



HAL
open science

Contribution à l'étude de l'évolution des génomes de champignons ectomycorhiziens du genre *Tuber* (Pézizomycètes) par génomique comparative

Thibaut Payen

► **To cite this version:**

Thibaut Payen. Contribution à l'étude de l'évolution des génomes de champignons ectomycorhiziens du genre *Tuber* (Pézizomycètes) par génomique comparative. Sylviculture, foresterie. Université de Lorraine, 2015. Français. NNT : 2015LORR0046 . tel-01751590

HAL Id: tel-01751590

<https://hal.univ-lorraine.fr/tel-01751590>

Submitted on 29 Mar 2018

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

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



AVERTISSEMENT

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

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

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

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

LIENS

Code de la Propriété Intellectuelle. articles L 122. 4

Code de la Propriété Intellectuelle. articles L 335.2- L 335.10

http://www.cfcopies.com/V2/leg/leg_droi.php

<http://www.culture.gouv.fr/culture/infos-pratiques/droits/protection.htm>

Thèse

Présentée pour l'obtention du titre de
Docteur de l'Université de Lorraine
En biologie végétale et forestière

Par Thibaut PAYEN

Contribution à l'étude de l'évolution des génomes de champignons ectomycorhiziens du genre *Tuber* (Pézizomycètes) par génomique comparative

Co-encadrée par Francis MARTIN et Claude MURAT

Présentée et soutenue publiquement le 4 mai 2015

Devant un jury composé de :

Rapporteurs :

Pr. Fabienne Malagnac, Professeure, Université Paris-Sud, Orsay

Pr. Christophe Roux, Professeur, Université Paul Sabatier, Toulouse

Examineurs :

Pr. Eric Gelhaye, Professeur, Université de Lorraine, Vandoeuvre lès Nancy

Dr. Francis Martin, Directeur de recherche, INRA, Champenoux (Directeur de thèse)

Remerciements

Je tiens tout d'abord à remercier Fabienne Malagnac, Christophe Roux et Eric Gelhaye pour avoir accepté de juger mon travail.

Je tiens également à exprimer toute ma gratitude à mes deux encadrants, Francis et Claude pour m'avoir accueilli au laboratoire mais également pour leur patience durant mon apprentissage de la biologie des champignons. Je leur suis également reconnaissant pour le sujet passionnant et leur soutien dans les moments difficiles.

Il va sans dire que je remercie chaleureusement Manue et Stéphane De Mita qui ont été d'une grande aide pour la bioinformatique chez les champignons et la génomique respectivement.

Un grand merci également à tous les gens du laboratoire qui ont toujours été bienveillants vis-à-vis de ma curiosité candide et de mes questions continuelles sur leurs manips.

Je tiens particulièrement à remercier, dans l'ordre chronologique, Sanjay pour m'avoir forcé dès mon arrivée au laboratoire à parler Anglais, et pour nos discussions sur la politique indienne et le cricket, Vincent pour son humour grinçant, et Cora et Laure, mes plus que collègues pour tous les moments géniaux et les soirées sushis.

Cette thèse a compté quelques colloques et conférences, l'occasion d'échanges passionnants. J'ai ainsi pu rencontrer des gens très agréables, comme l'équipe du chan IRC #bioinfo-fr, mais également Josefiene qui m'a soutenu avant ma première présentation en anglais, et Mathilde avec qui j'ai eu de longues discussions philosophiques à Séville.

Une thèse n'est possible que si les amis sont là pour vous soutenir, il est donc normal de les trouver ici. Je tiens à saluer Emeline et sa bonne humeur. Il faut également que je remercie mes anciens co-pensionnaires de la Pacat', en particulier Patrick pour sa capacité d'écoute hors norme, Johan pour les discussions des nuits entières sur la politique et l'avenir de la robotique, Nevor, qui même s'il travaille pour le Grand Capital est un mec adorable, Agathe parce que « miaou » et un grand merci à Anaïs pour tout ce qu'elle a fait.

Je tiens également à remercier mes parents, qui malgré les inquiétudes, ont toujours été là pour moi et ma cousine, Fanny, pour son aide, ses conseils, les sushis et les corrections.

Merci à tous !

Table des matières

1	Introduction	1
1.1.	Les champignons : des acteurs majeurs des écosystèmes et des sociétés humaines	1
1.2.	La génomique et les champignons	3
1.2.1.	La Levure, le premier eucaryote à avoir été séquencé	3
1.2.2.	L'évolution des génomes des champignons filamenteux	4
1.3.	Les Pézizomycètes	9
1.3.1.	La génomique chez les Pézizomycètes	9
1.4.	Objectifs de ma thèse	13
1.5.	Bibliographie	17
2	Eléments répétés dans les champignons filamenteux	21
3	Phylogénomique du genre <i>Tuber</i>	43
4	Etude de génomique comparative chez les truffes et autres Pézizomycètes	69
5	Etude du polymorphisme génomique de <i>Tuber melanosporum</i> par reséquençage d'isolats	129
6	Reconstruction de l'histoire évolutive des retrotransposons <i>Gypsy</i> dans <i>Tuber melanosporum</i>	151
7	Discussion générale et perspectives	189
7.1	Evolution des génomes au sein des Tuberales	191
7.2	Convergences évolutives entre les ectomycorhiziens Ascomycètes et Basidiomycètes	197
7.3	Comment les ressources génomiques vont impacter la filière trufficole	199
7.4	Conclusion	202
7.4	Bibliographie	205
8	Annexes	
8.1	Liste des publications	
8.2	Liste des participation à des congrès internationaux	
8.3	Publication annexe 1 : Fungal Biology: Compiling Genomes and Exploiting Them	
8.4	Publication annexe 2 : Patterns of Genomic Variation in the Poplar Rust Fungus <i>Melampsora Larici-Populina</i> Identify Pathogenesis-Related Factors.	

1 Introduction

1.1. Les champignons : des acteurs majeurs des écosystèmes et des sociétés humaines

Les champignons représentent une composante importante dans de nombreux écosystèmes de notre planète et ils jouent un rôle primordial dans les cycles biogéochimiques. De par l'établissement de symbioses avec les plantes (voir ci-dessous) ils contribuent aussi à maintenir la biodiversité végétale en aidant la croissance des plantes. Les champignons ont également fortement aidé au développement des civilisations humaines puisque la levure *Saccharomyces cerevisiae* permet de faire du pain, du vin et de la bière. D'autres espèces fongiques sont aussi couramment utilisées dans l'alimentation humaine comme le champignon de Paris (*Agaricus bisporus*), le shiitake (*Lentinula edodes*), le cèpe (*Boletus edulis*), le matsutake (*Tricholoma matsutake*) ou encore la truffe noire du Périgord (*Tuber melanosporum*). D'autres espèces sont par contre indésirables pour l'Homme comme *Blumeria graminis*, un pathogène du blé, ou *Candida albicans*, un pathogène humain.

Les champignons sont traditionnellement séparés par leurs modes de vie et en particulier leurs stratégies pour acquérir des sucres. L'acquisition du carbone se fait selon trois grands modes : l'exploitation du carbone d'autres organismes vivants (pathogènes), l'acquisition sur des organismes morts et leurs déchets (saprotrophes) et la symbiose avec des organismes vivants (mycorhiziens, lichens). La symbiose mycorhizienne est prédominante chez les plantes terrestres vasculaires puisque près de 95 % d'entre elles sont associées dans une symbiose de ce type (Smith and Read 2010). Les champignons mycorhiziens permettent l'augmentation de la croissance des plantes (Koch, Croll, and Sanders 2006) et sont donc des éléments importants pour améliorer leur productivité dans un contexte où les engrais chimiques sont de plus en plus décriés. Ceci a entraîné la création par des entreprises d'engrais biologiques à

base de champignons mycorhiziens (par ex. MYCOR par IF Tech utilisant *Rhizophagus irregularis*). Les champignons endomycorhiziens à arbuscule (AMF), comme *R. irregularis*, ont été proposés comme alternative aux engrais chimiques qui sont une nuisance environnementale et une ressource limitée (Ceballos et al. 2013). En fait, il existe plusieurs types de symbiose mycorhizienne : la symbiose AMF, la symbiose endomycorhizienne avec les orchidées (ORM), la symbiose endomycorhizienne avec les éricacées (ERM), la symbiose ectomycorhizienne (ECM) et la symbiose ectendomycorhizienne (Garbaye 2013). Ces différents types de symbioses impliquent des partenaires différents : par exemple des champignons appartenant à la division des Glomeromycota et des plantes herbacées ou de grandes cultures pour l'AMF et des ascomycètes et basidiomycètes associés à des arbres et arbustes pour l'ECM. En dehors des différences morphologiques il existe aussi des différences dans l'impact de ces symbioses sur les écosystèmes. Par exemple, une étude récente montre que les ECM favorisent la séquestration du carbone dans le sol comparé aux AMF, ce qui peut avoir un impact direct sur la réduction des gaz à effet de serre et donc un impact sur le changement climatique (Averill, Turner, and Finzi 2014). Dans ce contexte une meilleure compréhension de l'évolution et de l'établissement des différents types de symbiose mycorhizienne semble primordiale pour mieux connaître et utiliser tout le potentiel des champignons mycorhiziens.

D'autre part, malgré leur importance aussi bien pour les écosystèmes que pour les activités humaines, le règne fongique est encore mal connu. En effet, un indice basé sur le nombre de champignons associés aux différentes espèces de plantes permet d'estimer qu'il existerait environ 1,5 millions d'espèces de champignons (Hawksworth 1991) alors que seulement 70,000 ascomycètes et 30,000 basidiomycètes (les deux embranchements contenant les champignons avec des fructifications visibles) ont été décrites. Ce constat mène à un effort accru pour identifier les espèces et ainsi aider à la compréhension du règne fongique.

La révolution des techniques de séquençages (« High-Throughput sequencing » ou HTS) permet désormais d'aborder des questions évolutives, populationnelles et de biodiversité très fondamentales et jusque-là inenvisageables (Faure, Dominique, and Salamitou 2013). Il est ainsi possible via la métataxonomie ou l'ADN barre code d'appréhender la biodiversité fongique dans les sols forestiers (Buée et al. 2009), dans les sols serpentiniques (Daghino et al. 2012) ou bien la phyllosphère (Jumpponen and Jones 2009). Les HTS ont aussi permis de révolutionner la génomique en permettant d'appréhender la nature des changements fonctionnels et des réseaux de gènes impliqués, la distribution des mutations et de leurs effets, la structure de la recombinaison, ou encore la nature composite des génomes soumis à des régimes de sélection complexes. Toutefois la génomique et le séquençage des génomes fongiques ont commencé bien avant l'apparition des HTS au milieu des années 2000. Dans le paragraphe suivant un historique sur le séquençage des génomes fongiques est présenté.

1.2. La génomique et les champignons

1.2.1. La Levure, le premier eucaryote à avoir été séquencé

En 1976, le premier séquençage d'un génome est publié, il s'agissait du virus bactériophage MS21 (Fiers et al. 1976). Il a ensuite fallu attendre vingt ans pour que le premier génome d'une bactérie, *Haemophilus influenzae*, soit séquencé (Fleischmann et al. 1995). Peu de temps après le séquençage du premier génome eucaryote a eu lieu celui d'un champignon domestiqué de longue date par l'homme, la levure de boulanger, *S. cerevisiae* (Goffeau et al. 1996). Jusqu'au milieu des années 2000 la méthode de séquençage développé par Sanger a été la plus utilisée. Mais depuis maintenant dix ans, une révolution s'est opérée avec de nouvelles méthodes de séquençage qui ont démocratisé le séquençage et ont également permis le reséquencage d'isolats d'une même espèce. Cette nouvelle possibilité a permis la mise en

place d'études d'évolution expérimentales. Une étude de ce type par Lang et collègues (Lang et al. 2013) a, par exemple, démontré la prévalence chez *S. cerevisiae* de « l'auto-stop » des mutations neutres à faiblement délétère dans des populations ayant des mutations adaptatives, permettant ainsi à des mutations négatives pour les individus de s'implanter. D'autres projets, comme Génolevures (Sherman et al. 2009) étudient les relations entre différentes levures au niveau de leurs génomes en incluant notamment des analyses de synténie au sein des Hemiascomycètes, avec depuis 1999 un ajout régulier de nouveaux génomes de levures.

Les technologies de séquençage actuelles ont également permis le séquençage de nombreux génomes de champignons filamenteux (Figure 1) pour en couvrir la diversité. Nous sommes passés de l'ère des espèces modèles à l'ère du séquençage à grande échelle, permettant ainsi la mise en place d'études comparatives.

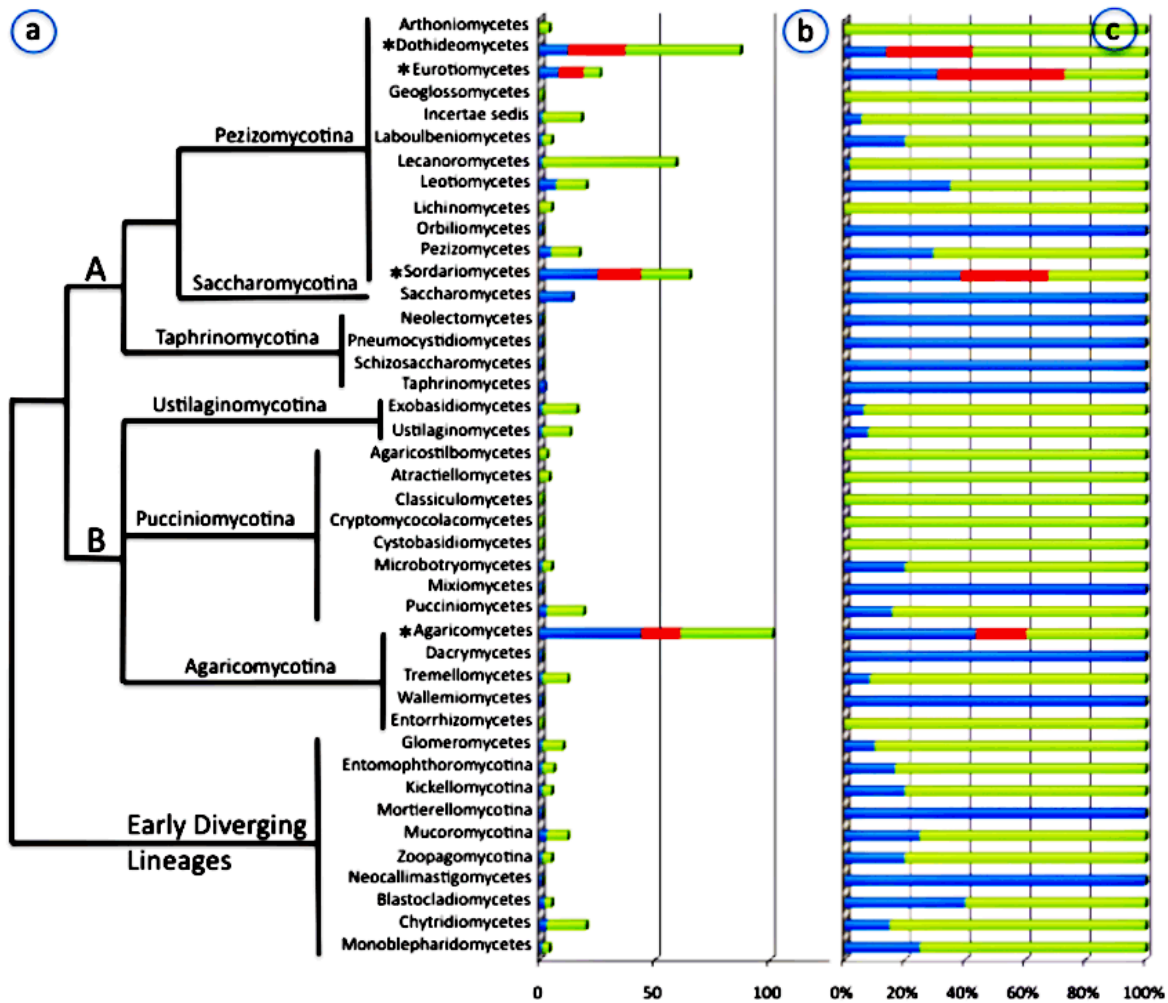


Figure 1 Echantillonnage au niveau de la famille des différents génomes de l'arbre phylogénétique des champignons tel que connu actuellement (a). La partie (b) représente le nombre de génomes par classe ou sous-phylum, alors que la partie (c) représente le pourcentage des différentes familles par classe et sous phylum, ayant été séquencés (bleu), en projet pour la première phase du projet 1000 génomes (rouge) et non encore échantillonnées (vert). Sur l'arbre, A représente les Ascomycètes et B les Basidiomycètes. (Source : <http://jgi.doe.gov/our-science/science-programs/fungal-genomics/1000-fungal-genomes/>)

1.2.2. L'évolution des génomes des champignons filamenteux

Les génomes eucaryotes sont une mosaïque d'éléments : gènes, régions régulatrices, séquences répétées et régions intergéniques. Grâce aux études de génomique comparative, l'évolution de ces différents compartiments génomiques peut être abordée. Ces analyses se placent dans la droite ligne des études de morphologie comparée puis de génétique comparée.

L'apport des HTS permet ainsi d'étudier les relations évolutives entre les espèces. L'exemple des modes de vie des champignons permet de voir que les caractéristiques morphologiques ne

se retrouvent pas forcément au niveau génomique, comme par exemple la séparation entre les pourritures brunes et les pourritures blanches (Riley et al. 2014). Cependant bien que la séparation par mode de vie ne suive pas l'évolution elle permet d'étudier les caractéristiques communes aux champignons ayant des styles de vie comparables. De telles études ont été réalisées chez les champignons saprotrophes, dégradant le bois afin d'y récupérer du carbone (Floudas et al. 2012). L'analyse réalisée par Floudas et collègues est basée sur les génomes de 31 champignons, et étudie les CAZymes (*Carbohydrate-Active Enzymes*) et les oxydoréductases qui sont deux familles de gènes connues pour être impliquées dans la dégradation du bois. Cette étude se base sur 12 génomes nouvellement séquencés et montre que parmi les 31 génomes utilisés il existe 4 lignages indépendants de pourritures brunes. Cette étude complétée par la publication d'une seconde étude focalisée sur un plus grand nombre de basidiomycètes (Riley et al. 2014) montre l'existence d'un gradient dans les répertoires de gènes des groupes traditionnels de pourritures brunes et de pourritures blanches. La génomique comparative permet donc d'apprécier l'évolution des espèces et ceci en lien avec leur écologie. En effet, dans le cadre du « Mycorrhizal Genome Initiative » coordonné par Francis Martin (Martin et al. 2011), il a été montré qu'un certain nombre de CAZymes permettant de dégrader la matière organique d'origine végétale (ex. Glycosyl hydrolase GH6 et GH7) sont absentes dans les génomes des champignons ECM (Martin et al. 2008; Martin et al. 2010; Kohler et al. 2015). Par contre, d'autres champignons mycorrhiziens comme *Tulasnella caulospora* et *Oidiodendron maius*, formant des symbioses ORM et ERM, sont très riches en CAZymes (Kohler et al. 2015). Il semble donc exister un lien entre contenu génique et stratégie écologique en regardant les régions codantes.

La taille des génomes est aussi très différente allant de quelques millions de paires de bases (Mpb) pour les levures à plus de 100 Mpb pour certains champignons filamenteux comme le

pathogène *B. graminis* ou la truffe noire, *T. melanosporum*. Cette taille est principalement corrélée avec le contenu en séquences répétées. Plus de détails sont donnés dans le chapitre bibliographique sur ces éléments (Chapitre 2). Le principal résultat de cette analyse bibliographique que l'on a effectuée montre que les champignons interagissant avec les plantes vivantes ont de plus gros génomes. Ce résultat a été récemment confirmé par une analyse réalisée chez les Amanites (Hess et al. 2014).

Les séquences répétées ne sont pas uniquement responsables de la taille des génomes mais ils ont également un rôle dans l'évolution de ceux-ci. Certaines études montrent les effets des séquences répétées sur l'évolution des génomes (Yoshida et al. 2009; Raffaele and Kamoun 2012; Grandaubert et al. 2014). L'étude de Yoshida et collègues (2009), basée sur le génome de *Magnaporthe oryzae*, analyse l'emplacement des effecteurs de ce champignon pathogène du riz. Les champignons pathogènes, comme tous les pathogènes, sont impliqués dans une course aux armements avec leurs plantes hôtes (Clay and Kover 1996) et sont donc soumis à une pression de sélection forte, en particulier sur leurs protéines sécrétées. Ainsi Yoshida et collègues (2009) montrent une préférence pour les gènes de protéines sécrétés pour les régions riches en éléments transposables, ces régions étant plus soumises à des réarrangements chromosomiques. L'article de revue de Raffaele et Kamoun met aussi en évidence de nombreux exemples de gènes codants pour des effecteurs se trouvant entre deux éléments transposables et pouvant ainsi facilement être éliminés du génome du pathogène en cas de reconnaissance par les défenses de la plante. Cet article présente également l'hypothèse du « jump or die » qui stipule que les pathogènes ayant une plus grande plasticité et donc une plus faible spécificité d'hôte ont une plus grande chance de survie car ils peuvent changer d'hôte en cas d'extinction d'une espèce. L'étude, plus récente, de Grandaubert et collègues (2014) étudie des espèces plus proches les unes des autres puisque cette étude se focalise sur un complexe d'espèces de champignons pathogènes des plantes. Dans cette étude 5 espèces

ayant des hôtes et des degrés de pathogénicité différents ont été séquencées et comparées. Au sein du complexe d'espèce *Leptosphaeria maculans*-*Leptosphaeria biglobosa*, cette étude montre une grande variabilité de la taille des génomes (de 30 à 45 Mb) et en contenu d'éléments transposables (de 3 à 33% du génome). L'espèce ayant été envahie par les éléments transposables, *Leptosphaeria maculans* 'brassicae' montre des signes de réarrangements chromosomiques qui sont probablement liés à la spéciation.

D'autres études existent sur des champignons comme celle portant sur différentes espèces de *Penicillium* associées à différents fromages (Cheeseman et al. 2014). Cette étude met en évidence un processus bien connu chez les procaryotes mais peu étudié chez les eucaryotes ; le transfert horizontal de matériel génétique. Ce processus qui est jugé comme rare, a eu lieu à plusieurs reprises chez *Penicillium* et s'est accompagné du transfert d'un grand nombre de gènes avec 250 gènes prédits dans la région *Wallabi*, détectée comme ayant subi le transfert horizontal. L'étude conclue que le transfert horizontal peut avoir des conséquences importantes pour l'adaptation des espèces fongiques à des nouveaux environnements.

Comme montré ci-dessus, la génomique comparative permet d'appréhender l'évolution des différents groupes fongiques aussi bien au niveau interspécifique qu'intraspécifique. Toutefois, certaines classes fongiques sont encore mal connues et peu étudiées au niveau génomique. C'est le cas des Pézizomycètes puisque seulement deux génomes ont été publiés à ce jour : la truffe noire (Martin et al. 2010) et *Pyronema confluens* (Traeger et al. 2013) et ce malgré le fait que cette classe fongique contient *Ascobolus immersus*, une espèce modèle en génétique depuis les années 1930 et qu'il est possible de retrouver des espèces ECM, ORM et saprotrophes chez les Pézizomycètes. Dans le paragraphe suivant la classe des Pézizomycètes sera présentée ainsi qu'un état des lieux des différentes connaissances sur leurs génomes.

1.3. Les Pézizomycètes

1.3.1. La génomique chez les Pézizomycètes

La classe des Pézizomycètes est une classe de champignons du sous-embranchement des Pezizomycotina contenant 200 genres et 1,683 espèces décrites (Kirk et al. 2008). Il s'agit de l'une des plus anciennes lignées ayant divergé dans ce sous-embranchement (Spatafora et al. 2006) et cette classe contient des champignons à la biologie très diverse avec des champignons saprotrophes et d'autres mycorrhiziens mais également avec des carpophores épigés ou hypogés. Ces deux caractéristiques ont la propriété intéressante de ne pas suivre la phylogénie et sont donc probablement des propriétés apparues de façons parallèles au sein même des Pézizomycètes (O'Donnell et al. 1997; Hansen, Laessoe, and Pfister 2001; Læssøe and Hansen 2007). Une étude de génomique comparative chez les Pézizomycètes permettrait donc d'aborder différentes questions sur le style de vie et la fructification de ces champignons. Comme déjà évoqué précédemment, à ce jour uniquement deux espèces de Pézizomycètes ont été séquencées : *T. melanosporum* (Martin et al. 2010) et *Pyronema confluens* (Traeger et al. 2013). Dans le cadre de différents projets (TuberEvol, projet conjoint entre l'INRA et le Génoscope ; le projet 1000 génomes fongiques avec le département de l'énergie aux états unis (JGI DOE) et le projet CSP2012 métatranscriptomique des sols forestiers avec le JGI-DOE) le séquençage de plusieurs espèces de Pézizomycètes est prévu. A ce jour huit génomes de Pézizomycètes sont disponibles. Parmi les Pézizomycètes hypogés, les truffes, dont *T. melanosporum* (Martin et al. 2010) a été le premier Pézizomycète séquencé, ont été étudiées de façon plus spécifique dans le cadre de cette thèse.

Les truffes sont des champignons ectomycorhiziens appartenant au genre *Tuber*. Le nombre d'espèces de ce genre a été estimé à 180 espèces dans le monde (Bonito et al. 2010) dont au moins 28 en Europe (Ceruti, Fontana, and Nosenzo 2003). Parmi celles-ci, certaines sont hautement appréciées pour leurs qualités gustatives comme la truffe noire du Périgord (*T.*

melanosporum) ou la truffe blanche du Piémont (*T. magnatum*) qui se vendent respectivement entre 500 et 1000€ le kilo et entre 2000 et 5000€ le kilo selon la production annuelle. Les caractéristiques des truffes leur permettant de pousser sur des sols trop pauvres pour l'agriculture traditionnelle ainsi que les prix élevés et la chute de la production naturelle ont mené à la création de truffières artificielles. Les premières truffières artificielles datent du XVIII^{ème} siècle (Hall, Yun, and Amicucci 2003), par récupération de plants auto-semés proches d'arbres producteurs par Joseph Talon. De nouvelles méthodes sont apparues en 1970 à l'INRA de Clermont-Ferrand avec la mise au point d'une technique de mycorhization de plants par des truffes noires du Périgord (Chevalier and Grente 1979) qui ont permis la mise en place de plantations productrices dès 5 ans. De nos jours en France, la production varie de 40 à 60 tonnes et est constituée à plus de 80% par des plantations artificielles. Cette domestication en parallèle avec l'utilisation de *T. melanosporum* comme modèle pour la mycorhization a mené à de nombreuses études sur la biologie de cette truffe afin de répondre à des problématiques aussi bien de rendement et d'adaptation aux conditions biotique et abiotique qu'à des questions plus fondamentales sur les gènes impliqués dans le goût, la plasticité du génome ou l'évolution de la truffe noire du Périgord et plus généralement du genre *Tuber*.

Le cycle de vie de *T. melanosporum* est illustré dans la Figure 2. Le développement sexuel des *Tuber* est contrôlé par un locus de compatibilité sexuelle avec 2 allèles MAT 1 et MAT 2 (Rubini et al. 2011). Les deux types sexuels sont capables de façon égale de former des ectomycorhizes (ECM) en interaction avec les arbres, cependant un arbre ne semble porter des mycorhizes que de l'un des types sexuels (Claude Murat et al. 2013). Les fructifications (gleba) sont formées d'un tissu haploïde du même génotype que les mycorhizes à sa proximité et portent des spores haploïdes dont chacune peut être MAT1 ou MAT2. Il a été démontré que d'autres *Tuber* comme *T. magnatum* (Paolocci et al. 2006) et *T. indicum* (Belfiori et al. 2013)

ont un système de reproduction avec 2 idiomorphes fonctionnant de la même façon que *T. melanosporum*.

T. melanosporum dépend des animaux pour la dispersion de ses spores. La forte odeur des truffes, venant de composés volatiles soufrés (Splivallo et al. 2011), est une méthode pour attirer les insectes et les gros animaux sauvages qui est utilisée par l'Homme pour chercher les fructifications à l'aide de cochons ou de chiens. Des études récentes (Pacioni et al. 2015) montrent également que les fructifications matures de *T. melanosporum* produisent des anandamides, qui sont des membres de la famille des endocannabinoïdes à des doses pouvant, associées à l'odeur forte des truffes, avoir un effet sur les récepteurs animaux, faisant de la truffe une espèce appréciée au sein des animaux. Au-delà de cet aspect d'attraction des animaux les endocannabinoïdes sont produits également chez les animaux en association avec les récepteurs. Les récepteurs semblent absents chez *T. melanosporum* pouvant indiquer une apparition avant la séparation entre les deux règnes et une apparition subséquente des récepteurs chez les animaux.

Des analyses de diversité génétique utilisant des marqueurs microsatellites (SSR) ont également été conduites chez *T. magnatum* (Rubini et al. 2005), *T. aestivum* (Molinier et al. 2013) et *T. melanosporum* (Riccioni et al. 2008; Claude Murat et al. 2013). Ces études ont mis en évidence un niveau de diversité génétique non négligeable surtout pour *T. aestivum* et *T. melanosporum* permettant d'analyser la structure génétique au niveau local (truffière) ou Européen. Après 18 ans de controverses sur le niveau de diversité génétique et la structuration des populations, les nouveaux marqueurs moléculaires mis au point grâce aux ressources génétiques disponibles pour la truffe noire ont permis de conclure que chaque truffe est génétiquement différent d'une autre (C Murat et al. 2011). De même il a été possible d'identifier jusqu'à 13 géotypes dans 50 m² sur une plantation de *T. melanosporum*.

Toutefois, les marqueurs microsatellites utilisés sont des marqueurs neutres et ils ne permettent donc pas d'études d'adaptation. C'est pourquoi dans le cadre de ma thèse nous avons généré une ressource de « single nucléotide polymorphism » (SNPs) permettant de densifier le génome en marqueurs mais aussi d'adresser la sélection et l'adaptation des truffes. De telles approches de génomique populationnelle ont été appliqués sur d'autres organismes auparavant comme par exemple pour identifier des régions génomiques impliquées dans l'adaptation de *Neurospora crassa* au froid (Ellison et al. 2011).

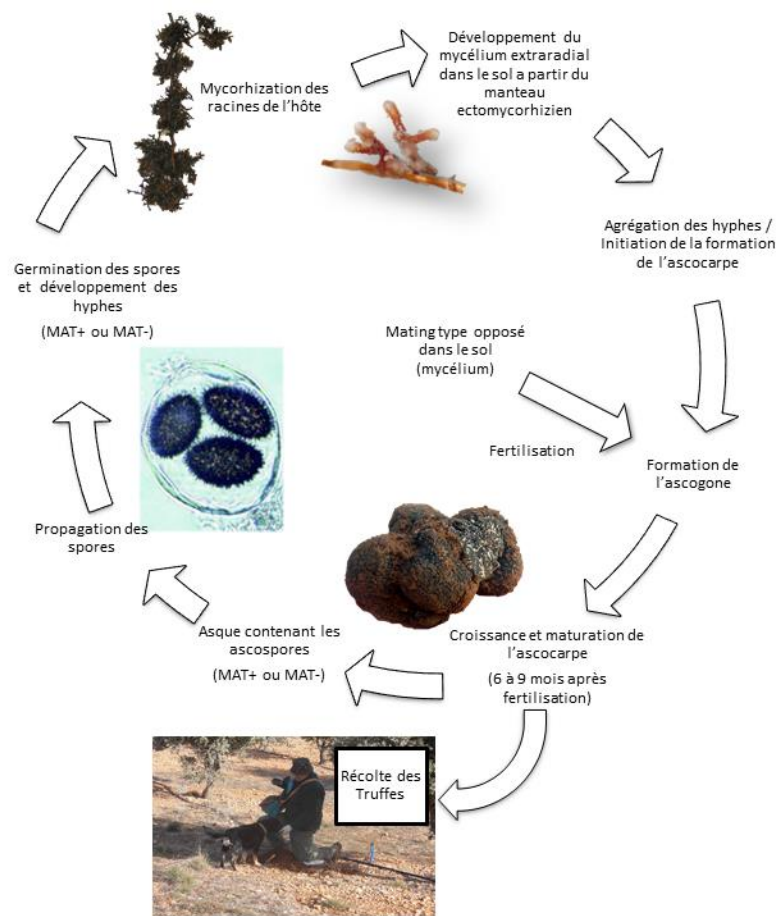


Figure 2 Cycle de vie de *Tuber melanosporum* d'après De la Varga Pastor 2013

1.4. Objectifs de ma thèse

Les travaux que j'ai réalisés durant ma thèse au sein de l'équipe « Ecogénomique des Interactions » de l'UMR Interactions Arbres/Micro-organismes visent à apporter une

meilleure compréhension de la structure et de l'évolution des génomes de Pézizomycètes et plus particulièrement de truffes en utilisant pour cela des ressources génétiques disponibles. Les travaux ont été réalisés au niveau interspécifique mais aussi intraspécifique par le reséquençage d'échantillons de la même espèce. Ce manuscrit de thèse est divisé en cinq chapitres. Tout d'abord deux chapitres introductifs, un sur les séquences répétées (**chapitre 2**) et l'autre sur les apports de la génomique et de la phylogénie pour la compréhension du cycle biologique et de l'évolution au sein des truffes (**chapitre 3**). Ces deux chapitres correspondent à des analyses bibliographiques publiées. Suivront ensuite trois chapitres de résultats portant sur :

- l'identification des caractéristiques spécifiques aux différentes espèces de Pézizomycètes d'une part et aux Pézizomycètes comparés aux autres espèces fongiques connues. Cette étude a été réalisée grâce à des outils de génomique comparative basés sur des outils préexistants et des scripts développés dans le cadre de ce travail (**Chapitre 4**),
- l'identification au sein de l'espèce *Tuber melanosporum* des polymorphismes de type SNPs grâce au reséquençage d'individus dans le but de les comparer à un génome de référence (**Chapitre 5**),
- la caractérisation de l'histoire, de la dynamique et de la diversité d'un type particulier d'éléments transposables, les rétrotransposons gypsy chez *T. melanosporum* (**Chapitre 6**).

Une discussion permettant de remettre en perspective ces différents travaux ainsi qu'une ouverture sur les perspectives ouvertes ont ensuite été réalisées par ces travaux et se trouvent en fin de manuscrit (**Chapitre 7**).

1.5. Bibliographie

- Averill, Colin, Benjamin L. Turner, and Adrien C. Finzi. 2014. "Mycorrhiza-Mediated Competition between Plants and Decomposers Drives Soil Carbon Storage." *Nature* 505 (7484): 543–45. doi:10.1038/nature12901.
- Belfiori, Beatrice, Claudia Riccioni, Francesco Paolocci, and Andrea Rubini. 2013. "Mating Type Locus of Chinese Black Truffles Reveals Heterothallism and the Presence of Cryptic Species within the *T. Indicum* Species Complex." *PLoS ONE* 8 (12): e82353. doi:10.1371/journal.pone.0082353.
- Bonito, Gregory M, Andrii P Gryganskyi, James M Trappe, and Rytas Vilgalys. 2010. "A Global Meta-Analysis of Tuber ITS rDNA Sequences: Species Diversity, Host Associations and Long-Distance Dispersal." *Molecular Ecology* 19 (22): 4994–5008. doi:10.1111/j.1365-294X.2010.04855.x.
- Buée, M., M. Reich, C. Murat, E. Morin, R. H. Nilsson, S. Uroz, and F. Martin. 2009. "454 Pyrosequencing Analyses of Forest Soils Reveal an Unexpectedly High Fungal Diversity." *New Phytologist* 184 (2): 449–56. doi:10.1111/j.1469-8137.2009.03003.x.
- Ceballos, Isabel, Michael Ruiz, Cristhian Fernández, Ricardo Peña, Alia Rodriguez, and Ian R. Sanders. 2013. "The In Vitro Mass-Produced Model Mycorrhizal Fungus, *Rhizophagus Irregularis*, Significantly Increases Yields of the Globally Important Food Security Crop Cassava." *PLoS ONE* 8 (8): e70633. doi:10.1371/journal.pone.0070633.
- Ceruti, A., A. Fontana, and C. Nosenzo. 2003. "European species of the genus *Tuber*. An historical revision." *Monografie XXXVII*, 467 pp.
- Cheeseman, Kevin, Jeanne Ropars, Pierre Renault, Joëlle Dupont, Jérôme Gouzy, Antoine Branca, Anne-Laure Abraham, et al. 2014. "Multiple Recent Horizontal Transfers of a Large Genomic Region in Cheese Making Fungi." *Nature Communications* 5 (January). doi:10.1038/ncomms3876.
- Chevalier, G., and J. Grente. 1979. "Application Pratique de La Symbiose Ectomycorhizienne: Production À Grande Échelle de Plants Mycorhizés Par La Truffe (*Tuber Melanosporum* Vitt.)." *Mushroom Science* 1979.
- Clay, Keith, and Paula X. Kover. 1996. "The Red Queen Hypothesis and Plant/Pathogen Interactions." *Annual Review of Phytopathology* 34 (1): 29–50. doi:10.1146/annurev.phyto.34.1.29.
- Daghino, Stefania, Claude Murat, Elisa Sizzano, Mariangela Girlanda, and Silvia Perotto. 2012. "Fungal Diversity Is Not Determined by Mineral and Chemical Differences in Serpentine Substrates." *PLoS ONE* 7 (9): e44233. doi:10.1371/journal.pone.0044233.
- De la Varga Pastor, Herminia. 2013. "Traceability of the Mycorrhizal Symbiosis in the Controlled Production of Edible Mushrooms."

1.5. Bibliographie

- Averill, Colin, Benjamin L. Turner, and Adrien C. Finzi. 2014. "Mycorrhiza-Mediated Competition between Plants and Decomposers Drives Soil Carbon Storage." *Nature* 505 (7484): 543–45. doi:10.1038/nature12901.
- Belfiori, Beatrice, Claudia Riccioni, Francesco Paolocci, and Andrea Rubini. 2013. "Mating Type Locus of Chinese Black Truffles Reveals Heterothallism and the Presence of Cryptic Species within the *T. Indicum* Species Complex." *PLoS ONE* 8 (12): e82353. doi:10.1371/journal.pone.0082353.
- Bonito, Gregory M, Andrii P Gryganskyi, James M Trappe, and Rytas Vilgalys. 2010. "A Global Meta-Analysis of Tuber ITS rDNA Sequences: Species Diversity, Host Associations and Long-Distance Dispersal." *Molecular Ecology* 19 (22): 4994–5008. doi:10.1111/j.1365-294X.2010.04855.x.
- Buée, M., M. Reich, C. Murat, E. Morin, R. H. Nilsson, S. Uroz, and F. Martin. 2009. "454 Pyrosequencing Analyses of Forest Soils Reveal an Unexpectedly High Fungal Diversity." *New Phytologist* 184 (2): 449–56. doi:10.1111/j.1469-8137.2009.03003.x.
- Ceballos, Isabel, Michael Ruiz, Cristhian Fernández, Ricardo Peña, Alia Rodriguez, and Ian R. Sanders. 2013. "The In Vitro Mass-Produced Model Mycorrhizal Fungus, *Rhizophagus Irregularis*, Significantly Increases Yields of the Globally Important Food Security Crop Cassava." *PLoS ONE* 8 (8): e70633. doi:10.1371/journal.pone.0070633.
- Ceruti, A., A. Fontana, and C. Nosenzo. 2003. "European species of the genus *Tuber*. An historical revision." *Monografie XXXVII*, 467 pp.
- Cheeseman, Kevin, Jeanne Ropars, Pierre Renault, Joëlle Dupont, Jérôme Gouzy, Antoine Branca, Anne-Laure Abraham, et al. 2014. "Multiple Recent Horizontal Transfers of a Large Genomic Region in Cheese Making Fungi." *Nature Communications* 5 (January). doi:10.1038/ncomms3876.
- Chevalier, G., and J. Grente. 1979. "Application Pratique de La Symbiose Ectomycorhizienne: Production À Grande Échelle de Plants Mycorhizés Par La Truffe (*Tuber Melanosporum* Vitt.)." *Mushroom Science* 1979.
- Clay, Keith, and Paula X. Kover. 1996. "The Red Queen Hypothesis and Plant/Pathogen Interactions." *Annual Review of Phytopathology* 34 (1): 29–50. doi:10.1146/annurev.phyto.34.1.29.
- Daghino, Stefania, Claude Murat, Elisa Sizzano, Mariangela Girlanda, and Silvia Perotto. 2012. "Fungal Diversity Is Not Determined by Mineral and Chemical Differences in Serpentine Substrates." *PLoS ONE* 7 (9): e44233. doi:10.1371/journal.pone.0044233.
- De la Varga Pastor, Herminia. 2013. "Traceability of the Mycorrhizal Symbiosis in the Controlled Production of Edible Mushrooms."

