,068 predicted genes corresponding to $1.04x10^7$ base pairs (bp) was analyzed, with an average of $1.71x10^3$ bp per gene alignment. Of these bp, a total of $1.94x10^5$ corresponded to polymorphic sites, with an average of 83.28 polymorphic sites per analyzed gene alignment. Other diversity indicators yielded averages per site for π of 0.0010 for Chernobyl strains and 0.0017 for the eastern reference strains, and for θ of 0.0005 for Chernobyl strains and 0.0008 for the eastern reference strains.

The average pairwise dN/dS ratios ranged between 0.336 and 0.359. Table 1 presents the dN/dS average values for all the pairwise estimates per Chernobyl strain and the

corresponding radiation measurement performed in the field at that collection site. We detected no sign of genomic degeneration in the Chernobyl area. Indeed, we found no significant correlation between the average pairwise dN/dS values and the radiation measurements at the different sampling sites in Chernobyl ($r = 0.068$, $n = 12$, $P = 0.83$, Figure 7). With this correlation, a sample size of more than 830 genomes would be required to detect a significant weak correlation. The mean dN/dS values were even significantly lower in the Chernobyl strains (mean dN/dS = 0.346) than in the reference Eastern group (mean dN/dS = 0.353) (Student's t-test, $t = 2.16$, $df = 14$, $P = 0.004$).

Discussion

The prevalence of the anther-smut disease caused by *M. lychnidis-dioicae* was much lower in more contaminated areas, influenced by an interaction with butterfly abundance. Radiation indeed reduced the abundance of pollinators, and at low radiation levels prevalence increased with abundance of butterflies, as expected given that spores are transmitted by pollinators (Roche *et al.* 1995). At high radiation levels however the opposite was found, prevalence decreasing with butterfly abundance. Perhaps butterflies are poorer pollinators at high radiation level because of abnormal behaviour. In fact, studies of birds, spiders, plants and insects have shown more abnormalities at higher radiation levels (Møller & Mousseau 2013). It is also possible that plants are more stressed at higher radiation levels and therefore display higher resistance levels. However, previous studies have in contrast found lower resistance of plants to biotic stress in Chernobyl (Dmitriev *et al.* 2011).

We found no evidence of reduced spore viability or elevated frequency of non-synonymous substitutions in *M. lychnidis-dioicae* as a function of radiation level. The magnitude of the correlation was far outside the mean estimate for effects of ionizing radiation on mutations at Chernobyl ($r = 0.67$, 95% confidence intervals 0.59 to 0.73, (Møller & Mousseau 2014). We even found significantly lower mean values of dN/dS in Chernobyl, which may be due to stronger selection in contaminated areas against individuals bearing mildly deleterious mutations. Previous studies have shown that elimination of inferior phenotypes occurred by more intense selection in Chernobyl

(*i.e.*, stronger purifying selection) (Ellegren *et al.* 1997; Moller 1993, 2002; Moller *et al.* 2013; Moller *et al.* 2012; Moller & Mousseau 2001, 2003; Moller *et al.* 2005a; Moller *et al.* 2005b).

The lower prevalence and lower abundance of pollinators in contaminated areas are unlikely to have biased dN/dS values, *i.e.* to have changed the rate of non-synonymous substitutions by themselves. Indeed, *M. lychnidis-dioicae* is highly selfing (Giraud *et al.* 2008), so the low prevalence in Chernobyl is unlikely to lead to increased selfing rates, which could result in higher accumulation of deleterious mutations on itself. In addition, a smaller effective population size could also lead to increased rates of deleterious mutation accumulation, but this is unlikely to have had any effect with standard mutation rates since 1986, as *M. lychnidis-dioicae* undergoes a single generation per year, with a sex event before each new plant infection (Giraud *et al.* 2008). In any case, these two effects would be expected to increase the dN/dS values in contaminated areas, while we observed the opposite.

Melanin has been invoked as a mechanism of radio-resistance in bacteria and fungi (Dadachova *et al.* 2007; Khajo *et al.* 2011 ; Tugay *et al.* 2011). The presence of melanin has however not been reported so far in *Microbotryum*, and very few homologs to genes known to be involved in the synthesis of melanin were found in its genome (Table 3). However, red pigments have been described in *M. lychnidis-dioicae* (at that time the fungus was named *Ustilago violacea*) (Will *et al.* 1984; Will & Reppe 1984). Other red-pigmented organisms have been found to be radio-resistant (Asker *et al.* 2007; Copeland *et al.* 2012; Su *et al.* 2014; Yuan *et al.* 2009), although red pigmentation is mostly known as a protection against UV radiation so far, including in *M. lychnidis-dioicae* (Will *et al.* 1984; Will & Reppe 1984). It may be that red pigmentation selected for UV-resistance also confers radio-resistance, a hypothesis that should be explored in future studies. Actually *M. lychnidis-dioicae* spores produced in anthers are extremely dark and resistant, and the fungus grows only small hyphae in the plants, probably remaining in the meristem it infects until flower development (Schafer *et al.* 2010).

A number of studies have concluded that there is evidence of adaptation to low dose radiation at Chernobyl, in particular in fungi. These range from proteomic analyses

(Danchenko *et al.* 2009; Klubicova *et al.* 2010) and studies of DNA methylation (Kovalchuk *et al.* 2003) to other physiological mechanisms (Klubicova *et al.* 2012; Kovalchuk *et al.* 2004). There is also evidence consistent with adaptation through the intra-cellular antioxidant glutathione (GSH; (Galvan *et al.* 2014). Perhaps the most clear-cut evidence of adaptation concerns resistance to radioactivity in generalist bacteria that are widely distributed across Europe (Ruiz-González *et al.* 2014).

In conclusion, our study reinforces the view that fungi, and even non-melanized fungi, can cope extremely well with high radiation levels. A previous study reported the emergence of a more virulent crop pathogen population in Chernobyl and this could be a consequence of selection for resistance to radiation (Dmitriev *et al.* 2011), although reduced host plant fitness may also play a role. Finally, our findings also suggest the existence of strong purifying selection in radiated areas.

Acknowledgements

We thank the GenoToul platform for sequencing. We acknowledge the ERC Starting Grant GenomeFun 309403. We thank Laetitia Giraud, Gilles Deparis, Brian Malave and Melissa Sheth for help with spore cultures and spore counts.

References

Alvarez LW, Alvarez W, Asaro F, Michel HW (1980) Extraterrestrial cause for the Cretaceous-Tertiary extinction: Experimental results and theoretical interpretation. *Science* **208**, 1095-1108.

Anisimova M, Liberles D (2012) Detecting and understanding natural selection. In: *Codon Evolution: mechanisms and models* (eds. G C, A.. S). Oxford University Press, Oxford.

Antonovics J, Hood M, Partain J, Heuhsen AM (2002) The ecology and genetics of a host-shift: *Microbotryum* as a model system. *Am. nat.* **160**, S40-S53.

Asker D, Beppu T, Ueda K (2007) Unique diversity of carotenoid-producing bacteria isolated from Misasa, a radioactive site in Japan. *Applied Microbiology and Biotechnology* **77**, 383-392.

Badouin H, Hood M, Gouzy J*, et al.* (2015) Chaos of rearrangements in the mating-type chromosomes of the anther-smut fungus *Microbotryum lychnidis-dioicae*. *Genetics*.

Bernasconi G, Antonovics J, Biere A*, et al.* (2009) Re-emergence of *Silene* as a classic model system in ecology and evolution. *Heredity* **103**, 5–14.

Charmasson S, Sarradin PM, Le Faouder A*, et al.* (2009) High levels of natural radioactivity in biota from deep-sea hydrothermal vents: a preliminary communication. *Journal of Environmental Radioactivity* **100**, 522-526.

charmHartl DL (1988) *A primer of population genetics, second edition* Sinauer Associates, Inc., Sunderland.

Chen DQ, Wei LX (1991) Chromosome aberration, cancer mortality and hormetic phenomena among inhabitants in areas of high background radiation in China. *Journal of Radiation Research* **32**, 46-53.

Cheriyan VD, Kurien CJ, Das B*, et al.* (1999) Genetic monitoring of the human population from high-level natural radiation areas of Kerala on the southwest coast of India. II. Incidence of numerical and structural chromosomal aberrations in the lymphocytes. *Radiation Research* **152**, S154-S158.

Cherry R, Desbruyères D, Heyraud M, Nolan C (1992) High levels of natural radioactivity in hydrothermal vent polychaetes. *Comptes Rendus de l'Academie des Sciences, Ser. 3* **315**, 21-26.

Copeland A, Zeytun A, Yassawong M*, et al.* (2012) Complete genome sequence of the orange-red pigmented, radioresistant *Deinococcus proteolyticus* type strain (MRPT). *Standards in Genomic Sciences* **6**, 240-250.

Dadachova E, Bryan RA, Huang X*, et al.* (2007) Ionizing radiation changes the electronic properties of melanin and enhances the growth of melanized fungi. *Public Library of Science One* **2**, e457.

Dadachova E, Casadevall A (2008) Ionizing radiation: how fungi cope, adapt, and exploit with the help of melanin. *Current Opinions in Microbiology* **11**, 525-531.

Danchenko M, Skultety L, Rashydov NM*, et al.* (2009) Proteomic analysis of mature soybean seeds from the Chernobyl area suggests plant adaptation to the contaminated environment. *Journal of Proteome Research* **8**, 2915-2922.

De Mita S, Siol M (2012) EggLib: processing, analysis and simulation tools for population genetics and genomics. *BMC Genet.* **13**, 27.
Dmitriev AP, Grodzinskii DM, Gushcha NI, Kryzhanovskaya MS (2011) Effect of chronic irradiation on plant resistance to biotic stress in 30-km Chernobyl nuclear power plant exclusion zone. *Russian Journal of Plant Physiology* **58**, 1062-1068.
Ellegren H, Lindgren G, Primmer CR, Møller AP (1997) Fitness loss and germline mutations in barn swallows breeding in Chernobyl. *Nature* **389**, 593-596.
Fiala-Médioni A, Alayse AM, Cahet G (1986) Evidence of in situ uptake and incorporation of bicarbonate and amino acids by a hydrothermal vent mussel. *J. Exp. Mar. Biol. Ecol.* **96**, 191-198.
Fontanillas E, Hood M, Badouin H*, et al.* (2015) Degeneration of the non-recombining regions in the mating type chromosomes of the anther smut fungi. *Mol Biol Evol* **32**, 928-943.
Forster L, Forster P, Lutz-Bonengel S, Willkomm H, Brinkmann B (2002) Natural radioactivity and human mitochondrial DNA mutations. *Proceedings of the National Academy of Science of the USA* **99**, 13950-13954.
Galvan I, Bonisoli-Alquati A, Jenkinson S*, et al.* (2014) Chronic exposure to low-dose radiation at Chernobyl favours adaptation to oxidative stress in birds. *Functional Ecology* **28**, 1387-1403.
Ghiassi-Nejad M, Mortazavi SMJ, Cameron JR, Niroomand-rad A, Karam PA (2002) Very high background radiation areas of Ramsar, Iran: Preliminary biological studies. *Health Physics* **82**, 87-93.
Giraud T (2004) Patterns of within population dispersion and mating of the fungus *Microbotryum violaceum* parasitising the plant *Silene latifolia*. ***Heredity*** **93**, 559-565.
Giraud T, Jonot O, Shykoff JA (2005) Selfing propensity under choice conditions in a parasitic fungus, Microbotryum violaceum, and parameters influencing infection success in artificial inoculations *Int. J. Plant Sci.* **166**, 649-657.
Giraud T, Yockteng R, Lopez-Villavicencio M, Refregier G, Hood ME (2008) The mating system of the anther smut fungus, *Microbotryum violaceum*: selfing under heterothallism. *Euk. Cell* **7**, 765-775.
Gladieux P, Ropars R, Badouin H*, et al.* (2014) Fungal evolutionary genomics provide insights into the mechanisms of adaptive divergence in eukaryotes. *Mol Ecol* **23**, 753-773.
Gladieux P, Vercken E, Fontaine M*, et al.* (2011) Maintenance of fungal pathogen species that are specialized to different hosts: allopatric divergence and introgression through secondary contact. *Mol. Biol. Evol.* **28**, 459-471.
Gwynn JP, Nalbandyan A, Rudolfsen G (2013) Po-210, Pb-210, K-40 and Cs-137 in edible wild berries and mushrooms and ingestion doses to man from high consumption rates of these wild foods *Journal of Environmental Radioactivity* **116**, 34-41.
Hayata I, Wang C, Zhang W*, et al.* (2004) Effect of high-level natural radiation on chromosomes of residents in southern China. *Cytogenetic and Genome Research* **104**, 237-239.
Hood ME, Antonovics J (2000) Intratetrad mating, heterozygosity, and the maintenance of deleterious alleles in *Microbotryum violaceum (=Ustilago violacea)*. *Heredity* **85**, 231-241.

Jiang T, Hayata I, Wang C*, et al.* (2000) Dose-effect relationship of dicentric and ring chromosomes in lymphocytes of individuals living in the high background radiation areas in China. *Journal of Radiation Research* **41 Suppl**, 63-68.

Jollivet D, Desbruyeres D, Ladrat C, Laubier L (1995a) Evidence for differences in the allozyme thermostability of deep-sea hydrothermal vent polychaetes (Alvinellidae): A possible selection by habitat. *Marine Ecology - Progress Series* **123**, 125-136.

Jollivet D, Desbruyeres D, Ladrat C, Laubier L (1995b) Evidence for differences in the allozyme thermostability of deep-sea hydrothermal vent polychaetes (Alvinellidae): a possible selection by habitat. *Marine Ecology Progress Series* **123**, 125-136.

Karam PA, Leslie SA (2005) Changes in terrestrial natural radiation levels over the history of life. In: *Natural Radiation Environment* (eds. McLaughlin JP, Simopoulos SE, Steinhausler F), pp. 107-117.

Keller S, Macheleidt J, Scherlach K*, et al.* (2011) Pyomelanin formation in *Aspergillus fumigatus* requires HmgX and the transcriptional activator HmgR but is dispensable for virulence. *PLoS One* **6**.

Khajo A, Bryan RA, Friedman M, al. e (2011) Protection of melanized *Cryptococcus neoformans* from lethal dose gamma irradiation involves changes in melanin's chemical structure and paramagnetism

PLoS One **6**, e25092.

Klubicova K, Danchenko M, Skultety L*, et al.* (2012) Soybeans grown in the Chernobyl area produce fertile seeds that have increased heavy metal resistance and modified carbon metabolism. *PLoS One* **7**.

Klubicova K, Danchenko M, Skultety L*, et al.* (2010) Proteomics analysis of flax grown in Chernobyl area suggests limited effect of contaminated environment on seed proteome. *Environmental Science & Technology* **44**, 6940-6946.

Kochupillai N, Verma IC, Grewal MS, Ramalingaswami V (1976) Downs-syndrome and related abnormalities in area of high background radiation in costal Kerala. *Nature* **262**, 60-61.

Kovalchuk I, Abramov V, Pogribny I, Kovalchuk O (2004) Molecular aspects of plant adaptation to life in the Chernobyl zone. *Plant Physiology* **135**, 357-363.

Kovalchuk I, Kovalchuk O, Kalck V*, et al.* (2003) Pathogen-induced systemic plant signal triggers DNA rearrangements. *Nature* **423**, 760-762.

Kovalchuk O, Dubrova YE, Arkhipov A, Hohn B, Kovalchuk I (2000) Germline DNA: Wheat mutation rate after Chernobyl. *Nature* **407**, 583-584.

Lachin L (1981).

Langfelder K, Streibel M, Jahn B, Haase G, Brakhage AA (2003) Biosynthesis of fungal melanins and their importance for human pathogenic fungi. *Fungal Genetics and Biology* **38**, 143-158.

Lehman AR (2006) *DNA repair* Elsevier, Amsterdam , The Netherlands.

Li L, Stoeckert CJ, Roos DS (2003) OrthoMCL: Identification of ortholog groups for eukaryotic genomes. *Genome Research* **13**, 2178-2189.

Lopez-Villavicencio M, Jonot O, Coantic A*, et al.* (2007) Multiple infections by the anther smut pathogen are frequent and involve related strains. *PloS Pathogens* **3**, e176.

Martincorena I, Campbell PJ (2015) Somatic mutation in cancer and normal cells. *Science* **349**, 1483-1489.
Mascanzoni D (2001) Long-term Cs-137 contamination of mushrooms following the Chernobyl fallout *journal of radioanalytical and nuclear chemistry* **249**, 245-249.
Mietelski JW, Dubchak S, Blazej S, al. e (2010) Cs-137 and K-40 in fruiting bodies of different fungal species collected in a single forest in southern Poland *Journal of Environmental Radioactivity* **101** 706-711.
Moller AP (1993) Morphology and sexual selection in the barn swallow *Hirundo rustica* in Chernobyl, Ukraine. *Proceedings of the Royal Society B-Biological Sciences* **252**, 51-57.
Moller AP (2002) Developmental instability and sexual selection in stag beetles from Chernobyl and a control area. *Ethology* **108**, 193-204.
Moller AP, Bonisoli-Alquati A, Mousseau TA (2013) High frequency of albinism and tumours in free-living birds around Chernobyl. *Mutation Research-Genetic Toxicology and Environmental Mutagenesis* **757**, 52-59.
Moller AP, Bonisoli-Alquati A, Rudolfsen G, Mousseau TA (2012) Elevated mortality among birds in Chernobyl as judged from skewed age and sex ratios. *PLoS One* **7**.
Moller AP, Mousseau TA (2001) Albinism and phenotype of barn swallows (*Hirundo rustica*) from Chernobyl. *Evolution* **55**, 2097-2104.
Moller AP, Mousseau TA (2003) Mutation and sexual selection: A test using barn swallows from Chernobyl. *Evolution* **57**, 2139-2146.
Moller AP, Mousseau TA (2009) Reduced abundance of insects and spiders linked to radiation at Chernobyl 20 years after the accident. *Biology Letters* **5**, 356-359.
Møller AP, Mousseau TA (2013) The effects of natural variation in background radioactivity on humans, animals and other organisms. *Biological Reviews* **88**, 226-254.
Møller AP, Mousseau TA (2014) Strong effects of ionizing radiation from Chernobyl on mutation rates. *Scientific Reports* **5**, 8363.
Møller AP, Mousseau TA (2015) Strong effects of ionizing radiation from Chernobyl on mutation rates. *Scientific Reports* **5**, 8363.
Moller AP, Mousseau TA, Milinevsky G*, et al.* (2005a) Condition, reproduction and survival of barn swallows from Chernobyl. *Journal of Animal Ecology* **74**, 1102-1111.
Moller AP, Surai P, Mousseau TA (2005b) Antioxidants, radiation and mutation as revealed by sperm abnormality in barn swallows from Chernobyl. *Proceedings of the Royal Society B-Biological Sciences* **272**, 247-252.
Muller HJ (1954) The manner of production of mutations by radiation. In: *Radiation Biology, Vol. 1: High Energy Radiation* (ed. Hollaender A), pp. 475-626. McGraw-Hill, New York.
Nadson GA, Philippov GS (1925) Influence des rayons x sur la sexualité et la formation des mutantes chez les champignons inferieurs (Mucorinées). *Comptes Rendus de la Societé de la Biologie et de ses Filiales* **93**, 473-474.
National Academy of Sciences – Natural Resources Council. Committee on the Biological Effects of Ionizing Radiation. (BEIR V). (1990) *Health Effects of Exposure to Low Levels of Ionizing Radiation* National Academy Press, Washington, DC.

Perlin M, Amselem J, Fontanillas E*, et al.* (2015) Sex and Parasites: Genomic and transcriptomic analysis of *Microbotryum lychnidis-dioicae*, the biotrophic and plant-castrating anther smut fungus. *BMC Genomics* **16**, 461.

Premi S, Srivastava J, Chandy SP, Ali S (2009) Unique signatures of natural background radiation on human Y chromosomes from Kerala, India. *PLoS One* **4**.

Roche BM, Alexander HH, Maltby AD (1995) Dispersal and disease gradients of anther-smut infection of *Silene alba* at different life stages. *Ecology* **76**, 1863-1871.

Ruiz-González MX, Czirják GÁ, Genevaux P*, et al.* (2014) Adaptation of feather-associated bacteria to different levels of background radiation near Chernobyl.

Schafer A, Kemler M, Bauer R, et al. (2010) The illustrated life cycle of Microbotryum on the host plant Silene latifolia. *BOTANY-BOTANIQUE* **88** 875-885.

Schmaler-Ripcke J, Sugareva V, Gebhardt P*, et al.* (2009) Production of pyomelanin, a second type of melanin, via the tyrosine degradation pathway in *Aspergillus fumigatus*. *Applied and Environmental Microbiology* **75**, 493-503.

Shravage BV, Dayananda KM, Patole MS, Shouche YS (2007) Molecular microbial diversity of a soil sample and detection of ammonia oxidizers from Cape Evans, McMurdo Dry Valley, Antarctica. *Microbiological Research* **162**, 15-25.

Su SY, Chen M, Teng C*, et al.* (2014) *Hymenobacter kanuolensis* sp nov., a novel radiation-resistant bacterium. *International Journal of Systematic and Evolutionary Microbiology* **64**, 2108-2112.

Suyama M, Torrents D, Bork P (2006) PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Research* **34**, W609-W612.

Thornton K (2003) libsequence: a C++ class library for evolutionary genetic analysis. *Bioinformatics* **19**, 2325-2327.

Tsai HF, Wheeler MH, Chang YC, Kwon-Chung KJ (1999) A developmentally regulated gene cluster involved in conidial pigment biosynthesis in *Aspergillus fumigatus*. *Journal of Bacteriology* **181**, 6469-6477.

Tugay T, Zhdanova NN, Zheltonozhsky V, Sadovnikov L, Dighton J (2006) The influence of ionizing radiation on spore germination and emergent hyphal growth response reactions of microfungi. *Mycologia* **98**, 521-527.

Tugay TI, Zheltonozhskaya MV, Sadovnikov LV, et al. (2011) Effects of ionizing radiation on the antioxydant system of microscopic fungi with radioadaptive properties found in the Chernobyl ecxlucion zone. *Health Physics* **101**, 375-382.

UNSCEAR (1988) *Sources, effects and risks of ionizing radiation* United Nations, New York.

Vercken E, Fontaine M, Gladieux P*, et al.* (2010) Glacial refugia in pathogens: European genetic structure of anther smut pathogens on *Silene latifolia* and *S. dioica*. *PloS Pathogens* **6**, e1001229.

von Sonntag C (2010) *Free-radical-induced DNA damage and its repair: A chemical perspective* Springer, Berlin, Germany.

Wang ZY, Boice JD, Wei LX*, et al.* (1990) Thyroid nodularity and chromosome aberrations among women in areas of high background radiation in China *Journal of the National Cancer Institute* **82**, 478-485.
Wheeler MH, Abramczyk D, Puckhaber LS*, et al.* (2008) New biosynthetic step in the melanin pathway of *Wangiella (Exophiala) dermatitidis*: Evidence for 2-Acetyl-1,3,6,8-tetrahydroxynaphthalene as a novel precursor. *Eukaryotic Cell* **7**, 1699-1711.
Will O, Newl N, Reppe C (1984) The photosensitivity of pigmented and non-pigmented strains of *Ustilago violacea*. *Curr Microbiol.* **10**, 295-302.
Will O, Reppe C (1984) Ultraviolet light sensitivity of carotene-and cytochrome-c-accumulating strains of the smut fungus *Ustilago violacea*. *Curr Microbiol.* **11**, 31-35.
Yuan ML, Zhang W, Dai SM*, et al.* (2009) *Deinococcus gobiensis* sp nov., an extremely radiation-resistant bacterium. *International Journal of Systematic and Evolutionary Microbiology* **59**, 1513-1517.
Zakharov IA (2005) Intratetrad mating and its genetic and evolutionary consequences. *Russian Journal of Genetics* **41**, 508-519.
Zhdanova NN, Tugay T, Dighton J, Zheltonozhsky V, McDermott P (2004) Ionizing radiation attracts soil fungi. *Mycological Research* **108**, 1089-1096.
Zhdanova NN, Zakharchenko VA, Vember VV, Nakonechnaya LT (2000) Fungi from Chernobyl: mycobiota of the inner regions of the containment structures of the damaged nuclear reactor. *Mycological Research* **104**, 1421-1426.

Table 1. Information on *Microbotryum lychnidis-dioicae* strains for which genomes have been analysed: strain ID, GPS coordinates, prevalence of the disease in the population, date of collection, spore viability and fertility (mean number of haploid sporidia germinating from 100 diploid teliospores after 48 hours), radiation level measure (µSv/h) at the collection site and mean non-synonymous substitution rate over synonymous substitution rates (dN/dS) in its genome compared to reference strains from uncontaminated areas belonging to the same eastern genetic cluster. For the strains from uncontaminated areas, the comparison was done with the three other reference strains. EMBL genome accession numbers are given, the study number being PRJEB11006.

Strain ID	Loca-tion	GPS coordi-nates	Preva-lence (Sam-ple size)	Date of collection	Mean ±SE spore count	Radia-tion (µSv/h)	Mean dN/ dS	Genome accession numbers
ch1 1101-1	Near Zalizia, Ukraine	N 51° 8' 60" E 30° 7' 12"	0.030 (299)	September 2013		1.234	0.347	ERS884833
ch2 1101-2	Near Zalizia, Ukraine	N 51° 8' 60" E 30° 7' 12"	0.030 (299)	September 2013		0.24	0.348	ERS884991
ch3 1101-3	Near Zalizia, Ukraine	N 51° 8' 60" E 30° 7' 12"	0.030 (299)	September 2013	34.30±1.47	0.235	0.345	ERS885064
ch4 1102-1	EcoCente r, Ukraine	N 51° 12' 36" E 30° 0' 0"	0.075 (40)	September 2013	38.92±1.48	0.196	0.347	ERS885276
ch5 1102-2	EcoCente r, Ukraine	N 51° 12' 36" E 30° 0' 0"	0.075 (40)	September 2013		0.165	0.336	ERS885329
1103-3	Vesniane , Ukraine	N51°18'600'' E 29°38'263''		September 2013	90.71± 3.38	3.56		
ch7 1105	Red Forest, Ukraine	N 51° 13' 48" E 30°2'23.999"	0.000 (78)	September 2013	75.68±3.20	21.03	0.351	ERS885496
ch8 1106	Hilton, Ukraine	N 51°32’378" E 21°10’427"	0.167	September 2013	54.90±2.29	0.08	0.346	ERS989710
1108	Dytiatki Ukraine	N 51°06’516" E 30°08'809''		September 2013	118.16±3.22	0.16		
1109	Ecopolis Ukraine	N51°23'355'' E30°04'225''		September 2013	30.69±0.98	0.05		
1161	Chernoby l Village	N 51°33’000" E 31°11'212''		June 2014	6.71 ±0.77	0.35		
ch11	Vesniane	N 51° 10' 48"	0.031	June 2014	24.48±1.54	8.35	0.342	ERS885569

1162	, Ukraine	E 30° 22'12"	(192)					
ch12 1163	Ivankov, Ukraine	N 51° 13' 48" E 30°1'11.999"	0.093 (216)	June 2014	53.49±3.03	0.08	0.350	ERS885640
ch13 1164	Voronkov, Ukraine	N 50° 7' 48" E 30° 31' 48"	0.167 (228)	June 2014	29.9± 1.68	0.03	0.352	ERS885695
ch14 1165-1	Near Chernobyl, Ukraine	N 51° 12' 36" E 30° 0' 0"	0.113 (53)	June 2014	41.53±1.78	0.05	0.344	ERS885752
ch15 1165-2	Near Cherno-byl, Ukraine	N 51° 12' 36" E 30° 0' 0"	0.113 (53)	June 2014		0.05	0.344	ERS885826
ch17 576	Teschow, Germany	N 53° 47' 17.408" E 12°37'51.02"		July 2007		0	0.483	ERS882975
ch19 925	Langoyene, near Oslo, Norway	N 59°52'18.642" E 10°43'25.212"		June 2010		0	0.359	ERS883109
ch20 443	Kraghede, Denmark	N 57° 12' 12.92' E 9° 59' 25.8"		August 2006		0	0.364	ERS881900
ch21 446	Tokay, Hungary	N 48° 7' 26.256" E 21° 25' 56.636"		October 2006		0	0.364	ERS882829

Table 2. ANOVA analysis of the effect of the abundance of butterflies, bumblebees and radiation level (μSv/h) on disease prevalence (all explanatory variables being log-transformed). The model was initially tested with all interactions and the non-significant ones were sequentially removed. The model had the statistics $F_{(4, 16)} = 10.41$, $r^2 = 0.65$, $P = 0.0002$.

	d.f.	**Sum of squares**	**F**	**P**	**Estimate (SE)**
Intercept			0.2401	0.6325	0.0042 (0.0086)
Log abundance bumblebees	1	0.0021	0.0657	0.8009	-0.0096 (0.0376)
Log abundance butterflies	1	0.0550	1.7409	0.2056	0.0643 (0.0487)
Log radiation	1	0.2059	5.5232	0.021	-0.-192 (0.0075)
Log radiation x log butterflies	1	0.2874	9.1027	0.0082	-0.1667 (0.0553)
Error	16	0.5056			

Table 3. Orthologs of the three main melanin synthesis pathways detected in the genomes of *Aspergillus fumigatus, A. niger* and *Microbotryum lychnidis-dioicae*.

Gene name	***A. fumigatus****	***A. niger***	***M. lychnidis-dioicae***	**Description**
DHN melanin pathway				
Abr2	**Afu2g17530 (Abr2)**	An01g13660 (McoB)	no ortholog	Fungal pigment MCO
	Afu1g15670	An01g14010 (McoA)	no ortholog	
	Afu4g14280	An03g03750 (McoC)	no ortholog	
		An04g10400 (McoO)	no ortholog	
		An05g02540 (McoP)	no ortholog	
Abr1	**Afu2g17540 (Abr1)** **Afu5g03790 (FetC)**	An14g05370 (BrnA) An01g11120 (McoE)	no ortholog	Fungal ferroxidase
		An01g08960 (McoH)	MVLG_01868	
		An15g05520 (McoK)		
Ayg1	**Afu2g17550 (Ayg1)**	An14g05350 (Ayg1)	no ortholog	
Arp2	**Afu2g17560 (Arp2)**	An02g00220	no ortholog	1,3,6,8-tetrahydroxynaphthalene reductase
Arp1	**Afu2g17580 (Arp1)**	An08g09920	no ortholog	Scytalone dehydratase
PKS1	**Afu2g17600 (Pks1)**	An03g05440	no ortholog	Polyketide synthase
	Afu4g00210 (EncA)	An04g09530	no ortholog	
	Afu4g14560	An09g05730 (FwnA)	no ortholog	
	Afu7g00160	An11g07310	no ortholog	
DOPA melanin pathway				
melC2	Afu3g01070	An01g09220 (MelC2)	no ortholog	Tyrosinase
		An03g00280	no ortholog	

melO		An12g01670	no ortholog	
		An09g02980	no ortholog	
	Afu4g14490	An12g05810 (McoJ)	MVLG_00670, MVLG_02184, MVLG_03092	Laccase
		An16g02020 (McoM)		
		An11g03580 (McoD)		
		An08g08450 (McoG)		
		An05g02340 (McoF)		
		An01g00860 (McoN)		
		An18g02690 (McoI)		
L-Tyrosine degradation pathway				
Tat	Afu2g13630	An02g05540	MVLG_06370	tyrosine aminotransferase
hppD	Afu2g04200	An11g02200	no ortholog	4-hydroxyphenylpyruvate dioxygenase
hmgA	Afu2g04220	An11g02180	no ortholog	homogentisate dioxygenase
fahA	Afu2g04230	An11g02170	MVLG_02428	fumarylacetoacetate hydrolase
maiA	Afu2g04240	An11g02160	no ortholog	maleylacetoacetate isomerase

**A. fumigatus* genes that belong to the DHN -melanin gene cluster (Tsai *et al.* 1999) are in bold; other genes share sequence similarity with the DHN-melanin genes.

Figure legends

Fig. 1. Sites used for sampling *Microbotryum lychnidis-dioicae* in Chernobyl, shown as white circles. Darker colors imply higher levels of background radiation.

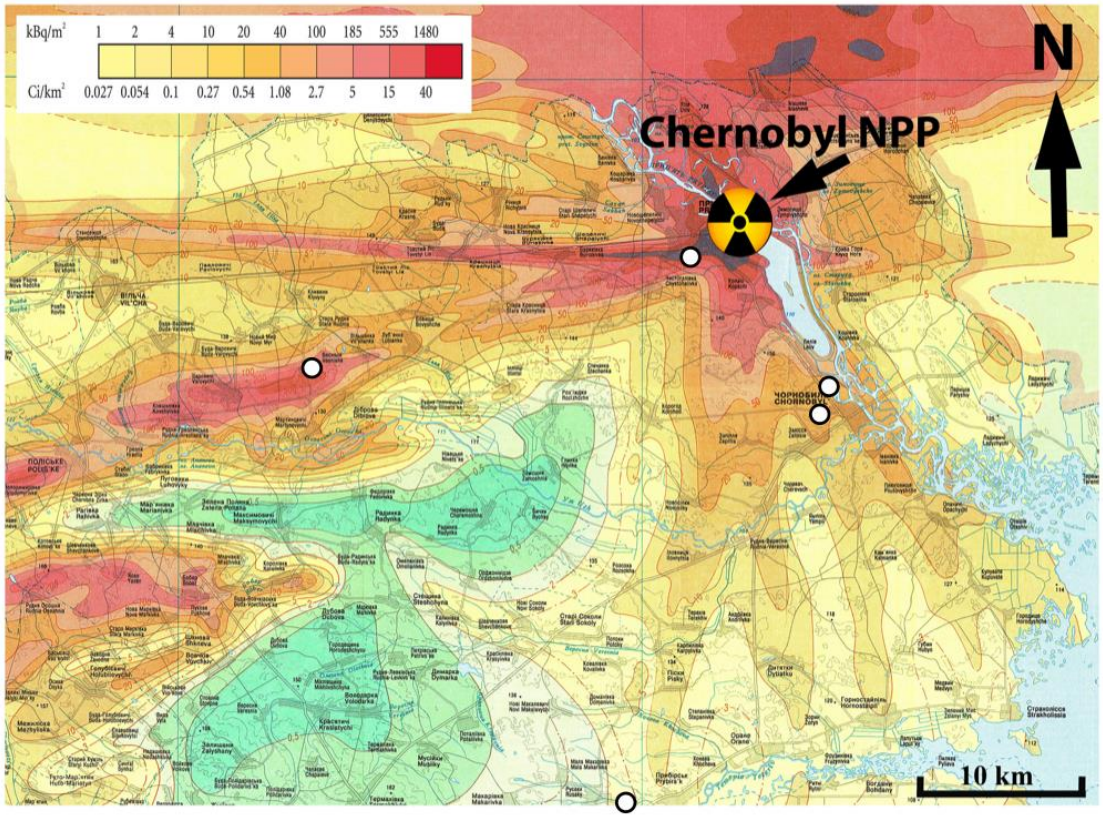


Fig. 2. Prevalence of the anther smut disease on *Silene latifolia* plotted against the logarithm of butterfly abundance.

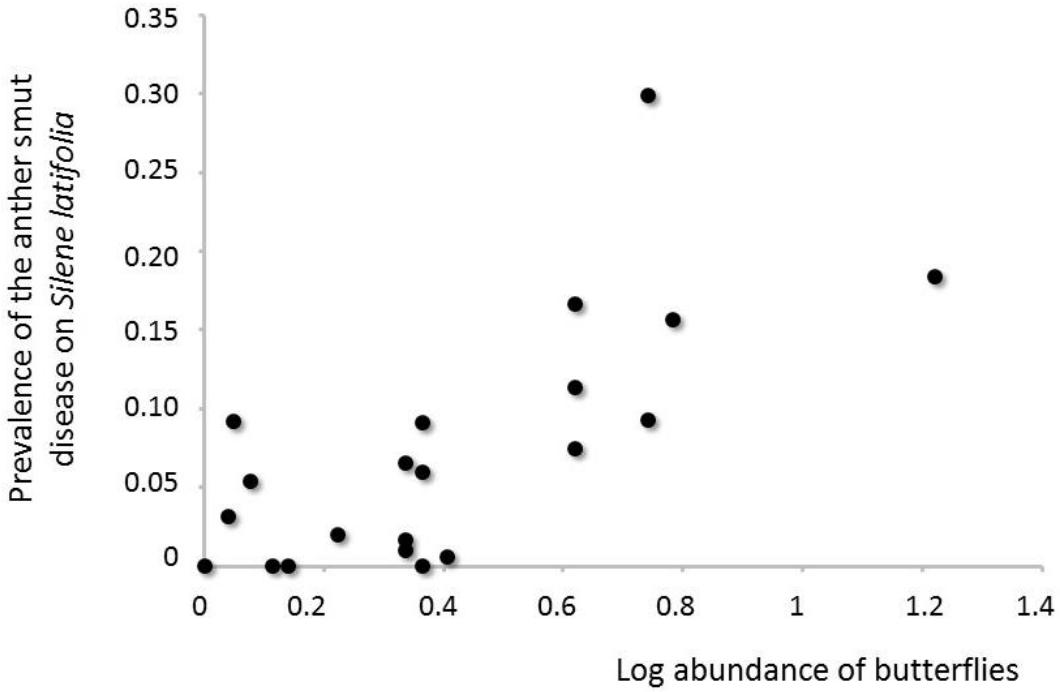


Fig. 3. Butterflies abundance plotted against the logarithm of the radiation level (μGy/h).

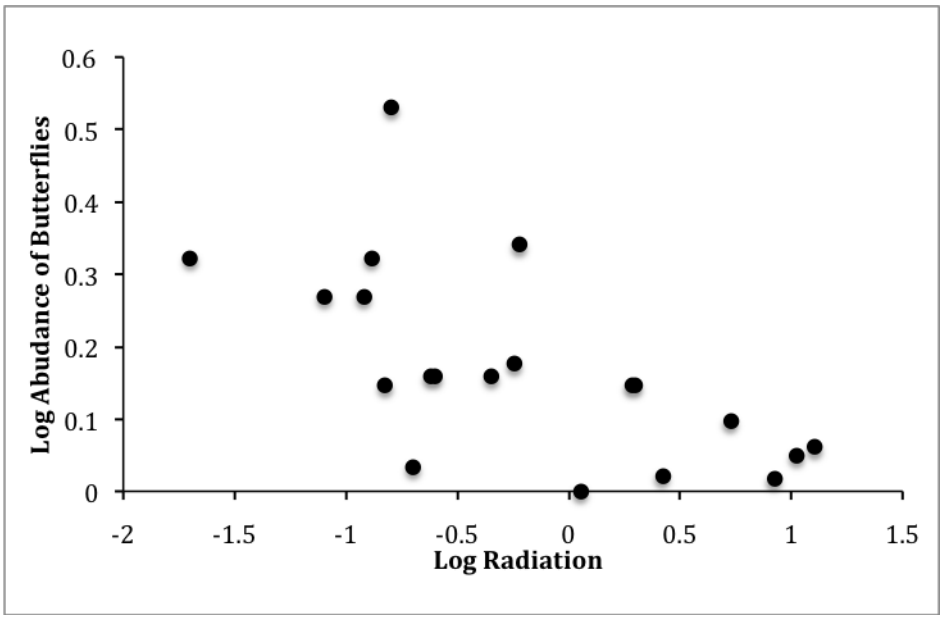


Fig. 4. Prevalence of the anther-smut disease on *Silene latifolia* plotted against the logarithm of radiation (μGy/h).

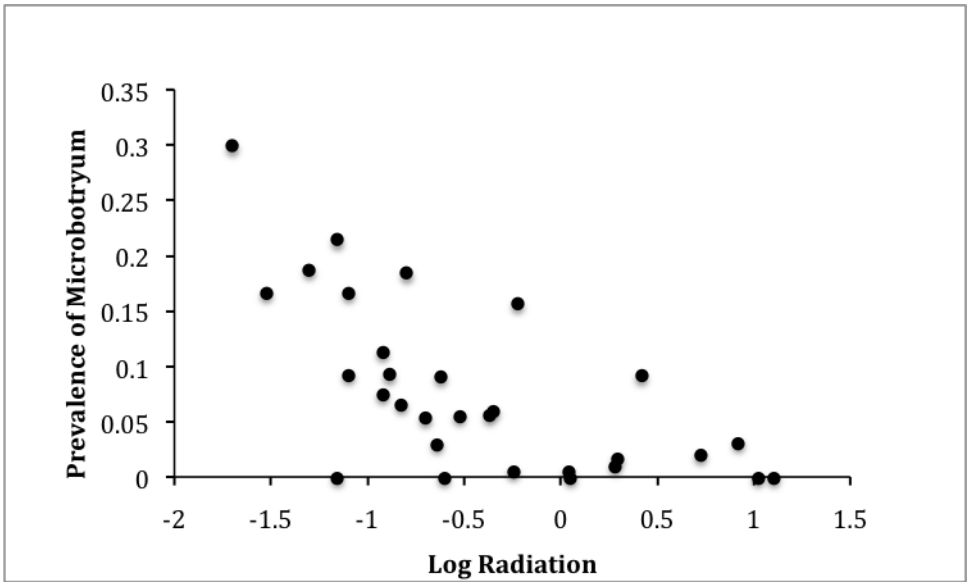


Fig. 5. Disease prevalence as a function of radiation level (μGy/h) and butterfly abundance (all log transformed).

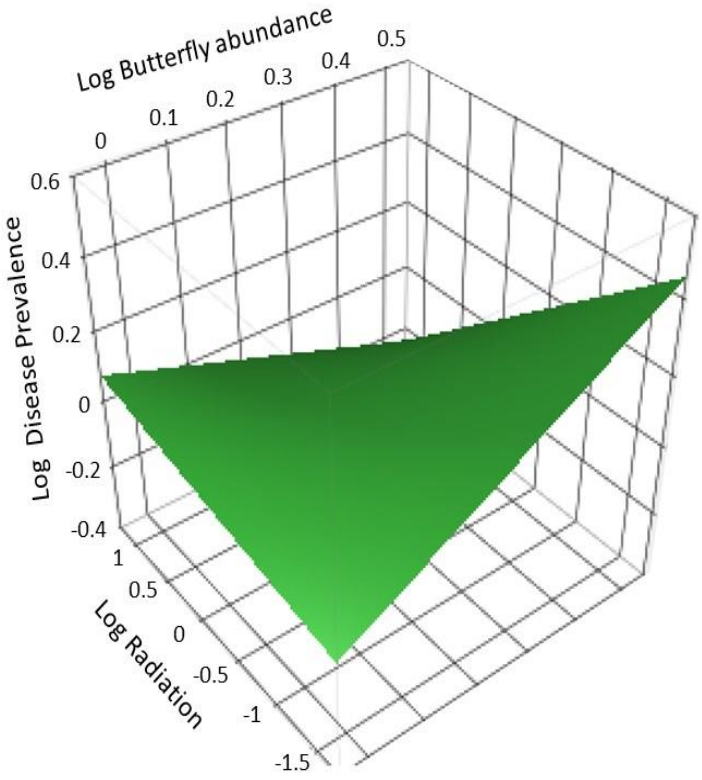


Fig. 6. Spore viability and fertility for *Microbotryum lychnidis-dioicae* strains from the Chernobyl area, estimated as the mean of haploid sporidia numbers (± standard error) growing from each of 100 diploid teliospores, plotted against log radiation level (μGy/h).

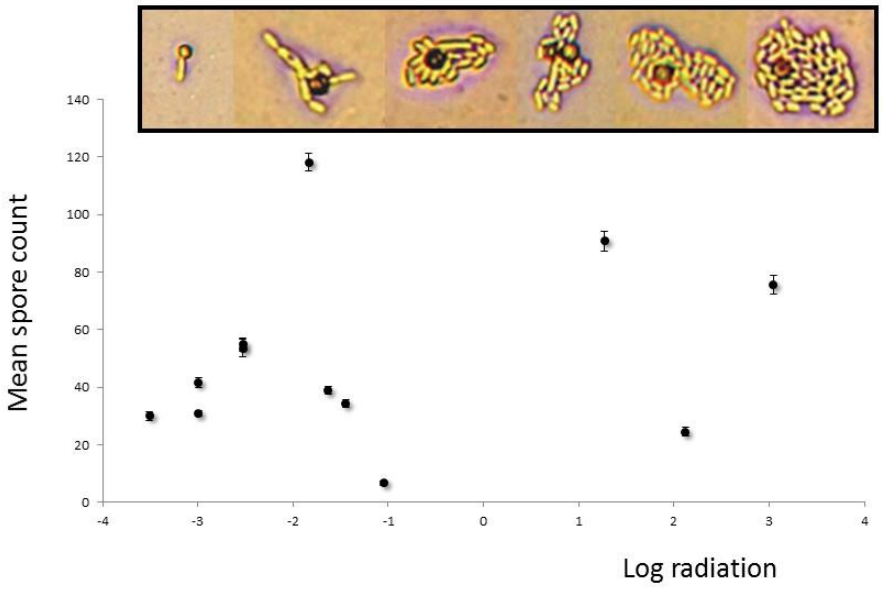
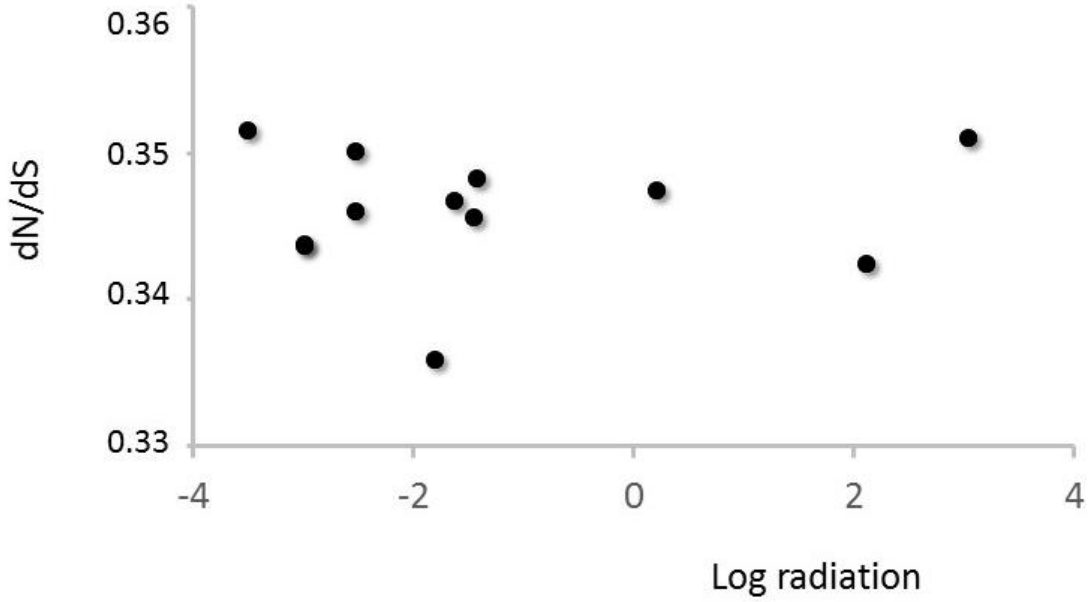


Fig. 7: Mean non-synonymous substitution rate over synonymous substitution rates (dN/dS) for *Microbotryum lychnidis-dioicae* strains from the Chernobyl area, plotted against log radiation level (μGy/h).



3.3 Discussion et Perspectives

L'analyse de populations issues de différents sites de Tchernobyl a montré que la prévalence de l'infection diminuait avec le taux de radiations, ce qui pourrait être causé par la baisse du nombre de pollinisateurs ou par une baisse de valeur sélective du champignon. On n'a pas observé de taux de dN/dS plus important dans les populations de Tchernobyl que dans le groupe témoin, contrairement à ce qui est observé chez un certain nombre d'autres espèces [90]. Le taux de dN/dS est même moins important que dans le groupe témoin, ce qui suggère que le champignon pourrait être résistant aux radiations, ou que l'augmentation de la pression de sélection purifiante suffit à contre-balancer l'augmentation du taux de mutations. Chez les champignons radio-résistants connus, la résistance semble causée par la synthèse de mélanine, mais la plupart des gènes nécessaires à la synthèse de la mélanine sont absents chez *M. lychnidis-dioicae*. La résistance aux radiations pourrait donc impliquer d'autres pigments.

4

Bases génétiques de l'adaptation chez *Microbotryum violaceum*

4.1 Introduction

Dans cette partie, nous nous sommes intéressés à l'adaptation chez les pathogènes, en prenant comme modèles les champignons biotrophes du complexe *M. violaceum*. Comprendre comment les pathogènes s'adaptent à leur environnement est très intéressant d'un point de vue conceptuel, pour mieux comprendre les mécanismes de la divergence adaptative et de la co-évolution, et c'est aussi primordial pour comprendre et prévoir l'émergence de nouvelles maladies. Cette partie comprend une synthèse sur les apports de la génomique à la compréhension des mécanismes de divergence adaptative chez les champignons, dans laquelle nous introduisons de manière assez générale la problématique de l'adaptation chez les champignons, entre autres chez les champignons pathogènes. Je me focaliserai donc dans la suite de cette introduction sur l'adaptation à l'hôte chez les champignons biotrophes, et chez les champignons du complexe *M. violaceum* en particulier.

4.1.1 L'adaptation chez les champignons pathogènes

Chez les champignons pathogènes, la plupart des espèces sont des spécialistes, spécifiques d'une seule espèce hôte ou d'un spectre d'hôte restreint. On trouve beaucoup d'espèces très proches infectant des hôtes différents, ces espèces de pathogènes ayant longtemps été cryptiques car

impossibles à distinguer sur le plan morphologique. Les spéciations ont souvent lieu par saut d'hôte [95], c'est-à-dire l'acquisition par une espèce de la capacité à infecter un nouvel hôte, suivie d'une spéciation entre la population présente sur l'hôte de départ et celle présente sur le nouvel hôte, avec perte de la capacité à infecter l'hôte ancestral. Les sauts d'hôtes induisent une sélection forte pour l'adaptation à un nouvel environnement, bien délimité et identifiable, et offrent donc une très bonne opportunité d'étudier la divergence adaptative.

Quand on s'intéresse à l'adaptation à l'hôte chez les pathogènes, deux grandes questions se posent :

- Qu'est-ce qui détermine la spécificité d'hôte et comment est-ce qu'une espèce acquiert la capacité à infecter un nouvel hôte ? Cette question est particulièrement importante pour comprendre l'apparition de nouvelles maladies.

- Comment est-ce qu'un pathogène co-évolue sur le long terme avec une espèce hôte donnée ?

Lorsqu'une espèce acquiert la capacité à infecter un nouvel hôte, cela peut entraîner une divergence, soit adaptative, en réponse à la pression pour s'adapter au nouvel hôte, soit consécutive d'un isolement reproductif, par exemple un isolement écologique créé par les différences d'habitat, les deux n'étant pas exclusifs l'un de l'autre. Au final, il peut y avoir un événement de spéciation. On n'a pas encore résolu la question de savoir, si, de manière générale, l'isolement reproducteur est une conséquence directe de la divergence adaptative, par exemple via des gènes qui ont des effets pléiotropes sur l'adaptation et la reproduction, ou une conséquence indirecte de la divergence adaptative, via des incompatibilités non directement adaptatives entre les génomes [96, 97]. Les incompatibilités génétiques désignent des relations épistatiques qui causent une baisse de valeur sélective. Le modèle d'incompatibilités le plus connu est celui des incompatibilités de Bateson-Dobzhansky-Müller (BDM), qui postule que des mutations sont fixées en allopatrie, et que des incompatibilités entre allèles de gènes différents se produisent dans les hybrides formés suite à un contact secondaire. Ces incompatibilités entraînent une baisse de valeur sélective des individus hybrides, et ainsi un isolement reproducteur [98, 96]. Dans ce modèle, l'isolement reproducteur est ainsi une conséquence indirecte de la divergence. Il est nécessaire d'étudier l'adaptation et la spéciation sur toute une diversité de modèles pour pouvoir tirer des conclusions générales sur les mécanismes de l'adaptation et les liens entre adaptation et spéciation.

Dans ce cadre, nous nous sommes intéressés à la question de l'adaptation et de la divergence adaptative dans le complexe *M. violaceum*, qui est un complexe de champignons biotrophes obligatoires, l'infection de la plante étant indispensable pour compléter le cycle de vie du champignon. Il existe trois grands types de champignons phytopathogènes : les champignons nécrotrophes qui tuent leurs hôtes et extraient leurs nutriments des tissus morts, les biotrophes qui infectent et exploitent leur hôtes sans les tuer, et les hémibiotrophes dont le cycle de vie inclut une phase biotrophe et une phase nécrotrophe [99]. De nombreuses études se sont intéressées à l'adaptation chez les champignons biotrophes, notamment d'un point de vue génomique (revues dans [100, 99, 101, 102]). La comparaison des génomes de champignon biotrophes et non biotrophes a révélé certaines caractéristiques qui seraient en lien avec le mode de vie biotrophe [103, 104]. En particulier, on trouve un large répertoire d'effecteurs potentiels, c'est-à-dire de petites protéines sécrétées qui pourraient être impliquées dans les interactions avec l'hôte et la virulence, et dont certaines sont sur-exprimées *in planta* [105]. Néanmoins, le plus souvent, l'implication de ces effecteurs dans le processus d'infection n'est pas démontrée. Les génomes des champignons biotrophes comportent peu de gènes codant pour des enzymes de dégradation des parois cellulaires, qui sont abondantes chez les champignons nécrotrophes, et au contraire montrent des expansions de familles de gènes codant pour des enzymes de modification de la paroi cellulaire, qui pourraient permettre aux champignons d'échapper au système immunitaire de l'hôte. On trouve aussi des expansions de familles de transporteurs membranaires impliqués dans l'absorption des nutriments. Les effecteurs potentiels sont parfois situés dans des régions génomiques particulières, comme des chromosomes facultatifs, au niveau des télomères ou dans des régions riches en éléments transposables, mais ils peuvent aussi être dispersés dans le génome comme chez *Ustilago maydis*. Ces effecteurs semblent de plus évoluer plus rapidement que le reste du génome, peut-être en lien avec leur localisation dans des régions "dynamiques" du génome ou avec la proximité d'éléments transposables [106].

4.1.2 L'adaptation dans le complexe *Microbotryum violaceum*

Les champignons du complexe *M. violaceum* sont des champignons biotrophes qui manipulent le système de reproduction de leur hôte, provoquant leur castration (cf. partie 1.2.1 p. 13). *Microbotryum violaceum* forme un complexe d'espèces où les spéciations ont le plus fréquemment lieu par saut d'hôte [22].

On pense que l'adaptation à la plante hôte serait le facteur de sélection le plus important dans l'évolution de ces champignons, mais des facteurs abiotiques comme la température pourraient également jouer un rôle. Plusieurs études se sont intéressées à l'adaptation et à la divergence adaptative chez les champignons du complexe *M. violaceum*. Il a été montré que des populations de *M. lychnidis-dioicae* étaient maladaptées à leur hôte *Silene latifolia* [29], ce qui suggère qu'il y aurait co-évolution entre *M. lychnidis-dioicae* et son hôte, mais que le pathogène serait en retard sur son hôte dans la "course aux armements", potentiellement à cause d'un taux de dispersion plus faible et/ou d'un effectif efficace plus faible [29].

Dans les populations naturelles, il y a une forte différenciation génétique entre espèces de *M. violaceum* et probablement très peu de flux de gènes [107, 19] même si de rares hybrides ont été détectés entre les espèces *M. lychnidis-dioicae* et *M. silenes-dioicae* qui parasitent respectivement *Silene latifolia* et *Silene dioica* [23]. L'isolement écologique induit par la différence d'habitat est seulement partiel, les espèces hôtes étant parfois présentes en sympatrie et la conjugaison ayant lieu à la surface des plantes [107]. Cela suggère que d'autres forces pourraient être impliquées dans l'isolement reproducteur. L'homogamie, c'est-à-dire d'union préférentielle de gamètes issues de la même espèce, ne semble pas jouer un rôle important [108, 109]. L'existence d'un isolement post-syngamie est par contre bien établie : il a en effet été montré que des hybrides inter-spécifiques obtenus expérimentalement présentaient une fertilité et une viabilité plus faible que les souches pures [21]. Le degré d'isolement reproducteur augmente avec la distance génétique [108], ce qui est cohérent avec l'accumulation progressive d'incompatibilités qui restreignent les possibilités d'hybridation. La baisse de valeur sélective des hybrides pourrait cependant aussi être liée à une mauvaise adaptation des génotypes hybrides aux plantes hôtes.

On ne connaît pas les bases moléculaires de l'infection par *M. violaceum*, ce qui rend difficile l'utilisation d'approches dites "gènes candidats", où on s'intéresse à des familles de gènes bien précises. Au début de ma thèse, on ne disposait pas d'outils de génétique fonctionnelle pour les champignons du complexe *M. violaceum*. Le modèle génétique le plus proche de *M. violaceum* est le champignon biotrophe *Ustilago maydis*, qui provoque la maladie du charbon du maïs, et où de nombreuses mutations affectant la virulence ont été identifiées. *Ustilago maydis* et *M. violaceum* ne sont néanmoins pas si proches, même s'ils ont longtemps été classés au sein de la même famille (*M. violaceum* était nommé *Ustilago violacea*, avant que *M. violaceum* ne soit déplacé dans une classe différente des basidiomycètes).

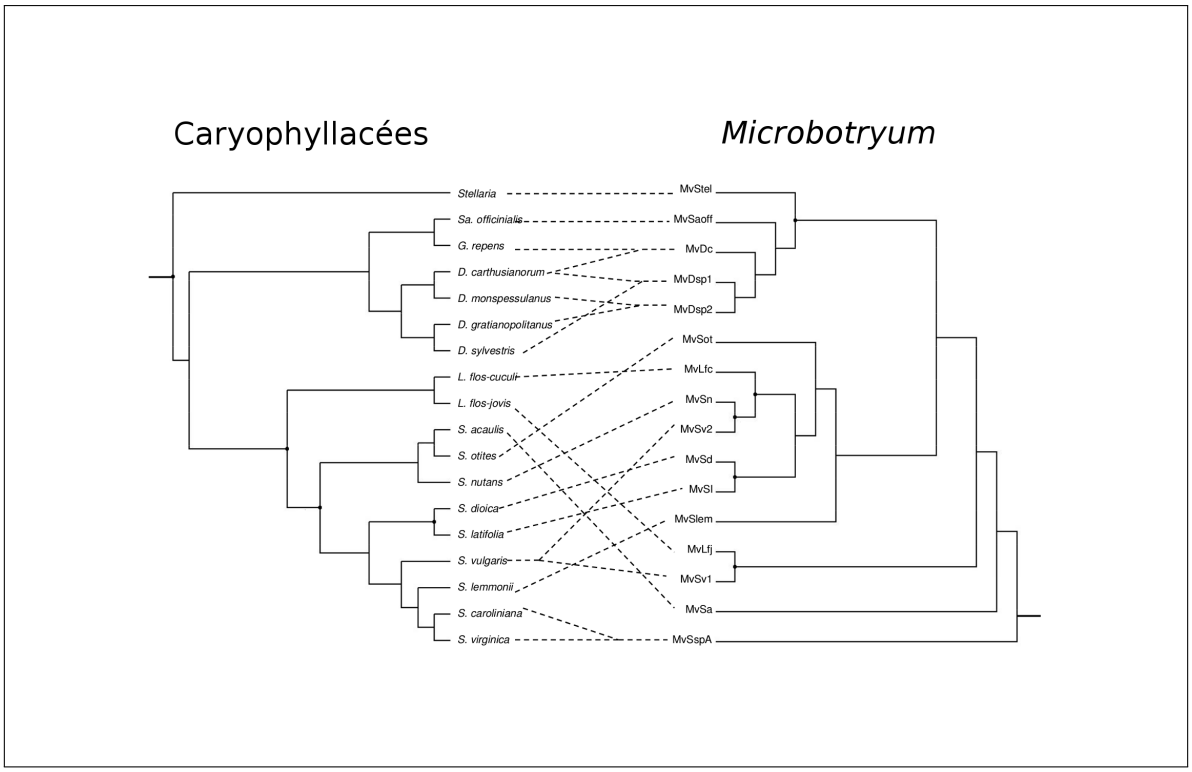


FIGURE 4.1 – Spéciations par sauts d'hôtes dans le complexe *Microbotryum violaceum*. La comparaison des phylogénies d'espèces du complexe *M. violaceum* et de leurs hôtes montre que les événements de spéciation ont le plus fréquemment lieu par saut d'hôte [22].

Quelques études ont essayé d'identifier les bases génétiques de l'adaptation à l'hôte dans le complexe *M. violaceum*. Une quarantaine de gènes candidats sous sélection diversifiante a été identifiée à partir des banques d'EST réalisées sur quatre espèces différentes du complexe *M. violaceum* [40]. Parmi ces gènes, certains possédaient des annotations en lien avec la synthèse de métabolites secondaires, la respiration en condition de stress ou la régulation de l'expression génétique, des fonctions compatibles avec un rôle dans les interactions hôte-pathogène. L'étude de 11 de ces gènes au sein de deux espèces du complexe *M. violaceum*, *M. lychnidis-dioicae* et *M. silenes-dioicae*, a montré que 9 d'entre eux étaient sous sélection purifiante en intra-espèce [41]. Cela suggère que des gènes différents pourraient être impliqués dans la divergence adaptative suite à un saut d'hôte, et dans la co-évolution à long terme avec un hôte donné. Mais rien ne prouve que les gènes détectés sous sélection diversifiante aient un rôle dans l'adaptation.

4.1.3 Questions et objectifs

Au début de ma thèse, un génome de référence venait tout juste d'être séquencé avec la technologie 454 dans le cadre d'un consortium pour l'espèce la plus étudiée du complexe *M. violaceum*, *M. lychnidis-dioicae*. Nous souhaitions générer des données de reséquençage pour étudier les bases génomiques de l'adaptation chez *M. lychnidis-dioicae*. En effet, une approche possible pour détecter des gènes ou régions génomiques impliqués dans l'adaptation est d'étudier le polymorphisme dans les populations. La sélection laisse des signatures particulières sur le polymorphisme : par exemple lorsqu'un allèle bénéfique est sélectionné positivement, la fréquence de cet allèle dans la population augmente, ainsi que celle des allèles non bénéfiques qui lui sont liés, créant une baisse locale de la diversité. C'est ce qu'on appelle un balayage sélectif.

Un certain nombre de méthodes ont été développées pour détecter des balayages sélectifs dans les génomes, à partir de l'analyse du niveau de diversité, du déséquilibre de liaison ou des spectres de fréquences alléliques [110, 111, 112, 113]. Ces méthodes supposent généralement que seule une petite portion du génome est affectée par la sélection, et cherchent donc à détecter des déviations par rapport à un modèle neutre. Des événements démographiques comme des goulots d'étranglements ou des expansions de tailles de populations peuvent également affecter le polymorphisme. Il n'est pas trivial de faire la différence entre les signatures laissées par la sélection et celles laissées par la démographie [114]. Une approche consiste à considérer que sélection n'affecte qu'une petite partie du génome alors que la démographie affecte l'ensemble du génome. On peut alors soit reconstituer l'histoire démographique de la population à partir de l'ensemble des données de polymorphisme et utiliser ce modèle démographique comme modèle neutre [115], ou simplement chercher des déviations par rapport au "fond génomique" (qui peut comporter la totalité des données de polymorphisme, ou un nombre restreint de loci pris au hasard) sans modéliser explicitement la démographie [112].

Disposer de données de polymorphisme sur l'ensemble du génome est aussi l'occasion d'étudier les variations du polymorphisme le long du génome et les facteurs qui influencent ces variations. Plusieurs formes d'auto-stop génétique sont en effet connues pour diminuer le niveau local de polymorphisme, comme l'auto-stop de sites liés à des allèles bénéfiques (souvent qualifié de balayage sélectif), ou l'élimination de sites liés à des allèles délétères ("background selection" en anglais) [116]. On s'attend à ce que la réduction de polymorphisme causée par l'auto-stop génétique soit plus forte dans les régions codantes car la pression de sélection est censée y être

plus élevée [117], et plus forte dans les régions de faible recombinaison [118]. D'autres facteurs peuvent aussi influencer le niveau de polymorphisme, comme le contenu en GC [119].

Pour étudier l'adaptation à l'hôte dans le complexe *M. violaceum*, nous nous sommes donc posés les questions suivantes :

- L'analyse du génome de référence peut-elle indiquer quels gènes pourraient a priori être impliqués dans les interactions avec l'hôte, et donc potentiellement l'adaptation à l'hôte ?
- Une approche de génomique des populations peut-être permettre de détecter des régions sous sélection au niveau intra-spécifique, et donc de nouveaux gènes candidats ?
- Y a-t-il un lien entre adaptation et architecture génomique, par exemple sous la forme d'"îlots de pathogénicité" regroupant de nombreux gènes candidats ?
- Comment varie la diversité le long du génome dans les espèces *M. lychnidis-dioicae* et *M. silenes-dioicae*, et par quels facteurs est-elle influencée ? En particulier, observe-t-on une corrélation positive entre diversité et taux de recombinaison et une corrélation négative entre diversité et densité en sites codants, qui pourraient indiquer que diverses formes d'auto-stop génétiques influencent le niveau de polymorphisme le long du génome ?

Dans ce cadre, j'ai participé à l'analyse du génome de référence (cf. 4.3 p. 140). J'ai mené une étude de génomique des populations sur deux espèces jumelles du complexe *M. violaceum*, *M. lychnidis-dioicae* et *M. silenes-dioicae* (cf. 4.4 p. 165). J'ai aussi participé à la rédaction d'un article de synthèse sur la génomique adaptative chez les champignons (cf. 4.2 p. 117).

4.2 Article 1 : Fungal evolutionary genomics provides insight into the mechanisms of adaptive divergence in eukaryotes

Dans les congrès sur la spéciation et la divergence adaptative, les champignons sont souvent peu représentés. En effet, la spéciation est le plus souvent étudiée dans des modèles animaux ou plantes, tels que les emblématiques pinsons de Darwin ou les plantes du genre *Mimulus*. Comparés aux champignons, les plantes et animaux offrent en effet des phénotypes plus variés

et plus souvent directement observables, et les espèces sont plus souvent délimitables à l'aide de critères uniquement morphologiques.

Les champignons représentent cependant des modèles extrêmement intéressants pour étudier les phénomènes de divergence adaptative et de spéciation chez les Eucaryotes. Beaucoup de champignons ont des génomes bien annotés, grâce à l'homologie avec des modèles génétiques comme *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe* ou *Neurospora crassa*. Le développement de la biologie moléculaire a permis la découverte de nombreuses espèces cryptiques adaptées à des environnements différents, et présentant différents niveaux de divergence et/ou d'introgression. Les champignons présentent aussi des avantages d'un point de vue expérimental, avec des petits génomes, des temps de génération courts, des phénotypes simples à mesurer et l'accès direct à la phase haploïde.

Nous avons donc écrit l'article de synthèse qui suit, dans le but de présenter l'intérêt des modèles champignons pour étudier la divergence adaptative et la spéciation. Pour cela, nous avons d'abord introduit de manière générale la problématique de l'adaptation chez les champignons, puis nous avons illustré les apports des études sur les champignons dans la compréhension de la divergence adaptative, et enfin pointé des directions d'études futures. Dans cet article, publié dans Molecular Ecology, j'ai rédigé une partie de la partie sur l'importance de l'hybridation et des introgressions dans la divergence adaptative.

MOLECULAR ECOLOGY

Molecular Ecology (2014) **23**, 753–773 doi: 10.1111/mec.12631

INVITED REVIEWS AND SYNTHESES

Fungal evolutionary genomics provides insight into the mechanisms of adaptive divergence in eukaryotes

PIERRE GLADIEUX,*†‡ JEANNE ROPARS,*† HÉLÈNE BADOUIN,*† ANTOINE BRANCA,*† GABRIELA AGUILETA,§¶ DAMIEN M. DE VIENNE,§¶** RICARDO C. RODRÍGUEZ DE LA VEGA,*† SARA BRANCO‡ and TATIANA GIRAUD*†

*Ecologie, Systématique et Evolution, UMR8079, University of Paris-Sud, Orsay 91405, France, †Ecologie, Systématique et Evolution, CNRS, UMR8079, Orsay 91405, France, ‡Department of Plant and Microbial Biology, University of California, Berkeley, CA 94720-3102, USA, §Center for Genomic Regulation (CRG), Dr, Aiguader 88, Barcelona 08003, Spain, ¶Universitat Pompeu Fabra (UPF), Barcelona 08003, Spain, **Laboratoire de Biométrie et Biologie Evolutive, Université Lyon 1, CNRS, UMR5558, Villeurbanne 69622, France

Abstract

Fungi are ideal model organisms for dissecting the genomic bases of adaptive divergence in eukaryotes. They have simple morphologies and small genomes, occupy contrasting, well-identified ecological niches and tend to have short generation times, and many are amenable to experimental approaches. Fungi also display diverse lifestyles, from saprotrophs to pathogens or mutualists, and they play extremely important roles in both ecosystems and human activities, as wood decayers, mycorrhizal fungi, lichens, endophytes, plant and animal pathogens, and in fermentation or drug production. We review here recent insights into the patterns and mechanisms of adaptive divergence in fungi, including sources of divergence, genomic variation and, ultimately, speciation. We outline the various ecological sources of divergent selection and genomic changes, showing that gene loss and changes in gene expression and in genomic architecture are important adaptation processes, in addition to the more widely recognized processes of amino acid substitution and gene duplication. We also review recent findings regarding the interspecific acquisition of genomic variation and suggesting an important role for introgression, hybridization and horizontal gene transfers (HGTs). We show that transposable elements can mediate several of these genomic changes, thus constituting important factors for adaptation. Finally, we review the consequences of divergent selection in terms of speciation, arguing that genetic incompatibilities may not be as widespread as generally thought and that pleiotropy between adaptation and reproductive isolation is an important route of speciation in fungal pathogens.