- Ellison, Christopher E, Charles Hall, David Kowbel, Juliet Welch, Rachel B Brem, N L Glass, and John W Taylor. 2011. "Population Genomics and Local Adaptation in Wild Isolates of a Model Microbial Eukaryote." *Proceedings of the National Academy of Sciences of the United States of America* 108 (7): 2831–36. doi:10.1073/pnas.1014971108.
- Faure, Denis, Joly Dominique, and Sylvie Salamitou. 2013. *Génomique Environnementale*. Les Cahiers Prospectives. CNRS.
- Fiers, W., R. Contreras, F. Duerinck, G. Haegeman, D. Iserentant, J. Merregaert, W. Min Jou, et al. 1976. "Complete Nucleotide Sequence of Bacteriophage MS2 RNA: Primary and Secondary Structure of the Replicase Gene." *Nature* 260 (5551): 500–507. doi:10.1038/260500a0.
- Fleischmann, R. D., M. D. Adams, O. White, R. A. Clayton, E. F. Kirkness, A. R. Kerlavage, C. J. Bult, et al. 1995. "Whole-Genome Random Sequencing and Assembly of Haemophilus Influenzae Rd." *Science* 269 (5223): 496–512. doi:10.1126/science.7542800.
- Floudas, Dimitrios, Manfred Binder, Robert Riley, Kerrie Barry, Robert a. Blanchette, Bernard Henrissat, a. T. Martinez, et al. 2012. "The Paleozoic Origin of Enzymatic Lignin Decomposition Reconstructed from 31 Fungal Genomes." *Science* 336 (6089): 1715–19. doi:10.1126/science.1221748.
- Garbaye, Jean. 2013. *La symbiose mycorhizienne: Une association entre les plantes et les champignons*. Editions Quae.
- Goffeau, A., B. G. Barrell, H. Bussey, R. W. Davis, B. Dujon, H. Feldmann, F. Galibert, et al. 1996. "Life with 6000 Genes." *Science* 274 (5287): 546–67. doi:10.1126/science.274.5287.546.
- Grandaubert, Jonathan, Rohan GT Lowe, Jessica L. Soyer, Conrad L. Schoch, Angela P. Van de Wouw, Isabelle Fudal, Barbara Robbertse, et al. 2014. "Transposable Element-Assisted Evolution and Adaptation to Host Plant within the Leptosphaeria Maculans-Leptosphaeria Biglobosa Species Complex of Fungal Pathogens." *BMC Genomics* 15 (1): 891. doi:10.1186/1471-2164-15-891.
- Hall, Ian R, Wang Yun, and Antonella Amicucci. 2003. "Cultivation of Edible Ectomycorrhizal Mushrooms." *Trends in Biotechnology* 21 (10): 433–38. doi:10.1016/S0167-7799(03)00204-X.
- Hansen, Karen, Thomas Laessoe, and Donald H. Pfister. 2001. "Phylogenetics of the Pezizaceae, with an Emphasis on Peziza." *Mycologia* 93 (5): 958. doi:10.2307/3761760.
- Hawksworth, D. L. 1991. "The Fungal Dimension of Biodiversity: Magnitude, Significance, and Conservation." *Mycological Research* 95 (6): 641–55. doi:10.1016/S0953-7562(09)80810-1.

- Hess, Jaqueline, Inger Skrede, Benjamin E. Wolfe, Kurt LaButti, Robin A. Ohm, Igor V. Grigoriev, and Anne Pringle. 2014. "Transposable Element Dynamics among Asymbiotic and Ectomycorrhizal *Amanita* Fungi." *Genome Biology and Evolution* 6 (7): 1564–78. doi:10.1093/gbe/evu121.
- Jumpponen, A., and K. L. Jones. 2009. "Massively Parallel 454 Sequencing Indicates Hyperdiverse Fungal Communities in Temperate *Quercus Macrocarpa* Phyllosphere." *New Phytologist* 184 (2): 438–48. doi:10.1111/j.1469-8137.2009.02990.x.
- Kirk, P. M., P. F. Cannon, D. W. Minter, and J. A. Stalpers. 2008. *Dictionary of the Fungi*. <http://www.publish.csiro.au/pid/5998.htm>.
- Koch, Alexander M., Daniel Croll, and Ian R. Sanders. 2006. "Genetic Variability in a Population of Arbuscular Mycorrhizal Fungi Causes Variation in Plant Growth." *Ecology Letters* 9 (2): 103–10. doi:10.1111/j.1461-0248.2005.00853.x.
- Kohler, Annegret, Alan Kuo, Laszlo G. Nagy, Emmanuelle Morin, Kerrie W. Barry, Francois Buscot, Björn Canbäck, et al. 2015. "Convergent Losses of Decay Mechanisms and Rapid Turnover of Symbiosis Genes in Mycorrhizal Mutualists." *Nature Genetics* advance online publication. doi:10.1038/ng.3223.
- Læssøe, Thomas, and Karen Hansen. 2007. "Truffle Trouble: What Happened to the Tuberales?" *Mycological Research*, New Bottles for Old Wine, 111 (9): 1075–99. doi:10.1016/j.mycres.2007.08.004.
- Lang, Gregory I., Daniel P. Rice, Mark J. Hickman, Erica Sodergren, George M. Weinstock, David Botstein, and Michael M. Desai. 2013. "Pervasive Genetic Hitchhiking and Clonal Interference in Forty Evolving Yeast Populations." *Nature* 500 (7464): 571–74. doi:10.1038/nature12344.
- Martin, Francis, A. Aerts, D. Ahrén, A. Brun, E. G. J. Danchin, F. Duchaussoy, J. Gibon, et al. 2008. "The Genome of *Laccaria Bicolor* Provides Insights into Mycorrhizal Symbiosis." *Nature* 452 (7183): 88–92. doi:10.1038/nature06556.
- Martin, Francis, D. Cullen, D. Hibbett, A. Pisabarro, J. W. Spatafora, S. E. Baker, and I. V. Grigoriev. 2011. "Sequencing the Fungal Tree of Life." *New Phytologist* 190 (4): 818–21. doi:10.1111/j.1469-8137.2011.03688.x.
- Martin, Francis, Annegret Kohler, Claude Murat, Raffaella Balestrini, Pedro M Coutinho, Olivier Jaillon, Barbara Montanini, et al. 2010. "Périgord Black Truffle Genome Uncovers Evolutionary Origins and Mechanisms of Symbiosis." *Nature* 464 (7291): 1033–38. doi:10.1038/nature08867.

- Molinier, Virginie, Claude Murat, Emmanuelle Morin, Armelle Gollotte, Daniel Wipf, and Francis Martin. 2013. "First Identification of Polymorphic Microsatellite Markers in the Burgundy Truffle, *Tuber Aestivum* (Tuberaceae)." *Applications in Plant Sciences* 1 (2): 1200220. doi:10.3732/apps.1200220.
- Murat, C, C Riccioni, B Belfiori, N Cichocki, J Labbé, E Morin, E Tisserant, F Paolocci, A Rubini, and F Martin. 2011. "Distribution and Localization of Microsatellites in the Perigord Black Truffle Genome and Identification of New Molecular Markers." *Fungal Genetics and Biology* 48 (6): 592–601. doi:10.1016/j.fgb.2010.10.007.
- Murat, Claude, Andrea Rubini, Claudia Riccioni, Herminia De la Varga, Emila Akroume, Beatrice Belfiori, Marco Guaragno, et al. 2013. "Fine-Scale Spatial Genetic Structure of the Black Truffle (*Tuber Melanosporum*) Investigated with Neutral Microsatellites and Functional Mating Type Genes." *New Phytologist* 199 (1): 176–87. doi:10.1111/nph.12264.
- O'Donnell, Kerry, Elizabeth Cigelnik, Nancy S. Weber, and James M. Trappe. 1997. "Phylogenetic Relationships among Ascomycetous Truffles and the True and False Morels Inferred from 18S and 28S Ribosomal DNA Sequence Analysis." *Mycologia* 89 (1): 48. doi:10.2307/3761172.
- Pacioni, Giovanni, Cinzia Rapino, Osvaldo Zarivi, Anastasia Falconi, Marco Leonardi, Natalia Battista, Sabrina Colafarina, et al. 2015. "Truffles Contain Endocannabinoid Metabolic Enzymes and Anandamide." *Phytochemistry*. Accessed January 7. doi:10.1016/j.phytochem.2014.11.012.
- Paolocci, Francesco, Andrea Rubini, Claudia Riccioni, and Sergio Arcioni. 2006. "Reevaluation of the Life Cycle of *Tuber Magnatum*." *Applied and Environmental Microbiology* 72 (4): 2390–93. doi:10.1128/AEM.72.4.2390-2393.2006.
- Raffaele, Sylvain, and Sophien Kamoun. 2012. "Genome Evolution in Filamentous Plant Pathogens: Why Bigger Can Be Better." *Nature Reviews. Microbiology* 10 (6): 417–30. doi:10.1038/nrmicro2790.
- Riccioni, Claudia, Beatrice Belfiori, Andrea Rubini, Valentina Passeri, Sergio Arcioni, and Francesco Paolocci. 2008. "Tuber *Melanosporum* Outcrosses: Analysis of the Genetic Diversity within and among Its Natural Populations under This New Scenario." *New Phytologist* 180 (2): 466–78. doi:10.1111/j.1469-8137.2008.02560.x.
- Riley, Robert, Asaf A. Salamov, Daren W. Brown, Laszlo G. Nagy, Dimitrios Floudas, Benjamin W. Held, Anthony Levasseur, et al. 2014. "Extensive Sampling of Basidiomycete Genomes Demonstrates Inadequacy of the White-Rot/brown-Rot Paradigm for Wood Decay Fungi." *Proceedings of the National Academy of Sciences* 111 (27): 9923–28. doi:10.1073/pnas.1400592111.

- Rubini, Andrea, Beatrice Belfiori, Claudia Riccioni, Emilie Tisserant, Sergio Arcioni, Francis Martin, and Francesco Paolocci. 2011. "Isolation and Characterization of MAT Genes in the Symbiotic Ascomycete Tuber *Melanosporum*." *New Phytologist* 189 (3): 710–22. doi:10.1111/j.1469-8137.2010.03492.x.
- Rubini, Andrea, Francesco Paolocci, Claudia Riccioni, Giovanni G. Vendramin, and Sergio Arcioni. 2005. "Genetic and Phylogeographic Structures of the Symbiotic Fungus Tuber *Magnatum*." *Applied and Environmental Microbiology* 71 (11): 6584–89. doi:10.1128/AEM.71.11.6584-6589.2005.
- Sherman, David J., Tiphaine Martin, Macha Nikolski, Cyril Cayla, Jean-Luc Souciet, and Pascal Durrens. 2009. "Génolevures: Protein Families and Synteny among Complete Hemiascomycetous Yeast Proteomes and Genomes." *Nucleic Acids Research* 37 (suppl 1): D550–D554. doi:10.1093/nar/gkn859.
- Smith, Sally E., and David J. Read. 2010. *Mycorrhizal Symbiosis*. Academic Press.
- Spatafora, Joseph W., Gi-Ho Sung, Desiree Johnson, Cedar Hesse, Benjamin O'Rourke, Maryna Serdani, Robert Spotts, et al. 2006. "A Five-Gene Phylogeny of Pezizomycotina." *Mycologia* 98 (6): 1018–28. doi:10.3852/mycologia.98.6.1018.
- Splivallo, Richard, Simone Ottonello, Antonietta Mello, and Petr Karlovsky. 2011. "Truffle Volatiles: From Chemical Ecology to Aroma Biosynthesis." *New Phytologist* 189 (3): 688–99. doi:10.1111/j.1469-8137.2010.03523.x.
- Traeger, Stefanie, Florian Altegoer, Michael Freitag, Toni Gabaldon, Frank Kempken, Abhishek Kumar, Marina Marcet-Houben, Stefanie Pöggeler, Jason E. Stajich, and Minou Nowrousian. 2013. "The Genome and Development-Dependent Transcriptomes of *Pyronema Confluens*: A Window into Fungal Evolution." *PLoS Genet* 9 (9): e1003820. doi:10.1371/journal.pgen.1003820.
- Yoshida, Kentaro, Hiromasa Saitoh, Shizuko Fujisawa, Hiroyuki Kanzaki, Hideo Matsumura, Kakoto Yoshida, Yukio Tosa, et al. 2009. "Association Genetics Reveals Three Novel Avirulence Genes from the Rice Blast Fungal Pathogen *Magnaporthe Oryzae*." *The Plant Cell Online* 21 (5): 1573–91. doi:10.1105/tpc.109.066324.

2 Eléments répétés dans les champignons filamenteux

Dans ce premier chapitre bibliographique intitulé « **Repeated Elements in Filamentous Fungi with a Focus on Wood-Decay Fungi** » et publié dans le livre « The Ecological Genomics of Fungi », nous avons décrit les différents types d'éléments répétés que l'on retrouve dans les génomes des champignons filamenteux. Dans ce chapitre nous avons tenté, à travers la littérature, de mettre en avant des questions concernant l'impact des éléments transposables sur les génomes des champignons. Notre analyse bibliographique permet de voir que les champignons vivants en interaction avec les plantes (pathogènes ou symbiotiques) ont tendance à avoir des génomes ayant une plus grande proportion en séquences répétées et principalement en éléments transposables. Plusieurs hypothèses peuvent expliquer cela comme : 1) l'acquisition d'éléments transposables par transfert horizontal de la plante aux champignons et 2) la théorie de l'« epi-transposon equilibrium » disant que suite à un stress (comme un changement d'hôte) la régulation des transposons est arrêtée ce qui provoque la colonisation du génome par ces éléments. En tout cas la forte présence de séquences répétées confère à ces génomes une plus grande plasticité et ainsi une plus grande capacité d'adaptation.

2 Repeated Elements in Filamentous Fungi with a Focus on Wood-Decay Fungi

Claude Murat¹, Thibaut Payen¹, Denis Petitpierre¹,
and Jessy Labbé²

¹UMR1136 INRA, Université de Lorraine, Lab of Excellence ARBRE, Centre INRA de Nancy, Champenoux, France

²BioSciences Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee

Introduction

In the last decade, the genome of several dozen filamentous fungi have been sequenced. Interestingly, vast diversity in genome size was observed (Fig. 2.1) with 14-fold differences between the 9 Mb of the human pathogenic dandruff fungus (*Malassezia globosa*; Xu, Saunders, et al., 2007) and the 125 Mb of the ectomycorrhizal black truffle of Périgord (*Tuber melanosporum*; Martin, Kohler, et al., 2010). Recently, Raffaele and Kamoun (2012) highlighted that the genomes of several lineages of filamentous plant pathogens have been shaped by repeat-driven expansion. Indeed, repeated elements are ubiquitous in all prokaryote and eukaryote genomes; however, their frequencies can vary from just a minor percentage of the genome to more than 60 percent of the genome. Repeated elements can be classified in two major types: satellites DNA and transposable elements. In this chapter, the different types of repeated elements and how these elements can impact genome and gene repertoire will be described. Also, an intriguing link between the transposable elements richness and diversity and the ecological niche will be highlighted.

Satellites DNA

Satellites are tandem repetitions of motifs ranging from one to thousands of nucleotides in length. The length of the motif is used to classify satellites in three groups: microsatellites (1–6 nucleotides), minisatellites (7–100 nucleotides), and satellites (more than 100 nucleotides). However this classification

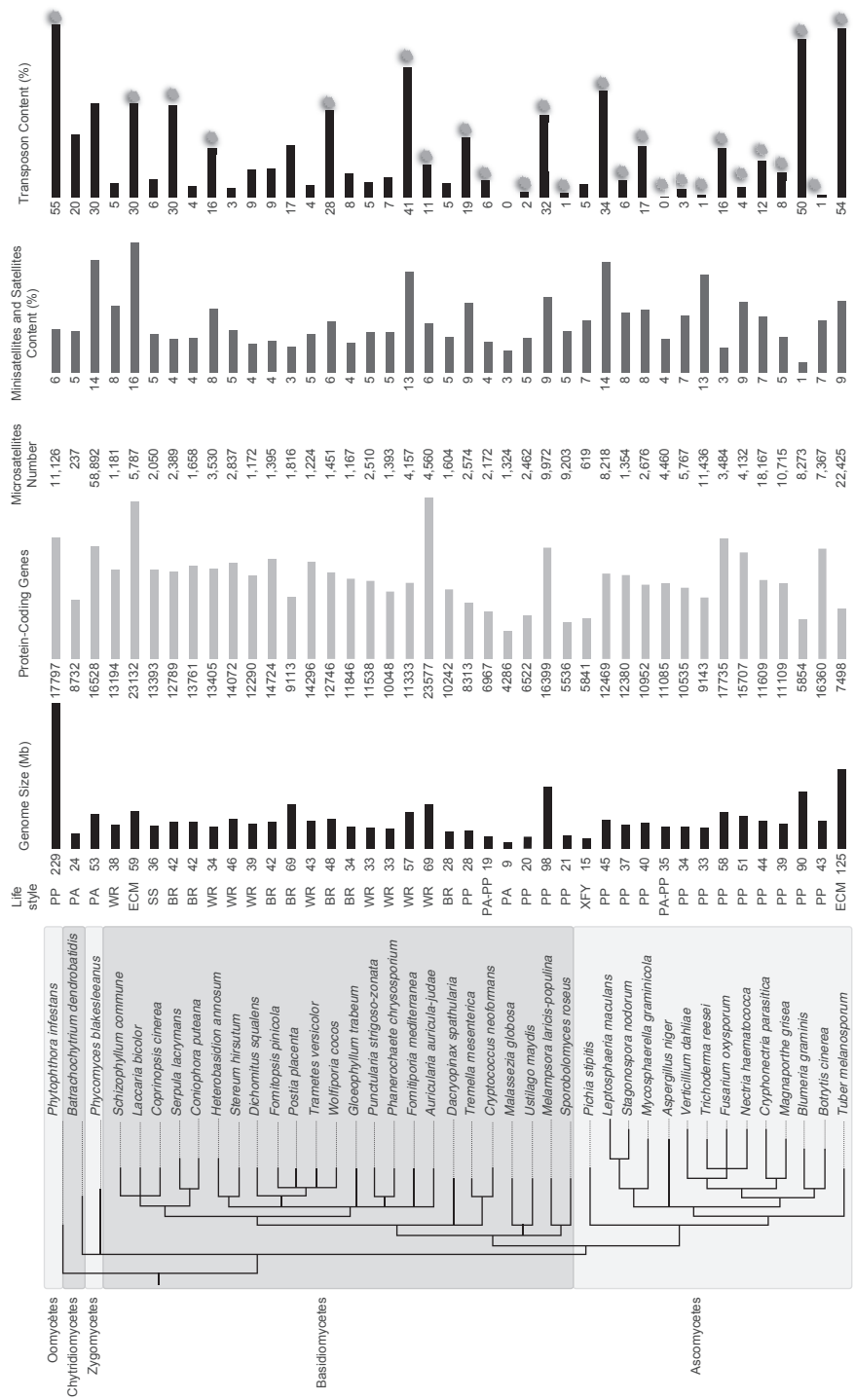


Figure 2.1 Main feature of some filamentous fungi sequenced belonging to plant pathogen, saprotroph, and mycorrhizal species. The phylogeny was generated using the Interactive Tree of Life (iTOL) with National Center for Biotechnology Information (NCBI) taxonomy identifiers (branch lengths are arbitrary). The life style (BR, brown rot; ECM, ectomycorrhizal; PA, animal pathogen; PP, plant pathogen; SS, soil saprotroph; WR, white rot; XFY, xylose fermenting yeast), genome size, protein content, number of microsatellites, percentage of genome coverage for minisatellites and satellites and transposable elements are indicated. The species interacting with living plants are indicated with a green star. Transposable element identification procedure is described in the supplementary data of Foudas, Binder, et al. (2012).

can vary according to authors (e.g., minisatellites is considered to be between 7 and 100 bp [Vergnaud & Denoeud 2000] or 10 and 50 bp [Jeffreys, Wilson, et al., 1985]).

Microsatellites

Microsatellites are the most studied tandem repeats, and several studies comparing microsatellite richness in filamentous fungi are available (Pannebakker, Niehuis, et al., 2010; Labbé, Murat, et al., 2011; Murat, Riccioni, et al., 2011). Murat, Riccioni, et al. (2011) compared the microsatellites of 48 fungal species highlighting large differences in the number of microsatellites among fungal species that vary from 224 elements for *Batrachochytrium dendrobatidis* to 56,846 elements for *Phycomyces blakesleeanus*. For all Ascomycetes and Basidiomycetes, microsatellites cover less than 1 percent of their genome at the exception of Mucoromycotina species: *Rhizopus oryzae* (1.5 percent) and *P. blakesleeanus* (2.5 percent), and Saccharomycotina species: *Lodderomyces elongisporus* (3.6 percent), *Candida tropicalis* (2.4 percent) and *Candida albicans* (1.9 percent). Interestingly the microsatellites do not seem to contribute significantly to genome size because no correlation between number and relative abundance of microsatellites with genome size was found (Murat, Riccioni, et al., 2011). Microsatellites are not distributed equally in the genome. In general, microsatellites seem more frequent in introns and intergenic regions, but for *L. bicolor*, they are also frequent in transposable elements (Labbé, Murat, et al., 2011). Although microsatellites may not have a clear effect on genome size, it has been shown that they modify protein function when localized in exons and can also modify gene expression when localized in promoter regulatory regions (Verstrepen, Jansen, et al., 2005; Riley & Krieger, 2009; Vincés, Legendre, et al., 2009; Rudd, Antoniwi, et al., 2010). Microsatellites in coding regions have been extensively studied within the context of human diseases, revealing abundant evidence on their contribution to neuronal diseases and cancers (Ashley & Warren, 1995). Microsatellite instability (MSI) is a hypermutable phenotype caused by the loss of DNA mismatch repair activity. MSI is detected in about 15 percent of all colorectal cancers (Boland & Goel, 2010). Microsatellite expansions or contractions in protein-coding regions can lead to a gain or loss of gene function via frameshift mutation or expanded toxic mRNA (Li, Korol, et al., 2004).

Verstrepen, Jansen, et al. (2005) found that in the genome of *Saccharomyces cerevisiae*, 75 percent of the gene models containing microsatellites in their coding region coded for cell surface proteins. Moreover, several gene models with microsatellites in their coding region have been implicated in plant infection of pathogenic fungi because they could be involved during the formation of intercellular hyphae or could conceivably

function as effectors during symptomless host plant colonization (Rudd, Antoniw, et al., 2010). Riley and Krieger (2009) highlighted that dinucleotide repeats in the untranslated region (UTR) of human genes are mainly involved in regulation of gene expression. This was shown for the microsatellites occurring in promoter regions of *S. cerevisiae* genes, where variations in repeat length can promote changes in expression (Vinces, Legendre, et al., 2009). In *T. melanosporum*, polymorphic microsatellites were found in the UTR of fruiting bodies and ectomycorrhiza-regulated genes, suggesting that they can have an effect on gene expression in fungi as in human, but no evidence of this effect is actually available (Murat, Riccioni, et al., 2011). The increasing interest of microsatellites studies and publications of the last years may come from their role in human disease and the multiplication of genomic resources.

Beside the potential effect of microsatellites on phenotype, microsatellites are also among the most popular molecular markers for population genetic in all organisms because they are assumed to be neutral and have a high level of polymorphism (Jarne & Lagoda, 1996). In human microsatellites, mutation rates of microsatellites are as frequent as 10^{-3} to 10^{-4} per locus per generation (Weber & Wong, 1993) compared with a rate of 10^{-8} per generation for single-nucleotide substitutions (Drake, Charlesworth, et al., 1998). For *T. melanosporum*, only the number of repetition for dinucleotide was correlated with the number of alleles. Dutech, Enjalbert, et al. (2007) found a similar result with a correlation of the mean repeat number for the dinucleotide and the number of alleles in several fungal species, in birds, insects, and fish but not in mammals and angiosperms. The mutation leading to the formation of a new allele can occur through different mechanisms, including errors during recombination, unequal crossing over and polymerase slippage during DNA replication, or from repair (Oliveira, Pádua, et al., 2006). If the accumulation of errors in a microsatellite continues indefinitely, then microsatellite will be large. This is not what was observed in the fungal genomes because longer microsatellites covered few hundred bp (e.g., 180, 204, 210, and 312 bp in *L. bicolor*, *P. blakesleeanus*, *Melampsora laricii-populina*, and *T. melanosporum*, respectively). There is therefore probably some selection acting against long microsatellites or a specific mechanism that eliminates long microsatellites. One hypothesis is that these sequences are particularly prone to big deletion or to single nucleotide substitution (Chambers & MacAvoy, 2000). This would generate small alleles that represent the “death” of the microsatellite under its life cycle (Hancock, Goldstein, et al., 1999; Chambers & MacAvoy, 2000). The smaller size of microsatellites in fungi suggest that in this group of organisms the phenomenon of single-sequence repeat (SSR) “death” did not allow the SSR to grow indefinitely, although in other eukaryotes some microsatellites spanning until 2 kbp have been observed (Sharma, Grover, et al., 2007).

The use of microsatellites to analyze fungal population genetics is not new but as highlighted by Dutech, Enjalbert, et al. (2007) microsatellites makers were not available for many species of fungi compared to other organisms. However, the availability of genome sequences coupled with new sequencing technology makes it easier now to characterize polymorphic microsatellites in non-model species. In fully sequenced genomes, microsatellites were characterized, using bioinformatic tools such as MISA or MAGELLAN, in several species such as *T. melanosporum* (Murat, Riccioni, et al., 2011), *L. bicolor* (Labbé, Murat, et al., 2011), and *M. laricii-populina* (Xhaard, Andrieux, et al., 2009). The new sequencing technologies become important to characterize microsatellites with two approaches using 454 pyrosequencing: enriched library sequencing or shotgun sequencing. The enriched library pyrosequencing was used by Malausa, Gilles, et al. (2011) to characterize microsatellites in insects, fungi, Oomycetes, and plants. However this requires information from enriched libraries, and this is why shotgun sequencing with 454 pyrosequencing seems more promising. Microsatellites can be identified directly in pyrosequencing reads without a preliminary assembly. This approach was used to characterize microsatellites in animals (e.g., termites, Singham, Vargo, et al., 2012), plants (e.g., *Acacia harpophylla*, Lepais & Bacles, 2011) and fungi (e.g., *Peltigera dolichorhiza* complex, Magain, Forest, et al., 2010; Burgundy Truffle [*T. aestivum*], Molinier, 2013). Beside pyrosequencing, Illumina sequencing was recently used to identify microsatellites in bird and snakes (Castoe, Poole, et al., 2012). Perl scripts named PAL_FINDER_v0.02.03 were used to extract reads with perfect microsatellites. The same pipeline PAL_FINDER_v0.02.03 allowed the design of primers to characterize new polymorphic markers with these microsatellites. These authors highlighted the usefulness of Illumina reads and concluded that Illumina “Seq-to-SSR” is effective, inexpensive, and reliable even for species that have few microsatellites loci.

Minisatellites and Satellites

Often, because of the lack of a precise definition of minisatellites, these structures are usually not annotated in genomic sequence data. There are few minisatellites studies relative to microsatellites focused studies. However, as for microsatellites, length variations of minisatellites have been found to be involved in several diseases such as diabetes, epilepsy, or cancer, and there is evidence for other contributions to genome function (see Buard & Jeffreys 1997; Vergnaud & Denoeud, 2000). The study of minisatellites in human genomics is not new because in 1985, Jeffreys, Wilson, et al. demonstrated that minisatellites with a repeat of 10 to 15 nucleotides could provide an individual-specific DNA “fingerprint” of general use in human genetic

analysis. In fungi few studies focused on minisatellites markers. Recently, Bally, Grandaubert, et al. (2010) defined a pipeline, FONZIE, aimed to provide a set of specific primer sequences for polymerase chain reaction (PCR) amplification of single-locus micro- and minisatellites markers. This pipeline was successfully used to characterize minisatellite markers for the pathogenic fungi *Leptosphaeria maculans* (Dilmaghani, Gladieux, et al., 2012). The advantage to studying minisatellites is that their size (motifs more than 7 nucleotides) allows the analysis of amplicons to be conducted directly on agarose gels, thereby reducing costs.

Unlike microsatellites, there is no study available, to our knowledge, comparing minisatellite and satellite patterns in fungal genomes. Using the software *tandem repeat finder* (Benson, 1999), minisatellites and satellites were identified in 40 fungal genomes belonging to Ascomycetes, Basidiomycetes, Zygomycetes, and Chytridiomycetes, as well as in one Oomycete (*Phytophthora infestans*) (see Fig. 2.1). In fungi, the number of minisatellites ranged from 1,631 for *B. dendrobatidis* to 110,404 for *T. melanosporum* (data not shown). The genome coverage of minisatellites ranged from 1.27 percent in *Blumeria graminis* to 13.4 percent in *L. maculans*. The number of satellites ranged from 28 for *M. globosa* to 3,772 for *L. bicolor* and the genome coverage was for all species less than 2 percent except for *L. bicolor* (4.17 percent). The majority (90 percent) of the motifs were smaller than 40 nucleotides and the most represented length was 21 nucleotides with a mean of 4,342 minisatellites of this size per genome (data not shown). However, the sequence of the 21 bp minisatellites is different from a species to another.

Unlike microsatellites, a correlation between the genome size and the number of minisatellites and satellites was found ($r^2=0,55$ and $r^2=0,53$, respectively). This suggests that minisatellites and satellites contribute to genome expansion; this seems particularly true for *L. bicolor*, *L. maculans*, *Fomitiporia mediterranea*, *Trichoderma reesei*, and *P. blakesleeanus* for which minisatellites and satellites represented more than 10 percent of their genome. The role of minisatellites and satellites in fungal genome is actually unknown. But that they could be important for gene regulation, such as microsatellites, and genomic rearrangement, such as transposable elements, can not be excluded.

In conclusion, satellites DNA are frequent in the fungal genomes and for some species this particular repeated sequences could represent more than 10 percent of the genome. Until now mainly microsatellites have been analyzed because of their usefulness as molecular markers. But the availability of many fungal species, as well as the easy tools existing to identify these sequences, make possible to define their pattern in fungal genomes now. Minisatellites and satellites could contribute to the expansion of genome size; another type of repeated sequences, transposable elements, will now be the focus.

Transposable Elements

The Different Families of Transposable Elements

Transposable elements (TEs) are short, mobile, conserved segments of DNA that can replicate and randomly insert copies within genomes of all species of the three domains of life: eubacteria, archaeabacteria, and eukaryotes. Although TEs were first identified in fungi in the yeast *S. cerevisiae* (Boeke, 1989) and known to exist in bacteria, plant, and animals since the 1970s, conventional genetic studies with *Ascobolus immersus* mutants established their existence in filamentous fungi beginning of the 1980s (Decaris, Francou, et al., 1978; Berg & Howe, 1989; Craig 2002). TEs are viewed as having an important influence on the evolution of eukaryote genomes and as central agents in the evolutionary restructuring of fungal genomes (Kidwell & Lisch, 2000). Given their abundance, these elements often constitute a large proportion of eukaryotic genomes (e.g., ~45% of the human genome, Lander, Linton, et al., 2001); 50 to 80 percent of some grass genomes (Meyers, Tingey, et al., 2001); and more than 50 percent in some fungal species (Martin, Kohler, et al., 2010; Spanu, Abbott, et al., 2010). Their dynamics include different mechanisms, such as transposition (normal or aberrant), ectopic recombination, horizontal transmission, amplification bursts, degradation, and epigenetic inactivation. Moreover, the examination of TE distribution in natural populations provided valuable information concerning ecological and epidemiological considerations (Daboussi & Capy, 2003). An enormous increase in the understanding of the biology of fungal TEs has occurred in the past decade because of the diversity of fungal research in organisms playing an important role in agriculture, medicine, and biotechnology and has been supported by the sequencing of more than 50 genomes. Here, the genomic features of TEs in filamentous fungi are reviewed, with a particular focus on their abundance, distribution, and importance in genome structure.

Eukaryotic TEs are divided into two classes, depending on their mode of transposition (for a review see Wicker, Sabot, et al., 2007; Nakayashidi, 2011): Class I elements or retroelements or also retrotransposons, which mobilize via a “copy-and-paste” mechanism that uses a RNA intermediate and class II elements or DNA transposons, which mobilize via a cut-and-paste mechanism that use a DNA intermediate. These two classes are composed of five major types: long terminal repeat (LTR) retrotransposons, non-LTR retrotransposons, cut-and-paste DNA transposons, rolling-circle DNA, and self-synthesizing DNA transposons. Each type of TE is composed of a number of superfamilies or clades based on length and target site features, with each superfamily consisting of numerous families. The retrotransposons (class I elements) are the most common TE in fungi (Boeke, Stoye, et al., 1997). As noted, retrotransposons can be classified into two types: LTR retrotransposons

and non-LTR retrotransposons (encompassing LINE elements), depending whether they possess or lack LTRs at both ends, tyrosine recombinase retroelements (YR; subdivided in three families, *DIRS*, *Ngaro*, and *VIPER*), Penelope-like retrotransposable elements, and short interspersed nuclear elements (SINEs). The LTR retrotransposons, which have a LTR at their extremities, have been divided into superfamilies: vertebrate retroviruses (Retroviridae), hepadnaviruses, caulimoviruses, *Ty1-Copia-like* (Pseudoviridae), *Ty3-Gypsy-like* (Metaviridae), and *Pao-BEL-like*, depending on their sequence similarity and the type of gene products they encode. The two main superfamilies of LTR retrotransposons found in fungi are *Gypsy* and *Copia*, which differ in the order of reverse transcriptase (RT), ribonuclease H (RH), and integrase (IN) domains in the virus-like polyprotein (POL; *Gypsy*: PR-RT-RH-INT, *Copia*: PR-INT-RT-RH). The DNA transposons (class II elements) have terminal inverted repeats (TIRs) or a rolling-circle replicon mechanism (e.g., *Helitrons* elements), similar to some known prokaryotic transposition mechanism or self-synthesizing DNA transposons (*Polintons*). Members of both classes are found in the genomes of filamentous fungi (Wicker, Sabot, et al., 2007).

Are Transposable Elements Impacting the Genome?

TEs have a remarkable potential to cause a variety of changes in the genome of their hosts. By transposing into or near genes, class I and class II TEs contribute to partial or total gene inactivation. Insertion may also place a gene under the control of TE regulatory sequences. Resulting from their ability to excise from a given site, class II transposons can generate a wide degree of variation in DNA sequence and phenotype. In addition, TEs have the ability to rearrange genomic information. DNA rearrangements may be local or associated with large-scale chromosomal modification. The range of transposon-associated genetic changes is well documented in many organisms (Kidwell & Lisch, 2002). Davière, Langin, et al. (2001) and Daboussi and Capy (2001) analyzed the impact of TEs in the rapid reorganization of the *Fusarium oxysporum* genome. Karyotypic variation is a common feature in natural fungal isolates, especially in those lacking the sexual cycle (Fierro & Martin, 1999; Kistler & Miao, 1992). Extensive analysis of chromosome length polymorphism has provided evidence that they include translocations, deletions of large chromosomal fragments, and much duplication. These studies showed that the high level of chromosome-length polymorphism of some chromosomes correlates with the high density of TEs (Davière, Langin, et al., 2001) and that the occurrence of chromosomal rearrangements is frequently associated with clustering of TEs on chromosomes (Hua-Van, Davière, et al., 2000). These findings suggest that they probably result from ectopic recombination

between TEs scattered throughout the genome. The range of karyotypic changes observed in some species during mitosis without phenotypic changes indicates that many of them are probably genetically neutral (Davière, Langini, et al., 2001), at least under laboratory conditions (Kistler & Miao, 1992). However, some rearrangements can be beneficial and may play an important role in the evolution of the host, as reported for wine yeast strains (Perez-Ortin, Querol, et al., 2002). Such events could lead to new gene linkages that may be advantageous for adaptation to new environments (e.g., the translocation-associated *Tox1* locus of *Cochliobolus heterostrophus* [Kodama, Rose, et al., 1999]) and the appearance of new virulent alleles in *Magnaporthe grisea* as a result of the rearrangements in unstable subtelomeric regions with nested repeated sequences (Orbach, Farrall, et al., 2000).

Filamentous fungi show a large variability in genome sizes (Raffaele & Kamoun, 2012). Filamentous fungi typically have small genomes in the 10- to 40-Mb range, usually with limited amounts of repetitive DNA (Baker, Thykaer, et al., 2008) and thus the Ascomycota and Basidiomycota appear to have a tendency toward streamlined genomes. The majority of these taxa contain no more than 10 to 15 percent repetitive DNA (Wöstemeyer & Kreibich, 2002). However some filamentous fungi are rich in non-coding DNA and display an irregular architecture, with an uneven distribution of genes and repetitive elements across and between chromosomes (Novikova, Fet, et al., 2009; Ma, van der Does, et al., 2010; Labbé, Murat, et al., 2012). Some species have genomes with an extremely high proportion of repetitive DNA, reaching 64 percent in *B. graminis* (Spanu, Abbott, et al., 2010). Typically, the expansion of filamentous fungal genomes can be largely accounted for by a proliferation of repetitive DNA. As seen previously, the satellites' DNA can contribute to expansion of genome size; however, this contribution is not sufficient to explain the entire genome size, and often TEs are the main responsible for genome expansion (see Fig. 2.1).

How Can Genomes Control Transposable Elements' Diversity and Expansion?

Diversity of TEs and their copy number depends on the evolutionary history of a particular species or a cluster of closely related species, their population structure, and ecological features. There are several main processes that could affect the copy number and diversity of TEs in fungal genomes: (a) stochastic loss of elements as described for mariner-like elements (Lohe, Mriyama, et al., 1995); (b) burst of transposition (e.g., in *T. melanosporum* two independent burst of Gypsy retrotransposons were highlighted [Martin, Kohler, et al., 2010]); (c) the limitation of copy number increase by natural selection, which removes deleterious insertions—the effect of deleterious insertion is difficult to evaluate but a

study was realized with *Drosophila melanogaster* to assess the effect of P element insertion on fitness (Mackay, Lyman, et al., 1992); (d) passive and active inactivation of repetitive sequences—these mechanisms will be described more in detail; and (e) self-regulation of transposition (decrease of the transposition rate when the copy number increases, Johnson, 2007)—this mechanism was proposed for MAGGY in *M. grisea* (Murata, Kadotani, et al., 2007). The population structure and dynamics, as well as mating mode and environmental conditions, also play an important role in the TEs' evolution. The transposition of the LTR retrotransposon MAGGY of *M. grisea* was shown to be particularly enhanced during mating and under abiotic stress (Eto, Ikeda, et al., 2001; Ikeda, Nakayashiki, et al., 2001). Similarly, numerous RTs and transposases were over-expressed in *T. melanosporum*–fruiting bodies, suggesting their activation during the sexual reproduction phase (Martin, Kohler, et al., 2010). Interestingly, Duplessis, Spanu, et al. (see Chapter 7) suggest that the frequency of sex during host infection for pathogenic fungi can impact the TE invasion.

The inactivation of repeated sequences is an important factor, which leads to the shifts in diversity and copy number of TEs, especially in fungi. The known mechanisms of repeat sequences inactivation include the repeat induced point (RIP) mutation, methylation induced premeiotically (MIP), and quelling. RIP was the first genome defense mechanism identified in eukaryotes, discovered in *Neurospora crassa* (Selker, Cambareri, et al., 1987). RIP occurs only during sexual cycle by introducing C:G to A:T mutations into two copies of duplications greater than about 400 bp with more than approximately 80 percent of nucleotide identity. In *Neurospora*, the RIP mutations are preferentially occurring in CpA dinucleotide (Cambareri, Jensen, et al., 1989). Actually RIP-like mechanisms were detected in *Podospora anserina* (Graia, Lospinet, et al., 2000), *M. grisea* (Ikeda, Nakayashiki et al., 2002), *L. maculans* (Idnurm & Howlett, 2003), and *Nectria haematococca* (Coleman, Rounsley, et al., 2009). Recently, Clutterbuck (2011) investigated the genome of 49 filamentous Ascomycetes to examine the evidence of multiple C/T transitions typical of RIP. The results highlighted that RIP-like activity varied greatly in extent of mutation as well as in dinucleotide context of C/T transition. Interestingly only *Chaetomium globosum* showed no evidence of directional mutation. In Basidiomycetes, RIP-like accumulation was described for *Puccinia graminis*, *M. larici-populina*, *Microbotrytium lychnidis-dioicae*, and *Rhodotorula graminis* (Hood, Katawczik, et al., 2005; Horns, Petit, et al., 2012) in which the target site seems to be the trinucleotide TpCpG. Horns, Petit, et al. (2012) did not find RIP-like hyperaccumulation mutation in four species of Agaromycotina and Ustilagomycotina, suggesting that RIP-like process is conserved within Puccinomycotina subphylum. Galagan and Selker (2004) highlighted that RIP not only impacted the genome via repeated sequences inactivation, but also that this mechanism was available to control gene duplication, which is considered as crucial for genome evolution. RIP

could also be a mechanism promoting gene divergence as for *L. maculans* effectors (Rouxelle, Grandaubert, et al., 2011).

The MIP was first described in *A. immersus* (Goyon & Faugeron, 1989). MIP was also detected in the Basidiomycete *Coprinopsis cinereus* (Freedman & Pukkila, 1993). MIP follows the same rules as RIP; that is, duplications are inactivated prior to meiosis, but results in cytosine methylation without mutation, and consequently this mechanism is reversible. These similarities suggested that RIP evolved from MIP (Selker, 2002). The methylation caused by MIP can block transcription elongation, resulting in gene silencing (Barry, Faugeron, et al., 1993). In *T. melanosporum*, the genes involved in RIP were not identified (Martin, Kohler, et al., 2010), although a strong preference for transition in CpG dinucleotide was observed (Clutterbuck, 2011). A possible explanation could be the presence of MIP that can increase the mutation of the methylated cytosines as documented for mammalian DNA (Krickler, Drake, et al., 1992).

The third mechanism is the quelling described in *N. crassa* that resembles posttranscriptional gene silencing in plant (Irelan & Selker, 1996). Quelling recognizes mRNA from repeated sequences in the vegetative tissues and targets them for degradation.

The Impact of Transposable Elements on Genes and Phenotypes

The investigation of the localization of TE in genomes can provide interesting information. The distribution of TE in the genome is different from a species to another; some species have TE located in clusters or nests of several hundred Kbp such as in the genome of *Trametes versicolor* (Floudas, Binder, et al., 2012). Often the regions rich in TEs are telomeric and centromeric. For example in *Agaricus bisporus*, 66 percent of the TEs are located in telomeric and centromeric regions (Foulongne-Oriol et al., unpublished). Other genomes have TEs all along their genome; this is the case for *F. mediterranea* (Floudas, Binder, et al., 2012), *T. melanosporum* (Martin, Kohler, et al., 2010) and *B. graminis* (Spanu, Abbott, et al., 2010). These last species have a high percentage of their genome corresponding to TEs (41.28 percent, 57.73 percent, and 64 percent for the three species, respectively). In general the TE rich regions are poor in genes but as will be seen, genes can be present in these regions impacting greatly their evolution and expression. In *Verticillium* spp, the observed biased TE insertion in gene-rich regions within an individual genome and the “patchy” distribution among different strains suggested that TE could be a major generator of *Verticillium* intra- and interspecific genomic variation (Amyotte, Tan, et al., 2012). One other example is the finding of particular TE associated with mating type locus in *Neurospora* spp. (Gioti, Mushegian, et al., 2012). This finding suggested that these elements could have contributed to the shift from heterothallic ancestors to homothallic

species by direct transposition of neighboring genes and facilitating unequal crossovers between unrelated intergenic regions of opposite mating types.

Repeat-rich genomic regions frequently coincide with synteny breakpoints, having evolved at accelerated rates compared with the rest of the genome. This is the case for the macrosynteny between *A. bisporus* and *C. cinereus* (Fig. 2 in Morin, Kohler, et al., 2012). Similarly, synteny breakpoints between *Sclerotinia sclerotiorum* and *Botrytis cinerea* are marked by an increased density of repetitive elements in *S. sclerotiorum* (Amselem, Cuomo, et al., 2011). For example, several synteny loss in species-specific secondary metabolism cluster loci appeared to be associated with the presence of TEs in one genome. Interestingly, between *Paxillus involutus* and *Pisolithus tinctorius*, some break of synteny are the result of a block of repeated sequences present at the same position in both species, suggesting that it appears in the common ancestor of these species (data not shown).

Such repeat rich regions tend to harbor genes that are implicated notably in virulence and host adaptation or as effector genes. For example in *L. maculans*, a plant pathogen, AT-rich blocks originated from RIP on repeated sequences, is enriched in effector-like (Rouxelle, Grandaubert, et al., 2011). The presence of effector-like genes in TE-rich regions for mildew species is discussed in Chapter 7. This particular environment induces rapid-sequence divergence and promotes the potentiality to adapt rapidly to new host constraint. TEs are also able to modify phenotype as shown for *Phytophthora ramorum* (Kasuga, Kozanitas, et al., 2012). These authors observed a burst of TE expression in oak isolates of *P. ramorum* in concomitance with phenotypic alterations, suggesting that TE derepression correlated with diversity in expression profiles leading to the phenotypic alteration.

TEs are generally distributed throughout the fungal genomes and could be major contributors to the genesis of new genes or to the adaptation of existing genes, notably via mechanisms such as molecular domestication, ectopic recombination, and gene retrotransposition. Molecular domestication, also known as the process of TE recruitment by the host genome, is the co-opted use by the organism of a function carried by a TE. Because TEs encode proteins that can, for example, bind, copy, break, join, or degrade nucleic acids, they have been repeatedly domesticated during eukaryotic evolution (Miller, McDonald, et al., 1999). As another mechanism, retrotransposon-mediated ectopic recombination results from the physical occurrence of retrotransposon insertions at particular sites in the genome and can imply various genomic rearrangements, such as duplications, deletions, and translocations. Gene retrotransposition is also another mechanism that can rearrange genes. Gene retrotransposition operates during the retrotransposition process itself and only duplicates gene sequences but no retrotransposons sequence. The genome sequences have revealed a lot of new information about the evolution of filamentous fungi and the genomic features that

underlie their success. Most strikingly, several lineages of filamentous fungi are remarkable in displaying an evolutionary trend toward bigger, TE-rich genomes.

Are Fungal Species Rich in Transposable Elements Interacting with Living Plants?

Recently, Raffaella and Kamoun (2012) discussed why some filamentous plant pathogens have convergent evolution toward large genomes infested with repetitive elements. For pathogenic fungi, the plasticity conferred by the TEs is thought to be adaptive because they increased recombination rate driven by TE activity. Consequently they adapt faster during coevolution with their host. These authors proposed that clade selection opposes the advantages conferred by smaller compact genomes because lineage with less adaptable genomes have an increased probability of extinction. Raffaella and Kamoun (2012) focused on pathogenic filamentous fungi and Oomycetes, but is it possible to draw similar conclusion with fungi having different life strategy? Interestingly, the two ectomycorrhizal fungal genomes, *L. bicolor* and *T. melanosporum*, have large genomes rich in repeated sequences (see Fig. 2.1). Recently, Floudas, Binder, et al. (2012) performed a comparative analysis of 31 fungal genomes suggesting that lignin-degrading peroxidases expanded in the lineage leading to the ancestor of the Agaricomycetes. To gain information about the effect of TEs on wood decay fungi genome size and to investigate a putative link between life style and TE richness, the repeated elements were characterized in these genomes (see supplementary data of Floudas, Binder, et al., 2012).

The TE genome coverage varied from 0 percent for *M. globosa* to 41.42 percent for the white rot *F. mediterranea* (see Fig. 2.1 and Table S5 in Floudas, Binder, et al., 2012). Repeated sequences have not fully disappeared from *M. globosa* genome because 3 percent of its genome corresponds to satellite DNA (see Fig. 2.1). For all the 31 genomes, a correlation between genome size and TE richness was found (see Fig. S4 in Floudas, Binder, et al., 2012). In most genomes, the Gypsy retrotransposons and not categorized elements are the most frequent (see Fig. S1 in Floudas, Binder, et al., 2012). In *F. mediterranea*, the Gypsy retrotransposons covered more than 20 percent of its genome and almost 30 percent of the *T. melanosporum* genome (Martin, Kohler, et al., 2010). To gain more information on Gypsy-like retrotransposons diversity in wood decay fungi, a specific identification of reverse transcriptase (RVT) was realized (Payen, Murat, et al., unpublished data). This analysis did not identify RVT in the genome of only five (*Aspergillus niger*, *Pichia stipidis*, *Stagonospora nodorum*, *T. reesei*, and *Ustilago maydis*) out of the 31 species. Almost 30 percent of the RVT identified have no homology with known Gypsy retrotransposon families and therefore can correspond to new families. For the other RVTs most of them belong to the Chromovirus clade with the exception of

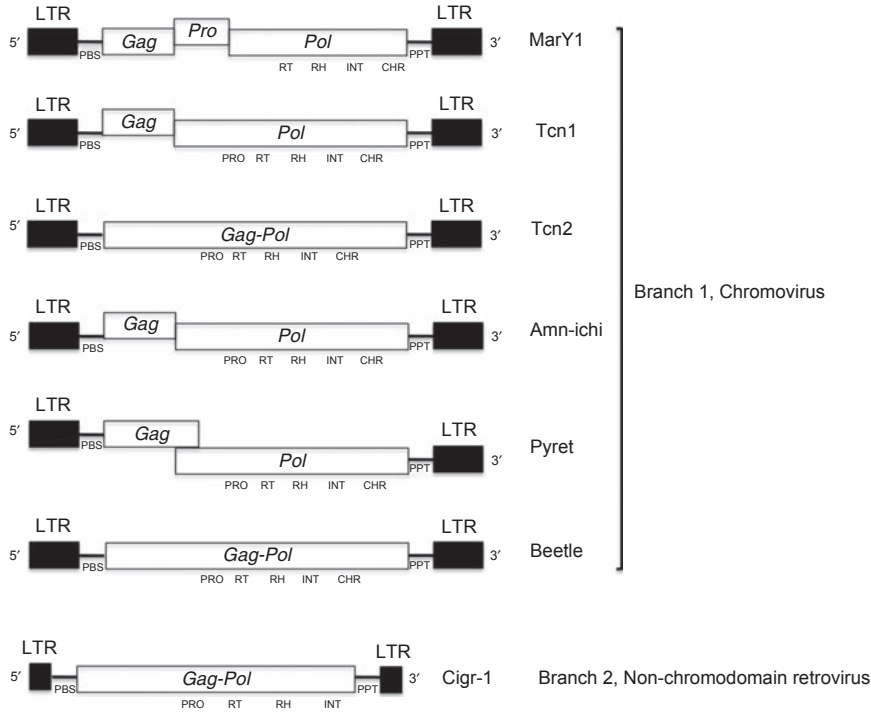


Figure 2.2 Structural organization of full-length LTR Gypsy retrotransposons found in filamentous fungi. The families presented are the most frequent found by RT screening in the genome species included in Floudas, Binder, et al. (2012). The gypsy retrotransposon diversity was assessed by RT identification using a RPS-Blast search (Altschul, 1990) with the reverse transcriptase 1 (RVT1) motif (pfam00078). For each species, the putative RVT1 sequences were isolated and clustered together using Usearch (Edgar, 2010) and 90 percent of similarity on at least 90 percent of the sequence (Gorinsek, Gubensek, et al., 2004). For each of these clusters a sequence was taken randomly. All clusters were aligned with the reference sequences coming from gypsyDB and others families (Gorinsek, Gubensek, et al., 2004). The alignment was done using Clustal Omega (Sievers, Wilm, et al., 2011). A phylogeny by neighbor joining was done using QuickTree (Howe, Bateman, et al., 2002) with a bootstrap value of 10,000. We have considered as belonging to a family of known Gypsy all the clusters supported by at least a bootstrap of 40 percent. CHR, chromodomain; INT, integrase; LTR, long terminal repeat; PBS, putative primer-binding site; PPT-polypurine tract; PRO, proteinase; RH, ribonuclease H; RT, reverse transcriptase.

Cigr-1 family, which presents in *M. laricii-populina* and *Cryptococcus neoformans* (Fig. 2.2). Interestingly, Cigr-1 was considered as a plant-animal lineage Gypsy retrotransposon family and elements of this family was not yet found in fungi (Sormacheva & Blinov, 2011). Additional investigations are needed to know if these Cigr-1 RVTs can get results from horizontal transfer because both species interact with plants. Indeed horizontal transfer was already suggested with LTR retrotransposons. Novikova, Smyshlyaev, et al.

(2010) proposed a horizontal transfer of Tcn1 gypsy retrotransposons between fungi and non-seed plant. Chromovirus were identified in all 26 genomes, and among them MarY1 family was the most frequent with 52.8 percent of the RVT sequences (see Fig. 2.2). MarY1 is a Chromovirus initially characterized in the genome of *Tricholoma matsutake* (Murata & Yamada, 2000) known to be widespread in fungus. The second frequent Gypsy retrotransposon family is Tcn2 known to be specific of Basidiomycetes (see Fig. 2.2; Novikova, Smyshlyaev, et al., 2010). Elements of this family were not identified in Ascomycetes. With the exception of *P. placenta*, all species with high number of RVTs (more than 200) interact with living plants. According to the present results and previous data on plant symbiotic and pathogenic fungi and Oomycetes, it seems that fungi rich in TEs interact with living plants (see Fig. 2.1; Martin, Aerts, et al., 2008; Haas, Kamoun, et al., 2009; Martin, Kohler, et al., 2010; Spanu, Abbott, et al., 2010; Raffaele & Kamoun 2012). How can this observation be explained?

One hypothesis is that species interacting with living plants need to evolve rapidly and TE can give plasticity to the genome. Recently, Zeh, Zeh, et al. (2009) proposed the “epi-transposon equilibrium” hypothesis in which TEs play the role of “punctuated equilibria.” The punctuated equilibrium means that the evolution proceeds through rapid morphological change and speciation followed by long-term stasis. These authors proposed that the punctuated equilibria result from an evolutionary tug-of-war between host genomes and TEs. According to the epi-transposon equilibrium hypothesis, stresses associated with climatic changes or colonization of new habitat or ecological niches result in TE reactivation, via disruption of epigenetic controls (e.g., MIP). TE can then rapidly modify the genome and gene expression, allowing the adaptation to the new conditions. It cannot be excluded that by changing their ecological niche (e.g., interaction with dead wood to living plant, transition between saprotrophic to symbiotic status) some fungal species activated the TE present in their genome. Interestingly for the two ectomycorrhizal fungi with genome sequenced to date, *L. bicolor* and *T. melanosporum*, RIP was not found. This absence of irreversible genome defense could explain the high proportion of TE in these genomes and the possibility to reactive their TE, and the question is if the RIP was not lost during evolution by ectomycorrhizal fungi to increase genome plasticity? A second hypothesis is that symbiotic fungi have a particular ecological niche linked with their host plants and therefore their populations could be more limited. These small populations could promote the TE expansion. Among the different mechanisms of TE invasion, the horizontal transfer is a possibility of new element to colonize a genome. Richards (2011) highlighted that the horizontal transfer could be linked with ecological niche; consequently it cannot be excluded that fungi interacting with living plants are subject to transfer of TE from the plant. This hypothesis needs to be taken in consideration for the future analyses. There is no doubt

that the genome projects aimed to sequence more mycorrhizal genomes as well as plant genomes will have more information on the possible link between TE abundance and the interaction with living plants.

Acknowledgments

We are grateful to Francis Martin, François Le Tacon, Emmanuelle Morin, Marc-Henri Lebrun, and Joëlle Amselem for the numerous critical discussions. We would like to thank also Francis Martin for the critical comments of the manuscript. This work was supported by grants from the National Institute of Agricultural Research, the Région Lorraine Council, Lab of Excellence ARBRE, ANR SYSTERRA SYSTRUF (ANR-09-STRA-10) and TUBEREVOL project of the Genoscope. The wood decay work conducted by the US Department of Energy Joint Genome Institute was supported by the Office of Science of the US Department of Energy under contract DE-AC02-05CH11231.

References

- Altschul S. 1990. Basic local alignment search tool. *J Mol Biol.* 215: 403–410.
- Amyotte SG, Tan X, et al. 2012. Transposable elements in phytopathogenic *Verticillium* spp.: Insights into genome evolution and inter- and intra-specific diversification. *BMC Genomics.* 13: 314.
- Amselem J, Cuomo CA, et al. 2011. Genomic analysis of the necrotrophic fungal pathogens *Sclerotinia sclerotiorum* and *Botrytis cinerea*. *PLoS Genet.* 7: e1002230.
- Ashley CT & Warren ST. 1995. Trinucleotide repeat expansion and human disease. *Annu Rev Genet.* 29: 703–728.
- Baker SE, Thykaer J, et al. 2008. Fungal genome sequencing and bioenergy. *Fungal Biol Rev.* 22: 1–5.
- Bally P, Grandaubert J, et al. 2010. FONZIE: An optimized pipeline for minisatellite marker discovery and primer design from large sequence data sets. *BMC Res Notes.* 3: 322.
- Barry C, Faugeron G, et al. 1993. Methylation induced premeiotically in *Ascobolus*: coextension with DNA repeat lengths and effect on transcript elongation. *Proc Natl Acad Sci USA* 90: 4557–4561.
- Benson G. 1999. Tandem repeats finder: A program to analyze DNA sequences. *Nucl Acids Res.* 27: 573–580.
- Berg DE & Howe MM. 1989. *Mobile DNA*. Washington DC: American Society for Microbiology Press.
- Boeke JD. 1989. Transposable elements in *Saccharomyces cerevisiae*. In *Mobile DNA* (eds. DE Berg & MM Howe), 335–374. Washington DC: American Society for Microbiology Press.
- Boeke JD, Stoye JP. 1997. Retrotransposons, endogenous retroviruses, and the evolution of retroelements, retroviruses. In *Retroviruses* (eds. JM Coffin, SH Hughes, et al.), 343–435. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.
- Boland CR & Goel A. 2010. Microsatellite instability in colorectal cancer. *Gastroenterology.* 138: 2073–2087.
- Buard J & Jeffreys A. 1997. Big, bad minisatellites. *Nat Genet.* 15:327–328.
- Cambareri EB, Jensen BC, et al. 1989. Repeat-induced G-C to A-T mutations in *Neurospora*. *Science.* 244: 1571–1575.

- Castoe TA, Poole AW, et al. 2012. Rapid Microsatellite Identification from Illumina Paired-End Genomic Sequencing in Two Birds and a Snake. *PLoS One*. 7(2): e30953.
- Chambers GK & MacAvoy ES. 2000. Microsatellites: Consensus and controversy. *Comp Biochem Physiol B Biochem Mol Biol*. 126: 455–476.
- Clutterbuck AJ. 2011. Genomic evidence of repeat-induced point mutation (RIP) in filamentous ascomycetes. *Fungal Genet Biol*. 48: 306–326.
- Coleman JJ, Rounsley SD, et al. 2009. The genome of *Nectria haematococca*: contribution of supernumerary chromosomes to gene expansion. *PLoS Genet*. 5: e1000618.
- Craig NL. 2002. *Mobile DNA II*. Washington, DC: American Society for Microbiology Press.
- Daboussi MJ & Capy P. 2003. Transposable elements in filamentous fungi. *Annu Rev Microbiol*. 57: 275–299.
- Davière JM, Langin T, et al. 2001. Potential role of transposable elements in the rapid reorganization of the *Fusarium oxysporum* genome. *Fungal Genet Biol*. 34(3): 177–192.
- Decaris B, Francou F, et al. 1978. Unstable ascospore color mutants of *Ascobolus immerses*. *Mol Gen Genet*. 81: 69–81.
- Dilmaghani A, Gladioux P, et al. 2012. Migration patterns and changes in population biology associated with the worldwide spread of the oilseed rape pathogen *Leptosphaeria maculans*. *Mol Ecol*. 21: 2519–2533.
- Drake JW, Charlesworth B, et al. 1998. Rates of spontaneous mutation. *Genetics*. 148: 1667–1686.
- Dutech C, Enjalbert J, et al. 2007. Challenges of microsatellite isolation in fungi. *Fungal Genet Biol*. 44: 933–949.
- Edgar RC. 2010. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics*. 26: 2460–2461.
- Eto Y, Ikeda K, et al. 2001. Comparative analyses of the distribution of various transposable elements in *Pyricularia* and their activity during and after the sexual cycle. *Mol Gen Genet*. 264: 565–577.
- Fierro F & Martin JF. 1999. Molecular mechanisms of chromosomal rearrangement in fungi. *Crit Rev Microbiol*. 25: 1–17.
- Floudas D, Binder M, et al. 2012. The oaleozoic origin of enzymatic lignin decomposition reconstructed from 31 fungal genomes. *Science*. 336: 1715–1719.
- Freedman T & Pukkila PJ. 1993. *De novo* methylation of repeat sequences in *Coprinus cinereus*. *Genetics*. 135: 357–366.
- Galagan JE & Selker EU. 2004. RIP: The evolutionary cost of genome defense. *Trends Genet*. 20: 417–423.
- Gioti A, Mushegian AA, et al. 2012. Unidirectional evolutionary transitions in fungal mating systems and the role of transposable elements. *Mol Biol Evol*. 29: 3215–3226.
- Gorinsek B, Gubensek F, et al. 2004. Evolutionary genomics of chromoviruses in eukaryotes. *Mol Biol Evol*. 21: 781–798.
- Goyon C & Faugeron G. 1989. Targeted transformation of *Ascobolus immersus* and *de novo* methylation of the resulting duplicated DNA sequences. *Mol Cell Biol*. 9: 2818–2827.
- Graia F, Lespinet O, et al. 2001. Genome quality control: RIP (repeat-induced point mutation) comes to *Podospora*. *Mol Microbiol*. 40: 586–595.
- Haas BJ, Kamoun S, et al. 2009. Genome sequence and analysis of the Irish potato famine pathogen *Phytophthora infestans*. *Nature*. 461: 393–398.
- Hancock JM, Goldstein DB, et al. 1999. Microsatellites and other simple sequences: genomic context and mutational mechanisms. In *Microsatellites: Evolution and Applications* (eds. DB Goldstein & C Schlotterer), 1–9. Oxford: Oxford University Press.
- Hood ME, Katawczik M, et al. 2005. Repeat-induced point mutation and the population structure of transposable elements in *Microbotryum violaceum*. *Genetics*. 170: 1081–1089.
- Horns F, Petit E, et al. 2012. Patterns of repeat-induced point mutation in transposable elements of basidiomycete fungi. *Genome Biol Evol*. 4: 240–247.
- Howe K, Bateman A, et al. 2002. QuickTree: Building huge Neighbour-Joining trees of protein sequences. *Bioinformatics*. 18: 1546–1547.

- Hua-Van A, Davière JM, et al. 2000. Genome organization in *Fusarium oxysporum*: Clusters of class II transposons. *Curr Genet.* 37:339–347.
- Jarne P & Lagoda PJJ. 1996. Microsatellites, from molecules to populations and back. *Trends Ecol Evol.* 11: 424–429.
- Idnurm A & Howlett BJ. 2003. Analysis of loss pathogenicity mutants reveals that Repeat-induced point mutation can occur in the Dothideomycete *Leptoshaeria maculans*. *Fungal Genet Biol.* 39: 31–37.
- Ikeda K, Nakayashiki H, et al. 2001. Heat shock, copper sulfate and oxidative stress activate the retro-transposon MAGGY resident in the plant pathogenic fungus *Magnaporthe grisea*. *Mol Genet Genomics.* 266: 318–325.
- Ikeda K, Nakayashiki H, et al. 2002. Repeat-induced point mutation (RIP) in *Magnaporthe grisea*: implication for its sexual cycle in the natural field context. *Mol Microbiol.* 45: 1355–1364.
- Irelan JT & Selker EU. 1996. Gene silencing in filamentous fungi: RIP, MIP and quelling. *J Genetics.* 3: 313–324.
- Jeffreys AJ, Wilson V, et al. 1985. Hypervariable “minisatellite” regions in human DNA. *Nature.* 314: 67–73.
- Johnson LJ. 2007. The genome strikes back: The evolutionary importance of defence against mobile elements. *Evol Biol.* 34: 121–129.
- Kasuga T, Kozanitas M, et al. 2012. Phenotypic diversification is associated with host-induced transposon derepression in the Sudden Oak death pathogen *Phytophthora ramorum*. *PLoS One.* 7(4): e34728.
- Kidwell MG & Lisch DR. 2000. Transposable elements and host genome evolution. *Trends Ecol Evol.* 15: 95–99.
- Kidwell MG & Lisch DR. 2002. Transposable elements as sources of genomic variation. In: *Mobile DNA II* (ed. NL Craig), 59–90. Washington, DC: American Society for Microbiology Press.
- Kistler HC & Mio VP. 1992. New modes of genetic change in filamentous fungi. *Annu Rev Phytopathol.* 30: 131–152.
- Kricker MC, Drake JW, et al. 1992. Duplication-targeted DNA methylation and mutagenesis in the evolution of eukaryotic chromosomes. *Proc Natl Acad Sci USA.* 89: 1075–1079.
- Kodama M, Rose MS, et al. 1999. The translocation-associated Tox1 locus of *Cochliobolus heterotrophus* is two genetic elements on two different chromosomes. *Genetics.* 151: 585–596.
- Labbé J, Murat C, et al. 2011. Survey and analysis of simple sequence repeats in the *Laccaria bicolor* genome, with development of microsatellite markers. *Curr Genet.* 57: 75–88.
- Labbé J, Murat C, et al. 2012. Characterization of transposable elements in the ectomycorrhizal fungus *Laccaria bicolor*. *PLoS One.* 7: e40197.
- Lander ES, Linton LM, et al. 2001. Initial sequencing and analysis of the human genome. *Nature.* 409: 860–921.
- Lepais O & Bacles CFE. 2011. Comparison of random and SSR-enriched shotgun pyrosequencing for microsatellite discovery and single multiplex optimization in *Acacia harpophylla* F. Muell, Ex Bebeth. *Mol Ecol Res.* 11: 711–724.
- Li Y-C, Korol AB, et al. 2004. Microsatellites within genes: Structure, function, and evolution. *Mol Biol Evol.* 21: 991–1007.
- Lohe AR, Mriyama EN, et al. 1995. Horizontal transmission, vertical inactivation and stochastic loss of mariner-like transposable elements. *Mol Biol Evol.* 12: 62–72.
- Ma L-J, van der Does HC, et al. 2010. Comparative genomics reveals mobile pathogenicity chromosomes in *Fusarium*. *Nature.* 464: 367–373.
- Mackay TFC, Lyman RF, et al. 1992. Effects of P-element insertions on quantitative traits in *Drosophila melanogaster*. *Genetics.* 130: 315–332.
- Magain N, Forrest LL, et al. 2010. Microsatellite primers in the *Peltigera dolichorhiza* complex (lichenized ascomycete, Peltigerales). *Am J Bot.* 97: e102–e104.
- Malausa T, Gilles A, et al. 2011. High-throughput microsatellite isolation through 454 GS-FLX Titanium pyrosequencing of enriched DNA libraries. *Mol Ecol Res.* 11: 638–644.

- Martin F, Aerts A, et al. 2008. The genome of *Laccaria bicolor* provides insights into mycorrhizal symbiosis. *Nature*. 452: 88–92.
- Martin F, Kohler A, et al. 2010. Périgord black truffle genome uncovers evolutionary origins and mechanisms of symbiosis. *Nature*. 464: 1033–1038.
- Meyers BC, Tingey SV, Morgante M. 2001. Abundance, distribution, and transcriptional activity of repetitive elements in the maize genome. *Genome Res*. 11: 1660–1676.
- Miller WJ, McDonald J F, et al. 1999. Molecular domestication—more than a sporadic episode in evolution. *Genetica*. 107: 197–207.
- Molinier V, Murat C, et al. 2012. First identification of polymorphic microsatellite markers in the Burgundy truffle, *Tuber aestivum* (Tuberaceae). *Application in Plant Sciences*. 1(2): 1200220. doi: <http://dx.doi.org/10.3732/apps.1200220>.
- Morin E, Kohler A, et al. 2012. Genome sequence of the button mushroom *Agaricus bisporus* reveals mechanisms governing adaptation to a humic-rich ecological niche. *Proc Natl Acad Sci USA*. 109: 17501–17506.
- Murat C, Riccioni C, et al. 2011. Distribution and localization of microsatellites in the Périgord black truffle genome and identification of new molecular markers. *Fungal Genet Biol*. 48: 592–601.
- Murata T, Kadotani N, et al. 2007. siRNA-dependent and -independent post-transcriptional cosuppression of the LTR-retrotransposon MAGGY in the phytopathogenic fungus *Magnaporthe grisea*. *Nucl Acid Res*. 35: 5987–5994.
- Murata H & Yamada A. 2000. marY1, a Member of the gypsy group of long terminal repeat retroelements from the ectomycorrhizal basidiomycete *Tricholoma matsutake*. *Appl Environ Microbiol*. 66: 3642–3645.
- Nakayashiki H. 2011. The trickster in the genome: Contribution and control of transposable elements. *Genes Cells*. 16: 827–841.
- Novikova O, Fet V, et al. 2009. Non-LTR retrotransposons in fungi. *Funct Integr Genomics*. 9: 27–42.
- Novikova O, Smyshlyaev G, et al. 2010. Evolutionary genomics revealed interkingdom distribution of Tcn1-like chromodomain-containing Gypsy LTR retrotransposons among fungi and plants. *BMC Genomics*. 11:231.
- Oliveira EJ, Pádua JG, et al. 2006. Origin, evolution and genome distribution of microsatellites. *Genet Mol Biol*. 29: 294–307.
- Orbach MJ, Farrall L, et al. 2000. A telomeric avirulence gene determines efficacy for the rice blast resistance gene Pi-ta. *Plant Cell*. 12: 2019–2032.
- Pannebakker BA, Niehuis O, et al. 2010. The distribution of microsatellites in the *Nasonia parasitoid* wasp genome. *Insect Mol Biol*. 19: 91–98.
- Perez-Ortin JE, Querol A, et al. 2002. Molecular characterization of a chromosomal rearrangement involved in the adaptive evolution of yeast strains. *Genome Res*. 12: 1533–1539.
- Raffaele S & Kamoun S. 2012. Genome evolution in filamentous plant pathogens: Why bigger can be better. *Nature Rev Microbiol*. 10: 417–430.
- Richards TA. 2011. Genome evolution: Horizontal movements in the fungi. *Curr Biol*. 21: 166–167.
- Riley DE & Krieger JN. 2009. UTR dinucleotide simple sequence repeat evolution exhibits recurring patterns including regulatory sequence motif replacements. *Gene*. 429: 80–86.
- Rouxelle T, Grandaubert J, et al. 2011. Effector diversification within compartment of the *Leptosphaeria maculans* genome affected by repeat-induced point mutation. *Nature Commun*. 2: doi:10.1038/ncomms1189.
- Rudd JJ, Antoniw J, et al. 2010. Identification and characterisation of *Mycosphaerella graminicola* secreted or surface-associated proteins with variable intragenic coding repeats. *Fung Genet Biol*. 47: 19–32.
- Selker EU, Cambareri EB, et al. 1987. Rearrangement of duplicated DNA in specialized cells of *Neurospora*. *Cell*. 51: 741–752.
- Selker EU. 2002. Repeat induced gene silencing in fungi. *Adv Genet*. 46: 439–450.

- Sharma PC, Grover A, et al. 2007. Mining microsatellites in eukaryotic genomes. *Trends Biotechnol.* 25: 490–498.
- Sievers F, Wilm A, et al. 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol.* 7: 539.
- Singham GV, Vargo EL, et al. 2012. Polymorphic microsatellite loci from an indigenous Asian fungus-growing termite, *Macrotermes gilvus* (Blattodea: Termitidae) and cross amplification in related taxa. *Environ Entomol.* 41: 426–431.
- Sormacheva ID & Blinov AG. 2011. LTR retrotransposons in plants. *Russ J Genet Appl Res.* 1: 540–564.
- Spanu PD, Abbott JC, et al. 2010. Genome expansion and gene loss in powdery mildew fungi reveal tradeoffs in extreme parasitism. *Science.* 330: 1543–1546.
- Vergnaud G & Denoeud F. 2000. Minisatellites: Mutability and genome architecture. *Genome Res.* 10: 899–907.
- Verstrepen KJ, Jansen A, et al. 2005. Intragenic tandem repeats generate functional variability. *Nature Genet.* 37: 986–990.
- Vinces MD, Legendre M, et al. 2009. Unstable tandem repeats in promoters confer transcriptional evolvability. *Science.* 324: 1213–1216.
- Weber JL & Wong C. 1993. Mutation of human short tandem repeats. *Human Mol Genet.* 2: 1123–1128.
- Wicker T, Sabot F, et al. 2007. A unified classification system for eukaryotic transposable elements. *Nature Rev Genet.* 8: 973–982.
- Wöstemeyer J & Kreibich A. 2002. Repetitive DNA elements in fungi (Mycota): Impact on genomic architecture and evolution. *Curr Genet.* 41: 189–198.
- Xhaard C, Andrieux A, et al. 2009. Characterization of 41 microsatellite loci developed from the genome sequence of the poplar rust fungus, *Melampsora larici-populina*. *Conservation Genet Resour.* 1: 21–25.
- Xu J, Saunders CW, et al. 2007. Dandruff-associated *Malassezia* genomes reveal convergent and divergent virulence traits shared with plant and human fungal pathogens. *Proc Natl Acad Sci USA.* 104: 18730–18735.
- Zeh DW, Zeh JA, et al. 2009. Transposable elements and an epigenetic basis for punctuated equilibria. *Bioessays.* 31: 715–726.

3 Phylogénomique du genre *Tuber*

Dans ce chapitre intitulé « **Truffle Phylogenomics: New Insights into Truffle Evolution and Truffle Life Cycle** » publié dans « *Advances in Botanical Research, Volume 70* » nous avons passé en revue les récentes découvertes sur l'évolution des truffes et leur répartition géographique au cours de l'histoire. Ce chapitre introduit également de façon succincte les travaux de génomique comparative au sein des Pézizomycètes en détaillant l'intérêt de ce type d'analyses. Dans ce chapitre nous avons mis en avant les questions qui nous semblaient pertinentes à étudier pour mieux connaître le cycle biologique des truffes et ainsi en améliorer la production. Ces questions sont :

1. Quels facteurs favorisent la germination des spores ?
2. Quelle est la taille des génotypes de truffes et leur durée de vie ?
3. Comment pouvons-nous promouvoir l'initiation de la reproduction sexuée ?
4. Comment préserver la truffe lors de sa croissance qui est liée à l'arbre ?
5. Quel est le rôle de la microfaune et des microorganismes dans le cycle biologique des truffes ?
6. Est-ce que les truffes sont capables de s'adapter aux changements climatiques ?
7. Quels parcours techniques permettent de favoriser le cycle biologique des truffes dans les truffières ?



Truffle Phylogenomics: New Insights into Truffle Evolution and Truffle Life Cycle

Thibaut Payen^{*}, Claude Murat^{*,1}, Gregory Bonito[†]

^{*}Institut National de la Recherche Agronomique (INRA), UMR 1136 INRA/Lorraine University, Interactions Arbres/Micro-organismes, Centre de Nancy, Champenoux, France

[†]Royal Botanic Gardens, Melbourne, South Yarra, Australia

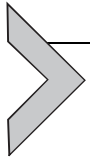
¹Corresponding author: e-mail address: claudemurat@nancy.inra.fr

Contents

1. Introduction	212
2. Truffle Diversity and Phylogeography	213
2.1 <i>Tuber</i> biodiversity	213
2.2 Phylogeography and origin of <i>Tuber</i>	213
3. Harnessing Genomes to Unravel the Black Périgord Truffle Life Cycle	217
3.1 Black Périgord truffle life cycle: An intimate relation between the fungi, plant, and climate	217
3.2 Intraspecific genetic diversity: From microsatellites to SNPs	219
3.3 Sexual reproduction: Where are both mating-type strains?	222
4. Comparative Genomic in Pezizomycetes	224
4.1 Sequencing the genome of <i>Tuber</i> spp.	224
4.2 Sequencing the genome of other Pezizomycetes	227
5. Concluding Remarks	229
Acknowledgements	230
References	230

Abstract

Truffles are ectomycorrhizal filamentous fungi belonging to genus *Tuber* naturally found across a range of climates in the Northern Hemisphere. Some truffle species such as Périgord black truffle (*T. melanosporum*) and the white truffle (*T. magnatum*) in Europe, the black Chinese truffle (*T. indicum*) in Asia, and the pecan truffle (*T. lyonii*) in North America are particularly famous due to the organoleptic qualities of their fruiting bodies and are among the world's most precious culinary delicacies. In this chapter, we summarize the recent studies pertaining to the phylogeography and biogeographic history of the genus *Tuber*, genomic analysis of the black truffle *Tuber melanosporum*, and comparative genomics towards understanding the life cycle, ecology, and evolution of *Tuber* species. Such understanding is critical to the anticipated future of truffle production in light of global change.



1. INTRODUCTION

True truffles are ectomycorrhizal filamentous fungi belonging to genus *Tuber* and living in symbiosis with numerous trees and shrubs. Truffles are naturally found across the Northern Hemisphere under a range of climates. Some European species including the burgundy truffle (*Tuber aestivum*), the Périgord black truffle (*T. melanosporum*), and the white piedmont truffle (*T. magnatum*) are particularly famous due to their sought after fruiting bodies (called ascocarps or truffles), which are among the world's most precious culinary delicacies. Economically important truffle species also occur in North America (i.e. the pecan truffle *T. lyonii*) and in Asia (i.e. the Chinese truffles *T. indicum* and *T. sinoaestivum*).

The cultivation of truffles was born with Joseph Talon (1793–1873) who told “*Si vous voulez récolter des truffes, semez des glands*” (If you want to harvest truffles, sow acorns). However, the link between truffles and trees was not fully understood until the first description of ectomycorrhizae by Frank (1889), who described ectomycorrhizae formed by *T. aestivum* and hypothesized the function of these symbiotic structures. After the grape phylloxera crisis of at the end of the nineteenth century, (Hall, Brown, & Zambonelli, 2007) grape vineyards in France were replaced with truffle orchards and the golden age of truffle production began.

During the twentieth century, truffle production in France declined from more than 1000 t to less than 100 t. This decline of truffle production has been attributed to several causes. The first rural desertification led to a decrease in the surface area of land devoted to truffle cultivation. Further, remaining truffle orchards were no longer managed; as the trees matured their canopies gradually closed altering microhabitats and water availability and leading to a decline in truffle production. Climate change has also been evoked to explain declining truffle yields (Büntgen et al., 2012) through alterations in regional rainfall and climate patterns.

The success in the mass production of seedlings colonized with *T. melanosporum* ectomycorrhizae began in the 1970s and has led to an increase in the area planted with truffles and their hosts (Chevalier, 1998; Chevalier & Grente, 1979). In France alone, 800–1000 ha of plantations have been established annually over the past 20 years. However, even with these efforts, truffle production has not reached the levels achieved at the beginning of the twentieth century. The truffle industry is currently requesting innovative tools to intelligently develop truffle cultivation into

a sustainable and fully profitable activity. In this chapter, we will present the contribution of comparative genomics towards understanding the truffle life cycle, ecology, and evolution. Such understanding is critical to the anticipated future of truffle production in light of global change. The most recent studies pertaining to the phylogeography and biogeographic history of the genus *Tuber* will also be presented.



2. TRUFFLE DIVERSITY AND PHYLOGEOGRAPHY

2.1. *Tuber* biodiversity

The genus *Tuber* is among the most diverse groups of fungi to produce belowground fruiting bodies known as truffles. *Tuber* is estimated to include a minimum of 180 species worldwide (Bonito, Gryganskyi, Trappe, & Vilgalys, 2010). Much of this diversity remains undescribed. The majority of newly described and undescribed diversity in *Tuber* resides in poorly studied clades. In particular, the spiny-spored Rufum clade and the aveolate-reticulate spored Maculatum and Puberulum clades are particularly speciose (Bonito, Gryganskyi, et al., 2010). Many (but not all) of the species in these groups produce smaller-sized fruiting bodies that are morphologically cryptic and lack any apparent culinary value, attributes that may explain why these species have been largely overlooked by mycologists. With renewed interest in *Tuber* diversity and with the aid of molecular tools for identification, an impressive number of new *Tuber* species have recently been described in Europe (Alvarado, Moreno, & Manjón, 2012; Božac, Širić, & Kos, 2012), North America (Bonito, Trappe, Rawlinson, & Vilgalys, 2010; Guevara et al., 2013), and Asia (Deng, Liu, Liu, & Wang, 2013; Fan, Cao, & Li, 2012a, 2012b; Fan, Cao, & Yu, 2012; Fan, Hou, & Li, 2012). Some of these newly described species are likely to have value as culinary species.