Keywords: *Coccidioides*, competition, effector, gene regulation, genetic incompatibilities, genomic islands, local adaptation, *Neurospora*, *Penicillium*, positive selection, *Saccharomyces*, yeast

Received 13 September 2013; accepted 4 December 2013

Introduction

Understanding the genetic and genomic processes behind adaptive phenotypes remains a holy grail in biology (Olson-Manning *et al.* 2012; Alfoeldi & Lindblad-Toh 2013). This exercise is not purely academic; it is also crucial for predicting the ways in which organisms will respond to global crises, such as climate change, changes in landscapes and ecosystems, the spread of invasive species, the emergence of pathogens, resistance to drugs and vaccines and increasing food demand. Key challenges currently include identifying the genes involved in ecologically relevant traits and understanding the nature, timing and architecture of the genomic changes governing the origin and

Correspondence: Tatiana Giraud, Fax: +33 1 69 15 46 97; E-mail: Tatiana.Giraud@u-psud.fr

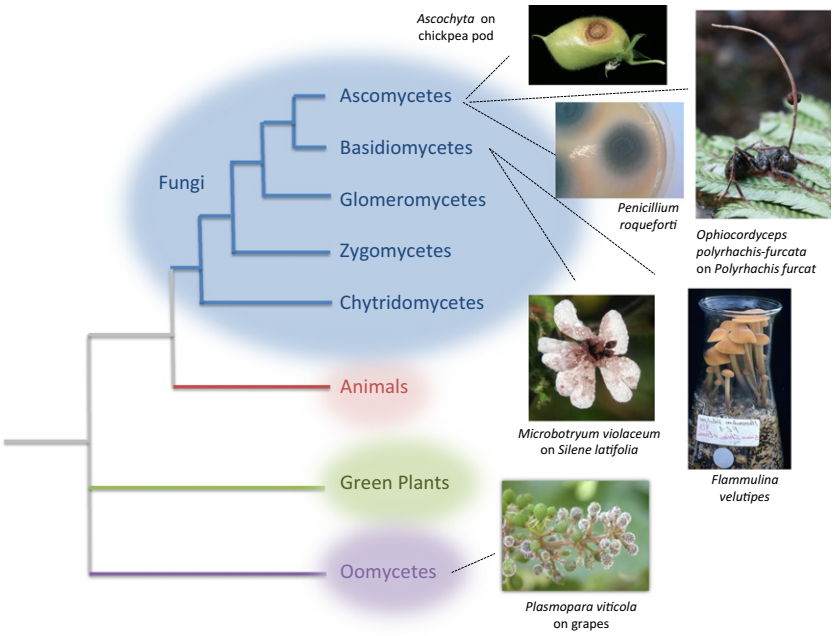


Fig. 1 The tree of life, including the main fungal clades, with illustrative images: for ascomycetes, *Ascochyta* on chickpea pod, *Penicillium roquefor* on a Petri dish, *Ophiocordyceps polyrhachis-furcata* on *Polyrhachis furcat*; for basidiomycetes, *Microbotryum violaceum* on *Silene lafolia* and *Flammulina velupes*; for oomycetes, *Plasmopara vicola* on grapes. Oomycetes were long considered to be fungi, because of convergent fungal-like traits, such as a filamentous form, for example, but they are actually more closely related to brown algae and are not considered in this review.

processes of local adaptation, lineage divergence and ecological speciation. These processes will be grouped together below under the umbrella term 'adaptive divergence'. The possibility of sequencing thousands of related genomes has opened up new possibilities for tackling these crucial issues from a new perspective, providing unprecedented insight into the genomic bases and processes of adaptive divergence (Luikart *et al.* 2003; Alfoeldi & Lindblad-Toh 2013; Fu & Akey 2013; Leducq 2014).

The sequencing of multiple genomes can reveal the relative prevalence of amino acid substitutions, genomic rearrangements, gene gains, gene losses and introgressions as sources of genomic change. However, it remains difficult to address these questions in most eukaryotes, due to their large genomes, phenotypic complexity, long generation times and the difficulty of validating computational inferences by setting up genetic transformation protocols. In this respect, fungi are ideal model organisms. They have a simple morphology, inhabit contrasting and well-identified ecological niches and display considerable diversity, including a huge variety of lifestyles (from saprotrophs to pathogens, mutualists and even predators). Fungi have small genomes and tend to have short generation times and to be amenable to experimental approaches, making it possible, for example, to validate gene functions or to conduct experimental evolution studies (Dettman *et al.* 2007; Schoustra *et al.* 2009; Ajouz *et al.* 2010; Anderson *et al.* 2010; Schoustra & Punzalan 2012; Barrick & Lenski 2013).

Furthermore, the hundreds of genomes available make it possible to use similarity data from public databases for gene annotation, based on information gained from the dissection of gene function in model fungi, such as the baker's yeast *Saccharomyces cerevisiae* and the orange bread mould *Neurospora crassa*. Systematic gene deletion and phenotype screening have been carried out in these models (Winzeler *et al.* 1999; Giaever *et al.* 2002; Colot *et al.* 2006; Hillenmeyer *et al.* 2008). More than 160 sequenced fungal genomes have been published to date and more than 250 are available in total. Although fungi are often considered to be microbes, they are actually phylogenetically close to animals (Fig. 1), thus sharing physiological complexity and similar modes of genomic architecture and evolution with metazoans and other eukaryotes (Tirosh *et al.* 2006; Stajich *et al.* 2009). Inferences drawn from fungi can therefore provide information that can be extrapolated to the genomic processes of adaptive divergence in eukaryotes (Tirosh *et al.* 2006). Fungi thus have considerable potential for use as tractable models of agronomic, medical, industrial and ecological importance (Stajich *et al.* 2009; Gladieux *et al.* 2010a).

There are millions of fungal species, spanning billions of years of evolution and displaying diverse lifestyles that evolved on multiple occasions in the tree of life (James *et al.* 2006), and playing extremely important roles in both ecosystems and human activities. The process of adaptive divergence can be studied in (i) domesticated fungi, used for the production of food and beverages and for bioindustry in general (Fay & Benavides 2005; Liti *et al.* 2009; Libkind *et al.* 2011), (ii) pathogens, which frequently adapt to new hosts, causing emerging diseases that have a major negative impact on human welfare, through agricultural and eco-

nomic losses, and threats to biodiversity (Fisher *et al.* 2012), (iii) symbiotic fungi, including endophytes, that protect their hosts against herbivory or enhance plant growth and resistance to stress (Zuccaro *et al.* 2011), and mycorrhizal fungi, which provide nitrogen and phosphate to wide range of plants, and (iv) saprotrophic fungi, the only organisms capable of breaking down lignin to any great extent (Floudas *et al.* 2012).

We review here recent insights into the genomic mechanism of adaptive divergence gleaned from fungi, ranging from adaptive polymorphism to speciation, and improving our understanding of these processes in eukaryotes. We first outline the various ecological sources of divergent selection acting on fungi, with the aim of providing an overview for nonmycologists of the selection agents promoting adaptive divergence in these organisms. We then review evidence for the roles of different types of genomic changes, showing that gene loss and changes to gene expression and genomic architecture are important mechanisms of adaptation, in addition to the more widely recognized processes of amino acid substitution and gene duplication. We also review interspecific sources of novel genomic variation on which selection can act and suggest that introgression, hybridization and horizontal gene transfer (HGT) are of much greater importance than is often suggested. In addition, we show that transposable elements can influence all these processes, serving as important adaptation factors. Finally, we review the evidence relating to the direct translation of divergent selection into reproductive isolation, through either pleiotropy between adaptation and reproductive isolation or the appearance of genetic incompatibilities between species due to rapid evolution. We argue that the contribution of classical epistatic incompatibilities, although real and important, may be less pervasive than currently thought. Specific fungal terms are defined in the glossary.

Sources and targets of ecological divergence

Ecological divergence in fungi may be due to abiotic and biotic factors, including intra- and interspecific interactions, or artificial selection (domestication).

Abiotic factors

Environmental factors, such as temperature and edaphic conditions, are drivers of ecological differentiation in diverse fungi. Differences in temperature regimes, for instance, maintain differentiated populations of crop pathogens, adapted to different climates or seasons (Enjalbert *et al.* 2005; Frenkel *et al.* 2010; Mboup *et al.* 2012), and also strongly affect the distribution of forest pathogens (Vacher *et al.* 2008). Climate change has been implicated in the recent emergence and probable expansion in the future of several diseases caused by fungi (Fisher *et al.* 2000; Bergot *et al.* 2004; Fabre *et al.* 2011). In *Neurospora crassa*, a population genomics approach revealed that temperature and latitude were important drivers of local adaptation. Indeed, genome scans have detected islands of genomic differentiation including genes involved in both the response to cold temperatures and the circadian cycle, and growth rate assays have confirmed differences in fitness between populations (Ellison *et al.* 2011). In addition, carotenoid accumulation in *Neurospora* has been correlated with latitude (Luque *et al.* 2012). Carotenoids protect against UV irradiation and their production is induced by light. Pigment levels may therefore be subject to selection and may underlie population differentiation in these fungi.

Edaphic factors have been found relevant in the niche requirements of forest mutualists and decomposers, with species showing clear preferences based on soil chemistry (Frankland 1998; Branco 2010). Humidity and other abiotic factors associated with elevation also act as strong selection agents in tree pathogens and symbionts, affecting their distributions (Cordier *et al.* 2012).

There may be trade-offs between different types of performance subject to selection by different abiotic agents. For instance, trade-offs may have led to different strategies in closely related pathogenic fungal species, some being better at overwintering, whereas others are better at sporulating late in the season (Giraud *et al.* 1997; Feau *et al.* 2012).

Pathogens and mutualists: the host as a selective agent

Unlike abiotic factors, biotic agents evolve and constitute discrete ecological niches in sympatry, therefore causing strong disruptive selection. This makes fungal pathogens, which are dependent on their living hosts, attractive systems for the study of adaptive divergence. Host shifts have been associated with multiple cases of recently diverged sibling species in numerous pathogens (Giraud *et al.* 2008; de Vienne *et al.* 2013), and ecosystems modified by humans continually provide examples of new fungal diseases emerging on new hosts (Anderson *et al.* 2004; Gladieux *et al.* 2010a; Fisher *et al.* 2012), providing ample opportunity to investigate the early stages of adaptive divergence. The strong selection imposed by hosts is conducive to ecological differentiation in fungal pathogens, through an increase in the frequency of locally advantageous alleles and prevention of the immigration of locally deleterious ancestral alleles (Giraud *et al.* 2010). Furthermore, specialist pathogens are more efficient in the arms race with their hosts (Whitlock 1996). Most fungal pathogens are indeed specialists, evolving by host shifts (de Vienne *et al.* 2013).

In pathogens of animals, virulence reflects a capacity to neutralize or resist the somatic adaptive immune system of the host (Upadhya *et al.* 2013) and, for pathogens of homeothermic animals, an ability to cope with temperatures of 37 °C or higher (which is usually a strong barrier for fungi). Amphibians, which are cold-blooded, act as hosts to several species of fungal pathogens; the recent emergence and spread of chytridiomycosis, caused by the chytrid *Batrachochytrium dendrobatidis,* has attracted much attention as it poses a threat to amphibians worldwide (Fisher *et al.* 2009).

By contrast to the situation in animals, plant defenses are based on the innate immunity of each cell and on systemic signals emanating from sites of infection (Jones & Dangl 2006). Plant immune responses can be broadly grouped into two major layers: responses triggered by general microbe-associated molecular patterns (*e.g.* chitin) and those triggered by isolate-specific pathogen effectors, recognized by specific nucleotide-binding and leucine-rich repeat (NB-LRR) resistance proteins (Schulze-Lefert & Panstruga 2011). The specific defense response typically involves localized cell death, which completely prevents microbial growth, thus exerting strong selection on fungi (Giraud *et al.* 2010). Over the last 100 years, plant breeders have mostly used resistance genes of major effect – usually encoding NB-LRR proteins – to control infectious fungal diseases, as they are associated with a phenotype that can be easily selected and display simple Mendelian inheritance (Crute & Pink 1996). However, with only a few exceptions, such resistances have rapidly broken down (McDonald 2010), providing spectacular examples of adaptive evolution (Brown 1994; Guerin *et al.* 2007; Terauchi & Yoshida 2010; Xhaard *et al.* 2011). More generally, widely distributed, high-density and genetically uniform populations, such as those of cultivated crops, regularly select for new pathogens (Stukenbrock & McDonald 2008). Population genetics and phylogenetics have provided evidence that many fungal pathogens have emerged through host shifts, host-range expansion or host tracking over the last 10 000 years, following the domestication of the affected crops (Couch *et al.* 2005; Munkacsi *et al.* 2006; Stukenbrock *et al.* 2007; Zaffarano *et al.* 2008; Frenkel *et al.* 2010; Gladieux *et al.* 2010b, 2011; Silva *et al.* 2012). By contrast, host plants do not seem to exert strong selective pressure on mycorrhizal fungi, and there are few examples of strict mycorrhizal host specificity (Bruns *et al.* 2002).

Ecological interactions between fungi

Ecological interactions between competitors also act as strong biotic selection agents and may be involved in processes of ecological divergence. Competition can favour the exploitation of underused resources, leading to character displacement and niche expansion, resource polymorphism and speciation (Bono *et al.* 2013). Intraspecific competition and interspecific competition in natural conditions have been documented in ectomycorrhizal and saprotrophic fungi (Boddy 2000; Kennedy 2010) and in fungal pathogens (Koskella *et al.* 2006; Lopez-Villavicencio *et al.* 2007; Staves & Knell 2010; López-Villavicencio *et al.* 2011). The mechanisms involved in competitive interactions include the production of toxins inhibiting competitors (Kaiserer *et al.* 2003; Sass *et al.* 2007), vegetative incompatibility and hyphal interference, resulting in the death of incompatible hyphae after somatic fusion (Glass *et al.* 2000; Silar 2005). Competition between fungi shapes the distribution of species in the field and can play an important role in species divergence. In addition, many of the virulence traits in fungal pathogens also have a function in competition outside of the host (Morris *et al.* 2009; Stergiopoulos *et al.* 2012). Predation can also play a selective role in divergence, and several authors have suggested that many emerging animal pathogens might actually have acquired virulence traits, such as resistance to phagocytosis, through their initial selection for the avoidance of predation by amoebas or nematodes (Casadevall *et al.* 2003; Greub & Raoult 2004).

Sexual selection and assortative mating

Sexual selection is another biotic source of divergence. It acts on traits involved in mate recognition, conditioning mate preference and influencing the genetic architecture passed on to progenies. The existence of strong consistent biases in nucleus fertilization in crosses between decomposers of the species *Schizophyllum commune* has suggested that sexual selection may be of relevance in fungi (Nieuwenhuis *et al.* 2010). Also, yeasts display a preference for higher levels of pheromone production, which might result from sexual selection (Jackson & Hartwell 1990; Rogers & Greig 2009). However, it remains unclear how widespread sexual selection is or how it promotes divergence in fungi.

Another important force underlying adaptive divergence is selection against maladaptive hybrids, that is, reinforcement or selection for assortative mating in sympatry. Several cases of reinforcement have been reported, particularly in toadstools (Kohn 2005; Giraud *et al.* 2008; Giraud & Gourbiere 2012), but also in ascomycetes of the genus *Neurospora* (Turner *et al.* 2011).

Domestication

Domestication is a specific case of adaptive divergence in response to artificial selection. Diverse fungi have

been domesticated for the fermentation of food products, such as wine, beer, bread (*Aspergillus oryzae, Saccharomyces sp.*), dried sausages (*Penicillium nalgiovense*), red fermented rice (*Monascus purpureus*) and cheeses (*Penicillium sp.*, Fig. 1). Additional characteristics have also been selected, including growth, mycelium thickness, colour or lipolytic and proteolytic activities in cheese *Penicillium* species. Other domesticated fungi (i) have organoleptic properties that are appreciated (e.g. the button mushroom *Agaricus bisporus*, the shiitake *Lentinula edodes*, the jelly ear *Auricularia auricula-judae*), (ii) produce secondary metabolites useful for medical purposes, such as penicillin (*Penicillium rubens*), or (iii) are a source of enzymes useful for the biodegradation of plant polysaccharides in biofuel production (*Trichoderma reesei*).

Genetic bases of adaptation: within-species changes

We have outlined above the different types of selection agents potentially involved in adaptive divergence in fungi. The genomic bases of such adaptations are beginning to be dissected in fungi, and we review below the exciting findings obtained to date, focusing first on intraspecific genomic changes, that is, mutations leading to amino acid substitutions, gene regulation and gene gains and losses.

Amino acid substitutions

Amino acid changes can occur due to errors in DNA replication or the addition of nucleotides after the insertion and excision of transposable elements (TEs) (Daboussi & Capy 2003), as shown, for instance, for genes controlling spore colour (Colot *et al.* 1998) and encoding effectors (Farman 2007; Rep & Kistler 2010).

Candidate gene analyses have shown that amino acid changes are pervasive in underlying fungal adaptation (reviewed in Aguileta *et al.* 2009; Stergiopoulos *et al.* 2007; Stergiopoulos & de Wit 2009; Stukenbrock & McDonald 2009). We focus here on the most recent findings stemming from the analysis of genomic sequence variation in natural populations/species. Genomic approaches have the advantage of not requiring a priori information on the genes involved in adaptation and can therefore be used to elucidate the functions involved in host/habitat specialization and the environmental factors driving adaptive divergence, through a 'reverse-ecology' approach.

Genes under diversifying selection have been identified from analyses of expressed sequence tags (ESTs) collected in closely related species with different ecological profiles. In the *Microbotryum* and *Botrytis* pathogen species complexes and in *Saccharomyces* yeasts, for instance, this approach has identified several dozens of genes that have been subject to recurrent selection, generating repeated adaptive amino acid changes in the same gene or even at the same site in different variants (Li *et al.* 2009; Aguileta *et al.* 2010, 2012). A large proportion of these genes under positive selection have been annotated as transmembrane proteins, putative secreted proteins or cell wall proteins presumably involved in transporter activities and establishing communication with the host cell and the external environment. Interestingly, most of the genes that have been through episodes of adaptive diversification between *Microbotryum* species specialized on different hosts appeared to have subsequently evolved under strong functional constraints in the lineages remaining specialized on a given host plant (Gladieux *et al.* 2013), indicating they are not involved in the arms race against their host.

Whole-genome sequencing in plant pathogens has revealed the existence of an arsenal of effectors expressed during infection, revolutionizing our understanding of plant–fungus interactions (reviewed in (de Jonge *et al.* 2011; Oliver 2012; Stergiopoulos & de Wit 2009). Effectors are generally defined as secreted proteins that manipulate host innate immunity, enabling infection to occur (Dodds & Rathjen 2010). Resequencing data for the wheat pathogens *Blumeria graminis* and *Zymoseptoria tritici* have identified several hundred putative effector genes displaying signatures of positive selection (Stukenbrock *et al.* 2011; Wicker *et al.* 2013). Such a pervasive rapid evolution of effectors in fungal pathogens is consistent with a permanent arms race, with increases in the complexity of recognition systems in hosts being matched by the development of new systems for escaping recognition in the pathogen. Other examples of adaptive amino acid change consistent with rampant coevolution with hosts or competitors in fungal lineages include toxin biosynthesis gene clusters (Ward *et al.* 2002; Carbone *et al.* 2007) and secreted proteins in the amphibian-killing fungus *B. dendrobatidis* (Farrer *et al.* 2013; Rosenblum *et al.* 2013).

More generally, with the availability of larger genome-wide data sets for multiple members of same genus or family, it is becoming increasingly possible to identify the gene-specific or genome-wide effects of selection associated with changes in life history traits or ecological strategies. For instance, comparison of the genomes of human pathogens of the genus *Coccidioides* genomes with those of closely related species has revealed lineage-specific accelerations in the amino acid substitution rate in a set of about 70 genes displaying functional enrichment in genes associated with biopolymer metabolic processes and RNA metabolic processes

(Sharpton *et al.* 2009). Footprints of positive selection, in terms of amino acid changes, have also been found in domesticated fungi, due to the strong selection exerted by humans. In the fungus used for sake production for instance, *A. oryzae*, one of the genes with the strongest signal of positive selection with respect to its wild relatives encodes a glutaminase catalysing the hydrolysis of L-glutamine, a widely used food flavour enhancer present at high levels in sake (Gibbons *et al.* 2012).

The genetic code itself can also provide unexpected adaptive amino acid changes. Many fungi have altered their genetic code to incorporate serine residues at sites at which leucine was previously incorporated. This may sound highly deleterious, but experiments in *Candida albicans* have shown that such misincorporation is well tolerated (Bezerra *et al.* 2013). Basic growth rates were lower, but many novel phenotypes were retrieved, including enhanced growth on novel media, drug resistance and response to human immune cells.

Gene family size changes, gene deletions

As amino acid changes, gene duplication was identified early on as a possible mechanism of adaptive innovation (Ohno 1970). The sequencing of a large number of fungal genomes has made it possible to identify gene families that have been expanded, but also gene families that have been reduced in particular lineages or that have become species specific (Ames *et al.* 2012; Han *et al.* 2013; Leducq 2014) (Fig. 2). Species- or population-specific genes may be derived from within-group innovation by such rapid divergence that homology is not recognized (Kellis *et al.* 2004) or may be obtained by horizontal or lateral gene transfer through the non-vertical acquisition of genome fragments containing coding sequences (Keeling 2009). The reduction or complete loss of gene families has also often been found to reflect ecological shifts (Casadevall 2008).

Gene family expansion. Expanded gene sets are widespread and have received special attention due to their role in adaptive divergence. Duplication allows functional diversification or an adaptive increase in enzyme production. For example, families of genes encoding secreted proteins (often regarded as potential effectors), nutrient transporters and enzymes acting on carbohydrates are commonly expanded in fungal obligate parasites and mutualist symbionts, such as mycorrhizal fungi (Martinez *et al.* 2004; Martin *et al.* 2008; Powell *et al.* 2008; Butler *et al.* 2009; Duplessis *et al.* 2011; Ohm *et al.* 2012). Families of membrane-bound transporters, including amino acid permeases and sugar transporters, allow the uptake of nutrients by pathogens and of carbohydrates by symbiotic fungi from their host plants in exchange for nitrogen and phosphate compounds. Expanded sets of effectors produced in large amounts during plant colonization are involved in the establishment of symbiosis or the development of disease, avoiding the triggering of host defences, whilst the fungus is growing within living cells (Martin *et al.* 2008; Zuccaro *et al.* 2011).

Gene family expansion has led not only to improvements in nutrient uptake and host infection in fungal pathogens and symbionts, but also to adaptations for more efficient catabolism and drug resistance. In the human pathogens *Malassezia globosa* and *Candida* yeasts and in the amphibian pathogen *B. dendrobatidis*, expanded families of proteases or hydrolases may play a role in the ability of the pathogens to degrade host tissues (Xu *et al.* 2007; Joneson *et al.* 2011). In white-rot* fungi, the gene families encoding the enzymes involved in wood decay (oxidases, peroxidases and hydrolytic enzymes) have been expanded (Martinez *et al.* 2004). In the human pathogen *Cryptococcus neoformans*, recent amplification of a gene encoding an arsenite efflux transporter has enhanced arsenite resistance, the degree of enhancement being correlated with the copy number of the repeat (Chow *et al.* 2012).

The expansion of gene families has also contributed to genomic innovation in domesticated fungal species, enhancing their enzyme production or particular aspects of metabolism. In the domesticated species *Penicillium rubens* (previously known as *P. chrysogenum*), the penicillin biosynthesis gene cluster shows footprints of tandem duplication events, whereas this cluster is present as a single copy in the wild type (Fierro *et al.* 1995). The genomic regions specific to the domesticated *A. oryzae*, which is used for sake or enzyme production and is derived from a plant pathogen, are enriched in genes involved in the synthesis of secondary metabolites, and specific expansions of genes encoding secreted hydrolytic enzymes, proteins involved in amino acid metabolism and amino acid/sugar uptake transporters have been reported (Machida *et al.* 2005). Expanded families of transporters have also contributed to the evolution of various degrees of sensitivity to drugs or anaerobic fermentation in yeasts (Dunn *et al.* 2005; Marcet-Houben *et al.* 2009; Lin & Li 2010).

A number of gene expansions have been facilitated by mobile genetic elements. TEs can acquire cellular genes or gene fragments between their terminal inverted repeats and replicate them throughout the genome (Manning *et al.* 2013). Such TE-mediated duplications have contributed to the amplification of toxin production and increased the copy number of effectors, and their diversification, in the pathogen *Pyrenophora tritici-repentis* (Manning *et al.* 2013). In natural yeast populations, transposon-mediated genome rearrangements

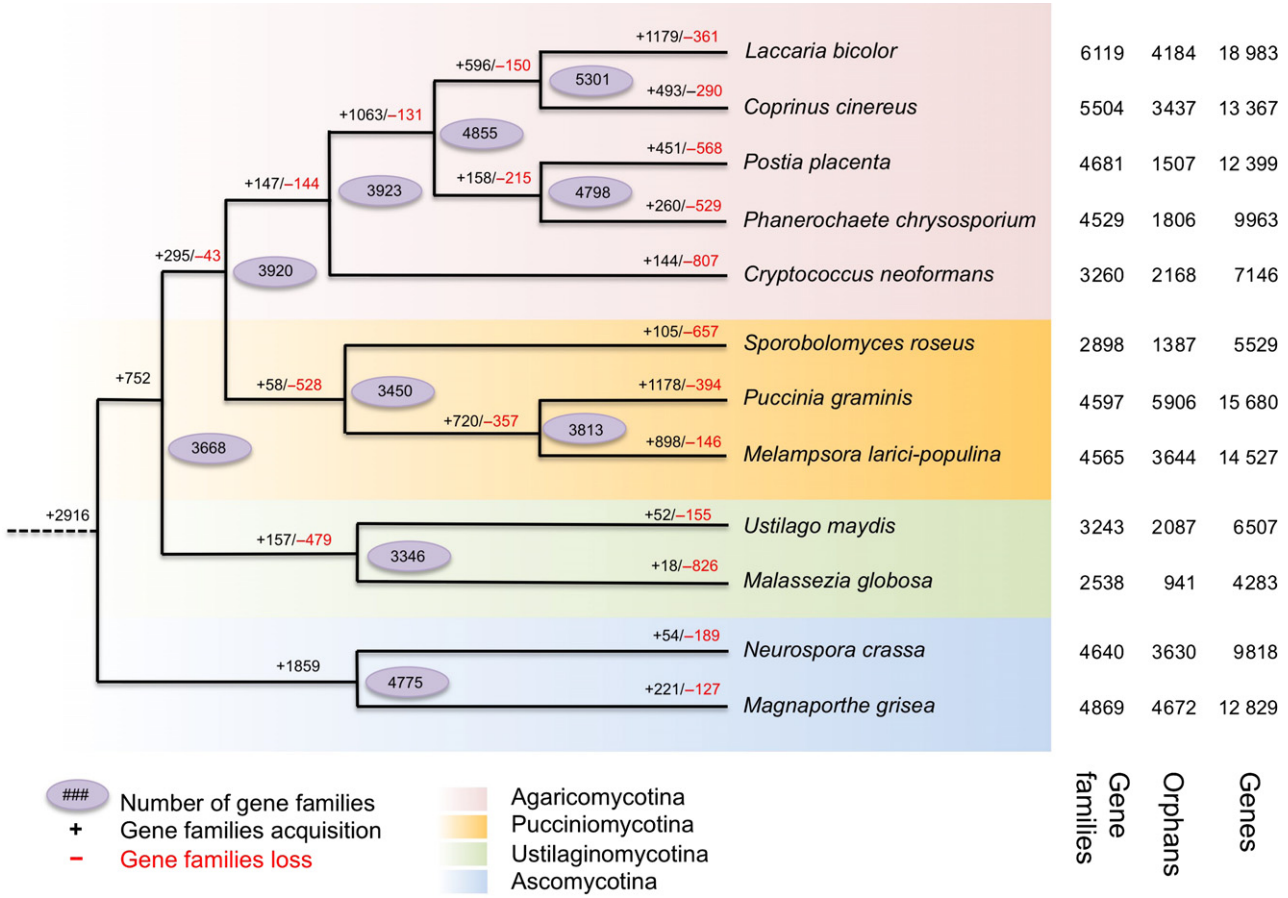


Fig. 2 Predicted pattern of gene family gain and loss in representative fungal genomes. The figure represents the total number of protein families in each species or node estimated by the Dollo parsimony principle. The numbers on the branches of the phylogenetic tree correspond to the numbers of expanded (left, black), contracted (right, red) or inferred ancestral (oval) protein families in each lineage, determined by comparison with the putative pan-proteome. For each species, the number of gene families, orphan genes and the total number of genes are indicated on the right. Reproduced with permission from Duplessis *et al.* (2011).

have resulted in the duplication of genes governing the regulation of the copper pathway, enhancing copper tolerance (Chang *et al.* 2013).

Whole-genome duplication has also created duplication events in yeasts, with about 500 duplicate genes known to have been retained in *Saccharomyces cerevisiae* (Kellis *et al.* 2004). Some of these duplicated genes diversified and became specialized, encoding, in particular, enzymes controlling metabolism in aerobic and hypoxic conditions, enabling *S. cerevisiae* to grow without oxygen, by fermenting glucose anaerobically (Wolfe 2004). In most cases of diverging paralogs, neofunctionalization commonly occurs where one copy evolved slowly, retaining the ancestral function, whereas the other copy evolved much more rapidly, acquiring a novel function (Kellis *et al.* 2004), as predicted before the advent of genome sequencing (Ohno 1970). By contrast, other duplicate genes have remained identical by gene conversion, conferring advantages through greater protein production, as in the case of ribosomal protein genes (Kellis *et al.* 2004).

Gene losses. Gene losses are a common landmark of the switch to an obligate biotrophic or symbiotic lifestyle in fungi, with a decrease in the numbers of plant cell wall degradation proteins, secreted toxins and genes controlling the machinery redundant with that of their hosts (Kamper *et al.* 2006; Martin *et al.* 2008; Spanu *et al.* 2010; Duplessis *et al.* 2011; McDowell 2011; Spanu 2012). The mutualist endophyte *Piriformospora indica*, for instance, lacks the pathways required for the synthesis of toxic secondary metabolites and some essential genes involved in nitrogen assimilation (Zuccaro *et al.* 2011). The human *Malassezia* pathogens have no fatty acid synthase gene, instead producing multiple secreted lipases, which they use to obtain fatty acids from human skin (Xu *et al.* 2007). An extreme case of genome reduction in pathogens is provided by the fungus-related

microsporidia. These obligate intracellular parasites have lost many genes, resulting in a high degree of dependence on the host genome (Peyretaillade *et al.* 2011). Cases of adaptive gene reduction have also occurred during the evolution of ectomycorrhizal biotrophy* and brown rot* saprotrophy from lignin-degrading white-rot* ancestors (Eastwood *et al.* 2011; Floudas *et al.* 2012). The reduction of specific protein families (lignin-degrading peroxidases) in these fungi represents a considerable saving in terms of the production of expensive machinery completely unnecessary for their lifestyles. Similarly, the biocontrol agent *Pseudozyma flocculosa*, which antagonizes powdery mildew, is not a plant pathogen, and it lacks the secreted proteins involved in the pathogenicity of its close relative *U. maydis* (Lefebvre *et al.* 2013).

Gene deletion is an important mechanism of adaptation not only because it saves energy due to the elimination of unnecessary enzymatic machineries, but also because it allows pathogens to avoid detection by host defences. For example, recent sequencing of the genome of the wheat powdery mildew *Blumeria graminis* revealed large deletions in some individuals, with most of the deleted genes identified as candidate effectors (Wicker *et al.* 2013). Another example is provided by the emergence of *Magnaporthe oryzae* on rice, which was probably facilitated by the loss of an effector (Couch *et al.* 2005), mediated by transposable elements.

Regulatory change

Looking at the presence of specific genes or alleles is not sufficient for a full understanding of adaptive divergence, as the specific regulation of genes may also play an important role (Wolbach *et al.* 2009; Spanu & Kämper 2010). Genes may be regulated by *cis*-acting modulators of single genes or *trans*-acting modulators of multiple genes. *Cis*-acting sites (i.e. the DNA-encoded nucleosome organization of promoters) evolve much more rapidly than coding sequences (Borneman *et al.* 2007; Tuch *et al.* 2008). In the budding yeast, *trans*-acting loci affecting the expression of up to 100 genes each have been identified, with related functions displaying coregulation (Brem *et al.* 2002).

A first line of evidence for the importance of gene regulation in adaptation comes from genomic scans looking for signs of selection. Analyses of EST data in closely related plant pathogens have identified several genes involved in *trans* regulation among those undergoing rapid diversification (Aguileta *et al.* 2010, 2012). In *S. cerevisiae*, selective sweeps have been identified on several *cis*-regulatory mutations downregulating the levels of interacting proteins involved in endocytosis in pathogenic strains, thereby increasing virulence (Fraser *et al.* 2009, 2013). More generally, genome-wide scans have revealed widespread positive selection, affecting expression levels in yeasts (Fraser *et al.* 2009; Bullard *et al.* 2010).

A second line of evidence supporting the role of regulatory changes in adaptation has been obtained from a combination of genome sequencing and expression data. These studies have revealed that genes encoding proteins involved in growth and general metabolism have conserved expression patterns, whereas those involved in responses to external and internal signals (e.g. stress response and effectors) or with nonessential functions display divergent patterns of expression between species (Tirosh *et al.* 2006; Thompson & Regev 2009). The proximal mechanism allowing such differences in the rate of expression to evolve is linked to promoter organization. Genes displaying divergent expression harbour promoters that are particularly sensitive to chromatin remodelling, a particular nucleosome organization and the presence of a TATA box (Tirosh *et al.* 2006; Field *et al.* 2009; Thompson & Regev 2009). These observations suggest that there is selection for stable promoters for essential and housekeeping genes, whereas the promoters for genes controlling responses to stresses or encoding effectors are more labile. This situation favours stability and robustness to mutation for essential genes, and the rapid evolution of stress-response and effector genes (Tirosh *et al.* 2006).

Some important changes in gene regulation have been found to be caused by the insertion of transposable elements. TE-mediated rearrangements have, for instance, been implicated in changes to the regulation of genes conferring sulphite resistance in yeast strains used for wine production (Perez-Ortin *et al.* 2002).

The exceptional genetic tractability of the yeast *S. cerevisiae* has made it possible to investigate the role of gene regulation in adaptation by experimental evolution experiments. After evolution over thousands of generations of cultures in various suboptimal media (glucose, sulphate or phosphate limitation) or with novel nutrient sources, adaptation has been shown to involve the massive remodelling of gene expression, demonstrating the importance of genetic regulation for adaptation (Ferea *et al.* 1999; Gresham *et al.* 2008).

Genomic architecture of adaptation

The genomic changes underlying adaptive phenotypes appear to be nonrandomly distributed within the genomes, with some regions more prone to the accumulation of changes. Genome sequencing in fungi has, indeed, revealed many cases of genomic heterogeneity, with gene-dense and gene-sparse repeat-rich regions, the latter often evolving more rapidly and carrying

genes under divergent selection between closely related species. This phenomenon has been referred to as 'two-speed genomes'. Gene-sparse regions may constitute genomic islands enriched in ecologically important genes (e.g. effectors), and displaying higher rates of evolution, in terms of presence/absence, copy number or nonsynonymous substitutions (Hatta *et al.* 2002; Gout *et al.* 2006; Cuomo *et al.* 2007; Fedorova *et al.* 2008; Martin *et al.* 2008; Ma *et al.* 2010; Schirawski *et al.* 2010; Stajich *et al.* 2010; Klosterman *et al.* 2011). The islands with high rates of evolution may be subtelomeric (Chuma *et al.* 2011), AT-rich isochore-like regions (Van de Wouw *et al.* 2010; Rouxel *et al.* 2011), small dispensable chromosomes (Coleman *et al.* 2009; Ma *et al.* 2010; Goodwin *et al.* 2011; Croll & McDonald 2012; Croll *et al.* 2013), dynamic gene clusters (Schirawski *et al.* 2010; Schardl *et al.* 2013) or regions of extensive chromosomal reshuffling (de Jonge *et al.* 2011).

The genes underlying adaptive phenotypes can thus be clustered on supernumerary chromosomes, such as those of the pathogen *Fusarium solani*, carrying genes involved in pathogenicity, antibiotic resistance and the use of unique carbon/nitrogen sources (Coleman *et al.* 2009). Similarly, genomic islands in the human pathogen *Aspergillus fumigatus*, which may be as large as 400 kb, contain mostly repetitive DNA and species-specific genes involved in metabolic processes conferring adaptation to specific ecological niches (Fedorova *et al.* 2008). The genomic islands of *A. fumigatus* are preferentially located in subtelomeric regions prone to frequent nonhomologous recombination between paralogs, duplications, sequence variability (Brown *et al.* 2010) and gene transfers (Kavanaugh *et al.* 2006). There may be selection for the location of effector genes in regions in which frequent transposable element movements and rearrangements induce high rates of mutation, as these genes frequently evolve under positive selection (Gout *et al.* 2006; Klosterman *et al.* 2011; Rouxel *et al.* 2011). Sequencing of the genomes of several strains of the cereal pathogen *Fusarium graminearum* has revealed that the gene-sparse regions with species-specific genes are also the most polymorphic within species (Cuomo *et al.* 2007).

Other types of 'genomic islands' involved in adaptation include those strongly differentiated between closely related species adapted to different ecological niches. Genomic islands displaying strong differentiation have been detected in closely related fungi by genome resequencing, which has made it possible to identify regions involved in adaptation to contrasting environments by a 'reverse-ecology' approach, that is, with no a priori candidate gene or assumed function (Li *et al.* 2008). A striking example is provided by the detection of a region displaying high levels of differentiation between *Neurospora* species, harbouring genes shown to be involved in adaptation to temperature (Ellison *et al.* 2011) (Fig. 3).

Chromosomal rearrangements constitute another kind of genomic architecture that can be involved in adaptation by changing gene regulation or triggering gene duplication for example (Adams *et al.* 1992; Perez-Ortin *et al.* 2002; Gresham *et al.* 2008). Transposable elements have been implicated in many adaptive chromosomal rearrangements (Gioti *et al.* 2012), and some have even been recruited for the promotion of regular mating-type switching (Butler *et al.* 2004; Barsoum *et al.* 2013). Genetic engineering in yeasts has even shown that chromosomal rearrangements can promote enhanced fitness in the absence of other polymorphisms (Colson *et al.* 2004; Teresa Avelar *et al.* 2013). Similarly, experimental evolution and genetic engineering in yeast have revealed that a whole-genome duplication and a frameshift mutation were sufficient to induce a dramatic change in behaviour, generating a fast-sedimenting, multicellular phenotype (Oud *et al.* 2013).

Another feature of the genomic architecture of adaptation in fungi is the clustering of genes involved in the same biosynthesis pathway, such as those encoding proteins involved in the production of secondary metabolites, in particular (Slot & Rokas 2010). The genes within clusters are coregulated, facilitating their coinduction in response to particular conditions, such as plant colonization or stresses, and preventing the accumulation of intermediate toxic compounds (McGary *et al.* 2013). Clustering has been reported, for example, for the genes responsible for mycotoxin production in *Fusarium* (Ma *et al.* 2010), secreted proteins involved in virulence in *Ustilago maydis* (Kamper *et al.* 2006), for the galactose utilization pathway (Slot & Rokas 2010) and for the genes involved in nitrate assimilation (Slot & Hibbett 2007). Conversely, cluster fragmentation may facilitate metabolic retooling and the subsequent host adaptation of plant pathogens (Bradshaw *et al.* 2013).

Genetic bases of adaptation: genomic novelties of trans-specific origin

We have reviewed above the genomic bases of adaptation resulting from genomic changes occurring within species, that is, amino acid substitutions, regulation, gene content and genomic architecture. In this section, we will focus on adaptations extending across species boundaries (i.e. hybridization and HGTs). Recent findings suggest that these sources of adaptive change may have made a substantial contribution to evolution, rather than being merely anecdotal, as initially thought (Roper *et al.* 2011).

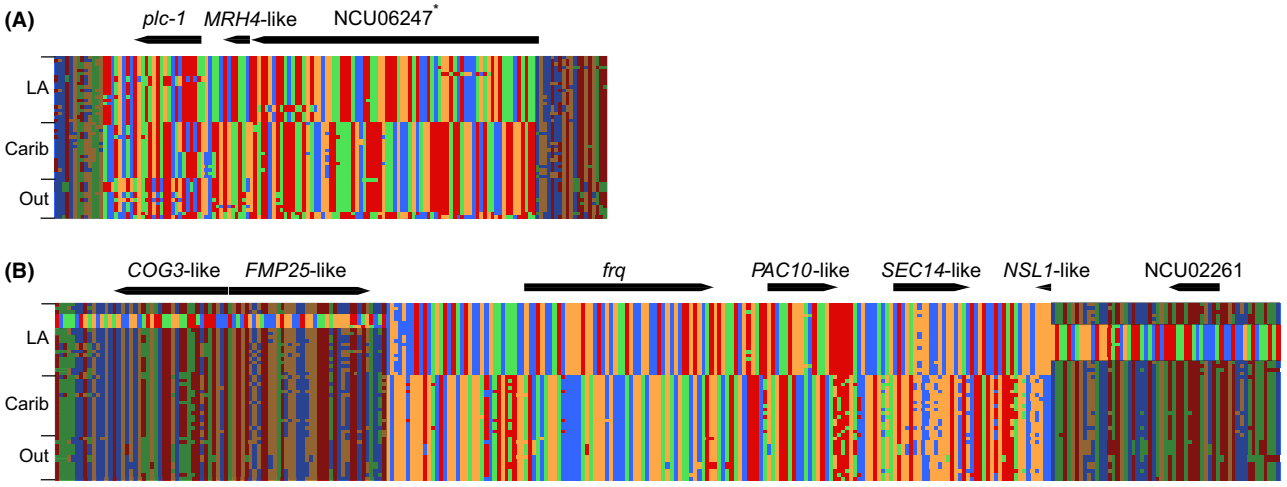


Fig. 3 Genomic islands of divergence in *Neurospora crassa*. Each column is a polymorphic site, and each row contains the genotype for a particular strain. The flanking regions surrounding the divergence outliers are shaded and accentuate the distinct patterns of nucleotide polymorphism within the divergence outlier regions. Strains are grouped by population of origin. LA, Louisiana; Carib, Caribbean (Florida, Haiti and the Yucatan); Out, outgroups from Central America, South America and Africa. The matrix in A is a 10-kb divergence island on chromosome 3. The matrix in B is a 27-kb divergence island on chromosome 7. Reproduced with permission from Ellison *et al.* (2011). MRH4-like RNA helicases are key factors in the microbial cold response. *frq* is a circadian oscillator gene frequency.

Hybridization and introgression

Hybridization is generally deleterious, as it breaks up adaptive combinations of alleles and may bring together incompatible alleles. However, introgression*, the transfer of genetic material between hybridizing taxa through backcrossing, can sometimes be important in adaptation, by incorporating innovation, as reported in several cases in fungi. For example, population genomic analyses of resequencing data have also revealed widespread recent introgressions between the human pathogens *Coccidioides immitis* and *C. posadasii* (Neafsey *et al.* 2010) (Fig. 4). The genomic regions introgressed in *C. immitis* may confer a selective advantage, as they are enriched in coding sequences, accounting for about 8% of the genes in *C. immitis*. In another human pathogen, *Cryptococcus neoformans*, a genomic fragment containing 14 genes from the variety *grubii*, has introgressed in the variety *neoformans* (Kavanaugh *et al.* 2006). The adaptive nature of this introgression is indicated by the spread of the island throughout almost the entire gene pool of the variety *neoformans* and by its duplication.

Widespread recent introgressions have also been identified in several *Saccharomyces* yeasts (Naumova *et al.* 2005; Liti *et al.* 2006). For instance, it has been suggested that a subtelomeric introgression between *S. paradoxus* and *S. cerevisiae* containing 12 genes is adaptive, based on the invasion of the entire European population of *S. paradoxus* by the introgressed fragments (Liti *et al.* 2006). One of the candidate genes for the selective advantage identified encodes resistance to toxin killing (Liti *et al.* 2006). Genome sequencing has also revealed the introgression of three large regions between the distantly related wine yeasts *S. cerevisiae* and *Zygosaccharomyces bailii*. Two of these regions are subtelomeric, and the three together encompass 34 genes involved in key wine fermentation functions (Novo *et al.* 2009). These recent introgressions might be due to the close association of the species concerned with human populations, as humans are known to have mixed previously isolated species and to exert strong selection for novel phenotypes.

Another type of adaptive introgression involves the regeneration, by introgression, of alleles that have accumulated deleterious mutations. In phylogenetic lineages of the filamentous ascomycete *Neurospora tetrasperma* for instance, comparative genomics studies have revealed the presence of a large introgression tract (>4 Mb) in the region of suppressed recombination of the mating-type chromosome (Sun *et al.* 2012). This introgression from freely recombining species of *Neurospora* has been shown to be associated with a lower level of degeneration of the mating-type chromosome and may therefore contribute to the reinvigoration of genomic regions subject to suppressed recombination.

Hybridization can even promote the emergence of new species, by creating transgressive phenotypes* allowing adaptation to new ecological niches (Brasier 2001; Schardl & Craven 2003). Examples of transgressive phenotypes in hybrids can also be found in domes-

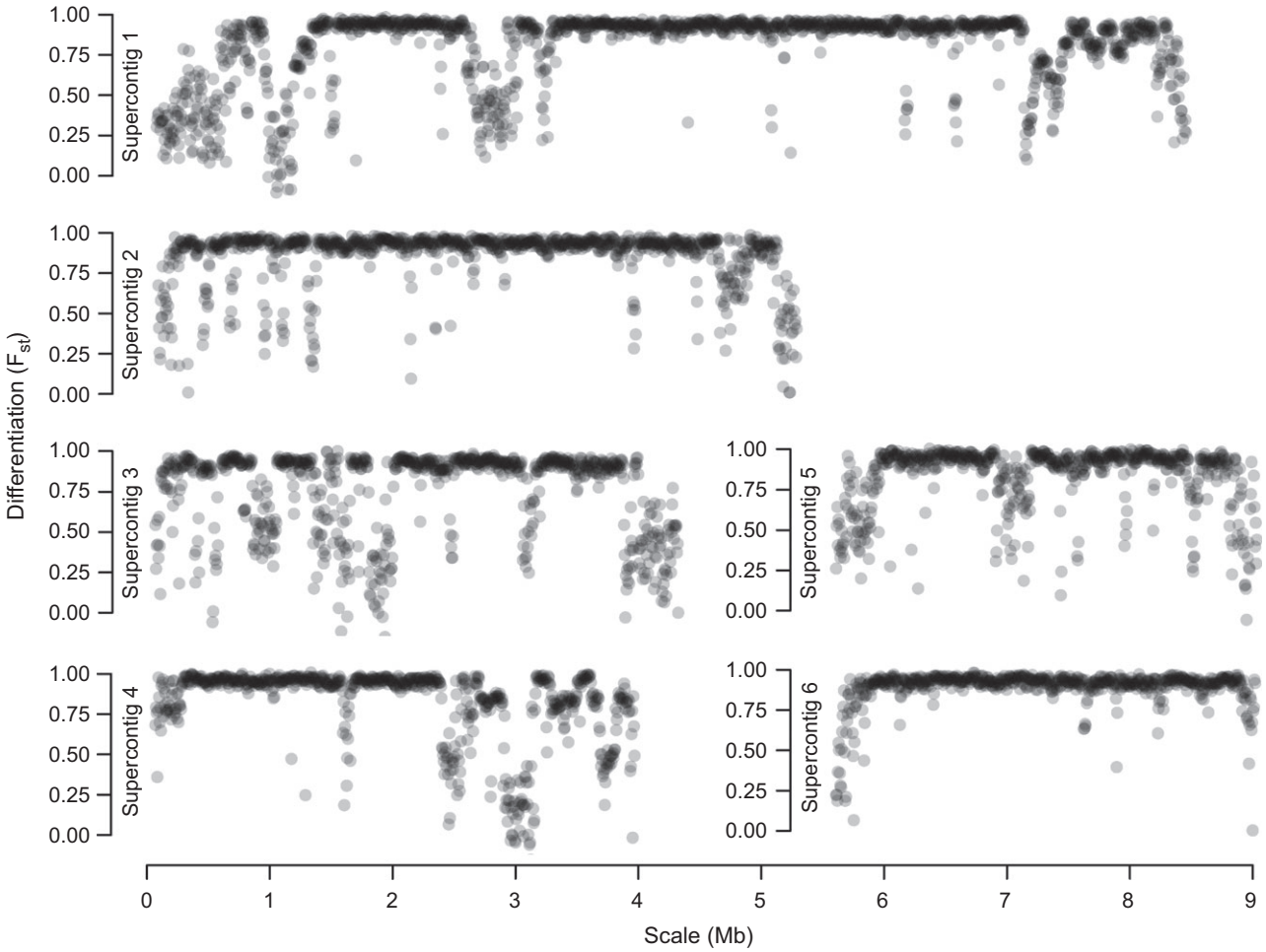


Fig. 4 Illustration of introgression between fungi: the case of the human pathogens *Coccidioides immitis* and *C. posadasii*. Sliding-window analysis of species divergence (F_{ST}). Each circle represents the F_{ST} calculated for a nonoverlapping 5-kb window. Low F_{ST} values indicate introgressed genomic regions. Reproduced with permission from Neafsey *et al*. (2010).

ticated fungi. The lager-brewing yeast, *S. pastorianus*, is an allotetraploid obtained by the fusion of *S. cerevisiae* with the cryotolerant species *S. eubayanus* (Libkind *et al.* 2011). The production of lager requires a particular low-temperature fermentation that can only be achieved efficiently by *S. pastorianus*. Transgressive phenotypes* have also been demonstrated in certain growth conditions for artificial interserotype hybrids of the human pathogen *C. neoformans* (Shahid *et al.* 2008). Together, these cases suggest that transgressive hybridization could promote adaptation at the margin of species distributions.

Horizontal gene transfer

Unlike introgression and hybridization, HGT results in the transmission of genetic material between species in the absence of sexual reproduction. HGT has been long acknowledged as a major driver of prokaryotic evolution and is increasingly recognized as a prominent source of adaption in eukaryotes (Keeling & Palmer 2008; Keeling 2009; Syvanen 2012). Within eukaryotes, fungi are the group for which the largest number of HGT events has been described. This may be explained by the large number of genomes sequenced and the performance of large-scale phylogenomic analyses specifically for the detection of HGT in fungi (Marcet-Houben & Gabaldon 2010). For instance, a systematic survey of 60 fungal genomes from various clades has shown that more than 700 genes of prokaryotic origin were acquired by HGT (Marcet-Houben & Gabaldon 2010). The finding that most of the known HGTs in eukaryotes involve prokaryotic genes probably reflects the abundance of prokaryotes in all environments (Keeling & Palmer 2008) and the relative ease of detecting such HGT events, due both to the wealth of bacterial genomes publicly available and the marked differences between eukaryotic and prokaryotic genomic features.

Eukaryote-to-eukaryote gene transfers may prove to be more frequent when more genome sequences become available. A number of HGT events have been facilitated by mobile genetic elements (Friesen *et al.* 2006). There are now many convincing examples of HGT conferring a selective advantage on the recipient fungal host, as outlined below.

HGT promoting pathogenicity on plants. Fungal pathogens have provided several examples of horizontal transfers, including the acquisition of new virulence factors by HGT. For example, the toxin-encoding gene *ToxA* was transferred in the 1940s from the wheat pathogen *Stagonospora nodorum* to another pathogenic fungus, *Pyrenophora tritici-repentis* (Friesen *et al.* 2006). This HGT, mediated by TEs, allowed *P. tritici-repentis* to acquire virulence on wheat. Another study on the same genus revealed the presence of numerous genes acquired by HGT from bacteria and plants and directly associated with virulence factors, that is, genes encoding proteins that interfere with plant-defence responses or enzymes that can degrade plant cell walls (Sun *et al.* 2013). Another example is provided by the acquisition, by fungal pathogens infecting cereal hosts, of novel virulence factors from bacteria also present in grains (Gardiner *et al.* 2012).

The acquisition of virulence factors by whole-chromosome HGT has also been reported in fungi. In *Fusarium oxysporum*, a whole 'pathogenicity chromosome' bearing a collection of genes encoding small proteins involved in virulence has been horizontally transferred between otherwise genetically isolated strains. The possibility of transferring a whole chromosome between strains was confirmed experimentally by simply incubating strains together (Ma *et al.* 2010). In *Alternaria alternata*, a small extra chromosome is present only in strains pathogenic on tomato. Genetic divergence patterns suggest that this chromosome was also acquired by HGT (Akagi *et al.* 2009). The acquisition of pathogenicity chromosomes by HGT can thus mediate the evolution of pathogenicity in fungi.

HGT of secondary metabolic clusters. In addition to horizontal transfers of single genes or whole chromosomes, several transfers of gene clusters have been described, involving, in particular, clusters controlling the biosynthesis of secondary metabolites. For instance, the ascomycetous mould *Trichoderma reesi* has acquired a cluster of nitrate assimilation genes from a distantly related fungus from the phylum Basidiomycota (Slot & Hibbett 2007). It has been suggested that this transfer facilitated a niche shift in fungi and may have been a key acquisition for the colonization of dry land by Dikarya, the group formed by ascomycete and basidiomycete fungi. Another example is provided by the intact transfer of a 23-gene cluster involved in the biosynthesis of a highly toxic secondary metabolite pathway between ascomycetes from different taxonomic classes: from a Eurotiomycete (*Aspergillus* species) to a Sordariomycete (*Podospora anserina*) (Slot & Rokas 2011). This HGT increased the toxicity of the receptor species and highlights the importance of HGT for metabolic diversity of fungi.

It has even been suggested that the organization of secondary metabolite genes into clusters in fungi is a consequence of their acquisition by HGT (Walton 2000; Slot & Rokas 2010), by analogy to the 'selfish operon' theory in prokaryotes (Lawrence & Roth 1996). If transfers are frequent, then the organization of genes into clusters is selectively advantageous for the cluster itself, because it increases its dispersal and its probability of survival. The presence of multiple virulence-related genes on pathogenicity chromosomes that are transferred horizontally may result from the same type of selection.

HGT and adaptation to extreme environments. Some HGTs in fungi have been shown to be involved in the adaptation of species to extreme environments. This is the case for the wine yeast *S. cerevisiae* EC1118, which can tolerate harsh wine fermentation conditions characterized by high sugar and alcohol content, hypoxia, low nitrogen, vitamin and lipid contents (Novo *et al.* 2009). Three genomic regions of foreign origin have been detected in this wine yeast strain; one of these regions was probably acquired by HGT from *Zygosaccharomyces bailii*, a yeast contaminant of wine that is known to be able to survive the entire fermentation process. Several prokaryotic glycosyl hydrolases have also been acquired via HGT by fungi inhabiting the rumen of herbivorous mammals (Garcia-Vallvé *et al.* 2000), allowing the colonization of this new environment in which cellulose and hemicellulose are the main carbon sources.

A striking example of HGT associated with domestication is provided by cheese-associated *Penicillium* species (Cheeseman *et al.* 2014), in which multiple, independent, recent transfers of a long genomic tract of about 500 kb have been detected. The screening of hundreds of different strains and species from diverse environments revealed the presence of the transferred regions only in *Penicillium* strains associated with dairy environments. This transfer might confer a competitive advantage in the complex and highly nutritive cheese environment, as it contains the gene encoding the *Penicillium* antifungal protein (Marx *et al.* 2008), which has been shown to have cytotoxic activity against various filamentous fungi in vitro (Kaiserer *et al.* 2003).

HGT thus seems to be a significant source of evolutionary novelty in fungi and is often associated with

major adaptive transitions, such as the acquisition of pathogenicity, the infection of new hosts and, more generally, adaptation to new environmental conditions. The publication of larger numbers of genome sequences may reveal that HGT is of general importance in eukaryotes.

Conversion of divergent adaptation into reproductive isolation

The nature of the mechanisms responsible for the transition between adaptive polymorphism and speciation remains one of the central questions in evolutionary biology. The mechanisms of reproductive isolation in fungi have been reviewed elsewhere (Kohn 2005; Giraud *et al.* 2008), including at the genomic level (Stukenbrock 2013; Leducq 2014). We focus here exclusively on the two principal phenomena converting adaptation directly into speciation: pleiotropy between adaptation and reproductive isolation, and Bateson–Dobzhansky–Muller incompatibilities (BDM), in which rapidly evolving genes cause deleterious interactions in hybrids.

Pleiotropy and linkage

When adaptive divergence occurs in the face of gene flow due to divergent selection, recombination impedes the construction of adaptive allelic combinations (Rice 1987), unless specific mechanisms keep the alleles together. These mechanisms include recombination suppression (Kirkpatrick & Barton 2006) and the evolution of assortative mating, with the most favourable situation for adaptive divergence being a locus pleiotropically inducing adaptation to a specific niche and assortative mating (Maynard Smith 1966). Such traits controlling both local adaptation and assortative mating have long been considered as exceptions, but might be more widespread than previously thought, as several cases have recently been reported (Servedio *et al.* 2011). A particular case of interest is host specificity in pathogenic ascomycetes: as these fungi mate within their host after growing and obtaining resources, mating can only occur between individuals adapted to the same host, so host specificity also controls assortative mating (Giraud *et al.* 2010). Ecological divergence in pathogenic ascomycetes is further facilitated by the small number of genes involved in host specificity, with a single toxin gene or virulence allele being able to confer the ability to colonize the host. This, combined with the billions of spores produced, provides opportunities for mutation and a high selective load (Giraud *et al.* 2010). One well-studied example is *Venturia*, the fungus responsible for apple scab, in which population divergence between hosts and within apple orchards in sympatry has been associated with a single virulence locus, causing instant adaptive speciation (Giraud *et al.* 2010; Gladieux *et al.* 2011). As host specificity can pleiotropically cause reproductive isolation, interfertility can be retained long after speciation, without the development of incompatibility (Giraud *et al.* 2008; Le Gac & Giraud 2008). A genome scan comparing closely related *Venturia* pathogens revealed no genes under selection other than the gene involved in the ability to infect the host plant and no incompatibility footprints, with the entire genetic structure explicable by the presence or absence of virulence alleles (Leroy *et al.* 2013). Such traits controlling host specificity and assortative mating pleiotropically are therefore not only conducive to adaptive divergence but also leave specific genomic footprints.

Genetic incompatibilities

Reproductive isolation can arise directly from genes responsible for adaptation to divergent ecological niches, but it can also evolve as a by-product of rapidly evolving genes in closely related species that will become incompatible when brought together in hybrids. Epistatic interactions between alleles at different loci that are deleterious in hybrids, causing intrinsic nonviability or sterility, constitute a form of BDM incompatibility often observed in hybridizing species (Orr & Turelli 2001). Alternatively, hybrid nonviability may result from poor suitability for growth in parental environments, also referred to as 'extrinsic isolation' (Egan & Funk 2009).

According to the BDM genetic incompatibility model, epistatic interactions at two or more loci in hybrids tend to be negative, causing sterility or nonviability (Coyne & Orr 2004). This model is widely accepted and has been supported both theoretically and experimentally (Presgraves 2010). However, only a few genes involved in such incompatibilities have been identified in a handful of model species (Presgraves 2010). Several of these 'speciation genes' were identified in fungi and, strikingly, many involve mitochondrial–nuclear incompatibilities and display rapid evolution (Lee *et al.* 2008; Anderson *et al.* 2010; Chou *et al.* 2010). One elegant study used chromosome replacement lines in two yeast species and identified a genetic incompatibility between an *S. bayanus* nuclear gene and *S. cerevisiae* mitochondria, resulting in F2 hybrid sterility (Lee *et al.* 2008). The rapid evolution of a mitochondrial gene impedes the interactions of its RNA with a protein encoded by the nucleus, preventing its translation. Another systematic screening of incompatibilities between yeast species also identified mitochondrial–nuclear incompatibilities (Chou *et al.* 2010). It has been suggested that some

incompatibilities in yeasts are more complex than mere interactions between two loci (Kao & Sherlock 2008).

BDM genetic incompatibilities are often considered in opposition to ecological speciation, but these two processes may actually be associated. The most detailed examples of this association have been provided by laboratory studies with saprotrophs. In a now classic experiment, the initial stages of speciation were provoked by growing replicate *Saccharomyces cerevisiae* populations under different suboptimal growth conditions, on high-salinity or low-glucose minimal medium (Dettman *et al.* 2007). After 500 generations, postzygotic isolation had evolved, in the form of a lower fitness of hybrids due to antagonistic epistasis, as a direct consequence of divergent adaptation. Whole-genome sequencing revealed that postzygotic isolation was caused by a BDM incompatibility between the evolved alleles of genes conferring higher fitness in high-salt and low-glucose media, respectively (Anderson *et al.* 2010). This work neatly demonstrates how divergent selection can drive the evolution of intrinsic genetic incompatibilities between populations and shows that adaptive divergence is not limited to the evolution of inherently ecological forms of reproductive isolation, such as ecological selection against migrants or hybrids (Nosil 2012).

Divergent adaptation was also found to promote reproductive isolation directly in an experiment in *Neurospora* (Dettman *et al.* 2008), indicating that this process may be widespread among fungi and illustrating the convenience of the experimental models available in this kingdom for the testing of evolutionary hypotheses. In this study, two incompatibility loci (*dfe* and *dma*) consistent with the BDM model were identified and shown to cause severe sexual reproduction defects in the hybrids of evolved populations, and in naturally occurring hybrids of *N. crassa* and *N. intermedia*.

BDM incompatibilities are powerful mechanisms known to be involved in isolation, but they may not be as widespread as generally thought. Some genome-wide analyses in fungi have failed to find BDM incompatibilities despite the screening of almost the entire genome by chromosome replacement experiments (Greig 2007; Kao & Sherlock 2008; Lee *et al.* 2008). Furthermore, studies investigating the shape of the decrease in fungal hybrid fitness as a function of genetic distance between species found no evidence of the snowball pattern expected under the hypothesis of a predominance of BDM genetic incompatibilities (Gourbiere & Mallet 2010; Giraud & Gourbiere 2012). The linear or slowing decrease in fitness appeared more consistent with reinforcement, sterility due to karyotypic variation or a lack of ecological suitability of hybrids (Gourbiere & Mallet 2010; Giraud & Gourbiere 2012).

Conclusion and future prospects

We have reviewed the numerous insights into the genomic processes of adaptive divergence in eukaryotes provided by studies on fungi. Many predictions have been validated, such as the role of gene duplication in novel functions, the positive impact of TEs on evolvability, the importance of changes in gene regulation, the clustering of some adaptive changes in particular genomic regions and the occurrence of BDM incompatibilities. Other processes, previously considered anecdotal, have been shown to feature prominently among the drivers of adaptive divergence. These processes include gene deletions, introgressions, changes in genomic architecture and HGTs. The strength and nature of ecologically based divergent selection or life cycle characteristics have also been shown to be important. The genomic heterogeneity in rates of evolution in some fungi, with regions differing in their susceptibility to mutations, may facilitate the resolution of an apparent 'conflict of interest' between different classes of genes. Isochore-like structures, for instance, make it possible to cope with trade-offs in which there is a need to maintain some functions under strong constraints, with others evolving rapidly in response to positive selection. This trade-off may also be resolved by gene regulation, with promoters differing in evolvability according to the type of function.

Increasingly affordable sequencing will continue to contribute to genomic studies of fungal adaptation and speciation, making it possible to address a new range of questions. Future challenges include understanding how the selective pressures on phenotypes are reflected in genome evolution and the extent to which coding and noncoding sequence differences between species are adaptive. Large-scale genome resequencing projects will also improve our understanding of local adaptation and diversification within species. Population genomics studies will make it possible to identify the genomic regions that have experienced selective sweeps without the need for a priori gene candidates, thereby revealing the relative importance in adaptation of regulatory changes versus changes in coding sequences, and, more generally, the genomic features underlying adaptive phenotypes. Dramatic improvements in the functional annotation of fungal genomes should facilitate real progress in the use of such reverse-ecology approaches, which are just beginning to be applied to fungi. Increasing the number of sequenced species will make it possible to assess the actual extent of eukaryote-to-eukaryote HGT. We are also just beginning to unravel the genomic architecture of adaptive divergence, and many processes occurring at the genome scale are still probably not understood or not even predicted.

Studies of the genomic processes allowing adaptive divergence in the face of gene flow are also required, for which fungal models will be ideal. In particular, domesticated fungi, recently emerged pathogens and experimentally evolved populations will make it possible to investigate the earliest stages of divergence in the face of gene flow, before they become confounded with other species differences. Experiments associated with genome sequencing will be required for investigations of the relative importance of BDM incompatibilities in speciation, the traits inducing pleiotropically adaptation and reproductive isolation, the intrinsic ecological barriers to gene flow (selection against immigrants and hybrids) and genomic rearrangements. In particular, the role of rearrangements as a protection against recombination for accumulating a suite of coadapted alleles allowing local adaptation in the face of gene flow is still debated (Navarro & Barton 2003), and this aspect could be investigated with fungal genome sequences. Improvements in assembly technology will be required for this, and optical mapping appears to be particularly promising in this respect (Hood *et al.* 2013).

New methods for analysing genomes are continually becoming available, making it possible to test new hypotheses. For instance, recently developed methods have made it possible to investigate the co-evolution of protein-encoding genes within genomes through the use of machine learning processes (de Vienne & Aze 2012) or to generate testable hypotheses from phylogenetic profiles for lineage-specific functional modules in the form of information about the correlated gain and loss of protein families (Pellegrini 2012). These approaches are based on a rationale of identifying gene families with correlated or coupled evolution. Advances in theoretical population genomics have also increased power for the inference of demographic parameters (Gutenkunst *et al.* 2009; Excoffier *et al.* 2013), determinations of the relative importance of natural selection and random genetic drift for shaping molecular evolution (Messer & Petrov 2013), and the identification of genes under natural selection (Eilertson *et al.* 2012). More generally, systems biology is another fascinating perspective for studying fungal genomes. Systems biology studies on genomes have mostly focused on deciphering molecular networks and biochemical pathways so far (Costanzo *et al.* 2010; Altelaar *et al.* 2013; Mitra *et al.* 2013); however, this approach could be useful for understanding genome evolution and adaptation and even for predicting evolutionary trajectories (Papp *et al.* 2011).

Acknowledgements

This work was supported by the ANR Grant 12-ADAP-0009-02-GANDALF and the ERC starting grant, GenomeFun 309403 to TG, the ANR FROMA-GEN Grant ANR-12-PDOC-0030 to AB, the National Science Foundation Grant DBI 1046115 to SB and Marie Curie postdoctoral fellowships to PG [FP7-PEOPLE-2010-IOF-No.273086], GA [FP7-PEOPLE-2010-IEF-No.274223] and RRDLV [FP7 COFUND PRES-SUD No. 246556]. We thank François Delmotte, Tobin Peever, the Mycology Laboratory BIOTEC and Jacques Guinberteau (INRA) for the pictures in Fig. 1.

References

Adams J, Puskas-Rozsa S, Simlar J, Wilke C (1992) Adaptation and major chromosomal changes in populations of *Saccharomyces cerevisiae*. *Current Genetics*, **22**, 13–19.

Aguileta G, Refrégier G, Yockteng Y, Fournier E, Giraud T (2009) Rapidly evolving genes in pathogens: methods for detecting positive selection and examples among fungi, bacteria, viruses and protozoa. *Infection, Genetics and Evolution*, **9**, 656–670.

Aguileta G, Lengelle J, Marthey S *et al.* (2010) Finding candidate genes under positive Selection in non-model species: examples of genes involved in host specialization in pathogens. *Molecular Ecology*, **19**, 292–306.

Aguileta G, Lengelle J, Chiapello H *et al.* (2012) Genes under positive selection in a model plant pathogenic fungus, *Botrytis*. *Infection Genetics and Evolution*, **12**, 987–996.

Ajouz S, Nicot PC, Bardin M (2010) Adaptation to pyrrolnitrin in *Botrytis cinerea* and cost of resistance. *Plant Pathology*, **59**, 556–566.

Akagi Y, Akamatsu H, Otani H, Kodama M (2009) Horizontal chromosome transfer, a mechanism for the evolution and differentiation of a plant-pathogenic fungus. *Eukaryotic Cell*, **8**, 1732–1738.

Alfoeldi J, Lindblad-Toh K (2013) Comparative genomics as a tool to understand evolution and disease. *Genome Research*, **23**, 1063–1068.

Altelaar AFM, Munoz J, Heck AJR (2013) Next-generation proteomics: towards an integrative view of proteome dynamics. *Nature Reviews Genetics*, **14**, 35–48.

Ames R, Money D, Ghatge V, Whelan S, Lovell S (2012) Determining the evolutionary history of gene families. *Bioinformatics*, **28**, 48–55.

Anderson PK, Cunningham AA, Patel NG, Morales FJ, Epstein PR, Daszak P (2004) Emerging infectious diseases of plants: pathogen pollution, climate change and agrotechnology drivers. *Trends in Ecology & Evolution* **19**, 535–544.

Anderson JB, Funt J, Thompson DA *et al.* (2010) Determinants of divergent adaptation and Dobzhansky-Muller interaction in experimental yeast populations. *Current Biology*, **20**, 1383–1388.

Barrick JE, Lenski RE (2013) Genome dynamics during experimental evolution. *Nature Reviews Genetics*, **14**, 827–839.

Barsoum E, Martinez P, Astrom SU (2013) Alpha 3, a transposable element that promotes host sexual reproduction. *Genes & Development*, **24**, 33–44.

Bergot M, Cloppet E, Perarnaud V, Deque M, Marcais B, Desprez-Loustau ML (2004) Simulation of potential range expansion of oak disease caused by *Phytophthora cinnamomi* under climate change. *Global Change Biology*, **10**, 1539–1552.

Bezerra A, Simoes J, Lee WP *et al.* (2013) Reversion of a fungal genetic code alteration links proteome instability with genomic and phenotypic diversification. *Proceedings of the National Academy of Sciences USA*, **110**, 11079–11084.

Boddy L (2000) Interspecific combative interactions between wood decaying basidiomycetes. *FEMS Microbiology & Ecology*, **31**, 185–194.

Bono L, Gensel C, Pfenning D, Burch C (2013) Competition and the origins of novelty: experimental evolution of niche-width expansion in a virus. *Biology Letters*, **9**, 1–5.

Borneman AR, Gianoulis TA, Zhang ZDD *et al.* (2007) Divergence of transcription factor binding sites across related yeast species. *Science*, **317**, 815–819.

Bradshaw RE, Slot JC, Moore GG *et al.* (2013) Fragmentation of an aflatoxin-like gene cluster in a forest pathogen. *New Phytologist*, **198**, 525–535.

Branco S (2010) Serpentine soils promote ectomycorrhizal fungal diversity. *Molecular Ecology*, **19**, 5566–5576.

Brasier CM (2001) Rapid evolution of introduced plant pathogens via interspecific hybridization. *BioScience*, **51**, 123–133.

Brem RB, Yvert G, Clinton R, Kruglyak L (2002) Genetic dissection of transcriptional regulation in budding yeast. *Science*, **296**, 752–755.

Brown JK (1994) Chance and selection in the evolution of barley mildew. *Trends in Microbiology*, **2**, 470–475.

Brown CA, Murray AW, Verstrepen KJ (2010) Rapid expansion and functional divergence of subtelomeric gene families in yeasts. *Current Biology*, **20**, 895–903.

Bruns T, Bidartondo M, Taylor D (2002) Host specificity in ectomycorrhizal communities: what do the exceptions tell us? *Integrative and Comparative Biology*, **42**, 352–359.

Bullard JH, Mostovoy Y, Dudoit S, Brem RB (2010) Polygenic and directional regulatory evolution across pathways in *Saccharomyces*. *Proceedings of the National Academy of Sciences*, **107**, 5058–5063.

Butler G, Kenny C, Fagan A, Kurischko C, Gaillardin C, Wolfe KH (2004) Evolution of the MAT locus and its Ho endonuclease in yeast species. *Proceedings of the National Academy of Sciences of the United States of America*, **101**, 1632–1637.

Butler G, Rasmussen MD, Lin MF *et al.* (2009) Evolution of pathogenicity and sexual reproduction in eight *Candida* genomes. *Nature*, **459**, 657–662.

Carbone I, Jakobek JL, Ramirez-Prado JH, Horn BW (2007) Recombination, balancing selection and adaptive evolution in the aflatoxin gene cluster of *Aspergillus parasiticus*. *Molecular Ecology*, **16**, 4401–4417.

Casadevall A (2008) Evolution of intracellular pathogens. *Annual Review of Microbiology*, **62**, 19–33.

Casadevall A, Steenbergen JN, Nosanchuk JD (2003) 'Ready made' virulence and 'dual use' virulence factors in pathogenic environmental fungi – the *Cryptococcus neoformans* paradigm. *Current Opinion in Microbiology*, **6**, 332–337.

Chang S-L, Lai H-Y, Tung S-Y, Leu J-Y (2013) Dynamic large-scale chromosomal rearrangements fuel rapid adaptation in yeast populations. *Plos Genetics*, **9**, e1003232.