2.2. Phylogeography and origin of *Tuber*

Recent studies aimed at assessing the phylogenetic diversity of *Tuber* and reconstructing the truffle family tree have given new understanding pertaining to the origin, evolution, and biogeographic history of *Tuber* (Bonito et al., 2013; Jeandroz, Murat, Wang, Bonfante, & Tacon, 2008). Based on extensive biogeographic sampling of *Tuber*, multigene datasets were generated for molecular dating and phylogenetic approaches. In these

studies, *Tuber* was resolved as monophyletic, and a common ancestor that was most likely an aboveground cup fungus was inferred. [Bonito et al. \(2013\)](#) used a concatenated dataset of 4 loci to resolve 11 major clades in *Tuber* ([Fig. 7.1](#)) and found a high degree of species-level endemism in the genus. While species may show high levels of endemism, *Tuber* clades including the Rufum, Melanosporum, Puberulum, Maculatum, and Macrosporium are distributed across the entire Northern Hemisphere (Europe, Asia, North America, Central America, and Northern Africa), and the Aestivum and Excavatum clades are distributed in both Europe and Asia. A few clades show continental endemism including the Gennadii and Multimaculatum clades in Europe, the Japonicum clade in Asia, and the Gibbosum clade in North America. While some *Tuber* species appear to be generalists in regard to their ectomycorrhizal hosts (e.g. *T. aestivum* and *T. indicum*), host preferences show higher-level patterns. For instance, most of the species in the Rufum, Excavatum, Aestivum, Maculatum, and Gennadii clades tend to be associated with angiosperm hosts, while those in the Japonicum and Gibbosum clades appear to prefer hosts in the Pinaceae. The Puberulum clade exhibits the widest range of host associations and many species are host generalists.

[Bonito et al. \(2013\)](#) carried out divergence dating analyses in BEAST using a relaxed clock model with an uncorrelated exponential prior distribution. They estimated the divergence of *Tuber* from other genera occurred during the early Cretaceous Period 156 Mya (fig. 5—node 4 in [Bonito et al., 2013](#)), with most of the extant subgeneric lineages present 65 Mya and experiencing major radiations during the Paleogene ([Bonito et al., 2013](#)). Estimates for the mean ages of the most recent common ancestor (MRCA) of the 11 resolved *Tuber* clades are Multimaculatum (121 Mya), Aestivum (101 Mya), Rufum (86 Mya), Melanosporum (79 Mya), Puberulum (65 Mya), Japonicum (46 Mya), Excavatum (43 Mya), Maculatum (67 Mya), Macrosporium (43 Mya), Gennadii (48 Mya), and Gibbosum (27 Mya). While there is still much uncertainty whether the origin of *Tuber* was European or Asian ([Bonito et al., 2013](#); [Jeandroz et al., 2008](#)), putative geographic origins of the most common ancestors for a majority of clades were resolved by [Bonito et al. \(2013\)](#) with North America as the ancestral area of the Gibbosum, Maculatum, Rufum, and Melanosporum clades; Europe as the ancestral area of the Aestivum, Excavatum, and Gennadii clades; and Asia as the ancestral area for the Japonicum clade.

Tuber is often considered to have a Laurasian distribution because species are naturally distributed across Northern latitudes in North America,

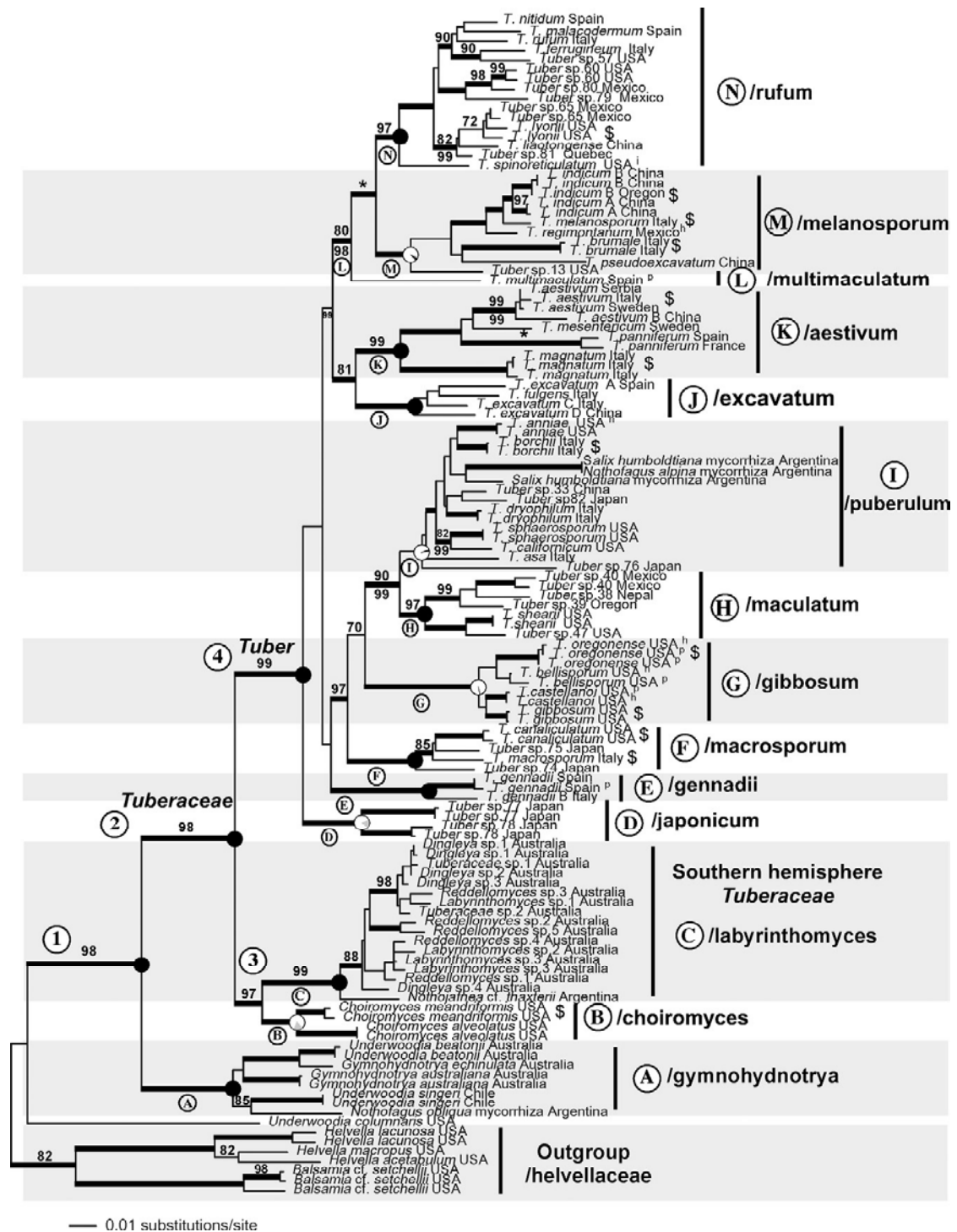


Figure 7.1 Maximum likelihood (ML) phylogenetic reconstruction of the *Tuberaceae* phylogeny based on ITS, 28S rDNA, EF1, and RPB2 genic loci. Thickened branches represent ML bootstrap support >70 and posterior probabilities of 100. ML bootstrap values above nodes are based on 1000 replicates. Posterior probabilities are presented below nodes. Thickened branches without numbers received maximum ML and Bayesian support values. Reconstructed ancestral host plant associations (based on maximum likelihood) are represented at internal nodes by circles, black for ancestors in symbiotic association with angiosperms, white for ancestors in symbiotic association with *Pinaceae*, and grey for ancestors in symbiotic association with angiosperms and *Pinaceae*. (Continued)

Europe, Asia, and Northern Africa. However, species of *Tuber* belonging to the *Puberulum* clade were recently detected as ectomycorrhizae on *Nothofagus* in Argentina (Bonito et al., 2013). While humans have dispersed fungi to many continents, these South American species appear to be phylogenetically novel and are likely native species whose ancestors migrated into South America from the North, putatively in association with an ectomycorrhizal host such as *Salix*. In addition to having wide host preferences, the *Puberulum* clade appears to have the widest geographic distribution of all *Tuber* clades.

With the growing interest in cultivating truffles *Tuber* is now being introduced into areas outside their native range. For instance, the European black truffle *Tuber melanosporum*, which is the most important cultivated *Tuber* species, has been introduced globally, including Sweden, South Africa, Morocco, Israel, China, United States, Canada, Mexico, Chile, Argentina, Australia, and New Zealand. An understanding of the ecological and physiological constraints of *T. melanosporum* is necessary to ensuring that these efforts are successful. Unfortunately, in the act of introducing *Tuber melanosporum* into these countries, other *Tuber* species have also been introduced outside their range. Given that unintentional introductions of *T. indicum* or *T. brumale* have been documented in nearly every country where mycologists have carried out molecular sequence-based assessments (e.g. the United States, Italy, New Zealand, Australia, and Canada), it is likely that nontarget *Tuber* introductions are a frequent occurrence where efforts to cultivate *T. melanosporum* have been made (Bonito, Trappe, Donovan, & Vilgalys, 2011; Guerin-Laguette, Cummings, Hesom-Williams, Butler, & Wang, 2013; Murat, Zampieri, Vizzini, & Bonfante, 2008). Biosecurity measures to mitigate against further unintentional introductions of *Tuber* are needed and should include molecular testing of inoculum used for inoculations. While specific primers have been designed for rapid identification of target species (Amicucci, Guidi, Zambonelli,

Figure 7.1—Cont'd Nodes supported by transitions in spore ornamentation from alveolate–reticulate to spiny are shown with an asterisk *. Economically important species are denoted by the symbol \$ after their name and geographic origin. The phylogeny is rooted with taxa from the *Helvellaceae* including species of epigeous *Helvella* and hypogeous *Balsamia*. Major lineages of *Tuber* and *Tuberaceae* are indicated to the right of the tree. The *Tuberaceae* form a monophyletic group, which is resolved as a sister group to a previously unrecognized Southern Hemisphere lineage (Gymnohydnotrya). Type specimens are denoted by the superscripts: h, holotype; i, isotype; and p, paratype. *Reprinted with permission from Bonito et al. (2013).*

Potenza, & Stocchi, 2000; Bonito, 2009; Paolocci, Rubini, Granetti, & Arcioni, 1999; Zampieri, Mello, Bonfante, & Murat, 2009), high-throughput sequenced-based assessments of inoculum slurries and root communities (i.e. amplicon sequencing by 454 pyrosequencing or Illumina) offer a more comprehensive strategy for ensuring inoculum quality for truffle tree nurseries and may allow for simultaneous assessments of overall genetic diversity levels or biogeographic origin of species in the inoculum.



3. HARNESSING GENOMES TO UNRAVEL THE BLACK PÉRIGORD TRUFFLE LIFE CYCLE

3.1. Black Périgord truffle life cycle: An intimate relation between the fungi, plant, and climate

The life cycle of the black Périgord truffle begins in late winter or early spring with the germination of haploid spores released by mature ascocarps. The growing haploid mycelium colonizes tree roots to form ectomycorrhizal symbiotic associations (Paolocci, Rubini, Riccioni, & Arcioni, 2006; Riccioni et al., 2008). In May or June, after plasmogamy (the fusion of haploid mycelia and the initial fertilization event), an ascogenous heterokaryotic tissue surrounded by homokaryotic maternal tissues gives birth to the ascocarp. This ascocarp primordium is linked to the ectomycorrhizae until the truffle fruiting body completely matures and requires carbon stored in the trunk or roots of the host plant for the truffle to grow and mature (Le Tacon et al., 2013). As recently demonstrated, truffles are heterothallic species and mating-type genes coding for sexual compatibility have been identified (see in the succeeding text). The young truffle, born in the late spring or beginning of summer, grows slowly during the summer and starts to mature in autumn. The crop in the Northern Hemisphere generally starts in mid-November and finishes mid-March.

The interaction between truffles and their host plant is critical since truffles need to be associated with a host plant to complete their life cycle by producing mature fruiting bodies. Indeed, truffle ascocarps need carbon to grow, and as demonstrated recently, nearly all of this carbon comes from the host plant photosynthate via ectomycorrhizae (Le Tacon et al., 2013). In this study, stable ^{13}C -labelling was used to demonstrate that the ascocarps accumulate host carbon until reaching complete maturity. *Tuber* ascocarps remained a carbon sink even several months after the end of photosynthetic carbon assimilation by the host and at low temperature. These results have important field implications. For instance, guidelines for tree pruning are not

yet available. Winter pruning could reduce carbon availability for maturing truffles. More detailed studies on the interaction and dialogue between truffle and host tree has been investigated through transcriptomic analyses (Kohler & Tisserant, 2014).

The sequenced black truffle genome is an exceptional resource for investigating the development and physiology of the black truffle (Kües & Martin, 2011). Transcriptomic analyses of tissues under different conditions including vegetative hyphal growth (Amicucci et al., 2011), cold stress (Zampieri et al., 2011), different stages in the life cycle (Montanini et al., 2011), and metal homeostasis involvement (Bolchi et al., 2011) have been investigated. Splivallo and colleagues (Splivallo, Ottonello, Mello, & Karlovsky, 2011) also highlighted putative ecological roles of truffle volatiles as chemical signals between the fungus, other soil organisms, and plant roots.

The long duration of the truffle life cycle makes truffle harvests sensitive to climatic conditions of the entire year and particularly sensitive to the water balance of summer months, which is often critical in the Mediterranean climate. It is currently understood and well documented that black truffle production is tightly linked to climate (Büntgen et al., 2012). The role of drought in diminishing truffle production is well established. For instance, in 1914, Pradel demonstrated the importance of supplementing water to truffle-producing trees during drought summers to maintain truffle production. More recently, Le Tacon, Delmas, Gleyze, and Bouchard (1982) determined that maintaining soil water potential above 3.5 during the summer months with irrigation increased truffle production by a factor of 18 in 1979–1980 and by more than 40 in 1980–1981. Moreover, the interannual variability in black truffle production observed over the past 25 years is highly correlated with fluctuations in rainfall (Le Tacon et al., 2014). The consequence of drought for truffle production could be important. For example, the 2003 drought caused a drop in French production to about 80%; most of the truffle grounds affected, without water, did not produce any truffles as a result of this drought. This suggests that models and guidelines for sustainable watering of truffières are needed. Drought severely impacts truffle growers' incomes and consequently the whole truffle industry, leading to unpredictable truffle production and a fluctuating truffle market. Drought is expected to increase in both frequency and severity in most European regions as a result of changes in the temporal distribution of rainfall and increase in potential evapotranspiration driven by global warming (Sheffield, Wood, & Roderick, 2012). Truffles may have nonlinear and inconsistent responses to drought, depending on which phase of their life

cycle is subjected to water stress. A better knowledge of climatic conditions and the effects of truffle growers' interventions (e.g. soil tilling, tree pruning, and watering) on the different phases of the truffle life cycle, such as the distribution of soil mycelium, is needed to anticipate the response of truffles to climate stresses.

In contrast to other ectomycorrhizal fungi (e.g. *Hebeloma cylindrosporum*) (Marmeisse et al., 2004), it is impossible to complete the truffle life cycle *in vitro*. Indeed, it is possible to obtain free-living mycelium of different *Tuber* spp., which grow slowly (Iotti, Amicucci, Stocchi, & Zambonelli, 2002), and ectomycorrhizae (Giomaro et al., 2002) can be produced, but to our knowledge, the sexual reproduction of *Tuber* has never been observed *in vitro*. The only documented production of ascocarps in control conditions was the fruiting of *T. maculatum* in vases associated with *Pinus strobus* (Fassi & Fontana, 1969). However, this was never replicated. The truffle life cycle is therefore challenging to study. Many unknowns regarding the basic biology and life cycle of truffles deserve further attention in order to increase truffle production. These include the following:

1. What factors promote and are responsible for spore germination?
2. How large are individual *Tuber* genets and what is their lifespan?
3. How can sexual reproduction of truffles be promoted?
4. How can we preserve and enhance ascocarp (fructification) growth?
5. Which is the role of microfauna and microbes on the truffle life cycle?
6. Are truffles able to adapt to climatic changes?
7. How can we enhance the truffle life cycle with management techniques?

A major advancement to our understanding of the truffle biology has come from the recently sequenced *T. melanosporum* genome (Martin et al., 2010). This effort has reaffirmed the importance of *T. melanosporum* as a model ectomycorrhizal ascomycete species. Molecular tools developed from this research are already filling many knowledge gaps, and more tools and insights pertaining to truffle biology and ecology are expected.

3.2. Intraspecific genetic diversity: From microsatellites to SNPs

The level of intraspecific genetic diversity of *T. melanosporum* has been debated since Bertault, Raymond, Berthomieu, Callot, and Fernandez (1998) argued that this species exhibits low levels of genetic diversity. More recently, Riccioni et al. (2008) used polymorphic microsatellites and questioned this conclusion. Thanks to the sequencing of the black truffle genome (Martin et al., 2010) and the development of highly polymorphic

microsatellite makers, we now have a better view of the intraspecific genetic diversity of this species, which has a genotypic diversity index close to its theoretical maximum (Murat et al., 2011). Several of the microsatellites developed detected more than 10 alleles and had a value of expected heterozygosity (i.e. estimating the fraction of all individuals who would be heterozygous for any randomly chosen locus) greater than 0.5 indicating that these SSRs are highly polymorphic. These polymorphic markers were used to characterize small-scale genetic diversity in two truffle orchards in Rollainville (Lorraine, France; 48°21'45" North and 5°44'23" East) and in Montemartano (Umbria, Italy; 42°44'00" North and 12 44'00" East) (Murat et al., 2013). In both truffle orchards, a pronounced spatial genetic structure was found with numerous small-sized genets. Indeed, the maximum distances between ramets of the same genet were 2.35 and 4.70 m in the Montemartano and the Rollainville truffle orchards, respectively. Most genets had a size of less than 1 m, and many were limited to one sampling core. In Rollainville, the sampling was carried out in a 60-m² area during two consecutive years and only three genets were shared out of the 10 and 13 genets found during the 2010–2011 and 2011–2012 seasons, respectively. One other interesting result was that the spatial distribution of the *T. melanosporum* genets on their hosts was nonrandom, resulting in field patches colonized by genets that shared the same mating types (see in the succeeding text). These results suggest that *Tuber* invests mainly in sexual reproduction. Indeed, it seems that the genet size depends more on competition and the time elapsed since the last disturbance than on other factors such as forest age (Douhan, Vincenot, Gryta, & Selosse, 2011). The management of truffle orchards (e.g. tree pruning, tillage, grass cutting, and searching for truffles with dogs) causes disturbance to create an open woodlands effect, and it is well known that *T. melanosporum* fruiting bodies are generally found in such ecosystems, while the presence of competing fungi in closed canopies depresses its fructification (Hall et al., 2007).

Using intersimple sequence repeat (ISSR) fingerprints, García-Cunchillos and colleagues (2014) confirmed high levels of genetic diversity in Spanish black truffle populations. They analysed 190 ascocarps from 23 natural populations distributed throughout Spain. Each ascocarp presented a unique genotype and Nei, Shannon, and Bayesian heterozygosity genetic diversity indices showed high levels of genetic diversity throughout the Iberian populations ($hN = 0.393$, $hS = 0.678$, and $HS = 0.418$). Interestingly, they identified strong genetic structure among the 23 Spanish populations as well as within geographic groups. These results confirm that southern

truffle populations in Spain include augmented levels' genetic diversity compared with northern truffle populations suggesting that the southern Iberian Peninsula may have served as a refuge for truffles during the last glaciation as proposed by [Riccioni et al. \(2008\)](#).

Classical population genetic analyses with microsatellites allowed the investigation on a very small proportion of the *Tuber* genome. With continual advances in high-throughput sequencing technologies, exhaustive investigation of the genomic variation is now feasible. Investigations of genetic variation in *Tuber* address questions pertaining to adaptation and genome plasticity of the species. Screening and identification of polymorphisms among strains of the same species is the most common way to identify genetic variation. Among all the polymorphisms, single nucleotide polymorphisms (SNPs) have gained much interest in the scientific community ([Ganal, Altmann, & Röder, 2009](#)). SNPs have been used for investigating the genetic origin of human diseases such as asthma and diabetes ([Moffatt et al., 2010](#); [Yamauchi et al., 2010](#)) and also in crop plants ([Ganal et al., 2009](#); [Rafalski, 2002](#)) to identify genomic regions involved in disease resistance in maize ([Poland, Bradbury, Buckler, & Nelson, 2011](#)). The identification of SNPs by genome resequencing was used for filamentous fungi (e.g. *Neurospora crassa* ([Ellison et al., 2011](#)), *Fusarium graminearum* ([Cuomo et al., 2007](#)), and *Coccidioides* spp. ([Neafsey et al., 2010](#)); see [Stukenbrock, 2014](#), in this book) but to our knowledge not yet for mycorrhizal species. To assess the overall genetic diversity of *T. melanosporum* and factors driving genome plasticity of this species the genome of six *T. melanosporum* genotypes were sequenced using Illumina technology and were compared to the reference genome ([Martin et al., 2010](#)). By comparing seven genomes, we identified more than 400,000 SNPs confirming that the genetic diversity of *T. melanosporum* is similar to that of other filamentous fungi (Payen Thibaut, Murat Claude, Gigant Anaïs, Morin Emmanuelle, De Mita Stéphane and Martin Francis, in preparation). SNPs were more frequent in repeated sequences although some SNPs were also identified in coding regions of more than 2000 genes. The generated SNPs provide a valuable resource for future population genetic and genomics studies. In this analysis, we also highlighted the role of transposable elements, which represents about 60% of the black truffle genome, in its genome plasticity. These first results are promising to investigations into black truffle adaptive capacities to climate change. A new sampling of isolates covering all European populations is currently underway to perform a robust and more comprehensive population genomic analysis.

3.3. Sexual reproduction: Where are both mating-type strains?

The interest of scientists in sexual reproduction of fungi dates back to the nineteenth century when Ehrenberg first reported sex in a fungal species (Ehrenberg, 1829). In filamentous ascomycetes, in addition to the widespread potential for asexual (vegetative) reproduction, two main types of sexual reproduction can be found: homothallism and heterothallism. Whereas homothallic fungi are self-fertile, heterothallic species depend on the presence of a partner of the opposite mating type to reproduce and are therefore self-sterile (Kronstad, 2007; Murtagh, Dyer, & Crittenden, 2000; Paoletti et al., 2007). Concerning mycorrhizal fungal species, in addition to their critical role in establishing and maintaining the complexity in agroforest ecosystems worldwide (Smith & Read, 2008), the production of edible fruiting bodies (e.g. boletes and truffles) by some of them makes the understanding of their sexual reproduction mode very attractive for mycologists.

In 1998, Bertault and colleagues concluded that *T. melanosporum* had “a very closed mating system, such as homothallism or even exclusive selfing”. This assertion was challenged by the data from microsatellites markers (Paolucci et al., 2006; Riccioni et al., 2008) and more recently by the identification of mating-type genes (Rubini et al., 2011, 2011). Microsatellite genotyping of ECM and asci, along with their surrounding sterile tissue in the ascocarps, the gleba, indicates that events of outcrossing are frequent in *T. melanosporum* (Riccioni et al., 2008). Conclusive evidence of sexual reproduction of *T. melanosporum* has been gained by looking at the structure and organization of the mating-type (*MAT*) locus in the sequenced genome of strain Mel28 (Martin et al., 2010). The *MAT* locus is a unique region of the fungal genome that governs the establishment of cell-type identity and orchestrates the sexual cycle (Fraser & Heitman, 2003). All known heterothallic ascomycetes have a single *MAT* locus with two alternative and dissimilar homologous sequences called idiomorphs (Metzenberg & Glass, 1990). One contains a *MAT* gene that encodes a protein with an α -box domain (*MAT1-1-1*), while the other contains an alternative *MAT* gene that encodes a high mobility group (HMG) protein (*MAT1-2-1*) (Debuchy, Berteaux-Lecellier, & Silar, 2010). Conversely, in homothallic filamentous ascomycetes, a single strain harbours both *MAT1-1-1* and *MAT1-2-1* genes. In Mel28 strain, only the *MAT1-2-1* gene was identified, whereas the second *MAT* gene was characterized in those strains that do not carry *MAT1-2-1*. Associated with outcrossing evidence produced with microsatellites, it was concluded that

T. melanosporum is a heterothallic species (Rubini, Belfiori, Riccioni, Arcioni, et al., 2011; Rubini, Belfiori, Riccioni, Tisserant, et al., 2011). Consequently, for *T. melanosporum* to complete its life cycle, the recognition of two truffle strains of opposite mating type is needed. Interestingly, both mating-type strains are not distributed randomly in the root systems of host plants in productive truffle orchards (Murat et al., 2013; Rubini, Belfiori, Riccioni, Arcioni, et al., 2011; Rubini, Belfiori, Riccioni, Tisserant, et al., 2011). Rather, patches colonized by strains of the same mating types are observed. This suggests that competition to colonize the root system may occur between the two mating types. Interestingly, both mating types can be detected in the soil in these patches. However, both mating types colonizing the same root systems were detected in some truffle orchards trees in Australia (Linde & Selmes, 2012). While the maternal strain producing the gleba can be detected in the surrounding ectomycorrhizae (Murat et al., 2013) and confirming the link between the ectomycorrhizae and ascocarps (Le Tacon et al., 2013), where does the paternal strain come from? To answer this question, studies are underway to identify the genotype of the paternal strain and investigate its distribution in the soil.

Recently, Belfiori, Riccioni, Paolocci, and Rubini (2013) identified mating-type genes in the Chinese truffle (*T. indicum*), a close relative to *T. melanosporum* (Fig. 7.1). Interestingly, they identified different sequences and the structure of the MAT idiomorphs between *T. indicum* strains belonging to different ITS classes suggested the presence of a complex of cryptic species. The mating-type idiomorphs of *T. indicum* and *T. melanosporum* are quite well conserved, and the sexual compatibility between the two species cannot be fully excluded. This is an important result since there is at least one documented case of the introduction of *T. indicum* into Italy (Murat et al., 2008). To limit possibilities of introgression between the two species, it is critical to avoid additional introductions of *T. indicum* into the natural range of *T. melanosporum*.

The characterization of mating-type genes identified by sequencing the black truffle genome is a very good example of a quick transfer from fundamental research to the truffle industry. Indeed, an international patent (no. WO2012/032098) was recently deposited for the characterization of the mating-type strains. This patent will soon be exploited to provide truffle industry with innovative tools. Indeed, promoting the initiation of sexual reproduction by truffle fungi is a critical point of interest, which should help to increase truffle production and yield. But as highlighted in the succeeding

text, the technical management (e.g. soil tilling, tree pruning, and watering) of truffle orchards is also important to allow ascocarps in the soil to develop and ripen over a duration of several months. This is particularly true in the context of climate change.



4. COMPARATIVE GENOMIC IN PEZIZOMYCETES

Comparative genomics allows for the assessment of relationships between genomes to understand the function and evolutionary processes acting upon them. The link between fungal life strategies (wood decay, symbiosis, and parasites) and genome structure (size, repeat sequences, and gene catalogue) has been a hot topic for several years (Martin et al., 2011). For example, biotrophic pathogens (Spanu et al., 2010), ectomycorrhizal symbionts (Martin et al., 2008, 2010), and endomycorrhizal fungi *Rhizophagus irregularis* (Tisserant et al., 2013) have a decreased repertoire of genes involved in the degradation of plant cell wall polysaccharides and in the biosynthesis of secondary metabolite toxins.

Until recently, *T. melanosporum* was the sole representative of the Pezizomycotina, a speciose and diverse clade of fungi, to have a genome sequence available (Martin et al., 2010). In 2013, Traeger and colleagues published the genome of *Pyronema confluens* a saprotrophic pezizomycete. Interestingly, *P. confluens* has a small genome of 50 Mb compared to the 125 Mbp genome of *T. melanosporum*. In order to gain information on the evolution in Pezizomycetes, and more precisely the Tuberaceae, the genome of new species is currently being sequenced. In the succeeding text, we will provide an update of the different genomic projects that are currently realized in Pezizomycetes.

4.1. Sequencing the genome of *Tuber* spp.

Genome sequencing of five other *Tuber* spp. is currently underway (Table 7.1). These species were chosen because they are economically important species and represent different clades within the *Tuber* phylogeny (Fig. 7.1). They include the following:

1. *T. aestivum*, known as the burgundy truffle, is harvested in all the European countries and is sold for between 100 and 400 €/kg. This species belongs to the Aestivum clade and is becoming popular due to its broad range of host preference, large range of soil types and climates that it can fruit, and availability as mycorrhized plants by nurseryman.

Table 7.1 List of the current genomic projects in the Pezizomycetes
% of

Species	Family	Genome size (Mb)	Genome repeated sequences	Material	Status of the project	Natural habitat	Project
<i>Ascobolus immersus</i>	Ascobolaceae	59.53	NA	Mycelium	Complete	Worldwide	1000 Fungal Genomes Project (JGI-DOE)
<i>Choitomyces venosus</i>	Tuberaceae	124	54	Fruiting body	Complete	Europe	Mycorrhizal Genomics Initiative (JGI-DOE)
<i>Pyronema confluens</i>	Pyrenomycetaceae	50	12	Mycelium	Published	France	Traeger et al. (2013)
<i>Terfezia boudieri</i>	Pezizaceae	63	NA	Mycelium	Pending	North Africa and southern Europe	Mycorrhizal Genomics Initiative (JGI-DOE)
<i>Tuber aestivum</i>	Tuberaceae	135.6	50	Fruiting body	Annotation pending	Europe	TuberEvolve (Genoscope-INRA)
<i>Tuber borchii</i>	Tuberaceae	NA	NA	Mycelium	Pending	Europe	Metatranscriptomics of Soil Forest Ecosystems (JGI-DOE)
<i>Tuber indicum</i>	Tuberaceae	NA	NA	Fruiting body	Pending	China	INRA/Kunming University
<i>Tuber lyonii</i>	Tuberaceae	NA	NA	Mycelium	Pending	North America	Duke University
<i>Tuber magnatum</i>	Tuberaceae	192.8	58	Fruiting body	Annotation pending	Italy and South-Central Europe	INRA—UNITO
<i>Tuber melanosporum</i>	Tuberaceae	124.95	58	Mycelium	Published	Southwest Europe	Martin et al. (2010)

NA, not available.

2. *T. borchii*, called the bianchetto, is harvested in many European countries but is most famous in Italy where it is commercialized and marketed for between 50 and 100 €/kg. This species belongs to the *Puberulum* clade and is characterized by its broad range of host associates, aggressive growth, and wide edaphic tolerances. *Tuber borchii* is now being cultivated on numerous continents.
3. *T. indicum*, known as the Chinese truffle, is phylogenetically and morphologically very close to *T. melanosporum* and belongs to the *Melanosporum* clade. It is mainly harvested in China but is mostly exported to Europe and the United States. This species has been inadvertently introduced into Italy (Murat et al., 2008) and the United States (Bonito et al., 2011). *T. indicum* is sold in Europe for between 20 and 100 €/kg.
4. *T. lyonii*, known as the pecan truffle, is native to North America. This species fruits naturally in pecan orchards and under oak trees in the southern and eastern United States but is found as far north as Quebec. This species has spiny spores and belongs to the *Rufum* clade. Recent efforts to cultivate this species are underway and the first fruitings from artificially colonized seedlings were produced this past year (*personal communication*, Tim Brenneman). *Tuber lyonii* sells for between \$200 and 300 per lb in the United States.
5. *T. magnatum*, known as the Piedmont white truffle, is harvested in Italy and central Europe (Croatia, Serbia, Slovenia, Romania, Bulgaria, and Hungary). It has also been found in southern France the past few years. This species has yet to be cultivated and is the most expensive truffle species sold at prices of 1000–3000 €/kg (but can reach more than 100,000 €/kg in auction sales).

The production of bulk mycelium for DNA isolation (or inoculation) is often difficult for *Tuber* spp. because they are fastidious to isolate and grow so slowly in pure culture. However, sufficient mycelium of *T. melanosporum* (G rard Chevalier—INRA Clermont-Ferrand), *T. borchii* (Alessandra Zambonelli and Mirco Iotti—University of Bologna), and *T. lyonii* (Gregory Bonito—Duke University) was grown to produce several grams of mycelium, enough to allow DNA and RNA to be extracted for genome and transcriptome sequencing. For genome sequencing of *T. indicum*, *T. aestivum*, and *T. magnatum*, nucleic acids were isolated from fresh young fruiting bodies. The gleba (internal part) of the truffle ascocarp is rich in microorganisms such as bacteria (Antony-Babu et al., 2014) and therefore extracted nucleic acids usually contain a mix belonging to *Tuber* and the

other microorganisms. When obtaining nucleic acids from fungal fruiting bodies, it is therefore necessary to exclude carry-over from associated bacteria prior to sequencing. In the first attempt to sequence the *T. magnatum* genome from a fruiting body, more than 90% of the 454 reads mapped to bacteria. More rigorous quality checks were necessary to avoid the huge bacterial contamination in further efforts. First, PCR with bacterial and fungal primers on serial dilutions of the DNA are done. Next, preliminary sequencing run is carried out with 454 GS Junior pyrosequencer to assess levels of bacterial contamination. Finally, samples deemed of sufficient quality for genome sequencing (i.e. with less than 10% of bacterial contamination) are sequenced and obtained reads are filtered using bacterial genomes as reference sequences to exclude bacterial reads.

Genome assemblies are currently available for *T. aestivum* and *T. magnatum* (Table 7.1). Interestingly, these two genomes are rather large with a size of 135 and 192 Mbp with 50% and 58% of repeated sequences, respectively. As for *T. melanosporum*, the gypsy-like retrotransposons are the most frequent repeat sequences in these two genomes. Preliminary analyses suggest that the same gypsy-like retrotransposon families are present in the three annotated *Tuber* genomes (*T. aestivum*, *T. magnatum*, and *T. melanosporum*) indicating that the invasion of these retrotransposons in *Tuber* genomes occurred early in the evolution of this genus. The structure of the *Tuber* and *Choiromyces venosus* (see in the succeeding text) genomes is conserved as shown also by large syntenic blocks (Fig. 7.2). The genome of *T. lyonii* has been sequenced but assemblies are not yet available. Comparisons between these four clades of *Tuber* should help address questions pertaining to host specificity, sexual and asexual reproduction, and development and growth of truffles.

4.2. Sequencing the genome of other Pezizomycetes

In the framework of the Mycorrhizal Genomics Initiative (Martin & Bonito, 2013; Martin et al., 2011), the genomes of two other truffles outside of the genus *Tuber* have been sequenced: *C. venosus* (<http://genome.jgi.doe.gov/Chove1/Chove1.home.html>) and *Terfezia boudieri* (<http://genome.jgi.doe.gov/Terbo1/Terbo1.home.html>). *Choiromyces venosus*, commonly called the pig truffle, belongs to the Tuberales and has a strong, distinctive odour. Gastronomic value of this whitish truffle varies throughout Europe, as it is consumed in a few countries (i.e. Germany, Hungary, and Sweden) but considered toxic in France and Italy. However, no records of *C. venosus*

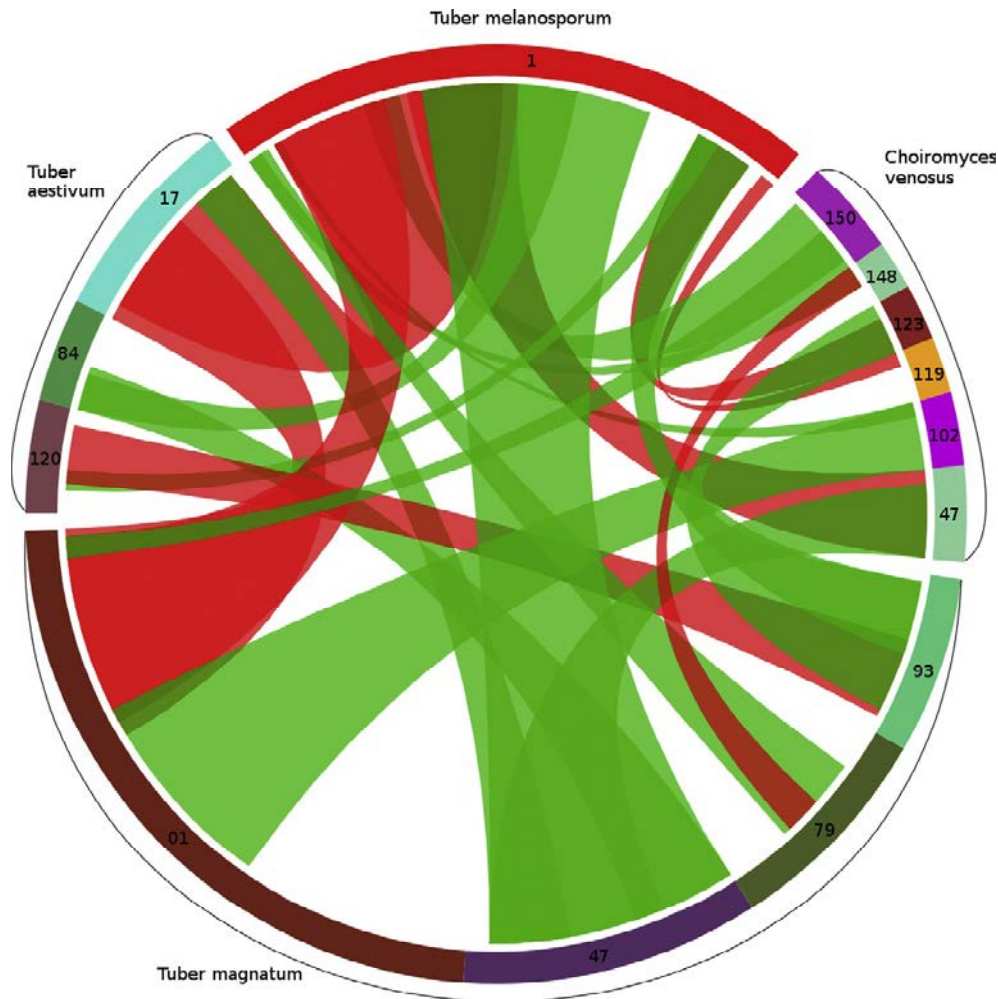


Figure 7.2 Example of synteny between the largest scaffold of *T. melanosporum* (scaffold 1) and its homologues in *T. aestivum*, *T. magnatum*, and *C. venosus*. The number of scaffold is indicated for each species. The synteny was realized using SyMAP version 4.0 (Soderlund, Bomhoff, & Nelson, 2011).

poisoning have ever been reported in northern Europe where it has consumed for more than a century (Wedén, Larsson, Burman, & Backlund, 2009). This symbiotic species is associated with deciduous and coniferous trees and has a preference for clay soils. The genome of *C. venosus* is similar to that of *Tuber* spp. in size (125 Mbp) and repeat content (54%) (Table 7.1). There are also large syntenic block between *C. venosus* and *Tuber* spp. (Fig. 7.2). The phylogenetic position of *Choiromyces* was for a time ambiguous since Percudani, Trevisi, Zambonelli, and Ottonello (1999) included this genus inside *Tuber* spp. However, more recent phylogenetic analyses clearly separate these two genera with strong statistical support (Fig. 7.1).

Terfezia boudieri, known as desert truffles, are harvested around the Mediterranean Sea and the Middle East in areas associated with *Helianthemum* spp. It belongs to the Pezizaceae and represents an independent origin of

the truffle lifestyle. Its genome has been sequenced. The comparison between *C. venosus*, *T. boudieri*, and *Tuber* spp. should provide new insights into the evolution of the truffle life strategy (e.g. hypogeous fruiting body formation).

Recently, through the “1000 Fungal Genomes Project” Initiative (JGI-DOE), the genome of *Ascobolus immersus*, belonging to the Ascobolaceae, has been sequenced. *Pyronema confluens* (Traeger et al., 2013), which belongs to the Pyronemataceae, has also been sequenced recently. Interestingly, two of these three last species have smaller-sized genomes than *Tuber* spp. and fewer repetitive elements (Table 7.1).



5. CONCLUDING REMARKS

The age of fungal genomics is well upon us. While mycorrhizal species such as *Tuber* may be fastidious to culture, various approaches have been made for obtaining sufficient quantities of nucleic acids for genome and transcriptome sequencing. These include growing cultures in dilute liquid media, across several agar plates, or by extracting nucleic acids directly from fresh young fruiting bodies. DNA from bacteria associated with truffle fruiting bodies can be problematic when obtaining genomic DNA from fruiting body materials; however, these bacteria are likely responsible for some of the organoleptic qualities of truffles and their development may be of interest to the study of the truffle microbiome. Several *Tuber* genomes are now sequenced including six species (*T. aestivum*, *T. borchii*, *T. indicum*, *T. lyonii*, *T. magnatum*, and *T. melanosporum*) and representing four of the 11 clades in *Tuber*. When fully assembled, these data should help resolve questions pertaining to truffle growth and development, ecological adaptability, origin, and evolution. Tools arising from these genomic projects should empower a new generation of truffle growers and researchers with the ability to simultaneously screen planted materials for contamination, diseases, genetic diversity, and origin of target species using high-throughput sequencing approaches. In addition, genomes of other Pezizomycetes are being sequenced and will offer new perspectives on the evolution of this particular class for which species with different life history strategies exist. We are convinced that this will lead to new insights and knowledge through the analysis of fruiting body development (hypogeous vs. epigeous), symbiosis-related genes, molecular crosstalk between fungus and host, the impact of repetitive elements on genome organization, and the consequence of intrahyphal and intraascocarp bacteria on fungal growth, function, and development.

ACKNOWLEDGEMENTS

We are grateful to Francis Martin and the Mycorrhizal Genomics Initiative consortium for access to unpublished genome data. Sequence data of *Choiromyces venosus* were produced by the U.S. Department of Energy Joint Genome Institute in collaboration with the user community. Sequence data of *Tuber aestivum* were produced by the Genoscope. Our research was financed by the “Institut National de la Recherche Agronomique” (INRA), the Région de Lorraine, the lab of excellence ARBRE (ANR-11-LABX-0002-01), and the ANR SYSTERRA SYSTRUF (ANR-09-STRA-10). The computations were performed at the INRA Nancy Ecogenomics facilities. Part of our research was sponsored by the Genomic Science Program of the U.S. Department of Energy, Office of Biological and Environmental Research (under contract DE-AC05-00OR22725).

REFERENCES

- Alvarado, P., Moreno, G., & Manjón, J. L. (2012). Comparison between *Tuber gennadii* and *T. oligospermum* lineages reveals the existence of the new species *T. cistophilum* (Tuberaceae, Pezizales). *Mycologia*, *104*(4), 894–910.
- Amicucci, A., Balestrini, R., Kohler, A., Barbieri, E., Saltarelli, R., Faccio, A., et al. (2011). Hyphal and cytoskeleton polarization in *Tuber melanosporum*: A genomic and cellular analysis. *Fungal Genetics and Biology*, *48*(6), 561–572.
- Amicucci, A., Guidi, C., Zambonelli, A., Potenza, L., & Stocchi, V. (2000). Multiplex PCR for the identification of white *Tuber* species. *FEMS Microbiology Letters*, *189*(2), 265–269.
- Antony-Babu, S., Deveau, A., Van Nostrand, J. D., Zhou, J., Le Tacon, F., Robin, C., et al. (2014). Black truffle-associated bacterial communities during the development and maturation of *Tuber melanosporum* ascocarps and putative functional roles. *Environmental Microbiology*. <http://dx.doi.org/10.1111/1462-2920.12294>.
- Belfiori, B., Riccioni, C., Paolocci, F., & Rubini, A. (2013). Mating type locus of Chinese black truffles reveals heterothallism and the presence of cryptic species within the *T. indicum* species complex. *PLoS One*, *8*(12), e82353.
- Bertault, G., Raymond, M., Berthomieu, A., Callot, G., & Fernandez, D. (1998). Trifling variation in truffles. *Nature*, *394*(6695), 734.
- Bolchi, A., Ruotolo, R., Marchini, G., Vurro, E., di Toppi, L. S., Kohler, A., et al. (2011). Genome-wide inventory of metal homeostasis-related gene products including a functional phytochelatin synthase in the hypogeous mycorrhizal fungus *Tuber melanosporum*. *Fungal Genetics and Biology*, *48*(6), 573–584.
- Bonito, G. (2009). Fast DNA-based identification of the black truffle *Tuber melanosporum* with direct PCR and species-specific primers. *FEMS Microbiology Letters*, *301*(2), 171–175.
- Bonito, G. M., Gryganskyi, A. P., Trappe, J. M., & Vilgalys, R. (2010). A global meta-analysis of *Tuber* ITS rDNA sequences: Species diversity, host associations and long-distance dispersal. *Molecular Ecology*, *19*(22), 4994–5008.
- Bonito, G., Smith, M. E., Nowak, M., Healy, R. A., Guevara, G., Cázares, E., et al. (2013). Historical biogeography and diversification of truffles in the tuberaceae and their newly identified southern hemisphere sister lineage. *PloS One*, *8*(1), e52765.
- Bonito, G., Trappe, J. M., Donovan, S., & Vilgalys, R. (2011). The Asian black truffle *Tuber indicum* can form ectomycorrhizas with North American host plants and complete its life cycle in non-native soils. *Fungal Ecology*, *4*(1), 83–93.
- Bonito, G., Trappe, J. M., Rawlinson, P., & Vilgalys, R. (2010). Improved resolution of major clades within *Tuber* and taxonomy of species within the *Tuber gibbosum* complex. *Mycologia*, *102*(5), 1042–1057.

- Božac, R., Širić, I., & Kos, I. (2012). Tuber decipiens: A new black Tuber species from Croatia. *Periodicum Biologorum*, 114(1), 125–128.
- Büntgen, U., Egli, S., Camarero, J. J., Fischer, E. M., Stobbe, U., Kausserud, H., et al. (2012). Drought-induced decline in Mediterranean truffle harvest. *Nature Climate Change*, 2(12), 827–829.
- Chevalier, G. (1998). The truffle cultivation in France: Assessment of the situation after 25 years of intensive use of mycorrhizal seedlings. In *Proceedings from the first international meeting on ecology, physiology and cultivation of edible mycorrhizal mushrooms* (pp. 3–4), Uppsala, Sweden.
- Chevalier, G., & Grente, J. (1979). Application pratique de la symbiose ectomycorhizienne: Production à grande échelle de plants mycorrhizés par la truffe (*Tuber melanosporum* Vitt.). *Mushroom Science*, 10, 483–505.
- Cuomo, C. A., Güldener, U., Xu, J.-R., Trail, F., Turgeon, B. G., Di Pietro, A., et al. (2007). The *Fusarium graminearum* genome reveals a link between localized polymorphism and pathogen specialization. *Science*, 317(5843), 1400–1402.
- Debuchy, R., Berteaux-Lecellier, V., & Silar, P. (2010). Mating systems and sexual morphogenesis in ascomycetes. In K. A. Borkovich & D. J. Ebbole (Eds.), *Cellular and molecular biology of filamentous fungi* (pp. 501–536). Washington: ASM Press. Retrieved from, <http://hal.archives-ouvertes.fr/hal-00480603>.
- Deng, X. J., Liu, P. G., Liu, C. Y., & Wang, Y. (2013). A new white truffle species, *Tuber panzhihuanense* from China. *Mycological Progress*, 12, 557–561.
- Douhan, G. W., Vincenot, L., Gryta, H., & Selosse, M.-A. (2011). Population genetics of ectomycorrhizal fungi: From current knowledge to emerging directions. *Fungal Biology*, 115(7), 569–597.
- Ehrenberg, C. G. (1829). Syzygites, eine neue Schimmeligattung: Nebst Beobachtungen über sichtbare Bewegung in Schimmeln. *Verhandlungen der Gesellschaft naturforschender Freunde zu Berlin*, 1, 98–109.
- Ellison, C. E., Hall, C., Kowbel, D., Welch, J., Brem, R. B., Glass, N. L., 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 United States of America*, 108(7), 2831–2836.
- Fan, L., Cao, J.-Z., & Li, Y. (2012a). *Tuber microsphaerosporum* and *Paradoxa sinensis* spp. nov. *Mycotaxon*, 120(1), 471–475.
- Fan, L., Cao, J.-Z., & Li, Y. (2012b). *Tuber sinosphaerosporum* sp. nov. from China. *Mycotaxon*, 122(1), 347–353.
- Fan, L., Cao, J.-Z., & Yu, J. (2012). Tuber in China: *T. sinopuberulum* and *T. vesicoperidium* spp. nov. *Mycotaxon*, 121(1), 255–263.
- Fan, L., Hou, C.-L., & Li, Y. (2012). Tuber microverrucosum and *T. huizeanum* two new species from China with reticulate ascospores. *Mycotaxon*, 122(1), 161–169.
- Fassi, B., & Fontana, A. (1969). Sintesi micorrizica tra *Pinus strobus* e *Tuber maculatum*. II. Sviluppo dei semenzali trapianti e produzione di ascocarpi. *Allionia*, 15, 115–120.
- Frank, B. (1889). Über die auf Wurzelsymbiose beruhende Ernährung gewisser Bäume durch unterirdische Pilze. *Berichte der Deutschen Botanischen Gesellschaft*, 3, 128–145.
- Fraser, J. A., & Heitman, J. (2003). Fungal mating-type loci. *Current Biology*, 13(20), R792–R795.
- Ganal, M. W., Altmann, T., & Röder, M. S. (2009). SNP identification in crop plants. *Current Opinion in Plant Biology*, 12(2), 211–217.
- García-Cunchillos, I., Sánchez, S., Barriuso, J. J., & Pérez-Collazos, E. (2014). Population genetics of the westernmost distribution of the glaciations-surviving black truffle *Tuber melanosporum*. *Mycorrhiza*, 24(1), 89–100.
- Giomaro, G., Sisti, D., Zambonelli, A., Amicucci, A., Cecchini, M., Comandini, O., et al. (2002). Comparative study and molecular characterization of ectomycorrhizas in *Tilia americana* and *Quercus pubescens* with *Tuber brumale*. *FEMS Microbiology Letters*, 216(1), 9–14.

- Guerin-Laguette, A., Cummings, N., Hesom-Williams, N., Butler, R., & Wang, Y. (2013). Mycorrhiza analyses in New Zealand truffières reveal frequent but variable persistence of *Tuber melanosporum* in co-existence with other truffle species. *Mycorrhiza*, 23(2), 87–98.
- Guevara, G., Bonito, G., Trappe, J. M., Cázares, E., Williams, G., Healy, R. A., et al. (2013). New North American truffles (*Tuber* spp.) and their ectomycorrhizal associations. *Mycologia*, 105(1), 194–209.
- Hall, I. R., Brown, G. T., & Zambonelli, A. (2007). In *Taming the truffle: The history, lore, and science of the ultimate mushroom: Vol. 304*. Oregon: Timber Press Portland.
- Iotti, M., Amicucci, A., Stocchi, V., & Zambonelli, A. (2002). Morphological and molecular characterization of mycelia of some *Tuber* species in pure culture. *New Phytologist*, 155(3), 499–505.
- Jeandroz, S., Murat, C., Wang, Y., Bonfante, P., & Tacon, F. L. (2008). Molecular phylogeny and historical biogeography of the genus *Tuber*, the “true truffles” *Journal of Biogeography*, 35(5), 815–829.
- Kohler, A., & Tisserant, E. (2014). Exploring the transcriptome of mycorrhizal interactions. In F. Martin (Ed.), *Genomics of plant-related fungi. Advances in Botanical Research*, 70, 23–78.
- Kronstad, J. W. (2007). Self-fertility: The genetics of sex in lonely fungi. *Current Biology*, 17(19), R843–R845.
- Kües, U., & Martin, F. (2011). On the road to understanding truffles in the underground. *Fungal Genetics and Biology*, 48(6), 555–560.
- Le Tacon, F., Delmas, J., Gleyze, R., & Bouchard, D. (1982). Influence du regime hydrique du sol et de la fertilisation sur la fructification de la truffe noire du Périgord (*Tuber melanosporum* Vitt.) dans le Sud-Est de la France [Vaucluse; irrigation, paillage]. *Acta Oecologia Oecologia Applicata*, 3, 291–306.
- Le Tacon, F., Zeller, B., Plain, C., Hossann, C., Bréchet, C., & Robin, C. (2013). Carbon transfer from the host to *Tuber melanosporum* mycorrhizas and ascocarps followed using a ¹³C pulse-labeling technique. *PLoS One*, 8(5), e64626.
- Le Tacon, F., Marçais, B., Courvoisier, M., Murat, C., Montpied, P., & Becker, M. (2014). Climatic variations explain annual fluctuations in French Périgord black truffle wholesale markets but do not explain the decrease in black truffle production over the last 48 years. *Mycorrhiza*, 24(1), 115–125.
- Linde, C. C., & Selmes, H. (2012). Genetic diversity and mating type distribution of *Tuber melanosporum* and their significance to truffle cultivation in artificially planted truffières in Australia. *Applied and Environmental Microbiology*, 78(18), 6534–6539.
- Marmeisse, R., Guidot, A., Gay, G., Lambilliotte, R., Sentenac, H., Combiér, J.-P., et al. (2004). *Hebeloma cylindrosporum*—a model species to study ectomycorrhizal symbiosis from gene to ecosystem. *New Phytologist*, 163(3), 481–498.
- Martin, F., Aerts, A., Ahrén, D., Brun, A., Danchin, E. G. J., Duchaussoy, F., et al. (2008). The genome of *Laccaria bicolor* provides insights into mycorrhizal symbiosis. *Nature*, 452(7183), 88–92.
- Martin, F., & Bonito, G. M. (2013). Ten years of genomics for ectomycorrhizal fungi: What have we achieved and where are we heading? In A. Zambonelli & B. Bonito (Eds.), *Soil Biology Series. Edible ectomycorrhizal mushrooms*. Berlin: Springer-Verlag.
- Martin, F., Cullen, D., Hibbett, D., Pisabarro, A., Spatafora, J. W., Baker, S. E., et al. (2011). Sequencing the fungal tree of life. *New Phytologist*, 190, 818–821.
- Martin, F., Kohler, A., Murat, C., Balestrini, R., Coutinho, P. M., Jaillon, O., et al. (2010). Périgord black truffle genome uncovers evolutionary origins and mechanisms of symbiosis. *Nature*, 464(7291), 1033–1038.
- Metzenberg, R. L., & Glass, N. L. (1990). Mating type and mating strategies in *Neurospora*. *Bioessays*, 12(2), 53–59.

- Moffatt, M. F., Gut, I. G., Demenais, F., Strachan, D. P., Bouzigon, E., Heath, S., et al. (2010). A large-scale, consortium-based genomewide association study of asthma. *New England Journal of Medicine*, 363(13), 1211–1221.
- Montanini, B., Levati, E., Bolchi, A., Kohler, A., Morin, E., Tisserant, E., et al. (2011). Genome-wide search and functional identification of transcription factors in the mycorrhizal fungus *Tuber melanosporum*. *New Phytologist*, 189(3), 736–750.
- Murat, C., Riccioni, C., Belfiori, B., Cichocki, N., Labbé, J., Morin, E., et al. (2011). Distribution and localization of microsatellites in the Perigord black truffle genome and identification of new molecular markers. *Fungal Genetics and Biology*, 48(6), 592–601.
- Murat, C., Rubini, A., Riccioni, C., De la Varga, H., Akroume, E., Belfiori, B., et al. (2013). Fine-scale spatial genetic structure of the black truffle (*Tuber melanosporum*) investigated with neutral microsatellites and functional mating type genes. *New Phytologist*, 199(1), 176–187.
- Murat, C., Zampieri, E., Vizzini, A., & Bonfante, P. (2008). Is the Perigord black truffle threatened by an invasive species? We dreaded it and it has happened! *New Phytologist*, 178(4), 699–702.
- Murtagh, G. J., Dyer, P. S., & Crittenden, P. D. (2000). Reproductive systems: Sex and the single lichen. *Nature*, 404(6778), 564.
- Neafsey, D. E., Barker, B. M., Sharpton, T. J., Stajich, J. E., Park, D. J., Whiston, E., et al. (2010). Population genomic sequencing of *Coccidioides* fungi reveals recent hybridization and transposon control. *Genome Research*, 20(7), 938–946.
- Paoletti, M., Seymour, F. A., Alcocer, M. J. C., Kaur, N., Calvo, A. M., Archer, D. B., et al. (2007). Mating type and the genetic basis of self-fertility in the model fungus *Aspergillus nidulans*. *Current Biology*, 17(16), 1384–1389.
- Paolocci, F., Rubini, A., Granetti, B., & Arcioni, S. (1999). Rapid molecular approach for a reliable identification of *Tuber* spp. ectomycorrhizae. *FEMS Microbiology Ecology*, 28(1), 23–30.
- Paolocci, F., Rubini, A., Riccioni, C., & Arcioni, S. (2006). Reevaluation of the life cycle of *Tuber magnatum*. *Applied and Environmental Microbiology*, 72(4), 2390–2393.
- Percudani, R., Trevisi, A., Zambonelli, A., & Ottonello, S. (1999). Molecular phylogeny of truffles (Pezizales: Terfeziaceae, Tuberaceae) derived from nuclear rDNA sequence analysis. *Molecular Phylogenetics and Evolution*, 13(1), 169–180.
- Poland, J. A., Bradbury, P. J., Buckler, E. S., & Nelson, R. J. (2011). Genome-wide nested association mapping of quantitative resistance to northern leaf blight in maize. *Proceedings of the National Academy of Sciences of the United States of America*, 108(17), 6893–6898.
- Rafalski, A. (2002). Applications of single nucleotide polymorphisms in crop genetics. *Current Opinion in Plant Biology*, 5(2), 94–100.
- Riccioni, C., Belfiori, B., Rubini, A., Passeri, V., Arcioni, S., & Paolocci, F. (2008). *Tuber melanosporum* outcrosses: Analysis of the genetic diversity within and among its natural populations under this new scenario. *New Phytologist*, 180(2), 466–478.
- Rubini, A., Belfiori, B., Riccioni, C., Arcioni, S., Martin, F., & Paolocci, F. (2011). *Tuber melanosporum*: Mating type distribution in a natural plantation and dynamics of strains of different mating types on the roots of nursery-inoculated host plants. *New Phytologist*, 189(3), 723–735.
- Rubini, A., Belfiori, B., Riccioni, C., Tisserant, E., Arcioni, S., Martin, F., et al. (2011). Isolation and characterization of MAT genes in the symbiotic ascomycete *Tuber melanosporum*. *New Phytologist*, 189(3), 710–722.
- Sheffield, J., Wood, E. F., & Roderick, M. L. (2012). Little change in global drought over the past 60 years. *Nature*, 491(7424), 435–438.
- Smith, S. E., & Read, D. J. (2008). *Mycorrhizal symbiosis* (3rd ed.). San Diego, CA: Academic Press.

- Soderlund, C., Bomhoff, M., & Nelson, W. M. (2011). SyMAP v3.4: A turnkey synteny system with application to plant genomes. *Nucleic Acids Research*, *39*(10), e68. <http://dx.doi.org/10.1093/nar/gkr123>.
- Spanu, P. D., Abbott, J. C., Amselem, J., Burgis, T. A., Soanes, D. M., Stüber, K., et al. (2010). Genome expansion and gene loss in powdery mildew fungi reveal tradeoffs in extreme parasitism. *Science*, *330*(6010), 1543–1546.
- Splivallo, R., Ottonello, S., Mello, A., & Karlovsky, P. (2011). Truffle volatiles: From chemical ecology to aroma biosynthesis. *New Phytologist*, *189*(3), 688–699.
- Stukenbrock, E. H. (2014). Speciation genomics of fungal plant pathogens. In F. Martin (Ed.), *Genomics of plant-related fungi. Advances in Botanical Research*, *70*, 397–423.
- Traeger, S., Altegoer, F., Freitag, M., Gabaldon, T., Kempken, F., et al. (2013). The genome and development-dependent transcriptomes of *Pyronema confluens*: A window into fungal evolution. *PLoS Genet*, *9*(9), e1003820. <http://dx.doi.org/10.1371/journal.pgen.1003820>.
- Tisserant, E., Malbreil, M., Kuo, A., Kohler, A., Symeonidi, A., Balestrini, R., et al. (2013). Genome of an arbuscular mycorrhizal fungus provides insight into the oldest plant symbiosis. *Proceedings of the National Academy of Sciences of the United States of America*. <http://dx.doi.org/10.1073/pnas.1313452110>.
- Wedén, C., Larsson, S., Burman, R., & Backlund, A. (2009). The edible truffle *Choiromyces venosus* and its use in Sweden. *Acta Botanica Yunnanica*, *31*(S16), 94–96.
- Yamauchi, T., Hara, K., Maeda, S., Yasuda, K., Takahashi, A., Horikoshi, M., et al. (2010). A genome-wide association study in the Japanese population identifies susceptibility loci for type 2 diabetes at UBE2E2 and C2CD4A–C2CD4B. *Nature Genetics*, *42*(10), 864–868.
- Zampieri, E., Balestrini, R., Kohler, A., Abbà, S., Martin, F., & Bonfante, P. (2011). The Perigord black truffle responds to cold temperature with an extensive reprogramming of its transcriptional activity. *Fungal Genetics and Biology*, *48*(6), 585–591.
- Zampieri, E., Mello, A., Bonfante, P., & Murat, C. (2009). PCR primers specific for the genus *Tuber* reveal the presence of several truffle species in a truffle-ground. *FEMS Microbiology Letters*, *297*(1), 67–72.

4 Etude de génomique comparative chez les truffes et autres Pézizomycètes

Dans ce paragraphe, j'ai comparé les génomes de trois truffes du genre *Tuber* à ceux de cinq autres Pézizomycètes dans un article en préparation pour la revue PNAS et intitulé « **The genome of truffles provide insights into the early diverging Ascomycetes evolution** ».

La classe des Pézizomycètes est composée de 200 genres et de 1683 espèces décrites. Parmi ces espèces, on retrouve des saprotrophes, des pathogènes et des symbiotiques ainsi que des champignons très connus comme les truffes et les morilles. Malgré leur intérêt, la structure et l'évolution des génomes des champignons de la classe des Pézizomycètes est encore mal connue. Le génome de la truffe noire *Tuber melanosporum* a été le premier génome de Pézizomycète à avoir été séquencé. Ce génome a comme caractéristiques une grande taille, une abondance en séquences répétées et un répertoire de gènes limité. Ces caractéristiques n'ont pas été retrouvées chez un autre Pézizomycète (*Pyronema confluens*) séquencé en 2013. Les objectifs de cette étude étaient de caractériser la structure des génomes de truffes et d'autres Pézizomycètes. Pour mener à bien ce projet, huit génomes de Pézizomycètes, dont trois de truffes, ont été comparés. Cette étude nous a permis d'élaborer un modèle pour l'évolution des génomes au sein des Tuberaceae qui ont subi une expansion de leur taille due aux éléments transposables et pour les *Tuber* une réduction de leur répertoire de gènes principalement concernant les enzymes dégradant la matière organique végétale (PCWDE). Au contraire, *Choiromyces venosus* un autre champignon appartenant aux Tuberaceae, a subi une expansion principalement de gènes orphelins. D'autre part, les champignons ectomycorhiziens du genre *Tuber* ont des convergences évolutives avec certains champignons ectomycorhiziens Basidiomycètes concernant principalement leurs capacités saprotrophiques, comme la perte d'un certain nombre de PCWDE. Toutefois, des divergences ont été trouvées entre les champignons ectomycorhiziens basidiomycètes et les truffes concernant les petites

protéines sécrétées sur exprimées dans la symbiose (MiSSP). Chez *T. melanosporum*, aucune MiSSP n'est très fortement surexprimée dans l'ectomycorhize et aucune des MiSSPs présentes n'est spécifique à *T. melanosporum*. Cette étude a permis de mieux connaître la structure des génomes chez les truffes et d'élaborer un modèle d'évolution de leurs génomes.

The genome of truffles provide new insights into the early diverging Ascomycetes evolution

Abstract

Truffles are fungi belonging to Pezizomycete class and forming ectomycorrhizal symbiosis with tree and shrubs. The Pezizomycetes constitute an early diverging lineage of Ascomycota composed of saprophytic, mycorrhizal and pathogens species found in soils, wood decays, leaves, roots and dung. Despite their interest, to date, only the genomes of *Tuber melanosporum* and *Pyronema confluens* have been published. Here we introduced the 145 Mbp and 192 Mbp haploid genomes of *T. aestivum* and *T. magnatum* as well as the genome of four additional Pezizomycetes: *Ascobolus immersus*, *Choiromyces venosus*, *Morchella conica* and *Terfezia boudieri*. The genome of truffles evolved by a size expansion due repeated sequences in the common ancestor of Tuberaceae about 150 Mya and a loss of the number of genes in the common ancestor of *Tuber* spp about 100 Mya. Large syntenic blocks were identified among *Tuber* spp, although the synteny is partially lost when the common ancestor is more ancient. Ectomycorrhizal Pezizomycetes (e.g., *Tuber* spp) have a reduced amount of genes encoding plant cell wall degrading enzymes (PCWDE), compared to saprotrophic Pezizomycetes (e.g. *A. immersus*). Nevertheless, *Tuber* species have retained a unique array of PCWDE, including a glycosyl hydrolase cellulase (GH6), lacking in basidiomycetes ECMs, suggesting they may possess limited ligninolytic capabilities. This study provides new insights in truffle and ECM symbiosis evolution, highlighting divergence and convergence between Basidiomycota and Ascomycota symbiotic species.

Keywords: Truffle, Pezizomycetes, comparative genomics, CAZymes

Significance

The fungi in the Pezizomycetes constitute an early diverging lineage of Ascomycota comprising saprophytic, mycorrhizal and pathogenic lifestyles. Pezizomycetes are characterized by the formation of apothecia containing operculate asci with forcible or passive spore discharge for epigeous or hypogeous species called truffles, respectively. Truffle lifestyle evolved several times from epigeous ascocarps during the evolution. Comparative genomics of Pezizomycetes aims to investigate the different fungal life strategies, the development of hypogeous versus epigeous ascocarps and the evolution of the particular organoleptic qualities of some species (e.g. truffles and morels). The present study highlights the evolutionary convergence (e.g., the reduction of plant cell wall enzymes and toxin-synthesizing genes) and divergences (e.g., lack of species-specific mycorrhizal induced small secreted proteins) between ectomycorrhizal truffles and basidiomycetes. It also provides novel insights on the truffle life cycle and Pezizomycetes evolution.

Introduction

The Pezizomycetes constitutes an early diverging lineage in Ascomycota (1) represented by 200 genus and 1,683 described species (2) either saprophytic, mycorrhizal or pathogens. This class is characterized by the formation of apothecia containing operculate asci with forcible spore discharge (3) although other forms exist such as hypogeous fungi. Fungi having subterranean sporocarps, called truffle, comprise many taxonomically unrelated species showing remarkable features of convergent evolution as a result of adaptation to this specialized habitat. Truffle ascomata contain asci with passive spore dispersal, which have multiple evolutionary origins in Pezizomycetes (3–5). They probably derive from epigeous fruiting bodies which have evolved as an adaptation to animal gazing or water stress (6, 7). Some Pezizomycetes species are well

known due to their edible fructifications such as truffles and morels. Truffles belonging to the genus *Tuber* (i.e. true truffles) form ectomycorrhizal symbiosis with trees and shrubs although those belonging to genus *Terfezia* (i.e. desert truffles) could form endomycorrhizal, ectendomycorrhiza and ectomycorrhizal depending on the nutrient condition (8). Mycorrhizal fungi are key players in forest and woodland ecosystems since they promote the growth of trees by helping them for nutrient uptake (9). True truffles are naturally distributed worldwide in the northern hemisphere with 180 estimated species (10). Some species such as the Burgundy truffle (*T. aestivum*), the Piémont white truffle (*T. magnatum*) and the Périgord black truffle (*T. melanosporum*) are among the most fungal delicacies due to the organoleptic quality of their fruiting bodies, also called ascocarps.

Despite the interest of Pezizomycetes for investigating the evolution of Fungi according to their life strategy, the development of hypogeous versus epigeous fruiting body formation and the evolution of the particular organoleptic qualities of some species (e.g. truffles and morels), to date, the genomes of only two Pezizomycetes have been published: *T. melanosporum* (11) and *Pyronema confluens* (12). The *T. melanosporum* genome sequencing (11) provided new tools and knowledge useful for truffle industry such as the mating type genes (13) and microsatellites markers (14). Recently, the resequencing of six *T. melanosporum* geographic accessions provided a resource of more than 500,000 single nucleotide polymorphisms and identified genomic regions putatively under selection (15). *T. melanosporum* has an atypical genome for a filamentous Ascomycete due to its large size (125 Mbp), its high proportion of repeated sequences (58%) and its low number of protein-coding genes (7,496 genes; (11)). These features are not shared by *P. confluens* that has a smaller genome (50 Mbp) and higher number of gene models (13,367 genes; (12)). The study of the evolution of truffle genomes

could therefore provide new insights on their ectomycorrhizal life style, the formation of their hypogeous fruiting body and the organoleptic qualities.

Here we introduced the *T. aestivum* and *T. magnatum* haploid genome assemblies and annotation of gene families, and compared them with the genome of *T. melanosporum*. In order to investigate the specificity of truffle genomes we compared them with *P. confluens* genome (12) and introduced also the genomes of another Tuberaceae (*Choiromyces venosus*) and three additional Pezizomycetes (*Ascobolus immersus*, *Morchella conica* and *Terfezia boudieri*). We focused on the genome structure, synteny, gene families and plant cell wall degrading enzymes (PCWDE) highlighting the specificity of truffle genomes.

Results and discussion

Phylogenomic

The ML tree obtained with the 2,338 gene families was consistent with the known phylogenies of Pezizomycetes (Fig. 1; (4, 16, 17)). Although the phylogenetic position of *Choiromyces* was sometimes included inside *Tuber* spp. (16), we confirmed *Choiromyces* is phylogenetically distinct from *Tuber* spp (Fig. 1; (17)). The Pezizomycete most recent common ancestor (MRCA) was estimated to 403 Mya to 537 Mya, this estimation is in accordance with previous estimations (18) confirming that Pezizomycete class is an early divergent class of Ascomycota. In the *Tuber* genus, our estimation of the divergence time are lower than previous analyses (17, 19): e.g. *T. aestivum*/*T. magnatum* MRCA is estimated between 33 Mya and 44 Mya although it was estimated between 57 Mya and 137 Mya by Jeandroz and colleagues (19) and between 53.1 Mya and 87.3 Mya by Bonito and colleagues (17). The interval is also lower in our study and this could be explained by the fact that our

phylogeny is based on 2,338 gene families where previous molecular clock analyses were realized with only a few genes.

Genome assembly, protein coding repertoires and synteny

To investigate the gene repertoires and the genome variability of truffles, we sequenced the genomes of *T. aestivum* and *T. magnatum*. The haploid genome size was 145 Megabases (Mbp) and 192 Mbp and the coding space 94% and 96% complete on the basis of conserved core eukaryotic single-copy genes (20) for *T. aestivum* and *T. magnatum*, respectively (Fig. 1; Table S1). Both genomes are larger than those of *T. melanosporum* (125 Mbp; (11)). The number of predicted gene model is comparable for all three *Tuber* spp from 9,344 in *T. aestivum* to 10,763 for *T. melanosporum* (Fig. 1; Table S1). The genome of four additional species of Pezizomycetes were produced showing important genome sizes variation among the Pezizomycetes (Fig. 1 and Table S1), from 48 Mbp in *M. conica* to 192 Mbp in *T. magnatum*. The number of protein coding genes is also highly variable from 9,344 in *T. aestivum* and 17,986 in *C. venosus* (Fig. 1 and Table S1). Interestingly, *C. venosus* presented a high number of gene models compared to *Tuber* spp, although this genus, belonging to Tuberaceae, is phylogenetically close to genus *Tuber* (Fig. 1; 11, 12). We may propose that the loss of numerous gene models occurred in the common ancestor of the three *Tuber* spp about 100 Mya and not in the common ancestor of Tuberaceae (Fig. 1 and Fig. S1). The OrthoMCL analysis showed that 3,916 gene families, composed between 4,163 and 4,349 genes for *T. aestivum* and *C. venosus* respectively, have at least one member in each of the eight Pezizomycetes genomes corresponding to the Pezizomycete core-genes (Fig. S2). The number of gene models with orthologs in all the three *Tuber* spp. and without orthologs in the other Pezizomycetes species is low (i.e. 112 to 139 for *T. aestivum* and *T. magnatum*, respectively) (Fig. S2). The same result is found for those with orthologs in all the four

Tuberaceae species and without orthologs in the other Pezizomycetes (i.e. 58 to 127 for *T. magnatum* and *C. venosus*, respectively) (Fig. S2). Most speciation events were associated to decrease in the number of gene families to contracting families with the exception of *C. venosus* which have more than twice as much families in expansion than in contraction (Fig. S3). The number of orphan gene models, i.e. gene models without orthologs in all the other Pezizomycetes species, varied from 5 % to 62 % of the gene models for *T. magnatum* and *A. immersus* respectively (Fig. S4). The high number of gene models for both *C. venosus* and *A. immersus* is therefore in part explained by a high proportion of orphan gene models (Fig. 1; Table S1).

At the genome scale, there is general sequence conservation among Pezizomycete genomes (Fig. S5). Using orthology relation, 144 two-genes blocks of synteny are observed between *A. immersus* and *P. confluens*, to 702 between *C. venosus* and *T. aestivum* for all pairs of species (Fig. 3C, S6). Larger blocks of synteny are observed between Tuberaceae (Fig. 3A, S6) up to 21 gene models, most without KOG definition, between *T. magnatum* and *T. melanosporum* (Fig. S6 and S7). The synteny tends to be lost for the genomes that have more ancient common ancestor (Fig. 3B and 3C).

Impact of repeated sequences

The number of tandem repeats and the genome coverage of transposable elements is higher in Tuberaceae than in other Pezizomycetes since it may correspond to 60 % of the *T. magnatum* genome (Fig. S8 and S9). Tuberaceae species presented a genome expansion compared to the other Pezizomycetes (Fig. 1; Fig. S1) explained by the repeated sequences content (R^2 Pearson =0.83, p-value=0.01) but not by the coding space (R^2 Pearson =-0.34, p-value=0.4; Fig. S10) and the tandem repeats (Fig. S11). Interestingly, the genome coding space is inversely correlated to genome size (R^2 Pearson =-0.92, p-value=0.001; Fig. S10). The gypsy

retrotransposons were the most frequent known transposable elements representing up to 46 % of the *T. magnatum* genome (Fig. S9). The transposable elements impact the whole genome sizes but also the mean exon size (R^2 Pearson =-0.88, p-value=0.01; Fig. S12). A similar result was found for ants (21) suggesting TE impact the genome by reducing the coding sequencing space.

How can we explain the differences in the repeated sequences content in the Pezizomycete genomes? Repeated sequences are controlled by defense mechanisms such as methylation induced premeiotically (MIP), a reversible control mechanism based on cytosine methylation, and repeat induced point mutation (RIP), an irreversible control mechanism by point mutations. The RIP indexes did not detect evidence of RIP since 1,000 bp window with TpA/ApT above 2 and (CpA+TpG)/(ApC+GpT) below 0.7 are low (Fig. S13). This results suggest that irreversible control of repeated sequences do not exist in Pezizomycetes as already proposed for *T. melanosporum* (11, 22, 23). On the contrary, evidence of methylation of TE for *T. melanosporum* suggested MIP like control mechanisms of repeated sequences (23). The invasion of the Tuberaceae genome by TE could be therefore the consequence of the interruption of this control due for example to environmental factors or change in the ecological niche (22) about 150 Mya (Fig. 1) when ectomycorrhizal host plants (such as Pinaceae) were already presents (24).

Truffle life cycle: sexual reproduction

In ascomycetes, homothallic and heterothallic species have a single MAT locus with two alternative genes (MAT1-1-1 and MAT1-2- 1)(25, 26). In heterothallic fungi, one haploid genome harbors on MAT gene. This is the case for *T. melanosporum* Mel28 strain that harbored only MAT1-2 (11, 13, 15). *T. aestivum* and *T. magnatum* harbored only MAT1-2 gene while *C. venosus* and *M. conica* harbored only MAT1-1 gene (Fig. S15). The MAT

genes were not identified in *A. immersus* and *T. boudieri*. This could be due to the low level of sequence conservation or because the MAT locus is not present in the genome assembly. Interestingly, *P. confluens*, a homothallic species, presented both MAT1-1 and MAT1-2 genes in its haploid genome (Fig. S15) (12). *C. venosus*, *M. conica*, *T. aestivum* and *T. magnatum* are therefore heterothallic as *A. immersus* that is known to be heterothallic (27) even if we were not able to identify MAT genes in its genome assembly. The understanding of truffle sexual reproduction is critical for improving truffle production (28) and the characterization of the MAT loci in *T. aestivum* and *T. magnatum* will provide us new tools for investigating these truffles *in situ* as we did for *T. melanosporum* (22)

The RNAseq data are available for *T. melanosporum* in free-living mycelium (FLM), ectomycorrhizas (ECM) and fruiting bodies (FB)(29). A total 351 gene models are highly up regulated (> 5 times) in FB compared to FLM. Of the induced genes 31.3% have orthologs in all the Pezizomycetes species, mainly gene models involved in metabolism, although 11.7 % are unique to *T. melanosporum*, mainly gene models with no KOG definition (Fig S16).

Truffle life cycle: symbiotic phase

The symbiotic phase is required for truffles to complete their life cycle (30). A total of 437 gene models are highly up regulated ECM compared to FLM of them 25.2% have orthologs in all the Pezizomycetes species, involved mainly in metabolism such as for basidiomycetes ECM (31). About 12 % of the up regulated gene models are unique to *T. melanosporum* and did not have KOG definition (Fig S17). This number of unique genes is similar to those found in symbiosis up regulated genes for basidiomycetes ECMs (7 to 38% of up regulated genes are unique depending of the species)(31).

The establishment of ECM symbiosis in *Laccaria bicolor* is associated with the expression of mycorrhiza-induced small secreted proteins (MiSSPs) (32). The number of secreted protein

vary from less than 216 in *T. boudieri* to 767 in *A. immersus* with the number of SSP varied from 144 for *T. aestivum* to 531 for *A. immersus* (Fig. S18). On the 437 gene models highly up regulated in ECM compared to FLM in *T. melanosporum*, 19 MiSSP were identified (Fig. S19). These MiSSP were not specific to *T. melanosporum* suggesting that they were more conserved than in ECM basidiomycetes where specific MiSSP are identified even when two species of the same genus are compared (e.g. *L. bicolor* and *L. amethystina*) (31).

Saprotrophism capacities in Pezizomycetes: diversity of CAZymes

Recently, Kohler and colleagues (2015) highlighted that ECM species (mainly basidiomycetes) have reduced complements of PCWDEs such as carbohydrate-active enzymes (CAZymes) In Pezizomycetes, the total number of CAZymes ranged from 233 for *T. magnatum* to 449 for *A. immersus* (Fig. S20). When we compared this total number of CAZymes according to the life strategy (Fig. S21), saprotroph species (i.e. *A. immersus*, *M. conica* and *P. confluens*) presented a higher number of enzymes than species forming mycorrhizal symbiosis (*Tuber* spp., *T. boudieri* and *C. venosus*) at the exception of *C. venosus* (Fig. S20 and S21). The loss of most of the GH, such as cellobiohydrolases GH6 and GH7 acting on crystalline cellulose, is a characteristic of ECM Basidiomycetes (31). For Pezizomycetes, there is effectively a lower number of GH acting on plant cell wall than saprotroph (Fig. S21). *Terfezia boudieri*, forming ecto-endomycorrhiza, has maintained one GH7 and GH6 was maintained in *C. venosus*, *M. conica*, *T. aestivum*, and *T. magnatum* (Fig. S22).

In the hierarchical clustering of the eight Pezizomycetes and based on the total number of CAZyme families (Fig. S23) or on the PCWDE (Fig. S24) the *Tuber* spp clustered together with *T. boudieri*, although *C. venosus* clustered with *P. confluens*. In order to have more information on the ECM evolution regarding CAZymes contents we decided to compare the

CAZyme repertoire in 103 species including the eight Pezizomycetes, ECM basidiomycetes, saprotrophs (brown and white rots) and pathogens fungi (Table S2). In the hierarchical clustering using these 103 species and based on the total number of CAZyme families (Fig. S25) the *Tuber* spp. clustered together with *T. boudieri*, saprotroph and pathogenic species. Interestingly, when only PCWDE are considered the *Tuber* spp clustered with ECM Basidiomycetes (Fig. 3). When only fungal cell wall degrading enzymes are considered the clustering is very different since all Pezizomycetes clustered together indicating these enzymes have a different evolution (Fig. S26). This analysis suggested an evolutionary convergence between ECM Basidiomycetes and *Tuber* spp. Interestingly, *C. venosus* clustered with saprotroph species and this species is also rich in CAZymes (Fig. 3 and Fig. S20-S26). *Choiromyces venosus*, is considered as a mycorrhizal species (33–35) and its natural abundance (i.e. ratio $^{15}\text{N}/^{13}\text{C}$) identified a typical symbiotic signature for its ascocarps. It seems therefore that as in *T. melanosporum* the carbon present in the *C. venosus* ascocarps is transferred from the plant via ectomycorrhizas. However, we cannot exclude that the soil mycelium can uptake carbon from organic matter.