Cheeseman K, Ropars J, Renault P *et al.* (2014) Multiple recent horizontal transfers of a large genomic region in cheesemaking fungi *Nature Communications*. in press.

Chou J-Y, Hung Y-S, Lin K-H, Lee H-Y, Leu J-Y (2010) Multiple molecular mechanisms cause reproductive isolation between three yeast species. *PLoS Biology*, **8**, e1000432.

Chow EWL, Morrow CA, Djordjevic JT, Wood IA, Fraser JA (2012) Microevolution of *Cryptococcus neoformans* driven by massive tandem gene amplification. *Molecular Biology and Evolution*, **29**, 1987–2000.

Chuma I, Isobe C, Hotta Y *et al.* (2011) Multiple translocation of the AVR-Pita effector gene among chromosomes of the rice blast fungus *Magnaporthe oryzae* and related species. *PLos Pathogens*, **7**, e1002147.

Coleman JJ, Rounsley SD, Rodriguez-Carres M *et al.* (2009) The genome of *Nectria haematococca*: contribution of supernumerary chromosomes to gene expansion. *Plos Genetics*, **5**, e1000618.

Colot V, Haedens V, Rossignol JL (1998) Extensive, nonrandom diversity of excision footprints generated by Ds-like transposon Ascot-1 suggests new parallels with V(D)J recombination. *Molecular and Cellular Biology*, **18**, 4337–4346.

Colot HV, Park G, Turner GE *et al.* (2006) A high-throughput gene knockout procedure for *Neurospora* reveals functions for multiple transcription factors. *Proceedings of the National Academy of Sciences of the United States of America*, **103**, 10352–10357.

Colson I, Delneri D, Oliver SG (2004) Effects of reciprocal chromosomal translocations on the fitness of *Saccharomyces cerevisiae*. *Embo Reports*, **5**, 392–398.

Cordier T, Robin C, Capdevielle X, Fabreguettes O, Desprez-Loustau ML, Vacher C (2012) The composition of phyllosphere fungal assemblages of European beech *(Fagus sylvatica)* varies significantly along an elevation gradient. *New Phytologist*, **196**, 510–519.

Costanzo M, Baryshnikova A, Bellay J *et al.* (2010) The genetic landscape of a cell. *Science*, **327**, 425–431.

Couch BC, Fudal I, Lebrun M-H *et al.* (2005) Origins of host-specific populations of the blast pathogen, *Magnaporthe oryzae*, in crop domestication with subsequent expansion of pandemic clones on rice and weeds of rice. *Genetics*, **70**, 613–630.

Coyne JA, Orr HA (2004) *Speciation*. Sinauer Associates, Sunderland, Massachusetts.

Croll D, McDonald BA (2012) The accessory genome as a cradle for adaptive evolution in pathogens. *PLos Pathogens*, **8**, e1002608.

Croll D, Zala M, McDonald BA (2013) Breakage-fusion-bridge cycles and large insertions contribute to the rapid evolution of accessory chromosomes in a fungal pathogen. *PLoS Genetics*, **9**, e1003567.

Crute IR, Pink DAC (1996) Genetics and utilization of pathogen resistance in plants. *Plant Cell*, **8**, 1747–1755.

Cuomo CA, Gueldener U, Xu JR *et al.* (2007) The *Fusarium graminearum* genome reveals a link between localized polymorphism and pathogen specialization. *Science*, **317**, 1400–1402.

Daboussi MJ, Capy P (2003) Transposable elements in filamentous fungi. *Annual Review of Microbiology*, **57**, 275–299.

Dettman JR, Sirjusingh C, Kohn LM, Anderson JB (2007) Incipient speciation by divergent adaptation and antagonistic epistasis in yeast. *Nature*, **447**, 585–588.

Dettman J, Anderson J, Kohn L (2008) Divergent adaptation promotes reproductive isolation among experimental populations of the filamentous fungus *Neurospora*. *BMC Evolutionary Biology*, **8**, 35.

Dodds PN, Rathjen JP (2010) Plant immunity: towards an integrated view of plant–pathogen interactions. *Nature Reviews Genetics*, **11**, 539.

Dunn B, Levine RP, Sherlock G (2005) Microarray karyotyping of commercial wine yeast strains reveals shared, as well as unique, genomic signatures. *BMC Genomics*, **6**, 53.

Duplessis S, Cuomo CA, Lin YC *et al.* (2011) Obligate biotrophy features unraveled by the genomic analysis of rust fungi. *Proceedings of the National Academy of Sciences of the United States of America*, **108**, 9166–9171.

Eastwood DC, Floudas D, Binder M *et al.* (2011) The plant cell wall decomposing machinery underlies the functional diversity of forest fungi. *Science*, **333**, 762–765.

Egan SP, Funk DJ (2009) Ecologically dependent postmating isolation between sympatric host forms of *Neochlamisus bebbianae* leaf beetles. *Proceedings of the National Academy of Sciences of the USA*, **106**, 19426–19431.

Eilertson KE, Booth JG, Bustamante CD (2012) SnIPRE: selection inference using a Poisson random effects model. *PLoS Computational Biology*, **8**, e1002806.

Ellison CE, Hall C, Kowbel D *et al.* (2011) Population genomics and local adaptation in wild isolates of a model microbial eukaryote. *Proceedings of the National Academy of Sciences of the USA*, **108**, 2831–2836.

Enjalbert J, Duan X, Leconte M, Hovmoller M, De Vallavieille-Pope C (2005) Genetic evidence of local adaptation of wheat yellow rust *(Puccinia striiformis f. sp tritici)* within France. *Molecular Ecology*, **14**, 2065–2073.

Excoffier L, Dupanloup I, Huerta-Sanchez E, Sousa VC, Foll M (2013) Robust demographic inference from genomic and SNP data. *PLoS Genetics*, **9**, e1003905.

Fabre B, Piou D, Desprez-Loustau ML, Marcais B (2011) Can the emergence of pine *Diplodia* shoot blight in France be explained by changes in pathogen pressure linked to climate change? *Global Change Biology*, **17**, 3218–3227.

Farman ML (2007) Telomeres in the rice blast fungus *Magnaporthe oryzae*: the world of the end as we know it. *FEMS Microbiology Letters*, **273**, 125–132.

Farrer RA, Henk DA, Garner TWJ, Balloux F, Woodhams DC, Fisher MC (2013) Chromosomal copy number variation, selection and uneven rates of recombination reveal cryptic genome diversity linked to pathogenicity. *PLoS Genetics*, **9**, e1003703.

Fay JC, Benavides JA (2005) Evidence for domesticated and wild populations of *Saccharomyces cerevisiae*. *PLoS Genetics*, **1**, e5.

Feau N, Lauron-Moreau A, Piou D, Marcais B, Dutech C, Desprez-Loustau ML (2012) Niche partitioning of the genetic lineages of the oak powdery mildew complex. *Fungal Ecology*, **5**, 154–162.

Fedorova ND, Khaldi N, Joardar VS *et al.* (2008) Genomic islands in the pathogenic filamentous fungus *Aspergillus fumigatus*. *Plos Genetics*, **4**, e1000046.

Ferea TL, Botstein D, Brown PO, Rosenzweig RF (1999) Systematic changes in gene expression patterns following adaptive evolution in yeast. *Proceedings of the National Academy of Sciences of the United States of America*, **96**, 9721–9726.

Field Y, Fondufe-Mittendorf Y, Moore IK *et al.* (2009) Gene expression divergence in yeast is coupled to evolution of DNA-encoded nucleosome organization. *Nature Genetics*, **41**, 438–445.

Fierro F, Barredo JL, Daez B, Gutierrez S, Fernandez FJ, Martan JF (1995) The penicillin gene cluster is amplified in tandem repeats linked by conserved hexanucleotide sequences. *Proceedings of the National Academy of Sciences of the United States of America*, **92**, 6200–6204.

Fisher MC, Koenig GL, White TJ, Taylor JW (2000) Pathogenic clones versus environmentally driven population increase: analysis of an epidemic of the human fungal pathogen *Coccidioides immitis*. *Journal of Clinical Microbiology*, **38**, 807–813.

Fisher MC, Garner TWJ, Walker SF (2009) Global emergence of *Batrachochytrium dendrobatidis* and amphibian chytridiomycosis in space, time, and host. *Annual Review of Ecology and Systematics*, **63**, 291–310.

Fisher MC, Henk DA, Briggs CJ *et al.* (2012) Emerging fungal threats to animal, plant and ecosystem health. *Nature*, **484**, 186–194.

Floudas D, Binder M, Riley R *et al.* (2012) The paleozoic origin of enzymatic lignin decomposition reconstructed from 31 fungal genomes. *Science*, **336**, 1715–1719.

Frankland J (1998) Fungal sucession – unravelling the unpredictable. *Mycological Research*, **102**, 1–15.

Fraser HB, Moses AM, Schadt EE (2009) Evidence for widespread adaptive evolution of gene expression in budding yeast. *Proceedings of the National Academy of Sciences of the United States of America*, **107**, 2977–2982.

Fraser HB, Levy S, Chavan A *et al.* (2013) Polygenic cis-regulatory adaptation in the evolution of yeast pathogenicity. *Genome Research*, **22**, 1930–1939.

Frenkel O, Peever TL, Chilvers MI *et al.* (2010) Ecological genetic divergence of the fungal pathogen *Didymella rabiei* on sympatric wild and domesticated *Cicer* spp. (Chickpea). *Applied and Environmental Microbiology*, **76**, 30–39.

Friesen TL, Stukenbrock EH, Liu Z *et al.* (2006) Emergence of a new disease as a result of interspecific virulence gene transfer. *Nature Genetics*, **38**, 953–956.

Fu W, Akey JM (2013) Selection and adaptation in the human genome. *Annual Review of Genomics and Human Genetics*, **14**, 467–489.

Garcia-Vallvé S, Romeu A, Palau J (2000) Horizontal gene transfer of glycosyl hydrolases of the rumen fungi. *Molecular Biology and Evolution*, **17**, 352–361.

Gardiner D, McDonald M, Covarelli L *et al.* (2012) Comparative pathogenomics reveals horizontally acquired novel virulence genes in fungi infecting cereal hosts. *Plos Pathogens*, **8**, e1002952.

Giaever G, Chu AM, Ni L *et al.* (2002) Functional profiling of the *Saccharomyces cerevisiae* genome. *Nature*, **418**, 387–391.

Gibbons JG, Salichos L, Slot JC *et al.* (2012) The evolutionary imprint of domestication on genome variation and function of the filamentous fungus *Aspergillus oryzae*. *Current Biology*, **22**, 1403–1409.

Gioti A, Mushegian AA, Strandberg R, Stajich JE, Johannesson H (2012) Unidirectional evolutionary transitions in fungal mating systems and the role of transposable elements. *Molecular Biology and Evolution*, **29**, 3215–3226.

Giraud T, Gourbiere S (2012) The tempo and modes of evolution of reproductive isolation in fungi. *Heredity*, **109**, 204–214.

Giraud T, Levis C, Fortini D, Leroux P, Brygoo Y (1997) RFLP markers show genetic recombination in *Botrytis cinerea* and transposable elements reveal two sympatric species. *Molecular Biology and Evolution*, **14**, 1177–1185.

Giraud T, Refregier G, de Vienne DM, Le Gac M, Hood ME (2008) Speciation in fungi. *Fungal Genetics and Biology*, **45**, 791–802.

Giraud T, Gladieux P, Gavrilets S (2010) Linking the emergence of fungal plant diseases with ecological speciation. *Trends in Ecology & Evolution*, **25**, 387–395.

Gladieux P, Byrnes E, Fisher M, Aguileta G, Heitman J, Giraud T (2010a) Epidemiology and evolution of fungal pathogens, in plants and animals. In: *Genetics and Evolution of Infectious Diseases* (ed. Tibayrenc M), pp. 59–106. Elsevier, London.

Gladieux P, Zhang X-G, Roldan-Ruiz I *et al.* (2010b) Evolution of the population structure of *Venturia inaequalis*, the apple scab fungus, associated with the domestication of its host. *Molecular Ecology*, **19**, 658–674.

Gladieux P, Guerin F, Giraud T *et al.* (2011) Emergence of novel fungal pathogens by ecological speciation: importance of the reduced viability of immigrants. *Molecular Ecology*, **20**, 4521–4532.

Gladieux P, Devier B, Aguileta G, Cruaud C, Giraud T (2013) Purifying selection after episodes of recurrent adaptive diversification in fungal pathogens. *Infection Genetics and Evolution*, **17**, 123–131.

Glass N, Jacobson D, Shiu P (2000) The genetics of hyphal fusion and vegetative incompatibility in filamentous Ascomycete fungi. *Annual Reviews of Genetics*, **2000**, 165–186.

Goodwin SB, Ben M'Barek S, Dhillon B *et al.* (2011) Finished genome of the fungal wheat pathogen *Mycosphaerella graminicola* reveals dispensome structure, chromosome plasticity, and stealth pathogenesis. *Plos Genetics*, **7**, e1002070.

Gourbiere S, Mallet J (2010) Are species real? The shape of the species boundary with exponential failure, reinforcement, and the "missing snowball". *Evolution*, **64**, 1–24.

Gout L, Fudal I, Kuhn ML *et al.* (2006) Lost in the middle of nowhere: the AvrLm1 avirulence gene of the Dothideomycete *Leptosphaeria maculans*. *Molecular Microbiology*, **60**, 67–80.

Greig D (2007) A screen for recessive speciation genes expressed in the gametes of F1 hybrid yeast. *PLoS Genetics*, **3**, e21.

Gresham D, Desai MM, Tucker CM *et al.* (2008) The repertoire and dynamics of evolutionary adaptations to controlled nutrient-limited environments in yeast. *PLoS Genetics*, **4**, e1000303.

Greub G, Raoult D (2004) Microorganisms resistant to free-living amoebae. *Clinical Microbiology Reviews*, **17**, 413–433.

Guerin F, Gladieux P, Le Cam B (2007) Origin and colonization history of newly virulent strains of the phytopathogenic fungus *Venturia inaequalis*. *Fungal Genetics and Biology*, **44**, 284–292.

Gutenkunst RN, Hernandez RD, Williamson SH, Bustamante CD (2009) Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *Plos Genetics*, **5**, e1000695.

Han M, Thomas G, Lugo-Martinez J, Hahn M (2013) Estimating gene gain and loss rates in the presence of error in genome assembly and annotation using CAFE 3. *Molecular Biology and Evolution*, **30**, 1987–1997.

Hatta R, Ito K, Hosaki Y *et al.* (2002) A conditionally dispensable chromosome controls host-specific pathogenicity in the fungal plant pathogen *Alternaria alternata*. *Genetics*, **161**, 59–70.

Hillenmeyer ME, Fung E, Wildenhain J *et al.* (2008) The chemical genomic portrait of yeast: uncovering a phenotype for all genes. *Science*, **320**, 362–365.

Hood ME, Petit E, Giraud T (2013) Extensive divergence between mating-type chromosomes of the anther-smut fungus. *Genetics*, **193**, 309–315.

Jackson CL, Hartwell LH (1990) Courtship in *Saccharomyces cerevisiae* – Both cell types choose mating partners by responding to the strongest pheromone signal. *Cell*, **63**, 1039–1051.

James TY, Kauff F, Schoch CL *et al.* (2006) Reconstructing the early evolution of Fungi using a six-gene phylogeny. *Nature*, **443**, 818–822.

Jones JDG, Dangl JL (2006) The plant immune system. *Nature*, **444**, 323–329.

Joneson S, Stajich JE, Shiu S-H, Rosenblum EB (2011) Genomic transition to pathogenicity in chytrid fungi. *Plos Pathogens*, **7**, e1002338.

de Jonge R, Bolton M, Thomma B (2011) How filamentous pathogens co-opt plants: the ins and outs of fungal effectors. *Current Opinion in Plant Biology*, **14**, 400–406.

Kaiserer L, Oberparleiter C, Weiler-GÃrz R, Burgstaller W, Leiter E, Marx F (2003) Characterization of the *Penicillium chrysogenum* antifungal protein PAF. *Archives of microbiology*, **180**, 204–210.

Kamper J, Kahmann R, Bolker M *et al.* (2006) Insights from the genome of the biotrophic fungal plant pathogen *Ustilago maydis*. *Nature*, **444**, 97–101.

Kao KC, Sherlock G (2008) Molecular characterization of clonal interference during adaptive evolution in asexual populations of *Saccharomyces cerevisiae*. *Nature Genetics*, **40**, 1499–1504.

Kavanaugh LA, Fraser JA, Dietrich FS (2006) Recent evolution of the human pathogen *Cryptococcus neoformans* by intervarietal transfer of a 14-gene fragment. *Molecular Biology and Evolution*, **23**, 1879–1890.

Keeling P (2009) Functional and ecological impacts of horizontal gene transfer in eukaryotes. *Current Opinion in Genetics & Development*, **19**, 613–619.

Keeling PJ, Palmer JD (2008) Horizontal gene transfer in eukaryotic evolution. *Nature Reviews Genetics*, **9**, 605–618.

Kellis M, Birren BW, Lander ES (2004) Proof and evolutionary analysis of ancient genome duplication in the yeast *Saccharomyces cerevisiae*. *Nature*, **428**, 617–624.

Kennedy P (2010) Ectomycorrhizal fungi and interspecific competition: species interactions, community structure, coexistence mechanisms, and future research directions. *New Phytologist*, **187**, 895–910.

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

Klosterman SJ, Subbarao KV, Kang SC *et al.* (2011) Comparative genomics yields insights into niche adaptation of plant vascular wilt pathogens. *Plos Pathogens*, **7**, e1002137.

Kohn LM (2005) Mechanisms of fungal speciation. *Annual Review of Phytopathology*, **43**, 279–308.

Koskella B, Giraud T, Hood ME (2006) Pathogen relatedness affects the prevalence of within-host competition. *The American Naturalist*, **168**, 121–126.

Lawrence JG, Roth JR (1996) Selfish operons: horizontal transfer may drive the evolution of gene clusters. *Genetics*, **143**, 1843–1860.

Le Gac M, Giraud T (2008) Existence of a pattern of reproductive character displacement in Basidiomycota but not in Ascomycota. *Journal of Evolutionary Biology*, **21**, 761–772.

Leducq J-B (2014) Ecological genomics of adaptation and speciation in Fungi. In: *Ecological Genomics: Ecology and the Evolution of Genes and Genomes* (eds Landry CR & Aubin-Horth N), pp. 49–72. Springer ScienceCBusiness Media, Dordrecht.

Lee H-Y, Chou J-Y, Cheong L, Chang N-H, Yang S-Y, Leu J-Y (2008) Incompatibility of nuclear and mitochondrial genomes causes hybrid sterility between two yeast species. *Cell*, **135**, 1065–1073.

Lefebvre F, Joly D, Labbé C *et al.* (2013) The transition from a phytopathogenic smut ancestor to an anamorphic biocontrol agent deciphered by comparative whole-genome analysis. *Plant Cell*, **25**, 1914.

Leroy T, Lemaire C, Dunemann F, Le Cam B (2013) The genetic structure of a *Venturia inaequalis* population in a heterogeneous host population composed of different *Malus* species. *BMC Evolutionary Biology*, **13**, 64.

Li YF, Costello JC, Holloway AK, Hahn MW (2008) "Reverse ecology" and the power of population genomics. *Evolution*, **62**, 2984–2994.

Li YD, Liang H, Gu Z *et al.* (2009) Detecting positive selection in the budding yeast genome. *Journal of Evolutionary Biology*, **22**, 2430–2437.

Libkind D, Hittinger CT, Valerio E *et al.* (2011) Microbe domestication and the identification of the wild genetic stock of lager-brewing yeast. *Proceedings of the National Academy of Sciences*, **108**, 14539–14544.

Lin Z, Li W-H (2010) Expansion of hexose transporter genes was associated with the evolution of aerobic fermentation in yeasts. *Molecular Biology and Evolution*, **28**, 131–142.

Liti G, Barton DBH, Louis EJ (2006) Sequence diversity, reproductive isolation and species concepts in *Saccharomyces*. *Genetics*, **174**, 839–850.

Liti G, Carter DM, Moses AM *et al.* (2009) Population genomics of domestic and wild yeasts. *Nature*, **458**, 337–341.

Lopez-Villavicencio M, Jonot O, Coantic A, Hood M, Enjalbert J, Giraud T (2007) Multiple infections by the anther smut pathogen are frequent and involve related strains. *PloS Pathogens*, **3**, e176.

López-Villavicencio M, Courjol F, Gibson A *et al.* (2011) Competition, cooperation among kin and virulence in multiple infections. *Evolution*, **65**, 1357–1366.

Luikart G, England PR, Tallmon D, Jordan S, Taberlet P (2003) The power and promise of population genomics: from genotyping to genome typing. *Nature Reviews Genetics*, **4**, 981–994.

Luque EM, Gutierrez G, Navarro-Sampedro L *et al.* (2012) A relationship between carotenoid accumulation and the distribution of species of the fungus *Neurospora* in Spain. *PLoS ONE*, **7**, e33658.

Ma L-J, van der Does HC, Borkovich KA *et al.* (2010) Comparative genomics reveals mobile pathogenicity chromosomes in *Fusarium*. *Nature*, **464**, 367–373.

Machida M, Asai K, Sano M *et al.* (2005) Genome sequencing and analysis of *Aspergillus oryzae*. *Nature*, **438**, 1157–1161.

Manning V, Pandelova I, Dhillon B *et al.* (2013) Comparative genomics of a plant-pathogenic fungus, Pyrenophora tritici-repentis, reveals transduplication and the impact of repeat elements on pathogenicity and population divergence. *G3*, **3**, 41–63.

Marcet-Houben M, Gabaldon T (2010) Acquisition of prokaryotic genes by fungal genomes. *Trends in Genetics*, **26**, 5–8.

Marcet-Houben M, Marceddu G, Gabaldon T (2009) Phylogenomics of the oxidative phosphorylation in fungi reveals extensive gene duplication followed by functional divergence. *BMC Evolutionary Biology*, **9**, 12.

Martin F, Aerts A, Ahren D *et al.* (2008) The genome of *Laccaria bicolor* provides insights into mycorrhizal symbiosis. *Nature*, **452**, 88–U87.

Martinez D, Larrondo LF, Putnam N *et al.* (2004) Genome sequence of the lignocellulose degrading fungus *Phanerochaete chrysosporium* strain RP78. *Nature Biotechnology*, **22**, 695–700.

Marx F, Binder U, Leiter E, Pocsi I (2008) The *Penicillium chrysogenum* antifungal protein PAF, a promising tool for the development of new antifungal therapies and fungal cell biology studies. *Cellular and Molecular Life Sciences*, **65**, 445–454.

Maynard Smith J (1966) Sympatric Speciation. *American Naturalist*, **100**, 637–650.

Mboup M, Bahri B, Leconte M, De Vallavieille-Pope C, Kaltz O, Enjalbert J (2012) Genetic structure and local adaptation of European wheat yellow rust populations: the role of temperature-specific adaptation. *Evolutionary Applications*, **5**, 341–352.

McDonald B (2010) How can we achieve durable disease resistance in agricultural ecosystems? *New Phytologist*, **185**, 3–5.

McDowell JM (2011) Genomes of obligate plant pathogens reveal adaptations for obligate parasitism. *Proceedings of the National Academy of Sciences of the United States of America*, **108**, 8921–8922.

McGary KL, Slot JC, Rokas A (2013) Physical linkage of metabolic genes in fungi is an adaptation against the accumulation of toxic intermediate compounds. *Proceedings of the National Academy of Sciences USA*, **110**, 11481–11486.

Messer PW, Petrov DA (2013) Frequent adaptation and the McDonald and Kreitman test. *Proceedings of the National Academy of Sciences USA*, **110**, 8615–8620.

Mitra K, Carvunis AR, Ramesh SK, Ideker T (2013) Integrative approaches for finding modular structure in biological networks. *Nature Reviews Genetics*, **14**, 719–732.

Morris CE, Bardin M, Kinkel LL, Moury B, Nicot PC, Sands DC (2009) Expanding the paradigms of plant pathogen life history and evolution of parasitic fitness beyond agricultural boundaries. *PLoS Pathogens*, **5**, e1000693.

Munkacsi AB, Kawakami S, Pan JJ *et al.* (2006) Genome-wide assessment of tandem repeat markers for biogeographical analyses of the corn smut fungus, *Ustilago maydis*. *Molecular Ecology Notes*, **6**, 221–223.

Naumova ES, Naumov GI, Masneuf-PomarÃ¨de I, Aigle M, Dubourdieu D (2005) Molecular genetic study of introgression between *Saccharomyces bayanus* and *S. cerevisiae*. *Yeast*, **22**, 1099–1115.

Navarro A, Barton NH (2003) Chromosomal speciation and molecular divergence – Accelerated evolution in rearranged chromosomes. *Science*, **300**, 321–324.

Neafsey DE, Barker BM, Sharpton TJ *et al.* (2010) Population genomic sequencing of *Coccidioides* fungi reveals recent hybridization and transposon control. *Genome Research*, **20**, 938–946.

Nieuwenhuis B, Debets A, Aanen D (2010) Sexual selection in mushroom-forming basidiomycetes. *Proceedings of the Royal Society B*, **278**, 152–157.

Nosil P (2012) *Ecological Speciation P9*. Oxford University Press, Oxford.

Novo M, Bigey F, Beyne E *et al.* (2009) Eukaryote-to-eukaryote gene transfer events revealed by the genome sequence of the wine yeast *Saccharomyces cerevisiae* EC1118. *Proceedings of the National Academy of Sciences*, **106**, 16333–16338.

Ohm R, Feau N, Henrissat BSC *et al.* (2012) Diverse lifestyles and strategies of plant pathogenesis encoded in the genomes of eighteen Dothideomycetes fungi. *PLoS Pathogens*, **8**, e1003037.

Ohno S (1970) *Evolution by Gene Duplication*. Springer-Verlag, New York.
Oliver R (2012) Genomic tillage and the harvest of fungal phytopathogens. *New Phytologist*, **196**, 1015–1023.
Olson-Manning CF, Wagner MR, Mitchell-Olds T (2012) Adaptive evolution: evaluating empirical support for theoretical predictions. *Nature Reviews Genetics*, **13**, 867–877.
Orr HA, Turelli M (2001) The evolution of postzygotic isolation: accumulating Dobzhansky-Muller incompatibilities. *Evolution*, **55**, 1085–1094.
Oud B, Guadalupe-Medina V, Nijkamp JF *et al.* (2013) Genome duplication and mutations in ACE2 cause multicellular, fast-sedimenting phenotypes in evolved *Saccharomyces cerevisiae*. *Proceedings of the National Academy of Sciences of the United States of America*, **110**, E4223–E4231.
Papp B, Notebaart RA, Pal C (2011) Systems-biology approaches for predicting genomic evolution. *Nature Reviews Genetics*, **12**, 591–602.
Pellegrini M (2012) Using phylogenetic profiles to predict functional relationships. *Methods in Molecular Biology*, **804**, 167–177.
Perez-Ortin JE, Querol A, Puig S, Barrio E (2002) Molecular characterization of a chromosomal rearrangement involved in the adaptive evolution of yeast strains. *Genome Research*, **12**, 1533–1539.
Peyretaillade E, El Alaoui H, Diogon M *et al.* (2011) Extreme reduction and compaction of microsporidian genomes. *Research in Microbiology*, **162**, 598–606.
Powell AJ, Conant GC, Brown DE, Carbone I, Dean RA (2008) Altered patterns of gene duplication and differential gene gain and loss in fungal pathogens. *BMC Genomics*, **9**, 15.
Presgraves DC (2010) The molecular evolutionary basis of species formation. *Nature Reviews Genetics*, **11**, 175.
Rep M, Kistler HC (2010) The genomic organization of plant pathogenicity in *Fusarium* species. *Current Opinion in Plant Biology*, **13**, 420–426.
Rice WR (1987) Speciation via habitat specialization: the evolution of reproductive isolation as a correlated character. *Evolutionary Biology*, **1**, 301–314.
Rogers DW, Greig D (2009) Experimental evolution of a sexually selected display in yeast. *Proceedings of the Royal Society B-Biological Sciences*, **276**, 543–549.
Roper M, Ellison C, Taylor JW, Glass NL (2011) Nuclear and genome dynamics in multinucleate ascomycete fungi. *Current Biology*, **21**, R786–R793.
Rosenblum E *et al.* (2013) Complex history of the amphibian-killing chytrid fungus revealed with genome resequencing data. *Proceedings of the National Academy of Sciences*, **110**, 9385–9390.
Rouxel T, Grandaubert J, Hane JK *et al.* (2011) Effector diversification within compartments of the *Leptosphaeria maculans* genome affected by Repeat-Induced Point mutations. *Nature Communications*, **2**, 202.
Sass V, Milles J, Kramer J, Prange A (2007) Competitive interactions of *Fusarium graminearum* and *Alternaria alternata* in vitro in relation to deoxynivalenol and zearalenone production. *Journal of Food Agriculture & Environment*, **5**, 257–261.
Schardl CL, Craven KD (2003) Interspecific hybridization in plant-associated fungi and oomycetes: a review. *Molecular Ecology*, **12**, 2861–2873.
Schardl C, Young C, Hesse U *et al.* (2013) Plant-symbiotic fungi as chemical engineers: multi-genome analysis of the Clavicipitaceae reveals dynamics of alkaloid loci. *PLoS Genetics*, **9**, e1003323.
Schirawski J, Mannhaupt G, Münch K *et al.* (2010) Pathogenicity determinants in smut fungi revealed by genome comparison. *Science*, **330**, 1546.
Schoustra S, Punzalan D (2012) Correlation of mycelial growth rate with other phenotypic characters in evolved genotypes of *Aspergillus nidulans*. *Fungal Biology*, **116**, 630–636.
Schoustra SE, Bataillon T, Gifford DR, Kassen R (2009) The properties of adaptive walks in evolving populations of fungus. *Plos Biology*, **7**, e1000250.
Schulze-Lefert P, Panstruga R (2011) A molecular evolutionary concept connecting nonhost resistance, pathogen host range, and pathogen speciation. *Trends in Plant Science*, **16**, 117–125.
Servedio MR, Van Doorn GS, Kopp M, Frame AM, Nosil P (2011) Magic traits in speciation: 'magic' but not rare? *Trends in Ecology & Evolution*, **26**, 389–397.
Shahid M, Han S, Yoell H, Xu J (2008) Fitness distribution and transgressive segregation across 40 environments in a hybrid progeny population of the human-pathogenic yeast *Cryptococcus neoformans*. *Genome*, **51**, 272–281.
Sharpton TJ, Stajich JE, Rounsley SD *et al.* (2009) Comparative genomic analyses of the human fungal pathogens *Coccidioides* and their relatives. *Genome Research*, **19**, 1722–1731.
Silar P (2005) Peroxide accumulation and cell death in filamentous fungi induced by contact with a contestant. *Mycological Research*, **109**, 137–149.
Silva DN, Talhinhas P, Cai L *et al.* (2012) Host jump drives rapid and recent ecological speciation of the emergent fungal pathogen *Colletotrichum kahawae*. *Molecular Ecology*, **21**, 2655–2670.
Slot JC, Hibbett DS (2007) Horizontal transfer of a nitrate assimilation gene cluster and ecological transitions in fungi: a phylogenetic study. *PLoS ONE*, **2**, e1097.
Slot JC, Rokas A (2010) Multiple GAL pathway gene clusters evolved independently and by different mechanisms in fungi. *Proceedings of the National Academy of Sciences of the United States of America*, **107**, 10136–10141.
Slot J, Rokas A (2011) Horizontal transfer of a large and highly toxic secondary metabolic gene cluster between fungi. *Current Biology*, **21**, 134–139.
Spanu P (2012) The genomics of obligate (and nonobligate) biotrophs. *Annual Reviews of Phytopathology*, **50**, 91–109.
Spanu P, Kämper J (2010) Genomics of biotrophy in fungi and oomycetes–emerging patterns. *Current Opinion in Plant Biology*, **13**, 409–414.
Spanu P, Abbott J, Amselem J *et al.* (2010) Genome expansion and gene loss in powdery mildew fungi reveal tradeoffs in extreme parasitism. *Science*, **330**, 1543–1546.
Stajich JE, Berbee ML, Blackwell M *et al.* (2009) The Fungi. *Current Biology*, **19**, R840–R845.
Stajich JE, Wilke SK, Ahren D *et al.* (2010) Insights into evolution of multicellular fungi from the assembled chromosomes of the mushroom *Coprinopsis cinerea (Coprinus cinereus)*. *Proceedings of the National Academy of Sciences of the United States of America*, **107**, 11889–11894.
Staves PA, Knell RJ (2010) Virulence and competitiveness: testing the relationship during inter- and intraspecific mixed infections. *Evolution*, **64**, 2643–2652.
Stergiopoulos I, de Wit PJGM (2009) Fungal effector proteins. *Annual Review of Phytopathology*, **47**, 233–263.

Stergiopoulos I, De Kock MJD, Lindhout P, De Wit PJGM (2007) Allelic variation in the effector genes of the tomato pathogen *Cladosporium fulvum* reveals different modes of adaptive evolution. *Molecular Plant-Microbe Interactions*, **20**, 1271–1283.

Stergiopoulos I, Kourmpetis YAI, Slot JC, Bakker FT, de Wit P, Rokas A (2012) *In silico* characterization and molecular evolutionary analysis of a novel superfamily of fungal effector proteins. *Molecular Biology and Evolution*, **29**, 3371–3384.

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

Stukenbrock EH, McDonald BA (2008) The origins of plant pathogens in agro-ecosystems. *Annual Review of Phytopathology*, **46**, 75–100.

Stukenbrock EH, McDonald BA (2009) Population genetics of fungal and oomycete effectors involved in gene-for-gene interactions. *Molecular Plant-Microbe Interactions*, **22**, 371–380.

Stukenbrock EH, Banke S, Javan-Nikkhah M, McDonald BA (2007) Origin and domestication of the fungal wheat pathogen *Mycosphaerella graminicola* via sympatric speciation. *Molecular Biology and Evolution*, **24**, 398–411.

Stukenbrock EH, Bataillon T, Dutheil JY *et al.* (2011) The making of a new pathogen: insights from comparative population genomics of the domesticated wheat pathogen *Mycosphaerella graminicola* and its wild sister species. *Genome Research*, **21**, 2157–2166.

Sun Y, Corcoran P, Menkis A, Whittle CA, Andersson SGE, Johannesson H (2012) Large-scale introgression shapes the evolution of the mating-type chromosomes of the filamentous ascomycete *Neurospora tetrasperma*. *PLoS Genetics*, **8**, e1002820.

Sun BF, Xiao JH, He S, Liu L, Murphy RW, Huang DW (2013) Multiple interkingdom horizontal gene transfers in *Pyrenophora* and closely related species and their contributions to phytopathogenic lifestyles. *PLoS ONE*, **8**, e60029.

Syvanen M (2012) Evolutionary implications of horizontal gene transfer. *Annual Review of Ecology and Systematics*, **46**, 341–358.

Terauchi R, Yoshida K (2010) Towards population genomics of effector-effector target interactions. *New Phytologist*, **187**, 929–939.

Teresa Avelar A, Perfeito L, Gordo I, Godinho Ferreira M (2013) Genome architecture is a selectable trait that can be maintained by antagonistic pleiotropy. *Nature Communication*, **4**, 2235.

Thompson DA, Regev A (2009) Fungal regulatory evolution: cis and trans in the balance. *Febs Letters*, **583**, 3959–3965.

Tirosh I, Weinberger A, Carmi M, Barkai N (2006) A genetic signature of interspecies variations in gene expression. *Nature Genetics*, **38**, 830–834.

Tuch BB, Galgoczy DJ, Hernday AD, Li H, Johnson AD (2008) The evolution of combinatorial gene regulation in fungi. *Plos Biology*, **6**, 352–364.

Turner E, Jacobson DJ, Taylor JW (2011) Genetic architecture of a reinforced, postmating, reproductive isolation barrier between *Neurospora* species indicates evolution via natural selection. *Plos Genetics*, **7**, e1002204.

Upadhya R, Campbell LT, Donlin MJ, Aurora R, Lodge JK (2013) Global transcriptome profile of *Cryptococcus neoformans* during exposure to hydrogen peroxide induced oxidative stress. *PLoS ONE*, **8**, e55110.

Vacher C, Vile D, Helion E, Piou D, Desprez-Loustau ML (2008) Distribution of parasitic fungal species richness: influence of climate versus host species diversity. *Diversity and Distributions*, **14**, 786–798.

Van de Wouw AP, Cozijnsen AJ, Hane JK *et al.* (2010) Evolution of linked avirulence effectors in *Leptosphaeria maculans* is affected by genomic environment and exposure to resistance genes in host plants. *Plos Pathogens*, **6**, e1001180.

de Vienne DM, Aze J (2012) Efficient prediction of co-complexed proteins based on coevolution. *PLoS ONE*, **7**, e48728.

de Vienne DM, Refregier G, Lopez-Villavicencio M, Tellier A, Hood ME, Giraud T (2013) Cospeciation vs host-shift speciation: methods for testing, evidence from natural associations and relation to coevolution. *New Phytologist*, **198**, 347–385.

Walton JD (2000) Horizontal gene transfer and the evolution of secondary metabolite gene clusters in fungi: an hypothesis. *Fungal Genetics and Biology*, **30**, 167–171.

Ward TJ, Bielawski JP, Kistler HC, Sullivan E, O'Donnell K (2002) Ancestral polymorphism and adaptive evolution in the trichothecene mycotoxin gene cluster of phytopathogenic *Fusarium*. *Proceedings of the National Academy of Sciences of the United States of America*, **99**, 9278–9283.

Whitlock MC (1996) The red queen beats the Jack-of-all-trades: the limitations of phenotypic plasticity and niche breadth. *The American Naturalist*, **148**, S65–S77.

Wicker T, Oberhaensli S, Parlange F *et al.* (2013) The wheat powdery mildew genome shows the unique evolution of an obligate biotroph. *Nature Genetics*, **45**, 1092–1096.

Winzeler EA, Shoemaker DD, Astromoff A *et al.* (1999) Functional characterization of the *S. cerevisiae* genome by gene deletion and parallel analysis. *Science*, **285**, 901–906.

Wolbach DJ, Thompson DA, Gasch AP, Regev A (2009) From elements to modules: regulatory evolution in Ascomycota fungi. *Current Opinion in Genetics & Development*, **19**, 1–8.

Wolfe K (2004) Evolutionary genomics: yeasts accelerate beyond BLAST. *Current Biology*, **14**, R392–R394.

Xhaard C, Fabre B, Andrieux A *et al.* (2011) The genetic structure of the plant pathogenic fungus *Melampsora larici-populina* on its wild host is extensively impacted by host domestication. *Molecular Ecology*, **20**, 2739–2755.

Xu J, Saunders CW, Hu P *et al.* (2007) Dandruff-associated *Malassezia* genomes reveal convergent and divergent virulence traits shared with plant and human fungal pathogens. *Proceedings of the National Academy of Sciences*, **104**, 18730–18735.

Zaffarano PL, McDonald BA, Linde CC (2008) Rapid speciation following host shifts in the plant pathogenic fungus *Rhynchosporium*. *Evolution*, **62**, 1418–1436.

Zuccaro A, Lahrmann U, Guldener U *et al.* (2011) Endophytic life strategies decoded by genome and transcriptome analyses of the mutualistic root symbiont *Piriformospora indica*. *Plos Pathogens*, **7**, e1002290.

All authors contributed to writing the review. TG supervised the writing.

4.3 Article 2 : Sex and parasites : genomic and transcriptomic analysis of *Microbotryum lychnidis-dioicae*, the biotrophic and plant-castrating anther smut fungus

L'article de synthèse précédent a montré que les champignons constituaient d'excellents modèles pour l'étude de la divergence adaptative chez les Eucaryotes. Etudier la génomique de l'adaptation au sein d'une espèce nécessite d'abord l'obtention d'un génome ou transcriptome de référence. L'étude du contenu en gènes et de l'architecture génomique peut renseigner sur les bases génomiques de l'adaptation. Le séquençage du génome de *M. lychnidis-dioicae*, l'espèce la plus étudiée du complexe *M. violaceum*, a été réalisé par un consortium auquel notre équipe a participé. Une souche haploïde de type sexuel a_1 a été séquencée avec la technologie 454.

L'article suivant, publié dans BMC Genomics, présente l'analyse de ce génome. Comme l'assemblage était très fragmenté (plus de 1000 scaffolds pour 26 Mb), et que les scaffolds n'étaient pas assignés à des chromosomes, l'assemblage de cet génome n'a pas permis d'étudier en détail l'architecture du génome, et en particulier celle du chromosome de type sexuel a_1 (cf. 2.1.3 p. 32). L'étude du génome s'est donc en grande partie focalisée sur l'analyse du contenu en gènes, particulièrement les gènes qui pourraient être impliqués dans le mode de vie biotrophe de *M. lychnidis-dioicae*. Nous avons également étudié les caractéristiques structurales de ce génome. Dans ce cadre, j'ai participé à l'analyse du contenu en GC. Les figures et tables supplémentaires sont en annexe A.3 (page 248).

Perlin et al. BMC Genomics (2015) 16:461
DOI 10.1186/s12864-015-1660-8

BMC Genomics

RESEARCH ARTICLE Open Access

Sex and parasites: genomic and transcriptomic analysis of *Microbotryum lychnidis-dioicae*, the biotrophic and plant-castrating anther smut fungus

Michael H Perlin[1*], Joelle Amselem[2,3†], Eric Fontanillas[4,5†], Su San Toh[1†], Zehua Chen[6†], Jonathan Goldberg[6], Sebastien Duplessis[7,8], Bernard Henrissat[9,10], Sarah Young[6], Qiandong Zeng[6], Gabriela Aguileta[11], Elsa Petit[4,5,9], Helene Badouin[4,5], Jared Andrews[1], Dominique Razeeq[1], Toni Gabaldón[11,12,13], Hadi Quesneville[2], Tatiana Giraud[4,5], Michael E. Hood[14†], David J. Schultz[1] and Christina A. Cuomo[6*]

Abstract

Background: The genus *Microbotryum* includes plant pathogenic fungi afflicting a wide variety of hosts with anther smut disease. *Microbotryum lychnidis-dioicae* infects *Silene latifolia* and replaces host pollen with fungal spores, exhibiting biotrophy and necrosis associated with altering plant development.

Results: We determined the haploid genome sequence for *M. lychnidis-dioicae* and analyzed whole transcriptome data from plant infections and other stages of the fungal lifecycle, revealing the inventory and expression level of genes that facilitate pathogenic growth. Compared to related fungi, an expanded number of major facilitator superfamily transporters and secretory lipases were detected; lipase gene expression was found to be altered by exposure to lipid compounds, which signaled a switch to dikaryotic, pathogenic growth. In addition, while enzymes to digest cellulose, xylan, xyloglucan, and highly substituted forms of pectin were absent, along with depletion of peroxidases and superoxide dismutases that protect the fungus from oxidative stress, the repertoire of glycosyltransferases and of enzymes that could manipulate host development has expanded. A total of 14 % of the genome was categorized as repetitive sequences. Transposable elements have accumulated in mating-type chromosomal regions and were also associated across the genome with gene clusters of small secreted proteins, which may mediate host interactions.

Conclusions: The unique absence of enzyme classes for plant cell wall degradation and maintenance of enzymes that break down components of pollen tubes and flowers provides a striking example of biotrophic host adaptation.

Keywords: *Microbotryum violaceum*, Anther smuts, CAZymes, Transposable elements, Mating-type chromosomes, Pathogen alteration of host development

* Correspondence: michael.perlin@louisville.edu; cuomo@broadinstitute.org
†Equal contributors
[1]Department of Biology, Program on Disease Evolution, University of Louisville, Louisville, KY 40292, USA
[6]Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA
Full list of author information is available at the end of the article

Background

Members of the genus *Microbotryum* are pathogenic fungi that have a global distribution and infect over nine families of host plants, with most replacing floral structures with fungal spores [1]. Within this genus, the anther-smut fungi of plants in the Caryophyllaceae, consist of many recognized and cryptic species, most possessing a very narrow host range, although some are more generalist [2]. Hybrid incompatibilities for the species examined in this complex leads to post-zygotic isolation in the forms of inviability and sterility [3–5]. The high rates of selfing and ecological specialization on different host plants are factors that should promote speciation in *Microbotryum* [6, 7]. The anther smut fungi thus allow examination of the ecology and evolution of host/pathogen interactions in "wild," non-agricultural environments [8, 9], where the genetically variable hosts provides an important contrast to the heavily studied, more monocultural hosts of agricultural systems. *Microbotryum* species also serve as a model for emerging infectious disease through host shifts [9, 10], for studying the evolution of mating systems, non-recombining mating-type chromosomes and sex chromosomes [11, 12], and for examining pathogens that alter the development of the host [13].

As obligate parasites, *Microbotryum* anther smut fungi must complete their life cycle in association with a host plant. Their fungal diploid teliospore masses give the flowers a dark, powdery appearance, thus the name "anther smut." Teliospores of *Microbotryum* are transported from diseased to healthy plants by direct transmission when plants are in close proximity [14] or by pollinating insects [15], where once deposited the diploid fungus germinates and undergoes meiosis to give rise to four haploid cells [6]. Each of these cells can bud off yeast-like sporidia on the plant surface. New infectious dikaryotic hyphae are rapidly produced by conjugation of two cells of opposite mating-type (a_1 and a_2) and enter the host tissue to grow endophytically until they reach the bud meristems and anthers [16]. Here the nuclei fuse (karyogamy) and teliosporogenesis occurs thus completing the life cycle (Fig. 1; [6, 17–19]).

The commandeering of insect pollinators for disease transmission is associated with significant pathogenic alterations of the host's floral morphology, particularly in relation to male and female structures. Although diseased plants are only slightly affected in vegetative morphology and survival [20], infection often results in complete host sterility, as no pollen is produced in the anthers and the ovary becomes rudimentary. Interestingly, in dioecious or gynodioecious species of *Silene* (*e.g.*, *S. latifolia* and *S. vulgaris*, respectively), ovaries of diseased female plants are aborted while a male morphology develops with spore-bearing

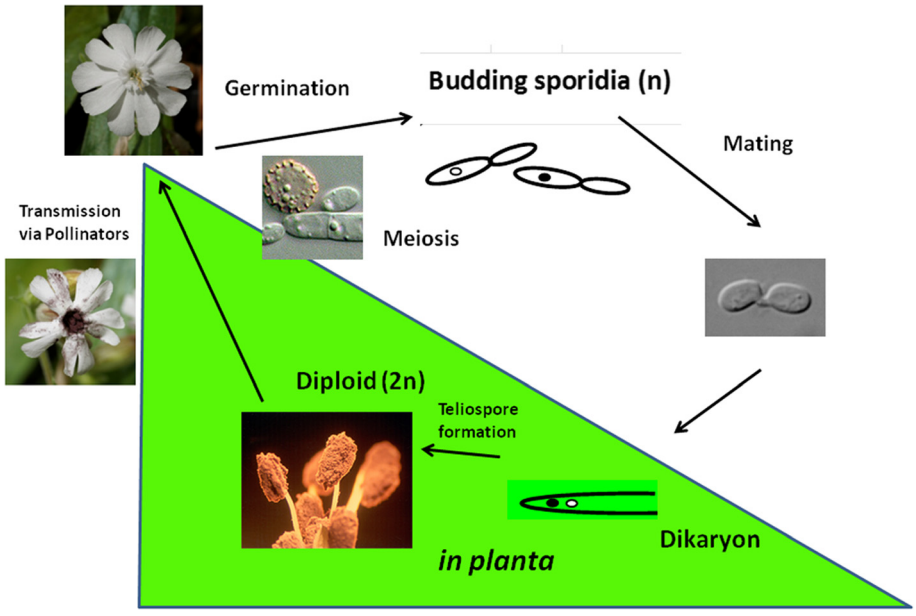


Fig. 1 Lifecycle of *Microbotryum lychnidis-dioicae*. The infection cycle for *M. lychnidis-dioicae*) is shown. Infection begins when diploid teliospores germinate on a suitable plant surface, and after meiosis, produce linear tetrads of haploid basidiospores. These cells may mate with cells of opposite mating-type (*i.e.*, *a1* or *a2*), either directly within the tetrad or after budding to yield free yeast-like sporidia. Mated cells form conjugation bridges and, upon receiving suitable cues from the host plant, develop into dikaryotic hyphae that penetrate the host tissue. The hyphae progress systemically through the plant and migrate to the flower primordia. There nuclear fusion (karyogamy) occurs, as the hyphae break up and develop into diploid teliospores that replace the pollen in the anthers of the developing flowers. Pollinator species are then able to transmit the spores to other uninfected flowers, thus completing the cycle

anthers. The basis for the changes in female plants is not fully understood. Studies have identified host genes expressed during infection of females that are also normally expressed in uninfected males (*MEN* genes [21]; *SLM2* [22]).

The formation of a hyphal dikaryon by mating between haploid cells is a prerequisite for infection in *Microbotryum*, and thus each newly diseased plant represents the completion of a sexual reproductive cycle. In heterothallic fungi, mating can occur only between haploid cells carrying different mating-types, which are controlled by alleles at one or two loci [23, 24]. In a few fungi, recombination suppression has evolved around the mating-type locus/loci [25–27], sometimes leading to structurally dimorphic chromosomes similar to the sex chromosomes in plants and animals [28, 29]. The suppression of recombination on mating-type chromosomes is expected to lead to degeneration; this can be manifested as transposable element accumulation [30], codon usage degeneration [31], accumulation of deleterious mutations [32], and eventually to chromosomal dimorphism, and thus the emergence of allosomes [12]. Such degeneration has in fact been observed on the mating-type chromosomes of some fungi, including the ascomycete *Neurospora tetrasperma* [31] and the basidiomycete *Microbotryum lychnidis-dioicae* [12, 30, 33], the most well-studied representative of the anther-smut fungi.

We sequenced the genome of haploid isolate Lamole p1A1 [11], of the a_1 mating type, to represent the *M. lychnidis-dioicae* species found in association with the perennial, dioecious host, *Silene latifolia.* RNA-Sequencing of distinct life cycle stages was incorporated to validate gene content and measure expression changes during infection. We identified gene family expansions that could play a role in plant infection by comparing *M. lychnidis-dioicae* to other basidiomycetes, taking advantage of increasing genome coverage of the Pucciniomycotina subphylum. The identification of genes that are induced or repressed during infection highlighted carbohydrate active enzymes (CAZymes) that may be involved in host cell degradation or manipulation of host development. Additionally, the *M. lychnidis-dioicae* genome is riddled with a diverse array of transposable elements (TEs), including a higher proportion of Helitron elements than found in the much larger and more highly-repetitive genomes of related rust fungi. Genome regions corresponding to the mating-type chromosomes of *M. lychnidis-dioicae* [11] are enriched for repetitive sequence. Further analysis of the sequence of the entire a_1 mating-type chromosome identified more than 300 genes linked with mating-type. Together, these findings provide an in-depth portrait of genetic architecture and adaptation in a specialized fungal plant pathogen.

Results

Genome sequence and content

The 25.2 Mb haploid genome of *M. lychnidis-dioicae* was sequenced using 454 technology, generating high coverage of three different-sized libraries (Additional file 1), and assembled using Newbler (Table 1). The assembly was comprised of 1,231 scaffolds where the average base was present in a scaffold of 185 kb and a contig of 50 kb (N50 measure, Table 1). Despite the large number of contigs, the assembly was a nearly complete representation of the sequenced genome, comprising 97 % of sequenced bases. The assembly included five scaffold ends with the typical fungal telomere repeat (TTAGGG), though three of these scaffolds were smaller than 1 kb in size.

High coverage strand-specific RNA-Seq, generated from three biological conditions (Additional file 2) assisted with the prediction of 7,364 protein coding genes and identified expression changes potentially important for the pathogenic lifecycle. Sampled conditions included two *in vitro* conditions, haploid cells grown on yeast peptone dextrose media (YPD) agar (referred to as rich) or on 2 % water agar (referred to as nutrient limited). These were compared to a sample from infected male plant tissue during the late stages of fungal development, where teliospores form on partially and fully opened smutted flowers [34, 35] (referred to as "MI-late"). Incorporation of RNA-Seq into predicted gene structures (Methods) defined UTRs for the vast majority (more than 6,100) predicted genes; the average length of 5′ and 3′ UTRs was 183 bases and 253 respectively. Coding sequences average 1,614 bases (median of 1,338 bases) in length and contain 5.6 exons; genes are separated by intergenic regions 502 bases in

Table 1 Genome statistics of nuclear genome and mating-type chromosome regions

	Nuclear genome	NRR[a] regions	PAR[b] regions
Assembly size (Mb)	26.1	1.86	0.38
Scaffolds (count)	1,231	85	2
Scaffold N50 (kb)	185	48	381
Contigs (count)	2,104	229	16
Contig N50 (kb)	50	13	45
GC content (%)	55.4	54.6	53.9
TE content (%)	14	41	13
Protein coding genes	7,364	350	99
Mean coding length	1,614	1,344	1,408
Median coding length	1,338	954	1,302
Mean exons/gene (count)	5.6	4.6	5.1
Mean intercds length (bp)[c]	1,181	2,600	1,861
tRNAs	134	5	2

[a]Non-recombining regions (NRR). [b]Pseudo-autosomal regions (PAR). [c]average length between coding sequence (cds) start and stop

length on average. The *M. lychnidis-dioicae* gene set has high coverage of a core eukaryotic gene set [36], highest in fact than any of the fungal gene sets used in comparative analysis (Additional file 3), suggesting the assembly includes a highly complete gene set.

The mitochondrial genome consisted of a single finished contig of 97 kb and included the canonical set of genes. These were the respiratory-related proteins of the NADH dehydrogenase family (*nad1-6* and *nad4L*), apocytochrome b (*cob*), cytochrome oxidases (*cox1-3*), the proteins related to ATP synthesis (*atp6*, *atp8* and *atp9*); ribosomal RNAs (*rns*, and sequence similar to *rnl*); ribosomal proteins (*rps3*); DNA polymerase (*dpo*) and 25 tRNAs. Homing endonucleases of the LAGLIDADG and GIY-YIG families and a maturase protein were located in mitochondrial intronic regions, including within three introns of *cox1*.

To examine the presence of AT-rich isochores and more generally the genome structure in GC composition in *M. lychnidis*-dioicae, we measured the fluctuation of GC percent along the assembly. Although the genome is not organized in discrete isochores as in some fungal genomes [37], we observe some large-scale variation (>100 kb) in base composition within chromosomes, as well as finer-scale fluctuations (Additional file 4A). The GC content was positively correlated with coding density, with the most significant correlations for 5 and 10 kb windows explaining 16.8 % of the variance (Additional file 4B). This correlation has been explained in other systems by biased codon usage toward GC-rich codons or alternatively biased gene conversion occurring more frequently in coding than in non-coding sequences [38]. In fact, an analysis of the preferred codons (*i.e.,* the most frequently used codons in the predicted genes), showed that 17 out of 18 had a GC base in the third position, which is the most degenerate position and therefore primarily influences the GC composition of genes (Additional file 5).

Shifts in transposable element type, location, and impact of RIP

The genome of *M. lychnidis-dioicae* contained diverse transposable elements (TEs), represented by 286 consensus elements, covering 14 % of the total assembly. The overall TE content was lower than of other species in Pucciniomycota; in *Puccinia graminis* f. sp. *tritici* or *Melampsora larici-populina*, in which TEs account for nearly 45 % of the assembled genomes, contributing to the expanded genome size of these fungi [39]. Among class I retrotransposon elements (36 % of TE sequences), Long Terminal Repeat (LTR) elements were the most common (28 % of TE sequences), in particular *Copia*-like elements (20 % of TE sequences), in agreement with prior studies of genome sampling and expressed sequence profiles [30, 40]. The remainder of class I elements consisted of Long Interspersed Nuclear Element (LINE) (7 % of TE sequences) and *Dictyostelium* Intermediate Repeat Sequence (DIRS) elements (1 % of TE sequences). Among class II DNA transposons (23 % of TE sequences), Terminal Inverted Repeat (TIR) and *Helitron* elements, which transpose by rolling-circle replication [41] account for 12 % and 10 % of TE sequences, respectively (Additional file 6, Table 2). The *Helitron* proportion was an order of magnitude higher than in the more repetitive genomes of other Pucciniomycota fungi *P. graminis* f. sp. *tritici* and *M. larici-populina*, for which a similar analysis characterized only 1 % of TE elements as *Helitrons* [39].

The TE categories varied significantly in their proximity to genes. A chi-squared test of heterogeneity found a significant difference in the TE content of regions nearby genes, comparing regions less than versus greater than 1 kb upstream and downstream of genes (upstream region p-value < 2.2e-16; downstream region p-value < 2.2e-08, Additional file 7). In particular, class II elements (TIR and *Helitrons*) were closer to genes than class I elements (LTR and LINE), with a greater enrichment upstream of genes compared to downstream (Additional file 7). In addition, there appeared to be an association between TE-rich regions in *M. lychnidis-dioicae* and genes for Small Secreted Proteins (SSPs), which can include effector proteins involved in host-pathogen interactions as suggested in some pathogen genomes [42, 43]. SSPs were indeed located nearer to TEs than the set of all other non-SSP genes (Chi-squared p-value < 5e-4; Additional file 8).

Hypermutation in TEs that resembles the genome defense, Repeat-Induced Point mutation (RIP), has previously been observed in the LTR elements (*copia*-like and *gypsy*-like elements) and *Helitron* transposons of *M. lychnidis-dioicae* [44, 45]. Some genes that appeared similar to those necessary for RIP in *Neurospora crassa* [46, 47] were found in the *M. lychnidis-dioicae* genome. These include a cytosine methyltransferase (MVLG_04160), but establishing the orthology with the RIP-essential *rid* gene

Table 2 Genome coverage of TE families (10,283 copies of 286 REPET consensus sequences)

TE copies	Copy number	Coverage relative to TE space		Coverage relative to assembly size	
LTR	1018	27.70	Class I 35.64	3.89	Class I 5.01
DIRS	57	1.34		0.19	
LINE	392	6.61		0.93	
SINE	0	0		0	
TIR	438	12.18	Class II 22.76	1.71	Class II 3.20
MITE	39	0.39		0.06	
Helitron	373	10.18		1.43	
Unknown	1942	41.60	41.60	5.85	5.85
Total	4259	100	100	14.06	14.06

from *N. crassa* versus other cytosine methyltransferases (*e.g.* Dim-2) would require further investigation. *M. lychnidis-dioicae* sequences similar to the Dim-5H3 histone methyltransferase that is essential for marking RIP regions in *N. crassa* (MVLG_02125, MVLG_05378) were also found.

Evaluation of dinucleotide signature at transition mutations in 179 TE families (2,298 genome copies, Additional file 9) revealed that 40 % of these TE copies exhibited elevated substitution rates that were particular to which nucleotide was 3′ to the cytosine (Methods, Additional file 10). Eighty per cent of TE copies with high frequencies of cytosine mutation showed a bias toward CpG dinucleotides, consistent with the "CpG effects" [48] of maintenance methylation known in eukaryotes [49–51], including fungi [45, 52] where CpG methylation of TEs have been shown in ascomycete and basidiomycete fungi [53]. Notably, the rate of CpG mutations varied according to TE order and superfamily examined: TE copies exhibiting a pattern of frequent transition at CpG sites appeared lower for class II DNA transposons (*Helitron*-type and TIR elements) than that of class I Retrotransposons (21 % and 58 % respectively) (Additional file 10). In addition, the GC content in TEs (54.4 %) was very close to GC content in genes (56.0 %) and genome (55.4 %).

In addition, the *M. lychnidis-dioicae* genome was found to contain the core RNAi machinery components, that may act to constrain proliferation of transposable elements, contrasting with some other Basidiomycetes that have lost this pathway [54]. The RNAi pathway components were identified based on similarity to known RNAi genes in fungi and validated by examining predicted functional domains (Methods). The genome of *M. lychnidis-dioicae* was also found to contain one copy of a RNA-dependent RNA polymerase (MVLG_02137), two copies of Argonaute (MVLG_06823 and MVLG_06899), and one copy of Dicer (MVLG_01202). All of these components of RNAi machinery were expressed under each of the conditions examined, suggesting the pathway is active across life cycle states, although one copy of Argonaute (MVLG_06823) was significantly more highly expressed (corrected p-value <0.01) during infection compared to nutrient limited and rich agar media.

Mating-type locus and chromosome

The central proteins involved in mating-type determination in basidiomycetes were found in *M. lychnidis-dioicae*. Orthologs of the two-component homeodomain transcription factor that functions in post-mating compatibility, HD1 (MVLG_07149) and HD2 (MVLG_07150), were assembled in a 14.1 kb region; as previously described HD1 and HD2 are adjacent and divergently transcribed in *M. lychnidis-dioicae* [55], similar to other fungi [56]. Both the mating pheromone receptor [57] and the homeodomain compatibility factor identified above are located at the ends of their respective supercontigs, such that the genomic proximity of these two essential mating-type-determining loci is unclear. The chromosomes bearing these two mating-type loci show suppressed recombination across most of their length, with only two small recombining regions at their ends, *i.e.,* pseudo-autosomal regions (PARs) [11, 12, 33, 55, 58]. The assembly scaffolds corresponding to the non-recombining regions (NRRs) and to the PARs were identified based on alignment to an optical map of the mating-type chromosomes [12, 33] and by performing additional sequencing of gel purified chromosomes (Methods). A total of 449 genes mapped to the a_1 mating-type chromosome, including 350 genes found on the NRRs and 99 genes found on the PARs (Table 1, Additional file 11). Other than the genes for the pheromone receptor, the homeodomain transcription factors, and the STE20 protein kinase, no other genes on the mating-type chromosome have a predicted function linked to mating in other systems.

Consistent with the expectation in regions of suppressed recombination (*e.g.*, [59]), the TE density (Fig. 2a) was several fold higher in the non-recombining region (NRR) of the a_1 mating-type chromosome (41 %) relative to the autosomes (9 %), confirming prior studies of a TE accumulation on the mating-type chromosomes as a whole [11, 12, 33]. The pseudoautosomal regions (PARs, supercontigs 37 and 43) displayed a TE content (~13 %) more similar to the estimate for autosomal regions than to the NRR of the mating-type chromosome. Gene density in the NRR of the mating-type chromosome (23 %), estimated as CDS density, was less than half the gene density of the autosomes (49 %) (Fig. 2b). The number of genes predicted per 10 kb positions also indicated a lower density in the NRR of the mating-type chromosome (1.7 genes) than in the autosomal partition (2.9 genes) (Fig. 2c). As with TE content, the PARs displayed gene density values (~40 % for CDS density and ~2.5 for genes per 10 kb) closer to the autosomal estimates than the NRR of the mating-type chromosome. The NRR of the mating-type chromosomes contained genes of the same distribution, with the exception of 2-fold elevated density of small secreted proteins (SSPs) and, in particular, a 5-fold enrichment of Cys-rich SSPs compared to the autosomes.

With regard to base pair composition, again the NRR exhibited a pattern distinct from the PARs or autosomes in GC content. The GC content in protein coding genes, irrespective of codon position, were similar among the NRR of the mating-type chromosome, PAR, and autosomes (Fig. 2d), as were the contents represented by TEs (Fig. 2e). However, in other sequences, representing

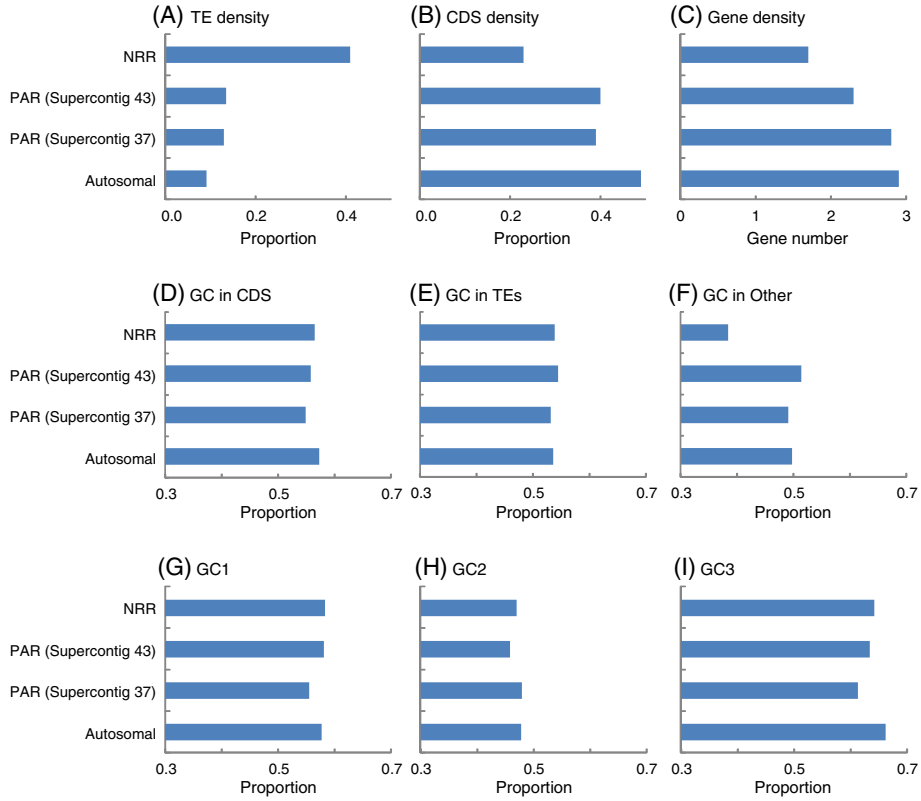


Fig. 2 Comparisons of sequence characteristics in non-recombining regions (NRR) of the mating-type chromosome, pseudoautosomal regions (PAR), and autosomes. The genomic regions are shown, with results for the two supercontigs ("37" and "43") corresponding to the two PARs presented separately. **a** Transposable element (TE) density is shown as the total length of TE sequences over the total length of DNA analyzed. **b, c** Gene density is shown as the proportion of the total length of coding region (CDS) over the total length of DNA analyzed, and number of putative genes identified per 10000 nucleotides, respectively. **d–f** Proportion GC base pair contents are shown for CDS, TEs, and for the remaining, predominantly intergenic regions. **g–i** Proportion GC base pair content for protein-coding genes are shown relative to first-, second-, and third-codon positions ("GC1", "GC2" and "GC3," respectively)

inter-genic regions not consisting of TEs, the NRR displayed markedly reduced GC content (Fig. 2f) while PARs and autosomes had similar GC levels. GC content of codon positions within protein-coding genes did not vary among the NRR, PARs, and autosomes (Fig. 2g–i), notably in that the third-codon position patterns was not reflective of intergenic GC composition variation across regions. This pattern as well as the lower GC content observed for the second-codon position compared to first or third positions is consistent with some prior research [60]. The correspondence analysis of codon usage (Additional file 12) among the non-recombining region of the mating-type chromosome, PAR and autosomal did not indicate obvious differences in codon usage.

Gene conservation and lineage-specific changes

The comparison of 7,364 predicted proteins of *M. lychnidis-dioicae* to those of diverse basidiomycetes revealed gene loss and gain patterns relevant in terms of the growth and pathogenesis of this organism. We included representatives of the three subphyla of basidiomycetes (Pucciniomycotina, Agaricomycotina, and Ustilaginomycotina), as well as three Ascomycota species as outgroups (Fig. 3, Additional file 13). Within the Pucciniomycotina, species compared included *M. lychnidis-dioicae* and two closely related Microbotryomycetes (*Sporobolomyces roseus* and *Rhodotorula glutinis*) and three other more distantly related species; within this group the two other plant pathogens (*P. graminis* f. sp. *tritici* and *M. larici-populina*) are biotrophic, like *M. lychnidis-dioicae*. These comparisons revealed 2,451 *Microbotryum* gene clusters representing 2,613 genes that were broadly conserved in the Basidiomycota (present in at least 13 of the examined 15 other basidiomycete genomes). The gene families specific to Pucciniomycotina, with orthologs present only in *M. lychnidis-dioicae* and/or the other Pucciniomycotina species, (Fig. 3, orange boxes) was composed of a small set of 224 predicted proteins from *M. lychnidis-dioicae*; the two rusts (*P. graminis* and *M. larici-populina*) share a larger number of Pucciniomycotina-specific proteins in part due

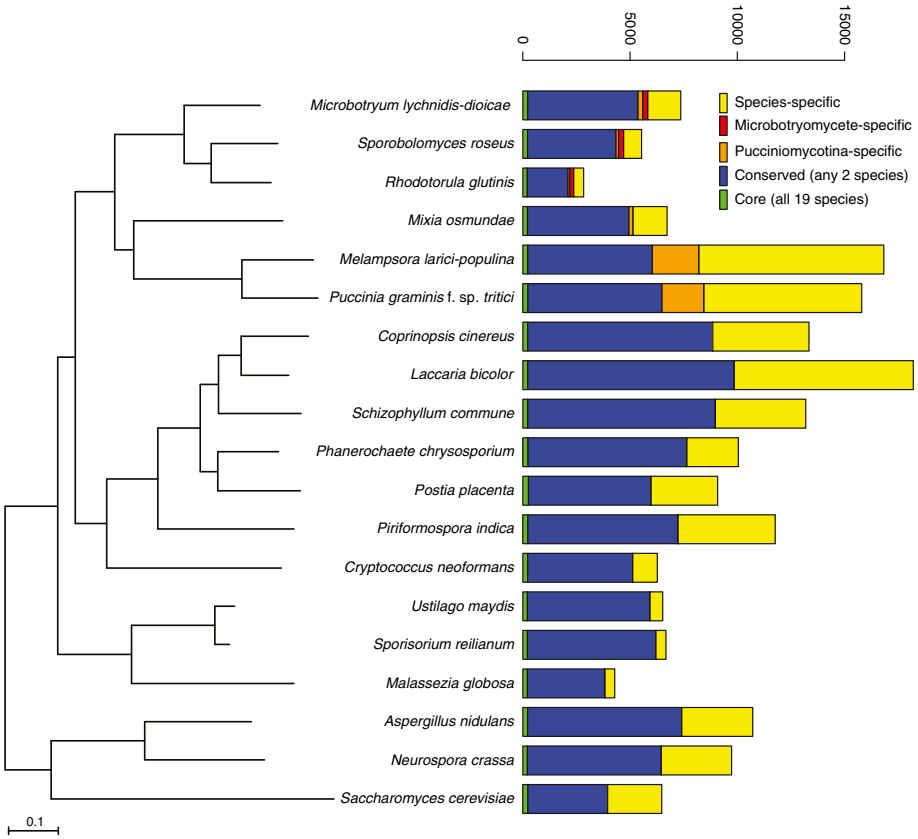


Fig. 3 Phylogenetic relationship and gene conservation of *Microbotryum lychnidis-dioicae* and 18 compared fungi. Phylogeny (left panel) is based on concatenated MUSCLE alignments of 80 single copy genes. Species phylogeny was inferred using RAxML (PROTCATWAG model) with 1,000 bootstrap replicates; all nodes were supported by at least 99 % of replicates. Ortholog conservation (right panel) highlights genes conserved in all species (core, green), genes conserved in at least two species (conserved, blue), genes unique to the Pucciniomycotina or Microbotryomycetes (orange and red, respectively) and genes unique to a given species (species-specific, yellow)

to their expanded genome size. Examining proteins conserved across the Microbotryomycetes, orthologs of 4,844 *M. lychnidis-dioicae* proteins were present in at least one other species, and of these 233 were specific to the Microbotryomycetes. A set of 190 gene duplications occurring specifically along the *M. lychnidis-dioicae* lineage were identified using phylogenetic analysis (Additional file 14, Additional file 15, phylomedb.org); these did not display functional enrichment for gene ontology (GO) term assignments, as GO terms were only assigned for 19 of the 190 genes. While most genes (70 %) were shared with occurrence in at least one other species, the remaining set of 1,534 genes appear specific to *M. lychnidis-dioicae*.