Materials and Methods

Genome Sequencing, Assembly, and Annotation. On the six genomes newly sequenced in this study, *A. immersus*, *C. venosus*, *M. conica* and *T. boudieri* were sequenced by the JGI, *T. aestivum* was sequenced by the Genoscope, and *T. magnatum* was sequenced in collaboration by UMR 1136 from Nancy INRA and Plant biology department of the University of Turin (Italy). The Table S1 present the sequencing strategy used for these six species. The methods used in this study are detailed in supplementary data.

Data Access. Genome assemblies and gene annotations are available on the JGI MycoCosm website (<http://jgi.doe.gov/fungi>) and on the INRA dedicated website (<http://mycor.nancy.inra.fr/IMGC/Pezizomycetes>).

Acknowledgements

The UMR1136 is supported by a grant overseen by the French National Research Agency (ANR) as part of the "Investissements d'Avenir" program (ANR-11-LABX-0002-01, Lab of Excellence ARBRE). This study benefited from ANR SYSTERRA SYSTRUF (ANR-09-STRA-10). INRA AIP bioresources The PhD fellow of Thibaut Payen is co-funded by Lorraine Region and the European Commission through the EcoFINDERS project (FP7-264465). We would like to thank Dr Francesco Paolocci for its help in the mating type genes characterization. We also thank the entire Pezizomycete pan-genomic consortium.

References

1. Spatafora JW, et al. (2006) A five-gene phylogeny of Pezizomycotina. *Mycologia* 98(6):1018–1028.
2. Kirk PM, Cannon PF, Minter DW, Stalpers JA (2008) *Dictionary of the Fungi* Available at: <http://www.publish.csiro.au/pid/5998.htm> [Accessed December 4, 2014].
3. Læssøe T, Hansen K (2007) Truffle trouble: what happened to the Tuberales? *Mycol Res* 111(9):1075–1099.
4. O'Donnell K, Cigelnik E, Weber NS, Trappe JM (1997) Phylogenetic Relationships among Ascomycetous Truffles and the True and False Morels Inferred from 18S and 28S Ribosomal DNA Sequence Analysis. *Mycologia* 89(1):48.
5. Hansen K, Laessoe T, Pfister DH (2001) Phylogenetics of the Pezizaceae, with an Emphasis on Peziza. *Mycologia* 93(5):958.
6. Thiers HD (1984) The Secotioid Syndrome. *Mycologia* 76(1):1.
7. Bruns TD, Fogel R, White TJ, Palmer JD (1989) Accelerated evolution of a false-truffle from a mushroom ancestor. *Nature* 339(6220):140–142.

8. Gutiérrez A, Morte A, Honrubia M (2003) Morphological characterization of the mycorrhiza formed by *Helianthemum almeriense* Pau with *Terfezia claveryi* Chatin and *Picoa lefebvrei* (Pat.) Maire. *Mycorrhiza* 13(6):299–307.
9. Smith SE, Read DJ (2010) *Mycorrhizal Symbiosis* (Academic Press).
10. Bonito GM, Gryganskyi AP, Trappe JM, Vilgalys R (2010) A global meta-analysis of Tuber ITS rDNA sequences: species diversity, host associations and long-distance dispersal. *Mol Ecol* 19(22):4994–5008.
11. Martin F, et al. (2010) Périgord black truffle genome uncovers evolutionary origins and mechanisms of symbiosis. *Nature* 464(7291):1033–8.
12. Traeger S, et al. (2013) The Genome and Development-Dependent Transcriptomes of *Pyronema confluens*: A Window into Fungal Evolution. *PLoS Genet* 9(9):e1003820.
13. Rubini A, et al. (2011) Tuber melanosporum: mating type distribution in a natural plantation and dynamics of strains of different mating types on the roots of nursery-inoculated host plants. *New Phytol* 189(3):723–735.
14. Murat C, et al. (2011) Distribution and localization of microsatellites in the Périgord black truffle genome and identification of new molecular markers. *Fungal Genet Biol* 48(6):592–601.
15. Payen T, et al. (2015) A survey of genome-wide single nucleotide polymorphisms through genome re-sequencing in the Périgord black truffle (*Tuber melanosporum* Vittad.). *Mol Ecol Resour*:n/a–n/a.
16. Percudani R, Trevisi A, Zambonelli A, Ottonello S (1999) Molecular phylogeny of truffles (Pezizales: Terfeziaceae, Tuberaceae) derived from nuclear rDNA sequence analysis. *Mol Phylogenet Evol* 13(1):169–180.
17. Bonito G, et al. (2013) Historical biogeography and diversification of truffles in the tuberaceae and their newly identified southern hemisphere sister lineage. *PLoS One* 8(1):e52765–e52765.
18. Prieto M, Wedin M (2013) Dating the Diversification of the Major Lineages of Ascomycota (Fungi). *PLoS ONE* 8(6):e65576.
19. Jeandroz S, Murat C, Wang Y, Bonfante P, Le Tacon F (2008) Molecular phylogeny and historical biogeography of the genus *Tuber*, the “true truffles.” *J Biogeogr* 35(5):815–829.
20. Parra G, Bradnam K, Korf I (2007) CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. *Bioinformatics* 23(9):1061–1067.
21. Schrader L, et al. (2014) Transposable element islands facilitate adaptation to novel environments in an invasive species. *Nat Commun* 5. Available at: <http://www.nature.com/ncomms/2014/141216/ncomms6495/full/ncomms6495.html> [Accessed February 19, 2015].

22. Murat C, et al. (2013) Fine-scale spatial genetic structure of the black truffle (*Tuber melanosporum*) investigated with neutral microsatellites and functional mating type genes. *New Phytol* 199(1):176–187.
23. Montanini B, et al. (2014) The black truffle methylome: non-exhaustive DNA methylation-mediated transposon silencing in a complex fungal genome with massive repeat element content. *Genome Biol* 15(7):411.
24. Rothwell GW, Mapes G, Stockey RA, Hilton J (2012) The seed cone *Eathiestrobus* gen. nov.: Fossil evidence for a Jurassic origin of Pinaceae. *Am J Bot* 99(4):708–720.
25. Metzberg RL, Glass NL (1990) Mating type and mating strategies in *Neurospora*. *Bioessays* 12(2):53–59.
26. Debuchy R, Berteaux-Lecellier V, Silar P (2010) Mating Systems and Sexual Morphogenesis in Ascomycetes. *Cellular and Molecular Biology of Filamentous Fungi* (ASM Press), pp 501 – 536. Available at: <http://hal.archives-ouvertes.fr/hal-00480603> [Accessed December 17, 2013].
27. G. Rizet (1939) Sur les spores dimorphes et l'hérédité de leurs caractères chez un nouvel *Ascobolus* hétérothallique. *Comptes Rendus Académie Sci* 208:1669.
28. Le Tacon F, et al. (2015) Certainties and uncertainties about the life cycle of the Périgord black truffle (*Tuber melanosporum* Vittad.). *Ann For Sci*:1–13.
29. Tisserant E, et al. (2011) Deep RNA sequencing improved the structural annotation of the *Tuber melanosporum* transcriptome. *New Phytol* 189(3):883–891.
30. Mello A, Murat C, Bonfante P (2006) Truffles: much more than a prized and local fungal delicacy. *FEMS Microbiol Lett* 260(1):1–8.
31. Kohler A, et al. (2015) Convergent losses of decay mechanisms and rapid turnover of symbiosis genes in mycorrhizal mutualists. *Nat Genet* advance online publication. Available at: <http://www.nature.com/ng/journal/vaop/ncurrent/full/ng.3223.html> [Accessed February 24, 2015].
32. Plett JM, et al. (2011) A secreted effector protein of *Laccaria bicolor* is required for symbiosis development. *Curr Biol CB* 21(14):1197–1203.
33. Maia LC, Yano AM, Kimbrough JW (1996) Species of Ascomycota forming ectomycorrhizae. *Mycotaxon* 57:371–390.
34. Comandini O, Contu M, Rinaldi AC (2006) An overview of *Cistus* ectomycorrhizal fungi. *Mycorrhiza* 16(6):381–395.
35. Wedén C, Larsson S, Burman R, Backlund A (2009) The edible truffle *Choiromyces venosus* and its use in Sweden. *Acta Bot Yunnanica* 31(S16):94–96.

Figure legends

Fig. 1. Phylogenomic relations, genome size and number of gene models for eight fungal genomes. The tree is a chronogram estimated with PATHd8 on the basis of a maximum-likelihood phylogeny constructed with 2,338 gene families. The ages were estimated with a relaxed molecular clock. The proportion of transposable elements (TE) and non-repeated sequences are indicated. The life strategy of the species is indicated: ECM, ectomycorrhizal; MYC, mycorrhizal; S, saprotroph

Fig. 2. Syntenic relation in the Pezizomycetes. Example of the *T. melanosporum* scaffold 1 synteny based on gene orthology among the four Tuberaceae species (A) or four other Pezizomycetes species (B). *Ascobolus immersus* is absent because there is no block syntenic with this scaffold. One syntenic block is considered only if at least two orthologous genes are presents in the same order in at least two genomes. The yellow, orange and red links mean the syntenic block are composed by less than 5, 5-10 and more than 10 genes, respectively. Gene models position are indicated in blue in the inner circle. The total number of syntenic block with at least two genes is presented in C. Above each bar the estimated age of the common ancestor between each pair of species is indicated in My. Abbreviations for taxon names are defined in supplementary Note.

Fig. 3. Hierarchical clustering based on plant cell wall degrading enzymes CAZyme families for 103 fungal genomes. The life strategy of the species is indicated: ECM, ectomycorrhizal; END, endophyte; MYC, mycorrhizal; ORM, orchid mycorrhizal; P, parasite; S, saprotroph. The Pezizomycete species are indicated by a red rectangle. Abbreviations for taxon names are defined in supplementary Note.

Fig. 1

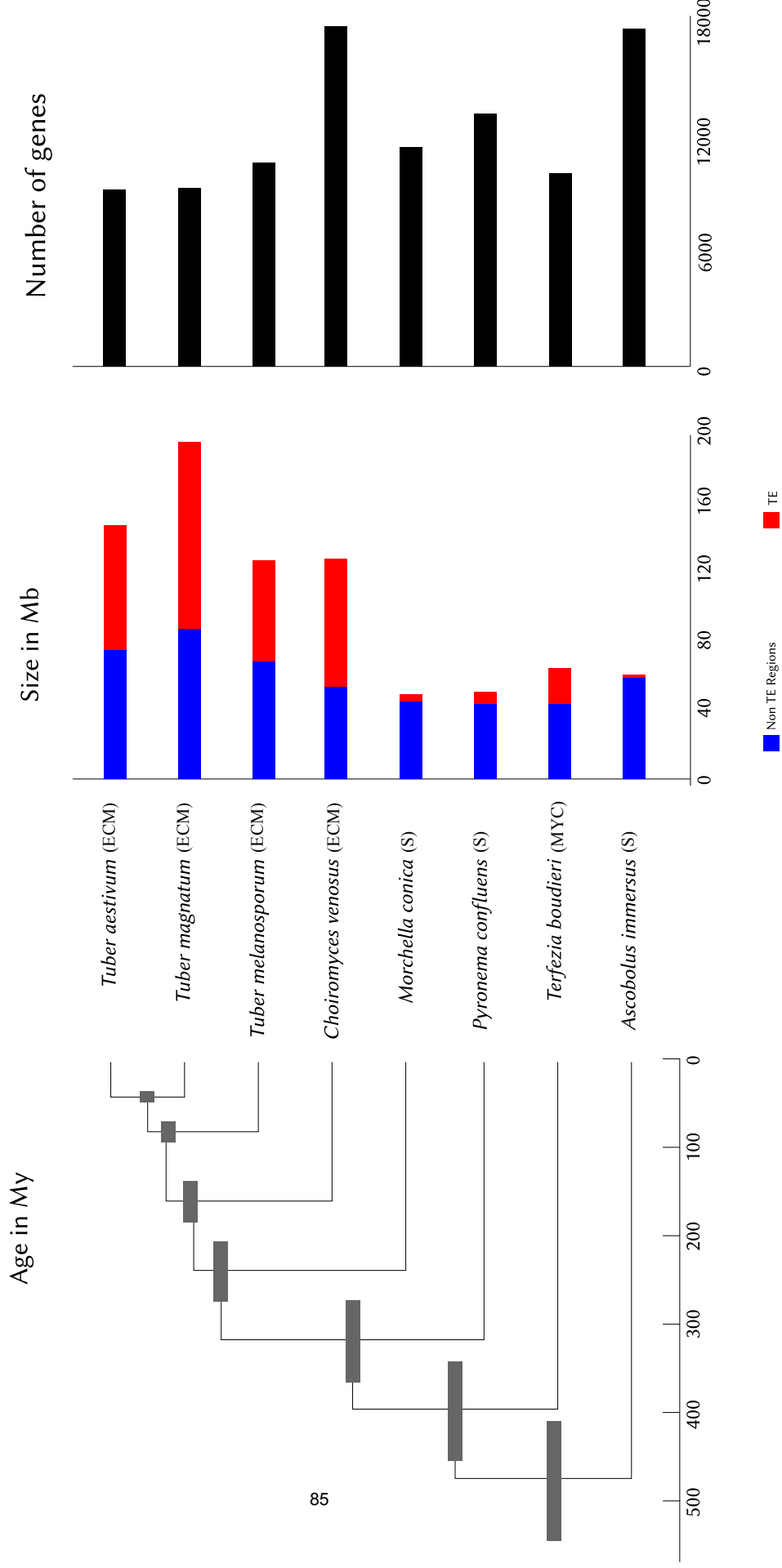
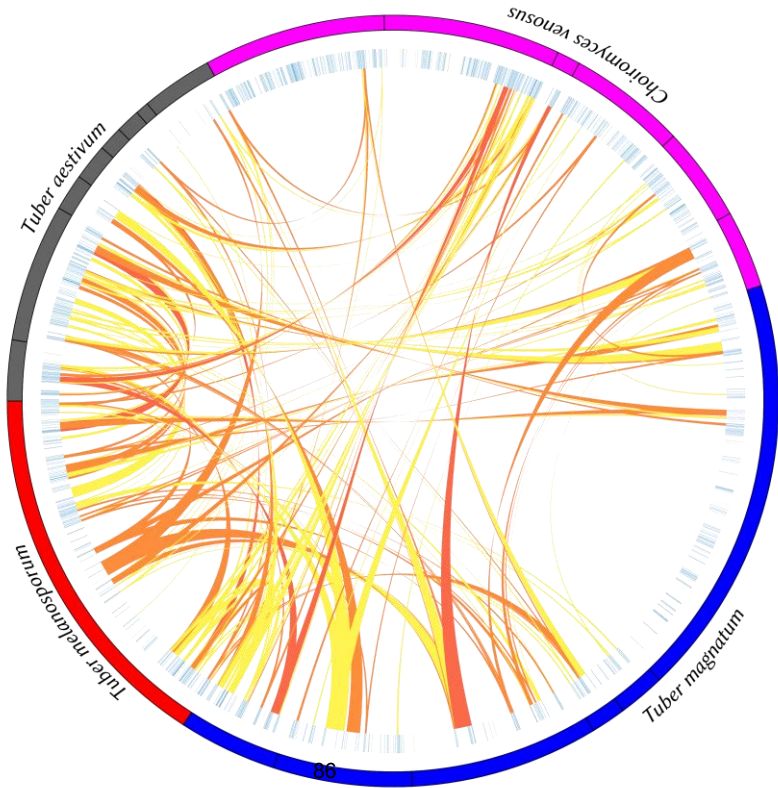


Fig. 2

A



B

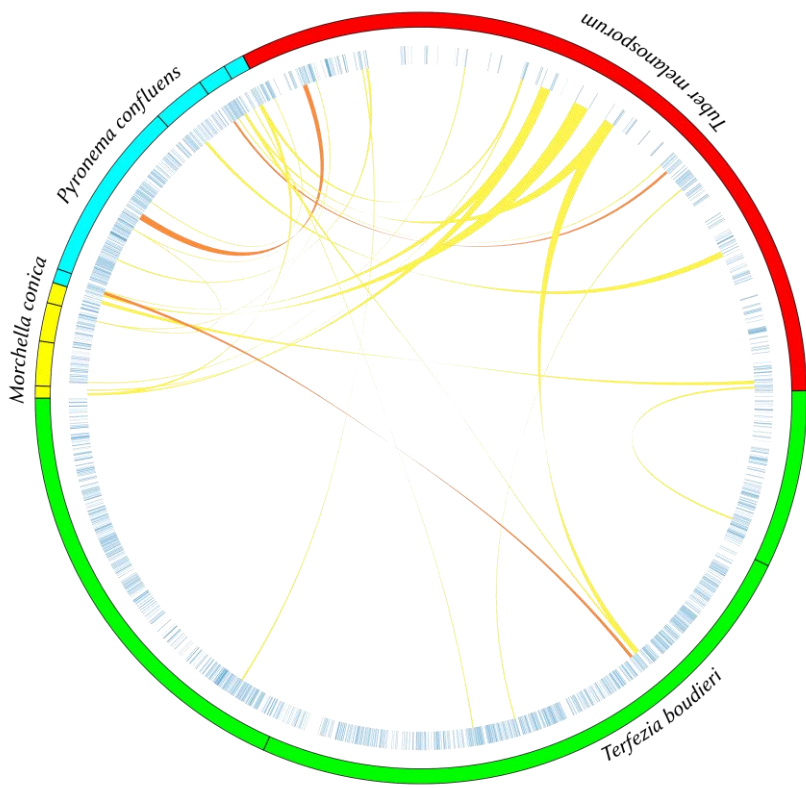
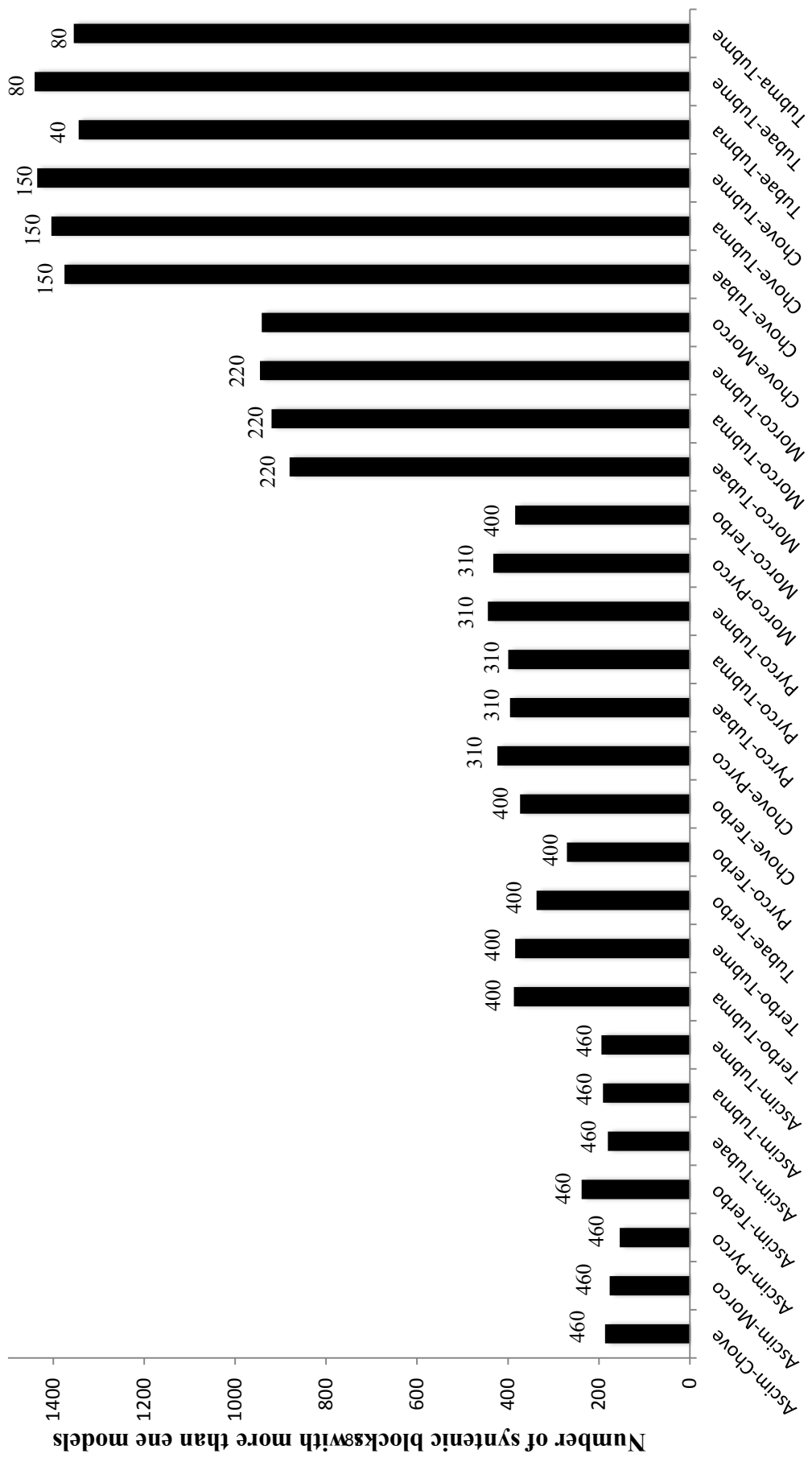
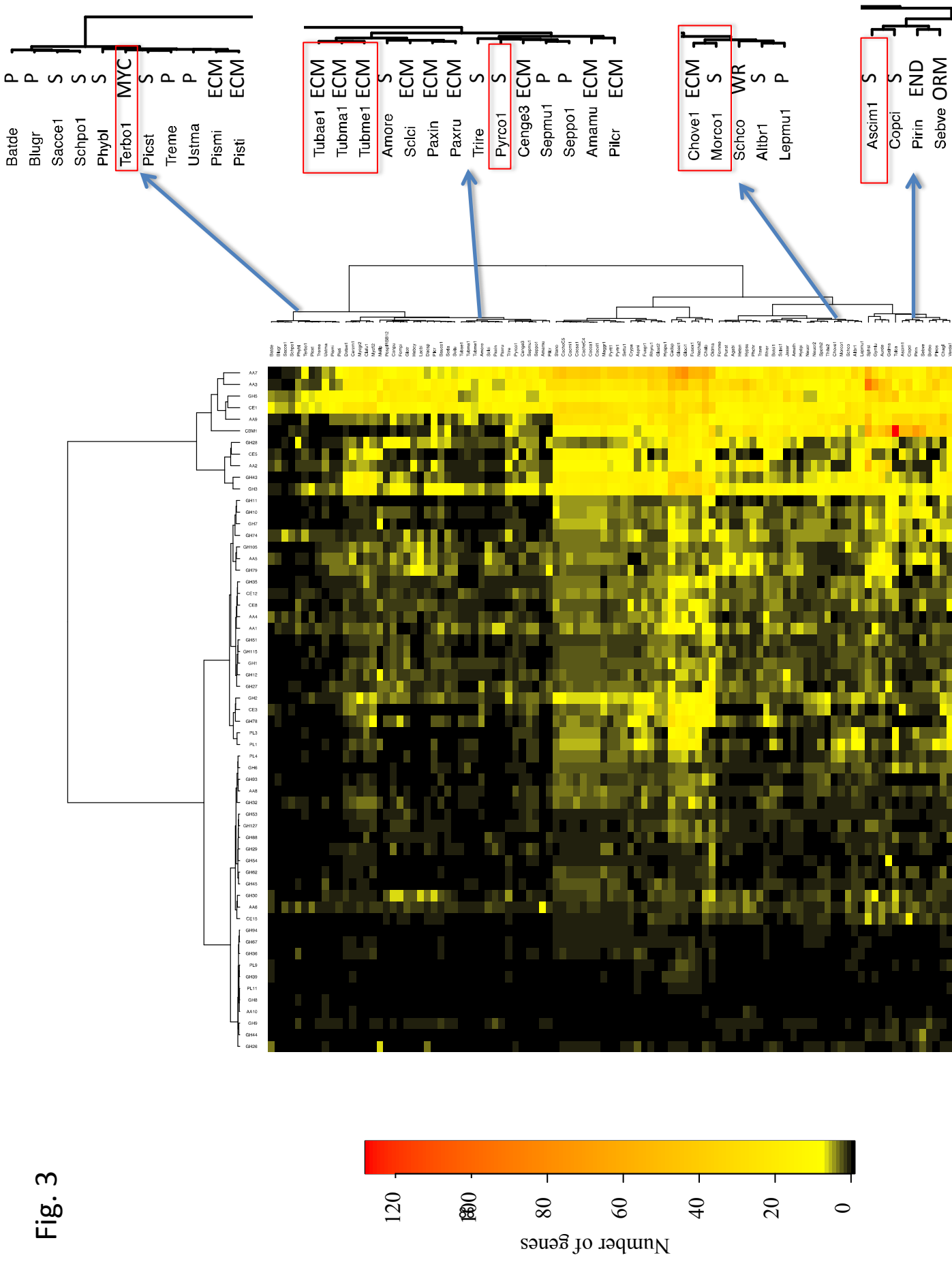


Fig. 2C





Supplementary notes

The genome of highly prized truffles provide new insights into the early diverging Ascomycetes evolution

- 1 Supplementary notes text
 - 1.1 Sequencing, assembly and annotation
 - 1.2 Repeated sequences
 - 1.3 Gene family analysis
 - 1.4 Phylogeny
 - 1.5 CAZymes annotation
 - 1.6 Secreted proteins identification
 - 1.7 Synteny
- 2 References
- 3 Supplementary tables

Supplementary notes text

1.1 Sequencing, assembly and annotation

On the eight genomes used for this study, two were previously published, *T. melanosporum* (1) and *P. confluens* (2), two were sequenced by the JGI in the context of the 1K Fungal Genome Project (*A. immersus* and *M. conica*), two were part of the MGI, aiming to study the mycorrhizal fungi evolution (*Terfezia boudieri* and *Choiromyces venosus*; (1)), *T. aestivum* was sequenced by the Genoscope in the frame of the TuberEvol project, and *T. magnatum* was sequenced in collaboration by UMR 1136 from Nancy INRA and Plant biology department of the University of Turin (Italy). The Table S1 present the sequencing strategy used for these eight species. Except *T. magnatum* all the sequencing and genome assembly was realized by JGI and Genoscope. For *T. magnatum* 454 and Illumina sequencing was realized by Beckman Genomics. Several genome assembly software were tested: SOAPdenovo (3), Velvet (4), CLC Genomics Workbench version 6.6 (<http://www.clcbio.com>) and ALLPATHS-LG (5). The Assemblaton Perl script was used to calculate the genome statistics. Except for ALLPATHS-LG, all assembly software gave low N50 values (few kb), a largest scaffold of less than 100 Kb and a very fragmented genome (data not shown). We decided to use ALLPATHS-LG since this is the only assembler that gave results that are of similar qualities with the assemblies of other Pezizomycetes (Table S1). ALLPATHS-LG requires at least one paired end Illumina library and one mate pairs with insert 3-5 Kb Illumina library. Since 454 reads cannot be used in All-path LG, wgsim (6) was used to simulate a paired end library using the 454 reads previously assembled with Newbler (7). After the genomic assembly, GapCloser with all raw reads (Illumina and 454) was used to close the gaps.

All sequenced genomes were associated with transcriptomic data obtained by RNAseq sometimes associated with 454 sequencing, permitting the annotation of their genes. Three different pipelines were used to do the annotation of the genes. Indeed *T. melanosporum* and *T. aestivum* have been annotated by the Génoscope pipeline in July 2014. For *T. melanosporum* this second version of the annotation lead to a change in the total number of genes compared to the version 1 (1). The genomes sequenced by the JGI as well as *T. magnatum* have been annotated using the JGI pipeline JGI Annotation Pipeline (8–10) which 1) detects and masks repeats and transposable elements, 2) predicts genes using a variety of methods, 3) characterizes each conceptually translated protein using a variety of methods, 4) chooses a ‘best’ gene model at each locus to provide a filtered working set, 5) clusters the filtered sets into draft gene families, and 6) creates a JGI Genome Portal with tools for public access and community-driven curation of the annotation (11). The gene model prediction for *P. confluens* have been done by AUGUSTUS, SNAP, and GeneMark-ES (2).

For all the genomes between 94 to 98 % of the conserved core eukaryotic single-copy genes (CEGMA) are present (12) (Table S1). Interestingly the three *Tuber* spp were sequenced by different strategies (Sanger for *T. melanosporum*, 454 for *T. aestivum* and 454/Illumina for *T. magnatum*) and the genome size, repeat content and number of protein-coding genes are similar even if the number of scaffolds differs from one species to another (Table 2). Besides, the *T. aestivum* and *T. melanosporum* protein-coding genes have been annotated by the Genoscope pipeline and *T. magnatum* have been annotated by the JGI pipeline and the number of genes is similar (Table S1).

1.2 Repeated sequences

RepeatScout (13) was used to identify *de novo* repetitive DNA in the eight genomes. The default parameters (with $l=15$) were used. The RepeatScout library was then filtered as follows: 1) all the sequences less than 100 bp in size were discarded; 2) repeats having less than ten copies in the genome were removed (as they may correspond to protein-coding gene families) and 3) repeats having significant hits to known proteins in UNIPROT (14) other than proteins known as belonging to transposable elements (TEs) were removed. The consensus sequences remaining were annotated manually by a TBLASTX search (15) against RepBase (16) to classify them in known TE families. To identify full length LTR retrotransposons, a second *de novo* search was performed with LTR_STRUC (17). The full-length candidates LTR retrotransposon sequences were checked for their homology using the TBLASTX algorithm against the sequences coming from the RepBase database. The number of repeat elements occurrences and the percent of genome coverage were assessed by masking the genome assembly using RepeatMasker (18) with the consensus sequences coming from the RepeatScout and LTR_STRUC pipelines

1.3 Gene family analysis

Protein sequences from the eight Pezizomycetes were clustered with orthoMCL (version 2.0.9; Li, Stoeckert, and Roos 2003) that used BLAST+ version 2.2.28+ (15) with an e-value threshold of $1e^{-5}$ and MCL version 12.5 (20) using an inflation parameter of 2.5. Using the clusters created with orthoMCL, an analysis of expansion in gene families have been performed using CAFE version 3.1 (21). The λ , corresponding to rate of birth/death of genes, was calculated by CAFE. This analysis was conducted using a unique λ value for all branches of the phylogenetic tree produced as described below. The rate of λ in our tree resulting from the CAFE analysis was $1.9e^{-2}$. In order to search for gene families specific to phylogenetic

groups (e.g. *Tuber* spp.) or ecological groups (e.g. mycorrhizal versus saprotrophs) a t-test for each cluster with a p-value of 0.01 corrected with Bonferroni methods was realized.

1.4 Phylogeny

The clusters produced by orthoMCL with a unique conserved protein in each of the eight species were selected to make a phylogeny. The protein sequences were aligned using Clustal Omega (22) and the highly conserved blocks were selected using Gblocks with all default parameters (23). The resulting alignment was used to build a maximum-likelihood tree with RaxML (24) with the Gamma Dayhoff model and 100 rounds of bootstrapping. The ML phylogeny was first midpoint rooted with *midpoint* function of R package *phangorn* (25). The smoothing parameter (called lambda) was estimated using *chronopl* function with the cross-validation enabled of R package *ape* (26). The lambda values ranging from 0 to 100,000 were tested to minimize the value of the cross-validation. The relaxed molecular clock was then realized with the function *chronos* of R package *ape*. The tree calibration was realized with the Tuberaeae most recent common ancestor estimated by Bonito et al. (27).

1.5 CAZymes annotation

Identification of carbohydrate active enzyme (CAZyme) genes in the eight Pezizomycetes genomes was achieved using the CAZy database (www.cazy.org) annotation pipeline (28). Briefly, the method combines BLAST and HMMer searches conducted against sequence libraries and HMM profiles made of the individual functional modules featured in the CAZy database. All positive hits were manually examined by human curators for final validation.

In order to compare the CAZymes content of the eight Pezizomycetes with those of other fungal species we decided to identify and classify the CAZymes for 100 fungal genomes (Annex 1) using the dbCAN HMMer-based classification system (29). In order to perform

this comparison we need to have all genomes analyzed by the same pipeline that's why we decided to use dbCAN instead of the expert CAZyme annotation described above. Indeed, as highlighted by Lombard and colleagues (30), a CAZyme analysis of a genome performed with one method usually cannot be compared with one done with another.

For this analysis, the fasta files with all predicted proteins sequences for each genome was scan using Hmmscan program in HMMER 3.0 package (31). Results were processed by the hmmscan-parser script supplied by the dbCAN. The following criteria suggested by dbCAN were applied: if alignment superior to 80 amino-acids an e-value inferior to $1e^{-5}$ was used otherwise an e-value inferior to $1e^{-3}$ was applied and finally a covered fraction of the HMM motif had to be superior to 30%.

1.6 Secreted proteins identification

The secreted proteins were predicted using combined characteristics: 1) the proteins were predicted as secreted if the presence of a signal peptide was detected with SignalP version 4.1 (32); 2) the protein subcellular localization was assigned as secretory pathway using TARGETP (version 1.1; (33) and as extracellular with WOLFPSORT version 0.2 (34), and 3) to filter out proteins that permanently reside in the lumen of the endoplasmic reticulum (ER), the proteins were scanned looking for the KDEL motif (Lys-Asp-Glu-Leu) in C-terminal (prosite accession 'PS00014') with ps_scan version 1.0 (35). Indeed, it is possible to distinguished proteins residing in the lumen of ER from newly synthesized secretory proteins by the presence of KDEL in C-terminal sequence. We also search for transmembrane helix with TMHMM version 2.0 (36) and predicted the proteins as secreted only if they had no helix or one overlapping the signal peptide. Finally, each predicted secreted protein has been classified as CAZyme, FOLyme, small-secreted protein (SSP), protease or lipase. A protein was classified as SSP if it is predicted as secreted and if it has a size below 300 amino acids.

Proteins were annotated as proteases if they show a good match with a MEROPS peptidase (30% identity on 50% of the protein) and if they contained the PFAM motif(s) characteristic of the protease family when existing. The same strategy has been applied to the lipases annotation using the Lipase Engineering Database (LED) (<http://www.led.uni-stuttgart.de/>).

1.7 Synteny

To conduct an analysis of the synteny in the eight genomes, two methods have been used.

In the first one, the command *promer* of the MUMmer package (37) was used to find blocks in the complete genomes with the one-against-one strategy using the six frames of amino acid translation of the genomes. Indeed the genomes were too divergent to use *nucmer* that analyze the genome at the nucleotide level except for *Tubercaceae* (data not shown). The command *mummerplot* was used to generate all the gnuplot files and gnuplot (38) was used to generate the final images.

A second analyze based on the gene models was also realized. Using the OMA program (39), we generated a list of true orthologous genes (at the opposite of orthoMCL that put several homologous genes of the same species together generating genes families) that were then used to generate a file used in *circos* (40) to generate a visualization of the synteny based on orthologous genes.

2 References

1. Martin F, et al. (2010) Périgord black truffle genome uncovers evolutionary origins and mechanisms of symbiosis. *Nature* 464(7291):1033–8.
2. Traeger S, et al. (2013) The Genome and Development-Dependent Transcriptomes of *Pyronema confluens*: A Window into Fungal Evolution. *PLoS Genet* 9(9):e1003820.
3. Luo R, et al. (2012) *SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler*. *Gigascience* 1: 18.
4. Zerbino DR, Birney E (2008) Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res* 18(5):821–9.
5. Gnerre S, et al. (2010) High-quality draft assemblies of mammalian genomes from massively parallel sequence data. *Proc Natl Acad Sci* 108(4):1513–1518.
6. Li H wgsim - Read simulator for next generation sequencing. Available at: <http://github.com/lh3/wgsim> [Accessed September 24, 2013].
7. Margulies M, et al. (2005) Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 437(7057):376–380.
8. Grigoriev IV, Martinez DA, Salamov AA (2006) Fungal genomic annotation. *Applied Mycology and Biotechnology*, Bioinformatics., ed Dilip K. Arora RMB and GBS (Elsevier), pp 123–142. Available at: <http://www.sciencedirect.com/science/article/pii/S1874533406800080> [Accessed February 26, 2015].
9. Kuo A, Bushnell B, Grigoriev IV (2014) Fungal Genomics: Sequencing and Annotation. *Adv Bot Res* 70:1–52.
10. Kohler A, et al. (2015) Convergent losses of decay mechanisms and rapid turnover of symbiosis genes in mycorrhizal mutualists. *Nat Genet* advance online publication. Available at: <http://www.nature.com/ng/journal/vaop/ncurrent/full/ng.3223.html> [Accessed February 24, 2015].
11. Grigoriev IV, et al. (2012) The Genome Portal of the Department of Energy Joint Genome Institute. *Nucleic Acids Res* 40(D1):D26–D32.
12. Parra G, Bradnam K, Korf I (2007) CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. *Bioinformatics* 23(9):1061–1067.
13. Price AL, Jones NC, Pevzner PA (2005) De novo identification of repeat families in large genomes. *Bioinformatics* 21 Suppl 1:i351–8.
14. UniProt Consortium (2008) The Universal Protein Resource (UniProt). *Nucleic Acids Res* 36(suppl 1):D190–D195.
15. Altschul S (1990) Basic Local Alignment Search Tool. *J Mol Biol* 215(3):403–410.
16. Jurka J, et al. (2005) Repbase Update, a database of eukaryotic repetitive elements. *Cytogenet Genome Res* 110(1-4):462–467.
17. McCarthy EM, McDonald JF (2003) LTR_STRUC: a novel search and identification program for LTR retrotransposons. *Bioinformatics* 19(3):362–367.
18. Smit AF, Hubley R, Green P (1996) *RepeatMasker Open-3.0*.

19. Li L, Stoeckert CJ, Roos DS (2003) OrthoMCL: Identification of Ortholog Groups for Eukaryotic Genomes. *Genome Res* 13(9):2178–2189.
20. Enright AJ, Dongen SV, Ouzounis CA (2002) An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Res* 30(7):1575–1584.
21. Han MV, Thomas GWC, Lugo-Martinez J, Hahn MW (2013) Estimating Gene Gain and Loss Rates in the Presence of Error in Genome Assembly and Annotation Using CAFE 3. *Mol Biol Evol* 30(8):1987–1997.
22. Sievers F, et al. (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* 7(539). Available at: <http://www.nature.com/doifinder/10.1038/msb.2011.75>.
23. Castresana J (2000) Selection of Conserved Blocks from Multiple Alignments for Their Use in Phylogenetic Analysis. *Mol Biol Evol* 17(4):540–552.
24. Stamatakis A (2006) RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22(21):2688–2690.
25. Schliep KP (2011) phangorn: phylogenetic analysis in R. *Bioinformatics* 27(4):592–593.
26. Paradis E, Claude J, Strimmer K (2004) APE: Analyses of Phylogenetics and Evolution in R language. *Bioinformatics* 20(2):289–290.
27. Bonito G, et al. (2013) Historical biogeography and diversification of truffles in the tuberaceae and their newly identified southern hemisphere sister lineage. *PLoS One* 8(1):e52765–e52765.
28. Cantarel BL, et al. (2009) The Carbohydrate-Active EnZymes database (CAZY): an expert resource for Glycogenomics. *Nucleic Acids Res* 37(suppl 1):D233–D238.
29. Yin Y, et al. (2012) dbCAN: a web resource for automated carbohydrate-active enzyme annotation. *Nucleic Acids Res* 40(W1):W445–W451.
30. Lombard V, Ramulu HG, Drula E, Coutinho PM, Henrissat B (2014) The carbohydrate-active enzymes database (CAZY) in 2013. *Nucleic Acids Res* 42(D1):D490–D495.
31. Eddy SR (2011) Accelerated Profile HMM Searches. *PLoS Comput Biol* 7(10):e1002195.
32. Petersen TN, Brunak S, Heijne G von, Nielsen H (2011) SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat Methods* 8(10):785–786.
33. Emanuelsson O, Nielsen H, Brunak S, von Heijne G (2000) Predicting Subcellular Localization of Proteins Based on their N-terminal Amino Acid Sequence. *J Mol Biol* 300(4):1005–1016.
34. Horton P, et al. (2007) WoLF PSORT: protein localization predictor. *Nucleic Acids Res* 35(suppl 2):W585–W587.
35. A G, E G, A B (2001) ScanProsite: a reference implementation of a PROSITE scanning tool. *Appl Bioinformatics* 1(2):107–108.
36. Melén K, Krogh A, von Heijne G (2003) Reliability Measures for Membrane Protein Topology Prediction Algorithms. *J Mol Biol* 327(3):735–744.
37. Kurtz S, et al. (2004) Versatile and open software for comparing large genomes. *Genome Biol* 5(2):R12.
38. Williams T, Kelley C, gnuplot development team *Gnuplot: an interactive plotting program* Available at: <http://gnuplot.info>.
39. Dessimoz C, et al. (2005) OMA, A Comprehensive, Automated Project for the Identification of Orthologs from Complete Genome Data: Introduction and First Achievements. *Comparative Genomics*, Lecture

Notes in Computer Science., eds McLysaght A, Huson DH (Springer Berlin Heidelberg), pp 61–72. Available at: http://link.springer.com/chapter/10.1007/11554714_6 [Accessed December 9, 2014].

40. Krzywinski M, et al. (2009) Circos: An information aesthetic for comparative genomics. *Genome Res* 19(9):1639–1645.

3 Supplementary tables

Table S1. Statistics of the eight Pezizomycetes genomes as well as the sequencing center, sequencing strategy and gene annotation pipeline.

Species	Abbreviation	Size (Mb)	# scaffold	N50 scaffold (Mb)	% N	% of TE	# of genes	# of 248 CEGMA genes	Sequencing center
<i>Ascobolus immersus</i>	<i>Ascim</i>	60	706	0.27	1.1	3.5	17877	238 (96 %)	JGI
<i>Choiromyces venosus</i>	<i>Chove</i>	124	1648	0.31	5.1	53	17986	240 (98 %)	JGI
<i>Morchella conica</i>	<i>Morco</i>	48	540	0.6	2.1	8.5	11600	235 (95 %)	JGI
<i>Pyronema confluens</i>	<i>Pyrco</i>	50	1588	0.14	0.1	14	13367	236 (95 %)	MPI
<i>Terfezia boudieri</i>	<i>Terbo</i>	63	516	0.78	5.7	32	10200	242 (98 %)	JGI
<i>Tuber aestivum</i>	<i>Tubae</i>	145	3244	0.44	9.1	49.5	9344	239 (96 %)	Génoscope
<i>Tuber magnatum</i>	<i>Tubma</i>	192	1283	1.81	13.4	58	9433	234 (94 %)	INRA/UNITO
<i>T. melanosporum</i>	<i>Tubme</i>	125	398	0.64	1.1	58	10763	239 (96 %)	Génoscope

Table S2. List and abbreviation of all the species used in this analysis.

Full species name	Abbreviation	Phylum	Life strategy
<i>Agaricus bisporus</i>	Agabi	Basidiomycota	saprophyte
<i>Alternaria brassicicola</i>	Altbr	Ascomycota	plant pathogen
<i>Amanita muscaria</i>	Amamu	Basidiomycota	ectomycorrhizal
<i>Amanita thiersii</i>	Amath	Basidiomycota	saprophyte
<i>Amorphotheca resiniae</i>	Amore	Ascomycota	saprophyte
<i>Ascobolus immersus</i>	Ascim	Ascomycota	saprophyte
<i>Ascocoryne sarcoides</i>	Ascsa	Ascomycota	saprophyte
<i>Aspergillus nidulans</i>	Aspni	Ascomycota	human pathogen
<i>Auricularia delicata</i>	Aurde	Basidiomycota	saprophyte
<i>Batrachochytrium dendrobatidis</i>	Batde	Chytridiomycota	animal pathogen
<i>Baudoinia compniacensis</i>	Bauco	Ascomycota	saprophyte
<i>Blumeria graminis</i>	Blugr	Ascomycota	plant pathogen
<i>Botryobasidium botryosum</i>	Botbo	Basidiomycota	saprophyte
<i>Botrytis cinerea</i>	Botci	Ascomycota	plant pathogen
<i>Cadophora sp.</i>	Cadsp	Ascomycota	endophyte
<i>Cenococcum geophilum</i>	Cenge3	Ascomycota	ectomycorrhizal
<i>Cercospora zeae-maydis</i>	Cerzm	Ascomycota	plant pathogen
<i>Chaetomium globosum</i>	Chagl	Ascomycota	saprophyte
<i>Chalara longipes</i>	Chalo	Ascomycota	saprophyte
<i>Choiromyces venosus</i>	Chove	Ascomycota	ectomycorrhizal
<i>Cladosporium fulvum</i>	Clafu	Ascomycota	plant pathogen
<i>Cochliobolus carbonum 26-R-1</i>	Cocca	Ascomycota	plant pathogen
<i>Cochliobolus heterostrophus C4</i>	CocheC4	Ascomycota	plant pathogen
<i>Cochliobolus heterostrophus C5</i>	CocheC5	Ascomycota	plant pathogen
<i>Cochliobolus miyabeanus WK-1C</i>	Cocmi	Ascomycota	plant pathogen
<i>Cochliobolus sativus ND90Pr</i>	Cocsa	Ascomycota	plant pathogen
<i>Cochliobolus victoriae FI3</i>	Cocvi	Ascomycota	plant pathogen
<i>Coniophora puteana</i>	Conpu	Basidiomycota	saprophyte
<i>Coprinopsis cinerea</i>	Copci	Basidiomycota	saprophyte
<i>Cryphonectria parasitica</i>	Crypa	Ascomycota	plant pathogen
<i>Dacryopinax sp.</i>	Dacsp	Basidiomycota	saprophyte
<i>Dothistroma septosporum NZE10</i>	Dotse	Ascomycota	plant pathogen
<i>Fomitiporia mediterranea</i>	Fomme	Basidiomycota	saprophyte
<i>Fomitopsis pinicola</i>	Fompi	Basidiomycota	saprophyte
<i>Fusarium graminearum</i>	Fusgr	Ascomycota	plant pathogen
<i>Fusarium oxysporum</i>	Fusox	Ascomycota	plant pathogen
<i>Galerina marginata</i>	Galma	Basidiomycota	saprophyte
<i>Gloeophyllum trabeum</i>	Glotr	Basidiomycota	Saprotrophic
<i>Glomerella acutata</i>	Gloac	Ascomycota	plant pathogen
<i>Glomerella cingulata</i>	Gloci	Ascomycota	saprophyte
<i>Glonium stellatum</i>	Glost	Ascomycota	Saprotrophic
<i>Gymnopus luxurians</i>	Gymlu	Basidiomycota	saprophyte
<i>Hebeloma cylindrosporum</i>	Hebcy	Basidiomycota	ectomycorrhizal

<i>Heterobasidion annosum</i>	Hetan	Basidiomycota	saprophyte
<i>Hydnomerulius pinastri</i>	Hydpi	Basidiomycota	saprophyte
<i>Hypholoma sublateritium</i>	Hypsu	Basidiomycota	saprophyte
<i>Hysterium pulicare</i>	Hypsu	Ascomycota	saprophyte
<i>Jaapia argillacea</i>	Jaaar	Basidiomycota	saprophyte
<i>Laccaria bicolor</i>	Lacbi	Basidiomycota	ectomycorrhizal
<i>Leptosphaeria maculans</i>	Lepmu	Ascomycota	plant pathogen
<i>Magnaporthe grisea</i>	Maggr	Ascomycota	plant pathogen
<i>Melampsora laricis-populina</i>	Mellp	Basidiomycota	plant pathogen
<i>Meliniomyces variabilis</i>	Melva	Ascomycota	endophyte
<i>Melinionyces bicolo</i>	Melbi	Ascomycota	ectomycorrhizien
<i>Morchella conica</i>	Morco	Ascomycota	saprophyte
<i>Mycosphaerella fijiensis</i>	Mycfi	Ascomycota	plant pathogen
<i>Mycosphaerella graminicola</i>	Mycgr	Ascomycota	plant pathogen
<i>Nectria haematococca</i>	Necha	Ascomycota	saprotrophe/pathogen.
<i>Neurospora crassa</i>	Neucr	Ascomycota	pathogen
<i>Oidiodendron maius</i>	Oidma	Ascomycota	ericoid mycorrhizal
<i>Paxillus involutus</i>	Paxin	Basidiomycota	ectomycorrhizal
<i>Paxillus rubicundulus</i>	Paxru	Basidiomycota	ectomycorrhizal
<i>Phanerochaete chrysosporium</i>	Phchr	Basidiomycota	saprophyte
<i>Phycomyces blakesleeanus</i>	Phybl	Mucoromucotina	saprophyte
<i>Pichia stipitis</i>	Picst	Ascomycota	saprophyte
<i>Piloderma croceum</i>	Pilcr	Basidiomycota	ectomycorrhizal
<i>Piriformospora indica</i>	Pirin	Basidiomycota	endophyte
<i>Pisolithus microcarpus</i>	Pismi	Basidiomycota	ectomycorrhizal
<i>Pisolithus tinctorius</i>	Pisti	Basidiomycota	ectomycorrhizal
<i>Pleurotus ostreatus</i>	Pleos	Basidiomycota	saprophyte
<i>Plicaturopsis crispa</i>	Plicr	Basidiomycota	saprophyte
<i>Postia placenta</i>	Pospl	Basidiomycota	saprophyte
<i>Punctularia strigosozonata</i>	Punst	Basidiomycota	saprophyte
<i>Pyrenophora teres f. teres</i>	Pyrtt	Ascomycota	necrotrophic
<i>Pyrenophora tritici-repentis</i>	Pytrr	Ascomycota	plant pathogen
<i>Pyronema confluens</i>	Pyrco	Ascomycota	saprophyte
<i>Rhizoscyphus ericae</i>	Rhier	Ascomycota	ericoid mycorrhizal
<i>Rhytidhysterium rufulum</i>	Rhyru	Ascomycota	saprophyte
<i>Saccharomyces cerevisiae</i>	Sacce	Ascomycota	saprophyte
<i>Schizophyllum commune</i>	Schco	Basidiomycota	saprophyte
<i>Schizosaccharomyces pombe</i>	Schpo	Ascomycota	saprophyte
<i>Scleroderma citrinum</i>	Scldi	Basidiomycota	ectomycorrhizal
<i>Sclerotinia sclerotiorum</i>	Scpsc	Ascomycota	plant pathogen
<i>Sebacina vermifera</i>	Sebve	Basidiomycota	orchid mycorrhizal
<i>Septoria musiva SO2202</i>	Sepmu	Ascomycota	plant pathogen
<i>Septoria populicola</i>	Seppo	Ascomycota	plant pathogen
<i>Serpula lacrymans</i>	Serla	Basidiomycota	saprophyte
<i>Setosphaeria turcica Et28A</i>	Settu	Ascomycota	plant pathogen
<i>Sphaerobolus stellatus</i>	Sphst	Basidiomycota	saprophyte

<i>Sporotrichum thermophile</i>	Spoth	Ascomycota	saprophyte
<i>Stagonosporaa nodorum</i>	Stano	Ascomycota	plant pathogen
<i>Suillus luteus</i>	Suilu	Basidiomycota	ectomycorrhizal
<i>Terfezia boudieri</i>	Terbo	Ascomycota	Mycorrhizal
<i>Thielavia terrestris</i>	Thite	Ascomycota	saprophyte
<i>Trametes versicolor</i>	Trave	Basidiomycota	saprophyte
<i>Tremella mesenterica</i>	Treme	Basidiomycota	saprophyte
<i>Trichoderma reesei</i>	Trire	Ascomycota	saprophyte
<i>Tuber aestivum</i>	Tubae	Ascomycota	ectomycorrhizal
<i>Tuber magnatum</i>	Tubma	Ascomycota	ectomycorrhizal
<i>Tuber melanosporum</i>	Tubme	Ascomycota	ectomycorrhizal
<i>Tulasnella calospora</i>	Tulca	Basidiomycota	orchid mycorrhizal
<i>Ustilago maydis</i>	Ustma	Basidiomycota	Plant pathogen
<i>Verticillium dahliae</i>	Verda	Ascomycota	plant pathogen

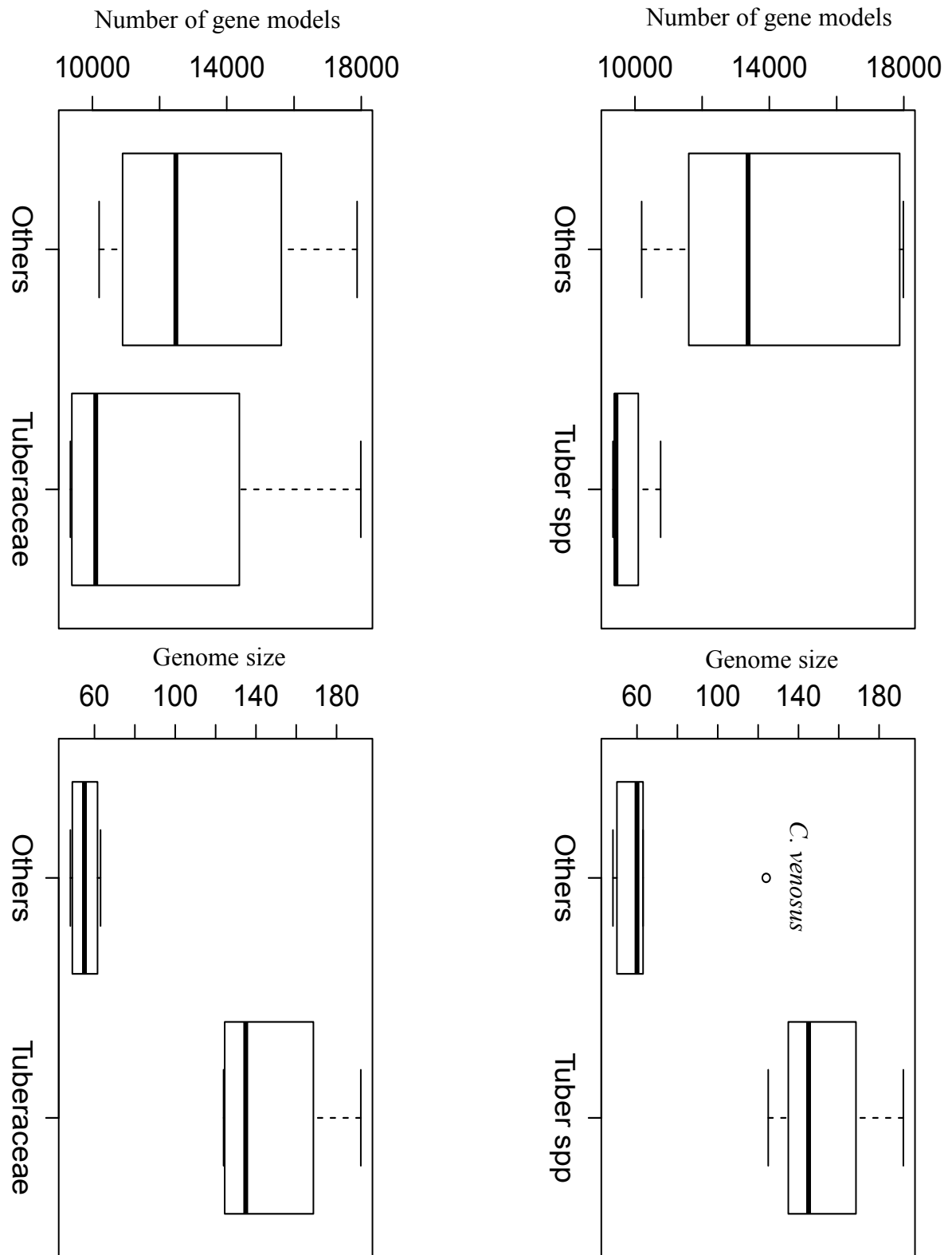


Fig. S1. Boxplots showing the number of gene models and genome size for Tuber spp (*T. aestivum*, *T. magnatum* and *T. melanosporum*) and others five Pezizomycetes and Tuberaceae (*C. venosus* and Tuber spp) and other four Pezizomycetes

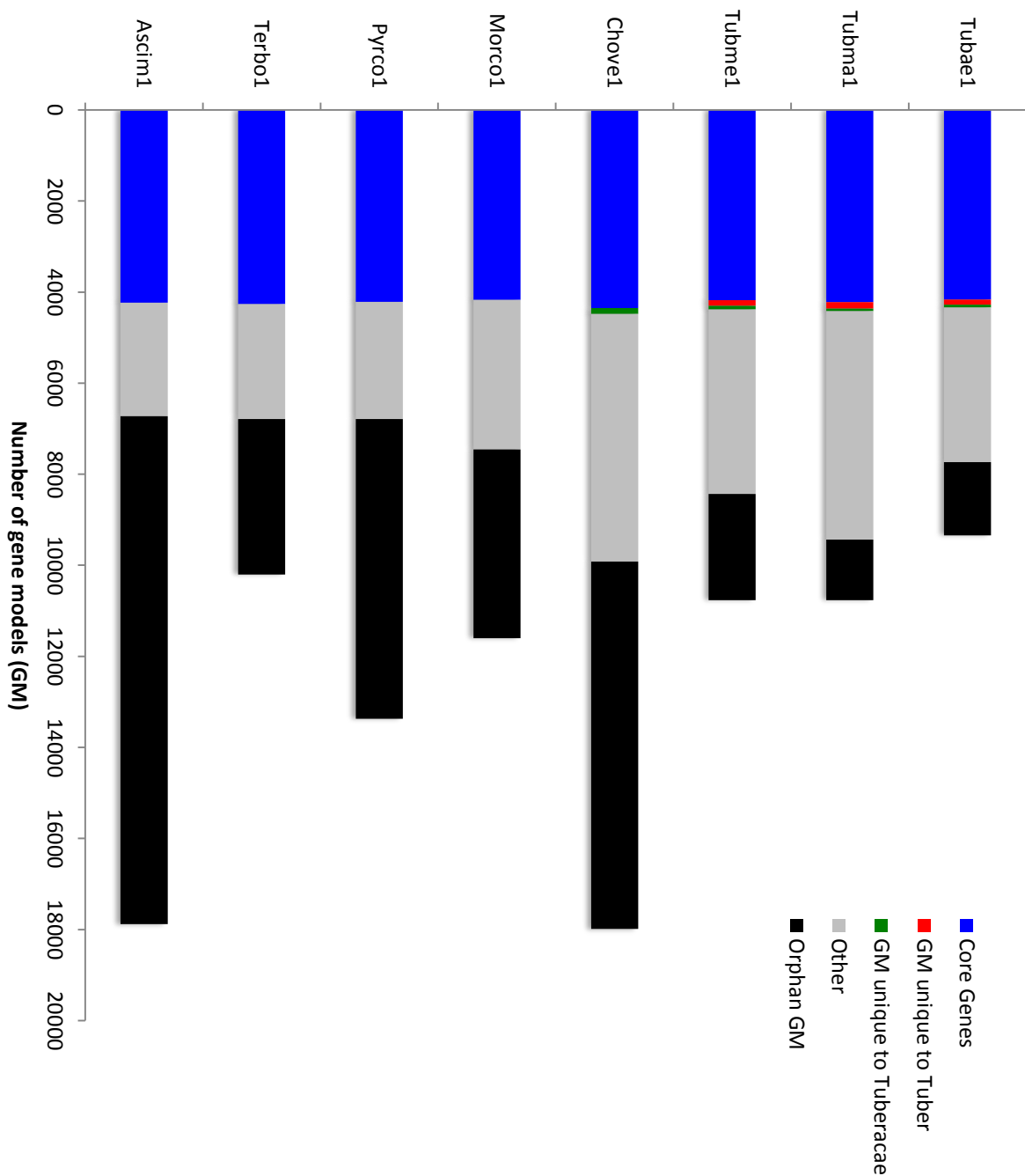


Fig. S2. Number of gene models (GM) for the eight Pezizomycete species that are found in all the species (core genes), only in Tuberaeae (GM unique to Tuberaeae), only in Tuber (GM unique to Tuber) or found in different species (others). The species abbreviation is found in the Supplementary notes.

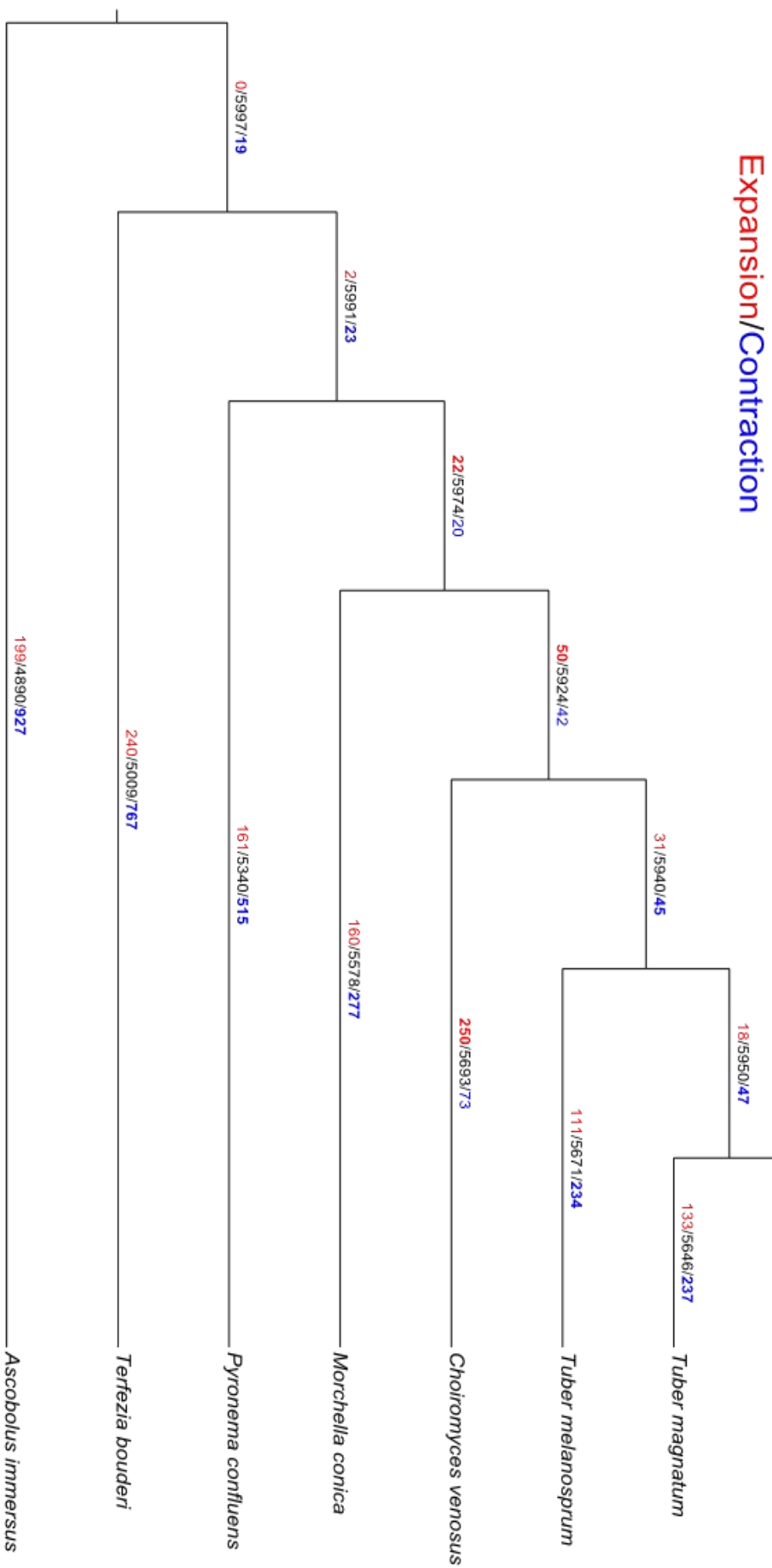


Fig. S3. The figure represents the total number of protein families in each species or node realized by CAFE. The numerals on branches show numbers of expanded (left, red), unchanged (middle, black) or contracted (right, blue) protein families along lineages by comparison to the putative pan-proteome.

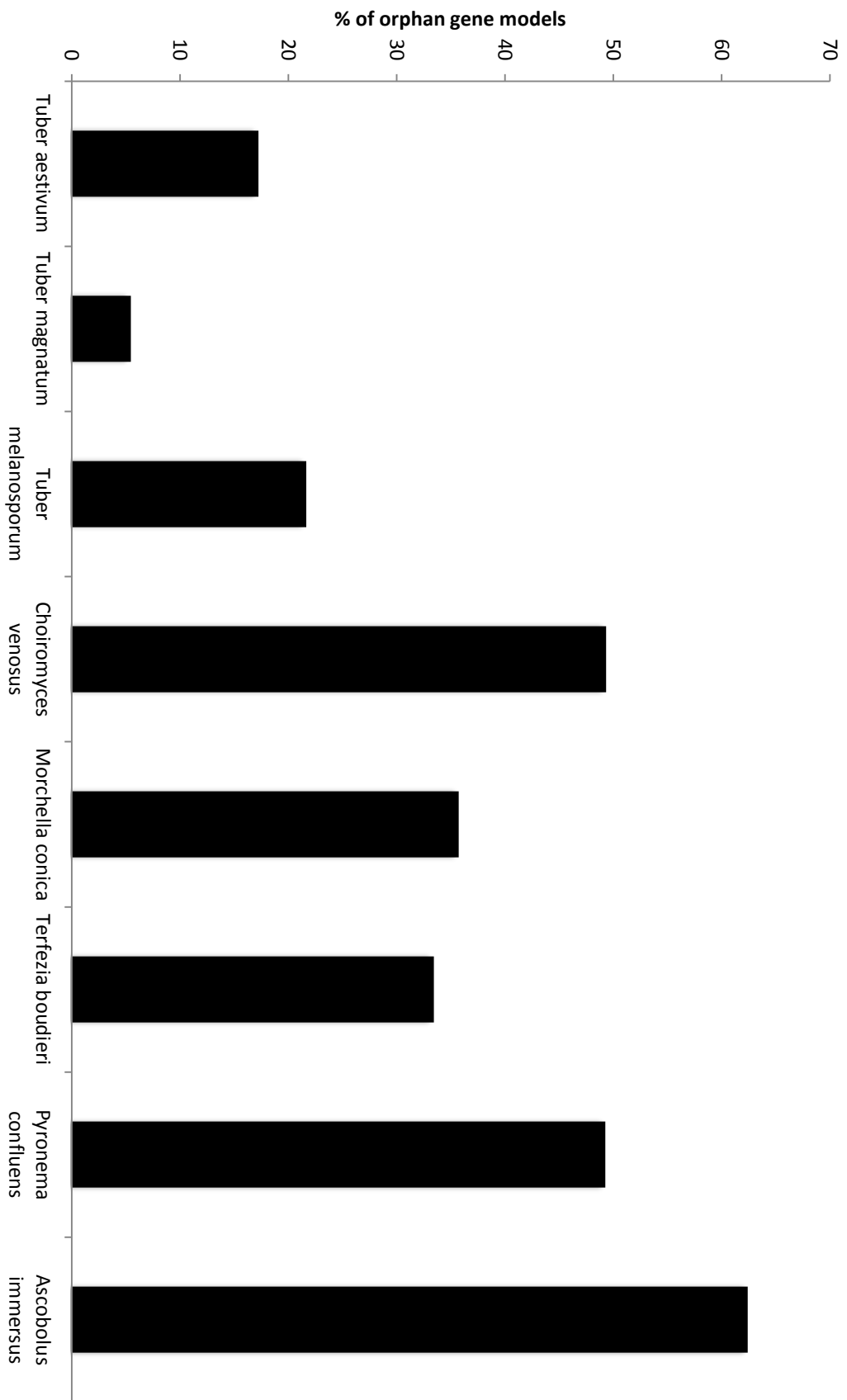


Fig. S4. Percentage of orphan gene models for the eight Pezizomycete genomes.

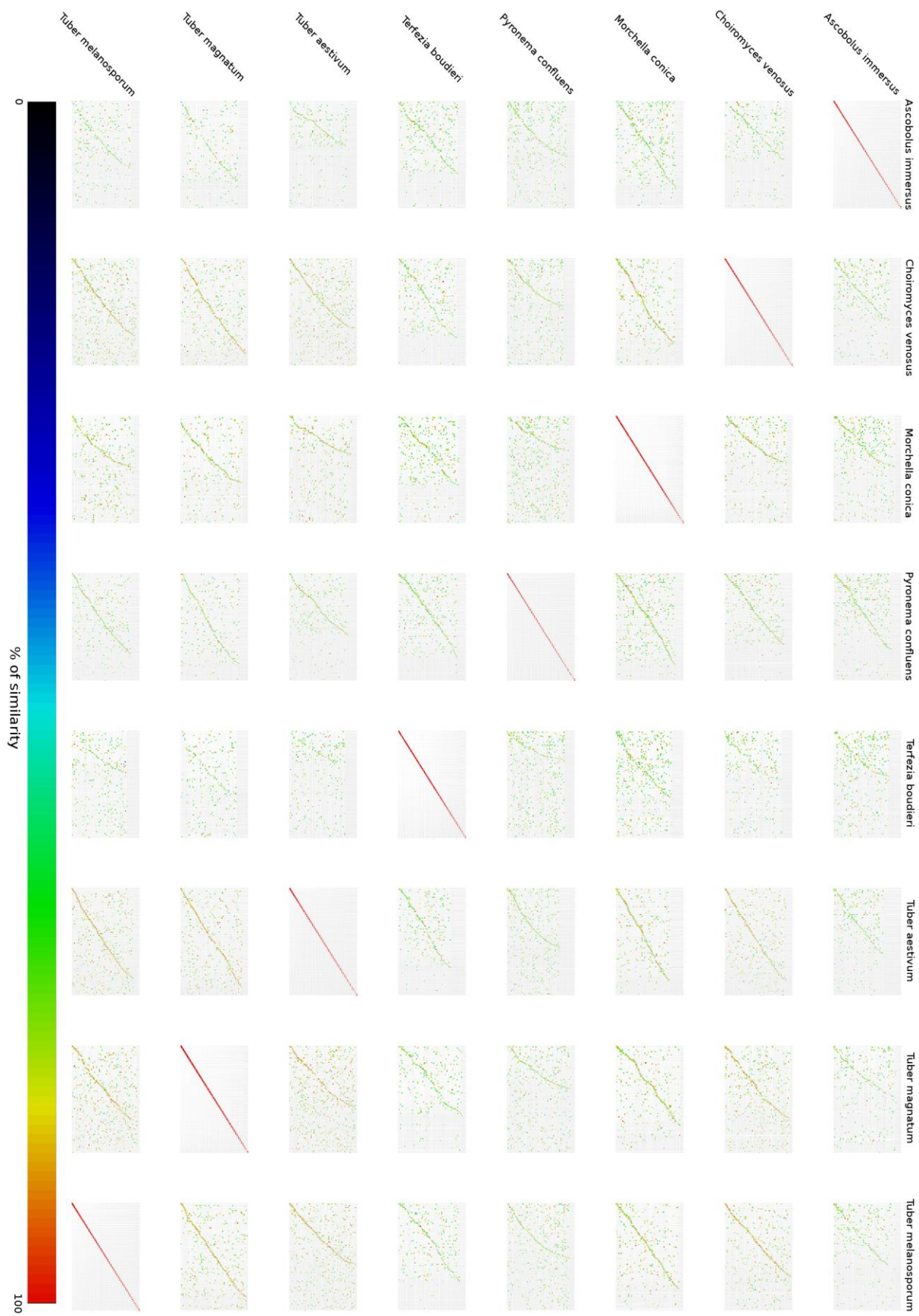


Fig. S5. Dot blot analysis of all the eight genomes based on whole genome nucleotide similarity.

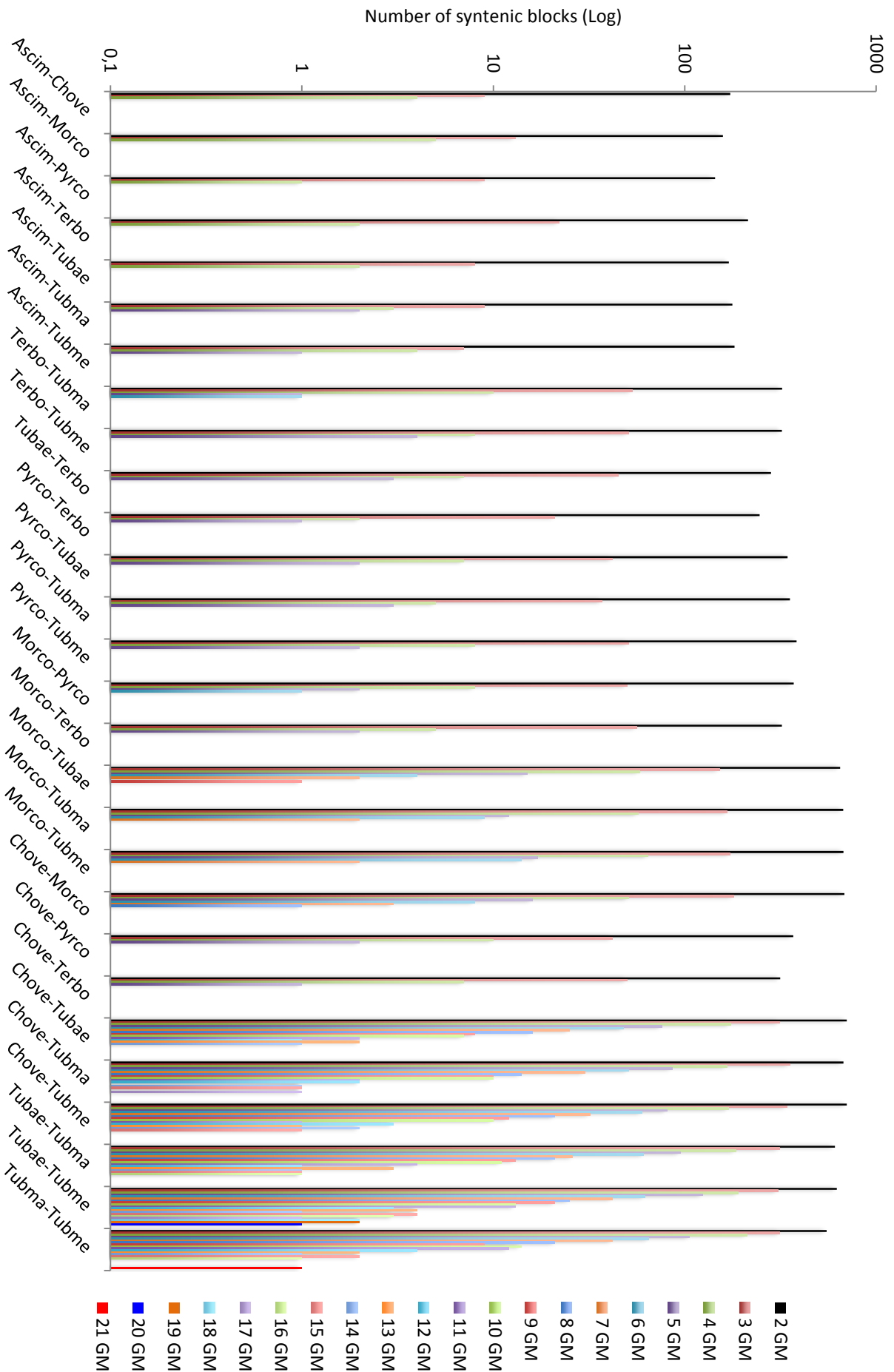


Fig. S6. Number of syntenic blocks for all pairs of species according to the number of gene models in the same order in each genome. The species name abbreviation is available in the Supplementary notes.

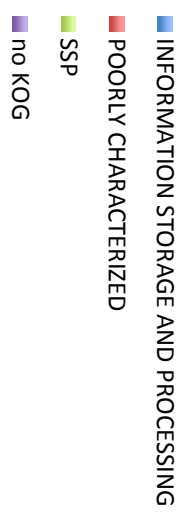
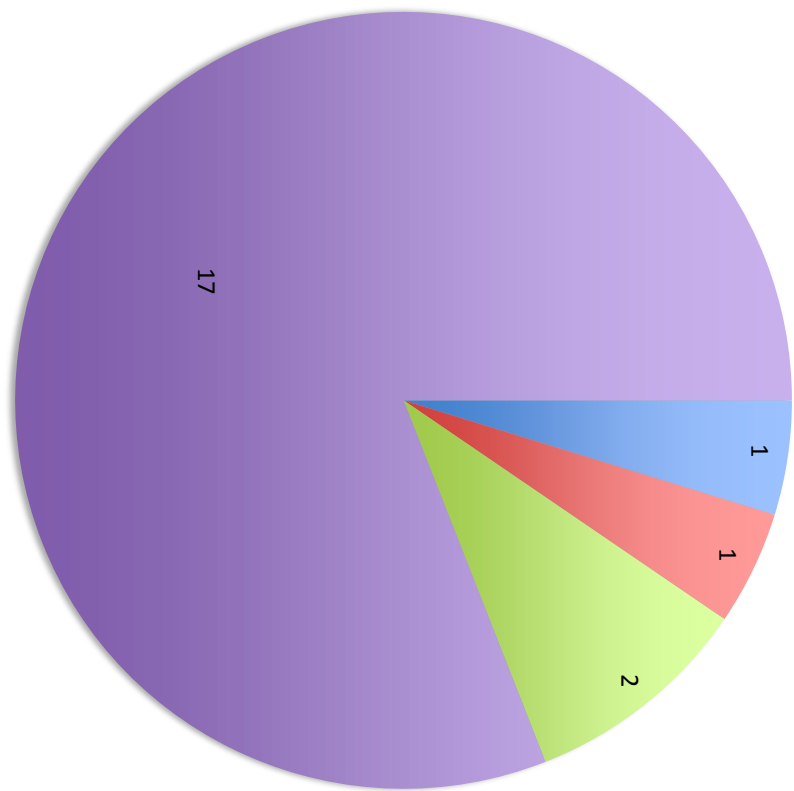


Fig. S7. KOG definition of the 21 gene models present in the largest syntenic block between *T. melanosporum* and *T. magnatum*.

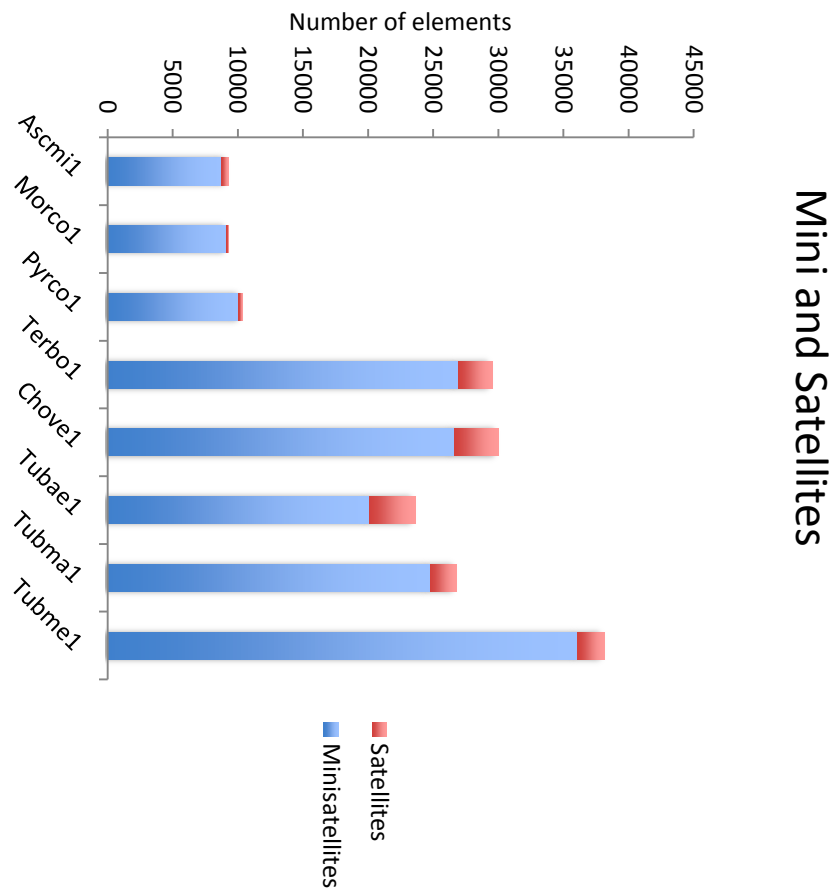
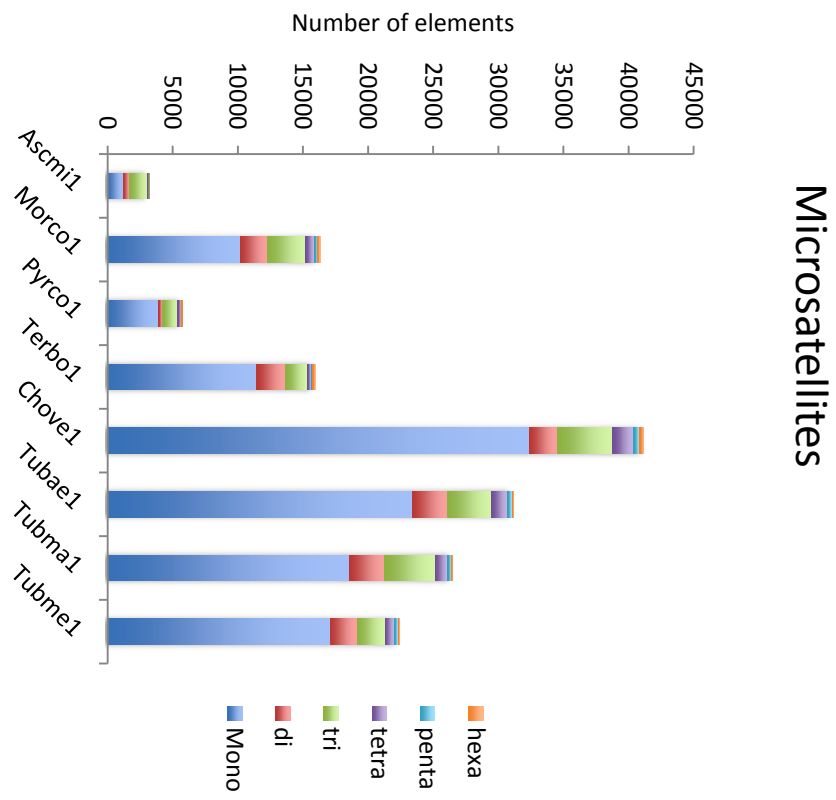


Fig. S8. Number of microsatellites, minisatellites and satellites for the eight Pezizomycetes. The species name abbreviation is available in the Supplementary notes.