The identification of enriched or depleted PFAM domains for *M. lychnidis-dioicae* compared to other fungi revealed significant differences in functional categories between these genomes. A total of six protein domains were significantly enriched or depleted (q-value < 0.05) in *M. lychnidis-dioicae* compared to the other basidiomycetes examined (Table 3, Additional file 16). Enriched domains include secretory lipases, and two domains of unknown function, DUF23 (glycosyl transferase 92) and DUF1034 (Fn3-like). DUF23 (PF01697) was present in five copies in *M. lychnidis-dioicae* and was only present otherwise in *Mixia osmundae* and *S. roseus* in our comparison; three of these five genes were mapped to the GT2 CAZY family expanded in *M. lychnidis-dioicae* (see below). Both *M. lychnidis-dioicae* and the rusts contain a large number of proteins with the DUF1034 domain; 8 of the 10 *M. lychnidis-dioicae* proteins with a DUF1034 domain also contained a subtilase family protease domain. Phylogenetic analysis of proteins with the DUF1034 domain suggested that independent gene family expansions occurred in different species; the rusts formed a separate clade from *Microbotryum* that is further subdivided, mostly along species lines (Additional file 17). Domains depleted in *M. lychnidis-dioicae* relative to other basidiomycetes were also identified, including Cytochrome p450, NACHT, and F-box domains (Table 3, Additional file 16). A more narrow comparison

Table 3 Expanded or depleted PFAM domains in *Microbotryum lychnidis-dioicae*

PFAM domain	*M. lychnis-dioicae*	*S. roseus*	*R. glutinis*	*M. osmundae*	*M. larici-populina*	*P. graminis-tritici*	*S. reilianum*	*U. maydis*	*M. globosa*	Agaricomycetes[c] (7)	Basidiomycete comparison[a]		Pucciniales comparison[b]	
											p-value	*q*-value	*p*-value	*q*-value
PF00067.15 Cytochrome P450	10	7	5	14	29	17	15	20	6	95	1.43E-11	6.49E-08	1.28E-02	1
PF05729.5 NACHT	1	2	1	1	1	1	1	1	0	44	6.20E-08	1.41E-04	4.51E-01	1
PF01697.20 Glycosyltransferase family 92	5	1	0	1	0	0	0	0	0	0	1.65E-05	2.50E-02	1.99E-02	1
PF06280.5 Fn3-like (DUF1034)	10	0	0	0	5	9	1	1	0	1	2.54E-05	2.88E-02	1.00E + 00	1
PF00646.26 F-box	7	18	12	12	9	11	11	10	3	43	5.20E-05	4.33E-02	2.72E-02	1
PF03583.7 Secretory lipase	7	0	0	0	0	0	3	2	6	0	5.72E-05	4.33E-02	1.76E-03	3.07E-01
PF00734.11 CBM1 Fungal cellulose binding	0	0	3	0	0	0	0	0	0	21	1.21E-04	7.10E-02	6.60E-02	1
PF02816.11 Alpha kinase	0	0	0	0	79	39	0	0	0	5	1.25E-04	7.10E-02	2.80E-27	1.27E-23
PF07690.9 MFS1 Major Facilitator Superfamily	119	110	26	64	90	72	104	98	30	129	4.11E-02	1	2.06E-08	4.67E-05
PF01753.11 zf-MYND finger	7	11	26	3	7	4	2	2	1	41	4.61E-04	2.33E-01	5.16E-08	7.82E-05
PF01083.15 Cutinase	0	0	1	4	21	9	3	4	0	2	7.90E-02	1	6.49E-07	5.89E-04
PF01670.9 Glycosyl hydrolase family 12	0	0	0	10	14	3	0	0	0	1	1.70E-01	1	1.08E-06	6.56E-04
PF11327.1 DUF3129	0	0	0	4	13	10	0	0	0	1	2.62E-01	1	1.08E-06	6.56E-04
PF00097.18 Zinc finger, C3HC4 type	23	11	9	26	22	93	28	24	12	24	6.82E-01	1	1.16E-06	6.56E-04
PF00080.13 Copper/zinc superoxide dismutase	0	0	0	2	6	18	0	0	0	1	2.62E-01	1	1.94E-06	9.79E-04
PF00098.16 Zinc knuckle	10	6	1	6	11	59	8	7	5	11	7.63E-01	1	6.31E-06	2.87E-03
PF06609.6 Fungal trichothecene efflux pump	16	16	5	6	6	3	8	9	2	12	4.10E-02	1	1.18E-05	4.86E-03
PF12013.1 DUF3505	0	0	0	1	4	16	0	10	0	0	2.65E-01	1	2.15E-05	8.14E-03
PF03101.8 FAR1 DNA-binding domain	0	0	0	1	14	4	1	1	0	1	4.09E-01	1	6.65E-05	2.32E-02
PF00083.17 Sugar (and other) transporter	54	56	13	31	46	34	56	46	14	61	3.28E-01	1	1.71E-04	5.54E-02
PF07738.6 Sad1/UNC-like	2	1	2	3	24	7	2	1	2	2	5.87E-01	1	2.45E-04	7.42E-02

[a]M. lychnis-dioicae compared to all other Basidomycetes; [b] Microbotryales (M. lychnis-dioicae, S. roseus, R. glutinis) compared to other Pucciniales (M. larici-populina, P. graminis tritici, M. osmundae); [c]Agaricomycetes represent average of the 7 species in this group; see Additional file 11 for counts per species

of the three Microbotryomycetes to the other species within the Pucciniomycotina identified additional expansions and depletions common to the species in this lineage. A fungal trichothecene efflux pump (PF06609) gene family was enriched in *M. lychnidis-dioicae* and *S. roseus*, with 16 copies in each genome (Table 3). By contrast, the alpha-kinase family that is highly expanded in the rusts [39] was absent in the Microbotryomycetes. The cutinase domain shows a similar conservation pattern; while multiple cutinase genes are found in other biotrophic pathogens, no copies were detected in *M. lychnidis-dioicae*. At lower levels of significance, the CBM1 cellulose binding domain was detected as absent from all species in the Pucciniomycotina with the exception of *R. glutinis* (Table 3, Additional file 16). More specific analysis of these enriched and depleted domains is presented below.

The expansion of the secretory lipases appeared specific to *M. lychnidis-dioicae* within the Pucciniomycotina (Table 3). Among all other basidiomycete genomes compared, secretory lipases are also highly represented in *Malassezia globosa* (Ustilaginomycotina), a skin fungus associated with human dandruff and dependent on its host for lipids. *Malassezia globosa* has an additional gene family expansion associated with lipid acquisition; this species has 6 copies of phospholipase C, whereas *M. lychnidis-dioicae* contains only a single phospholipase C protein. Unlike *M. globosa*, *M. lychnidis-dioicae* does not depend on lipids for growth, and contains a predicted fatty acid synthase (MVLG_04698). The secretory lipase family is also present at lower copy number in the two Ustilaginomycotina corn smuts, *Sporisorium reilianum* and *Ustilago maydis*, of which the latter responds to lipids, including corn oils, as part of a developmental switch [61]. A phylogenetic analysis of the secretory lipases in this comparison revealed that the *M. lychnidis-dioicae* lipases have undergone a lineage specific expansion, as in *M. globosa* (Fig. 3a). Most lipases were predicted to be secreted including three of the seven *M. lychnidis-dioicae* lipases. However four of the seven genes appeared partial based on alignment of the protein sequences; the missing 5′end from two genes deleted the region containing a secretion signal in paralogous copies. Further refinement of the assembly or transcripts is needed to identify the full length version of these genes or confirm if they are perhaps pseudogenes (see below), and establish their relative location in the genome.

Previous work in *M. lychnidis-dioicae* has shown that mating mixtures of haploid cells produce hyphae in response to phytols and to tocopherols [62]. We therefore exposed haploid sporidial cells (p1A1 or p1A2 strains) or mated cells (p1A1 and p1A2 mixed together) to various lipids or oils, including commercially available corn oil, (+/−)-α-tocopherol, and phytol. Each stimulated filamentous growth of mated mixtures (Fig. 4b–e), but had no observable effect on haploid cells (unmated). The observation that mated cells also respond to corn oil supports the hypothesis that lipid response may be important for the development of this species. This is possible, as lipids are likely present on the host meristem [63]. Moreover, we find that at least three lipase genes are differentially expressed when exposed to phytols (a constituent of chlorphyll), compared to similarly treated cells in the absence of phytols (see below; Additional file 18).

An important group of proteins identified as being expanded or depleted in *M. lychnidis-dioicae* are those predicted to be involved in cell wall modifications, both in terms of fungal cell walls, but also as might affect host plants. Families of structurally-related carbohydrate catalytic and carbohydrate-binding modules (or functional domains) are described in the CAZy database (www.cazy.org) [64]. Such enzymes break down, modify, or build glycosidic bonds. The assignment of the predicted proteins derived from a genome to CAZy families helps to shed light on the particular glycobiological features of an organism [65]. A total of 236 *M. lychnidis-dioicae* protein models were mapped to protein families in the CAZy database based on sequence conservation (percentage identity over CAZy domain length) (Additional file 19). The CAZy family profile of *M. lychnidis-dioicae* was then compared to that recently published for 33 basidiomycetes [66], in order to identify expanded and reduced families (Table 4, Additional file 20).

With 98 candidate glycosyltransferases (GTs), *M. lychnidis-dioicae* has more than any of the 33 basidiomycetes used in the comparison (average = 69.5; min = 41; max = 95). Both *M. lychnidis-dioicae* and *Puccinia graminis* contain a candidate fucosyltransferase, which is not present in the other basidiomycetes surveyed. This is compatible with the known presence of fucose in the cell wall of *Microbotryum* [67, 68]. Two other families expanded in *M. lychnidis-dioicae* include alpha-mannosyltransferases (GT32 and GT62), suggesting that the cell wall includes a larger fraction of alpha-mannan than in other species. In fungi, the synthesized cell wall carbohydrates are frequently remodeled by the action of dedicated glycoside hydrolases and transglycosidases that are found in distinct CAZy glycosyl hydrolase (GH) families. A notable feature of *M. lychnidis-dioicae* is that it has a reduced number β-1,3-glucan cleaving or modifying enzymes of families (GH16, GH72, GH81, and GH128).

Examination of the other GH families and of the other categories involved in carbohydrate breakdown (PL, CE, AA and ancillary CBM) revealed that *M. lychnidis-dioicae* completely lacked cellulases (no GH6, GH7, GH8, GH9, GH12, GH44, GH45, nor any GH5 with highest sequence similarity to characterized cellulases) (Table 4, Additional file 20). In addition no cellulose-targeting lytic

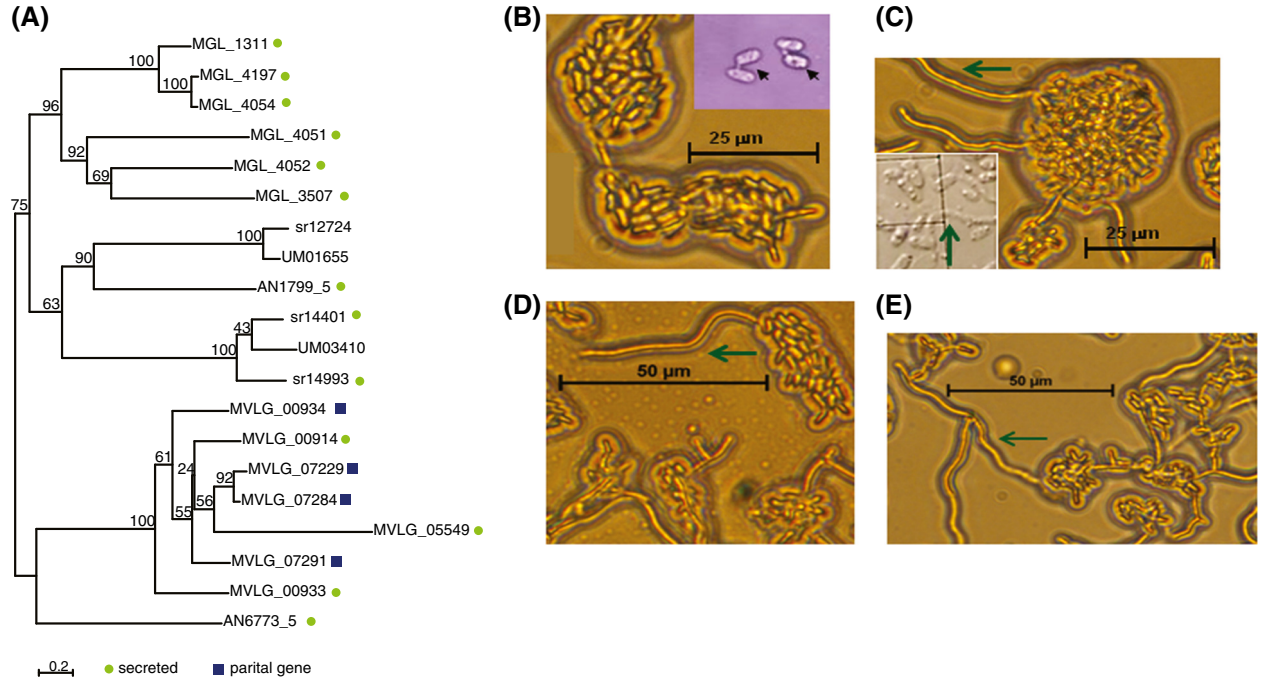


Fig. 4 Expansion of lipase gene family in *Microbotryum lychnidis-dioicae* and *Malassezia globosa*. Both *M. lychnidis-dioicae* and *M. globosa* contain a higher number of proteins with a lipase domain (PF03583.7) relative to the other fungi examined. **a** Phylogenetic tree of proteins with this domain. Protein sequences were aligned with MUSCLE, and a phylogeny was inferred from this alignment using RAxML (PROTCAT model, DAYHOFF matrix). A total of 1000 bootstrap replicates were performed, and the percent of replicates shown is on the tree nodes. All *M. lychnidis-dioicae* proteins are indicated by their respective MVLG designation. Similar lipase genes from other species are included, with the gene prefix denoting the species as follows: AN, *Aspergillus nidulans*; UM, *Ustilago maydis*; sr, *Sporisorium relianum*; MGL, *Malasezzia globosa*. Scale corresponds to substitutions per site. **b–e** Treatment of mated *M. lychnidis-dioicae* cells with various lipids. **b** water-treated mated cell control (Inset: higher magnification showing cells with conjugation bridge, black arrows); (**c**) mated cells treated with commercially-available corn oil (Inset: higher magnification of filamentation); D, mated cells treated with α-tocopherol; (**e**), mated cells treated with phytol. The green arrows show areas of filamentation emanating from mated cells after treatment with each specific type of lipid. **b** and **c** size bars, 25 µm; (**d**) and (**e**), size bars, 50 µm

polysaccharide monooxygenase of family AA9, nor any broad specificity β-glucanase of family GH131, nor any cellulose-binding module of family CBM1 could be found. This clearly shows that *M. lychnidis-dioicae* does not interact with nor digests cellulose during its interaction with plants, a finding confirmed by the failure of *M. lychnidis-dioicae* to grow on cellulose as a sole carbon source (Additional file 21). *Microbotryum lychnidis-dioicae* also completely lacks xylanase (no GH10, GH11, GH30), xyloglucanase (no GH74), and the enzymes for the cleavage of side-chains of xylan, xyloglucan and rhamnogalacturonan (no GH29, GH95, GH51, GH115), indicating that these cell wall polymers are not a carbon source for the fungus (Table 4, Additional file 20). Consistent with this prediction, *M. lychnidis-dioicae* failed to grow on xylan as a sole carbon source (Additional file 21).

Microbotryum lychnidis-dioicae was able to grow on pectin as a sole carbon source (Additional file 21),

Table 4 Selected CAZY expansions and depletions

	GT total	GH total	Beta glucan modification[a]	Cellulose related[b]	Xylan related[c]	GH26 beta-mannanases	Pectin/pectate lyases
Microbotryum lychnidis-dioicae	98	82	4	0	0	4	0
Average basidiomycete[d]	69.5	179.7	28.9	37.3	14.0	0.5	2.9
Piriformospora indica	65	194	31	98	41	1	12
Ustilago maydis	58	100	26	4	8	1	1
Puccinia graminis f. tritici	81	154	11	15	5	4	4
Melampsora laricis-populina	84	169	14	23	9	3	4

[a]GH16,GH72,GH81,GH128. [b]GH6, GH7, GH8, GH9, GH12, GH44, GH45, GH131, AA9, and CBM1. [c]GH10, GH11, GH30, GH29, GH95, GH51, GH115, GH74. [d]Average count for 33 basidiomycete genomes; see Additional file 20

although it does not break down pectin by the action of pectin/pectate lyases, as these enzymes are also absent from the genome (Table 4, Additional file 20). Instead, the genome harbors a suite of six family GH28 enzymes, which cleave polygalacturonic acid after its methylester groups have been removed by the action of six family CE8 pectin methylesterases. This CE8 family is present at high numbers in the two rust fungi and *M. lychnidis-dioicae*; the copy number amplification in *M. lychnidis-dioicae* appears to be due to tandem duplication, with one array of two genes and a second array of four genes. Four of the six CE8 copies have a predicted secretion signal, supporting a potential role in interacting with the host plant. Compared to 33 other basidiomycetes, *M. lychnidis-dioicae* stands out in having a significant expansion of its enzymatic arsenal for the breakdown of β-mannan, a polysaccharide present throughout plants but more abundant in flowers, siliques and stems [69]. *Microbotryum lychnidis-dioicae* encodes four candidate β-mannanases of family GH26 and a comparison with biochemically characterized enzymes shows that 10 out of its 19 GH5 enzymes also target β-mannan (the other *M. lychnidis-dioicae* GH5 enzymes target β-1,3-glucans (5 proteins), glucocerebrosides (3 proteins) and β-1,6-glucan (1 protein)). This β-mannan digestion arsenal is augmented by the presence of a GH2 enzyme, which shows a strong relatedness to characterized β-mannosidases.

Homogalacturonan is a major component (60 %) of plant pectin and the degradation pathway is required in several stages of plant development. One of these stages is anther dehiscence when pollen grains are released; this process requires pectinesterases and polygalacturonases. As indicated above, a total of six CE8 family pectin methylesterases are found in *M. lychnidis-dioicae*, of which four are predicted secreted proteins (MVLG_02682, 04072, 04073, 4074). Part two of the pathway requires polygalacturonase; a total of six *M. lychnidis-dioicae* proteins contain the polygalacturonase GH28 (PF00295) domain, of which MVLG_02498 is highly induced (over 1,000 fold) in MI-late. The homogalactorunan degradation pathway of *M. lychnidis-dioicae* may thus perform a similar role as pollen in anther dehiscence when the flowers bloom, since during teliospore formation of *M. lychnidis-dioicae*, host pollen is no longer available to perform that function.

Multiple classes of transporters are expanded in *M. lychnidis-dioicae* (Table 3), enabling uptake of diverse substrates. Major facilitator transporters, sugar transporters, and the fungal trichothecene efflux pump are present at high copy number relative to other Pucciniomycotina. A fungal trichothecene efflux pump, *TRI12*, was first described in *Fusarium sporotrichioides* as part of the gene cluster involved in trichothecene biosynthesis [70]; trichothecenes are a group of mycotoxins produced by various species of fungi. As the TRI12 domain is present at high copy number in *M. lychnidis-dioicae* and other Basidiomycetes are not known to produce trichothecenes, this suggests that this domain may have a role in transporting other small molecules.

Sugar transporters also play an important role in virulence of biotrophic plant pathogens, such as *Ustilago maydis* and several species of rust fungi. Specifically, a plasma membrane-localized sucrose transporter (Srt1) in *U. maydis* facilitated direct utilization of sucrose, thus eluding the plant defense mechanism [71]. The HeXose Transporter 1 (Hxt1) gene in the rust fungus *Uromyces fabae* is localized to haustoria to take advantage of that structure for sugar uptake [72]. The *M. lychnidis-dioicae* genome contains a total of 26 potential sugar transporters, with multiple high identity matches to Srt1 and Hxt1, which may fulfill similar roles.

One additional contrast between *M. lychnidis-dioicae* and the two plant pathogenic rust fungi examined suggests a difference in the relative importance of response to superoxides. Reactive oxygen species (ROS), including superoxides or H_2O_2 produced by the host plant, are a canonical part of the defense response to pathogens. *Microbotryum lychnidis-dioicae* is depleted in domains for Peroxidase (PF01328) and copper/zinc superoxide dismutase (PF00080); the two other Microbotryomycetes also lack proteins with these domains. Despite the reduced repertoire of such predicted proteins in *M. lychnidis-dioicae* relative to the rust fungi, six proteins (MVLG_00980, MVLG_03089, MVLG_03931, MVLG_02439, MVLG_03568, and MVLG_04684) were identified as containing peroxidase 2 (PF01328), peroxidase (PF00141), redoxin (PF08534), or Glutathione peroxidase (PF00255) domains. Of these predicted proteins, only MVLG_03089 was differentially expressed under the conditions examined and was up-regulated in MI-late relative to growth *in vitro* in rich medium. In addition, four predicted proteins with iron/manganese superoxide dismutase domains (PF02777 and PF00081), glutaredoxin (PF00462) or catalase (PF00199) domains were found (MVLG_00659, MVLG_06630, MVLG_06939, MVLG_04131). Finally, pathway analysis via MetaCyc predictions (http://fungicyc.broadinstitute.org/) suggests that *M. lychnidis-dioicae* contains components of the glutathione-mediated detoxification pathway: Glutathione transferase (EC 2.5.1.18: MVLG_05985, MVLG_04790) and membrane alanyl aminopeptidase (EC 3.4.11.2: MVLG_03673). However, there appears to be a missing component (3.4.19.9) in this pathway to facilitate formation of an intermediate of a glutathione-toxin conjugate. Biochemical and functional analyses will be required to determine the importance of these predicted enzymes in the ability of the pathogen to survive and flourish in its host.

Secreted proteins (SP) and candidate effectors

A total of 279 secreted proteins (SPs) were predicted in *M. lychnidis-dioicae* and their expression and conservation examined to identify candidates for interacting with the host (Table 5, Additional file 22). Among the 71 SPs that were smaller than 250 amino acids (small secreted proteins, SSPs), 46 were species specific in our comparative set and further do not share sequence similarity (e-value <1e-3) with any protein in the NCBI protein database. SSPs indeed often appear species-specific, likely because they co-evolve rapidly in an arms race with their hosts [43]. Notably, 48 of the SSPs were significantly up-regulated during plant infection (MI-late compared to rich media), but were not differentially expressed when comparing expression on rich and nutrient limited agar (Table 5), suggesting that these SSPs may play a specific role during plant infection.

Several cysteine-rich multigene families were identified among predicted secreted proteins. In some cases these families include tandemly duplicated genes; the MVLG_04105 family contains 4 members predicted to be SSPs (MVLG_04105, 04106, 04107, and 04096), three of which are adjacent in the genome on the mating-type chromosome (see below). Although these proteins lack PFAM domains, two of these are induced during infection. An additional family of Cys-rich proteins with nine members has a subset of six clustered in the genome (MVLG_05513, MVLG_05514, MVLG_05515, MVLG_05533, MVLG_05534, MVLG_05538). Seven of the nine proteins in this family were predicted to be secreted, yet their expression was highly variable, with two up-regulated in nutrient limited conditions and two down-regulated during infection. A small subset of Cysteine-rich proteins contains known protein domains. Two proteins (MVLG_02283 and MVLG_02288) contain the Cysteine-rich secretory protein family domain (PF00188). In addition, a total of 9 proteins contain the fungal-specific Cysteine rich CFEM domain (PF05730). All of these CFEM proteins were predicted to be secreted; four of these were significantly induced and three were repressed in MI-late relative to rich and nutrient limited agar.

To identify genes that could provide a mechanism for linking flower development to fungal development, we compared expression of the predicted secreted proteins with the *S. latifolia* EST library produced from flowers [73]. A total of 37 genes share sequence similarity with *S. latifolia* ESTs; ten secreted proteins matched plant ESTs with at least an e-value of e-19. One *S. latifolia* EST (09F02) showed similarity (BlastX, e-value < 7e-24) with two *M. lychnidis-dioicae* proteins (MVLG_02043 and MVLG_02936); the best match, MVLG_02043, encodes a predicted secreted gamma-glutamyltranspeptidase (GGT). Another EST (33C05) shared sequence similarity with two *M. lychnidis-dioicae* proteins (MVLG_00083, MVLG_02276); these in turn share similarity with the expansin family of "ripening related" proteins and were down-regulated during either growth on nutrient limited agar or late in infection (MI-late). The precise function of plant expansins is poorly defined at the molecular level, and the predicted function of the similar fungal proteins is even less well established. However, one such expansin-related protein in *Laccaria bicolor* was recently found to be expressed specifically in the extracellular matrix (ECM) of symbiotic tissues and localized within the fungal cell wall [74].

Table 5 Properties of predicted secreted proteins

Protein length	SP count	Induced in MI late	Repressed in MI late	Induced in water	Repressed in water	FPKM > 1	Highly expressed only in MI late
100	15	4	0	0	0	11	3
150	23	5	0	1	0	20	4
200	15	5	0	2	1	14	
250	18	5	1	1	0	14	3
400	62	8	10	8	3	59	2
500	56	8	7	5	2	55	
600	33	4	2	7	0	33	
700	23	7	1	0	0	23	
800	7	0	2	1	0	7	
900	6	1	0	0	0	6	1
1000	13	2	2	0	1	13	
2000	6	0	1	1	0	6	
3000	2	0	0	0	0	2	
Total	279	49	26	26	7	263	

Two cysteine-rich secreted proteins (MVLG_02288 and MVLG_02283) matched a *S. latifolia* flower EST annotated as having similarity with the plant PR-1 class of pathogenesis related proteins (PRs); these are proteins defined as encoded by the host plant but induced only in pathological or related situations (possibly of non-pathogenic origin). To be included among the PRs, a protein must be induced upon infection but not necessarily in all pathological conditions [75]. Another *S. latifolia* gene of interest, *SLM2,* is expressed in the stamens of smut infected flowers but not uninfected flowers [22]; four *M. lychnidis-dioicae* proteins showed blast similarity with e-value less than e-6. All of the hits were hypothetical proteins that contained the SRF-type transcription factor domain (PF00319), and three of the four were predicted to be targeted to the nucleus (MVLG_04297, MVLG_06278, and MVLG_07052).

Response to oxidative environments

Laccase-like multi-copper oxidase proteins are capable of degrading phenolic compounds like polymeric lignin and humic substances [76] and may be involved in the interacton of fungal pathogens with their host plants. Four proteins in *M. lychnidis-dioicae* contain the three multi-copper oxidase (MCO) domains (PF07731, PF07732, PF00394). Two MCO proteins were predicted to be secreted, and another MCO was predicted to have a GPI-anchor to the membrane. The fourth MCO was a membrane-anchored protein (MVLG_03092), with the N-terminus of the polypeptide outside the cell.

The glyoxal oxidase catalyses the oxidation of aldehydes to carboxylic acid and is an essential component of the extracellular lignin degradation pathway of the root rot fungus, *Phanerochaete chrysosporium*. Of a total of seven *M. lychnidis-dioicae* proteins that contain the glyoxal oxidase N-terminal domain (PF07250), two are predicted to be secreted and four have a predicted GPI-anchor. While two of these glyoxal oxidase genes are adjacent in the genome, they do not form a gene cluster with any MCO as observed at the lignin peroxidase gene cluster in *P. chrysosporium*.

Genes similar to plant hormone synthesis genes

Since *M. lychnidis-dioicae* infection of female *S. latifolia* hosts can alter normal flower development so as to produce pseudomale flowers, we investigated whether the genome might contain genes for pathways that could be associated with such changes. The predicted *M. lychnidis-dioicae* protein database was examined for components of biosynthesis pathways of eight plant hormones (Additional file 23), as well as for other signaling pathways that could have an impact on host gene expression or development. Based on sequence similarity to components for these pathways from plants and microbes, as well as additional confirmation of some complete pathways using MetaCyc predictions (http://fungicyc.broadinstitute.org/), we found evidence that *M. lychnidis-dioicae* encodes enzymes that could participate in hormone biosynthetic pathways, such as polyamine biosynthesis pathways, which produces compounds known to play developmental and stress-response roles in plant physiology [77] (see proposed model in Fig. 5). If these enzymes are, in fact, promoting hormone biosynthesis pathways, one possible explanation is that precursors for these pathways are provided by the plant. Alternatively, the potential components of these pathways we have identified are used by the fungus for functions other than manipulating host development. For example, cytokinin degradation could be mediated by a predicted FAD-oxidase (MVLG_04134), if this enzyme can function as a cytokinin dehydrogenase (EC 1.5.99.12); however, this predicted protein most closely resembles other fungal D-lactate dehydrogenases based on sequence similarity. Similarly, a predicted 2β-dioxygenase (EC1.14.11.13) (MVLG_00840) could be involved in gibberellin inactivation via hydroxylation, although this protein falls more generally into the 2OG-Fe(II) oxygenase superfamily.

Another potential set of pathways for plant signaling involves the production of glycerol lipids, such as diacylglycerol (DAG) or triacylgycerol (TAG). Many studies have demonstrated the importance of compounds like DAG in mammalian signaling [78], and DAG is important in pollen tube elongation in some plant species [79]. Such pathways often involve the action of phospholipase C; a predicted phospholipase C gene (MLVG_07108) and potential phospholipase A (MVLG_03207, MVLG_03384, MVLG_4789) and D (MLVG_01917, MVLG_03610) homologues are found in the genome. Therefore, the organism contains the proteins necessary for synthesizing 1,2-diacylglycerol, 1,2-diacyl-*sn*-glycerol-3-phosphate, 2-lysophosphatidylcholine, and 1-lysophosphatidylcholine. Based on MetaCyc prediction, *M. lychnidis-dioicae* possesses the requisite components of the CDP-diacylglycerol biosynthesis I pathway. In addition, a number of the secretory lipases whose PFAM domain is enriched in *M. lychnidis-dioicae* (PF03583, see above) are implicated in these phospholipase pathways.

Gene expression changes during infection and in response to lipids

To focus on genes potentially important for infection and interaction with the host, we identified genes whose expression was altered in MI-late. A total of 1,254 genes were differentially expressed in MI-late compared to either rich or nutrient limited agar; of these, a common set of 307 genes were induced in MI-late and 126 were repressed in MI-late compared to both other conditions (corrected p-value <0.001, Additional file 24, Methods). Of the 138 genes in both comparisons with a predicted

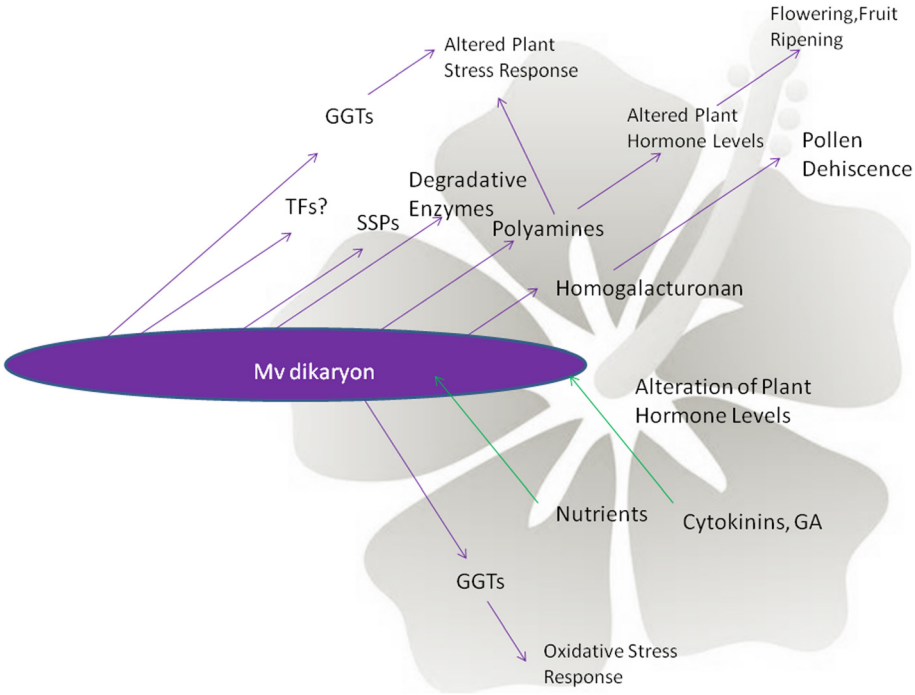


Fig. 5 Model of *Microbotryum lychnidis-dioicae* interactions with its host. The potential pathways identified in *M. lychnidis-dioicae* based on inventory of the genome were used to predict products potentially secreted or taken up by the fungus that could affect host development (see text for more detailed description). GA, gibberellic acid

PFAM domain, transporter domains were most frequently observed; a total of 20 of the 138 functionally assigned proteins corresponded to transporters, with MFS, sugar, OPT, and amino transporters each represented by two or more genes (Additional file 25). Carbohydrate active enzymes, kinases and transcription factors were also highly represented in these MI-late induced genes.

Several different classes of transporters were transcriptionally induced during infection, potentially promoting the uptake of small molecules and nutrients from the host plant. Significant enrichments include domains found in MFS transporters (Fisher's exact test, corrected p-value < 0.001) and sugar transporters ($p < 0.005$), the largest classes of transporters in *M. lychnidis-dioicae* and many fungi. The oligopeptide transporter protein was also enriched at a lower level of significance ($p < 0.1$) and four predicted proteins with this domain (MVLG_03106, MVLG_07217, MVLG_03161, and MVLG_00149) were up-regulated in MI-late. An MFS domain-containing protein of note up-regulated in MI-late was a nitrite transporter (TIGR00886.2; MVLG_00642); this gene is linked to two other genes associated with nitrate assimilation, a nitrite reductase (MVLG_00638) and a nitrate reductase (MVLG_00637). All three genes involved in nitrate assimilation were significantly up-regulated during infection.

In evaluating the expression of the 279 proteins predicted to be secreted, 48 were induced *in planta*. Several cysteine-rich secreted proteins were induced during infection. A pair of linked cysteine rich small secreted proteins (MVLG_04106 and MVLG_04107) was induced during infection. Four other proteins (MVLG_00115, MVLG_00802, MVLG_00815, and MVLG_00859) with the Cys-rich CFEM domain-containing family (PF05730) were significantly induced in infection. These proteins are good candidates for being potential effectors, based on proteins with similar properties that have been shown to be effectors in other systems [80].

Cell wall degrading enzymes may play a role in the infection, particularly during the stage where the fungus causes necrosis of host plant tissue. A total of nine glycoside hydrolases were up-regulated in MI-late; among GH families, GH28 polygalacturonase proteins are mostly highly enriched during infection ($p < 0.08$). Polygalacturonase is required for the second part of the homogalacturonan pathway implicated in pollen dehiscence. MVLG proteins that contain a glyoxal oxidase domain (PF07250) were also significantly enriched ($p < 0.009$) in genes induced during MI-late.

Mated cells of *M. lychnidis-dioicae* respond to corn oil, in addition to the other lipids previously reported [45]. This supports the hypothesis that lipid response may be important for the development of this species.

Although we observed no alteration of phenotype for un-mated haploid cells treated with lipids, three predicted cytoplasmic proteins with a secretory lipase domain (MVLG_07229, 07284, and 07291) were highly induced in haploids grown on nutrient limited agar compared to rich media or MI-late infections. To validate the expression levels from RNA-Seq data, relative levels for lipase genes were measured for mated and unmated cells, grown in nutrient limited, rich media, or treated with phytol by qRT-PCR (Additional file 18). Notably, when mated haploid cells were treated with phytol, after 12 h treatment, two genes (MVLG_00914, MVLG_05549) displayed substantial up-regulation, while the remaining three either increased slightly or decreased (Additional file 18). This could reflect priming of cells that are ready to mate for plant cues that would ultimately lead to stable dikaryon formation after successful mating.

Discussion

This analysis highlights genomic features of *M. lychnidis-dioicae* that reflect its particular lifecycle. *Microbotryum lychnidis-dioicae* grows as a biotroph, similar to the rust fungi, for most of the plant infection cycle. Notably in its capacity as a castrating pathogen, *M. lychnidis-dioicae* also causes necrosis that appears to be limited to developing flowers. Our analysis revealed that gene content contains a different profile than purely biotrophic or necrotrophic plant pathogens. This included a global number of CAZymes that is larger than other biotrophic fungi (Additional file 20); however, at the same time it shows complete loss of many CAZyme families that target the plant cell wall. This is consistent with a primarily intercellular colonization pattern of host apical meristems [81], though it raises questions about how appressorium penetration is accomplished.

Necrotic growth stages may be enabled by the subtilases, laccases, and copper radical oxidases, which are ligninolytic enzymes [82–84]; subtilases, in particular may play a key role in regulating the activities of laccase [76]. By contrast, certain CAZymes present in the biotrophic rust fungi are absent in *M. lychnidis-dioicae*; in particular the absence of cutinases suggests that penetration of the plant surface by *M. lychnidis-dioicae* does not require cutin degradation. This may reflect the fact that the normal portal of entry for infection is via the flower, a tissue that poses less of a barrier to the fungus. Of the plant polysaccharides that constitute a carbon source for many fungi, *M. lychnidis-dioicae* has lost the ability to digest cellulose, xylan, xyloglucan, and the highly substituted forms of pectin (rhamnogalacturonan). Retention of enzymes that breakdown polygalacturonic acid and β-mannan, components of pollen tubes and flowers, respectively, illustrates the high degree of specialization of this fungus.

Whereas smut fungi generally cause necrosis as a space-making process during sporulation [85], *Microbotryum* anther smuts are more aptly characterized as a growth-altering parasite [86]. At the stage of anther development, *M. lychnidis-dioicae* causes abortion of pollen production, and replacement by the diploid teliospores for dispersal by pollinator species *[*87*]*. Moreover, in the female, atrophy of pistils occurs during infection. There are a number of candidate pathways that might participate in pollen tube elongation or in blocking this process. Aspartyl proteases are involved in pollen tube elongation or prevention by *Metschnikowia reukaufii* [88]; seven candidate aspartyl proteases are predicted in *M. lychnidis-dioicae*. Additionally, consistent with transcriptional up-regulation of some component enzymes for the homogalactorunan degradation pathway of *M. lychnidis-dioicae*, this pathway may take over the role of pollen in anther dehiscence when the flowers bloom, since during teliospore formation of *M. lychnidis-dioicae*, host pollen is no longer available to perform that function Fig. 5.

Finally, since *M. lychnidis-dioicae* appears to lack a large repertoire for dealing with host-generated defenses that utilize reactive oxygen species (ROS) one could expect that the interaction with its host normally does not elicit such host responses or the fungus actively down-regulates them. Paradoxically, *M. lychnidis-dioicae* also appears to secrete some proteins that would serve to bring on this plant response by serving as prooxidants (*e.g.*, secreted gamma-glutamyltranspeptidases). Further elucidation of the roles of the predicted peroxidases superoxide dismutases, and glutathione-mediated detoxification pathway components will require functional analyses to evaluate their biological importance, if any, in the interaction of the pathogen with its hosts.

Secretory lipases represent one of the most significantly expanded gene families in *M. lychnidis-dioicae* compared to the other fungi examined. While haploid cells show no outward phenotype alteration in response to lipids like phytol, several secretory lipases of *M. lychnidis-dioicae* in haploid cells are up-regulated on nutrient limited agar. We propose that regulation of secretory lipases primes haploid cells so that, after mating, they can respond to the appropriate plant-derived cues (including lipids) to progress to the next developmental stage, stable dikaryotic hyphae. In fact, most of the secretory lipases we investigated were up-regulated in mated cells and when such cells were exposed to phytol (Additional file 18).

The identification of genes induced during infection (MI-late) suggests their involvement in host invasion and evasion of physical and chemical defense systems of the

plant. A role for CAZymes, including pectin methylesterases and GHs, during plant infection in *M. lychnidis-dioicae* is further supported by increased transcription in the MI-late sample. Analysis of the gene expression profile of both wheat stem and poplar rust (*P. graminis* and *M. larici-populina*) also found that many CAZyme genes related to cell wall degradation were up-regulated during plant infection [39]. In addition to the CAZymes, *M. lychnidis-dioicae* shows significant induction of diverse transporters during plant infection, which may be critical for uptake of small molecules during biotrophic growth.

As in other plant pathogenic fungi, candidate effectors in *M. lychnidis-dioicae* were predicted based on predicted localization, expression during infection, and sequence conservation. Notably, SSPs are located closer to TEs than other protein coding genes, suggesting that this could impact SSP expression or duplication. TEs probably play a role in the expansion of such a family. Indeed they contribute to genome rearrangements and gene duplications [89]. In *Fusarium oxysporum* f. sp. *lycopersici,* effector genes are present on chromosomes or regions enriched for DNA transposons [90]. Some secreted proteins are predicted to act on host cell walls and proteins, either for the remodeling of host development in the flower or for acquisition of additional nutrients by the fungus.

In assembling sequence of the a_1 mating-type chromosome, we characterized how the content of this allosome differs from autosomal regions, contrasting the non-recombining regions of the mating-type chromosome with the pseudo-autosomal regions (PARs) capable of recombination and with autosomes. Both the lower gene density and higher transposable element content in the non-recombining region of the mating-type chromosome relative to autosomal regions are consistent with a reduced efficiency of purifying selection due to the suppression of recombination, as occur on non-recombining sex chromosomes [59]. The two recombining PARs of the mating-type chromosome displayed TE content and gene density more similar to autosomes than the non-recombining part of the mating-type chromosome. By contrast, we observed no difference in codon usage nor in the GC content at third codon positions between autosomes and the mating-type chromosome, though this has been observed in the non-recombining part of the fungal mating-type chromosome of *Neurospora tetrasperma* [31]. Overall the maintenance of homologous meiotic pairing and recombination in PAR regions may render them more similar to autosomes than allosomes with respect to evolutionary forces of selection and drift. However, their physical linkage to the non-recombining region of mating-type chromosomes suggests intermediate modes of evolution [91].

Conclusion

Altogether, this study provides an in-depth genomic portrait of a fungal castrating, biotrophic plant pathogen reflecting its unique life cycle. In particular, the unique absence of enzyme classes for plant cell wall degradation and maintenance of enzymes that break down components of pollen tubes and flowers provides a striking example of biotrophic host adaptation. In addition, while there are fewer enzymes to digest cellulose, xylan, xyloglucan, and highly substituted forms of pectin, as well as proteins that could protect the fungus from oxidative stress, the repertoire of predicted cell wall modifying enzymes and those that could manipulate host development has expanded (see model in Fig. 5). Given the place of *M. lychnidis-dioicae* in a large species-complex with a vast host species pool, the insights from this genomic and transcriptomic analysis combined with comparative approaches with other members of the *Microbotryum* species complex will be most informative on the evolutionary processes involved in a radiation and specialization on a wide array of plant species from different genera.

Methods

Microbotryum lychnidis-dioicae lineage(s) and *Silene latifolia* host(s)

The *focal* lineage of *M. lychnidis-dioicae* for this work is the most studied in the context of disease ecology ("Lamole strain": GenBank I00-15Lamole.1; [9, 11]) and belongs to the recently-refined species designation *M. lychnidis-dioicae*, parasitizing *Silene latifolia*. From the original isolate, haploid sporidial strains were generated via micromanipulation of the meiotic products from a single tetrad, yielding the strains Lamole p1A1 and p1A2 that differ in electrophoretic karyotypes only in the mating-specific chromosome. For the work in this report, the haploid p1A1 strain (mating-type *a1*) was used as the source of the *focal* genome. Additionally, the *focal* genome contains size-heteromorphic sex chromosomes that share many features with sex chromosomes in plant and animal systems [11]. The corresponding a_2 strain, p1A2 was used together with its partner strain p1A1 in plant infections and in RNA-Seq analysis.

High molecular weight DNA preparation

Microbotryum lychnidis-dioicae Lamole p1A1 was grown on yeast peptone dextrose media (YPD; 1 % yeast extract, 10 % dextrose, 2 % peptone, 1.5 % agar) at room temperature for 5 days and ultimately extracted using a phenol chloroform isoamyl extraction method [92]. Harvested fungal cells were ground into fine powder using liquid nitrogen and resuspended in OmniPrep Genomic Lysis Buffer (G-Biosciences, cat no: 786–136) according to manufacturer's recommended tissue to reagent ratio.

The sample was heated in a 55-60 °C water bath for 15 min after extensive vortexing. Chloroform was added to the sample after allowing it to cool to room temperature. Using wide bore tips thereafter, 3–4 extractions using phenol chloroform isoamyl (25:24:1) solution were performed, followed by a final extraction with chloroform isoamyl (24:1) solution. Nucleic acid was then precipitated and the pellet was rinsed twice with ice-cold 70 % ethanol and then air-dried. The pellet was rehydrated using Tris-EDTA buffer (pH 8.0) (100 μl per 100 mg of ground tissue powder used) and treated with RNase (*Longlife* RNase, 5 mg/ml; G-Biosciences); 1 μl of RNase was added for every 100 μl of TE buffer used.

RNA isolation

Haploid cells

Haploid fungal cells of either Lamole p1A1 or p1A2 grown separately under rich conditions for 5 days on yeast peptone dextrose media (YPD; 1 % yeast extract, 10 % dextrose, 2 % peptone, 2 % agar) at room temperature were harvested for RNA extraction. RNAs were checked for quality using a Bioanalyzer (Agilent). The RNAs were then pooled in equal quantity (in terms of mass) based on the Bioanalyzer quantification. The same procedure was also performed for the haploid cells grown separately on 2 % water agar for 2 days, to compare the gene expression when haploid cells were subjected to nutrient free environment without the mating partner. Again, haploid cell samples, p1A1 and p1A2, were extracted as independent samples, and then mixed in equal proportion for RNA sequencing.

Host plant infection for RNA-Seq (MI-late stage)

Silene latifolia seeds (harvested in Summer 2009 from Lamole, Italy) were sterilized and hydrated by soaking them in a sterilizing solution (40 % household bleach, 20 % absolute ethanol and 1 drop of Triton X-100 as surfactant per 50 ml of solution) and washing five times in sterile distilled water, for 2 min per wash with constant agitation. Each seed was then individually planted in closed milk jars on sterile 0.3 % phytagar (Life Technologies), 0.5× MS (Murashige and Skoog) salts (Sigma-Aldrich) and 0.05 % MES (2-(N-morpholino)ethanesulfonic acid) buffer (Brand). Each jar was placed at 4 °C for 5 days to synchronize germination. The jars were then transferred to a 20 °C growth chamber with 13 h of fluorescent light daily. Germination starts within 3 days with the appearance of the radicle. When the seedlings were 15 days old, they were transplanted into 2″ square pots filled with Sunshine MVP Professional Growing Mix (Sun Gro Horticulture Canada Ltd, cat no. 02392868) soil and replaced into the growth chamber. Humidity was kept high initially using dome covers and flood trays. Seedlings were gradually exposed to chamber environment for increasing amounts of time daily in order for the seedling to harden and adapt to the lower humidity. The plants were transplanted to 4″ round pots when they began to bolt at about 30 days old. They were further transplanted into 7″ round pots when they had almost attained maximum height or when the volume of soil was not sufficient to provide hydration requirement for the plant. The plants were watered every other day with 100-ppm fertilizer (Peters Professional® 15-16-17 Peat-Lite Special, Formula no: S12893).

To infect the host plants, mated cells were prepared as follows. Haploid cells grown on rich media were harvested and resuspended in distilled water, adjusted to a concentration of 1×10^9 cells/ml in equal proportion before being spotted onto nutrient-free solid agar media (2 % agar) in 50 μl spots. The plates were allowed to dry and then incubated at 14 °C for about 48 h. Cells were inspected for conjugation tubes under the microscope and then 5 μl of 1×10^6 cells/ml resuspended in distilled water with anionic surfactant was pipetted onto the floral meristem when the cotyledon was fully developed (11–12 days). Infection was determined by the consistent blooming of fully smutted flowers. The floral buds were staged according to previous literature [35] under a dissecting scope (Nikon, Model: SMZ-U) and parts of the floral buds were measured with a glass stage micrometer (Imaging Research, Inc.).

Tissue originating from host plants was collected in RNAlater RNA stabilizing reagent (QIAGEN, cat no: 76106) and left at 4 °C overnight until sufficient tissue had been collected for the RNA extraction. The solution was removed before storing the sample at −80 °C. For infected male plants, we collected floral tissue from buds ranging in size from 4 mm to fully open smutted flowers. These tissue samples were pooled to yield the source for 'MI-late' RNA used in RNA-Seq analysis. Thus, they provide a pooled average picture of gene expression for this size range of infected tissue.

All RNA samples were extracted using the RNeasy Plant Mini Kit (QIAGEN, cat no: 74904) according to the manufacturer's instructions. DNase treatment was performed using Ambion's TURBO DNA-free (Applied Biosystems, cat no: AM1907), also according to manufacturer's instruction. For quality assessment before Illumina sequencing, 5 μg of DNase-treated RNA was reverse transcribed with SuperScript III First Strand Synthesis System for RT-PCR (Life Technologies, cat no: 18080–051). PCR was performed using TaKaRa Ex Taq Hot-Start DNA Polymerase (Takara, cat. no: RR001B) using 25 μl reaction volume. To check for DNA contamination and possible inhibitory substances in the RNA, three sets of housekeeping primers (Eurofins/MWG/Operon) were used. Amplification of a region of the *S. latifolia* partial *wdr1x* gene for a putative WD-repeat protein (GenBank IDs Y18519, Aj310656) was used to assess host cDNA and contaminating genomic DNA; the forward primer, 5′- CTCTG

CTGGAGGTGGAACAT-3′ and reverse primer, 5′- AG CACTGAACACCCCAACTT-3′; in this case a 253 bp fragment would be produced for cDNA, vs. a 335 bp fragment for genomic DNA. Targeting the *M. lychnidis-dioicae mepA* gene, we used as forward primer, 5′- CT TTTGCGTAGGAAGAATGC-3′ and as reverse primer, 5′- AGCACTGAACACCCCAACTT-3′; this combination yielded a 532 bp fragment from cDNA, compared with a 1039 bp fragment from genomic DNA. The other primer combination targeted the *M. lychnidis-dioicae* beta-tubulin gene, with forward primer, 5′- CGGACACCGT TGTCGAGCCT -3′, and reverse primer, 5′- TGAGGT CGCCGTGAGTCGGT-3′, yielding a 150 bp fragment from cDNA compared with a 215 bp fragment from genomic DNA. The PCR program was 30 s at 94 °C, 30 s at 60 °C and 1 min at 72 °C for 35 cycles. RNA quality was also evaluated using an Agilent BioAnalyzer; all samples had RNA integrity number scores of at least 7.8, indicating highly intact RNA.

Treatment of cells with lipids

Haploid fungal cells of Lamole p1A1 and p1A2 were grown separately under rich conditions for 5 days on YPD at room temperature, then harvested into sterile distilled water. The concentration was adjusted and resuspended in equal proportions in each type of medium, to achieve a final concentration of 1×10^9 cells/ml.

To allow better solubility of the lipids, 50 % ethanol was used as the solvent for the lipids. We used 1 % corn oil (Carlini), (±)-α-tocopherol (Sigma-Aldrich, cat no: T3251-5G), and phytol (Sigma-Aldrich, cat no: P3647), dissolved in the solvent and used as the resuspension media for the fungal cells. The mixtures were then spotted in 50 μl spots onto 2 % water agar and allowed to mate for 2 days at 14 °C. The cells were then observed under the microscope for conjugation tubes and filamentous structures. The solvent served as the control media to ensure that changes in phenotype were not due to the ethanol present.

Genome and transcriptome sequencing, assembly, and annotation

For genome sequencing, we constructed three libraries (Additional file 1) with different insert sizes and sequenced each using 454 Technology. The reads were assembled with Newbler (version MapAsmResearch-04/19/2010-patch-08/17/2010). The total assembly size of 26.1 Mb in scaffolds includes 99.6 % of bases of at least Q40 quality; gaps encompass 3.45 % of the total scaffold length.

For RNA-Seq, we purified polyA RNA and constructed a strand-specific library for each sample as previously described [93, 94] and sequenced each with Illumina technology generating 76 base paired reads. Across the three libraries, 96 % of reads met the Illumina Passing Filter (PF) quality threshold. Read alignment rates to the genome varied between three libraries; for the rich and nutrient limited samples, 90 % or 89 % of reads aligned respectively; for the MI-late sample only 23 % of reads aligned. This was expected as these samples also contain the host *Silene* RNAs. To assemble transcripts for use in annotation, RNA-Seq reads were aligned to the assembly with Blat, and then assembled using Inchworm [95] in the genome-guided mode.

To predict genes, we first generated a high confidence training set of 775 transcripts of at least 900 nt using Genemark [96] and the assembled RNA-Seq data. This was used to train Augustus [97] and Glimmerhmm [98]. RNA-Seq data was processed by PASA [99] to generate longer transcripts, and ORFs of at least 600 nt were predicted. Available ESTs from Genbank and Microbase were also utilized. EVM [99] was then used to select a preliminary gene set from the *ab initio* gene calls (Augustus, Genemark, GlimmerHmm, and SNAP), Genewise [100], ESTs, PASA ORFs, and the training set, with the highest weight given to the RNA-Seq based PASA ORFs. The EVM gene set was compared to the PASA ORFs, and non-repeat genes found only in the PASA set were added to the gene set. Finally, PASA was run on the final gene set to all updates of all gene structures with the RNA-Seq and incorporate alternatively spliced transcripts. Genes likely corresponding to repeats were filtered out using TransposonPSI (requiring 1e-10 and 30 % overlap), PFAM domains, Blast similarity to repetitive elements and 7 or more Blast hits to other genes in the set. Genes with flagged features (proteins ≤50 aa, internal exons ≤6 nt, introns ≤20 nt, introns ≥1500 nt, exons spanning gaps in the assembly, internal stop codons, overlapping other coding sequences, overlapping ncRNAs (tRNAs, rRNA, or other)) were manually reviewed and corrected where supported by the evidence. Gene names were assigned with the locus prefix MVLG.

The completeness of the gene set was evaluated by examining the conservation and completeness of core eukaryotic genes (CEGs, [36]). We compared the gene set of *M. lychnidis-dioicae* and of the 18 other fungal genomes used in comparisons to the CEGMA set, and identified Blast hits above and below the recommended 70 % coverage threshold (Additional file 3). A tool for streamlined analysis and visualization of conservation of CEGs is available on SourceForge (http://sourceforge.net/projects/corealyze/).

Differential expression analysis

We used differential expression analysis scripts in the Trinity pipeline [95, 101] to process RNA-Seq data generated from the three conditions (nutrient limited, rich, and MI-late). Briefly, we first extracted protein coding

gene sequences from the *M. lychnidis-dioicae* genome sequence based on coordinates of gene models, and added 100 bases of flanking sequence on each side to approximate UTRs. Then the RNA-Seq reads from each of the three samples were aligned to the extracted coding sequences using bowtie [102]. The alignment files were used to quantify transcript abundances by RSEM [103]. Differential gene expression analysis was conducted using edgeR with TMM normalization [104, 105] using a corrected p-value [106] cutoff of 1e-3. In comparing all pairs of the three conditions, a total of 1,413 genes were differentially expressed across the comparisons (Additional file 24).

TE detection and annotation

Two pipelines from REPET package (http://urgi.versailles.inra.fr/tools/REPET) were run on the *M. lychnidis-dioicae* contigs. The TEdenovo pipeline [107] was used to search for repeats in the genome. The first step uses Blaster with the following parameters [identity > 90 %, HSP (High Scoring segments Pairs) length >100b & <20Kb, *e*-value ≤ 1e-300]. HSPs found were clustered by 3 different methods: Piler [108], Grouper [109] and Recon [110]. Multiple alignments (MAP) of 20 longest members of each cluster (918 clusters) containing at least 3 members were used to derive a consensus. Consensus sequences were then classified based on their structure and similarities against Repbase Update (v15.11) [111] before removing redundancy (Blaster + Matcher). Consensus sequences without any known structure or similarity were classified as "Unknown".

The library of 425 classified consensuses provided by the TEdenovo pipeline was used to annotate TE copies in the whole genome using TEannot pipeline [109]. Annotation is based on 3 methods (Blaster, Censor, RepeatMasker). HSPs provided were filtered and combined. Three methods (TRF, Mreps and RepeatMasker) were also used to annotate SSR. TE duplicates and SSR were then removed. Finally a "long join procedure" [107] was used to address the problem of nested TEs. This procedure finds and connects fragments of TEs interrupted by other TEs inserted more recently to build a TE copy. The nesting patterns of such insertion must respect the three constraints: fragments must be co-linear (both on the genome and the same TE consensus reference), have the same age and separated by younger TE insertion. The identity percentage with the reference consensus is used to estimate the age of a copy. Using results of this first TEannot pipeline, we filtered out 111 consensus sequences without full-length copy in the genome. A copy may be built using one or more fragments joined by the TEannot long join procedure. We ran a second TEannot using the 306 consensus elements remaining after filtering out TE consensus without any full-length copy.

We also used gene prediction and proceeded to manual curation in order to improve TE annotation. We removed TE copies of consensuses that were identified as host genes. Indeed, these consensuses built from family of repeats containing at least 3 members and classified as unknown by the TEdenovo pipeline has been predicted as host genes belonging to multigenic families. We also filtered out TE copies not satisfying the criteria (identity > 0.8 & length > 150 & identity*length > 150) and those corresponding to low complexity region of the consensus Mivi-B-R219-Map20_classI-LINE-incomp very highly represented in the genome included in predicted genes. The few copies just over these thresholds were manually removed, depending on their location in genes and evidence of the gene (PFAM domain not related to TEs).

Search for signature of transition type (C to T) mutation bias

We performed pairwise alignments between each copy and respective consensus to finally provide multiple alignments for each family (consensus) using in-house scripts. TE copies with less than 80 % of identity with consensus and smaller than 400 bp were filtered out. We also filtered out TE families with less than 5 sequences in the multiple alignments. RIPCAL [112] was run on each multiple alignment to count both potential single mutations (transitions and transversions) and di-nucleotide target used in all possible transitions. Results were analysed using in-house R scripts to select most reliable mutated copies (if transition rate > 2 * transversion rate). For 40 % of copies exhibiting a transition mutation bias (of 2298 total copies, 179 consensus families (Additional file 10)), we considered that dinucleotide targets ($CA + TG^1$, $CC + GG^1$, $CG + CG^1$, $CT + AG^1$; ^{1} for reverse complement), were preferentially used if they represent a minimum of 30 % of the addition of the four possible. We expect 25 % of each if they are equiprobable.

Measurement of distance between genes and TEs

We computed the distance from each gene to the closest TE (case 1), or from each TE to the closest gene (case 2) using getDistance.py from S-MART package [113]. Only distances up to 10 kb were considered. For case 1, we also compared the subset of genes encoding predicted secreted proteins with the set of all other genes for different classes of distance intervals. For the case 2, we compared different TE categories in two classes of distance intervals (<1kbp and >1kbp). A chi^2 test of homogeneity (Pearson's chi-squared) was computed to test that the observed difference between the sets did not occur by chance (p-value < 0.05). The graphics and statistical test were performed using in-house R-Scripts.

Identification of the mating-type chromosome supercontigs

Using the same haploid genotype from which the whole genome was sequenced, DNA enriched for the a_1 mating-type chromosome was isolated from agarose gels after pulsed-field electrophoresis. With this technique, the isolated bands could include of small amounts of autosomal fragments that co-migrate with mating-type chromosomes, though these preparations have been shown to be strongly enriched for mating-type chromosome DNA [12]. The isolated DNA was amplified by whole genome amplification (REPLI-g kit, QIAGEN). The DNA was sequenced using 2- and 5 kb-insert size mate-paired libraries and 454 technology version Titanium (www.roche.com). Assembly of non-duplicated reads and excluding autosomal contamination yielded ~20-fold coverage.

The assembly was compared to the a_1 whole haploid genome sequence using NUCmer (http://mummer.sourceforge.net/) to validate assemblies and identify scaffolds corresponding to the mating-type chromosome in the whole genome assembly. The regions corresponding to autosomal contamination were identified by uneven and low read coverage and were excluded from further analyses; mitochondrial DNA was also excluded (GenBank NC_020353). All whole-genome scaffolds with more than 20-fold depth of the enriched mating-type chromosome sequence were confidently assigned to the mating-type chromosomes (Additional file 11). It was not possible to anchor the scaffolds onto the available optical map of the a_1 mating-type chromosome [12] due to the small sizes of the contigs relative to spacing of the restriction enzyme cut sites in the map. Annotation for mapped supercontigs regions was parsed from the genome-level annotation.

TE content was assessed using de novo TE annotation as described above. TE content was compared between the nonrecombining part of the mating-type chromosome (Table 1), the PARs, and the autosomes. GC content was compared between the non-recombining region of the mating-type chromosome, the PARs and the autosomes. Significance of the mean difference was assessed using a *t*-test and a nonparametric Wilcoxon test. In a second step, the GC content at the 3rd codon positions was inspected separately in identified coding regions. Mean GC contents at the 3rd codon positions were compared between the CDS on the non-recombining part of the mating-type chromosome, the PARs and the autosomal coding regions. All GC content analyses were conducted using in-house python and R scripts.

Prediction of the secretome

To predict a high confidence set of secreted proteins, results from several different software tools were integrated. These include TargetP1.1 [114], SignalP3.0, [115], SignalP4.0 (http://www.cbs.dtu.dk/services/SignalP/) [116], TMHMM2.0 [117], PredGPI [118], Phobius [119], NucPred [120], Prosite [121], and WoLF PSORT [122]. A subset of these tools were used to first exclude proteins as not secreted if they had transmembrane domains (two or more, from TMHMM or Phobius), an ER retention signal (0.00014 from Prosite), GPI anchor (specificity of >99.5 % using the general model of PredGPI), or nuclear localization (>0.8 threshold in NucPred). Secreted proteins were then predicted based on passing four of the six thresholds examined (TargetP secreted localization, SignalP3.0 NN Dscore > 0.43, SignalP3.0 HMM Sprob > 0.8, SignalP4.0 D-score > 0.45, WoLFPsort 'Extr' listed as major neighbor, or Phobius secreted localization).

Additional criteria were used for ambiguous predictions. If both TMHMM and Phobius agreed on the existence of 1–2 transmembrane (TM) domains in the protein, the protein was excluded from the probable secretome pool. If a protein was predicted to have a lowly probable GPI linkage (PredGPI specificity >99.0 % and <99.5 %) and a TM predicted by TMHMM and/or Phobius around the same region, this served as corroborating evidence for GPI anchorage to the membrane.

Where evidence conflicted or was insufficient for determining secretome status, BLASTp and Pfam domains were used to establish probable orthologs, followed by referencing the UniProtKB [123, 124] and FunSecKB [125] database for confirmation of localization of the orthologs, where available. Out of 7,360 predicted proteins, 6,899 proteins were excluded from the pool of the secretome based on the criteria described above. Of the remainder, 189 proteins had no contradictory calls in the positive prediction for SP. Another 272 went through further confirmation, of which 182 of these were confirmed to be non-SP and 63 were SP. Of the rest, 27 of them could not be finalized due to lack of ortholog matches in the NCBI database and lack of conserved domain for reference.

Gene clustering and comparative analysis

We compared *M. lychnidis-dioicae* to 18 other fungi (Additional file 13) that sample the three subphyla in Basidiomycota, including 5 other Pucciniomycotina, 7 Agaricomycotina, 3 Ustilaginomycotina, as well as 3 Ascomycota outgroups. For *M. lychnidis-dioicae* and the 18 other fungal genomes, we identified ortholog clusters using OrthoMCL [126] version 1.4 with a Markov inflation index of 1.5 and a maximum e-value of 1×10^{-5}. Two genomes, *R. glutinis* and *P. placenta*, are missing more broadly conserved orthologs than the other genomes; examining the 961 *Microbotryum* gene clusters with an ortholog missing in just one other genome, the number of missing clusters in any one Basidiomycete genome ranged from 1 to 34 with the exception of *R. glutinis* and *P. placenta*, missing 410 and 393 of these

highly conserved clusters, respectively. PFAM domains within each gene were identified using Hmmer3 [127], and gene ontology terms were assigned using BLAST2GO [128].

To examine gene duplication history, the phylome, or complete collection of phylogenetic trees for each gene in a genome, was reconstructed for *Microbotryum lychnidis-dioicae* and 19 other fungi, including those used for OrthoMCL (Additional file 13) and *Serpula lacrymans*. Phylomes were reconstructed using the previously described pipeline [129]. All trees and alignments have been deposited in PhylomeDB [129] and can be browsed online (www.phylomedb.org, phylome code 180). Trees were scanned to detect and date duplication events [130].

RNAi components from other other fungi were used as Blast queries to find homologs in *M. lychnidis-dioicae*; the queries used include *U. hordei* RdRp (CCF48827.1), *C. neoformans* Ago1 (XP_003194007), and *N. crassa* Dcl2 (Q75CC1.3) and Dcl1 (Q758J7.1). The putative function was confirmed by examining protein domains. The identified domains for each protein include: Piwi, PAZ and DUF1785 found in both copies of Argonaute (MVLG_06823, MVLG_06899); DEAD/DEAH helicase, double-stranded RNA binding, and RNAseIII (MVLG_01202). Sugar transporters were identified based on homology to the *Ustillago maydis* Srt1t transporter (Genbank: XP_758521) and the *Uromyces viciae-fabae* Hxt1 (Genbank: CAC41332).

The *M. lychnidis-dioicae* protein models corresponding to carbohydrate-active enzymes were assigned to families of glycoside hydrolases (GH), polysaccharide lyases (PL), carbohydrate esterases (CE), carbohydrate-binding modules (CBM), auxiliary activities (AA) and glycosyltransferases (GT) listed by the CAZy database [64], exactly as previously done for the analyses of dozens of fungal genomes [39, 66, 131, 132].

Data access

The assembly and annotation of *M. lychnidis-dioicae* was submitted to GenBank under accession number AEIJ01000000.

Additional files

Additional file 1: is a table providing Genome sequencing statistics.

Additional file 2: is a table showing RNA-Seq read statistics.

Additional file 3: is a figure showing Conservation of core eukaryotic (CEGMA) genes.

Additional file 4: is a figure Correlation between GC content and gene density.

Additional file 5: is a figure presenting Preferred codons for the different amino-acids.

Additional file 6: is a figure of Frequency of transposable element classes.

Additional file 7: is a figure displaying Proximity of TE copies for main orders (Class I LTR and LINE, Class II TIR and Helitrons) to the closest gene.

Additional file 8: is a figure presenting Comparison of TE proximity for secreted proteins compared to all other genes.

Additional file 9: is a table that shows TE elements used in RIPCAL analysis.

Additional file 10: is a figure that shows Frequency of mutation transition types in different TE classes.

Additional file 11: is a table presenting Genes predicted on non-recombining mating type regions, pseudoautosomal regions ("PAR") and autosomes.

Additional file 12: is a figure presenting Codon usage in autosomal and mating-type-specific genome regions.

Additional file 13: is a table showing Source of genome data used in comparative analysis.

Additional file 14: is a table that shows Gene family duplications from Phylome analysis.

Additional file 15: is a figure displaying Phylome for *M. lychnidis-dioicae*.

Additional file 16: is a figure of the Phylogenetic tree of DUF1034 proteins.

Additional file 17: is a table with Species counts of enriched or depleted protein domains.

Additional file 18: is a table that presents qRT-PCR validation of secretory lipase expression.

Additional file 19: is a table with CAZymes content comparisons.

Additional file 20: is a table that shows Phylogenetic comparison of CAZyme content.

Additional file 21: is a figure showing growth of *M. lychnidis-dioicae* on different sole carbon sources, including dextrose, cellulose, xylan, and pectin [133].

Additional file 22: is a table of the Full set of predicted secreted proteins.

Additional file 23: is a table showing Sequences similar to plant hormone-related genes.

Additional file 24: is a table with Genes differentially expressed between nutrient limited, rich, and MI-late samples.

Additional file 25: is a table of PFAM domains enriched in MI-late induced genes.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

MHP, DJS and CAC designed the research project. MHP, SY, EF, ZC, and CAC performed assembly and mapping. JG and QZ annotated the genome. MHP, JP, PW, EP, MEH, ZC, SST, and CAC analyzed the genomic sequences and transcriptome data. SST, SD, JA, EF, EP, JG, HB, BH, MHP, GA, TGE and CAC performed data analyses. SST, JMA, and DR conducted experimental analysis of secretory lipases and validation via qRT-PCR of RNA-Seq predictions. CAC, MHP, SST, JA, EF, TG, SD, BH and MEH wrote the paper. All authors approved the final version.

Acknowledgements

We acknowledge the Broad Institute Sequencing Platform for generating all DNA and RNA sequence described here, and Sinéad Chapman for coordinating the sequencing. We thank Mark Lawrence for sharing the *Rhodotorula glutinis* ATCC 204091 genome and annotation methods prior to publication. This project was supported by NSF award #0947963 to MHP, DJS, and CAC, the ANR-09-BLAN-064 and ERC GenomeFun 309403 grants to TG, and BIO2012-37161 and NPRP 5-298-3-086 to TGE. We would also like to thank Vincent Lombard, Elodie Drula and Anthony Levasseur for their help with the day-to-day development of the CAZy database.

Author details
[1]Department of Biology, Program on Disease Evolution, University of Louisville, Louisville, KY 40292, USA. [2]Institut National de la Recherche Agronomique (INRA), Unité de Recherche Génomique Info (URGI), Versailles, France. [3]Institut National de la Recherche Agronomique (INRA), Biologie et gestion des risques en agriculture (BIOGER), Thiverval-Grignon, France. [4]Ecologie, Systématique et Evolution, Bâtiment 360, Université Paris-Sud, F-91405 Orsay, France. [5]CNRS, F-91405 Orsay, France. [6]Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA. [7]INRA, UMR 1136, Interactions Arbres-Microorganismes, Champenoux, France. [8]UMR 1136, Université de Lorraine, Interactions Arbres-Microorganismes, Vandoeuvre-lès-Nancy, France. [9]Centre National de la Recherche Scientifique (CNRS), UMR7257, Université Aix-Marseille, 13288 Marseille, France. [10]Department of Biological Sciences, King Abdulaziz University, Jeddah, Saudi Arabia. [11]Centre for Genomic Regulation (CRG), Barcelona, Spain. [12]Universitat Pompeu Fabra (UPF), Barcelona, Spain. [13]Institució Catalana d'Estudis Avançats (ICREA), Barcelona, Spain. [14]Department of Biology, Amherst College, Amherst, MA 01002, USA.

Received: 18 December 2014 Accepted: 28 May 2015
Published online: 16 June 2015

References

1. Kemler M, Lutz M, Göker M, Oberwinkler F, Begerow D. Hidden diversity in the non-caryophyllaceous plant parasite members of *Microbotryum* (Pucciniomycotina: Microbotryales). Syst Biodivers. 2009;7:297–306.
2. De Vienne DM, Hood ME, Giraud T. Phylogenetic determinants of potential host shifts in fungal pathogens. J Evol Biol. 2009;22:2532–41.
3. Le Gac M, Hood ME, Fournier E, Giraud T. Phylogenetic evidence of host-specific cryptic species in the anther smut fungus. Evol Int J Org Evol. 2007;61:15–26.
4. Le Gac M, Hood ME, Giraud T. Evolution of reproductive isolation within a parasitic fungal species complex. Evol Int J Org Evol. 2007;61:1781–7.
5. De Vienne DM, Refregier G, Hood ME, Guigue A, Devier B, Vercken E, et al. Hybrid sterility and inviability in the parasitic fungal species complex *Microbotryum*. J Evol Biol. 2009;22:683–98.
6. Giraud T, Yockteng R, Lopez-Villavicencio M, Refregier G, Hood ME. Mating system of the anther smut fungus *Microbotryum violaceum*: selfing under heterothallism. Eukaryot Cell. 2008;7:765–75.
7. Gibson AK, Hood ME, Giraud T. Sibling competition arena: selfing and a competition arena can combine to constitute a barrier to gene flow in sympatry. Evol Int J Org Evol. 2012;66:1917–30.
8. Alexander HM. An experimental field study of anther-smut disease of *Silene alba* caused by *Ustilago violacea*: genotypic variation and disease incidence. Evolution. 1989;43:835–47.
9. Antonovics J, Hood ME, Partain J. The ecology and genetics of host shift: *Microbotryum* as a model system. Am Nat. 2002;160:S40–53.
10. Refrégier G, Le Gac M, Jabbour F, Widmer A, Shykoff JA, Yockteng R, et al. Cophylogeny of the anther smut fungi and their caryophyllaceous hosts: prevalence of host shifts and importance of delimiting parasite species for inferring cospeciation. BMC Evol Biol. 2008;8:100.
11. Hood ME, Antonovics J, Koskella B. Shared forces of sex chromosome evolution in haploid-mating and diploid-mating organisms: *Microbotryum violaceum* and other model organisms. Genetics. 2004;168:141–6.
12. Hood ME, Petit E, Giraud T. Extensive divergence between mating-type chromosomes of the anther-smut fungus. Genetics. 2013;193:309–15.
13. Hughes CF, Perlin MH. Differential expression of mepA, mepC and smtE during growth and development of *Microbotryum violaceum*. Mycologia. 2005;97:605–11.
14. Roche BM, Alexander HM, Maltby AD. Dispersal and disease gradients of anther-smut infection of *Silene Alba* at different life stages. Ecology. 1995;76:1863–71.
15. Jennersten O. Butterfly visitors as vectors of Ustilago violacea spores between caryophyllaceous plants. Oikos. 1983;40:125–30.
16. Akhter S, Antonovics J. Use of internal transcribed spacer primers and fungicide treatments to study the anther-smut disease, *Microbotryum violaceum* (=*Ustilago violacea*), of white campion *Silene alba* (=*Silene latifolia*). Int J Plant Sci. 1999;160:1171–6.
17. Kokontis J, Ruddat M. Promotion of Hyphal growth in *Ustilago-Violacea* by host factors from *Silene-Alba*. Arch Microbiol. 1986;144:302–6.
18. Kokontis JM, Ruddat M. Enzymatic hydrolysis of Hyphal growth factors for *Ustilago violacea* isolated from the host plant *Silene alba*. Bot Gaz. 1989;150:439–44.
19. Scutt CP, Kamisugi Y, Sakai F, Gilmartin PM. Laser isolation of plant sex chromosomes: studies on the DNA composition of the X and Y sex chromosomes of *Silene latifolia*. Genome Natl Res Counc Can Genome Cons Natl Rech Can. 1997;40:705–15.
20. Alexander HM, Antonovics J. Spread of anther-smut disease (*Ustilago violacea*) and character correlations in a genetically variable experimental population of *Silene alba*. J Ecol. 1995;83:783–94.
21. Robertson SE, Li Y, Scutt CP, Willis ME, Gilmartin PM. Spatial expression dynamics of Men-9 delineate the third floral whorl in male and female flowers of dioecious Silene latifolia. Plant J Cell Mol Biol. 1997;12:155–68.
22. Kazama Y, Koizumi A, Uchida W, Ageez A, Kawano S. Expression of the floral B-function gene SLM2 in female flowers of Silene latifolia infected with the smut fungus *Microbotryum violaceum*. Plant Cell Physiol. 2005;46:806–11.
23. Billiard S, López-Villavicencio M, Devier B, Hood ME, Fairhead C, Giraud T. Having sex, yes, but with whom? Inferences from fungi on the evolution of anisogamy and mating types. Biol Rev Camb Philos Soc. 2011;86:421–42.
24. Billiard S, López-Villavicencio M, Hood ME, Giraud T. Sex, outcrossing and mating types: unsolved questions in fungi and beyond. J Evol Biol. 2012;25:1020–38.
25. Fraser JA, Heitman J. Evolution of fungal sex chromosomes. Mol Microbiol. 2004;51:299–306.
26. Fraser JA, Hsueh YP, Findley KM, Heitman J. Evolution of the Mating-Type Locus: The Basidiomycetes. In: Heitman J, Kronstad J, Taylor J, Casselton L, editors. Sex in Fungi. Washington, D.C: ASM Press; 2007. p. 19–34.
27. Menkis A, Jacobson DJ, Gustafsson T, Johannesson H. The mating-type chromosome in the filamentous ascomycete *Neurospora tetrasperma* represents a model for early evolution of sex chromosomes. PLoS Genet. 2008;4, e1000030.
28. Hood ME. Dimorphic mating-type chromosomes in the fungus *Microbotryum violaceum*. Genetics. 2002;160:457–61.
29. Hood ME, Antonovics J. Intratetrad mating, heterozygosity, and the maintenance of deleterious alleles in *Microbotryum violaceum* (=*Ustilago violacea*). Heredity. 2000;85(Pt 3):231–41.
30. Hood ME. Repetitive DNA in the automictic fungus *Microbotryum violaceum*. Genetica. 2005;124:1–10.
31. Whittle CA, Sun Y, Johannesson H. Degeneration in codon usage within the region of suppressed recombination in the mating-type chromosomes of *Neurospora tetrasperma*. Eukaryot Cell. 2011;10:594–603.
32. Whittle CA, Johannesson H. Evidence of the accumulation of allele-specific non-synonymous substitutions in the young region of recombination suppression within the mating-type chromosomes of *Neurospora tetrasperma*. Heredity. 2011;107:305–14.
33. Fontanillas E, Hood ME, Badouin H, Petit E, Barbe V, Gouzy J, et al. Degeneration of the non-recombining regions in the mating-type chromosomes of the anther-smut fungi. Mol Biol Evol. 2014;32(4):928–43. doi:10.1093/molbev/msu396. Epub 2014 Dec 21.
34. Farbos I, Oliveira M, Negrutiu I, Mouras A. Sex organ determination and differentiation in the dioecious plant *Melandrium album* (*Silene latifolia*): a cytological and histological analysis. Sex Plant Reprod. 1997;10:155–67.
35. Grant S, Hunkirchen B, Saedler H. Developmental differences between male and female flowers in the dioecious plant *Silene latifolia*. Plant J. 1994;6:471–80.
36. Parra G, Bradnam K, Korf I. CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. Bioinforma Oxf Engl. 2007;23:1061–7.
37. Rouxel T, Grandaubert J, Hane JK, Hoede C, van de Wouw AP, Couloux A, et al. Effector diversification within compartments of the *Leptosphaeria maculans* genome affected by repeat-induced point mutations. Nat Commun. 2011;2:202.
38. Duret L, Galtier N. Biased gene conversion and the evolution of mammalian genomic landscapes. Annu Rev Genomics Hum Genet. 2009;10:285–311.
39. Duplessis S, Cuomo CA, Lin Y-C, Aerts A, Tisserant E, Veneault-Fourrey C, et al. Obligate biotrophy features unraveled by the genomic analysis of rust fungi. Proc Natl Acad Sci. 2011;108:9166–71.
40. Yockteng R, Marthey S, Chiapello H, Gendrault A, Hood ME, Rodolphe F, et al. Expressed sequences tags of the anther smut fungus, *Microbotryum violaceum*, identify mating and pathogenicity genes. BMC Genomics. 2007;8:272.
41. Kapitonov VV, Jurka J. Rolling-circle transposons in eukaryotes. Proc Natl Acad Sci U S A. 2001;98:8714–9.
42. Haas BJ, Kamoun S, Zody MC, Jiang RH, Handsaker RE, Cano LM, et al. Genome sequence and analysis of the Irish potato famine pathogen *Phytophthora infestans*. Nature. 2009;461:393–8.

43. Gladieux P, Ropars J, Badouin H, Branca A, Aguileta G, de Vienne DM, et al. Fungal evolutionary genomics provides insight into the mechanisms of adaptive divergence in eukaryotes. Mol Ecol. 2014;23:753–73.
44. Hood ME, Katawczik M, Giraud T. Repeat-induced point mutation and the population structure of transposable elements in *Microbotryum violaceum*. Genetics. 2005;170:1081–9.
45. Horns F, Petit E, Yockteng R, Hood ME. Patterns of repeat-induced point mutation in transposable elements of basidiomycete fungi. Genome Biol Evol. 2012;4:240–7.
46. Cambareri EB, Jensen BC, Schabtach E, Selker EU. Repeat-induced G-C to A-T mutations in *Neurospora*. Science. 1989;244:1571–5.
47. Selker EU, Cambareri EB, Jensen BC, Haack KR. Rearrangement of duplicated DNA in specialized cells of *Neurospora*. Cell. 1987;51:741–52.
48. Walser J-C, Furano AV. The mutational spectrum of non-CpG DNA varies with CpG content. Genome Res. 2010;20(7):875–82. doi:10.1101/gr.103283.109. Epub 2010 May 24.
49. Fryxell KJ, Moon W-J. CpG mutation rates in the human genome are highly dependent on local GC content. Mol Biol Evol. 2005;22:650–8.
50. Jiang C, Zhao Z. Directionality of point mutation and 5-methylcytosine deamination rates in the chimpanzee genome. BMC Genomics. 2006;7:316.
51. Morton BR, Bi IV, McMullen MD, Gaut BS. Variation in mutation dynamics across the maize genome as a function of regional and flanking base composition. Genetics. 2006;172:569–77.
52. Amselem J, Lebrun M-H, Quesneville H. Whole genome comparative analysis of transposable elements provides new insight into mechanisms of their inactivation in fungal genomes. BMC Genomics. 2015;16:141.
53. Zemach A, McDaniel IE, Silva P, Zilberman D. Genome-wide evolutionary analysis of eukaryotic DNA methylation. Science. 2010;328:916–9.
54. Nicolás FE, Torres-Martínez S, Ruiz-Vázquez RM. Loss and retention of RNA interference in fungi and parasites. PLoS Pathog. 2013;9, e1003089.
55. Petit E, Giraud T, de Vienne DM, Coelho MA, Aguileta G, Amselem J, et al. Linkage to the mating-type locus across the genus *Microbotryum*: insights into nonrecombining chromosomes. Evol Int J Org Evol. 2012;66:3519–33.
56. Gillissen B, Bergemann J, Sandmann C, Schroeer B, Bölker M, Kahmann R. A two-component regulatory system for self/non-self recognition in Ustilago maydis. Cell. 1992;68:647–57.
57. Devier B, Aguileta G, Hood ME, Giraud T. Ancient trans-specific polymorphism at pheromone receptor genes in basidiomycetes. Genetics. 2009;181:209–23.
58. Votintseva AA, Filatov DA. Evolutionary strata in a small mating-type-specific region of the smut fungus Microbotryum violaceum. Genetics. 2009;182:1391–6.
59. Bachtrog D. Accumulation of Spock and Worf, two novel non-LTR retrotransposons, on the neo-Y chromosome of Drosophila miranda. Mol Biol Evol. 2003;20:173–81.
60. Elhaik E, Landan G, Braur D. Can GC content at third-codon positions be used as a proxy for isochore composition? Mol Biol Evol. 2009;26:1829–33.
61. Klose J, de Sá MM, Kronstad JW. Lipid-induced filamentous growth in *Ustilago maydis*. Mol Microbiol. 2004;52:823–35.
62. Castle AJ. Isolation and Identification of *α* -Tocopherol as an Inducer of the Parasitic Phase of *Ustilago violacea*. Phytopathology. 1984;74:1194.
63. Suh MC, Samuels AL, Jetter R, Kunst L, Pollard M, Ohlrogge J, et al. Cuticular lipid composition, surface structure, and gene expression in arabidopsis stem epidermis. Plant Physiol. 2005;139:1649–65.
64. Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B. The carbohydrate-active enzymes database (CAZy) in 2013. Nucleic Acids Res. 2014;42(Database issue):D490–5.
65. Cantarel BL, Coutinho PM, Rancurel C, Bernard T, Lombard V, Henrissat B. The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. Nucleic Acids Res. 2009;37(Database issue):D233–8.
66. Riley R, Salamov AA, Brown DW, Nagy LG, Floudas D, Held BW, et al. Extensive sampling of basidiomycete genomes demonstrates inadequacy of the white-rot/brown-rot paradigm for wood decay fungi. Proc Natl Acad Sci U S A. 2014;111:9923–8.
67. Prillinger H, Deml G, Dörfler C, Laaser G, Lockau W. A contribution to the systematics and evolution of higher fungi - yeast-types in the basidiomycetes, part II: microbotryum-type. Bot Acta. 1991;104:5–17.
68. Roeijmans H, Prillinger H, Umile C, Sugiyama J, Nakase T, Boekhout T (1998) Analysis of carbohydrate composition of cell walls and extracellular carbohydrates. In: Kurtzman C, Fell JW, Boekhout T (ed) The Yeasts - A Taxonomic Study. Volume 1. 5th edition. Elsevier, p. 103–105.
69. Liepman AH, Nairn CJ, Willats WGT, Sørensen I, Roberts AW, Keegstra K. Functional genomic analysis supports conservation of function among cellulose synthase-like a gene family members and suggests diverse roles of mannans in plants. Plant Physiol. 2007;143:1881–93.
70. Alexander NJ, McCormick SP, Hohn TM. TRI12, a trichothecene efflux pump from *Fusarium sporotrichioides*: gene isolation and expression in yeast. Mol Gen Genet. 1999;261:977–84.
71. Wahl R, Wippel K, Goos S, Kämper J, Sauer N. A novel high-affinity sucrose transporter is required for virulence of the plant pathogen *Ustilago maydis*. PLoS Biol. 2010;8, e1000303.
72. Voegele RT, Struck C, Hahn M, Mendgen K. The role of haustoria in sugar supply during infection of broad bean by the rust fungus *Uromyces fabae*. Proc Natl Acad Sci U S A. 2001;98:8133–8.
73. Moccia MD, Oger-Desfeux C, Marais GA, Widmer A. A White Campion (*Silene latifolia*) floral expressed sequence tag (EST) library: annotation, EST-SSR characterization, transferability, and utility for comparative mapping. BMC Genomics. 2009;10:243.
74. Veneault-Fourrey C, Commun C, Kohler A, Morin E, Balestrini R, Plett J, et al. Genomic and transcriptomic analysis of Laccaria bicolor CAZome reveals insights into polysaccharides remodelling during symbiosis establishment. Fungal Genet Biol FG B. 2014;72:168–81.
75. Antoniw JF, Ritter CE, Pierpoint WS, Loon LCV. Comparison of three pathogenesis-related proteins from plants of two cultivars of tobacco infected with TMV. J Gen Virol. 1980;47:79–87.
76. Baldrian P. Fungal laccases - occurrence and properties. FEMS Microbiol Rev. 2006;30:215–42.
77. Takahashi T, Kakehi J-I. Polyamines: ubiquitous polycations with unique roles in growth and stress responses. Ann Bot. 2010;105:1–6.
78. Brose N, Betz A, Wegmeyer H. Divergent and convergent signaling by the diacylglycerol second messenger pathway in mammals. Curr Opin Neurobiol. 2004;14:328–40.
79. Dong W, Lv H, Xia G, Wang M. Does diacylglycerol serve as a signaling molecule in plants? Plant Signal Behav. 2012;7:472–5.
80. Stergiopoulos I, de Wit PJGM. Fungal effector proteins. Annu Rev Phytopathol. 2009;47:233–63.
81. Bauer R, Oberwinkler F, Vanky K. Ultrastructural markers and systematics in smut fungi and allied taxa. Can J Bot. 1997;75:1273–314.
82. Palmieri G, Bianco C, Cennamo G, Giardina P, Marino G, Monti M, et al. Purification, characterization, and functional role of a novel extracellular protease from *Pleurotus ostreatus*. Appl Environ Microbiol. 2001;67:2754–9.
83. Palmieri G, Cennamo G, Faraco V, Amoresano A, Sannia G, Giardina P. Atypical laccase isoenzymes from copper supplemented *Pleurotus ostreatus* cultures. Enzyme Microb Technol. 2003;33:220–30.
84. Whittaker MM, Kersten PJ, Cullen D, Whittaker JW. Identification of catalytic residues in glyoxal oxidase by targeted mutagenesis. J Biol Chem. 1999;274:36226–32.
85. Luttrell ES. Tissue replacement diseases caused by fungi. Annu Rev Phytopathol. 1981;19:373–89.
86. Cashion NL, Luttrell ES. Host parasite relationships in Karnal bunt of wheat. Phytopathology. 1988;78:75–84.
87. Schäfer AM, Kemler M, Bauer R, Begerow D. The illustrated life cycle of *Microbotryum* on the host plant *Silene latifolia*. Botany. 2010;88:875–85.
88. Eisikowitch D, Lachance MA, Kevan PG, Willis S, Collins-Thompson DL. The effect of the natural assemblage of microorganisms and selected strains of the yeast *Metschnikowia reukaufii* in controlling the germination of pollen of the common milkweed *Asclepias syriaca*. Can J Bot. 1990;68:1163–5.
89. Biémont C. A brief history of the status of transposable elements: from junk DNA to major players in evolution. Genetics. 2010;186:1085–93.
90. Schmidt SM, Houterman PM, Schreiver I, Ma L, Amyotte S, Chellappan B, et al. MITEs in the promoters of effector genes allow prediction of novel virulence genes in Fusarium oxysporum. BMC Genomics. 2013;14:119.
91. Otto SP, Pannell JR, Peichel CL, Ashman T-L, Charlesworth D, Chippindale AK, et al. About PAR: the distinct evolutionary dynamics of the pseudoautosomal region. Trends Genet TIG. 2011;27:358–67.
92. Luo H, Perlin MH. The gamma-tubulin-encoding gene from the basidiomycete fungus, *Ustilago violacea*, has a long 5′-untranslated region. Gene. 1993;137:187–94.
93. Levin JZ, Yassour M, Adiconis X, Nusbaum C, Thompson DA, Friedman N, et al. Comprehensive comparative analysis of strand-specific RNA sequencing methods. Nat Methods. 2010;7:709–15.

94. Parkhomchuk D, Borodina T, Amstislavskiy V, Banaru M, Hallen L, Krobitsch S, et al. Transcriptome analysis by strand-specific sequencing of complementary DNA. Nucleic Acids Res. 2009;37, e123.
95. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol. 2011;29:644–52.
96. Borodovsky M, Lomsadze A, Ivanov N, Mills R. Eukaryotic gene prediction using GeneMark.hmm. Curr Protoc Bioinforma. 2003;Chapter 4:Unit4 6.
97. Stanke M, Steinkamp R, Waack S, Morgenstern B. AUGUSTUS: a web server for gene finding in eukaryotes. Nucleic Acids Res. 2004;32(Web Server issue):W309–12.
98. Majoros WH, Pertea M, Salzberg SL. TigrScan and GlimmerHMM: two open source *ab initio* eukaryotic gene-finders. Bioinforma Oxf Engl. 2004;20:2878–9.
99. Haas BJ, Salzberg SL, Zhu W, Pertea M, Allen JE, Orvis J, et al. Automated eukaryotic gene structure annotation using EVidenceModeler and the program to assemble spliced alignments. Genome Biol. 2008;9:R7.
100. Birney E, Clamp M, Durbin R. GeneWise and Genomewise. Genome Res. 2004;14:988–95.
101. Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, et al. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nat Protoc. 2013;8:1494–512.
102. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 2009;10:R25.
103. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics. 2011;12:323.
104. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2009;26:139–40.
105. Kadota K, Nishiyama T, Shimizu K. A normalization strategy for comparing tag count data. Algorithms Mol Biol AMB. 2012;7:5.
106. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc Ser B Methodol. 1995;57:289–300.
107. Flutre T, Duprat E, Feuillet C, Quesneville H. Considering transposable element diversification in *de novo* annotation approaches. PLoS ONE. 2011;6, e16526.
108. Edgar RC, Myers EW. PILER: identification and classification of genomic repeats. Bioinformatics. 2005;21 Suppl 1:i152–8.
109. Quesneville H, Bergman CM, Andrieu O, Autard D, Nouaud D, Ashburner M, et al. Combined evidence annotation of transposable elements in genome sequences. PLoS Comput Biol. 2005;1:166–75.
110. Bao Z, Eddy SR. Automated *de novo* identification of repeat sequence families in sequenced genomes. Genome Res. 2002;12:1269–76.
111. Jurka J, Kapitonov VV, Pavlicek A, Klonowski P, Kohany O, Walichiewicz J. Repbase update, a database of eukaryotic repetitive elements. Cytogenet Genome Res. 2005;110:462–7.
112. Hane JK, Oliver RP. RIPCAL: a tool for alignment-based analysis of repeat-induced point mutations in fungal genomic sequences. BMC Bioinformatics. 2008;9:478.
113. Zytnicki M, Quesneville H. S-MART, a software toolbox to aid RNA-seq data analysis. PLoS ONE. 2011;6, e25988.
114. Emanuelsson O, Brunak S, von Heijne G, Nielsen H. Locating proteins in the cell using TargetP, SignalP and related tools. Nat Protoc. 2007;2:953–71.
115. Bendtsen JD, Nielsen H, von Heijne G, Brunak S. Improved prediction of signal peptides: SignalP 3.0. J Mol Biol. 2004;340:783–95.
116. Petersen TN, Brunak S, von Heijne G, Nielsen H. SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat Methods. 2011;8:785–6.
117. Krogh A, Larsson B, von Heijne G, Sonnhammer EL. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J Mol Biol. 2001;305:567–80.
118. Pierleoni A, Martelli PL, Casadio R. PredGPI: a GPI-anchor predictor. BMC Bioinformatics. 2008;9:392.
119. Käll L, Krogh A, Sonnhammer ELL. Advantages of combined transmembrane topology and signal peptide prediction–the Phobius web server. Nucleic Acids Res. 2007;35(Web Server issue):W429–32.
120. Brameier M, Krings A, MacCallum RM. NucPred–predicting nuclear localization of proteins. Bioinformatics. 2007;23:1159–60.
121. Sigrist CJA, Cerutti L, de Castro E, Langendijk-Genevaux PS, Bulliard V, Bairoch A, et al. PROSITE, a protein domain database for functional characterization and annotation. Nucleic Acids Res. 2010;38(Database issue):D161–6.
122. Horton P, Park K-J, Obayashi T, Fujita N, Harada H, Adams-Collier CJ, et al. WoLF PSORT: protein localization predictor. Nucleic Acids Res. 2007;35(Web Server issue):W585–7.
123. UniProt Consortium. Ongoing and future developments at the Universal Protein Resource. Nucleic Acids Res. 2011;39(Database issue):D214–9.
124. UniProt Consortium. Reorganizing the protein space at the Universal Protein Resource (UniProt). Nucleic Acids Res. 2012;40(Database issue):D71–5.
125. Lum G, Min XJ. FunSecKB: the Fungal Secretome KnowledgeBase. Database J Biol Databases Curation. 2011;2011:bar001.
126. Li L, Stoeckert CJ, Roos DS. OrthoMCL: identification of ortholog groups for eukaryotic genomes. Genome Res. 2003;13:2178–89.
127. Eddy SR. Accelerated profile HMM searches. PLoS Comput Biol. 2011;7, e1002195.
128. Conesa A, Gotz S, Garcia-Gomez JM, Terol J, Talon M, Robles M. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics. 2005;21:3674–6.
129. Huerta-Cepas J, Capella-Gutierrez S, Pryszcz LP, Denisov I, Kormes D, Marcet-Houben M, et al. PhylomeDB v3.0: an expanding repository of genome-wide collections of trees, alignments and phylogeny-based orthology and paralogy predictions. Nucleic Acids Res. 2011;39(Database issue):D556–60.
130. Huerta-Cepas J, Gabaldón T. Assigning duplication events to relative temporal scales in genome-wide studies. Bioinforma Oxf Engl. 2011;27:38–45.
131. Ohm RA, Feau N, Henrissat B, Schoch CL, Horwitz BA, Barry KW, et al. Diverse lifestyles and strategies of plant pathogenesis encoded in the genomes of eighteen Dothideomycetes fungi. PLoS Pathog. 2012;8, e1003037.
132. Floudas D, Binder M, Riley R, Barry K, Blanchette RA, Henrissat B, et al. The Paleozoic origin of enzymatic lignin decomposition reconstructed from 31 fungal genomes. Science. 2012;336:1715–9.
133. Holliday R. Ustilago Maydis. In: King RC, editor. Handbook of Genetics. New York: Plenum Press; 1974. p. 575–95.

4.4 Article 3 : Identifying genomic regions under selection using population genomics in the anther-smut fungus

L'analyse du génome de référence de *M. lychnidis-dioicae* a permis de détecter un certain nombre de familles de gènes ayant subi des expansions chez *M. lychnidis-dioicae*, et qui pourraient être impliquées dans le mode de vie biotrophe du champignon et l'adaptation à son hôte.

Dans l'étude suivante, nous avons réalisé une étude de génomique des populations dans deux espèces jumelles du complexe *M. violaceum*, à savoir les espèces *M. lychnidis-dioicae* et *M. silenes-dioicae* qui parasitent respectivement *Silene latifolia* et *Silene dioica*. Pour cela nous avons reséquencé les génomes d'une soixantaine d'individus et détecté le polymorphisme par rapport à l'assemblage PacBio de la souche de référence présentée dans le chapitre sur les chromosomes de type sexuel (cf. 2.3 page 55).

Dans cette étude, nos objectifs étaient :

- D'étudier la présence d'introgressions entre *M. lychnidis-dioicae* et *M. silenes-dioicae*, une étude précédente ayant détecté la présence de quelques hybrides dans les populations naturelles [23], et, le cas échéant, d'étudier le contenu et la localisation des régions introgressées.

- De rechercher des régions sous sélection en prenant en compte autant que possible la structure génétique, afin de déterminer si les balayages sélectifs sont fréquents chez *M. lychnidis-dioicae* et *M. silenes-dioicae*, et d'identifier des gènes candidats dans les interactions avec l'hôte.

- De tirer parti du fait d'avoir un assemblage quasi-complet pour étudier les variations du polymorphisme le long du génome, et d'évaluer si ces variations sont liées à une architecture génomique particulière, ou à certaines caractéristiques du génome comme le contenu en GC ou le taux de recombinaison.

Dans cette étude, j'ai généré une grande partie des données de reséquençage, détecté les polymorphismes, et effectué la plupart des analyses. Le manuscrit est en cours de finalisation et n'a pas encore été soumis. Les figures et tables supplémentaires sont en annexe A.4 (page 264).

Identifying genomic regions under selection using population genomics in the anther-smut fungus

Hélène Badouin[1], Pierre Gladieux[1,2], Jérôme Gouzy[3,4], Sophie Siguenza[3,4], Gabriela Aguileta[1], Alodie Snirc[1] , Stéphanie Le Prieur[1], Antoine Branca[1]#, Tatiana Giraud[1]#§

Both authors supervised the work

[1] Ecologie Systématique Evolution, Univ. Paris-Sud, CNRS, AgroParisTech, Université Paris-Saclay 91400 Orsay, France

[2] INRA BGPI

[3] INRA, Laboratoire des Interactions Plantes-Microorganismes (LIPM), UMR441, Castanet-Tolosan, F-31326, France.

[4] CNRS, Laboratoire des Interactions Plantes-Microorganismes (LIPM), UMR2594, Castanet-Tolosan, F-31326, France

§Corresponding author: Tatiana Giraud tatiana.giraud@u-psud.fr

key words: *Microbotryum violaceum*, effectors, selective sweeps, GC content, selfing, linked selection

Abstract

Unraveling the genes underlying the adaptive phenotypes, the genomic processes underlying adaptation and the evolutionary forces shaping diversity patterns along genomes are important goals in biology. There is still little data to evaluate these questions beyond a handful of model species. Here, we sequenced 53 genomes of the two most extensively studied species of anther-smut fungi, *Microbotryum lychnidis-dioicae* and *M. silenes-dioicae,* parasitizing *Silene latifolia* and *S. dioica*, respectively. We found no traces of introgression along the genomes of these two sympatric sister species despite lack of premating isolation, indicating strong selection against hybrids in nature across the whole genome. Polymorphism was negatively correlated with levels of linkage disequilibrium along the genome, as expected in cases of recurrent selective sweep or background selection. Within each species we identified several genomic regions under selection, with footprints of selective sweeps, some including genes upregulated in *planta* and with putative functions making them good candidates for being effectors of plant-pathogen interactions. Striking examples include a glycoside hydrolase, pectin lyases and an extracellular membrane protein with CFEM domain, these functions being known to be involved in fungal pathogenicity. The selective sweeps appeared scattered along genomes and in different locations in the two *Microbotryum* species. Looking at patterns of divergence between the two sister species, we found that the genes upregulated *in planta* accumulated more non-synonymous substitutions than other genes, supporting the view that many of them are likely effectors involved in interactions with the host plant. This study may thus finally provide clues on the genes involved in plant - pathogen interaction, that have remained elusive for long in this otherwise well-studied system. The identification of putative effectors will foster future functional and evolutionary studies, in the plant and in the anther smut pathogens, being model species of a natural plant-pathogen association. In addition, this study has even more general interest, showing that selective sweeps are widespread in natural plant pathogens, contributing to a broader picture of the occurrence and frequency of selection in natural populations and illustrating the ability of genome scans to detect adaptive events.

Introduction

Adaptation has been studied extensively and successfully on phenotypes and candidate genes. However, we are just beginning to be able to investigate the general processes underlying adaptation at the whole genome level and the importance of adaptation in shaping genomic patterns. Reaching a new level of understanding of the genomic processes of adaptation requires in particular the following questions to be addressed: How many and what kinds of genes are important in adaptation? How are they distributed along genomes? Are major adaptation events frequent or recent enough to be detected in genomes? What is the influence of intrinsic genomic features and genomic architecture on local genomic diversity? Do introgressions between species play a major role in adaptation? There is still little data to evaluate these issues beyond a handful of model species (*e.g.*, [1–9]). In particular, most of examples of selection detected using genome-wide polymorphism come from studies on very few model species, such as humans, drosophila, maize, *Arabidopsis* or malaria parasites [10–14]. We therefore need studies on more diverse organisms for eventually drawing generalities on the forces shaping diversity along genomes and for understanding the genomic architecture of adaptation.

Selective sweeps, caused by recent positive selection, leave traces on the pattern of polymorphism in the genome and various methods have been developed for detecting these footprints. First, statistics based on allele frequency spectrum (AFS) look for deviations from neutral expectations in the form of an excess of low frequency variants (positive or negative selection) or of intermediate frequency variants (balancing selection) [15]. Identifying polymorphisms as ancestral or derived give further power to detect selection through the derived allele frequency spectrum (dAFS). An excess of derived alleles in the shape of the dAFS indicates positive selection while an excess of ancestral alleles reflects negative selection [16]. The drawback of methods based on allele frequency spectra is that they are usually sensitive to past demographic events. Statistics contrasting local AFS to background AFS have therefore been developed, detecting outliers compared to the background, genome-wide allele frequency spectrum that integrates the impact of demography (CLR) [17]. Second, accumulation of non-synonymous substitutions at a higher rate than synonymous substitutions is widely used as evidence of positive selection [18]. Contrasting levels and non-synonymous and synonymous

polymorphism and divergence to test for a departure from neutrality can be even more powerful than DN/DS for detecting recurrent adaptive evolution [19]. Recurrent selective sweeps also leave footprints in genomes in terms of patterns of polymorphism. Due to hitchhiking during selective sweeps or background selection, polymorphism and recombination rate are in general positively correlated along chromosomes [20]. GC content can also be positively correlated with levels of diversity along genomes [21], because local enrichment in GC often occurs in regions where recombination or gene conversion occurs frequently [22]. Conversely, a negative correlation between GC content and polymorphism is expected at equilibrium when GC-biased gene conversion occurs [23,24].

Introgression between species can play an important role in adaptation and leave footprints in genomes in the form of genomic islands of low divergence between species; examples in fungi include wine yeasts [25,26] and *Coccidioides* human pathogens [27]. Fungi in fact present great potential as tractable models for studying adaptation and have great agronomic, medical, industrial, and ecological importance [28–31]. Although fungi are often considered to be microbes, they are very similar to animals in terms of evolution, forming the Opisthokonta clade with them. The inferences drawn from fungi can therefore provide information that can be extended to the processes of genomic adaptation in eukaryotes. In the study of eukaryotic adaptive divergence, fungi present many advantages [29,30], such as small genomes and high abundance of complexes of sibling species adapted to different hosts or habitats. Indeed, fungi have just started to be successfully used to study adaptation [30,32], but there have been few studies so far analysing multiple fungal genomes within species for studying selection and introgression, beyond a few exceptions (*e.g.*[32,27,33]).

The *Microbotryum* fungi, causing anther-smut disease in perennial plants of the Caryophyllaceae family [34], are particularly good models for addressing these questions. Indeed, they represent some of the best studied plant pathogens in natural ecosystems [35]. The ecology of the disease has been intensively studied (*e.g.* [36–41]) and *Microbotryum* fungi belong to a complex of sibling species specialized on different hosts in sympatry [42,43]. The fungi castrate their host plants, producing their spores in anthers in the place of pollen and inducing ovaries abortion. The different *Microbotryum* species display strong host specificity [44], no interspecific mate

discrimination [45], and show post-mating reproductive isolation increasing with genetic distance [44,45]. *Microbotryum* fungi undergo an obligate sex event before each new plant colonization [46].

One crucial lack of knowledge in this model system is the genetics of the interaction between the pathogens and their host plants. No gene-for-gene relationship has been detected despite intensive studies for several decades. Quantitative resistance nevertheless varies among plant families in all the species examined [47,48,38,49,41]. The species *M. lychnidis-dioicae* has been shown to be locally maladapted [50], *i.e.*, its host *Silene latifolia* is locally adapted, being more resistant to its local pathogens than to allopatric ones. This suggests that coevolution may be occurring in natural populations. However, little is known about the genes involved in the interaction, neither from the fungi nor the plant side.

Population genomics in *Microbotryum* may help detecting candidate genes relevant for coevolution with their host plants. As pathogenic and castrating pathogens, *Microbotryum* fungi are ideal models as theoretical studies have shown that these features are the most favorable for observing footprints of coevolution in genomes [51,52].

Here, we focused on the two most-studied species of anther smut fungi, the sister species *M. lychnidis-dioicae* and *M. silenes-dioicae*, parasitizing *S. latifolia* and *S. dioica*, respectively; their divergence has been estimated to have taken place ca. 420,000 years ago [53].

In the lab, hybrids between *M. lychnidis-dioicae* and *M. silenes-dioicae* are viable and fertile [44,45,54,55] and both species can infect both host plants [44,54]. In natural populations, hybrids have in fact been detected, although they were rare [53], despite largely overlapping geographical distributions in Europe [56] and lack of increased premating reproductive isolation in sympatry [57]. These previous studies using a dozen of microsatellite markers could however probably only detect early generation hybrids and could not address the question of whether some genomic regions were more or less permeable to persisting introgression.

Studies within the *M. violaceum* species complex using microsatellites have furthermore detected a strong geographical population structure, especially in *M. lychnidis-dioicae*, revealing footprints of ancient glacial refugia, as three genetic clusters found, respectively, in western Europe, Italy and eastern Europe, respectively [56]. Little admixture has been found between clusters based on microsatellites [58,56,59].

Therefore, the specific questions we addressed here were the following: 1) what is the degree of genome-wide long term introgression between *M. lychnidis-dioicae* and *M. silenes-dioicae*? What is the degree of genome-wide gene flow between geographic clusters within *M. lychnidis-dioicae*? 2) Can we detect genomic regions under selection in the genomes of the anther smut fungi *M. lychnidis-dioicae* and *M. silenes-dioicae*? Are they clustered in particular genomic locations? Do they involve the same genomic regions in both *Microbotryum* species? 3) Can we identify candidate effectors, *i.e.*, proteins involved in the interaction with the host plant?

For these goals, we sequenced the genomes of multiple individuals of *M. lychnidis-dioicae* and of *M. silenes-dioicae* sampled across Europe, as well as one individual of an outgroup species. Because tests of selection can be biased by a strong genetic structure [60], we first analyzed the population structure of these species. In order to detect genes potentially involved in host-pathogen co-evolution or adaptation to different host plants, tests of selection were then run along the genomes, looking for selective sweeps and genes with high rates of non-synonymous substitutions. The tests of selection were run both genome-wide and focusing on candidate genes possibly involved in interaction with the host plant, *i.e.*, secreted proteins and genes up-regulated *in planta* during infection in *M. lychnidis-dioicae* in *S. latifolia* (Perlin et al. 2015). We then tested whether polymorphism was negatively correlated with levels of linkage disequilibrium along the genome, as expected in cases of recurrent selective sweep or background selection. We used linkage disequilibrium as proxy of recombination rate, as it has been shown to be strongly correlated to recombination rate, in particular in fungi (Croll et al., 2015) We also tested the influence on diversity levels along genomes of other genomic features, such as gene density and GC content.

Results

SNP calling

We sequenced at 100X depth of coverage the genomes of 34 individuals of *M. lychnidis-dioicae*, 19 individuals of *M. silenes-dioicae* from across Europe, and one individual of an outgroup species, *M. violaceum s.l.* infecting *S. flos-cuculi* using Illumina HiSeq 2000 paired-end reads (Figure 1, Supp Table 1). We sequenced a single individual per locality, as previous studies showed little variability at this scale [56,53]. We excluded individuals identified as hybrids in previous studies using microsatellites, as these were likely early generation hybrids that would therefore provide little information on long-term introgression levels. Sequenced genomes were diploid, except six of them that were haploid (See Materials and Methods). This yielded a total number of 18.5 to 77.4 million reads of 100 bp.

Reads were mapped to the reference genome of *M. lychnidis-dioicae* p1A1 lamole (CLDV0100001-22), with finished, ungapped chromosomes or chromosomal arms [61], to call single nucleotide polymorphisms (SNPs). The percentage of paired reads mapped in a proper pair, *i.e.*, in the expected orientation and distance, was 40 % for the outgroup, and ranged otherwise from 59.1 to 94.9 % (Supp Table 2). SNPs were called with VarScan2 [62]. Because the mating-type chromosomes in *M. lychnidis-dioicae* exhibit suppressed recombination on 90% of their lengths [63,64,61], we could not apply the same tests as on autosomes to look at what factors influenced the variation in diversity or at footprints of selection. We therefore focused on autosomes in the present study.

After filtering, we obtained 203,347 bi-allelic autosomal polymorphic positions within *M. lychnidis-dioicae* and 30,296 within *M. silenes-dioicae*. Because *M. lychnidis-dioicae* and *M. silenes-dioicae* exhibit high rates of selfing [65], we expected low rates of heterozygous SNPs in autosomes in diploid genomes, except in recently admixed individuals. The percentages of heterozygous SNPs per strain ranged from 4 to 9% (Supp Table 3), with the exception of three strains in which heterozygosity reached 14%, 34% and 38%, respectively. The percentages of heterozygous SNPs (4-9%) were similar between diploid and haploid genomes. This suggested that most heterozygous sites probably did not represent genuine heterozygosity, but instead were

artefacts due to repetitive DNA or recently duplicated regions. A previous study similarly found unexpected heterozygous SNPs despite stringent filtering in highly inbred nematodes [66]. Therefore, we only considered homozygous SNPs in subsequent analyses and treated each individual as haploid, except for the inference of genetic structure where genuine heterozygous SNPs may help for the detection of admixed individuals. In fact, the three genomes showing an elevated number of heterozygous SNPs may represent footprints of recent admixture.

Genetic Structure in Microbotryum lychnidis-dioicae and *M. silenes-dioicae*

Population genetics inferences can be biased by an unrecognized genetic subdivision [60]. It is therefore important to delimit genetic clusters. To infer population structure and detect putative hybrids between *M. lychnidis-dioicae* (MvSl) and *M. silenes-dioicae* (MvSd), we ran *fastStructure* [67] for a number of clusters (K) ranging from 1 to 10. Although *Microbotryum* fungi are highly selfing, outcrossing and effective recombination occur frequently enough for such assignment algorithms to be able to accurately identify genetic clusters [53,56,68]. We assessed optimal numbers of clusters based both on the statistics provided by *fastStructure* and visual examination, looking for K values at which further increasing K did not yield any additional, well-defined, non-admixed genetic cluster. The optimal number of clusters in the total dataset pooling MvSl and MvSd was K=3, with a cluster corresponding to MvSd, and MvSl being divided into two clusters, one in western Europe and one in Italy and eastern Europe (Figure 1). At none of the K value did we detect admixed strains between the two species.

For inferring more finely the population structure within MvSl, we ran *fastStructure* on the SNPs of MvSl alone (Supp Figure 1). At K=3, a clear genetic structure was observed, with well-separated clusters in western, eastern and southern Europe, respectively (Figures 1A and C), as previously found using microsatellite data [56]. The optimal number of clusters was four, with a further split of the western cluster, separating a large cluster of 17 strains, hereafter called the north-western cluster, from three strains, two of them appearing admixed (Figure 1B and 1E). We also performed a principal component analysis (PCA) and reconstruct a neighbor joining (NJ) tree, both of which do not rely on any assumption about mating system or gene flow, and help visualize levels of differentiation and relationships between clusters (Figure 1D). In *M.*

lychnidis-dioicae, the PCA confirmed the strong genetic structure, the first two axes of the PCA explaining 38.9% and 13.3% of the variance, respectively. On the NJ tree, the eastern and Italian MvSl clusters appeared the closest one to each other and the most diverse, while the western MvSl cluster displayed less variability. This supports previous conclusions that the MvSl north-western cluster experienced a bottleneck during the post-glaciation recolonization [56]. The NJ tree also illustrated the strong differentiation between MvSl and MvSd and the particularity of the three strains from the green western cluster in Figure 1E, warranting their separation from the north western cluster for subsequent analyses. We obtained similar results for all clustering analyses keeping only SNPs in linkage equilibrium.

In the fastStructure analysis, four strains appeared as likely resulting from admixture between clusters (Figure 1E), *i.e.*, two strains located in Austria and admixed between the Italian and eastern clusters, and two strains located in France, admixed between the North- and South-western clusters. Out of the three strains displaying the highest numbers of heterozygous SNPs, one was detected as admixed between the North-western and South-western clusters. The two other highly heterozygous strains, located in Central and South-western France, did not appear intermediate either using FastStructure, PCA or discriminant PCA (dPCA, result not shown). Nevertheless, the relatively high level of heterozygosity, combined with their geographical location, intermediate between the north- and southwest clusters, suggested that these strains may also result from recent admixture. In subsequent analyses, we conservatively excluded the six potentially admixed strains, and considered four groups: *M. silenes-dioicae* (19 individuals), and three clusters of *M. lychnidis-dioicae*, *i.e.*, the Italian cluster (7 individuals), the eastern cluster (5 individuals) and the northwestern cluster (15 individuals). We did not further consider the South-western cluster as it included only three strains (Figure 1D).

Genome-wide levels of diversity and divergence

Differentiation between genetic groups was measured with F_{ST} [69]. We also computed d_{xy}, *i.e.*, the average number of differences between sequences of different populations [70], and the number of fixed differences between populations. Median nucleotidic diversity (Pi) and are presented in Table 1. Median Pairwise F_{ST} and fixed divergence are presented in Table 2.

Median diversity per bp was low, ranging from 2.22e-5 in MvSd to 1.1e-3 in the MvSl Italian cluster; such a low diversity is in agreement with previous results based on microsatellite markers [56]. The lower diversity in MvSd is also consistent with previous estimates of a smaller effective population size in MvSd than in MvSl [53].

Pairwise F_{ST} was very high between MvSd and each MvSl cluster (0.93 to 0.96, Table 2), and also between MvSl clusters, although to a lower extent (0.56 to 0.74, Table 2). Furthermore, there were few shared polymorphisms and many fixed differences between clusters, most polymorphisms being private; there were 2,590 shared polymorphisms between the species and the median Tajima's D was -1.03 in MvSd, and -0.2, 0.02 and 0.2 in the eastern, North-western and Italian MvSl clusters, respectively. Because a negative Tajima's D can be due to population expansion after a bottleneck, these values are also consistent with the view that MvSd and the MvSl western and eastern clusters experienced stronger range contraction and expansion because of the last glaciation than the MvSl Italian cluster [56].

Variation of polymorphism, divergence and linkage disequilibrium along the genome

We analyzed the patterns of linkage disequilibrium (LD) using r^2, *i.e.*, the coefficient of correlation between pairs of SNPs [71]. In MvSl, the background LD level was high, reflecting the strong underlying structure. Within the North-western cluster, the background LD was lower, allowing genomic patterns to be analyzed. LD decayed more rapidly in MvSd than in North-Western MvSl, r^2 decreasing below 0.2 after 67 kb in MvSd and 96 kb in North-Western MvSl (Supp Figure 2). LD was stronger around centromeres, and some additional large blocks of linkage disequilibrium were found on chromosomal arms, in particular on contigs S04, S06 and S08 in North-western MvSl. In MvSd, a few large blocks of strong LD were detected, around centromeres and elsewhere, for example on contig S10.

Variations of polymorphism within each of the four genetic groups, as well as pairwise F_{ST} and mean linkage disequilibrium within the MvSl North-western cluster, are shown in Figure 2. Statistics were computed using non-overlapping sliding windows of 50 or 100 kb, in each of the three MvSl clusters and in MvSd. Window sizes were chosen in order to have enough variation

per window, and thus limit noise, and also to account for the observed LD decay. Within-cluster polymorphism levels along chromosomes were significantly correlated between MvSl clusters (Supp Table 4, *e.g.,* Spearman's r = -0.46, P < 0.0001 using 100 kb windows between the Italian and North-western clusters), and between MvSl and MvSd (Spearman's r = -0.41, P < 0.0001 using 100 kb windows between MvSd and the MvSl Italian cluster). Near centromeric regions, nucleotidic diversity appeared lower while mean linkage disequilibrium appeared higher than elsewhere along the genome. The diversity measured as Pi and linkage disequilibrium were negatively correlated along chromosomes in all clusters of MvSl but not in MvSd (Supp Table 6, *e.g.,* Spearman's r = -0.39, P < 0.0001 in the North-western MvSl cluster).

F_{ST} between species was high all along the genome and we did not observe regions with very low F_{ST}. This, together with the F_{ST} values close to 1 (see paragraph above and Table 2), indicated low levels of gene flow and the absence of particular genomic regions that would be more permeable to gene flow. F_{ST} between MvSl clusters remained high (see paragraph above and Table 2), but we did not observe clear F_{ST} peaks that could indicate regions of local adaptation.

Looking for footprints of selection in candidate genes

For detecting genes under positive selection, we first used a candidate gene approach, analysing specific gene categories potentially involved in fungal plant pathogenicity, mainly the genes up-regulated *in planta* compared to *in vitro* conditions previously identified in *M. lychnidis-dioicae* in *S. latifolia* [72]. Further gene categories were also tested for positive selection (*i.e.,* genes encoding secreted proteins, major facilitator superfamily and sugar transporters), but did not yield significant results (Supplementary text). For each category of genes of interest, we tested whether they were under positive selection, *i.e.,* accumulated more non-synonymous substitutions than expected under neutral evolution, as estimated using synonymous substitution rates; we thus tested whether the gene category of interest had higher P_N/P_S within cluster (*i.e.,* the ratio of the numbers of non-synonymous over synonymous polymorphisms, normalized by the number of non-synonymous and synonymous sites, respectively) and/or higher DN/DS between clusters (*i.e.,* ratio of the number of non-synonymous over synonymous differences),

than the remaining gene set. We also tested for an enrichment of those genes in the tails of the distribution of the neutrality index, defined as $(DN/P_N)/(DS/P_S)$, which measures the strength and direction of departure from neutrality [27].

The genes up-regulated *in planta* showed a significantly higher DN/DS than other genes when computed between MvSd and each MvSl cluster (mean = 0.67 versus 0.51 in other genes, FDR = 3.42e-4, between North-western MvSl and MvSd), but not between MvSl clusters. P_N/P_S was slightly higher, but not significantly so. In addition, genes up-regulated *in planta* were enriched in the category of genes with DN/DS higher than 1 when computing the statistics between MvSd and MvSl clusters (FDR < 0.05), but not between MvSl clusters. The 21 secreted proteins genes up-regulated *in planta* showed a higher non-synonymous diversity (0.0018040 versus 0.0006 for the remaining gene set in the MvSl Italian cluster) but similar synonymous diversity as other genes; their DN/DS was higher than other genes, but not significantly so.

In order to determine if the genes up-regulated *in planta* evolved more frequently under adaptive evolution than other genes, we looked for an enrichment of the different categories of genes of interest in the tails of NI distribution. Enrichment in the left tail of the distribution (low NI, $DN/DS > P_N/P_S$) would indicate more frequent positive selection in the candidate categories than other genes, while enrichment in the right tail (high NI, $DN/DS < P_N/P_S$) would indicate more frequent balancing selection. We found a twofold, significant enrichment of genes up-regulated *in planta* in the left tail of the distribution (FDR < 0.05, one-sided Fisher's exact test, in comparisons between North-western MvSl *vs.* MvSd and eastern MvSl *vs.* MvSd, p-value < 0.05 between MvSl Italy and MvSd but non-significant considering FDR). The distribution of the neutrality index was shifted toward low NI in genes up-regulated *in planta* compared to other genes (Supp Figure 4).

Looking for potential effectors, we examined the Pfam annotations of the genes up-regulated *in planta* that showed the lowest neutrality index values (Table 4). Interesting annotations known to be involved in fungal pathogenicity (see discussion) included two major facilitators, two fungal-specific cysteine-rich CFEM domains, one of them secreted and the other anchored to the membrane, a secreted aspartic peptidase, a sugar transporter, a secreted multi-copper oxidase, a glyoxal oxidase and a ferritin.

Whole Genome Scan for Selection

In order to detect genes under selection without *a priori* candidates, we looked for footprints of selective sweeps in the two largest clusters, *i.e.,* MvSd and the MvSl North-western cluster. We ran the software SweeD, which implements a composite likelihood ratio (CLR) test using the site frequency spectrum for detecting selective sweeps. This algorithm compares the local site frequency spectrum to the site frequency spectrum computed on the whole alignment, for detecting shifts to low and high frequency variants. We computed CLR tests every 10 kb, and detected outliers with a FDR of 0.05 to identify putative selective sweeps. Several dozens of outliers regions were detected in MvSl North-western cluster and were scattered along the genome. Fewer sweeps were found in MvSd, and in different locations compared to those in MvSl (Figure 3). Outliers regions spanned 10 to 100 kb. They were not enriched in putatively secreted proteins, genes up-regulated *in planta*, or any specific gene ontology terms (results not shown). To find candidate genes for interaction with the host, we examined the expression patterns and putative functions of the genes located at the center of the sweeps (Supp Figure 3). Among the interesting candidates, we found a gene upregulated *in planta* with a CFEM domain, a clusters of transporters with OPT and major facilitator domains and a secreted lipase. OPT and major facilitator domains are expanded in *M. lychnidis-dioicae*, and could be involved in nutrient uptake [72]. Proteins with CFEM domains could be involved in interaction with the host and act as virulence factors [72]. In several cases, we found at the center of the sweeps genes upregulated *in planta* but without putative functions assigned (Supp Figure 3).

Because genes that co-evolve with the host, such as effectors interacting with the host defense system, are expected to show high levels of non-synonymous substitutions [73], we also analyzed patterns of synonymous and non-synonymous substitutions between clusters and between species. Indeed, genes that are involved in adaptation to different host plants are expected to be under positive selection between species but not necessarily within species; in fact, a dozen of genes that showed signatures of positive selection between *Microbotryum* species adapted to different hosts [74] were found under purifying selection within species [53].

We also performed McDonald and Kreitman tests on all genes between MvSd and each MvSl cluster. The tests of McDonald and Kreitman (MK tests) [19] contrast levels and non-synonymous and synonymous polymorphism and divergence to test for a departure from neutrality in non-synonymous sites, which can be even more powerful than DN/DS for detecting recurrent adaptive evolution [19]. Only a few dozen of tests were significant at the 0.05 level, and none after applying correction for multiple testing.

Determinants of polymorphism variation along the genome

We observed a strong negative correlation between polymorphism and per-base number of fixed differences between each MvSl cluster and MvSd, showing that the less polymorphic regions tended to accumulate more fixed differences (Supp Table 6, *e.g.,* Spearman's $r = -0.40$, $P < 0.001$ at 100 kb in the MvSl Italian cluster).

We used multiple linear regression models for testing what variables impacted the levels of within-species diversity along the genomes (using either total (Pi) or synonymous (Pis) diversity). In the case of genetic hitchhiking of sites linked to sites under selection [75], the selection of an adaptive allele reduces diversity on longer fragments when recombination rate is low, leading to a positive correlation between polymorphism and recombination rate [76], and thus a negative correlation between local diversity and linkage disequilibrium. For the same reason, we would also expect a negative correlation between polymorphism and the density of coding sites, where selection mostly occurs [77]. Because genetic hitchhiking of sites linked to sites under selection is expected to occur more frequently in cases where non-synonymous substitutions (that may be deleterious or adaptive) are frequent (as in angiosperms, [78]), we also examined the effect of DN/DS, *i.e.,* the ratio between the number of non-synonymous substitutions and the number of synonymous substitutions. GC content was also used as an explanatory variable for levels of diversity, as GC enrichment has been observed in regions where recombination occurs frequently or in contrast GC-biased gene conversion can lead to reduced variability associated with high GC content. We thus included the following variables in the models, for each cluster and window size: mean linkage disequilibrium (LD), GC content,

CDS density and pairwise DN/DS (*i.e.*, number of synonymous substitutions over number of non-synonymous substitutions).

Regarding the models for explaining the total genetic diversity (Pi), the mean linkage disequilibrium explained the highest part of variance. Its effect was negative and it was significant in all MvSl clusters but not in MvSd. GC content had a significant negative effect in 6 out of the 8 models. The effects of density in coding sites and DN/DS were not significant. The models explained 8.3 to 27.1 % of the variance depending on the cluster, variables and windows size considered (Table 3). The Figure 4 illustrates the relationships between polymorphism levels and GC content or linkage disequilibrium. We then built similar models to explain synonymous diversity. Linkage disequilibrium remained the most influential explanatory variables for synonymous diversity. GC content had a significant negative effect on diversity in 6 out of the 8 models. In models separating the effects of GC content at the first two (GC_{12}) or third codon position (GC_3), GC_{12} had a stronger effect than GC_3 (Supp Table 7). DN/DS had a significant negative effect in a single model (MvSl North-western cluster, 100 kb windows). The models explained from 5.1 to 25.6 % of variance.

Discussion

Low linkage disequilibrium levels indicate effective recombination despite high selfing rates

The analyses of linkage disequilibrium showed that the r^2 decreased below 0.2 after less than 100 kb in both *Microbotryum* species. This indicates that effective recombination regularly occurs in nature despite the high rates of intra-tetrad mating. Selfing rates were in fact previously found to be below 100%, with estimates based on microsatellite markers ranging between 0.90 and 0.95 in *M. lychnidis-dioicae* and *M. silenes-dioicae* [56]. Therefore, while this level of selfing is high enough to greatly depress individual heterozygosity, it is low enough to thoroughly mix haplotypes, allowing performing structure and selection analyses assuming panmixia.

The linkage disequilibrium level showed some degree of heterogeneity along the genome, with some peaks of higher LD around centromeres and elsewhere. Centromeres are indeed known to be regions with little recombination. The LD peaks on chromosome arms did not correspond to selective sweeps, as could have been expected in cases of large genetic hitchhiking. This again reinforces the view that recombination is effective despite high selfing rates. The LD heterogeneity along chromosome arms is thus likely due to hotspots and cold spots of recombination [79].

High levels of differentiation between species and lack of introgression

We detected no gene flow between the species *M. lychnidis-dioicae* and *M. silenes-dioicae* in whole genome sequences, while previous studies using microsatellite markers had detected some admixed genotypes [53]. The hybrids identified using microsatellite were however rare. Low levels of hybridization between those two species could be in part due to the different habitats of the plants and different pollinator guilds [80,81]. In addition, the detected hybrids were likely of early generations to be detectable using a dozen of markers. We chose here on purpose to sequence non-admixed microsatellite genotypes to be able to assess whether long-term gene flow occurred beyond first-generation hybrids between the two sympatric sister species parasitizing

sister plant species. This was not the case, indicating that introgression does not persist beyond one or two generations. Experimental crosses had previously shown lack of pre-mating isolation among *Microbotryum* species, even when co-occurring in the very same populations [45,57] but with increasing hybrid inviability and sterility with genetic distance [45,44]. F1 hybrids between the two sister species studied here, *M. lychnidis-dioicae* and *M. silenes-dioicae*, performed well in the lab [45,55,57], but F2 hybrids produced by F1 selfing had already mostly returned to homozygosity based on three dozens of microsatellite markers [44]. Our results here definitely show that no introgression persist in nature, in any genomic regions, further arguing for a very strong genome-wide selection, likely promoted by the scattering of genes involved in interactions with the host across the genome. This is consistent with the finding of selective sweeps and genes up-regulated in planta all along the genomes and not in particular locations (Figure 2). Selfing also acts as a strong barrier to inter-specific gene flow in *Microbotryum* because of the systematic competition between hybrids and selfed progeny during plant infection [59].

High levels of differentiation between Microbotryum lychnidis-dioicae clusters

We found very high levels of differentiation between MvSl clusters, indicating low levels of gene flow. This is likely also due to high levels of selfing, as well as to lack of long distance dispersal in this insect-borne fungus. F_{ST} between MvSl clusters was variable along chromosomes, but no particular pattern could be detected; some genomic regions showed drops in F_{ST} between MvSl clusters, but they did not correspond to the detected selective sweeps or to genes upregulated in planta with a low NI (Figure 2), except maybe on the scaffold S04. Variation in F_{ST} is not necessary due to different levels of gene flow but can alternatively can be due to linked selection and recombination rates: less recombination drives fixation of different alleles more rapidly between populations or species [82].

Levels of diversity along chromosomes indicate hitchhiking during selection

We found a strong negative effect of LD on diversity along the genomes, which suggests that recurrent linked selection due to either selective sweeps or background selection indeed plays a role in shaping diversity within *Microbotryum* populations [20]. There was a strong negative correlation between polymorphism and per-base number of fixed differences between each MvSl cluster and MvSd, showing that the less polymorphic regions tended to accumulate more fixed differences. This is also in agreement with recurrent selective sweep and background selection. Indeed, lower recombination rates increase the probability of mutation fixation within species at any given site by background selection [83]. However, we did not observe a negative effect of gene density on polymorphism, as would be expected under a model of linked selection if genes were the primary targets of selection [77].

We found a negative correlation between polymorphism level and GC content, which is expected at equilibrium if GC-biased gene conversion occurs [24]. Indeed, in such a case the GC-biased conversion acts as selection in maintaining lower levels of diversity [24]. GC-biased conversion has been shown to occur in fungi at recombination hotspots [84]. A negative correlation between polymorphism level and GC content can also be due to selection for GC-rich codons. In fact, in *M. lychnidis-dioicae,* GC content was found positively correlated with coding density [72], which can be due to biased codon usage toward GC-rich codons or biased gene conversion occurring more frequently in coding than in non-coding sequences [85]. An analysis of the preferred codons (*i.e.*, the most frequently used codons in the predicted genes) indeed showed that 17 out of 18 had a GC base in the third position [72], which is the most degenerate position and therefore primarily influences the GC composition of genes. However, our findings that GC_3 was less often significantly correlated with polymorphism than GC_{12} do not support these hypotheses, although we are not aware of any alternative explanation accounting for a significant negative relationship between GC content and polymorphism.

Detection of genes under selection

We detected particularly high genetic diversity in the genes upregulated *in planta*, indicating that these accumulated more non-synonymous substitutions than other genes, which supports the view that they are likely involved in interaction with the host plant. The shift in the neutrality index between the genes upregulated *in planta* and the remaining genes was however weak,

which may be due to the fact that only a fraction of them are involved in the host-pathogen arms race. We therefore focused on the genes upregulated *in planta* with a low value of the neutrality index, *i.e.*, those with highest non-synonymous substitutions, and found several interesting putative functions. Among them, CFEM proteins are cysteine-rich with extracellular domains and particularly good candidates for being fungal effectors [72,86]; a glyoxal oxidase has been found to be required for pathogenicity in *Ustilago maydis* [87]. Multi-copper oxidases and ferritin can protect against host-induced oxidative stress [88]. Aspartic peptidase are necessary for pollen tube elongation and may be necessary for anther dehiscence in flowers infected by *Microbotryum* [72]. Pectin lyases are also known to be involved in pathogenicity [89], as well as glycoside hydrolases [89].

The selective sweeps involved different genomic regions between the two *Microbotryum* species and were scattered along the genome. Moreover, the genes upregulated *in planta*, and those coding putatively secreted proteins did not seem to be clustered. Thus, there does not seem to be "genomic islands" of genes involved in pathogenicity in *Microbotryum*, contrary to what has been observed in other pathogenic fungi [90]. Furthermore, we detected fewer selective sweeps in MvSd than in MvSl. This may reflect a true difference in the number of selective sweeps, or could be due to a lower power for detecting selective sweeps in MvSd because of a lower genetic diversity. Within the selective sweeps, we did not find an enrichment of genes upregulated *in planta*, encoding putatively secreted proteins nor having particular gene ontology. However, it is unlikely that all the genes present in those regions are the target of positive selection. Some interesting putative functions were found for genes in the center of some sweeps, including a CFEM-domain gene upregulated *in planta*, putatively extracellular and membrane-anchored, a cluster of OPT and major facilitator transporters and a secreted lipase. OPT and major facilitator domains are found in membrane transporters, are expanded in *Microbotryum* and could be involved in nutrient uptake [72]. Secreted lipases are also expanded and could be involved in penetration of the plant by the fungi through the wax of the cuticle. Some genes located at the center of sweeps were upregulated *in planta* but did not have any annotation. Fungal putative effectors, such as small secreted proteins upregulated during infection, often lack annotations, which would be related to the fact that they evolve rapidly and that sequence homology is difficult to establish, even between close species [91].

Conclusions

The population genomics approach used here allowed identifying several strong selective sweeps in two *Microbotryum* species, including genes with relevant putative functions and being up-regulated *in planta*. This study may thus finally provide clues on the genes involved in host-plant interaction, that have remained elusive for long in this otherwise well-studied system. The identification of putative effectors will foster future functional and evolutionary studies, in the plant and in the anther smut pathogens, both being model species [35]. This will also be allowed by recent studies on changes in *S. latifolia* gene expression in infected versus healthy plants [72,92]. In addition, this study has even more general interest, showing that selective sweeps are widespread in a natural plant pathogen, while previous genome scans supported a view of neutralist evolution in some organisms, including humans and *Arabidopsis* [11]. Our findings therefore contribute to a broader picture of the occurrence of frequency of selection in natural populations and to the usefulness of whole genome scans for detecting adaptive events.

Materials and Methods

Sample collection, DNA preparation and sequencing

Samples were collected in different locations in Europe (Sup. Table 1). Diploid spores from anthers of a single flower were spread on Petri dishes on PDA medium at 23°C for a few days. On nutritive media, the diploid spores undergo meiosis and replicate clonally. A given flower bears diploid spores from a single individual [93]. Therefore, the harvested haploid sporidia on PDA represented thousands of meiotic products of a single diploid individual. Most sequenced strains were thus diploid. For a subset of the strains a single haploid clone of a given mating type was however isolated for their genome to be sequenced (Table x). For DNA extraction, cells were harvested from PDA medium and stored at -20°C until use. Most DNAs were extracted using the following method: cells were resuspended in a CTAB buffer, frozen in liquid nitrogen, and then crushed with glass beads to break cell walls. Samples were lysed at 60 °C during four hours with a RNase treatment. DNA was then purified with a solution of chloroform-

isoamylalcool (24:1), precipitated in isopropanol and washed twice with ethanol 70%. The dry pellet was resuspended in desionised water (20-40 µl). Some DNAs were extracted using the Macherey-Nagel NucleoSpin Soil kit #740780.250 following manufacturer's instructions and resuspended in desionised water (100 µl). DNA quality was assessed by measuring ratio of 230/260 and 280/260 nm with a NanoDrop 2000 spectrophotometer (Thermo Scientific), and double-strained DNA concentration was measured with a Qubit 2.0 fluorometer. Preparation of DNA libraries and sequencing were performed by Eurofins or on the INRA Genotoul platform. Paired-end libraries of 2x100 bp fragments with an insert size of 300 bp were prepared with Illumina TruSeq Nano DNA Library Prep Kits, and sequencing was performed on a HiSeq2000 Illumina sequencer, at a 100X coverage on average.

Determination of mating-type

Even some strains for which no intentional mating-type selection was done could grow as haploids, as alleles linked to mating types and deleterious in the haploid state are frequent in natural populations [94–96,46]. These alleles can be maintained in natural populations because the haploid state is virtually nonexistent in nature, as intra-tetrad mating is predominant [34,68,46]. Therefore, to determine if the sequenced genomes were diploid or haploid, and in this later case of what mating-type, we assembled each genome with SOAPdenovo and velvet [97,98], and the sequences of the pheromone receptor for a_1 and a_2 mating-types [99] were blasted against these assemblies.

Reads mapping, SNP calling and filtering

Reads were mapped against the almost finished PacBio reference genome of the Lamole strain [61] using the glint software (Faraut T. and Courcelle E.; http://lipm-bioinfo.toulouse.inra.fr/download/glint/, unpublished) with parameter set as follows: minimum length of the hsp >= 90, with ≤ 5 mismatches, no gap allowed, only best-scoring hits taken into account. Pairs that were not properly mapped, ie in expected relative orientation and distance,

were removed. Depending on the mating-type of each strain, the reference used included either one of the mating-type chromosomes, or both.

Variants were called with *VarScan v2.3* [62] with the following parameters : coverage >= 20, variant reads >= 10, average quality >= 30, minor allele frequency >= 0.3. Several filters were applied. Low complexity regions were masked. SNPs close to indels were removed with the *VarScan filter* function. SNPs were filtered on maximum coverage. For this, per base coverage was computed with *bamtools* [100] and coverage histograms were generated with *bedtools* [101]. For each strain and each chromosome, we masked positions whose coverage was higher than five times the peak of the Gaussian distribution. We masked SNPs showing a significant strand bias (Fisher's exact test, p-value 0.05). For subsequent analyses, only bi-allelic SNPs were retained, with a maximum rate of missing data of 0.2.

To assess the quality of SNP calling, we mapped 300X of illumina paired reads of the exact same strain as the Lamole reference strain on the reference genome. We analyzed the proportion of heterozygous SNPs in the haploid strains for sanity check, focusing on autosomes. In subsequent analyses, we only analyzed SNPs located on autosomes.

Genetic structure

To analyze genetic structure in the dataset, we selected all autosomal SNPs, including heterozygous SNPs which could be useful to detect recent hybrids. We performed a pluri-component analysis (PCA) and a discriminant analysis of pluri-component (dAPC) with the R package *adegenet* [102]. *fastStructure* [67] was run for K=1 to K=10 (ten runs per K value). We also performed a dendogram with the neighbor joining method, using the *nj* function of the R package *ape* with default parameters [103]. To find potentially admixed strains that were not detected by these methods, we analyzed the percentage of heterozygous SNPs of each strain in regard with its geographical location. In subsequent population genetic analyses, we excluded potentially admixed strains.

Statistics of population genetics

Our dataset included haploid and diploid strains. Since *Microbotryum* species have a very high rate of selfing, we expected a low frequency of heterozygous SNPs. In fact, except in admixed individuals, the frequency of heterozygous SNPs was similar in haploid and diploid strains, indicating that most of the heterozygous SNPs probably represent artifacts due to copy number variations or misassembled paralogs. Therefore, we only kept homozygous positions and treated all individuals as haploids. Statistics of population genetics were computed with *egglib* [104] and libsequence [105] in coding sequences (CDS). After excluding transposable elements and genes that did not pass quality filters, between 7,059 and 7,594 genes were included in the analyses depending on the cluster considered. Mean synonymous diversity ranged from 0.0003 (MvSd) to 0.0012 (MvSl Italian cluster). Between 3,191 CDS (MvSl eastern cluster) and 5,009 CDS (MvSl Italian cluster) had at least one SNP within clusters. We computed the following statistics within each cluster : nucleotide diversity (Pi) , Theta estimator of Watterson (thetaW) [106], Tajima's D [15], Fay and Wu statistics as standardized by Zeng (Z) [16]. We also measured differentiation between populations with F_{ST} [69], and two measures of divergence, d_{xy}, which represents the mean number of differences between sequences of different populations, and the number of fixed divergence between populations. This was done along sliding windows of 100 and 50 kb. For statistics based on the derived site-frequency spectrum, we used the individual of *M. violaceum sens. lat.* infecting *Lychnis flos-cuculi* for orienting polymorphisms. Diversity, d_{xy} and fixed divergence were normalized by the number of genotyped sites, defined as the number of sites that passed filters for at least 80% of the individuals in a given cluster or pair of clusters.

Linkage disequilibrium (LD) was computed as r^2, the coefficient of correlation between a pair of SNPs, with *rsq* [105], excluding singletons SNPs. To compute a mean LD by window, 10 SNPs were selected randomly from each window of 100 kb and 5 SNPs for each window of 50 kb, and the mean r^2 was computed. This was averaged on 10 random selections to get a value of mean LD per window. LD decay with physical distance was evaluated by fitting the observed r^2 values to the decay function from Hill and Weir [107] with a non-linear model.

Detection of genes under selection and analysis of gene categories

Functional annotation and detection of specific gene categories

Genes were assigned to functional categories using InterProScan v5 [108]. Pfam annotations were used to detect genes of the major facilitator superfamily (PF07690) and sugar transporters (PF00083). Both categories correspond to transmembrane proteins highly represented in in *Microbotryum*, and that may play a role in pathogenesis [72].

Putative secreted proteins were identified as proteins carrying a signal peptide using Phobius [109] or SignalP v4.1 [110] and without any detected transmembrane structures (TMHMM [111]).

Expression data

Perlin et al. [72] generated RNA-seq data in three different conditions: *in vitro* cultures, in low and rich nutrient conditions, respectively, and *in vivo* late infection, corresponding to infection of flower buds by the fungi. Analysis of differential expression by Perlin et al. yielded a list of 1,432 genes that were differentially expressed between at least two conditions, including 307 genes up-regulated during infection compared to both in vitro conditions, 208 up-regulated in low nutrient *in vitro* and 59 up-regulated in rich nutrients (false discovery rate FDR < 0.001). We built a correspondence table between the transcripts assembled by the Broad Institute used by Perlin et al., and the gene models of the PacBio assembly. For this goal, transcripts were mapped using GMAP [112] on the PacBio assembly with default parameters, and the corresponding gene models were retrieved using the BedIntersect program of the bedtools suite [101].

Detection of genes evolving under selection in specific gene categories

For analyses of selection on specific gene categories, we did not take into account the genes annotated as putative transposable elements or genes with fewer than 90% of sites passing filters. One tailed student's tests (FDR < 0.05) were used to detect an increase in PN/PS or DN/DS in specific gene categories. We also performed tests of McDonald and Kreitman and computed the number of non-synonymous and synonymous polymorphism and divergence (PN, PS, DN and DS). Fisher's exact tests were performed and FDR correction applied on p-values to correct for multiple testing.

The neutrality index, defined as (DN/DS)/(PN/PS), was computed by adding one pseudo-count to each class of mutation in the contingency table to assure that NI was defined for all genes. We performed an enrichment analysis for categories of interest (genes up-regulated *in planta*, secreted proteins, major facilitator superfamily) on the tails of the NI distribution (5 % and 95 % quantiles), with one-tailed Fisher exact tests. FDR corrections were applied to correct for multiple testing. The proportion of non-synonymous substitutions fixed by positive selection, alpha, defined as 1 - (PN.DS)/(PS.DN) [113], was estimated by summing PN, DS, PS and DN on all genes.

Determinants of genetic diversity

To assess the influence of several genomic traits on the patterns of genetic diversity along sliding windows, we measured total and synonymous diversity with the Pi and PiS statistics, and tested several explanatory variables: density in CDS, GC content, mean LD and DN/DS. Mean LD was used as a proxy of recombination rate, as no genetic map is available for *Microbotryum*. For the MvSl Italian and eastern clusters, background LD was very strong due to the small numbers of strains, so we used the mean LD in the North-western cluster for all MvSl clusters, assuming similar patterns of recombination rates along the genome among clusters. Correlations between genetic diversity and those variables were first assessed with Spearman tests. We also performed multiple linear regressions. For this, normality of the distributions was visually assessed, and a log-transformation was applied to mean LD and DN/DS. Normality and homoscedasticity of the residuals were visually assessed. All statistical analyses were performed with R version 3.1.0 [114], in non-overlapping windows of 50 and 100 kb. Windows with less than 5 genes or 10 genes passing quality filters in 50 kb or 100kb, respectively, were not taken into account. Using the theta statistics instead of Pi yielded similar results.

Genome Scan for detecting selective Sweeps

Selective sweeps were searched for using sweeD [115], implementing a composite likelihood ratio (CLR) test based on the SweepFinder algorithm [17]. CLR were computed every 10 kb for each contig. Outliers were detected with an R script provided by the developers of sweeD (http://pop-gen.eu/wordpress/software/sweed), using a FDR of 0.05.

Acknowledgements

We thank the GenoToul platform for sequencing. T. Giraud acknowledges the ANR grant ANR-12-ADAP-0009 and the ERC Starting Grant GenomeFun 309403. We thank all the strain collectors (Table X). We thank Olivier Tenaillon for helpful comments on a previous version of the manuscript. We thank Gilles Deparis and Laetitia Giraud for technical help.

We thank Gabriel Marais and Olivier Tenaillon for helpful discussions and Michael Hood for strains and help all along the project.

Author contributions

T. G. designed the study; T. G. and A.B. supervised the study; H.B., A.S. and S.L.P. performed the experiments; H.B., J.G., S.S., G.A. and P.G. analysed the data; H.B. and T.G. wrote the manuscript with contributions by all other authors. T.G. and J.G. contributed to the funding of the study.

Data availability

Numbers for public database accession: ACCESSION NUMBERS. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to tatiana.giraud@upsud.fr.

Bibliography

1. Granka JM, Henn BM, Gignoux CR, Kidd JM, Bustamante CD, Feldman MW. Limited Evidence for Classic Selective Sweeps in African Populations. Genetics. 2012;192: 1049–1064. doi:10.1534/genetics.112.144071

2. Hancock AM, Brachi B, Faure N, Horton MW, Jarymowycz LB, Sperone FG, et al. Adaptation to Climate Across the Arabidopsis thaliana Genome. Science. 2011;334: 83–86. doi:10.1126/science.1209244

3. Jones BL, Raga TO, Liebert A, Zmarz P, Bekele E, Danielsen ET, et al. Diversity of Lactase Persistence Alleles in Ethiopia: Signature of a Soft Selective Sweep. Am J Hum Genet. 2013;93: 538–544. doi:10.1016/j.ajhg.2013.07.008

4. Morris GP, Ramu P, Deshpande SP, Hash CT, Shah T, Upadhyaya HD, et al. Population genomic and genome-wide association studies of agroclimatic traits in sorghum. PNAS. 2013;110: 453–458. doi:10.1073/pnas.1215985110

5. Rubin C-J, Megens H-J, Barrio AM, Maqbool K, Sayyab S, Schwochow D, et al. Strong signatures of selection in the domestic pig genome. PNAS. 2012;109: 19529–19536. doi:10.1073/pnas.1217149109

6. Rubin C-J, Zody MC, Eriksson J, Meadows JRS, Sherwood E, Webster MT, et al. Whole-genome resequencing reveals loci under selection during chicken domestication. Nature. 2010;464: 587–591. doi:10.1038/nature08832

7. Svetec N, Pavlidis P, Stephan W. Recent strong positive selection on Drosophila melanogaster HDAC6, a gene encoding a stress surveillance factor, as revealed by population genomic analysis. Mol Biol Evol. 2009;26: 1549–1556. doi:10.1093/molbev/msp065

8. Tian F, Stevens NM, Buckler ES. Tracking footprints of maize domestication and evidence for a massive selective sweep on chromosome 10. PNAS. 2009;106: 9979–9986. doi:10.1073/pnas.0901122106

9. Udpa N, Ronen R, Zhou D, Liang J, Stobdan T, Appenzeller O, et al. Whole genome sequencing of Ethiopian highlanders reveals conserved hypoxia tolerance genes. Genome Biol. 2014;15: R36.

10. Clark RM, Linton E, Messing J, Doebley JF. Pattern of diversity in the genomic region near the maize domestication gene tb1. PNAS. 2004;101: 700–707. doi:10.1073/pnas.2237049100

11. Hernandez RD, Kelley JL, Elyashiv E, Melton SC, Auton A, McVean G, et al. Classic Selective Sweeps Were Rare in Recent Human Evolution. Science. 2011;331: 920–924. doi:10.1126/science.1198878

12. Lamason RL, Mohideen M-APK, Mest JR, Wong AC, Norton HL, Aros MC, et al. SLC24A5, a Putative Cation Exchanger, Affects Pigmentation in Zebrafish and Humans. Science. 2005;310: 1782–1786. doi:10.1126/science.1116238

13. Nair S, Williams JT, Brockman A, Paiphun L, Mayxay M, Newton PN, et al. A Selective Sweep Driven by Pyrimethamine Treatment in Southeast Asian Malaria Parasites. Mol Biol Evol. 2003;20: 1526–1536. doi:10.1093/molbev/msg162

14. Sattath S, Elyashiv E, Kolodny O, Rinott Y, Sella G. Pervasive Adaptive Protein Evolution Apparent in Diversity Patterns around Amino Acid Substitutions in Drosophila simulans. PLoS Genet. 2011;7: e1001302. doi:10.1371/journal.pgen.1001302

15. Tajima F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics. 1989;123: 585–595.

16. Zeng K, Fu Y-X, Shi S, Wu C-I. Statistical Tests for Detecting Positive Selection by Utilizing High-Frequency Variants. Genetics. 2006;174: 1431–1439. doi:10.1534/genetics.106.061432

17. Nielsen R, Williamson S, Kim Y, Hubisz MJ, Clark AG, Bustamante C. Genomic scans for selective sweeps using SNP data. Genome Res. 2005;15: 1566–1575. doi:10.1101/gr.4252305

18. Yang Z. Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. Mol Biol Evol. 1998;15: 568–573.

19. McDonald JH, Kreitman M. Adaptive protein evolution at the Adh locus in Drosophila. Nature. 1991;351: 652–654. doi:10.1038/351652a0

20. Cutter AD, Payseur BA. Selection at Linked Sites in the Partial Selfer Caenorhabditis elegans. Mol Biol Evol. 2003;20: 665–673. doi:10.1093/molbev/msg072

21. Alföldi J, Di Palma F, Grabherr M, Williams C, Kong L, Mauceli E, et al. The genome of the green anole lizard and a comparative analysis with birds and mammals. Nature. 2011;477: 587–591. doi:10.1038/nature10390

22. Duret L, Eyre-Walker A, Galtier N. A new perspective on isochore evolution. Gene. 2006;385: 71–74. doi:10.1016/j.gene.2006.04.030

23. Glémin S, Clément Y, David J, Ressayre A. GC content evolution in coding regions of angiosperm genomes: a unifying hypothesis. Trends Genet. 2014;30: 263–270. doi:10.1016/j.tig.2014.05.002

24. Marais G. Biased gene conversion: implications for genome and sex evolution. Trends Genet. 2003;19: 330–338. doi:10.1016/S0168-9525(03)00116-1

25. Marsit S, Mena A, Bigey F, Sauvage F-X, Couloux A, Guy J, et al. Evolutionary Advantage Conferred by an Eukaryote-to-Eukaryote Gene Transfer Event in Wine Yeasts. Mol Biol Evol. 2015;32: 1695–1707. doi:10.1093/molbev/msv057

26. Novo M, Bigey F, Beyne E, Galeote V, Gavory F, Mallet S, et al. Eukaryote-to-eukaryote gene transfer events revealed by the genome sequence of the wine yeast Saccharomyces cerevisiae EC1118. PNAS. 2009;106: 16333–16338. doi:10.1073/pnas.0904673106

27. Neafsey DE, Barker BM, Sharpton TJ, Stajich JE, Park DJ, Whiston E, et al. Population genomic sequencing of Coccidioides fungi reveals recent hybridization and transposon control. Genome Res. 2010;20: 938–946. doi:10.1101/gr.103911.109

28. Anderson PK, Cunningham AA, Patel NG, Morales FJ, Epstein PR, Daszak P. Emerging infectious diseases of plants: pathogen pollution, climate change and agrotechnology drivers. Trends Ecol Evol. 2004;19: 535–544. doi:10.1016/j.tree.2004.07.021

29. Stajich JE, Berbee ML, Blackwell M, Hibbett DS, James TY, Spatafora JW, et al. The Fungi. Curr Biol. 2009;19: R840–R845. doi:10.1016/j.cub.2009.07.004

30. Gladieux P, Ropars J, Badouin H, Branca A, Aguileta G, de Vienne DM, et al. Fungal evolutionary genomics provides insight into the mechanisms of adaptive divergence in eukaryotes. Mol Ecol. 2014;23: 753–773. doi:10.1111/mec.12631

31. Giraud T, Gladieux P, Gavrilets S. Linking emergence of fungal plant diseases and ecological speciation. Trends Ecol Evol. 2010;25: 387–395. doi:10.1016/j.tree.2010.03.006

32. Ellison CE, Hall C, Kowbel D, Welch J, Brem RB, Glass NL, et al. Population genomics and local adaptation in wild isolates of a model microbial eukaryote. Proc Natl Acad Sci U S A. 2011;108: 2831–2836. doi:10.1073/pnas.1014971108

33. Fraser BA, Weadick CJ, Janowitz I, Rodd FH, Hughes KA. Sequencing and characterization of the guppy (Poecilia reticulata) transcriptome. BMC Genomics. 2011;12: 202. doi:10.1186/1471-2164-12-202

34. Hood ME, Mena-Ali JI, Gibson AK, Oxelman B, Giraud T, Yockteng R, et al. Distribution of the anther-smut pathogen Microbotryum on species of the Caryophyllaceae. New Phytol. 2010;187: 217–229. doi:10.1111/j.1469-8137.2010.03268.x

35. Bernasconi G, Antonovics J, Biere A, Charlesworth D, Delph LF, Filatov D, et al. Silene as a model system in ecology and evolution. Heredity. 2009;103: 5–14. doi:10.1038/hdy.2009.34

36. Antonovics J, Alexander HM. Epidemiology of anther-smut infection of Silene alba (= S. latifolia) caused by Ustilago violacea: patterns of spore deposition in experimental populations. Proc R Soc Lond B Biol Sci. 1992;250: 157–163.

37. Thrall PH, Biere A, Antonovics J. Plant Life-History and Disease Susceptibility–The Occurrence of Ustilago Violacea on Different Species within the Caryophyllaceae. J Ecol. 1993;81: 489–498. doi:10.2307/2261527

38. Biere A, Antonovics J. Sex-specific costs of resistance to the fungal pathogen Ustilago violacea (Microbotryum violaceum) in Silene alba. Evolution. 1996;50: 1098–1110. doi:10.2307/2410650

39. Shykoff J, Bucheli E. Pollinator Visitation Patterns, Floral Rewards and the Probability of Transmission of Microbotryum-Violaceum, a Venereal-Disease of Plants. J Ecol. 1995;83: 189–198. doi:10.2307/2261557

40. Biere A, Honders SC. Anther Smut Transmission in Silene latifolia and Silene dioica: Impact of Host Traits, Disease Frequency, and Host Density. Int J Plant Sci. 1998;159: 228–235.

41. Carlsson-Granér U, Thrall PH. The spatial distribution of plant populations, disease dynamics and evolution of resistance. Oikos. 2002;97: 97–110. doi:10.1034/j.1600-0706.2002.970110.x

42. De VIENNE DM, Hood ME, Giraud T. Phylogenetic determinants of potential host shifts in fungal pathogens. J Evol Biol. 2009;22: 2532–2541. doi:10.1111/j.1420-9101.2009.01878.x

43. Le Gac M, Hood ME, Fournier E, Giraud T. Phylogenetic Evidence of Host-Specific Cryptic Species in the Anther Smut Fungus. Evolution. 2007;61: 15–26. doi:10.1111/j.1558-5646.2007.00002.x

44. De Vienne DM, Refrégier G, Hood ME, Guigue A, Devier B, Vercken E, et al. Hybrid sterility and inviability in the parasitic fungal species complex Microbotryum. J Evol Biol. 2009;22: 683–698. doi:10.1111/j.1420-9101.2009.01702.x

45. Le Gac M, Hood ME, Giraud T. Evolution of reproductive isolation within a parasitic fungal species complex. Evol Int J Org Evol. 2007;61: 1781–1787. doi:10.1111/j.1558-5646.2007.00144.x

46. Giraud T, Yockteng R, Lopez-Villavicencio M, Refregier G, Hood ME. Mating system of the anther smut fungus Microbotryum violaceum: Selfing under heterothallism. Eukaryot Cell. 2008;7: 765–775. doi:10.1128/EC.00440-07

47. Cafuir L, Antonovics Janis, Hood ME. Tissue Culture and Quantification of Individual-Level Resistance to Anther-Smut Disease in Silene vulgaris. Int J Plant Sci. 2007;168: 415–419. doi:10.1086/511754

48. Carlsson-Granér U. Anther-Smut Disease in Silene dioica: Variation in Susceptibility Among Genotypes and Populations, and Patterns of Disease Within Populations. Evolution. 1997;51: 1416–1426. doi:10.2307/2411194

49. Alexander HM, Antonovics J, Kelly AW. Genotypic Variation in Plant Disease Resistance–Physiological Resistance in Relation to Field Disease Transmission. J Ecol. 1993;81: 325–333. doi:10.2307/2261502

50. Kaltz O, Gandon S, Michalakis Y, Shykoff JA. Local maladaptation in the anther-smut fungus Microbotryum violaceum to its host plant Silene latifolia: Evidence from a cross-inoculation experiment. Evolution. 1999;53: 395–407. doi:10.2307/2640776

51. Ashby B, Gupta S. Parasitic castration promotes coevolutionary cycling but also imposes a cost on sex. Evolution. 2014;68: 2234–2244. doi:10.1111/evo.12425

52. Tellier A, Moreno-Gámez S, Stephan W. Speed of Adaptation and Genomic Footprints of Host–Parasite Coevolution Under Arms Race and Trench Warfare Dynamics. Evolution. 2014;68: 2211–2224. doi:10.1111/evo.12427

53. Gladieux P, Vercken E, Fontaine MC, Hood ME, Jonot O, Couloux A, et al. Maintenance of Fungal Pathogen Species That Are Specialized to Different Hosts: Allopatric Divergence and Introgression through Secondary Contact. Mol Biol Evol. 2011;28: 459–471. doi:10.1093/molbev/msq235

54. Gibson AK, Refrégier G, Hood Michael E., Giraud T. Performance of a Hybrid Fungal Pathogen on Pure-Species and Hybrid Host Plants. Int J Plant Sci. 2014;175: 724–730. doi:10.1086/676621

55. Van Putten WF, Biere A, Van Damme JMM, May G. Intraspecific competition and mating between fungal strains of the anther smut microbotryum violaceum from the host plants silene latifolia and s. dioica. Evolution. 2003;57: 766–776. doi:10.1554/0014-3820(2003)057[0766:ICAMBF]2.0.CO;2

56. Vercken E, Fontaine MC, Gladieux P, Hood ME, Jonot O, Giraud T. Glacial Refugia in Pathogens: European Genetic Structure of Anther Smut Pathogens on Silene latifolia and Silene dioica. PLoS Pathog. 2010;6: e1001229. doi:10.1371/journal.ppat.1001229

57. Refrégier G, Hood ME, Giraud T. No Evidence of Reproductive Character Displacement between Two Sister Fungal Species Causing Anther Smut Disease in Silene. Int J Plant Sci. 2010;171: 847–859. doi:10.1086/655867

58. Giraud T, Yockteng R, Marthey S, Chiapello H, Jonot O, Lopez-Villavicencio M, et al. PERMANENT GENETIC RESOURCES: Isolation of 60 polymorphic microsatellite loci in EST libraries of four sibling species of the phytopathogenic fungal complex Microbotryum. Mol Ecol Resour. 2008;8: 387–392. doi:10.1111/j.1471-8286.2007.01967.x

59. Gibson AK, Hood ME, Giraud T. Sibling competition arena: selfing and a competition arena can combine to constitute a barrier to gene flow in sympatry. Evol Int J Org Evol. 2012;66: 1917–1930. doi:10.1111/j.1558-5646.2011.01563.x

60. Huber CD, Nordborg M, Hermisson J, Hellmann I. Keeping It Local: Evidence for Positive Selection in Swedish Arabidopsis thaliana. Mol Biol Evol. 2014;31: 3026–3039. doi:10.1093/molbev/msu247

61. Badouin H, Hood ME, Gouzy J, Aguileta G, Siguenza S, Perlin MH, et al. Chaos of Rearrangements in the Mating-Type Chromosomes of the Anther-Smut Fungus Microbotryum lychnidis-dioicae. Genetics. 2015; genetics.115.177709. doi:10.1534/genetics.115.177709

62. Koboldt DC, Zhang Q, Larson DE, Shen D, McLellan MD, Lin L, et al. VarScan 2: Somatic mutation and copy number alteration discovery in cancer by exome sequencing. Genome Res. 2012;22: 568–576. doi:10.1101/gr.129684.111

63. Hood ME. Dimorphic mating-type chromosomes in the fungus Microbotryum violaceum. Genetics. 2002;160: 457–461.

64. Hood ME, Petit E, Giraud T. Extensive Divergence Between Mating-Type Chromosomes of the Anther-Smut Fungus. Genetics. 2013;193: 309–315. doi:10.1534/genetics.112.146266

65. Hood ME, Antonovics J. Mating within the meiotic tetrad and the maintenance of genomic heterozygosity. Genetics. 2004;166: 1751–1759.

66. Rödelsperger C, Neher RA, Weller AM, Eberhardt G, Witte H, Mayer WE, et al. Characterization of Genetic Diversity in the Nematode Pristionchus pacificus from Population-Scale Resequencing Data. Genetics. 2014;196: 1153–1165. doi:10.1534/genetics.113.159855

67. Raj A, Stephens M, Pritchard JK. fastSTRUCTURE: Variational Inference of Population Structure in Large SNP Data Sets. Genetics. 2014;197: 573–589. doi:10.1534/genetics.114.164350

68. Giraud T, Jonot Odile, Shykoff JA. Selfing Propensity under Choice Conditions in a Parasitic Fungus, Microbotryum violaceum, and Parameters Influencing Infection Success in Artificial Inoculations. Int J Plant Sci. 2005;166: 649–657. doi:10.1086/430098

69. Hudson RR, Slatkin M, Maddison WP. Estimation of levels of gene flow from DNA sequence data. Genetics. 1992;132: 583–589.

70. Takahata N, Nei M. Gene Genealogy and Variance of Interpopulational Nucleotide Differences. Genetics. 1985;110: 325–344.

71. Hill WG, Robertson A. Linkage disequilibrium in finite populations. Theor Appl Genet. 1968;38: 226–231. doi:10.1007/BF01245622

72. Perlin MH, Amselem J, Fontanillas E, Toh SS, Chen Z, Goldberg J, et al. Sex and parasites: genomic and transcriptomic analysis of Microbotryum lychnidis-dioicae, the biotrophic and plant-castrating anther smut fungus. BMC Genomics. 2015;16. doi:10.1186/s12864-015-1660-8

73. Stukenbrock EH, McDonald BA. Population Genetics of Fungal and Oomycete Effectors Involved in Gene-for-Gene Interactions. MPMI. 2009;22: 371–380. doi:10.1094/MPMI-22-4-0371

74. Aguileta G, Lengelle J, Marthey S, Chiapello H, Rodolphe F, Gendrault A, et al. Finding candidate genes under positive selection in Non-model species: examples of genes involved in host specialization in pathogens. Mol Ecol. 2010;19: 292–306. doi:10.1111/j.1365-294X.2009.04454.x

75. Charlesworth B, Morgan MT, Charlesworth D. The effect of deleterious mutations on neutral molecular variation. Genetics. 1993;134: 1289–1303.

76. Nabholz B, Sarah G, Sabot F, Ruiz M, Adam H, Nidelet S, et al. Transcriptome population genomics reveals severe bottleneck and domestication cost in the African rice (Oryza glaberrima). Mol Ecol. 2014;23: 2210–2227. doi:10.1111/mec.12738

77. Payseur BA, Nachman MW. Gene Density and Human Nucleotide Polymorphism. Mol Biol Evol. 2002;19: 336–340.

78. Tiley GP, Burleigh G. The relationship of recombination rate, genome structure, and patterns of molecular evolution across angiosperms. BMC Evol Biol. 2015;15: 194. doi:10.1186/s12862-015-0473-3

79. Croll D, Lendenmann MH, Stewart E, McDonald BA. The Impact of Recombination Hotspots on Genome Evolution of a Fungal Plant Pathogen. Genetics. 2015; genetics.115.180968. doi:10.1534/genetics.115.180968

80. Goulson D, Jerrim K. Maintenance of the Species Boundary between Silene dioica and S. latifolia (Red and White Campion). Oikos. 1997;79: 115–126. doi:10.2307/3546096

81. Putten WF van, Elzinga Jelmer A., Biere A. Host Fidelity of the Pollinator Guilds of Silene dioica and Silene latifolia: Possible Consequences for Sympatric Host Race Differentiation of a Vectored Plant Disease. Int J Plant Sci. 2007;168: 421–434. doi:10.1086/511050

82. Burri R, Nater A, Kawakami T, Mugal CF, Olason PI, Smeds L, et al. Linked selection and recombination rate variation drive the evolution of the genomic landscape of differentiation across the speciation continuum of Ficedula flycatchers. Genome Res. 2015; doi:10.1101/gr.196485.115

83. Manthey JD, Klicka J, Spellman GM. Chromosomal patterns of diversity and differentiation in creepers: a next-gen phylogeographic investigation of Certhia americana. Hered Edinb. 2015;115: 165–172. doi:10.1038/hdy.2015.27

84. Lesecque Y, Mouchiroud D, Duret L. GC-Biased Gene Conversion in Yeast Is Specifically Associated with Crossovers: Molecular Mechanisms and Evolutionary Significance. Mol Biol Evol. 2013;30: 1409–1419. doi:10.1093/molbev/mst056

85. Duret L, Galtier N. Biased Gene Conversion and the Evolution of Mammalian Genomic Landscapes. Annu Rev Genomics Hum Genet. 2009;10: 285–311. doi:10.1146/annurev-genom-082908-150001

86. Kulkarni RD, Thon MR, Pan H, Dean RA. Novel G-protein-coupled receptor-like proteins in the plant pathogenic fungus Magnaporthe grisea. Genome Biol. 2005;6: R24. doi:10.1186/gb-2005-6-3-r24

87. Leuthner B, Aichinger C, Oehmen E, Koopmann E, Müller O, Müller P, et al. A H2O2-producing glyoxal oxidase is required for filamentous growth and pathogenicity in Ustilago maydis. Mol Genet Genomics. 2005;272: 639–650. doi:10.1007/s00438-004-1085-6

88. Festa RA, Thiele DJ. Copper at the Front Line of the Host-Pathogen Battle. PLoS Pathog. 2012;8: e1002887. doi:10.1371/journal.ppat.1002887

89. Ma Z, Song T, Zhu L, Ye W, Wang Y, Shao Y, et al. A Phytophthora sojae Glycoside Hydrolase 12 Protein Is a Major Virulence Factor during Soybean Infection and Is Recognized as a PAMP. Plant Cell. 2015;27: 2057–2072. doi:10.1105/tpc.15.00390

90. Rouxel T, Grandaubert J, Hane JK, Hoede C, van de Wouw AP, Couloux A, et al. Effector diversification within compartments of the Leptosphaeria maculans genome affected by Repeat-Induced Point mutations. Nat Commun. 2011;2: 202. doi:10.1038/ncomms1189

91. Stergiopoulos I, de Wit PJGM. Fungal Effector Proteins. Annu Rev Phytopathol. 2009;47: 233–263. doi:10.1146/annurev.phyto.112408.132637

92. Zemp N, Tavares R, Widmer A. Fungal Infection Induces Sex-Specific Transcriptional Changes and Alters Sexual Dimorphism in the Dioecious Plant Silene latifolia. PLoS Genet. 2015;11: e1005536. doi:10.1371/journal.pgen.1005536

93. López-Villavicencio M, Jonot O, Coantic A, Hood ME, Enjalbert J, Giraud T. Multiple Infections by the Anther Smut Pathogen Are Frequent and Involve Related Strains. PLoS Pathog. 2007;3: e176. doi:10.1371/journal.ppat.0030176

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

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

96. Oudemans PV, Alexander HM, Antonovics J, Altizer S, Thrall PH, Rose L. The Distribution of Mating-Type Bias in Natural Populations of the Anther-Smut Ustilago violacea on Silene alba in Virginia. Mycologia. 1998;90: 372–381. doi:10.2307/3761395

97. Luo R, Liu B, Xie Y, Li Z, Huang W, Yuan J, et al. SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. Gigascience. 2012;1: 18.

98. Zerbino DR, Birney E. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res. 2008;18: 821–829. doi:10.1101/gr.074492.107

99. Devier B, Aguileta G, Hood ME, Giraud T. Ancient Trans-specific Polymorphism at Pheromone Receptor Genes in Basidiomycetes. Genetics. 2009;181: 209–223. doi:10.1534/genetics.108.093708

100. Barnett DW, Garrison EK, Quinlan AR, Strömberg MP, Marth GT. BamTools: a C++ API and toolkit for analyzing and managing BAM files. Bioinformatics. 2011;27: 1691–1692. doi:10.1093/bioinformatics/btr174

101. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 2010;26: 841–842. doi:10.1093/bioinformatics/btq033

102. Jombart T. adegenet: a R package for the multivariate analysis of genetic markers. Bioinformatics. 2008;24: 1403–1405. doi:10.1093/bioinformatics/btn129

103. Paradis E, Claude J, Strimmer K. APE: analyses of phylogenetics and evolution in R language. Bioinformatics. 2004;20: 289–290.

104. Mita SD, Siol M. EggLib: processing, analysis and simulation tools for population genetics and genomics. BMC Genet. 2012;13: 27. doi:10.1186/1471-2156-13-27

105. Thornton K. Libsequence: a C++ class library for evolutionary genetic analysis. Bioinformatics. 2003;19: 2325–2327.

106. Watterson GA. On the number of segregating sites in genetical models without recombination. Theor Popul Biol. 1975;7: 256–276. doi:10.1016/0040-5809(75)90020-9

107. Hill WG, Weir BS. Variances and covariances of squared linkage disequilibria in finite populations. Theor Popul Biol. 1988;33: 54–78.

108. Zdobnov EM, Apweiler R. InterProScan–an integration platform for the signature-recognition methods in InterPro. Bioinformatics. 2001;17: 847–848.

109. Käll L, Krogh A, Sonnhammer ELL. A Combined Transmembrane Topology and Signal Peptide Prediction Method. J Mol Biol. 2004;338: 1027–1036. doi:10.1016/j.jmb.2004.03.016

110. Petersen TN, Brunak S, von Heijne G, Nielsen H. SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat Meth. 2011;8: 785–786. doi:10.1038/nmeth.1701

111. Krogh A, Larsson B, von Heijne G, Sonnhammer EL. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J Mol Biol. 2001;305: 567–580. doi:10.1006/jmbi.2000.4315

112. Wu TD, Watanabe CK. GMAP: a genomic mapping and alignment program for mRNA and EST sequences. Bioinformatics. 2005;21: 1859–1875. doi:10.1093/bioinformatics/bti310

113. Rand DM, Kann LM. Excess amino acid polymorphism in mitochondrial DNA: contrasts among genes from Drosophila, mice, and humans. Mol Biol Evol. 1996;13: 735–748.

114. R Core Team. R: A Language and Environment for Statistical Computing [Internet]. Vienna, Austria: R Foundation for Statistical Computing; 2014. Available: http://www.R-project.org

115. Pavlidis P, Živković D, Stamatakis A, Alachiotis N. SweeD : Likelihood-Based Detection of Selective Sweeps in Thousands of Genomes. Mol Biol Evol. 2013;30: 2224–2234. doi:10.1093/molbev/mst112

Tables:

Table 1: Composition and polymorphism of clusters; number of individuals, number of SNPs, median per-base total nucleotidic diversity and median Tajima's D in non-overlapping 100 kb windows.

		M. lychnidis-dioicae		
	M. silenes-dioicae	North-West	Italy	East
Number of individuals	19	15	7	5
Number of homozygous SNPs	28000	50000	63000	32000
Pi per bp	2.22E-04	6.33E-04	1.10E-03	6.30E-04
Tajima's D	-1.03	0.0166	0.208	-0.2

-Table 2: Pairwise F_{ST} and number of fixed differences per bp between clusters of ***Microbotryum lychnidis-dioicae*** **(MvSl) and *M. silenes-dioicae* (MvSd).**

pairwise comparison	F_{ST}	Number of fixed differences (per bp)
MvSd and North-western MvSl	0.936	0.0044
MvSd and Italian MvSl	0.948	0.00415
MvSd and eastern MvSl	0.961	0.00456
North-western and Italian MvSl	0.725	0.000796
North-western and eastern MvSl	0.744	0.0011
Italian and eastern MvSl	0.562	0.000724

-Table 3: Multiple linear regressions models for explaining total diversity in the clusters of *Microbotryum lychnidis-dioicae* (MvSl) and in *M. silenes-dioicae* (MvSd), using 50 kb and 100 kb non-overlapping sliding windows. P-values : * < 0.05, ** < 0.01, *** < 0.001

Window size	Cluster	Multiple R-squared	GC content	Linkage disequilibrium	Density in coding sites	DN/DS
100 kb	North-western MvSl	0.1449	-0.00007484*	-0.0002069***	3.02E-04	-5.35E-05
	Italian MvSl	0.2711	-0.0001456***	-0.0003662***	5.61E-04	-6.56E-05
	eastern MvSl	0.08304	-7.59E-05	-0.000223**	-2.95E-06	5.56E-05
	MvSd	0.1548	-8.60E-05	-2.30E-05	-7.49E-05	-2.31E-05
50 kb	North-western MvSl	0.08747	-0.00009373***	-0.0001233***	2.40E-04	2.12E-05
	Italian MvSl	0.2131	-0.0001685***	-0.0003208***	3.09E-04	-5.92E-06
	eastern MvSl	0.08962	-0.0001109***	-0.0002096***	2.52E-05	-4.43E-05
	MvSd	0.1161	-0.00006061***	3.87E-05	-1.99E-04	-1.72E-07

Table 4: Genes up-regulated *in planta* with neutrality index values within the 5% lowest quantile between *Microbotryum silenes-dioicae* and in *M. lychnidis-dioicae* clusters. Only genes with Pfam annotations are included.

geneID	secreted	Pfam	NI MvSd - Italian MvSl	NI MvSd – North-western MvSl	NI MvSd – eastern MvSl
MvSlA1A2r3c_S01g02320	no	PF13813:MBOAT	0.200	0.429	0.400
MvSlA1A2r3c_S01g02877	no	PF07690:MFS	0.267	0.133	0.185
MvSlA1A2r3c_S01g03130	no	PF06687:SUR7	0.400	0.150	0.133
MvSlA1A2r3c_S01g03137	no	PF00735:Septin	0.167	0.500	0.250
MvSlA1A2r3c_S01g03169	no	PF00069:Pkinase	0.133	0.222	0.222
MvSlA1A2r3c_S02g03488	no	PF00067:p450	0.667	0.188	0.500
MvSlA1A2r3c_S02g03492	no	PF07690:MFS	0.250	0.250	0.100
MvSlA1A2r3c_S02g03550	no	PF13668:Ferritin	0.286	0.750	0.167
MvSlA1A2r3c_S02g03711	yes	PF05730:CFEM	0.235	0.111	0.250
MvSlA1A2r3c_S03g04914	no	PF03556:Cullin_binding	1.000	1.000	0.167
MvSlA1A2r3c_S04g05345	yes	PF07732:Cu-oxidase	0.417	0.167	0.143
MvSlA1A2r3c_S06g07277	no	PF00149:Metallophos	0.222	0.360	0.042
MvSlA1A2r3c_S07g07393	no	PF08030:NAD_binding	0.100	0.033	0.296
MvSlA1A2r3c_S07g07622	no	PF08241:Methyltransf_11	0.333	0.800	0.111
MvSlA1A2r3c_S07g07824	yes	PF00026:Asp	0.123	0.405	0.143
MvSlA1A2r3c_S08g07940	no	PF00083:Sugar_tr	0.125	0.125	0.100
MvSlA1A2r3c_S08g08275	no	PF01062:Bestrophin	0.375	0.083	0.500
MvSlA1A2r3c_S08g08431	no	PF13465:zf-H2C2	0.165	0.421	0.194
MvSlA1A2r3c_S09g08688	no	PF05730:CFEM	0.083	0.125	0.125
MvSlA1A2r3c_S09g08902	no	PF07250:Glyoxal_oxid_N	0.333	0.200	0.889
MvSlA1A2r3c_S10g09040	no	PF00642:zf-CCCH	0.364	0.078	0.455
MvSlA1A2r3c_S10g09280	no	PF03169:OPT	0.225	0.133	0.200
MvSlA1A2r3c_S14g10582	no	PF08240:ADH_N;PF0010	0.167	0.167	0.500
MvSlA1A2r3c_S16g10956	no	PF00107:ADH_zinc_N	0.167	0.200	1.000

Supplementary results

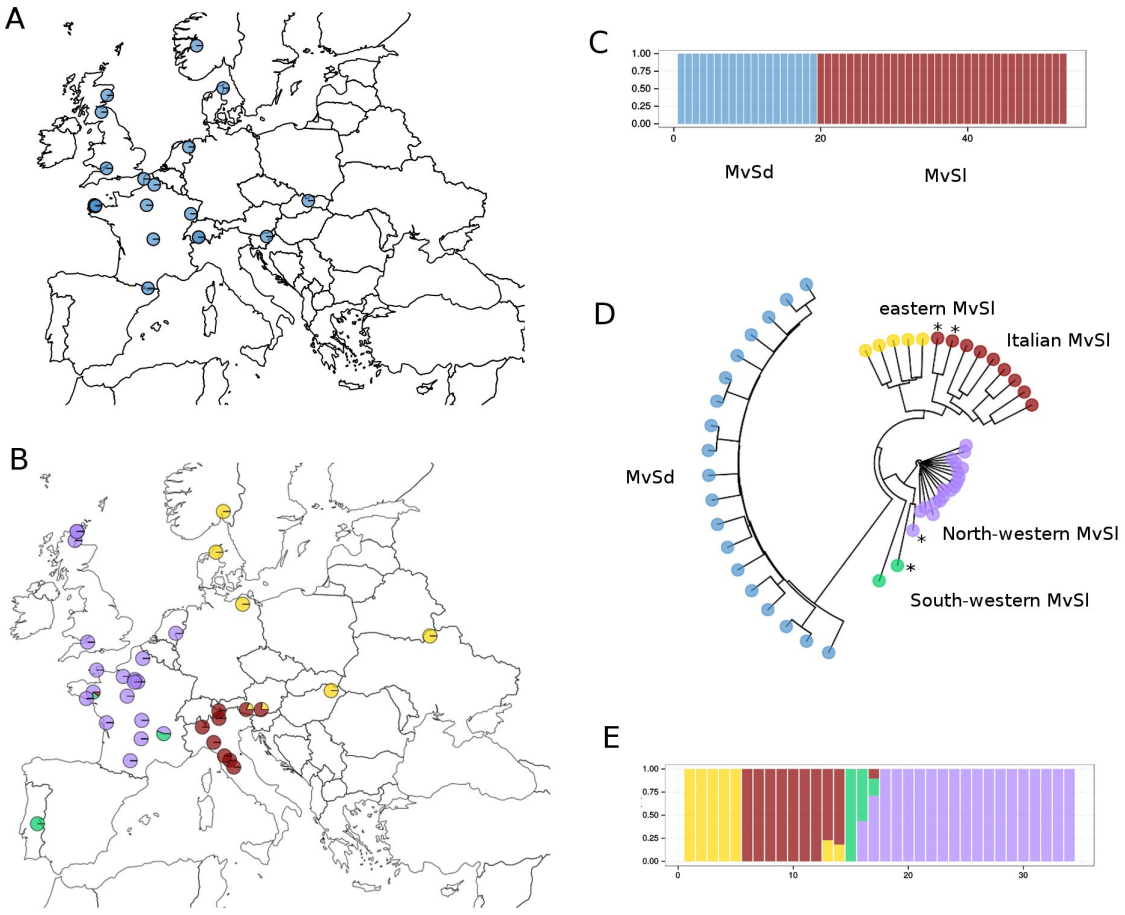
Looking for footprints of selection

In addition to the genes upregulated in planta presented in the main text, we also looked for signs of positive selection in additional gene categories that were thought a priori to possibly include genes involved in host-pathogen interaction, *i.e.*, i) genes encoding secreted proteins, that may serve as effectors during plant infection, in particular, but not only, the small secreted proteins {Stergiopoulos et De Wit, 2009; Sperschneider et al, 2015}, and ii) genes belonging to functional categories that are highly represented in the *Microbotryum* genome and may play a role in pathogenicity, *i.e.*, the major facilitator superfamily (PF07690) and sugar transporters (PF00083) (Perlin et al., 2015).

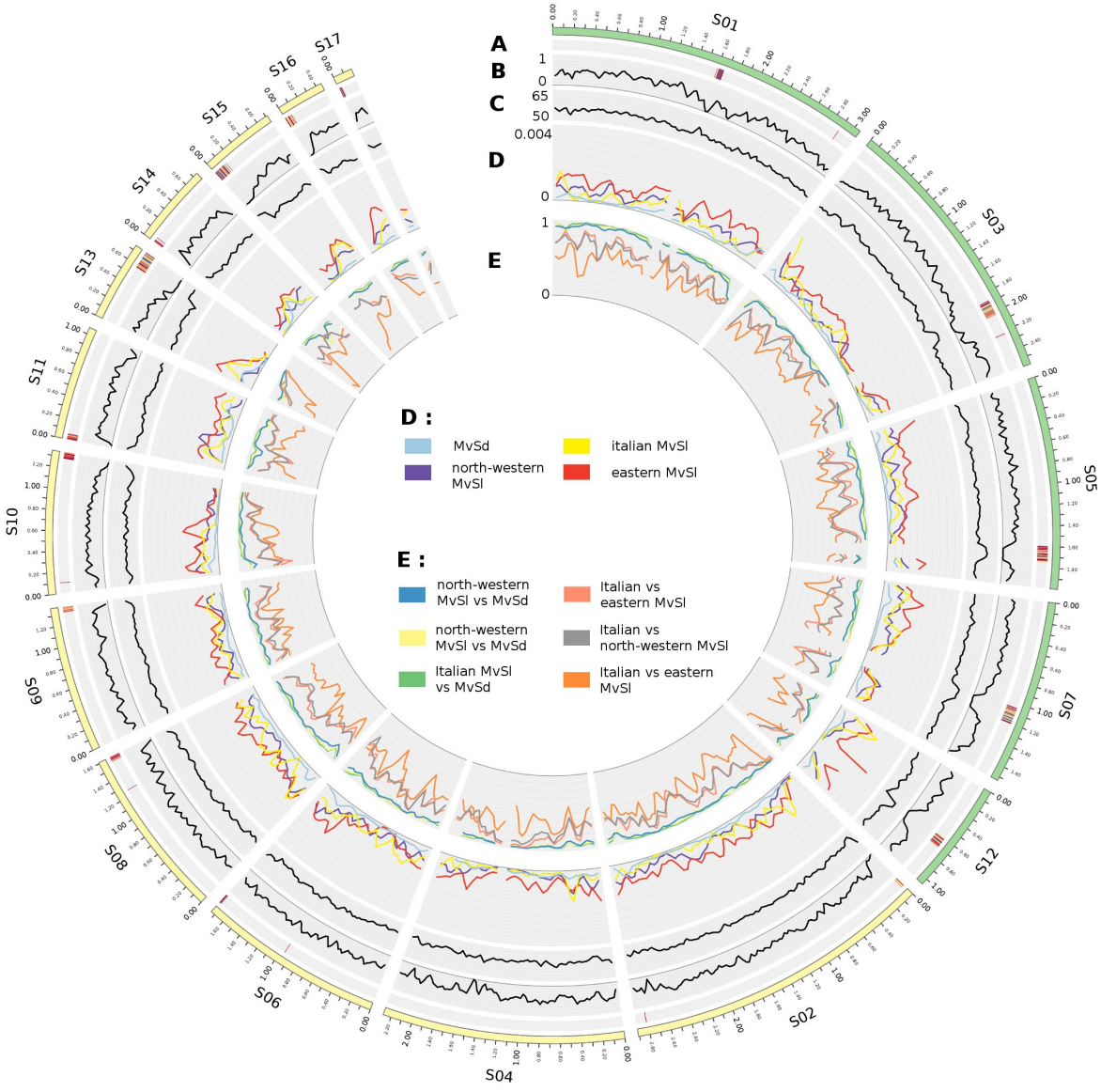
The genes belonging to the major facilitator superfamily showed an elevated DN/DS, although this was only significant between North-western MvSl and MvSd after FDR (mean = 0.60 versus 0.51 in other genes, FDR = 0.031). Sugar transporters had lower DN/DS and P_N/P_S than the remaining gene set, but not significantly so. We then examined genes encoding secreted proteins, 185 of which were predicted on autosomes. The total diversity was significantly higher in secreted proteins than in other genes in North-western MvSl and in MvSd (mean = 7.602e-4 in SP versus 4.966e-4 in other genes in North-western MvSl, False discovery rate (FDR) = 0.033 Mann-Whitney test). P_N/P_S and DN/DS were slightly higher in secreted proteins than in other genes, but not significantly so. When restricting the analysis to the 65 small secreted proteins (shorter than 250 amino acids), we could not detect any significant increase in diversity either.

When looking at the neutrality index, none of those genes categories were enriched in the tails of the distribution, contrary to the genes up-regulated *in planta* which were enriched in low NI genes.

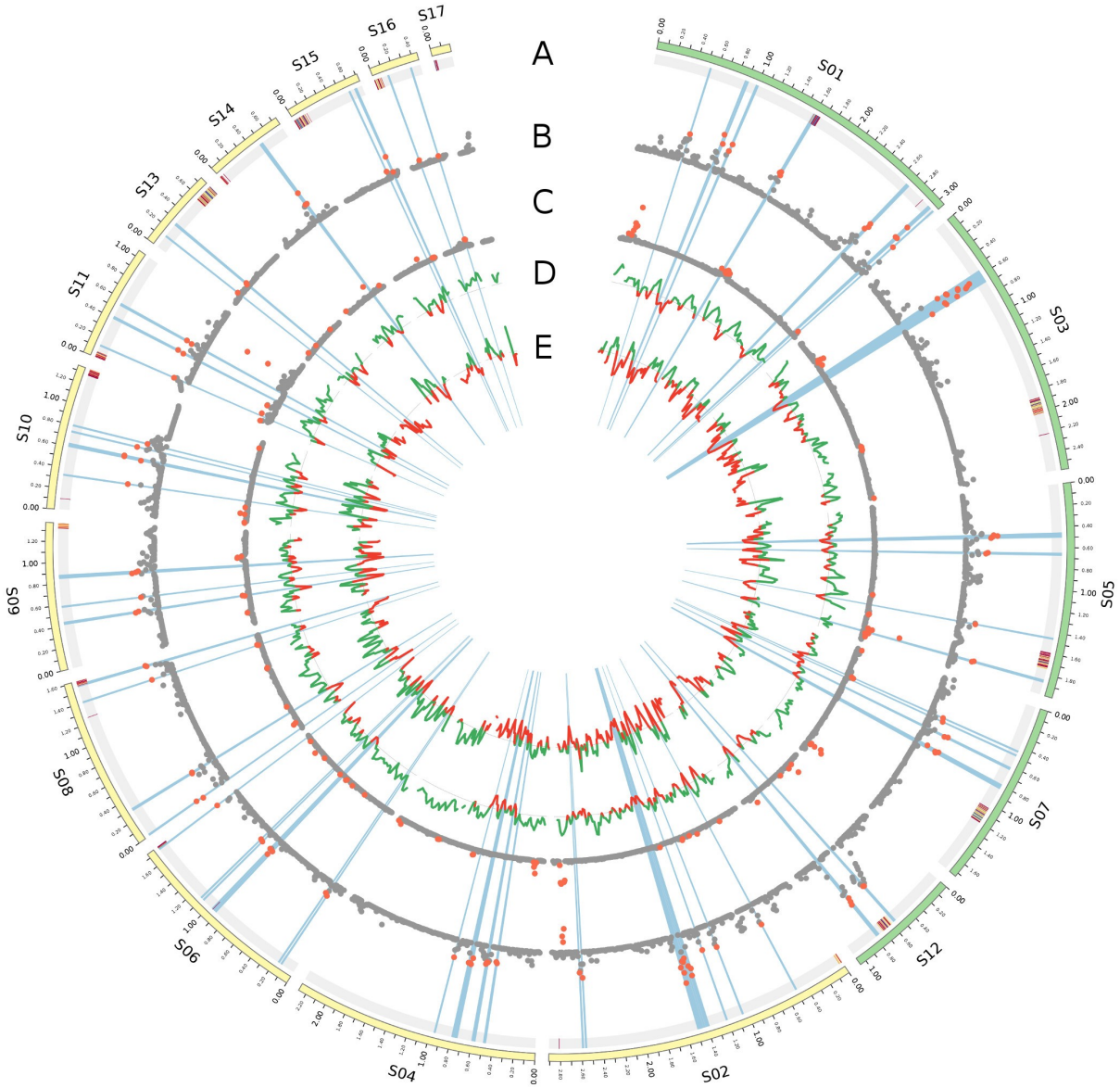
-Figure 1: Population genetic structure in *Microbotryum lychnidis-dioicae* (MvSl) and *M. silenes-dioicae* (MvSd). A) Map of sampled individuals in MvSd, in which no genetic subdivision could be detected. B) Map of samples individuals in MvSl, with cluster membership for K=4 represented by colors. C) Cluster membership for K=2 in the structure analysis on all MvSl and MvSd individuals, represented as vertical bars. D) Neighbor-joining tree of MvSl and MvSd individuals. E) Cluster membership for K=4 within MvSl, in the structure analysis on MvSl individuals, represented as vertical bars.



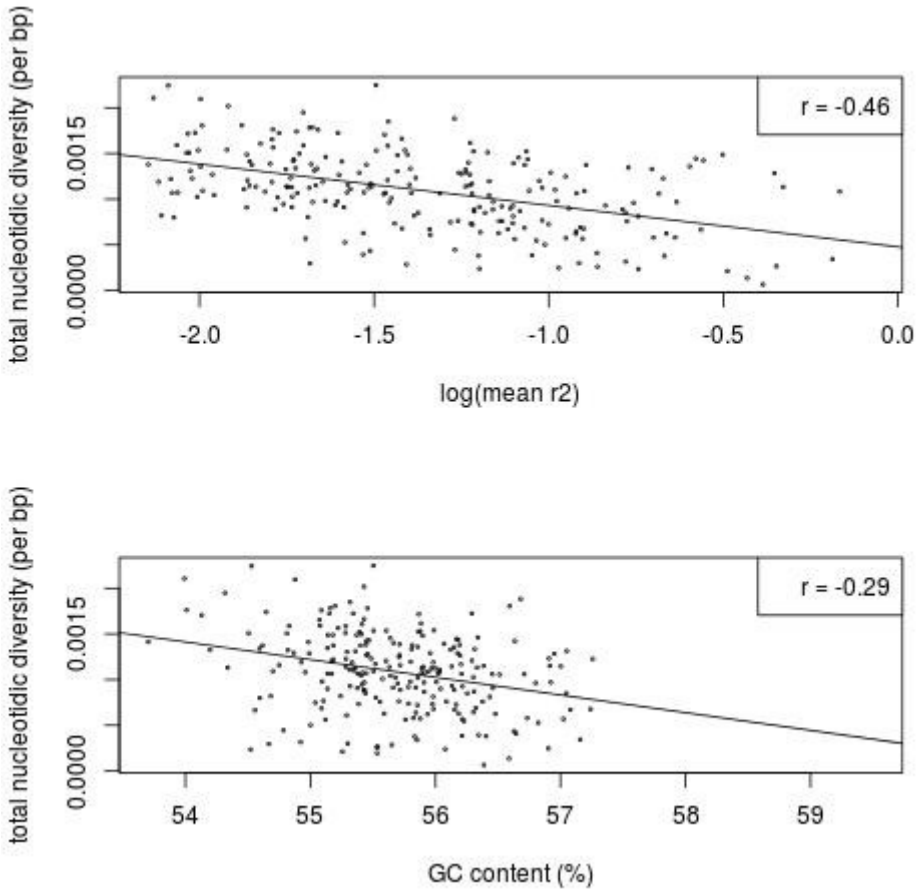
-Figure 2 : Nucleotide diversity (Pi) and Fst along the genome in *Microbotryum lychnidis-dioicae* (MvSl) clusters and in *Microbotryum silenes-dioicae* (MvSd). A : location of centromeric repeats. B : Density in coding sites. C : GC content (%). D : Pi on 100 kb sliding-window. E – Pairwise Fst.



-Figure 3: Composite likelihood ratio tests along the genome in *Microbotryum lychnidis-dioicae* north-western cluster and in *Microbotryum silenes-dioicae.* A : location of centromeric repeats. B : CLR in north-western MvSl, with outliers in red and sweeps highlighted in blue. C : CLR in MvSd. D : Fay and Wu's H in north-western MvSl. E : Tajima's D in north-western MvSl. Positive values are in green and negative ones in red.



-Figure 4: Correlations between nucleotidic diversity and A) log of mean linkage disequilibrium (r^2) or B) GC content, in the Italian cluster of *Microbotryum lychnidis-dioicae*, using 100 kb windows. The regression lines are drawn in black. Spearman's r coefficients are indicated.



4.5 Discussion et Perspectives

Synthèse des résultats

Dans ce travail sur l'adaptation chez *M. violaceum*, nous avons dégagé des caractéristiques du génome de *M. lychnidis-dioicae* qui sont en lien avec le mode de vie biotrophe obligatoire de ce champignon. Nous avons montré que le génome de *M. lychnidis-dioicae* possédait un répertoire d'effecteurs potentiels sous forme de petites protéines sécrétées. Ces effecteurs sont dispersés dans le génome, contrairement à d'autres champignons biotrophes comme *Leptosphaeria maculans* où les effecteurs sont regroupés dans des régions particulières [120], mais on trouve de petits clusters d'effecteurs, notamment du fait de duplications en tandem, et les petites protéines sécrétées sont souvent trouvées à proximité d'éléments transposables.

Certaines familles de gènes ont subi des expansions, comme les glycosyde hydrolases, qui pourraient jouer un rôle dans le remodelage de la paroi cellulaire du champignon, lui permettant d'échapper au système immunitaire de la plante, ou de manipuler le développement de la plante. Des familles de transporteurs qui pourraient être impliquées dans l'assimilation de nutriments ont également subi des expansions, comme chez d'autres champignons biotrophes [99]. Les gènes codant pour des enzymes de dégradation de la paroi cellulaire de la plante sont également absents ou présents en moins grand nombre que chez d'autres basidiomycètes non biotrophes. Des lipases sécrétées pourraient jouer un rôle dans la perception des lipides à la surface des méristèmes floraux.

Notre étude de génomique des populations a révélé l'absence d'introgressions entre *M. lychnidis-dioicae* et *M. silenes-dioicae*. Cela n'exclut pas la présence de quelques hybrides dans les populations naturelles [23], mais cette hybridation ne donne pas lieu à une introgression pérenne entre *M. lychnidis-dioicae* et *M. silenes-dioicae*. L'absence d'introgression pourrait être liée à la présence d'incompatibilités de type Bateson-Dobzhansky-Muller ou à une forte pression de sélection contre les hybrides sur tout le génome, les effecteurs potentiels étant distribués sur tout le génome chez *M. lychnidis-dioicae*, et des balayages sélectifs étant détectés dans de multiples régions du génome chez *M. lychnidis-dioicae*. Dans les balayages sélectifs, plusieurs gènes candidats ressortent particulièrement, comme un cluster de lipases sécrétées et un cluster de transporteurs de type OPT (famille en expansion chez *M. lychnidis-dioicae*. On trouve aussi plusieurs petites protéines fortement sur-exprimées dans la plante qui pourraient agir comme

effecteurs pendant l'infection. Nous avons de plus observé que les gènes surexprimés dans la plante chez *M. lychnidis-dioicae* présentaient un taux plus fort de substitutions non-synonymes entre *M. lychnidis-dioicae* et *M. silenes-dioicae* que les autres gènes, ce qui suggèrent qu'une partie d'entre eux pourraient être impliqués dans l'adaptation.

Nous avons trouvé que le polymorphisme le long du génome était positivement corrélé au taux de recombinaison, mesuré par le déséquilibre de liaison local, et inversement corrélé au GC12, mais pas au GC3. La corrélation positive entre taux de recombinaison et polymorphisme suggère que la sélection liée joue un rôle dans les variations du polymorphisme le long du génome. Nous avons aussi observé une corrélation négative entre le contenu en GC et le polymorphisme, mais cette corrélation n'est valable que pour le GC12 ou le GC total, et pas pour le GC3. Le GC3 n'est d'ailleurs pas corrélé au GC12 chez *M. lychnidis-dioicae* (cf. 4.3 p. 140).

Dans cette étude, nous nous sommes concentrés sur l'analyse des autosomes, d'une part car le nombre de souches disponibles étaient différents pour les autosomes et les chromosomes de type sexuel (pour certaines souches il est impossible de récupérer les deux chromosomes de type sexuel dans les cultures de la phase haploïde à cause de l'existence d'allèles haplo-délétères liés à un des deux types sexuels), et d'autre part car les tests utilisés supposent l'existence de recombinaison. L'analyse du polymorphisme sur les chromosomes de type sexuel sera réalisée ultérieurement.

Perspectives sur la génomique de l'adaptation dans le complexe *M. violaceum*

Notre étude de génomique des populations est le premier volet d'une étude plus large de l'adaptation dans le complexe *M. violaceum*. Dans le cadre de ma thèse, j'ai participé à la génération de données pour réaliser des analyses de génomique comparative entre plusieurs espèces de *Microbtryum*. Nous avons séquencé les génomes d'une douzaine d'espèces du complexe *M. violaceum* avec des banques illumina paired-end et mate-pair. L'analyse de ces données est en cours. Nos objectifs sont :

- D'une part d'identifier des gènes sous sélection diversifiante entre espèces et de réaliser des analyses d'enrichissement fonctionnel pour détecter quelles fonctions seraient les plus importantes dans l'adaptation.
- De reconstituer l'histoire des familles de gènes au sein du complexe et par rapport à

d'autres espèces des Pucciniomycètes, afin d'identifier les expansions et réductions qui pourraient être impliqués dans l'adaptation chez les champignons du complexe *M. violaceum* (spécialisation dans l'infection d'un tissu particulier, les bourgeons floraux, et manipulation du développement floral), et des expansions ou pertes de gènes lignées-spécifiques qui pourraient expliquer la spécificité d'hôte au sein du complexe *M. violaceum*.

La grande faiblesse des approches de scans génomiques, que ce soit par des approches de génomique des populations ou de génomique comparative, est souvent le manque de "validation" du rôle des gènes candidats dans l'adaptation. Selon le domaine de la biologie auquel on appartient, le terme "validation" prend un sens très différent, les biologistes moléculaires comprenant " l'observation d'un phénotype chez un mutant avec perte ou gain de fonction", et certains biologistes évolutifs "QTL ou étude de génétique d'association". Pour les gènes candidats identifiés chez *M. lychnidis-dioicae* et *M. silenes-dioicae*, nous avons deux pistes de validation :

Nous avons lancé une étude de l'hybridation interspécifique entres plusieurs espèces du complexe *M. violaceum*. Nous avons généré des hybrides interspécifiques expérimentaux entre espèces plus ou moins éloignées du complexe *M. violaceum*, sur trois ou quatre générations. Pour chaque croisement, les générations successives ont été générées soit par backcross soit par autofécondation, et inoculées sur les deux plantes hôtes des génotypes parentaux. Notre objectif est d'identifier des régions plus ou moins perméables aux flux de gènes, et de regarder si ces régions correspondent aux balayages détectés dans notre scan génomique.

En particulier, une partie du protocole consiste à effectuer des backcross entre hybrides et génotypes parentaux, et à inoculer ces croisements sur l'espèce de plante à laquelle le génotype parental du backcross n'est pas adapté (par exemple *Silene latifolia* pour un backcross entre une souche hybride et le parent *M. silenes-dioicae* adapté à *Silene dioica*). Nous espérons qu'au bout de quelques générations on n'aura conservé de l'espèce adaptée à cette plante hôte que les allèles essentiels à la capacité à infecter une espèce hôte donnée. Cependant, on aura aussi exclu les allèles impliquées dans d'éventuelles incompatibilités de type Bateson–Dobzhansky–Muller. Pour distinguer adaptation et incompatibilités, nous avons également inoculés les générations successives d'hybrides sur une espèce de Silène très susceptible à l'infection par différentes espèces de *M. violaceum*, probablement car cette espèce, annuelle, n'est pas infectée dans la nature par *M. violaceum* et n'a pas développé de résistance à ce champignon. On espère que la comparaison entre hybrides inoculés sur cette espèce susceptible et sur les autres plantes hôtes

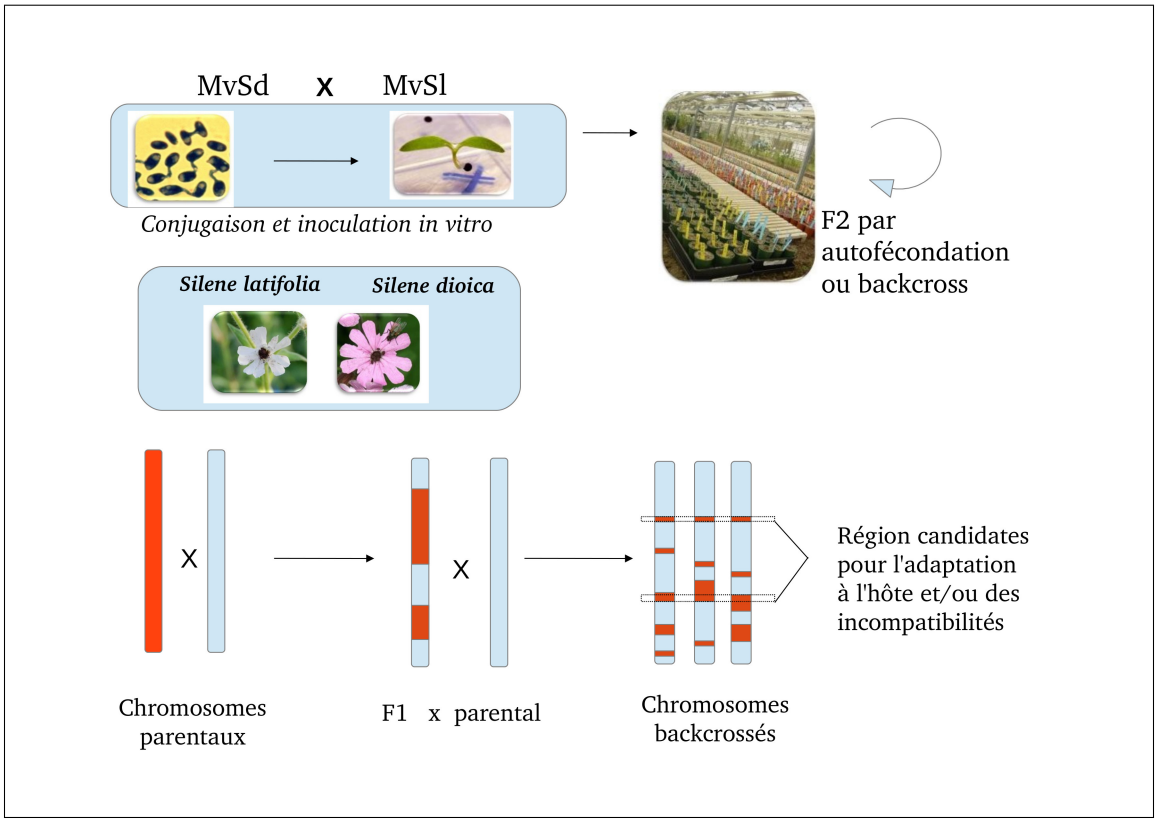


FIGURE 4.2 – Hybridation expérimentale dans le complexe *Microbotryum violaceum*. Des hybrides expérimentaux inter-spécifiques ont été réalisés entre *M. lychnidis-dioicae* et *M. silenes-dioicae* et *M. lychnidis-dioicae* et *M. violaceum sensus stricto* infectant *Lychnis flos-cuculi*. Les souches des générations successives ont été obtenues par auto-fécondation ou par backcross avec un des génotypes parentaux, et inoculées sur chacune des espèces hôtes infectées par les espèces parentes. Dans les croisements de backcross où un hybride est croisé avec un génotype parental et inoculé sur la plante non-hôte de ce génotype parental, on s'attend à ce que les allèles provenant du génotype adapté à la plante et impliqués dans l'adaptation à l'hôte soient conservés préférentiellement.

permettra de faire la différence entre régions impliquées dans l'adaptation et régions impliquées dans des incompatibilités.

Nous avons également réalisé des croisements témoins intra-spécifiques entre souches de différents groupes génétiques de *M. lychnidis-dioicae* qui devraient nous permettre d'obtenir une carte génétique pour cette espèce, et d'étudier plus finement le lien entre polymorphisme et taux de recombinaison.

Une autre approche pour "valider" le rôle des gènes et régions candidats est celle de la validation fonctionnelle. Un protocole de transformation a très récemment été développé chez *M. lychnidis-dioicae* par un de nos collaborateurs et sera bientôt publié. Un mutant est en cours de génération pour un gène candidat trouvé dans notre scan génomique, qui sur-exprimerait la protéine grâce à un protocole utilisant Agrobacterium [121]. Ce gène est induit dans la plante, situé dans une région de balayage sélectif, possède un domaine CFEM, qui est associé à des facteurs de virulence chez d'autres champignons, et est localisé hors de la cellule avec un ancrage à la membrane. On pourra tester l'effet de la virulence de ce mutant chez *M. lychnidis-dioicae*. A terme, l'enjeu sera de développer des protocoles de transformation plus rapides, et qui permettent d'invalider ou de surexprimer plusieurs gènes à la fois. Plusieurs pistes sont envisageables, entre particulier le développement d'un protocole de transformation CRIPS/CAS9 chez *M. lychnidis-dioicae* [122].

5

Conclusion générale

Dans cette thèse, j'ai présenté des travaux sur l'évolution des chromosomes de types sexuels dans le complexe d'espèces *M. violaceum*, la recherche de traces de dégénérescence dans les génomes de populations de *M. lychnidis-dioicae* de la région de Tchernobyl, et l'adaptation à l'hôte chez *M. lychnidis-dioicae* et *M. silenes-dioicae*.

Nous avons mis fin à la controverse sur la taille de la région non-recombinante des chromosomes de types sexuels de *M. lychnidis-dioicae*, en montrant que ceux-ci étaient non-recombinants sur environ 90% de leur longueur. Nous avons de plus mis en évidence un niveau exceptionnel de réarrangements entre les deux types sexuels, et des centaines de pertes de gènes. Nous avons également montré que les chromosomes de types sexuels de plusieurs espèces du complexe *M. violaceum* présentaient une accumulation indépendante de mutations non-synonymes, probablement délétères, et d'éléments transposables. Ces études ouvrent la voie à une comparaison de la structure des chromosomes de types sexuels de différentes espèces du complexe *M. violaceum* et de groupes externes, qui devrait permettre de mieux comprendre l'histoire de la suppression de recombinaison et la dynamique des régions non-recombinantes chez *M. violaceum*.

Nous avons étudié la dégénérescence génomique dans les génomes de populations de *M. lychnidis-dioicae* de la région de Tchernobyl. Nous n'avons pas détecté d'accumulation de mutations délétères à l'échelle du génome dans les populations exposées aux radiations. Ces résultats suggèrent que *M. violaceum* pourrait être résistant aux radiations, et/ou que la sélection pourrait être plus forte sur le site de Tchernobyl.

Nous nous sommes enfin intéressés à l'adaptation à l'hôte en effectuant une étude de génomique des populations dans deux espèces jumelles, *M. lychnidis-dioicae* et *M. silenes-dioicae*. Nous avons détecté des balayages sélectifs, surtout chez *M. lychnidis-dioicae*, et utilisé des données d'expression pour identifier des gènes qui pourraient être impliqués dans les interactions avec l'hôte dans ces régions de balayage sélectif. Nous avons montré que les gènes sur-exprimés dans la plante présentaient en moyenne un taux de substitutions non-synonymes entre les deux espèces jumelles plus fort que pour les autres gènes, ce qui suggère qu'une partie de ces gènes pourraient être impliqués dans l'adaptation. Nous avons enfin montré qu'il existait une corrélation négative entre polymorphisme et déséquilibre de liaison le long du génome, ce qui pourrait indiquer que l'auto-stop génétique influence le niveau de polymorphisme le long du génome. Cette étude constitue une des premières analyses de l'architecture génomique de la sélection chez un champignon pathogène dans un génome quasiment entièrement assemblé. Les génomes d'hybrides expérimentaux en cours de séquençage dans notre équipe (cf. 4.5, page 212) vont permettre de réaliser un premier test sur l'implication de ces balayages sélectifs dans l'adaptation à l'hôte. Des expériences de transformations sont également en cours en collaboration pour valider la fonction d'un gène candidat.

Dans cette thèse, j'ai étudié de manière relativement séparée les questions de l'évolution des régions non-recombinantes et celle de l'adaptation à l'hôte dans le complexe *M. violaceum*. Une question importante est évidemment de savoir si les centaines de gènes situés dans les régions non-recombinantes jouent un rôle important dans l'infection de la plante, ou au contraire plutôt dans la phase haploïde ou la conjugaison. En effet, l'absence de recombinaison peut maintenir des combinaisons d'allèles co-adaptés, même si elle doit aussi freiner leur apparition. De plus, l'analyse du génome de *M. lychnidis-dioicae* a montré un enrichissement de petites protéines sécrétées riches en cystéine sur le chromosome a_1 (cf. 4.3 p. 140), ce type de protéines étant des effecteurs potentiels [123].

Pour avancer dans la compréhension des bases moléculaires de l'infection par *M. violaceum*, et étudier le lien entre adaptation et régions non-recombinantes, on pourra étudier plus finement l'expression des gènes au cours de différents stades de l'infection. Cela pourra notamment indiquer s'il existe des clusters de gènes co-régulés, et si des gènes situés dans les régions non-recombinantes jouent un rôle lors de l'infection.

Cette thèse renforce le statut des chromosomes de types sexuels de champignons comme modèles d'allosomes complémentaires aux chromosomes sexuels. En effet, nos travaux ont contribué à montrer que les chromosomes de types sexuels des champignons possèdent des convergences remarquables avec les chromosomes sexuels. Mais la théorie des gènes à effets antagonistes invoquée pour expliquer l'existence de des chromosomes sexuels ne peut pas s'appliquer aux chromosomes de types sexuels. Il est donc extrêmement intéressant d'étudier si des forces évolutives différentes sont à l'œuvre dans ces deux systèmes, ou si une théorie plus générale suffit à rendre compte de l'existence à la fois des chromosomes sexuels et des chromosomes de types sexuels.

Plus largement, cette thèse illustre le fait que la technologie PacBio rend maintenant possible l'obtention d'un génome de référence de qualité pour la plupart des espèces, et ce à un faible coût, à condition de disposer d'échantillons d'ADN de qualité. Cette rupture technologique va grandement faciliter les études de génomique comparative et de génomique des populations dans une grande diversité de modèles. Cela devrait permettre de mieux comprendre les mécanismes généraux de l'adaptation, et il est probable que cela permettra aussi de découvrir de nombreuses régions non-recombinantes. Dans ce cadre, les défis des années à venir résident d'une part dans la capacité à développer des modèles pour interpréter les variations observées entre génomes, et d'autre part dans le développement et la mise en œuvre des outils pour tester à grande échelle les gènes candidats détectés par des approches de génomique comparative ou génomique des populations.

Bibliographie

[1] Mark Kirkpatrick, Toby Johnson, and Nick Barton. General Models of Multilocus Evolution. *Genetics*, 161(4) :1727–1750, January 2002.

[2] James F. Crow and Motoo Kimura. Evolution in Sexual and Asexual Populations. *The American Naturalist*, 99(909) :439–450, 1965.

[3] Peter D. Keightley and Sarah P. Otto. Interference among deleterious mutations favours sex and recombination in finite populations. *Nature*, 443(7107) :89–92, September 2006.

[4] B. Charlesworth. The evolution of sex chromosomes. *Science*, 251(4997) :1030–1033, January 1991.

[5] Brian Charlesworth and Deborah Charlesworth. The degeneration of Y chromosomes. *Philosophical Transactions of the Royal Society of London B : Biological Sciences*, 355(1403) :1563–1572, November 2000.

[6] Roberta Bergero and Deborah Charlesworth. The evolution of restricted recombination in sex chromosomes. *Trends in Ecology & Evolution*, 24(2) :94–102, February 2009.

[7] Peter H. Thrall, Arjen Biere, and Janis Antonovics. Plant Life-History and Disease Susceptibility–The Occurrence of Ustilago Violacea on Different Species within the Caryophyllaceae. *Journal of Ecology*, 81(3) :489–498, September 1993.

[8] Michael E Hood, Jorge I Mena-Alí, Amanda K Gibson, Bengt Oxelman, Tatiana Giraud, Roxana Yockteng, Mary T K Arroyo, Fabio Conti, Amy B Pedersen, Pierre Gladieux, et al. Distribution of the anther-smut pathogen Microbotryum on species of the Caryophyllaceae. *New Phytol*, 187(1) :217–229, July 2010.

[9] Wakana Uchida, Sachihiro Matsunaga, Ryuji Sugiyama, Yusuke Kazama, and Shigeyuki Kawano. Morphological development of anthers induced by the dimorphic smut fungus

Microbotryum violaceum in female flowers of the dioecious plant Silene latifolia. *Planta*, 218(2) :240–248, October 2003.

[10] Michael E. Hood and Janis Antonovics. Intratetrad mating, heterozygosity, and the maintenance of deleterious alleles in Microbotryum violaceum (=Ustilago violacea). *Heredity*, 85(3) :231–241, September 2000.

[11] Tatiana Giraud, Roxana Yockteng, Manuela López-Villavicencio, Guislaine Refrégier, and Michael E. Hood. Mating System of the Anther Smut Fungus Microbotryum violaceum : Selfing under Heterothallism. *Eukaryot Cell*, 7(5) :765–775, May 2008.

[12] Angela Maria Schäfer, Martin Kemler, Robert Bauer, and Dominik Begerow. The illustrated life cycle of Microbotryum on the host plant Silene latifolia. *Botany*, 88(10) :875–885, October 2010.

[13] Manuela López-Villavicencio, Odile Jonot, Amélie Coantic, Michael E Hood, Jérôme Enjalbert, and Tatiana Giraud. Multiple Infections by the Anther Smut Pathogen Are Frequent and Involve Related Strains. *PLoS Pathog*, 3(11), November 2007.

[14] Goldschmidt V. Vererbungsversuche mit den biologischen Artendes Antherenbrandes (Ustilago violacea Pers.). *Z. Bot.*, (21) :1 :90, 1928.

[15] E. D. Garber and M. Ruddat. Transmission Genetics of Microbotryum violaceum (Ustilago violacea) : A Case History. In Joan W. Bennett and Geoffrey M. Gadd Allen I. Laskin, editor, *Advances in Applied Microbiology*, volume 51 of *Advances in Applied Microbiology*, pages 107–127. Academic Press, 2002.

[16] Michael E Hood. Dimorphic mating-type chromosomes in the fungus Microbotryum violaceum. *Genetics*, 160(2) :457–461, February 2002.

[17] Michael E. Hood, Janis Antonovics, and Britt Koskella. Shared Forces of Sex Chromosome Evolution in Haploid-Mating and Diploid-Mating Organisms. *Genetics*, 168(1) :141–146, September 2004.

[18] Antonina A. Votintseva and Dmitry A. Filatov. Evolutionary Strata in a Small Mating-Type-Specific Region of the Smut Fungus Microbotryum violaceum. *Genetics*, 182(4) :1391–1396, January 2009.

[19] E. Bucheli, B. Gautschi, J. A. Shykoff, and others. Host-specific differentiation in the anther smut fungus Microbotryum violaceum as revealed by microsatellites. *Journal of Evolutionary Biology*, 13(2) :188–198, 2000.

[20] Mickael Le Gac, Michael E. Hood, Elisabeth Fournier, and Tatiana Giraud. Phylogenetic Evidence of Host-Specific Cryptic Species in the Anther Smut Fungus. *Evolution*, 61(1) :15–26, 2007.

[21] D. M. De Vienne, G. Refrégier, M. E. Hood, A. Guigue, B. Devier, E. Vercken, C. Smadja, A. Deseille, and T. Giraud. Hybrid sterility and inviability in the parasitic fungal species complex Microbotryum. *Journal of Evolutionary Biology*, 22(4) :683–698, April 2009.

[22] Guislaine Refrégier, Mickaël Le Gac, Florian Jabbour, Alex Widmer, Jacqui A Shykoff, Roxana Yockteng, Michael E Hood, and Tatiana Giraud. Cophylogeny of the anther smut fungi and their caryophyllaceous hosts : Prevalence of host shifts and importance of delimiting parasite species for inferring cospeciation. *BMC Evol Biol*, 8 :100, March 2008.

[23] Pierre Gladieux, Elodie Vercken, Michael C. Fontaine, Michael E. Hood, Odile Jonot, Arnaud Couloux, and Tatiana Giraud. Maintenance of Fungal Pathogen Species That Are Specialized to Different Hosts : Allopatric Divergence and Introgression through Secondary Contact. *Mol Biol Evol*, 28(1) :459–471, January 2011.

[24] Michael E Hood, Janis Antonovics, and Hilary Heishman. Karyotypic similarity identifies multiple host-shifts of a pathogenic fungus in natural populations. *Infection, Genetics and Evolution*, 2(3) :167–172, February 2003.

[25] Matthias Lutz, Markus Göker, Marcin Piatek, Martin Kemler, Dominik Begerow, and Franz Oberwinkler. Anther smuts of Caryophyllaceae : Molecular characters indicate host-dependent species delimitation. *Mycol Progress*, 4(3) :225–238, August 2005.

[26] Martin Kemler, Markus Göker, Franz Oberwinkler, and Dominik Begerow. Implications of molecular characters for the phylogeny of the Microbotryaceae (Basidiomycota : Urediniomycetes). *BMC Evol Biol*, 6 :35, April 2006.

[27] Jacqui A. Shykoff and Oliver Kaltz. Effects of the Anther Smut Fungus Microbotryum violaceum on Host Life-History Patterns in Silene latifolia (Caryophyllaceae). *International Journal of Plant Sciences*, 158(2) :164, March 1997.

[28] Jacqui A. Shykoff and Oliver Kaltz. Phenotypic Changes in Host Plants Diseased by Microbotryum violaceum : Parasite Manipulation, Side Effects, and Trade-offs. *International Journal of Plant Sciences*, 159(2) :236–243, March 1998.

[29] Oliver Kaltz, Sylvain Gandon, Yannis Michalakis, and Jacqui A. Shykoff. Local maladaptation in the anther-smut fungus Microbotryum violaceum to its host plant Silene latifolia : evidence from a cross-inoculation experiment. *Evolution*, pages 395–407, 1999.

[30] G. Bernasconi, J. Antonovics, A. Biere, D. Charlesworth, L. F. Delph, D. Filatov, T. Giraud, M. E. Hood, G. a. B. Marais, D. McCauley, et al. Silene as a model system in ecology and evolution. *Heredity*, 103(1) :5–14, April 2009.

[31] J. Antonovics. The effect of sterilizing diseases on host abundance and distribution along environmental gradients. *Proceedings of the Royal Society B : Biological Sciences*, 276(1661) :1443–1448, April 2009.

[32] J Antonovics, TJ Newman, and BJ Best. Spatially explicit studies on the ecology and genetics of population margins. In *SPECIAL SYMPOSIA-BRITISH ECOLOGICAL SOCIETY*, volume 14, pages 97–116. British Ecological Society, 1997.

[33] Ilkka A Hanski and Oscar E Gaggiotti. *Ecology, genetics and evolution of metapopulations*. Academic Press, 2004.

[34] Lorenza Buono, Manuela López-Villavicencio, Jacqui A. Shykoff, Alodie Snirc, and Tatiana Giraud. Influence of Multiple Infection and Relatedness on Virulence : Disease Dynamics in an Experimental Plant Population and Its Castrating Parasite. *PLoS One*, 9(6), June 2014.

[35] Eric Fontanillas, Michael E. Hood, Hélène Badouin, Elsa Petit, Valérie Barbe, Jérôme Gouzy, Damien M. de Vienne, Gabriela Aguileta, Julie Poulain, Patrick Wincker, et al. Degeneration of the Nonrecombining Regions in the Mating-Type Chromosomes of the Anther-Smut Fungi. *Mol Biol Evol*, page msu396, December 2014.

[36] Jessica Stapley, Julia Reger, Philine G. D. Feulner, Carole Smadja, Juan Galindo, Robert Ekblom, Clair Bennison, Alexander D. Ball, Andrew P. Beckerman, and Jon Slate. Adaptation genomics : the next generation. *Trends Ecol. Evol. (Amst.)*, 25(12) :705–712, December 2010.

[37] Roxana Yockteng, Sylvain Marthey, Hélène Chiapello, Annie Gendrault, Michael E Hood, François Rodolphe, Benjamin Devier, Patrick Wincker, Carole Dossat, and Tatiana Giraud. Expressed sequences tags of the anther smut fungus, Microbotryum violaceum, identify mating and pathogenicity genes. *BMC Genomics*, 8 :272, August 2007.

[38] Benjamin Devier, Gabriela Aguileta, Michael E. Hood, and Tatiana Giraud. Ancient Trans-specific Polymorphism at Pheromone Receptor Genes in Basidiomycetes. *Genetics*, 181(1) :209–223, January 2009.

[39] Michael E. Hood, Elsa Petit, and Tatiana Giraud. Extensive Divergence Between Mating-Type Chromosomes of the Anther-Smut Fungus. *Genetics*, 193(1) :309–315, January 2013.

[40] G. Aguileta, J. Lengelle, S. Marthey, H. Chiapello, F. Rodolphe, A. Gendrault, R. Yockteng, E. Vercken, B. Devier, M. C. Fontaine, et al. Finding candidate genes under positive selection in Non-model species : examples of genes involved in host specialization in pathogens : LOOKING FOR GENES INVOLVED IN HOST SPECIALIZATION. *Molecular Ecology*, 19(2) :292–306, January 2010.

[41] Pierre Gladieux, Benjamin Devier, Gabriela Aguileta, Corinne Cruaud, and Tatiana Giraud. Purifying selection after episodes of recurrent adaptive diversification in fungal pathogens. *Infection, Genetics and Evolution*, 17 :123–131, July 2013.

[42] Sophia Ahmed, J. Mark Cock, Eugenie Pessia, Remy Luthringer, Alexandre Cormier, Marine Robuchon, Lieven Sterck, Akira F. Peters, Simon M. Dittami, Erwan Corre, et al. A Haploid System of Sex Determination in the Brown Alga Ectocarpus sp. *Current Biology*, 24(17) :1945–1957, September 2014.

[43] James A. Fraser and Joseph Heitman. Evolution of fungal sex chromosomes. *Molecular Microbiology*, 51(2) :299–306, January 2004.

[44] C. A. Whittle and H. Johannesson. Evidence of the accumulation of allele-specific non-synonymous substitutions in the young region of recombination suppression within the mating-type chromosomes of Neurospora tetrasperma. *Heredity*, 107(4) :305–314, October 2011.

[45] Douglas C. Boyes and June B. Nasrallah. Physical linkage of the SLG and SRK genes at the self-incompatibility locus of Brassica oleracea. *Molec. Gen. Genet.*, 236(2-3) :369–373, January 1993.

[46] J. L. Campbell and B. C. Turner. Recombination block in the Spore killer region of Neurospora. *Genome*, 29(1) :129–135, February 1987.

[47] John Wang, Yannick Wurm, Mingkwan Nipitwattanaphon, Oksana Riba-Grognuz, Yu-Ching Huang, DeWayne Shoemaker, and Laurent Keller. A Y-like social chromosome causes alternative colony organization in fire ants. *Nature*, 493(7434) :664–668, January 2013.

[48] Mathieu Joron, Lise Frezal, Robert T. Jones, Nicola L. Chamberlain, Siu F. Lee, Christoph R. Haag, Annabel Whibley, Michel Becuwe, Simon W. Baxter, Laura Ferguson, et al. Chromosomal rearrangements maintain a polymorphic supergene controlling butterfly mimicry. *Nature*, 477(7363) :203–206, September 2011.

[49] Frédéric Veyrunes, Paul D. Waters, Pat Miethke, Willem Rens, Daniel McMillan, Amber E. Alsop, Frank Grützner, Janine E. Deakin, Camilla M. Whittington, Kyriena Schatzkamer, et al. Bird-like sex chromosomes of platypus imply recent origin of mammal sex chromosomes. *Genome Res.*, 18(6) :965–973, January 2008.

[50] Doris Bachtrog, Mark Kirkpatrick, Judith E. Mank, Stuart F. McDaniel, J. Chris Pires, William Rice, and Nicole Valenzuela. Are all sex chromosomes created equal ? *Trends in Genetics*, 27(9) :350–357, September 2011.

[51] Gabriel A. B. Marais, Michael Nicolas, Roberta Bergero, Pierre Chambrier, Eduard Kejnovsky, Françoise Monéger, Roman Hobza, Alex Widmer, and Deborah Charlesworth. Evidence for degeneration of the Y chromosome in the dioecious plant Silene latifolia. *Curr. Biol.*, 18(7) :545–549, April 2008.

[52] Sylvain Billiard, Manuela López-Villavicencio, Benjamin Devier, Michael E. Hood, Cécile Fairhead, and Tatiana Giraud. Having sex, yes, but with whom ? Inferences from fungi on the evolution of anisogamy and mating types. *Biological Reviews*, 86(2) :421–442, May 2011.

[53] Alexander Idnurm, Michael E. Hood, Hanna Johannesson, and Tatiana Giraud. Contrasted patterns in mating-type chromosomes in fungi : Hotspots versus coldspots of recombination. *Fungal Biology Reviews*, 2015.

[54] Joseph E. Ironside. No amicable divorce ? Challenging the notion that sexual antagonism drives sex chromosome evolution. *Bioessays*, 32(8) :718–726, August 2010.

[55] Butler G. The evolution of MAT : the Ascomycetes. In *Sex in Fungi : Molecular Determination and Evolutionary Implications*, pages 3–18. ASM Press, Washington, D.C, 2007.

[56] Audrius Menkis, David J. Jacobson, Tim Gustafsson, and Hanna Johannesson. The Mating-Type Chromosome in the Filamentous Ascomycete Neurospora tetrasperma Represents a Model for Early Evolution of Sex Chromosomes. *PLoS Genet*, 4(3) :e1000030, March 2008.

[57] Christopher E. Ellison, Jason E. Stajich, David J. Jacobson, Donald O. Natvig, Alla Lapidus, Brian Foster, Andrea Aerts, Robert Riley, Erika A. Lindquist, Igor V. Grigoriev, et al. Massive Changes in Genome Architecture Accompany the Transition to Self-Fertility in the Filamentous Fungus Neurospora tetrasperma. *Genetics*, 189(1) :55–69, September 2011.

[58] James A. Fraser and Joseph Heitman. Fungal mating-type loci. *Current Biology*, 13(20) :R792–R795, October 2003.

[59] Klaus B. Lengeler, Deborah S. Fox, James A. Fraser, Andria Allen, Keri Forrester, Fred S. Dietrich, and Joseph Heitman. Mating-Type Locus of Cryptococcus neoformans : a Step in the Evolution of Sex Chromosomes. *Eukaryot Cell*, 1(5) :704–718, October 2002.

[60] Guus Bakkeren and James W. Kronstad. Linkage of mating-type loci distinguishes bipolar from tetrapolar mating in basidiomycetous smut fungi. *Proceedings of the National Academy of Sciences*, 91(15) :7085–7089, 1994.

[61] Pierre Grognet, Frédérique Bidard, Claire Kuchly, Laetitia Chan Ho Tong, Evelyne Coppin, Jinane Ait Benkhali, Arnaud Couloux, Patrick Wincker, Robert Debuchy, and Philippe Silar. Maintaining two mating types : structure of the mating type locus and its role in heterokaryosis in Podospora anserina. *Genetics*, 197(1) :421–432, May 2014.

[62] Janis Antonovics and Joseph Y. Abrams. Intratetrad Mating and the Evolution of Linkage Relationships. *Evolution*, 58(4) :702–709, April 2004.

[63] Alan W. Day. Communication through fimbriae during conjugation in a fungus. *Nature*, 262(5569) :583–584, August 1976.

[64] Alan W. Day. Mating Type and Morphogenesis in Ustilago violacea. *Botanical Gazette*, 140(1) :94–101, March 1979.

[65] L. Xu, E. Petit, and M. E. Hood. Variation in mate-recognition pheromones of the fungal genus Microbotryum. *Heredity*, August 2015.

[66] James A Fraser, Stephanie Diezmann, Ryan L Subaran, Andria Allen, Klaus B Lengeler, Fred S Dietrich, and Joseph Heitman. Convergent Evolution of Chromosomal Sex-Determining Regions in the Animal and Fungal Kingdoms. *PLoS Biol*, 2(12), December 2004.

[67] Gabriel Marais. Biased gene conversion : implications for genome and sex evolution. *Trends in Genetics*, 19(6) :330–338, June 2003.

[68] Steve Rozen, Helen Skaletsky, Janet D. Marszalek, Patrick J. Minx, Holland S. Cordum, Robert H. Waterston, Richard K. Wilson, and David C. Page. Abundant gene conversion between arms of palindromes in human and ape Y chromosomes. *Nature*, 423(6942) :873–876, June 2003.

[69] Eduard Kejnovsky, Roman Hobza, Zdenek Kubat, Alex Widmer, Gabriel A. B. Marais, and Boris Vyskot. High intrachromosomal similarity of retrotransposon long terminal repeats : Evidence for homogenization by gene conversion on plant sex chromosomes ? *Gene*, 390(1–2) :92–97, April 2007.

[70] Bruce T. Lahn and David C. Page. Four Evolutionary Strata on the Human X Chromosome. *Science*, 286(5441) :964–967, October 1999.

[71] Michael E. Hood and Janis Antonovics. Mating Within the Meiotic Tetrad and the Maintenance of Genomic Heterozygosity. *Genetics*, 166(4) :1751–1759, January 2004.

[72] J. L. Abbate and M. E. Hood. Dynamic linkage relationships to the mating-type locus in automictic fungi of the genus Microbotryum. *J. Evol. Biol.*, 23(8) :1800–1805, August 2010.

[73] D. C. Schwartz, X. Li, L. I. Hernandez, S. P. Ramnarain, E. J. Huff, and Y. K. Wang. Ordered restriction maps of Saccharomyces cerevisiae chromosomes constructed by optical mapping. *Science*, 262(5130) :110–114, October 1993.

[74] Luigi Faino, Michael F. Seidl, Erwin Datema, Grardy C. M. van den Berg, Antoine Janssen, Alexander H. J. Wittenberg, and Bart P. H. J. Thomma. Single-Molecule Real-Time Sequencing Combined with Optical Mapping Yields Completely Finished Fungal Genome. *mBio*, 6(4) :e00936–15, January 2015.

[75] Elsa Petit, Tatiana Giraud, Damien M. de Vienne, Marco A. Coelho, Gabriela Aguileta, Joëlle Amselem, Jonathan Kreplak, Julie Poulain, Frédérick Gavory, Patrick Wincker, et al. Linkage to the Mating-Type Locus Across the Genus Microbotryum : Insights into Nonrecombining Chromosomes. *Evolution*, 66(11) :3519–3533, November 2012.

[76] B. P. S. Nieuwenhuis, S. Billiard, S. Vuilleumier, E. Petit, M. E. Hood, and T. Giraud. Evolution of uni- and bifactorial sexual compatibility systems in fungi. *Heredity*, 111(6) :445–455, December 2013.

[77] James A. Fraser and Joseph Heitman. Evolution of fungal sex chromosomes. *Molecular Microbiology*, 51(2) :299–306, January 2004.

[78] Banu Metin, Keisha Findley, and Joseph Heitman. The mating type locus (MAT) and sexual reproduction of Cryptococcus heveanensis : insights into the evolution of sex and sex-determining chromosomal regions in fungi. *PLoS Genet.*, 6(5) :e1000961, May 2010.

[79] James J. Bull. Sex Chromosomes in Haploid Dioecy : A Unique Contrast to Muller's Theory for Diploid Dioecy. *The American Naturalist*, 112(983) :245–250, January 1978.

[80] Carrie A. Whittle, Antonina Votintseva, Kate Ridout, and Dmitry A. Filatov. Recent and Massive Expansion of the Mating-Type-Specific Region in the Smut Fungus Microbotryum. *Genetics*, 199(3) :809–816, March 2015.

[81] Michael E. Hood, Melanie Katawczik, and Tatiana Giraud. Repeat-Induced Point Mutation and the Population Structure of Transposable Elements in Microbotryum violaceum. *Genetics*, 170(3) :1081–1089, July 2005.

[82] Michael H Perlin, Joelle Amselem, Eric Fontanillas, Su San Toh, Zehua Chen, Jonathan Goldberg, Sebastien Duplessis, Bernard Henrissat, Sarah Young, Qiandong Zeng, et al. Sex and parasites : genomic and transcriptomic analysis of Microbotryum lychnidis-dioicae, the biotrophic and plant-castrating anther smut fungus. *BMC Genomics*, 16(1), December 2015.

[83] Timothy Y. James, Prayook Srivilai, Ursula Kües, and Rytas Vilgalys. Evolution of the Bipolar Mating System of the Mushroom Coprinellus disseminatus From Its Tetrapolar Ancestors Involves Loss of Mating-Type-Specific Pheromone Receptor Function. *Genetics*, 172(3) :1877–1891, January 2006.

[84] Michael E. Hood, Molly Scott, and Mindy Hwang. Breaking linkage between mating compatibility factors : Tetrapolarity in *Microbotryum* : LINKAGE BETWEEN MATING COMPATIBILITY FACTORS. *Evolution*, 69(10) :2561–2572, October 2015.

[85] Hélène Badouin, Michael E. Hood, Jérôme Gouzy, Gabriela Aguileta, Sophie Siguenza, Michael H. Perlin, Christina A. Cuomo, Cécile Fairhead, Antoine Branca, and Tatiana Giraud. Chaos of Rearrangements in the Mating-Type Chromosomes of the Anther-Smut Fungus Microbotryum lychnidis-dioicae. *Genetics*, page genetics.115.177709, June 2015.

[86] T. M. Maia, S. T. Lopes, J. M. G. C. F. Almeida, L. H. Rosa, J. P. Sampaio, P. Goncalves, and M. A. Coelho. Evolution of Mating Systems in Basidiomycetes and the Genetic Architecture Underlying Mating-Type Determination in the Yeast Leucosporidium scottii. *Genetics*, 201(1) :75–89, September 2015.

[87] Yu Sun, Pádraic Corcoran, Audrius Menkis, Carrie A. Whittle, Siv G. E. Andersson, and Hanna Johannesson. Large-Scale Introgression Shapes the Evolution of the Mating-Type Chromosomes of the Filamentous Ascomycete Neurospora tetrasperma. *PLoS Genet*, 8(7) :e1002820, July 2012.

[88] Gilean A. T. McVean and Brian Charlesworth. The Effects of Hill-Robertson Interference Between Weakly Selected Mutations on Patterns of Molecular Evolution and Variation. *Genetics*, 155(2) :929–944, January 2000.

[89] Carlos D. Bustamante, Rasmus Nielsen, Stanley A. Sawyer, Kenneth M. Olsen, Michael D. Purugganan, and Daniel L. Hartl. The cost of inbreeding in Arabidopsis. *Nature*, 416(6880) :531–534, April 2002.

[90] Anders P. Møller and Timothy A. Mousseau. Low-dose radiation, scientific scrutiny, and requirements for demonstrating effects. *BMC Biology*, 11(1) :92, August 2013.

[91] Hans Ellegren, Gabriella Lindgren, Craig R. Primmer, and Anders Pape Moller. Fitness loss and germline mutations in barn swallows breeding in Chernobyl. *Nature*, 389(6651) :593–596, October 1997.

[92] Anders Pape Møller and Timothy Alexander Mousseau. Biological Indicators of Ionizing Radiation in Nature. In Robert H. Armon and Osmo Hänninen, editors, *Environmental Indicators*, pages 871–881. Springer Netherlands, 2015. DOI : 10.1007/978-94-017-9499-2_49.

[93] Ekaterina Dadachova and Arturo Casadevall. Ionizing radiation : how fungi cope, adapt, and exploit with the help of melanin. *Current Opinion in Microbiology*, 11(6) :525–531, December 2008.

[94] Nelli N. Zhdanova, Valentina A. Zakharchenko, Valeriya V. Vember, and Lidiya T. Nakonechnaya. Fungi from Chernobyl : mycobiota of the inner regions of the containment structures of the damaged nuclear reactor. *Mycological Research*, 104(12) :1421–1426, December 2000.

[95] D. M. de Vienne, G. Refrégier, M. López-Villavicencio, A. Tellier, M. E. Hood, and T. Giraud. Cospeciation vs host-shift speciation : methods for testing, evidence from natural associations and relation to coevolution. *New Phytologist*, 198(2) :347–385, April 2013.

[96] H. A. Orr. The population genetics of speciation : the evolution of hybrid incompatibilities. *Genetics*, 139(4) :1805–1813, January 1995.

[97] Jerry A. Coyne and H. Allen Orr. *Speciation*. W.H. Freeman, January 2004.

[98] Theodosius Dobzhansky. *Spermatogenesis in Pure and Hybrid Drosophila Pseudo-obscura*. J. Springer, 1934.

[99] Sarah Maria Schmidt and Ralph Panstruga. Pathogenomics of fungal plant parasites : what have we learnt about pathogenesis ? *Current Opinion in Plant Biology*, 14(4) :392–399, August 2011.

[100] Pietro Spanu and Jörg Kämper. Genomics of biotrophy in fungi and oomycetes — emerging patterns. *Current Opinion in Plant Biology*, 13(4) :409–414, August 2010.

[101] Eric Kemen and Jonathan D. G. Jones. Obligate biotroph parasitism : can we link genomes to lifestyles ? *Trends in Plant Science*, 17(8) :448–457, August 2012.

[102] Pietro D. Spanu. The Genomics of Obligate (and Nonobligate) Biotrophs. *Annual Review of Phytopathology*, 50(1) :91–109, 2012.

[103] F. Martin, A. Aerts, D. Ahrén, A. Brun, E. G. J. Danchin, F. Duchaussoy, J. Gibon, A. Kohler, E. Lindquist, V. Pereda, et al. The genome of Laccaria bicolor provides insights into mycorrhizal symbiosis. *Nature*, 452(7183) :88–92, March 2008.

[104] Sébastien Duplessis, Christina A. Cuomo, Yao-Cheng Lin, Andrea Aerts, Emilie Tisserant, Claire Veneault-Fourrey, David L. Joly, Stéphane Hacquard, Joëlle Amselem, Brandi L. Cantarel, et al. Obligate biotrophy features unraveled by the genomic analysis of rust fungi. *Proc. Natl. Acad. Sci. U.S.A.*, 108(22) :9166–9171, May 2011.

[105] Jörg Kämper, Regine Kahmann, Michael Bölker, Li-Jun Ma, Thomas Brefort, Barry J. Saville, Flora Banuett, James W. Kronstad, Scott E. Gold, Olaf Müller, et al. Insights from the genome of the biotrophic fungal plant pathogen Ustilago maydis. *Nature*, 444(7115) :97–101, November 2006.

[106] Thierry Rouxel, Jonathan Grandaubert, James K. Hane, Claire Hoede, Angela P. van de Wouw, Arnaud Couloux, Victoria Dominguez, Véronique Anthouard, Pascal Bally, Salim Bourras, et al. Effector diversification within compartments of the Leptosphaeria maculans genome affected by Repeat-Induced Point mutations. *Nat Commun*, 2 :202, February 2011.

[107] Jacqui A. Shykoff, Anne Meyhöfer, and Erika Bucheli. Genetic isolation among host races of the anther smut fungus Microbotryum violaceum on three host plant species. *International journal of plant sciences*, 160(5) :907–916, 1999.

[108] Mickael Le Gac, Michael E. Hood, and Tatiana Giraud. Evolution of Reproductive Isolation Within a Parasitic Fungal Species Complex. *Evolution*, 61(7) :1781–1787, 2007.

[109] Britta Büker, Elsa Petit, Dominik Begerow, and Michael E Hood. Experimental hybridization and backcrossing reveal forces of reproductive isolation in Microbotryum. *BMC Evol Biol*, 13 :224, October 2013.

[110] F. Tajima. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, 123(3) :585–595, January 1989.

[111] Kai Zeng, Yun-Xin Fu, Suhua Shi, and Chung-I. Wu. Statistical Tests for Detecting Positive Selection by Utilizing High-Frequency Variants. *Genetics*, 174(3) :1431–1439, January 2006.

[112] Pavlos Pavlidis, Daniel Živković, Alexandros Stamatakis, and Nikolaos Alachiotis. SweeD : Likelihood-Based Detection of Selective Sweeps in Thousands of Genomes. *Mol Biol Evol*, 30(9) :2224–2234, January 2013.

[113] Wolfgang Stephan. Signatures of positive selection : from selective sweeps at individual loci to subtle allele frequency changes in polygenic adaptation. *Mol Ecol*, pages n/a–n/a, July 2015.

[114] Rasmus Nielsen. Molecular signatures of natural selection. *Annu. Rev. Genet.*, 39 :197–218, 2005.

[115] Benoit Nabholz, Gautier Sarah, François Sabot, Manuel Ruiz, Hélène Adam, Sabine Nidelet, Alain Ghesquière, Sylvain Santoni, Jacques David, and Sylvain Glémin. Transcriptome population genomics reveals severe bottleneck and domestication cost in the African rice (Oryza glaberrima). *Mol Ecol*, 23(9) :2210–2227, May 2014.

[116] B. Charlesworth, M. T. Morgan, and D. Charlesworth. The effect of deleterious mutations on neutral molecular variation. *Genetics*, 134(4) :1289–1303, January 1993.

[117] Bret A. Payseur and Michael W. Nachman. Gene Density and Human Nucleotide Polymorphism. *Mol Biol Evol*, 19(3) :336–340, January 2002.

[118] David J. Begun and Charles F. Aquadro. Levels of naturally occurring DNA polymorphism correlate with recombination rates in D. melanogaster. *Nature*, 356(6369) :519–520, April 1992.

[119] Ines Hellmann, Kay Prüfer, Hongkai Ji, Michael C. Zody, Svante Pääbo, and Susan E. Ptak. Why do human diversity levels vary at a megabase scale ? *Genome Res.*, 15(9) :1222–1231, January 2005.

[120] I. Fudal, S. Ross, L. Gout, F. Blaise, M. L. Kuhn, M. R. Eckert, L. Cattolico, S. Bernard-Samain, M. H. Balesdent, and T. Rouxel. Heterochromatin-Like Regions as Ecological Niches for Avirulence Genes in the Leptosphaeria maculans Genome : Map-Based Cloning of AvrLm6. *MPMI*, 20(4) :459–470, March 2007.

[121] Caroline B. Michielse, Paul J. J. Hooykaas, Cees A. M. J. J. van den Hondel, and Arthur F. J. Ram. Agrobacterium-mediated transformation as a tool for functional genomics in fungi. *Curr Genet*, 48(1) :1–17, May 2005.

[122] Mariana Schuster, Gabriel Schweizer, Stefanie Reissmann, and Regine Kahmann. Genome editing in Ustilago maydis using the CRISPR–Cas system. *Fungal Genetics and Biology*, 2015.

[123] Ioannis Stergiopoulos and Pierre J.G.M. de Wit. Fungal Effector Proteins. *Annual Review of Phytopathology*, 47(1) :233–263, 2009.

Liste des figures

Annexe A

Figures et tables supplémentaires

A.1 Degeneration of the nonrecombining regions in the mating-type chromosomes of the anther-smut fungi

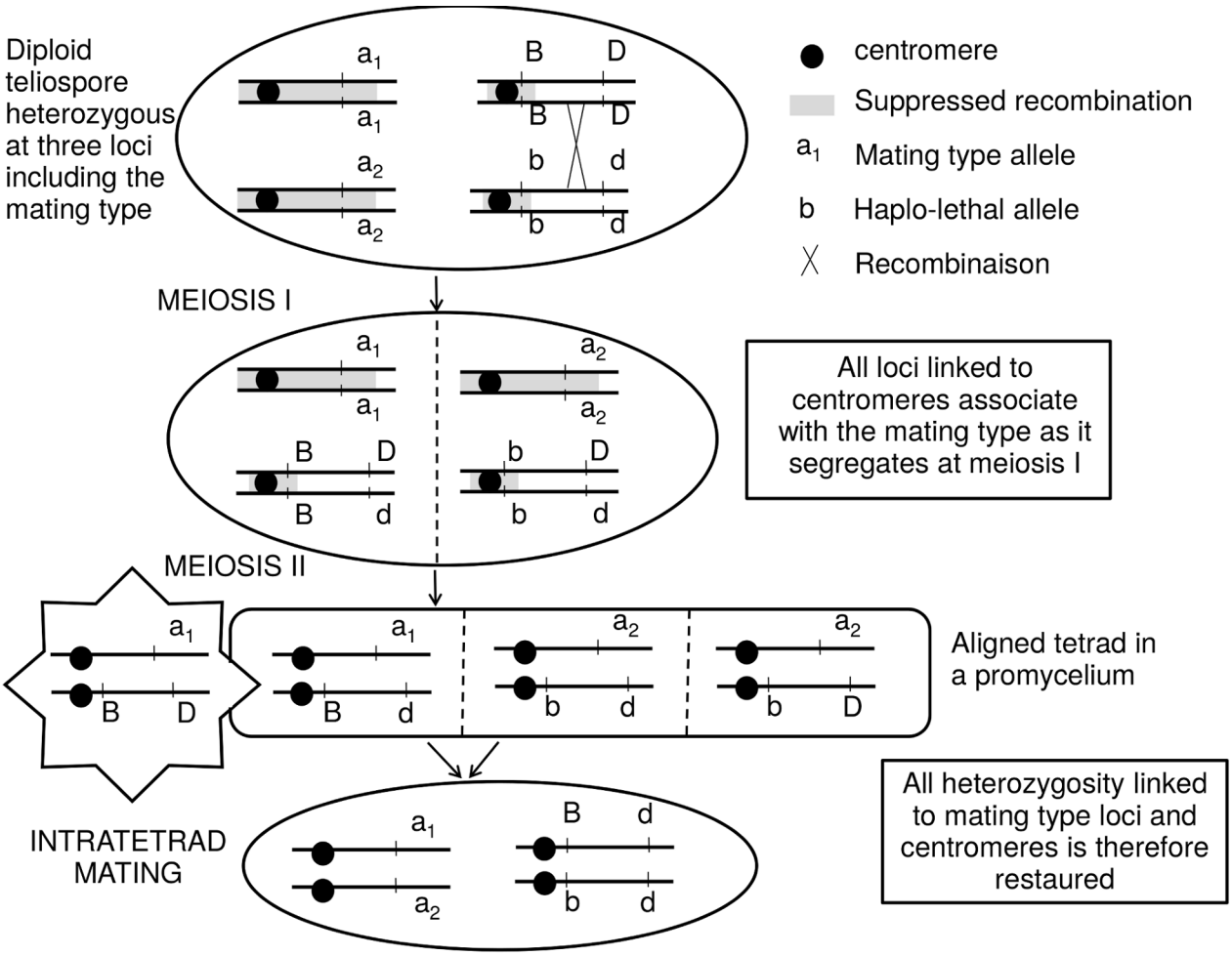


Figure S1: Automixis in *Microbotryum*. As all loci linked to the centromere segregate at Meiosis 1 with the mating type, all heterozygozity at these loci are restored by intratetrad mating, occurring necessarily between cells of different mating type alleles. This shelters deleterious loci liked to centromeres as well as those linked to the mating type. Figure after Hood and Antonovics 2000.

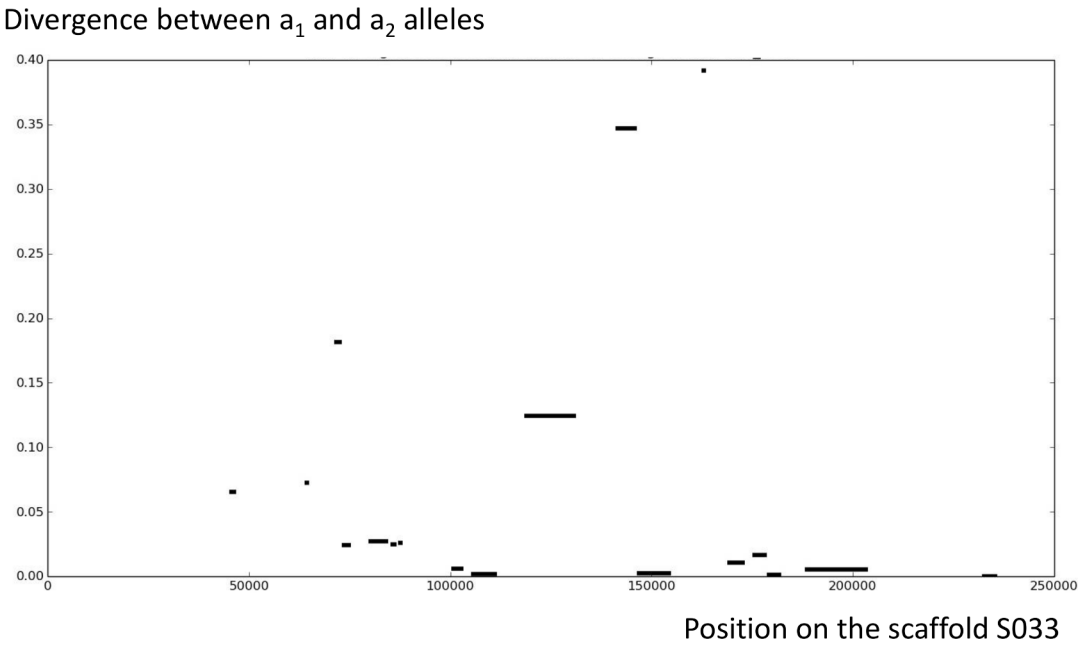


Figure S2. Heterogeneity of synonymous divergence between the a_1 and a_2 alleles of the mating-type chromosomes in their non-recombining regions (NRRs). Example of a scaffold (S033) in the NNRs with heterogeneity of synonymous divergence between the a_1 and a_2 alleles of *Microbotryum lychnidis-dioicae* exons (non transposable elements). The raw divergence is shown for alignable fragments of more than 1000 bp along the physical length of the scaffold.

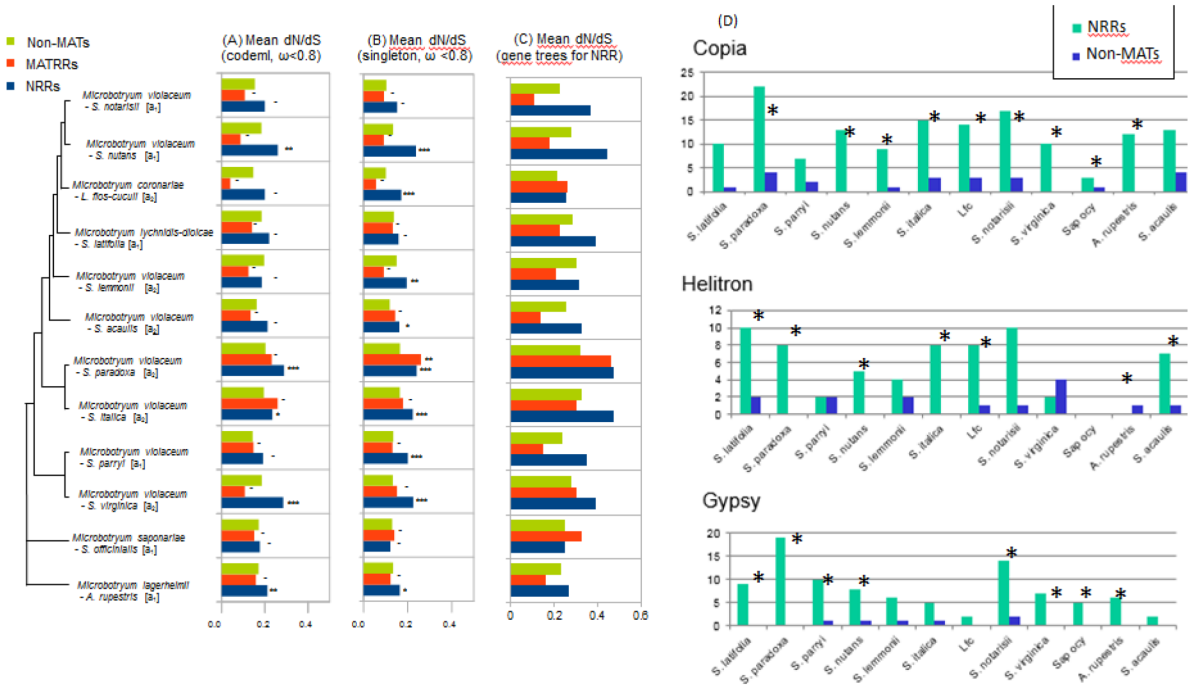


Figure S3. Degeneration evidence in non-recombining regions of the mating-type chromosomes in *Microbotryum*, in coding sequences and as accumulation of transposable elements. (A) and (B) Degeneration estimated by comparing mean non-synonymous vs synonymous substitutions (ω=dN/dS) across coding sequences (CDS) located respectively in non-mating-type chromosomes (non-MAT), NRRs (mating-type chromosome non-recombining regions) and MATRRs (mating-type chromosome recombining regions) in twelve haploid *Microbotryum* genomes (*i.e.*, either of a_1 or a_2 mating type), loci displaying dN/dS > 0.8 (i.e., nearly neutral and positively selected) being discarded from this analysis, (A) with the full frequency spectrum of substitutions used to compute dN/dS through a free-ratio branch model (Codeml, PAML), (B) with only singleton substitutions taken into account for focusing on recent events, or (C) using gene trees for NRR genes instead of species tree. (D) Transposable element (TE) contents of three families (Copia, Gypsy and Helitron) across contigs assigned respectively to non-mating-type chromosomes (non-MATs) and NRRs (mating-type chromosome non-recombining regions), in twelve haploid *Microbotryum* genomes (*i.e.*, either of a_1 or a_2 mating type). Results represent the proportion of sequences corresponding to TEs detected among 1,000 random 200-pb fragments in each compartment (non-MATs and NRR). Asterix indicate significant differences between non-mating-type chromosomes (non-MATs) and NRRs according to the *Z* test.

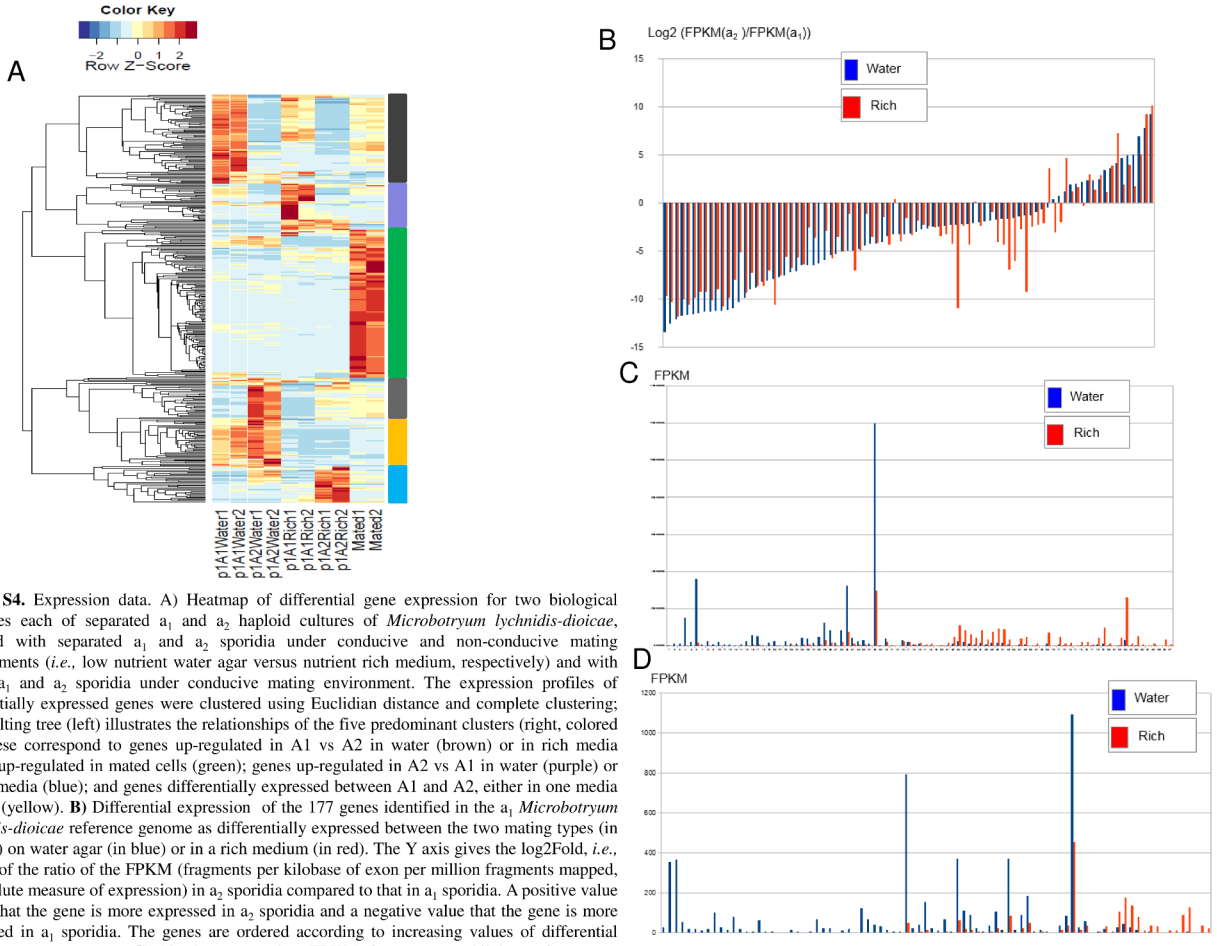


Figure S4. Expression data. A) Heatmap of differential gene expression for two biological replicates each of separated a_1 and a_2 haploid cultures of *Microbotryum lychnidis-dioicae*, assessed with separated a_1 and a_2 sporidia under conducive and non-conducive mating environments (*i.e.*, low nutrient water agar versus nutrient rich medium, respectively) and with mixed a_1 and a_2 sporidia under conducive mating environment. The expression profiles of differentially expressed genes were clustered using Euclidian distance and complete clustering; the resulting tree (left) illustrates the relationships of the five predominant clusters (right, colored bars these correspond to genes up-regulated in A1 vs A2 in water (brown) or in rich media (pink); up-regulated in mated cells (green); genes up-regulated in A2 vs A1 in water (purple) or in rich media (blue); and genes differentially expressed between A1 and A2, either in one media or both (yellow). **B)** Differential expression of the 177 genes identified in the a_1 *Microbotryum lychnidis-dioicae* reference genome as differentially expressed between the two mating types (in a_2 vs a_1) on water agar (in blue) or in a rich medium (in red). The Y axis gives the log2Fold, *i.e.*, the log of the ratio of the FPKM (fragments per kilobase of exon per million fragments mapped, an absolute measure of expression) in a_2 sporidia compared to that in a_1 sporidia. A positive value means that the gene is more expressed in a_2 sporidia and a negative value that the gene is more expressed in a_1 sporidia. The genes are ordered according to increasing values of differential expression on water agar. **C)** Absolute expression (FPKM, fragments per kilobase of exon per million fragments mapped) of the 91 genes identified in the a_1 *Microbotryum lychnidis-dioicae* reference genome as differentially expressed between the two mating types on water agar, in a_1 (in blue) or in a_2 (in red), and located on autosomes. The genes are ordered as in the Figure S3B. D) Absolute expression (FPKM, fragments per kilobase of exon per million fragments mapped) of the 86 genes identified in the a_1 *Microbotryum lychnidis-dioicae* reference genome as differentially expressed between the two mating types on water agar, in a_1 (in blue) or in a_2 (in red), and located on the-mating type chromosome. The genes are ordered as in the Figure S3B.

A.2 Chaos of rearrangements in the mating-type chromosomes of the anther-smut fungus *Microbotryum lychnidis-dioicae*.

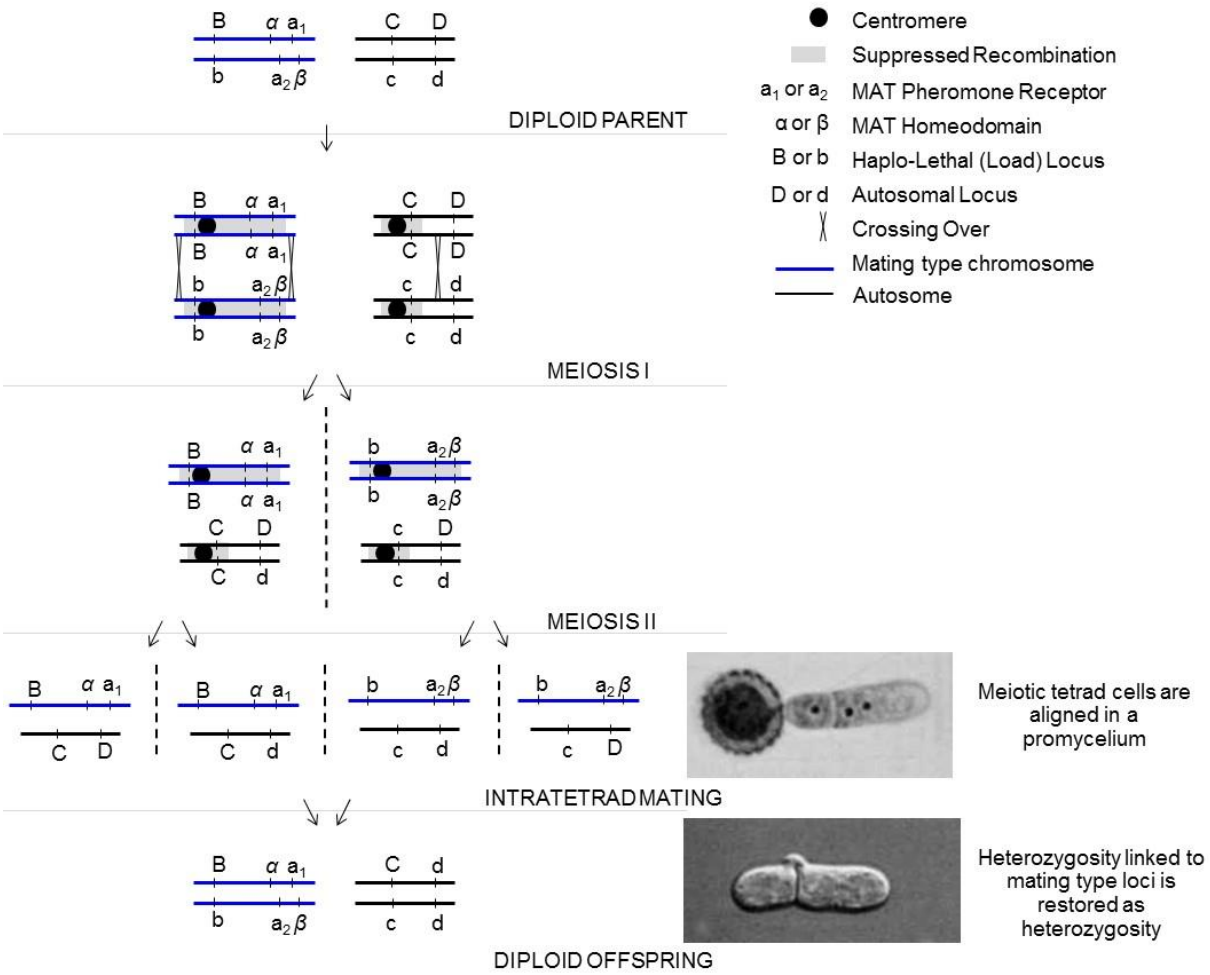


Figure S1: **Meiosis and intratetrad-mating (automixis) in *Microbotryum lychnidis-dioicae.*** A scenario with 2N number of 4 is shown, with a mating type (blue) and an autosomal (black) pair of homologous chromosomes. Mating type (MAT) is linked to the centromere. Mating type and linked loci (locus B) segregate at Meiosis I. Heterozygosity at loci linked to mating type, or that segregates a Meiosis I linked to other centromeres (locus C), is therefore restored by mating between cells from the same meiotic tetrad. This can shelters deleterious load loci linked to the mating type while purging heterozygosity at loci that segregate at Meiosis II due to crossing over (as shown for locus D).

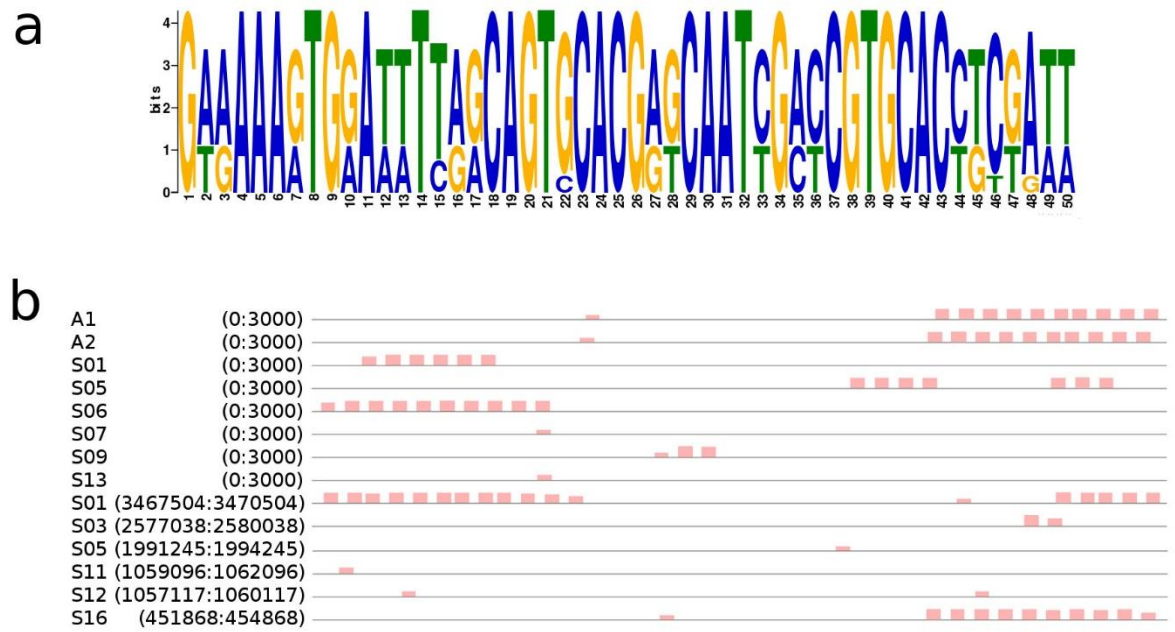


Figure S2: **Identification of sub-telomeric motifs in *Microbotryum lychnidis-dioicae*.** a) Consensus motif. b) Location in the 3000 bp of putative chromosomes edges. The genomic coordinates in bp are indicated in brackets.

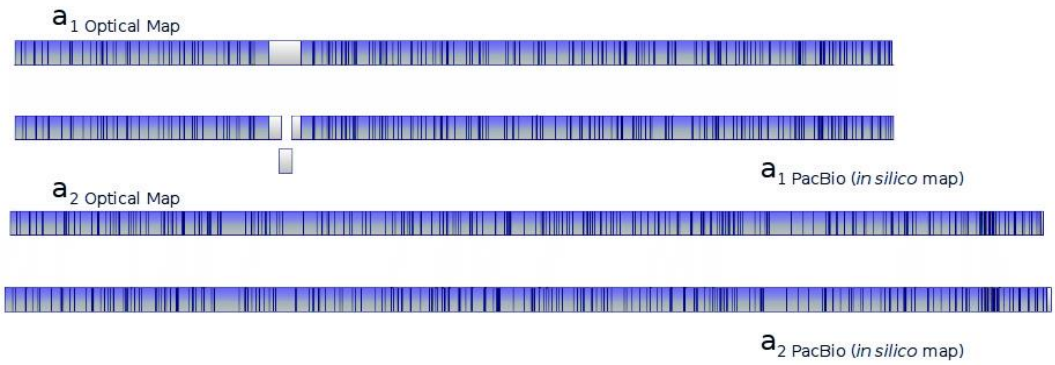


Figure S3: **Correspondence between the optical map and the PacBio assembly of the *Microbotryum lychnidis-dioicae* mating-type chromosomes.** The vertical lines within the chromosomes indicate restriction sites. Alignments compare restriction fragment sizes on single DNA molecules and from *in silico* sequences. Regions highlighted in blue indicate aligned optical map and *in silico* restriction site distributions, while regions in grey could not be aligned (*i.e.,* a_1 centromeres and a_2 telomeres).

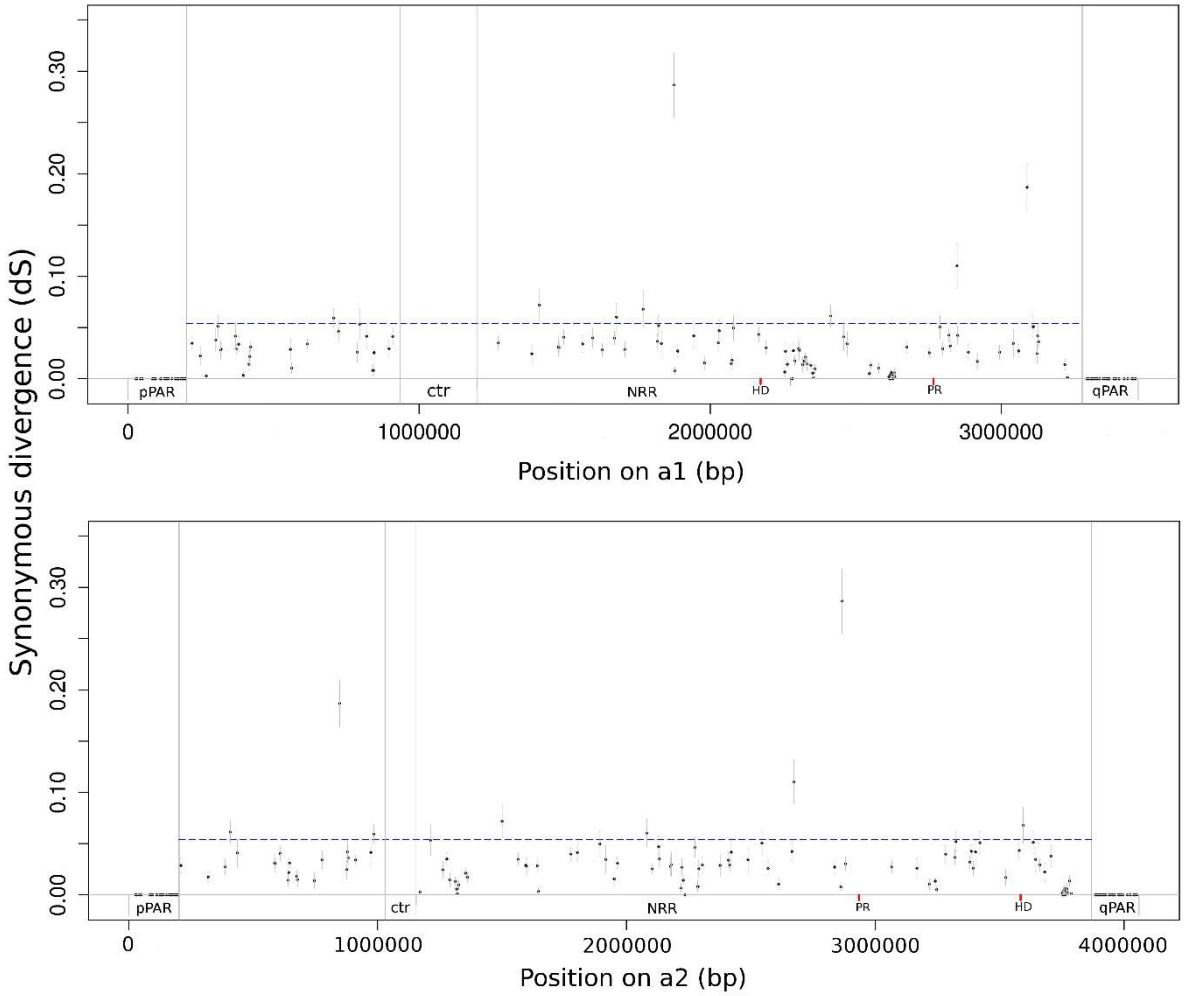


Figure S4: **Divergence between the *Microbotryum lychnidis-dioicae* mating-type chromosomes.** Synonymous divergence dS ± SE is plotted against the genomic coordinates of the a_1 (A) or a_2 (B) alleles, for predicted genes with a gap-free codon alignment of more than 1000 bp. The boundaries between the pseudo-autosomal region (PAR) and the non-recombining regions (NRR) are indicated, as well as the locations of the mating-type loci (*PR*: pheromone receptor gene; *HD*: homeodomain genes). The mean value of dS in the NRR is shown as blue dotted lines.

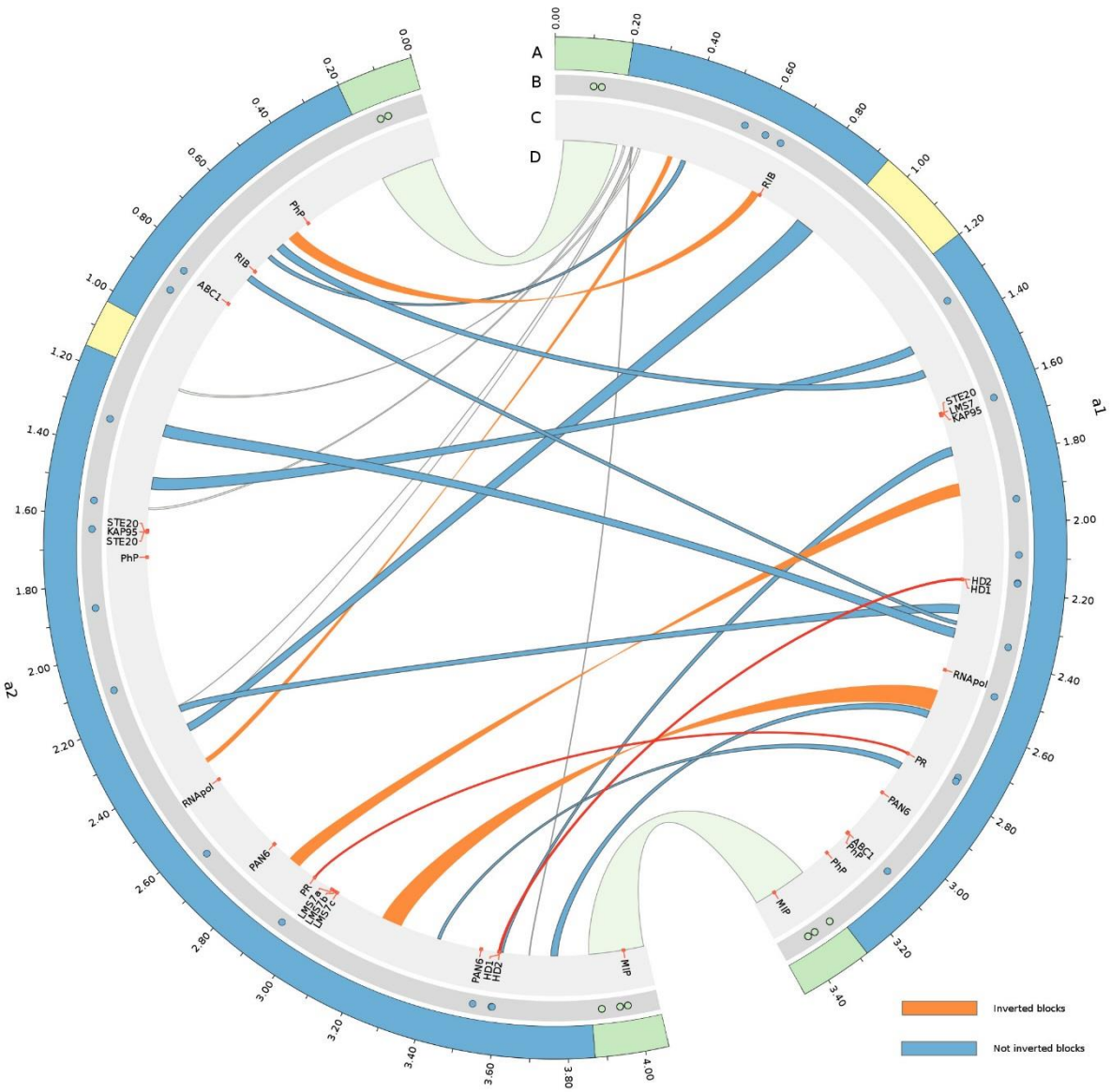


Figure S5: **Rearrangements between the main syntenic blocks of genes common to a_1 and a_2 *Microbotryum lychnidis-dioicae* mating-type chromosomes.** Tracks A to D show the location of different genomic elements, as follows: A – Structure of the chromosomes, with the pseudo-autosomal regions (PARs) in green, the non-recombining regions (NRRs) in blue, and the centromeres in yellow. B – Location of loci shown to be linked (blue circles) or unlinked (white circles) to mating-type by previous segregation analyses in *M. lychnidis-dioicae* {Abbate, 2010 #87;Petit, 2012 #86;Votintseva, 2009 #84}. C – Location of the genes related to the mating-type function: pheromone receptor and homeodomain genes (in red), the other genes likely involved in mating (*STE12, STE20*, and the precursors of pheromones, *PhP*) and the genes located around the pheromone receptor gene in the closely related *Sporidiobolus salmonicolor* {Coelho, 2010 #156} (*KAP95, RNAPol, RIB* and *ABC1*). D – Links between syntenic blocks of shared genes larger than 10 kb.

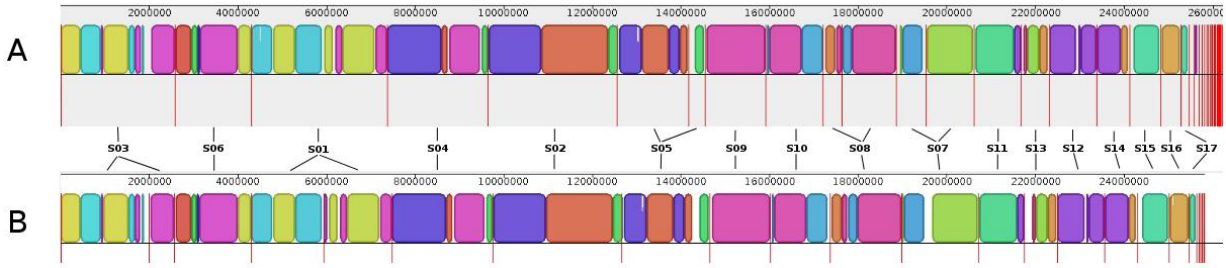


Figure S6: Autosomal contigs from the a_1 and a_2 assemblies, aligned with Mauve {Darling, 2010 #146}, illustrating their synteny. The contigs of a_1 (A) and a_2 (B) are separated by vertical red lines and the locally collinear blocks (LCBs) are represented as colored blocs. A couple of a_1 contigs were split into two contigs in the a_2 assembly (i.e., at position 2MB and 6MB), and conversely (at 14MB and 18MB). In these cases where an autosomal contig was larger in the a_1 or in the a_2 assembly, the largest contig was selected to resolve the final reference assembly (Figure 2). Contig names are indicated as in Figure 2, except the S18 that is one of the small contigs at right. The few short degenerate contigs at right were disregarded, likely resulting from the incorrect editing of reads.

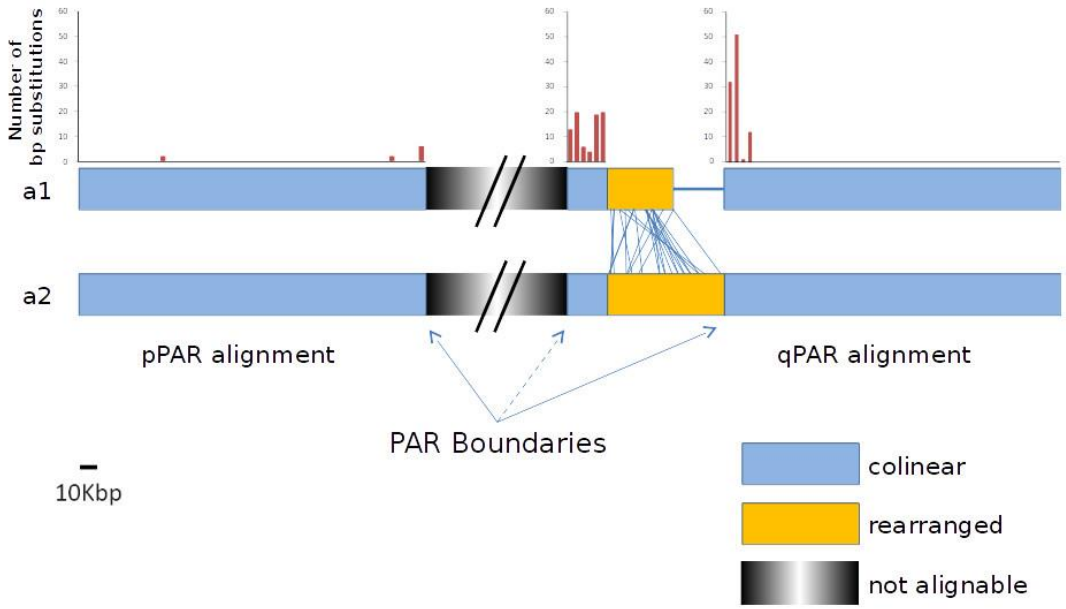


Figure S7: **Map of the base-pair substitutions in pseudo-autosomal regions of the *Microbotryum lychnidis-dioicae* mating-type chromosomes**. A collinear region of 23 kb in the non-recombining region is identified close to the qPAR boundary.

A.3 Sex and parasites : genomic and transcriptomic analysis of *Microbotryum lychnidis-dioicae*, the biotrophic and plant-castrating anther smut fungus

Additional file 1. Genome sequencing statistics.

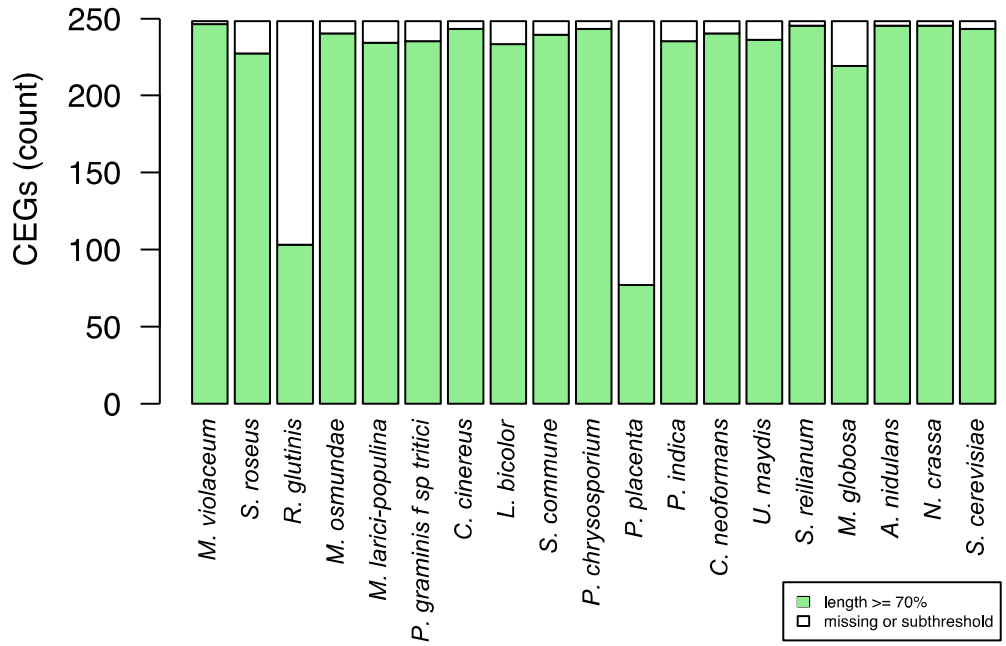
454 Library Insert	Sequencing Type	Molecular Barcode	Reads (count)	Average Read Length (bases)	PF Bases (count)	Sequence Depth (Fold)
Fragment	Single read	Madrid	1,145,609	371.3	425,393,315	16.36
2.7 kb	Paired read	Bratislava	510,499	280.2	143,013,765	5.50
5.5 kb	Paired read	Cairo	686,469	162.2	111,354,758	4.28
5.5 kb	Paired read	Budapest	233,405	189.1	44,129,203	1.70
Total			2,575,982		723,891,041	27.84

Additional file 2. RNA-Seq read statistics.

Sample	Reads	PF* Reads	PF Reads aligned**	PF Reads aligned (%)**
Haploid rich	17,025,508	16,239,792	14,613,444	89.99%
Haploid nutrient limited	19,275,930	18,335,352	16,366,386	89.26%
MI-late	20,165,846	19,575,594	4,441,268	22.69%
Total	56,467,284	54,150,738	35,421,098	

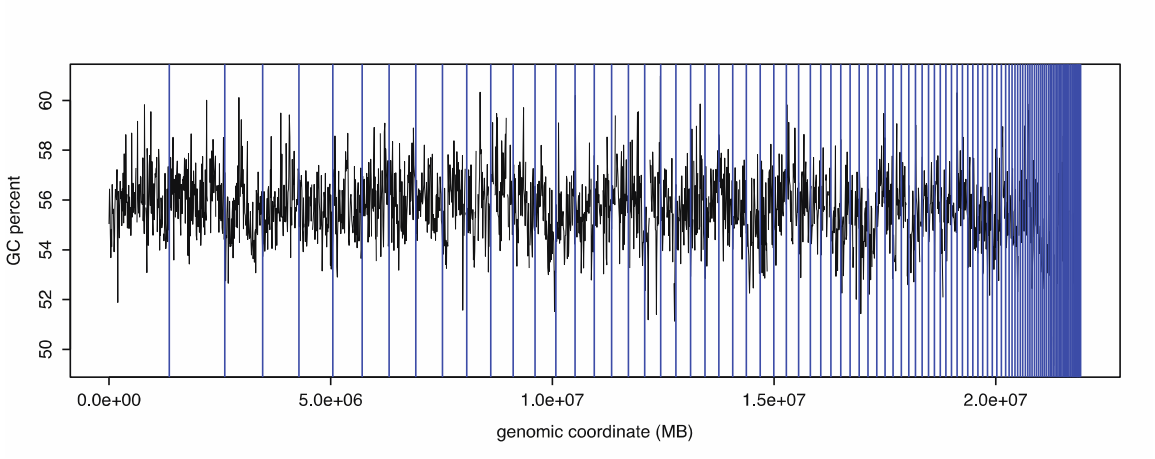
*Illumina Passing Filter high quality criteria.
**Number and percent of PF reads aligned to *M. lychnidis-dioicae* genome

Additional File 3. Conservation of core eukaryotic genes (CEGs) set across *M. lychnidis-dioicae* and other fungal genomes. The percent coverage of genes with significant Blast similarity is shown for alignments above and below the recommended 70% coverage threshold, which can indicate partial gene structures.

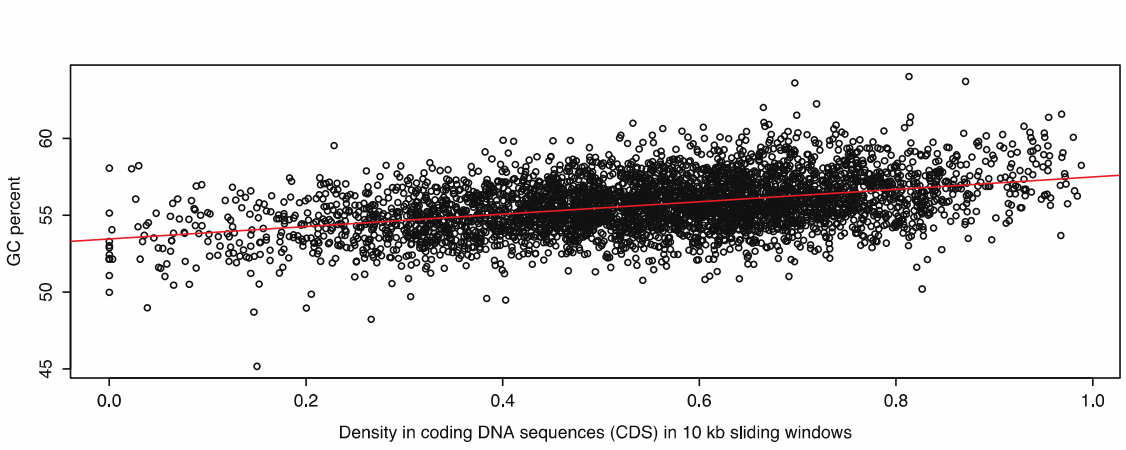


Additional file 4. Correlation between GC content and gene density.
A . Using the assembly generated from 454 sequencing, GC percent was computed in non-overlapping windows of 10 kbp and plotted against genomics coordinates. Vertical blue lines represent the limits between scaffolds. B. Gene density was also measured in the same windows as in A. There was a significant positive correlation (p-value < 2.3e-05) between density in (CDS) and GC percent, explaining 16.8 % of the variance.

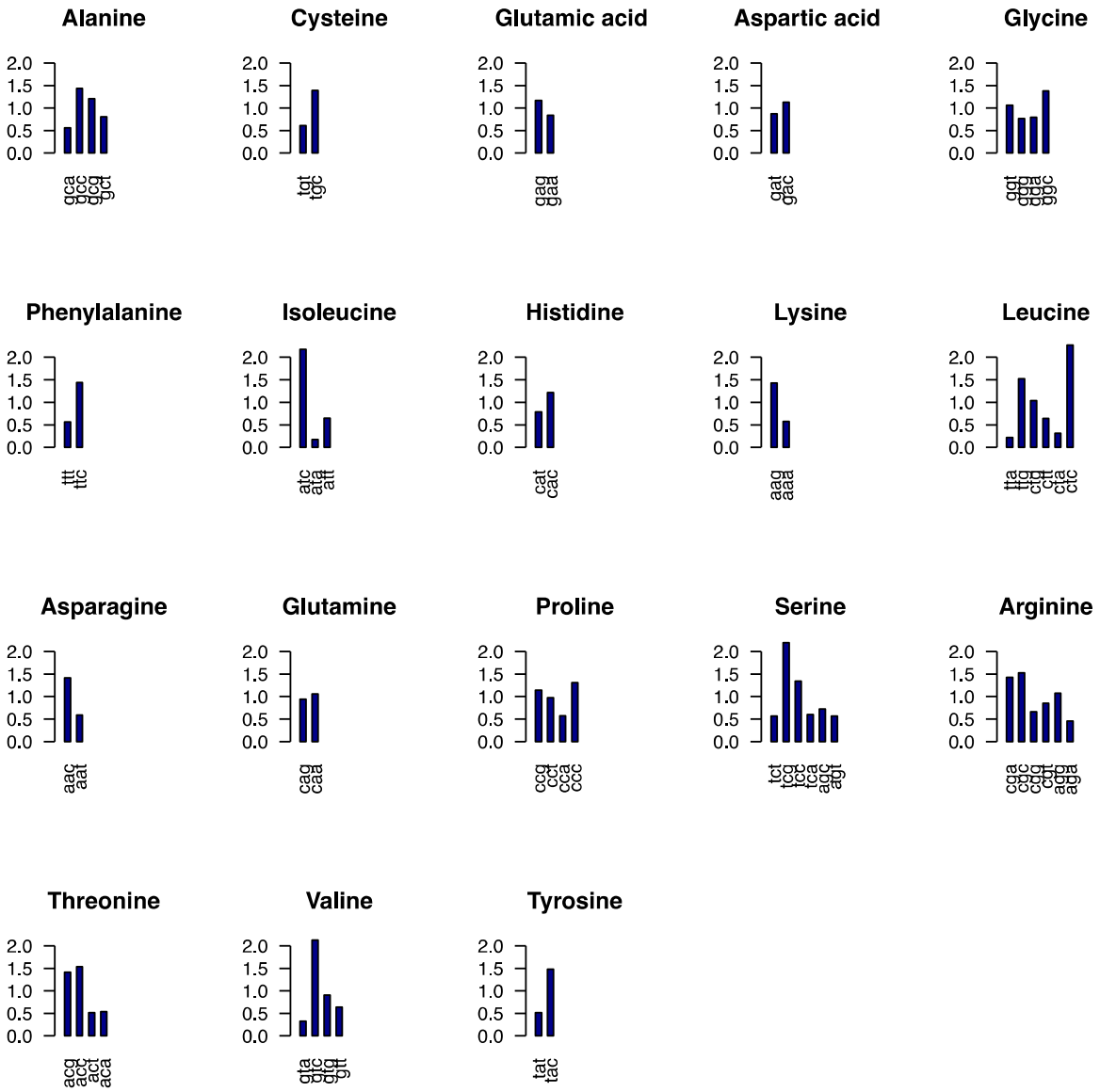
A



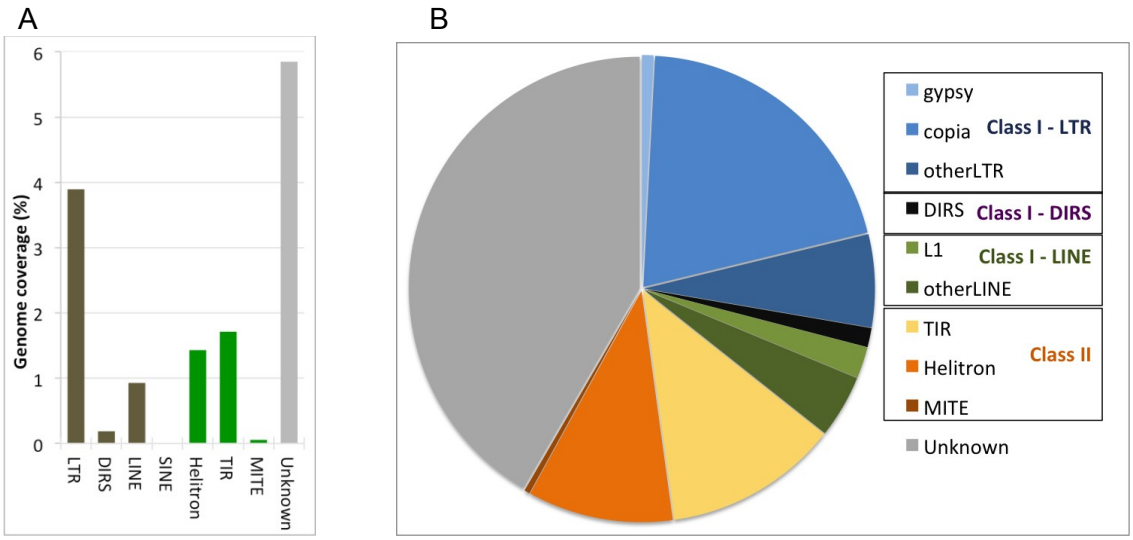
B



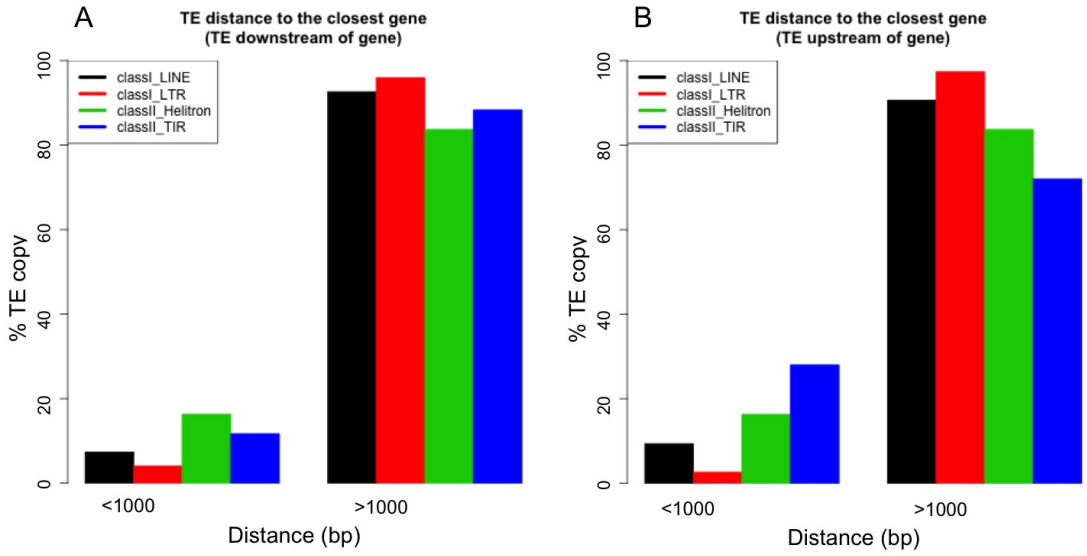
Additional file 5. Preferred codons for each amino-acid. The frequencies of the different codons are shown for each amino-acids. The most frequent codons often show higher GC content.



Additional file 6. **Distribution of transposable elements in *M. lychnidis-dioicae* genome according to their TE classification.** A. Genome coverage (%) of TEs according to their order; B. TE space coverage (%) of TEs according to superfamilies of the major orders (LTR, LINE, DIRS retro-transposons, TIR, Helitron and MITE DNA transposons).

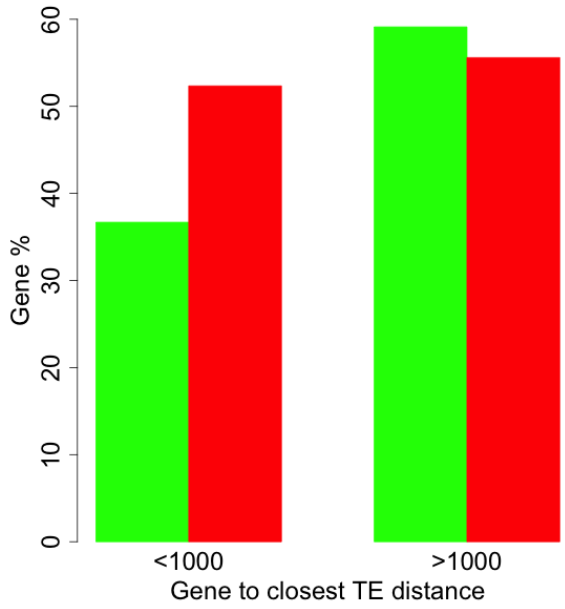


Additional file 7. Proximity of TE copies for main orders (Class I LTR and LINE, Class II TIR and Helitrons) to the closest gene. Two classes of distance were compared: < 1kb and > 1kb. A. Upstream regions of genes (Pearson's Chi-squared test p-value < 2.2e-16) and B. Downstream regions of genes (Pearson's Chi-squared test p-value < 2.2e-08)

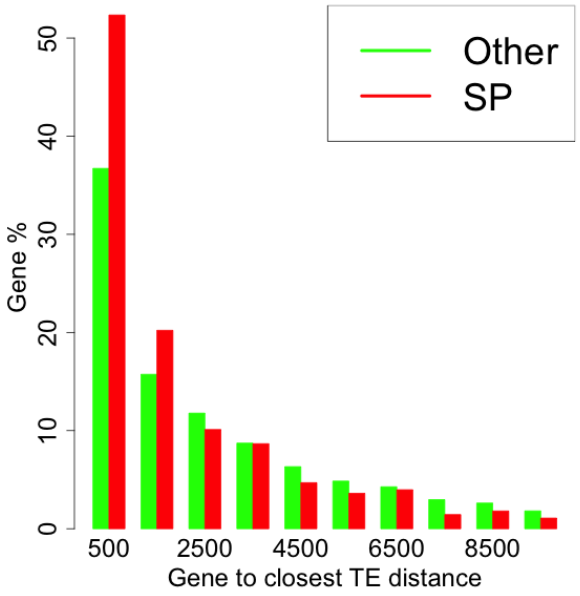


Additional file 8. Comparison of TE proximity for secreted proteins compared to all other genes. The percent of genes (Gene %) is shown for either A. Two classes of distance 0-1kb and >1kb (Pearson's Chi-squared test p-value = 5e-4) and B. Ten classes of distance from 0-1 kb, to 9-10 kb (Pearson's Chi-squared test p-value = 1e-2).

A



B

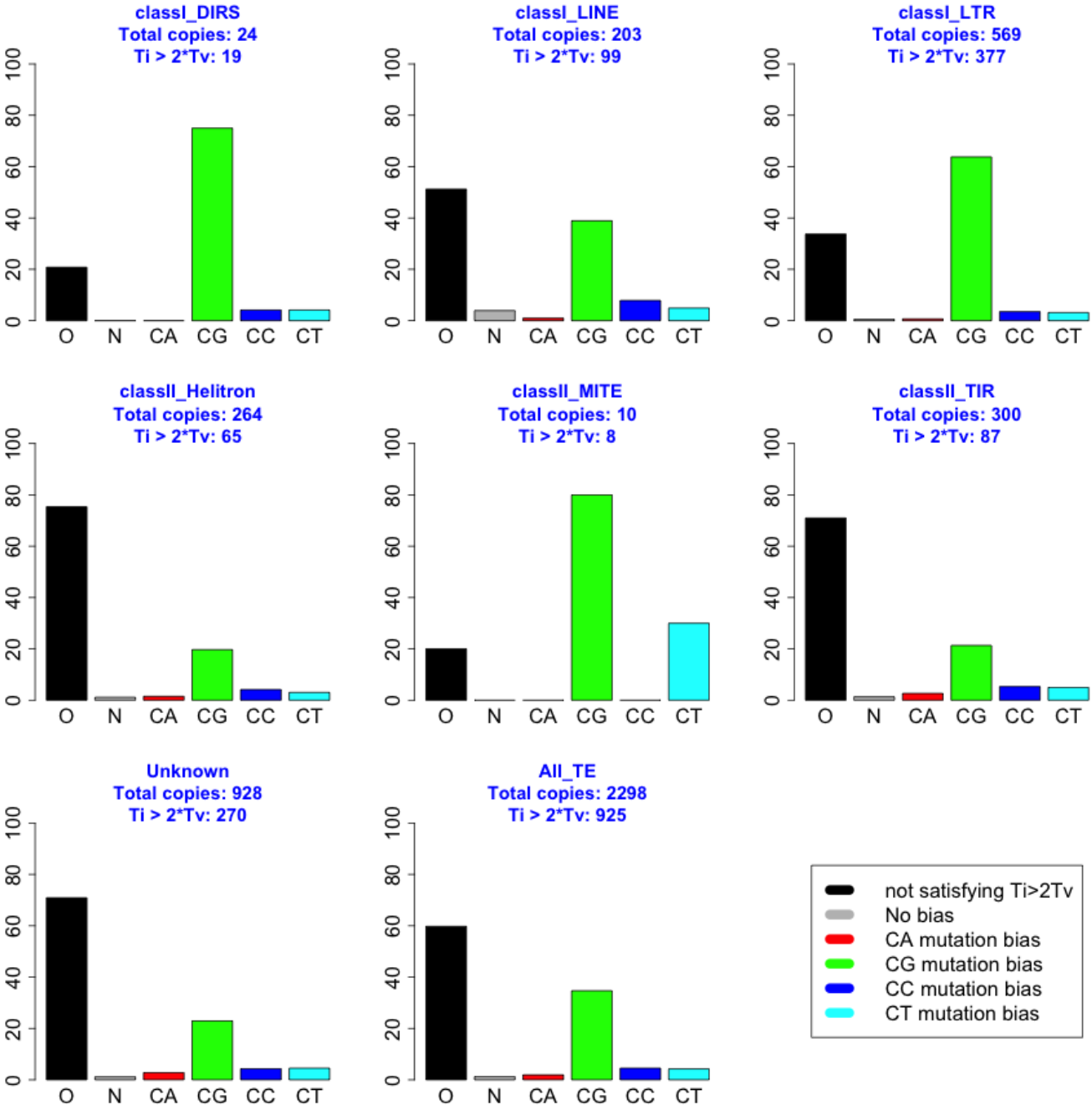


Additional file 9. TE used in RIPCAL analysis. 2,298 copies from 179 TE families were used in RIPCAL calculations. GC% is calculated on whole sequences used.

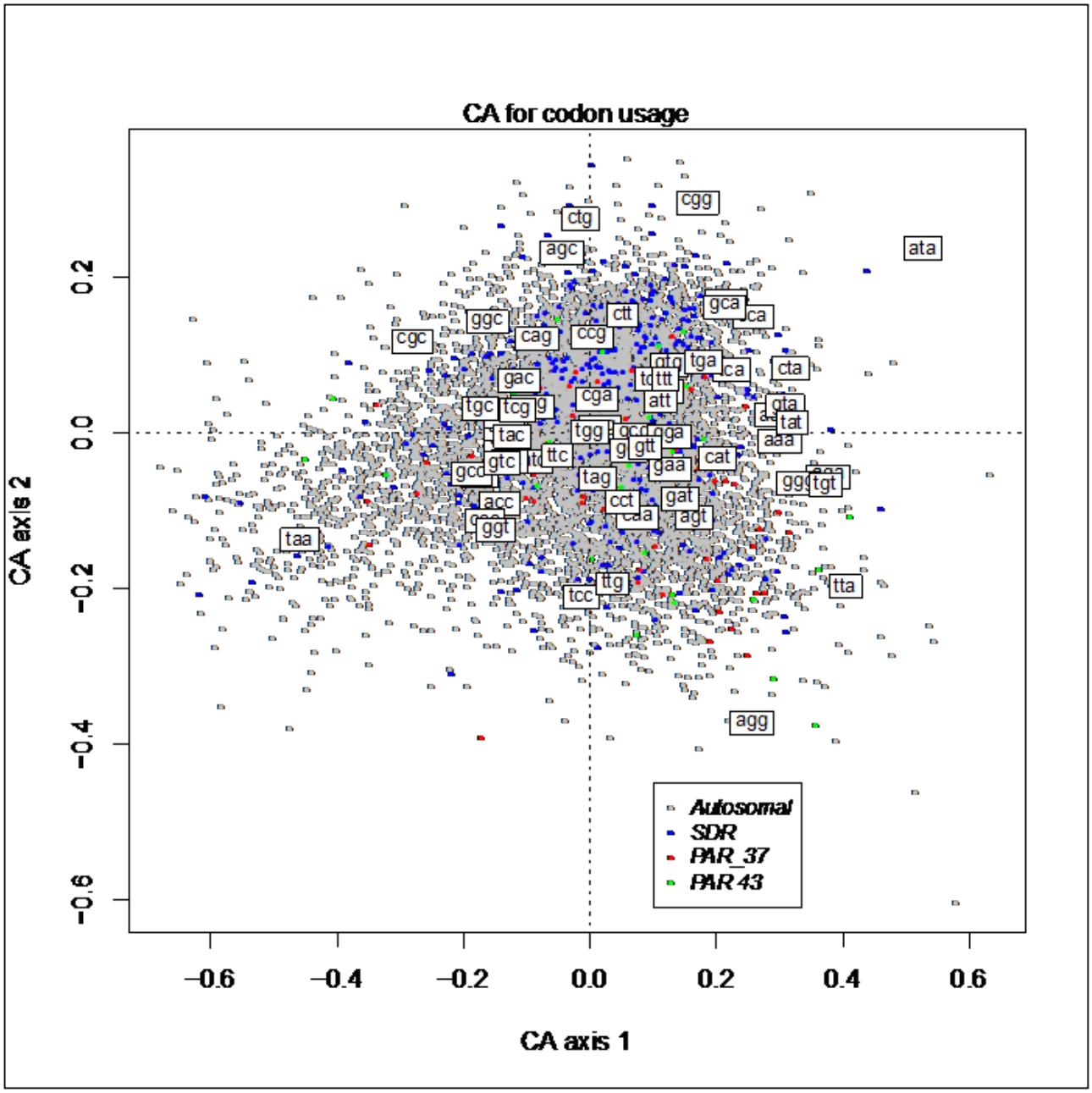
LTR families	LTR copies	DIRS families	DIRS copies	LINE families	LINE copies	TIR families	TIR copies	MITE families	MITE copies	Helitron families	Helitron copies	Unknown families	Unknown copies
37	569	2	24	11	203	18	300	1	10	15	264	95	928

GC% copies	GC% genes	GC% genome
54.4	56.04	55.43

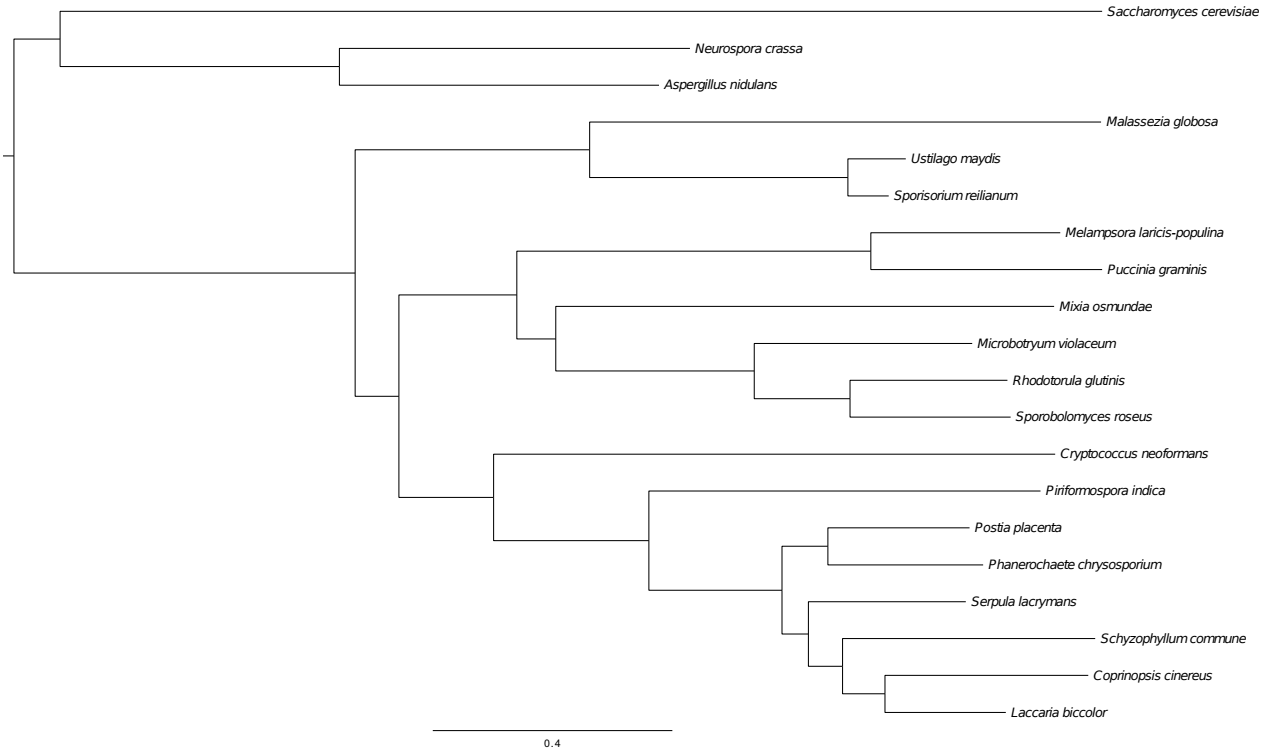
Additional file 10. Frequency of mutation transition types in different TE classes. For each TE type, the frequency of di-nucleotide substitutions are plotted; these are measured between each copy (total of 2,298 copies) and the highest GC content sequence in a multiple alignment of the TE consensus and the corresponding genome copies. Colored bars: percentage of copies with expected RIP-like mutation (if Transition rate > 2 * Transversion rate) and DI-nucleotide preferentially used >30% in CN->TN and (cNG -> cNA) mutations. Black bar: percentage of copies without expected RIP-like (Transition rate < 2 * Transversion rate). Note that the CA+CG+CC+CT percentage of copy could be over than 100% if copies exhibit more than one bias.



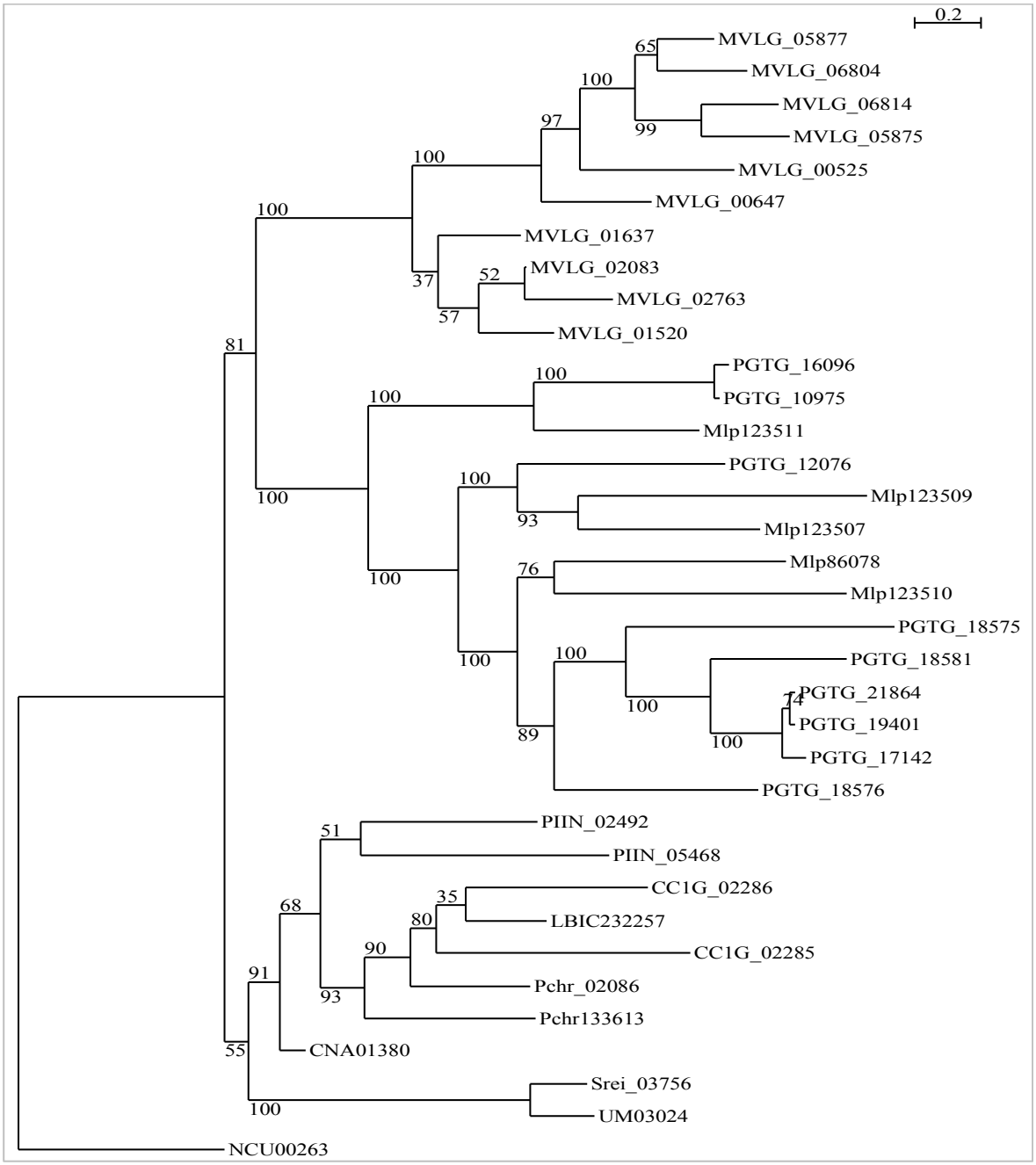
Additional file 12. Codon adaptation of autosomal and sex-specific genome regions.
Correspondence analysis of synonymous codon usage, which compares sets of genes belonging to non-recombining mating type regions, pseudoautosomal regions ("PAR") and autosomes.



Additional file 15. Phylome for *Microbotryum lychnidis-dioicae*. Maximum likelihood tree of 20 fungal species including *M. lychnidis-dioicae*, and other basidiomycete and ascomycete species. The tree has been rooted by the midpoint method for clarity. The alignments of 52 proteins with a single ortholog in each of the 20 studied species were concatenated into a single trimmed alignment of 48,538 positions. The phylome in this study was reconstructed using the pipeline described in (Huerta-Cepas et al. 2011) and can be browsed on-line (www.phylomedb.org, phylome code 180).



Additional file 17. Phylogenetic tree of DUF1034 proteins. Protein sequences were aligned with MUSCLE, and a phylogeny was inferred from this alignment using RAxML with the following settings: substitution model: PROTCAT; Matrix name: DAYHOFF; algorithm: (d) Hill-climbing-default and *N. crassa* as the outgroup. A total of 1,000 bootstrap replicates were performed, with the option Bootstrap random seed (b) 12345, and the percent of replicates shown on the tree nodes. All *M. lychnidis-dioicae* representatives are indicated by the MVLG_ prefix; other prefixes correspond to other species as follows: NCU, *N. crassa*; Srei, *S. reilianum*; UM, *U. maydis*; CNA, *C. neoformans*; CC1G, *C. cinereus*; Pchr, *P. chrysosporium*; LBIC, *L. bicolor*; PIIN, *P. indica*; PGTG, *P. graminis*; Mlp, *M. larici-populina*.

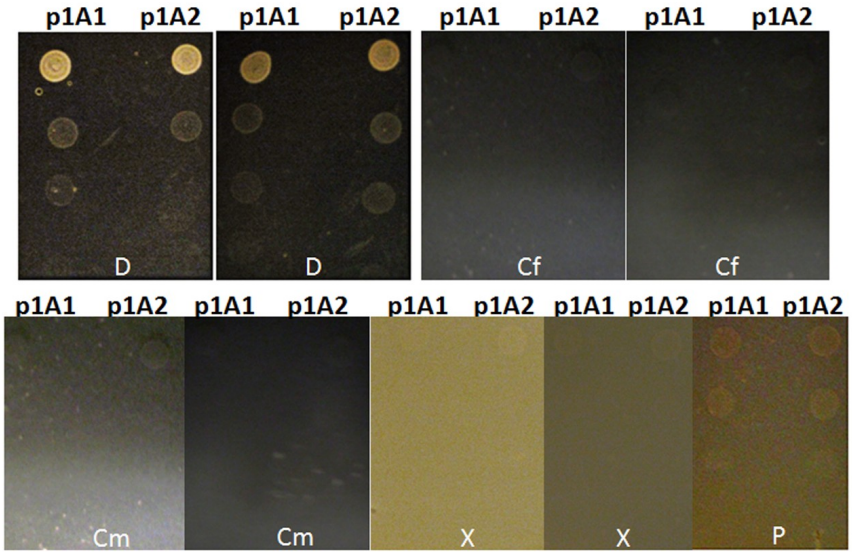


Additional file 18. qRT-PCR validation of secretory lipase expression

Strain(s)	Condition	Gene	Log fold Change vs. p1A1 Rich	Direction
p1A1	Water	MVLG_07291	0.77	Up
		MLVG_00914	0.327	Up
		MLVG_05549	1.922	Up
		MLVG_07229 & 07284[a]	1.899	Up
p1A2	Water	MVLG_07291	1.84	Up
		MLVG_00914	-1.308	Down
		MLVG_05549	0.327	Up
		MLVG_07229 & 07284	2.034	Up
Mated (p1A1 x p1A2)	Water	MVLG_07291	1.424	Up
		MLVG_00914	-1.098	Down
		MLVG_05549	0.672	Up
		MLVG_07229 & 07284	1.828	Up
Mated (p1A1 x p1A2)	Phytol	MVLG_07291	0.913	Up
		MLVG_00914	1.2	Up
		MLVG_05549	1.499	Up
		MLVG_07229 & 07284	2.098	Up
Mated (p1A1 x p1A2)	MI-late	MVLG_07291	-1.897	Down
		MLVG_00914	-1.772	Down
		MLVG_05549	0.77	Up
		MLVG_07229 & 07284	-0.564	Down

[a]The primers used could not distinguish between MVLG_07229 & 07284; so results are presented here for both, assuming their expression is similar.

Additional file 21. Growth of *M. lychnidis-dioicae* on different carbon sources, including dextrose, cellulose, xylan, and pectin. Strains p1A1 and p1A2 were grown overnight in HSS liquid medium [133] containing 50 mM $(NH_4)_2SO_4$ and 2% dextrose. Cells were collected by centrifugation and washed in sterile distilled H_2O, then resuspended in sterile distilled H_2O to an A_{600} of 1. Onto each type of agar, 10 ml samples were spotted of undiluted, and 10-fold, 100-fold, and 1000-fold dilutions of each strain. Growth tests were conducted in duplicate for each strain on each type of medium and plates were incubated at 26˚C for 96 h. Each agar plate consisted of HSS medium containing 50 mM $(NH_4)_2SO_4$ and 2% of the indicated carbon source: D, dextrose; Cf, cellulose, fibrous (Spectrum Chemical); Cm, cellulose, microcrystalline (Spectrum Chemical); X, xylan (TCI America); P, pectin (TCI America).

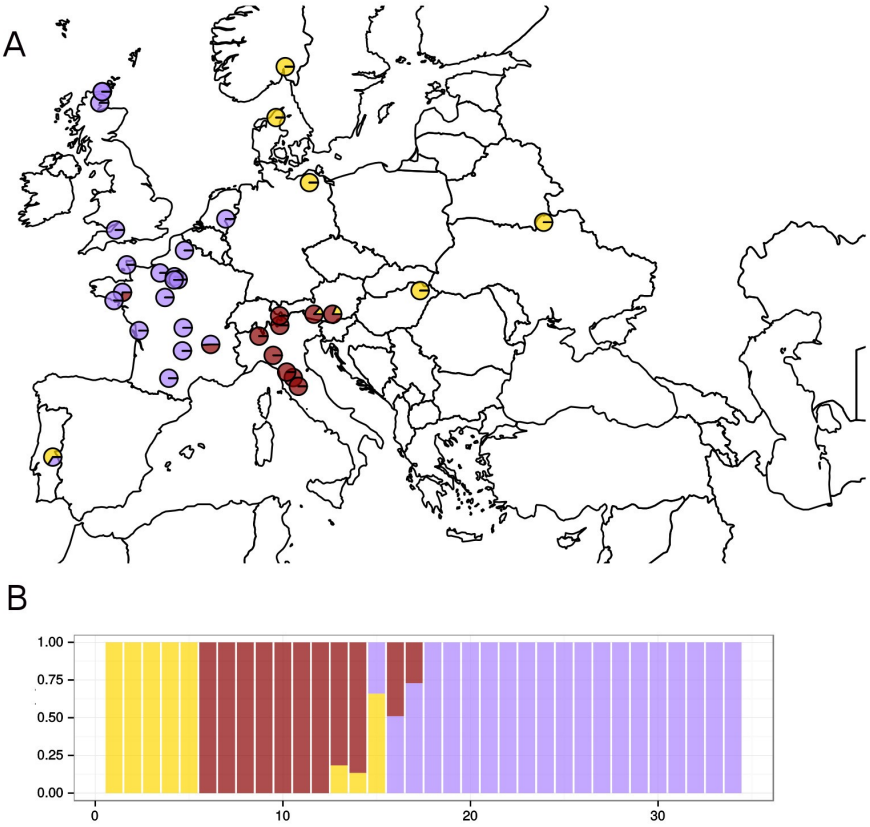


Additional file 25. PFAM domains enriched in MI-late induced genes.

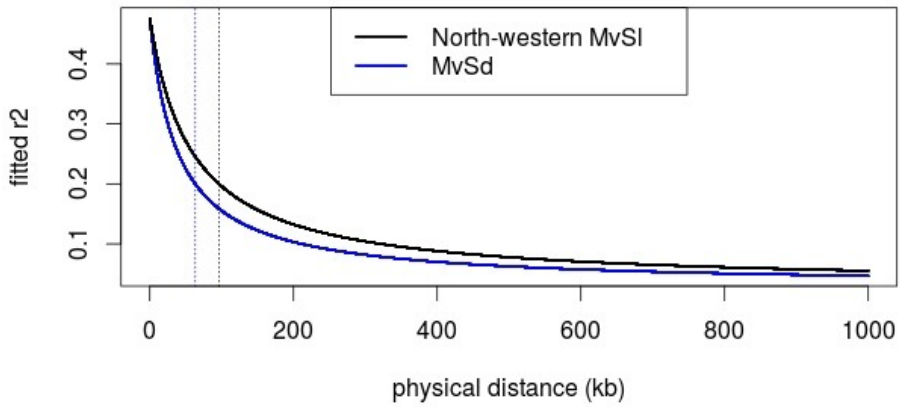
PFAM domain	MI-late Induced genes	Other genes	Fisher pvalue	Corrected pvalue
PF07690.9 Major Facilitator Superfamily	15	113	8.42E-07	0.0011
TIGR00879 MFS transporter, sugar porter (SP) family	6	14	7.96E-06	0.0051
PF07250.4 Glyoxal oxidase N-terminus	4	4	2.83E-05	0.0091
PF09118.4 Domain of unknown function (DUF1929)	4	4	2.83E-05	0.0091
PF00295.10 Glycosyl hydrolases family 28	3	3	0.0003	0.0838
PF00128.17 Alpha amylase, catalytic domain	3	4	0.0006	0.0962
PF01590.19 GAF domain	3	4	0.0006	0.0962
PF03169.8 OPT oligopeptide transporter protein	4	12	0.0006	0.0962
PF05978.9 Ion channel regulatory protein UNC-93	2	0	0.0007	0.0962

A.4 Identifying genomic regions under selection using population genomics in the anther-smut fungus

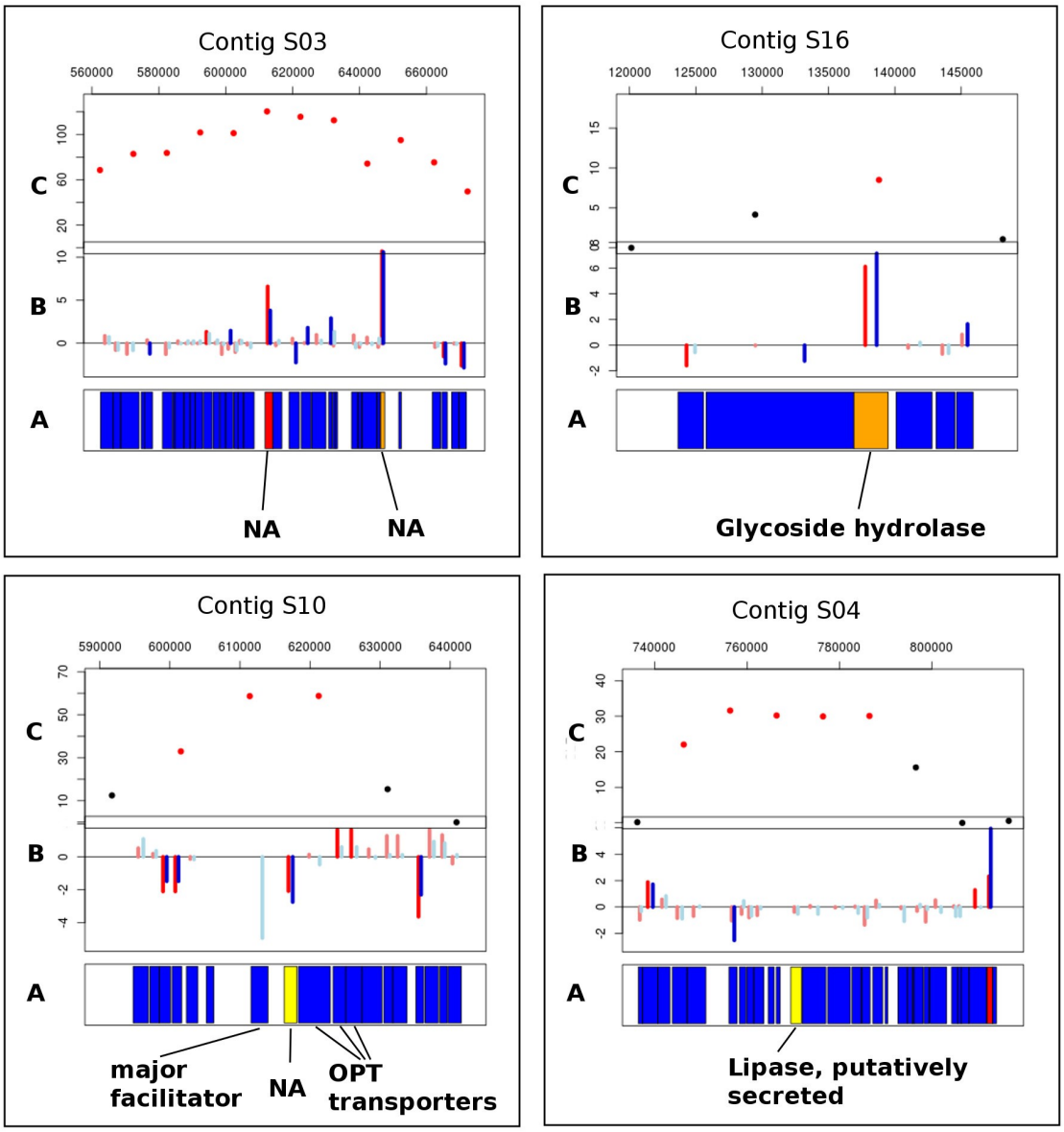
-Supp Figure 1: Population genetic structure in *Microbotryum lychnidis-dioicae* (MvSl) for K = 3. A) Map of sampled individuals in MvSl, with cluster membership for K=3, represented by colors. B) Cluster membership for K=3 in MvSl individuals, represented as vertical bars.



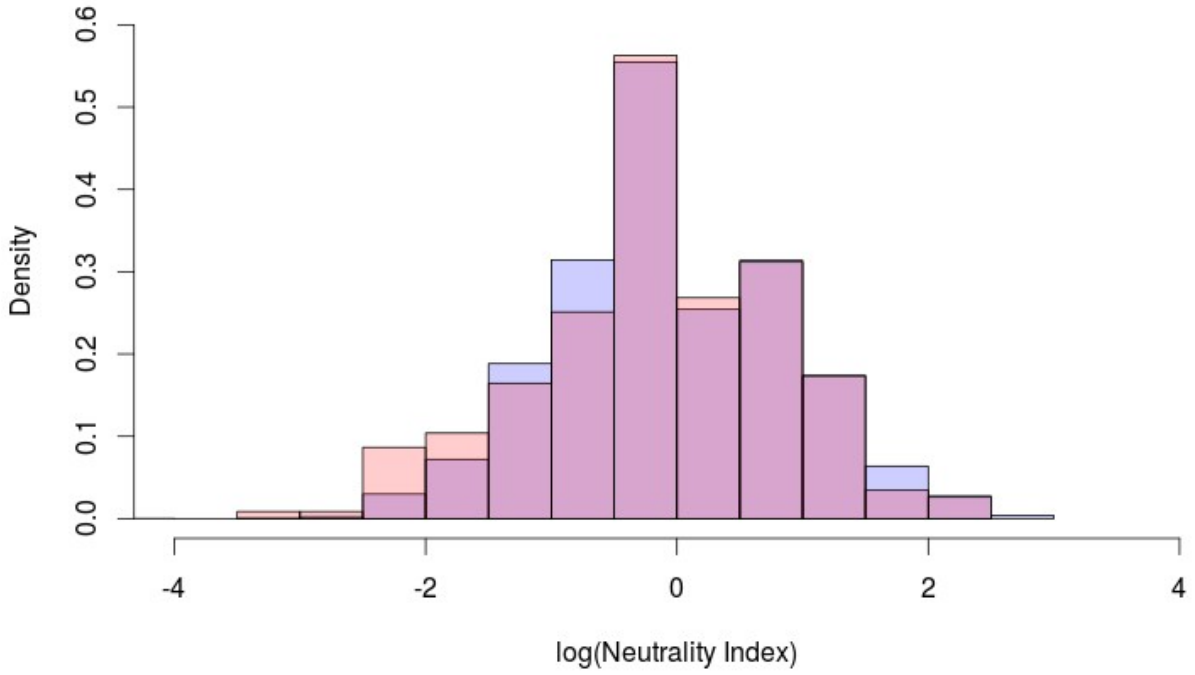
-Supp Figure 2: Decay of linkage disequilibrium (r^2) with physical distance in the North-western cluster of *Microbotryum lychnidis-dioicae* and in *M. silenes-dioicae*. Vertical dotted lines indicate the distance at which fitted r^2 decreases below 0.2 in each group.



-Supp Figure 3: Detail of 4 putative several selective sweeps in MvSl. A : location of gene models, indicated putative secretion signals (yellow); upregulation *in planta* compared to expression in the haploid stage cultivated *in vitro* in in rich or nutrient-limited medium (red); both putative secretion signal and upregulation *in planta* (orange). B : Expression pattern of genes, indicated as the log2(FPKM) of *in planta* versus *in vitro* rich mediua (blue) or *in planta* versus nutrient-limited conditions (red). Brighter shades indicate significant differential expression between *in planta* and each *in vitro* condition.



-Supp Figure 4: Distribution of the neutrality index (log10(NI)) for the genes up-regulated *in planta* (red) compared to other genes (blue).



Supp Table 1 : Samples used in the study.

sample ID	species	host species	country	mating-type
MvLf-1062-A1	M. sensus lato	Lychnis flos-cuculi	Russia	A1
MvSd-1030	M. silenes-dioicae	Silene dioica	Danemark	A1 A2
MvSd-1034	M. silenes-dioicae	Silene dioica	France	A1 A2
MvSd-1056	M. silenes-dioicae	Silene dioica	United Kingdom	A1 A2
MvSd-336-01-A1	M. silenes-dioicae	Silene dioica	France	A1
MvSd-578-2-A2	M. silenes-dioicae	Silene dioica	France	A2
MvSd-637	M. silenes-dioicae	Silene dioica	Netherlands	A1 A2
MvSd-69-05-A2	M. silenes-dioicae	Silene dioica	Switzerland	A2
MvSd-707-1-A2	M. silenes-dioicae	Silene dioica	France	A2
MvSd-72-A2	M. silenes-dioicae	Silene dioica	Switzerland	A2
MvSd-849-7	M. silenes-dioicae	Silene dioica	Slovenia	A1 A2
MvSd-851-6-A2	M. silenes-dioicae	Silene dioica	Slovakia	A2
MvSd-900-1-A1	M. silenes-dioicae	Silene dioica	Norway	A1
MvSd-932-2-A2	M. silenes-dioicae	Silene dioica	France	A2
MvSd-937-2-A2	M. silenes-dioicae	Silene dioica	France	A2
MvSd-949-2-A1	M. silenes-dioicae	Silene dioica	United Kingdom	A1
MvSd-IT02-32-2-17A-A2	M. silenes-dioicae	Silene dioica	Italy	A2
MvSd-sp002-A1	M. silenes-dioicae	Silene dioica	France	A1
MvSd-sp003-A1	M. silenes-dioicae	Silene dioica	France	A1
MvSd-335-H3-A1	M. lychnidis-dioicae	Silene latifolia	Italy	A1
Lamole strain	M. lychnidis-dioicae	Silene latifolia	Italy	A1 A2
MvSl-00-10-A2	M. lychnidis-dioicae	Silene latifolia	Italy	A2
MvSl-100-6-A1	M. lychnidis-dioicae	Silene latifolia	Italy	A1
MvSl-1005	M. lychnidis-dioicae	Silene latifolia	Portugal	A1 A2
MvSl-1040	M. lychnidis-dioicae	Silene latifolia	United Kingdom	A1 A2
MvSl-1049	M. lychnidis-dioicae	Silene latifolia	France	A1 A2
MvSl-1067	M. lychnidis-dioicae	Silene latifolia	France	A1 A2
MvSl-1134	M. lychnidis-dioicae	Silene latifolia	France	A1 A2
MvSl-1140-3	M. lychnidis-dioicae	Silene latifolia	France	A1 A2
MvSl-140-01-A2	M. lychnidis-dioicae	Silene latifolia	France	A2
MvSl-141-01-A2	M. lychnidis-dioicae	Silene latifolia	France	A2
MvSl-40-01-A2	M. lychnidis-dioicae	Silene latifolia	France	A2
MvSl-443-2	M. lychnidis-dioicae	Silene latifolia	Danemark	A1 A2
MvSl-446-2-A2	M. lychnidis-dioicae	Silene latifolia	Hungary	A2
MvSl-462-3	M. lychnidis-dioicae	Silene latifolia	United Kingdom	A1 A2
MvSl-466-3-A1	M. lychnidis-dioicae	Silene latifolia	France	A1
MvSl-576-A2	M. lychnidis-dioicae	Silene latifolia	Germany	A2
MvSl-641-A1	M. lychnidis-dioicae	Silene latifolia	Netherlands	A1
MvSl-660	M. lychnidis-dioicae	Silene latifolia	France	A1 A2
MvSl-661-A2	M. lychnidis-dioicae	Silene latifolia	France	A2
MvSl-687-5-A2	M. lychnidis-dioicae	Silene latifolia	France	A2
MvSl-699	M. lychnidis-dioicae	Silene latifolia	Ukraine	A1 A2
MvSl-728-4	M. lychnidis-dioicae	Silene latifolia	France	A1 A2
MvSl-781	M. lychnidis-dioicae	Silene latifolia	Italy	A1 A2
MvSl-830-2-A1	M. lychnidis-dioicae	Silene latifolia	Italy	A1
MvSl-856-2-A1	M. lychnidis-dioicae	Silene latifolia	France	A1
MvSl-925	M. lychnidis-dioicae	Silene latifolia	Norway	A1 A2
MvSl-933-1	M. lychnidis-dioicae	Silene latifolia	France	A1 A2
MvSl-973	M. lychnidis-dioicae	Silene latifolia	France	A1 A2
MvSl-974	M. lychnidis-dioicae	Silene latifolia	France	A1 A2
MvSl-979-4	M. lychnidis-dioicae	Silene latifolia	Austria	A1 A2
MvSl-980	M. lychnidis-dioicae	Silene latifolia	Austria	A1 A2
MvSl-I00-3_A2	M. lychnidis-dioicae	Silene latifolia	Italy	A2
MvSl-IOA-A1	M. lychnidis-dioicae	Silene latifolia	Italy	A1

Supp Table 2 : Mapping statistics

Individual	total number of reads	number of reads mapped as proper pairs	% of reads mapped as proper pairs
MvLf-1062-A1	27787088	11118322	0.4001254827
MvSd-1030	37174614	30568652	0.8222991098
MvSd-1034	24468480	19404892	0.7930566999
MvSd-1056	48670486	39803254	0.8178109008
MvSd-336-01-A1	43009516	36166164	0.840887491
MvSd-578-2-A2	25484586	21076410	0.8270257951
MvSd-637	43258018	35685554	0.8249465798
MvSd-69-05-A2	34637376	29985550	0.8656992377
MvSd-707-1-A2	44193006	36790780	0.8325023195
MvSd-72-A2	51820932	43143700	0.8325535326
MvSd-849-7	37404044	32876310	0.8789506824
MvSd-851-6-A2	32059056	27738724	0.8652383277
MvSd-900-1-A1	41877776	34762018	0.830082715
MvSd-932-2-A2	33696174	28725156	0.8524752988
MvSd-937-2-A2	38720758	33763548	0.8719753885
MvSd-949-2-A1	36285024	30638176	0.8443752442
MvSd-IT02-32-2-17A-A2	38523646	33172436	0.8610928467
MvSd-unk002-A1	45955680	38661034	0.8412678041
MvSd-unk003-A1	33322234	27241284	0.8175107347
MvSd-335-H3-A1	36395320	30260324	0.8314344811
MvSl-1064-A1A2	77394012	67832526	0.8764570313
MvSl-00-10-A2	37438808	35535562	0.9491638195
MvSl-100-6-A1	35287294	32195298	0.9123765058
MvSl-1005	31341354	26861886	0.8570748411
MvSl-1040	39333798	34053190	0.8657488402
MvSl-1049	37508314	32605384	0.8692841806
MvSl-1067	42603318	36866800	0.8653504405
MvSl-1134	30775016	27743988	0.9015101081
MvSl-1140-3	35765854	21164520	0.591752122
MvSl-140-01-A2	77394012	68228618	0.881574895
MvSl-141-01-A2	49616102	30182292	0.6083164695
MvSl-40-01-A2	41959070	31425656	0.7489597839
MvSl-443-2	47353584	41367636	0.8735903918
MvSl-446-2-A2	35045644	30905880	0.881875077
MvSl-462-3	38378000	33788836	0.8804220126
MvSl-466-3-A1	42324746	31927230	0.7543395535
MvSl-576-A2	32815136	24734728	0.7537597284
MvSl-641-A1	42456520	35878100	0.8450551293
MvSl-660	33357788	29279164	0.8777309814
MvSl-661-A2	40886926	36405274	0.8903891185
MvSl-687-5-A2	21379934	19241822	0.8999944527
MvSl-699	56481778	50516934	0.8943934803
MvSl-728-4	41546562	35942958	0.8651247244
MvSl-781	33563666	30380030	0.9051463568
MvSl-830-2-A1	40781214	37435968	0.9179709069
MvSl-856-2-A1	18535170	16310718	0.8799875048
MvSl-925	52447336	47298248	0.9018236503
MvSl-933-1	34942576	30059102	0.8602428739
MvSl-973	44983496	40199164	0.8936425039
MvSl-974	45444592	40078610	0.8819225399
MvSl-979-4	49158814	40428798	0.822411989
MvSl-980	39102282	35300060	0.9027621457
MvSl-I00-3_A2	39555786	37234266	0.9413102296
MvSl-IOA-A1	32339534	29533188	0.9132224354

Supp Table 3 : Number of SNPs per sample

Strain	Total number of SNPs	Number of Heterozygous SNPs	% of heterozygous SNPs	Isolated from monospores
MvSd-1030	145554	9195	6.32	No
MvSd-1034	137220	5930	4.32	No
MvSd-1056	147840	11058	7.48	No
MvSd-336-01-A1	148122	10457	7.06	No
MvSd-578-2-A2	144096	9446	6.56	No
MvSd-637	146207	10942	7.48	No
MvSd-69-05-A2	146827	9842	6.7	No
MvSd-707-1-A2	148382	10485	7.07	No
MvSd-72-A2	148902	10281	6.9	No
MvSd-849-7	144416	8734	6.05	No
MvSd-851-6	146267	10032	6.86	No
MvSd-900-1-A1	145119	9483	6.53	No
MvSd-932-2-A2	147398	9134	6.2	No
MvSd-937-2-A2	148668	9991	6.72	No
MvSd-949-2-A1	148502	9529	6.42	No
MvSd-IT02-32-2-17A-A2-1141	146287	8972	6.13	Yes
MvSd-unk002-A1	148861	10361	6.96	No
MvSd-unk003	147304	11133	7.56	No
MvSd-335-H3-A1	148560	9496	6.39	Yes
MvSl-1064	298	235	78.86	No
MvSl-00-10-A2	28814	1971	6.84	Yes
MvSl-100-6-A1	36002	2553	7.09	No
MvSl-1005	75283	3244	4.31	No
MvSl-1040	72868	4962	6.81	No
MvSl-1049	75435	10770	14.28	No
MvSl-1067	74039	5445	7.35	No
MvSl-1134	72528	4706	6.49	No
MvSl-1140-3	72024	4527	6.29	No
MvSl-140-01-A2	73840	5098	6.9	No
MvSl-141-01-A2	68292	4224	6.19	No
MvSl-40-01-A2	73645	5262	7.15	Yes
MvSl-443-2	62591	4466	7.14	No
MvSl-446-2	61738	5592	9.06	No
MvSl-462-3	73175	5712	7.81	No
MvSl-466-3	73806	6050	8.2	No
MvSl-576	60928	5620	9.22	No
MvSl-641-A1	73432	5863	7.98	No
MvSl-660	80102	27598	34.45	No
MvSl-661-A2	72871	5300	7.27	No
MvSl-687-5	70906	27138	38.27	No
MvSl-699	60804	3558	5.85	No
MvSl-728-4	72962	5627	7.71	No
MvSl-781	33760	2223	6.58	No
MvSl-830-2-A1	34761	2435	7	No
MvSl-856-2-A1	70308	5280	7.51	No
MvSl-925	62380	3824	6.13	No
MvSl-933-1	72639	4948	6.81	No
MvSl-973	73254	5150	7.03	No
MvSl-974	73253	5206	7.11	No
MvSl-979-4	45427	3195	7.03	No
MvSl-980	44913	4129	9.19	No
MvSl-I00-3_A2-1138	27479	2024	7.37	Yes
MvSl-IOA-A1	17808	1350	7.58	Yes

Supp Table 4 : Correlation between nucleotidic diversity (Pi) per bp in MvSd and MvSl clusters

Pi on 100 kb				Pi on 50 kb			
Spearman's rho	MvSl Italy	MvSl East	MvSd	Spearman's rho	MvSl Italy	MvSl East	MvSd
MvSl North West	0.464***	0.423***	0.262***	MvSl North West	0.404***	0.392***	0.249***
MvSl Italy		0.422***	0.408***	MvSl Italy		0.387***	0.362***
MvSl East			0.321***	MvSl East			0.324***

Supp Table 5 : Spearman's rho and significance level for correlation of nucleotitic diversity (Pi) and GC content, local linkage disequilibrium (LD), density in coding sites, DN/DS and number of fixed differences between MvSl and MvSd.

A – on sliding windows of 100 kb

	GC content	Density in CDS	mean LD	DN/DS	per base number of fixed differences with MvSd / MvSl	GC12	GC3
MvSl North West	-0.171**	0.169*	-0.392***	0.0015	-0.392***	-0.324***	-0.0535
MvSl Italy	-0.286***	0.121*	-0.463***	-0.0459	-0.397***	-0.328***	-0.08
MvSl East	-0.142*	0.075	-0.335***	0.0048	-0.4***	-0.167*	-0.009
MvSd	-0.301***	0.049	-0.215**	-0.0364	-0.292***	-0.265***	-0.203**

B – on sliding windows of 50 kb

	GC content	Density in CDS	mean LD	DN/DS	per base number of fixed differences with MvSd / MvSl	GC12	GC3
MvSl North West	-0.147**	0.155***	-0.239***	0.0131	-0.324***	-0.242***	-0.047
MvSl Italy	-0.267***	0.102*	-0.375***	-0.0049	-0.306***	-0.26***	-0.101*
MvSl East	-0.151**	0.0859	-0.281***	-0.018	-0.37***	-0.138**	-0.0235

Supp table 6 : Spearman's rho and significance level for correlation of nucleotitic diversity (Pi) and the number of fixed differences in 100 kb sliding window s :

	MvSl North West	MvSl Italy	MvSl East	MvSd
MvSl North West - Italy	-0.487***	-0.299***	-0.0952	-0.0459
MvSl North West - East	-0.467***	-0.0677	-0.452***	-0.0588
MvSl East - Italy	-0.07	-0.0796	-0.436***	0.0102
MvSd - MvSl Italy	-0.264***	-0.397***	-0.185**	-0.292***
MvSd - MvSl North-West	-0.392***	-0.149*	-0.166*	-0.228***
MvSd - MvSl East	-0.184**	-0.119	-0.4***	-0.201**

Supp Table 7 : Multiple linear regressions models for explaining synonymous diversity in the clusters of *Microbotryum lychnidis-dioicae* (MvSl) and in *M. silenes-dioicae* (MvSd), using 50 kb and 100 kb non-overlapping sliding windows. P-values : * < 0.05, ** < 0.01, * < 0.001**

A – Total GC content as explanatory variable

		r2	GC content	LD	Density in coding sites	DN/DS
100 kb	MvSl North West	0.132	-4.49E-05	-0.0002789***	-1.22E-04	-0.000347**
	MvSl Italy	0.2562	-0.0002011***	-0.0003944***	5.40E-04	-1.70E-04
	MvSl East	0.0987	-9.10E-05	-0.0002428***	1.13E-04	7.80E-05
	MvSd	0.1094	-0.00009452***	5.24E-05	-2.38E-04	2.46E-05
50 kb	MvSl North West	0.05143	-0.00008949**	-0.0001606**	-9.51E-05	-9.03E-05
	MvSl Italy	0.16	-0.0001448***	-0.0003779***	3.59E-04	-1.04E-04
	MvSl East	0.07545	-0.00008484**	-0.0002369***	1.13E-04	-5.28E-05
	MvSd	0.07207	-0.00007021***	1.71E-05	-1.60E-04	2.72E-06

B – GC12 and GC3 as explanatory variables

window size	cluster	r2	Fixed diver-gence	GC12	GC3	LD	Density in coding sites	DN/DS
100 kb	MvSl North West	0.25	-0.1702***	-3.34E-05	-1.53E-05	-0.0002842*	0.0007528*	-0.0002842*
	MvSl Italy	0.29	-0.1588***	-6.48E-05	-2.11E-05	-0.0003372***	0.0008601*	-1.85E-04
	MvSl East	0.25	-0.2199***	1.40E-05	-8.25E-06	-0.0001863**	6.52E-04	3.15E-05
	MvSd	0.22	-0.07153***	-0.00002871*	-0.00001414*	-3.34E-05	1.22E-05	9.81E-05
50 kb	MvSl North West	0.14	-0.1366***	-0.0000505*	-1.57E-05	-0.0001477**	3.73E-04	-1.03E-04
	MvSl Italy	0.21	-0.1361***	-0.00005545*	-1.60E-05	-3.465e-04**	7.10E-04	-1.13E-04
	MvSl East	0.15	-0.1466***	-2.51E-05	-8.90E-06	-0.0002094****	4.73E-04	-2.32E-05
	MvSd	0.19	-8.86E-02	-1.69E-05	-0.00001419**	-2.49E-06	-6.35E-05	-9.70E-07

Title : Evolutionary genomics in the *Microbotryum* fungi : host adaptation and mating-type chromosomes

Keywords : population genomics, mating-type chromosomes, host adaptation

Abstract : Understanding how species adapt to their environment is a major goal in evolutionary biology. Scientists aim to identity the genes underlying key adaptive traits, but also to understand more broadly adaptive processes and phenomena that allow or prevent optimal adaptation. Non-recombining regions are particular for these aspects. They can indeed protect adaptive combinations of alleles from recombination, and conversely, suppressed recombination can lead to degeneration, such as accumulation of deleterious mutations or genes losses. Even sexually-reproducing organisms often possess large non-recombining regions associated with sex ou mating-type determination. In this thesis, I therefore studied signatures of adaptation and degeneration in genomes of plant pathogenic fungi. Fungi of the species complex *Microbotryum violaceum*, with dozens of host-specific sibling species causing anther-smut disease in the Caryophyllaceae family, are particularly good models for addressing the question of the genomic processes involved in host adaptation. Moreover, they possess size-dimorphic, partly non-recombining mating-type chromosomes. To study the evolution of mating-type chromosomes, degeneration and host-adaptation in the *M. violaceum* species complex, we used a genomic approach. Using PacBio sequencing, we obtained a complete assembly of the mating-type chromosomes of the species *M. lychnidis-dioicae*. We showed that the non-recombining regions span 90 % of the mating-type chromosomes, exhibit an exceptional level of rearrangements between the two mating-types, and that hundreds of genes are in a hemizygous state and were therefore probably lost in one of the two mating-type chromosomes. Moreover, comparing a dozen of species of the *M. violaceum* complex revealed an accumulation of non-synonymous substitutions and of transposable elements in mating-type chromosomes. We also studied degeneration in the context of ionizing radiations, by analysing populations of *M. lychnidis-dioicae* exposed to different radiation levels in the Chernobyl area. We did not detect any increase in the rate of non-synonymous mutations compared to the control group or with radiation, which suggests that the fungi is radio-resistant or that selection is higher in the Chernobyl area. Lastly, we resequenced dozens of genomes of two sibling species of *M. violaceum* in order to study host adaptation. Analysing polymorphism patterns, we found several selective sweeps along the genome, at different locations in the two species. We identified candidate genes for host-adaptation in the regions of selective sweeps, based on their expression pattern and on their putative functions. In addition, genes up-regulated *in planta* exhibited a higher rate of non-synonymous substitutions than other genes, suggesting that many of them are likely involved in host adaptation.

This work paves the way to a larger comparison of genomes of different species of the *M. violaceum* species complex, in order to reconstruct the history of recombination suppression on the mating-type chromosomes on the one hand, and to study the genetic bases of adaptation to different hosts in a complex of phylogenetically close species on the other hand.

université PARIS-SACLAY

ÉCOLE DOCTORALE
Sciences du végétal : du gène à l'écosystème

Titre : Génomique évolutive chez les champignons *Microbotryum*: adaptation à l'hôte et chromosomes de types sexuels

Mots clés : Génomique des populations, chromosomes de types sexuels, adaptation à l'hôte

Résumé : Comprendre comment les espèces s'adaptent à leur environnement est un des buts majeurs de la biologie évolutive. Il s'agit d'identifier les gènes responsables des caractères adaptatifs, mais aussi de comprendre les mécanismes généraux de l'adaptation, et des phénomènes empêchant une adaptation optimale. Les régions non-recombinantes sont particulières pour ces aspects. En effet, elles peuvent protéger de la recombinaison des combinaisons d'allèles favorables, et inversement, la suppression de recombinaison peut entraîner une dégénérescence, comme une accumulation de mutations délétères ou des pertes de gènes. Même les organismes à reproduction sexuée possèdent cependant souvent de larges régions non-recombinantes, associées à la détermination du sexe génétique ou du type sexuel. Dans cette thèse, j'ai ainsi étudié les traces d'adaptation et de dégénérescence dans des génomes de champignons pathogènes de plantes. Les champignons du complexe d'espèces *Microbotryum violaceum*, qui causent la maladie du charbon des anthères chez les Caryophyllacées et comptent des dizaines d'espèces spécifiques d'hôtes différents, sont d'excellents modèles pour l'étude des processus génomiques de l'adaptation. Ils possèdent de plus des chromosomes de types sexuels non-recombinants sur une partie de leur longueur. Pour étudier l'évolution des chromosomes de types sexuels, la dégénérescence et l'adaptation à l'hôte dans le complexe *M. violaceum,* nous avons adopté diverses approches de génomique. En utilisant la technologie PacBio, nous avons obtenu un assemblage complet des chromosomes de types sexuels pour l'espèce *M. lychnidis-dioicae.* Nous avons montré que la région non-recombinante s'étend sur 90 % des chromosomes de types sexuels, présente un niveau de réarrangements exceptionnel entre les deux types sexuels, et que des centaines de gènes sont présents à l'état hémizygote et ont donc probablement été perdus dans un type sexuel. De plus, la comparaison des génomes d'une douzaine d'espèces de *M. violaceum* a montré une accumulation de mutations non-synonymes et d'éléments transposables dans les chromosomes de types sexuels. Nous avons aussi étudié la dégénérescence dans le contexte de l'exposition aux radiations ionisantes, en analysant des populations de *M. lychnidis-dioicae* exposées à différents niveaux de radiation dans la région de Tchernobyl. Nous n'avons pas détecté d'augmentation du taux de mutations non-synonymes par rapport au groupe témoin, ce qui suggère que le champignon est radio-résistant ou que la sélection est plus forte dans la région de Tchernobyl. Enfin, pour étudier l'adaptation à l'hôte, nous avons reséquencé des dizaines des génomes de deux espèces sœurs de *M. violaceum.* L'analyse du polymorphisme a révélé des balayages sélectifs tout le long des génomes et à des localisations différentes entre les deux espèces. Nous avons identifié un certain nombre de gènes candidats pour l'adaptation à l'hôte dans ces régions de balayages sélectifs, sur la base de leur expression *in planta* et de leurs annotations. Les gènes sur-exprimés dans la plante montraient d'autre part un taux de substitutions non-synonymes entre les deux espèces sœurs plus élevé que les autres gènes, ce qui suggère qu'une bonne partie pourrait être impliquée dans l'adaptation à l'hôte.

Ces travaux ouvrent la voie à une comparaison des génomes de différentes espèces du complexe *M. violaceum,* d'une part pour reconstituer l'histoire de la suppression de recombinaison dans les chromosomes de types sexuels, et d'autre part pour étudier les bases génétiques de l'adaptation à différents hôtes dans un complexe d'espèces phylogénétiquement proches.