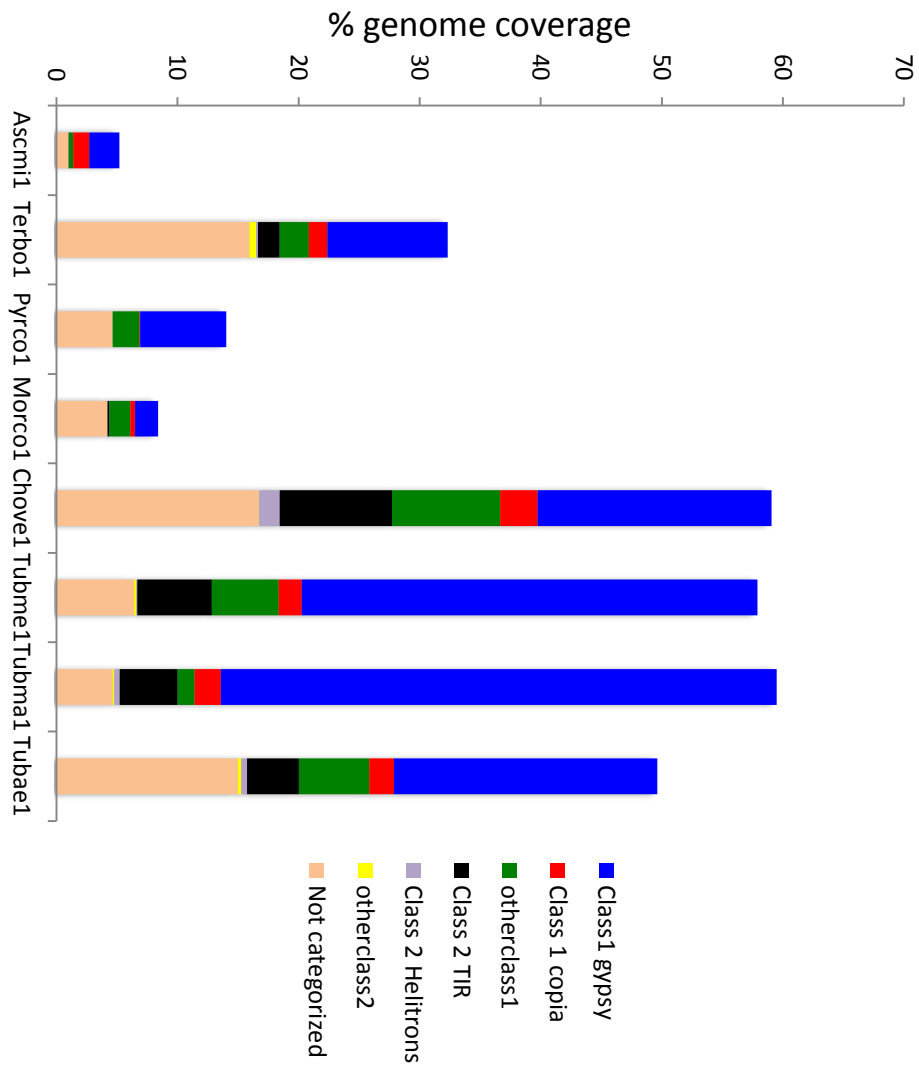


Fig. S9. Proportion of transposable elements in the eight Pezizomycetes. The species name abbreviation is available in the Supplementary notes.

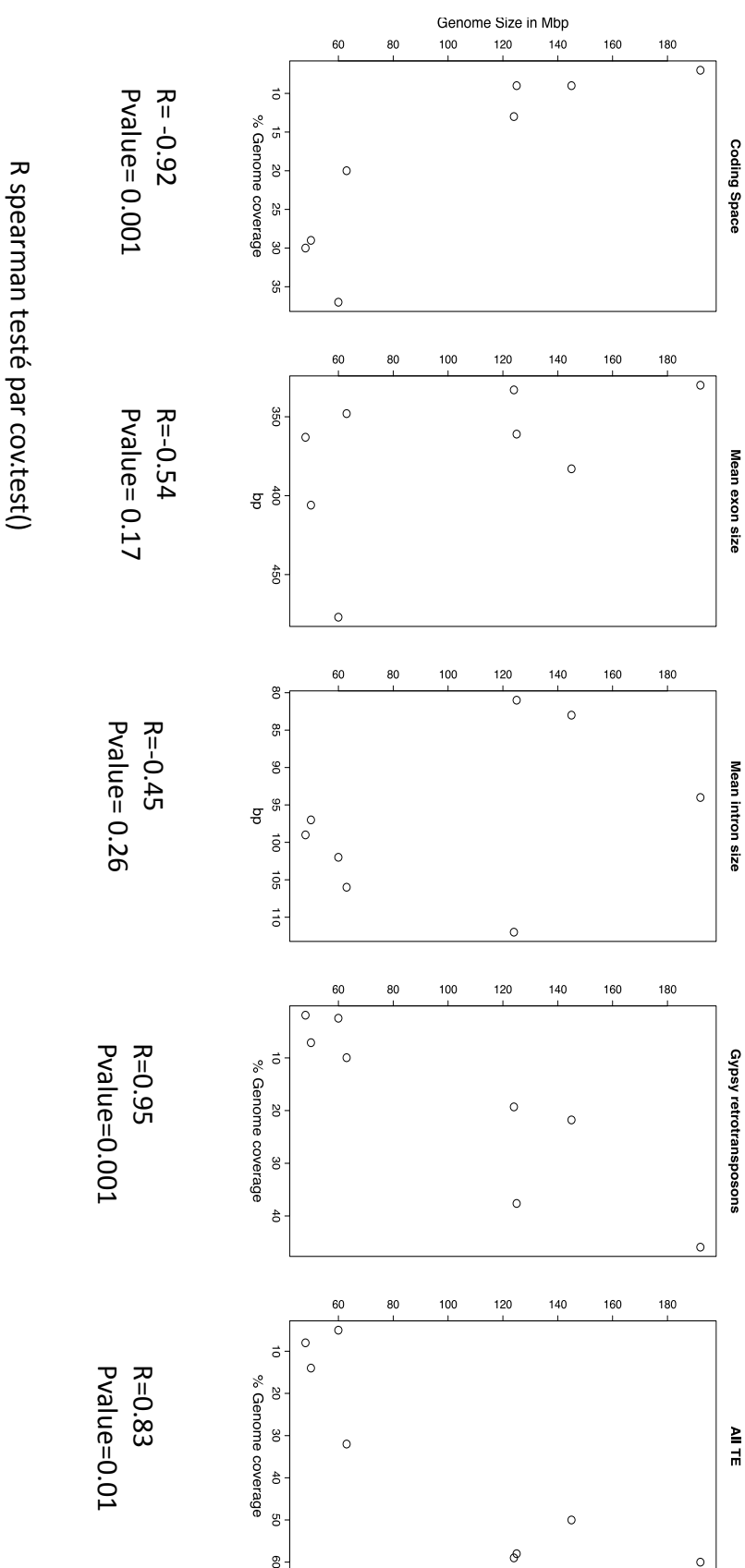


Fig. S10. Plots of the genome size according to the % of coding space, the mean exon size, the mean intron size, the % of gypsy retrotransposons and all transposable elements (TE). The significance of the correlation was tested by Spearman test in R using cov.test() function of stats package.

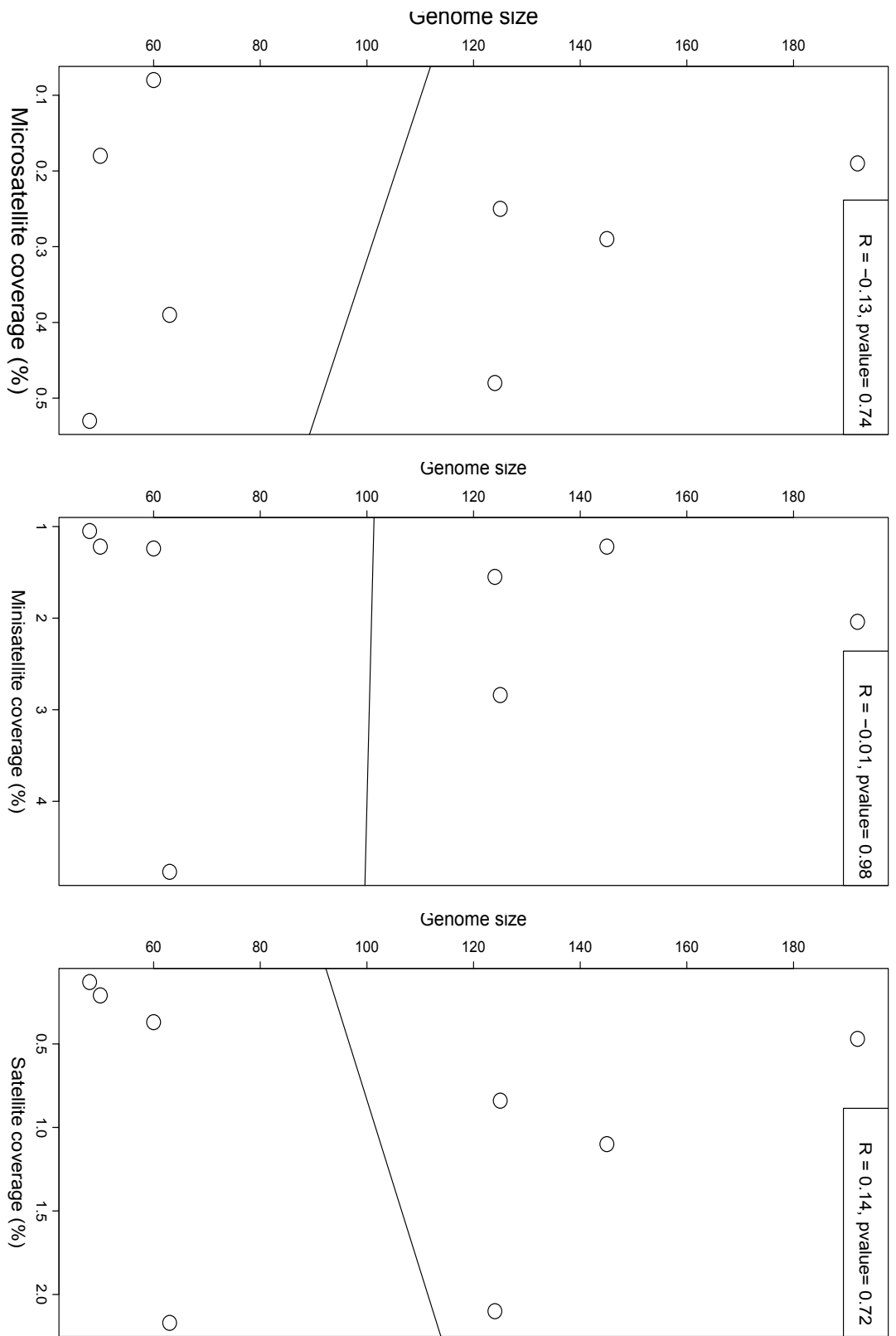


Fig. S11. Plots of the genome size according to the % microsatellites, minisatellites and satellites. The significance of the correlation was tested by Spearman test in R using `cov.test()` function of stats package.

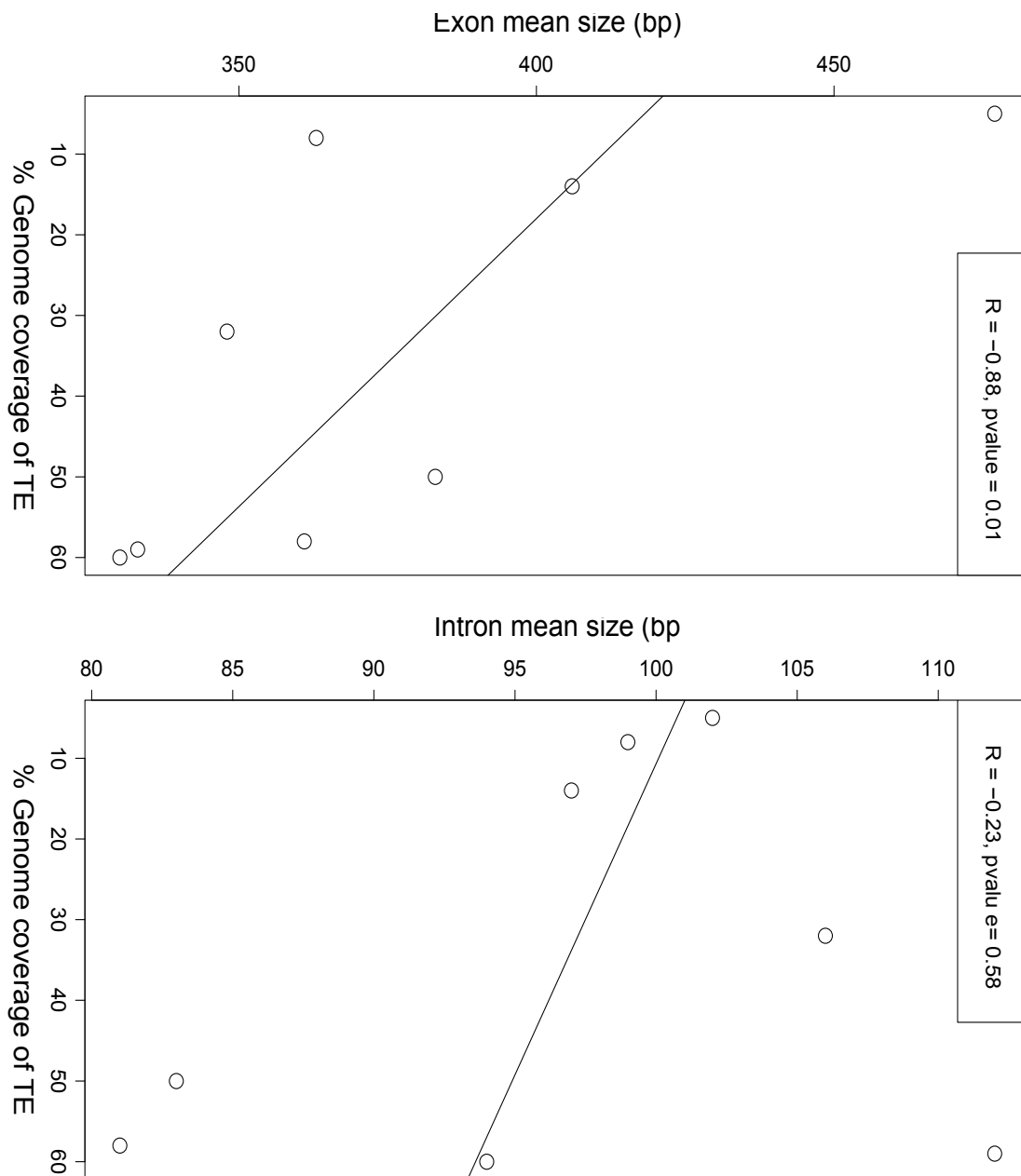


Fig. S12. Plots of the % of genome coverage of transposable elements (TE) and exon mean size and intron mean size. The significance of the correlation was tested by Spearman test in R using `cov.test()` function of stats package.

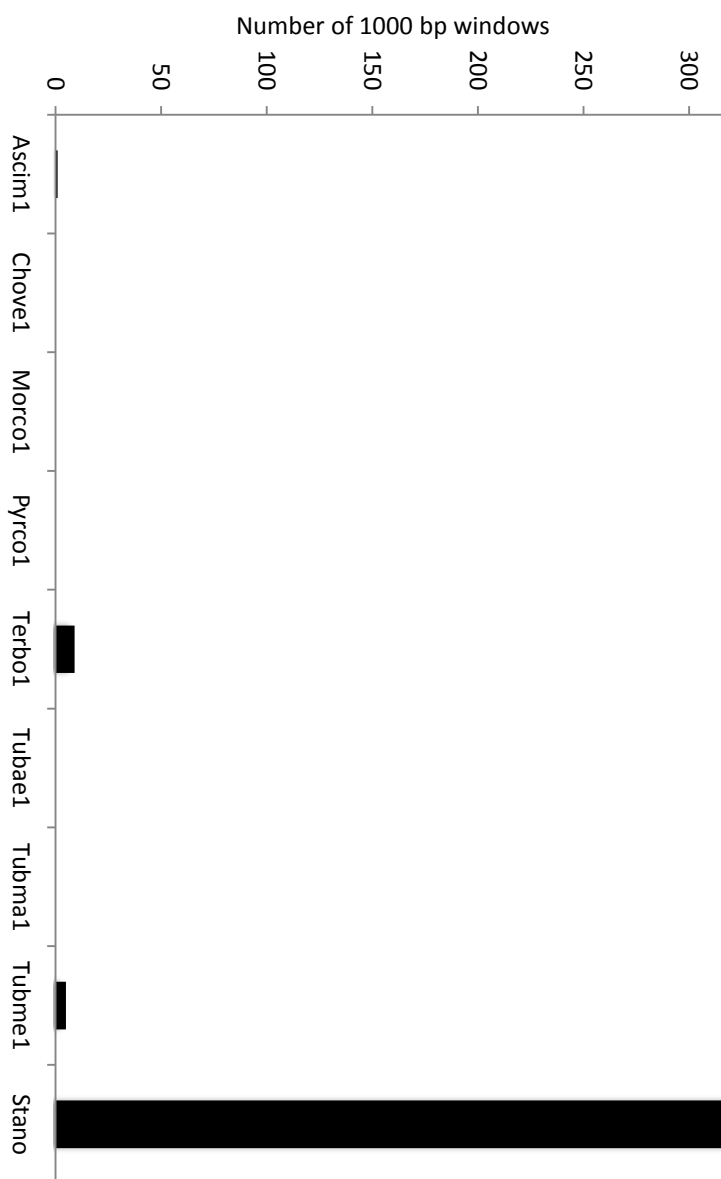


Fig. S13. Estimation of RIP (Repeated Induced Points) indexes in 1000 bp windows. The number of windows with both $TpA/ApT > 2$ and $(CpA+TpG)/(ApC+GpT) < 0.7$ are indicated. *Stagonospora nodorum* was included in this analysis since this species has a RIP defense mechanism.

	SCAFFOLD	MAT 1-1-1	MAT 1-2-1
<i>Tuber melanosporum</i>	247		X
<i>Pyronema confluens</i>	329	X	
	381		X
<i>Choiromyces venosus</i>	159	X	
<i>Morchella conica</i>	48	X	
<i>Tuber aestivum</i>	16		X
<i>Tuber magnatum</i>	4		X
<i>Ascobolus immersus</i>	-	-	-
<i>Terfezia boudieri</i>	-	-	-

Fig. S14. Presence of mating type genes in the eight Pezizomycetes. The number of the scaffold is indicated when found

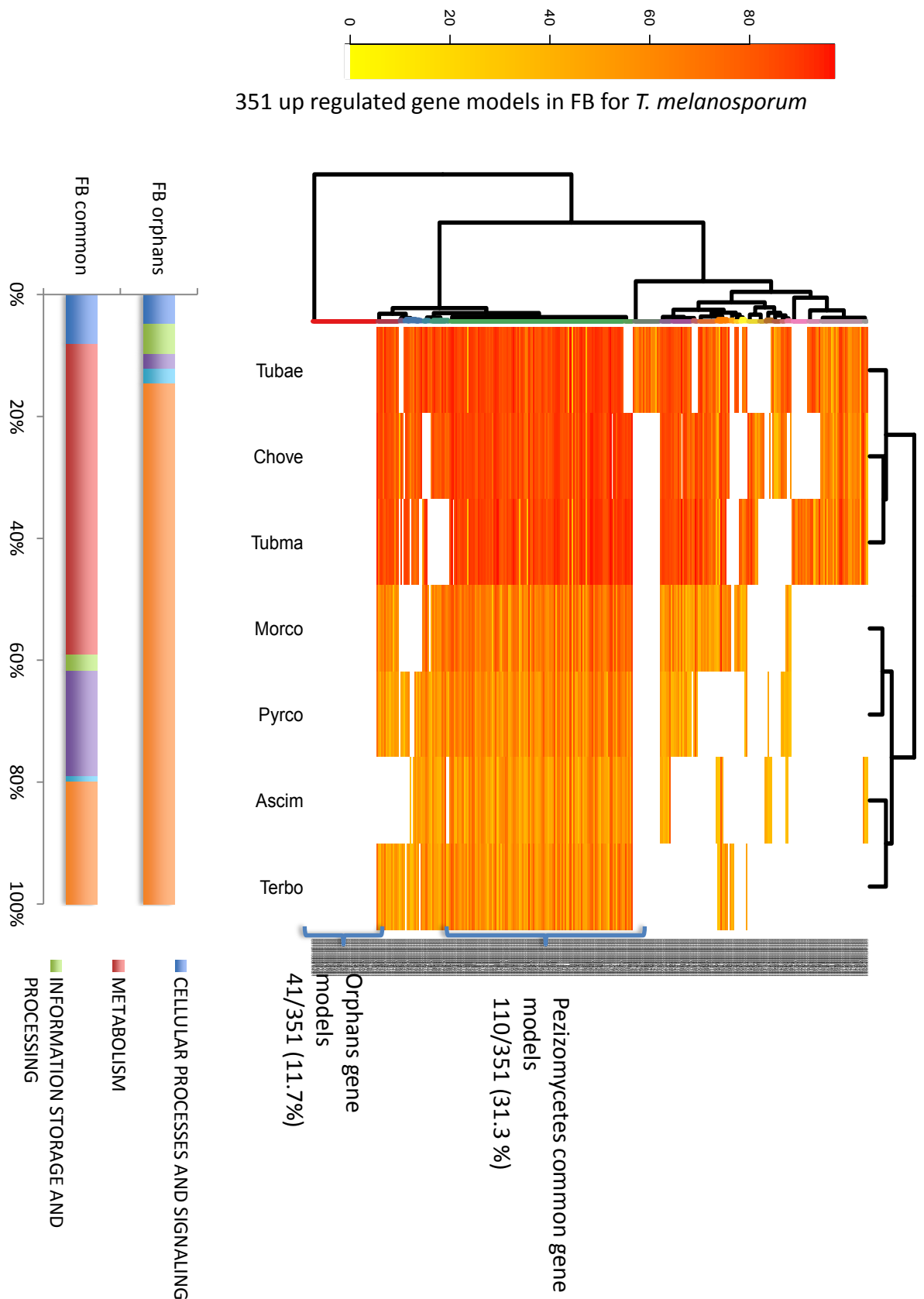


Fig. S15. Presence and sequence similarity of fruiting bodies upregulated genes from *T. melanosporum* in 7 Pezizomycete genomes. The heat map depicts a double-hierarchical clustering of 351 fruiting bodies upregulated *T. melanosporum* genes based on their percentage sequence identity (color scale at left) with their orthologs (if any) in selected fungal species (columns). KOG definitions of *T. melanosporum* orphan genes and Pezizomycete common genes are shown. The species abbreviation is found in the Supplementary notes.

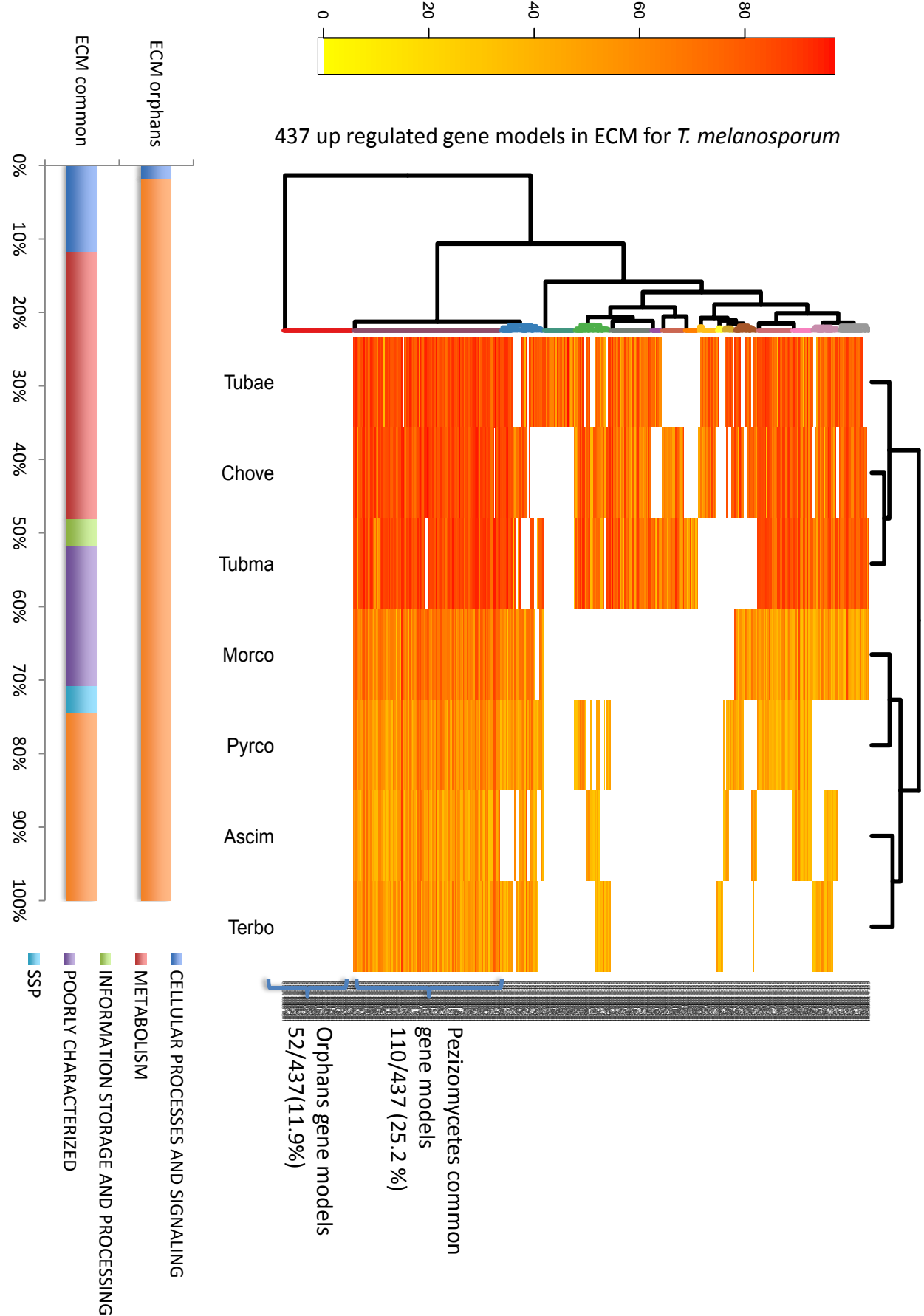


Fig. S16. Presence and sequence similarity of symbiosis-upregulated genes from *T. melanosporum* in 7 Pezizomycete genomes. The heat map depicts a double-hierarchical clustering of 437 symbiosis-upregulated *T. melanosporum* genes based on their percentage sequence identity (color scale at left) with their orthologs (if any) in selected fungal species (columns). KOG definitions of *T. melanosporum* orphan genes and Pezizomycete common genes are shown. The species abbreviation is found in the Supplementary notes.

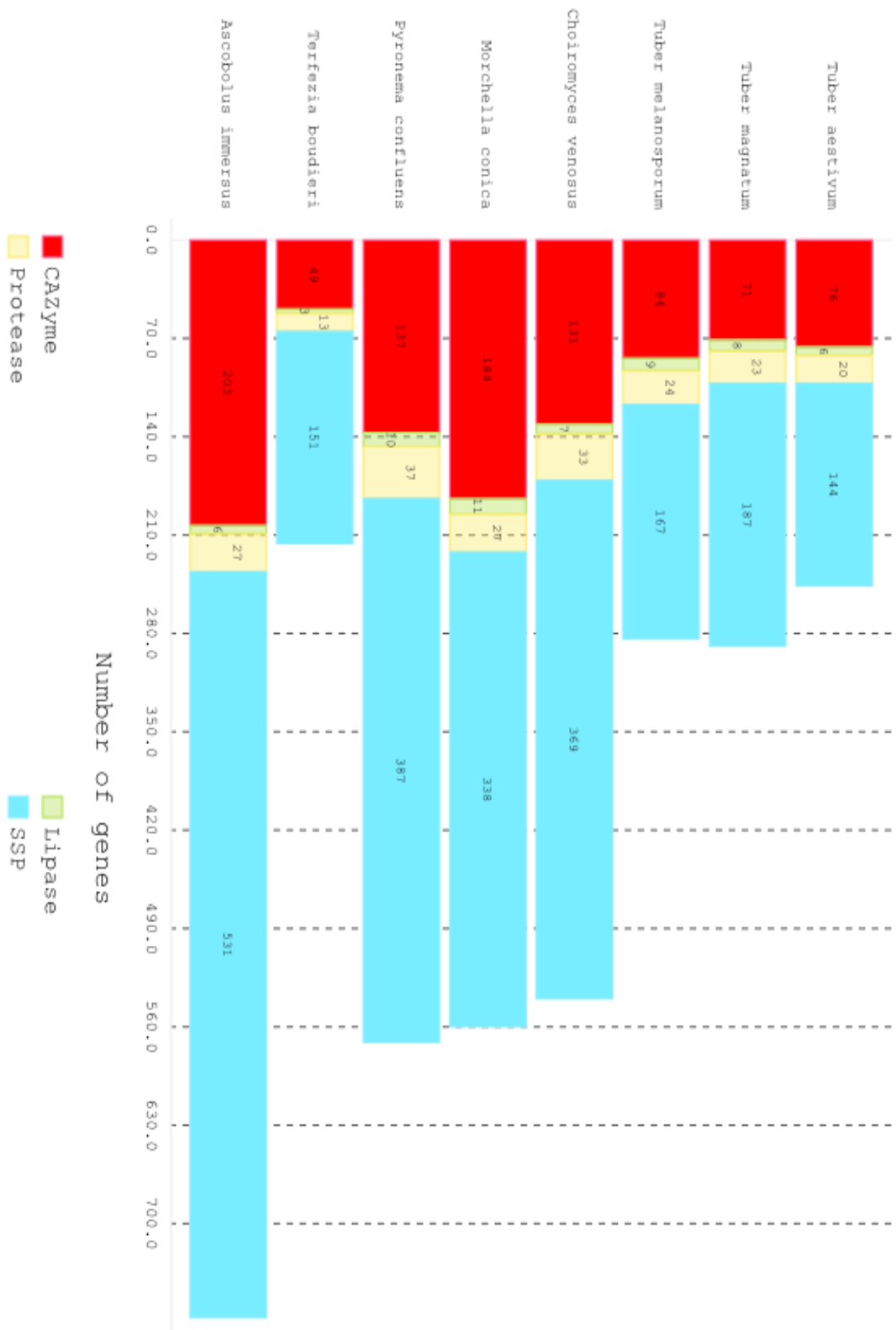
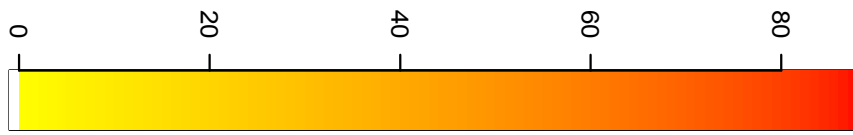


Fig. S17. Secretome of the 8 Pezizomycete species. The number of genes defined as CAZyme, protease, lipase and small secreted proteins (SSP) is indicated.



19 *T. melanosporum* MiSSP

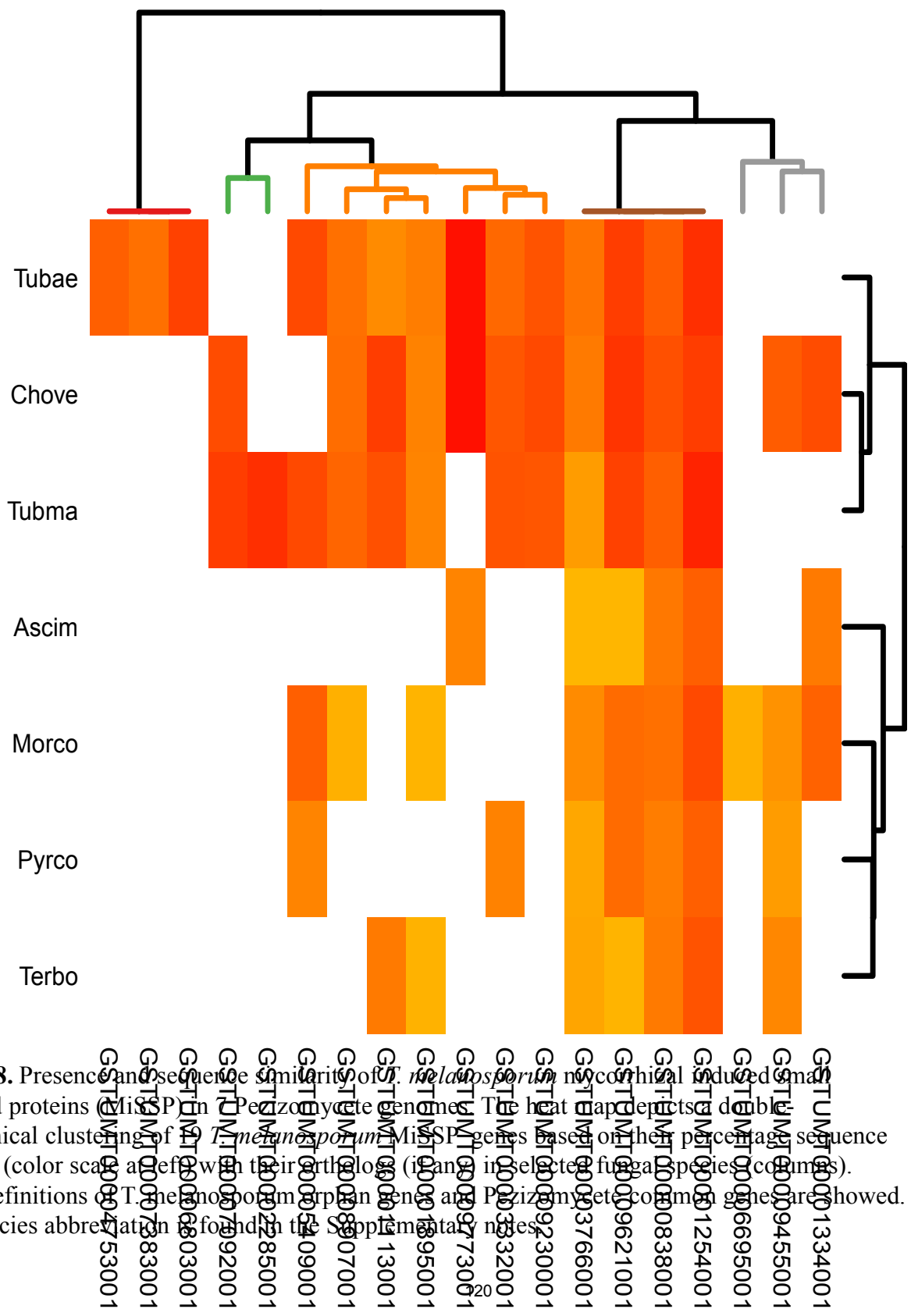


Fig. S18. Presence and sequence similarity of *T. melanosporum* mycelial induced secreted proteins (MiSSP) in 7 Perizomyetes genomes. The heat maps depicts a double hierarchical clustering of *T. melanosporum* MiSSP genes based on their percentage sequence identity (color scale at left) with their orthologs (in red) in selected fungal species (columns). KOG definitions of *T. melanosporum* perianth and Perizomyetes genome genes are shown. The species abbreviation is found in the Supplemental notes.

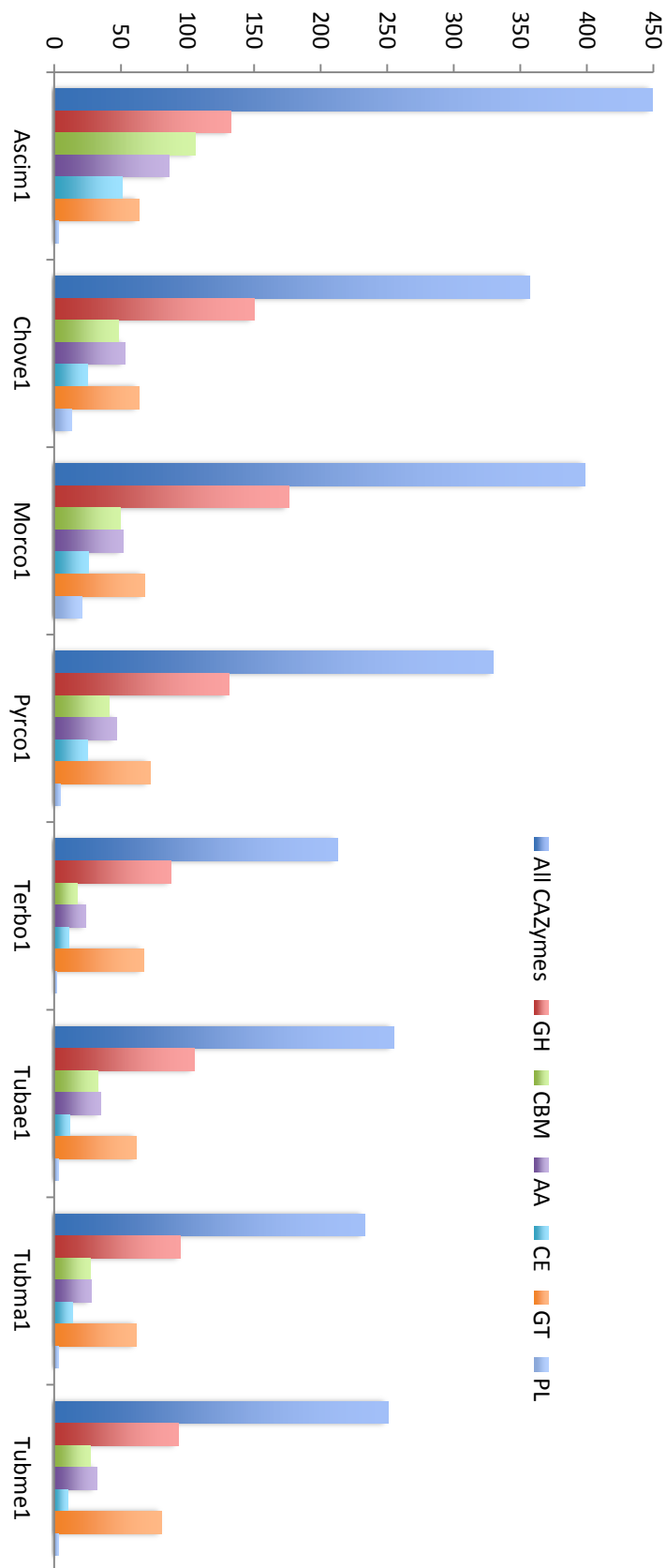


Fig. S19. Number of CAZymes in the 8 Pezizomycetes. GH, glycosyl hydrolase; CBM, carbohydrate-binding modules; AA, auxiliary activities ; CE, carbohydrate esterases; GT, glycosyl transferases ; PL, polysaccharide lyases. The species abbreviation is found in the Supplementary notes.

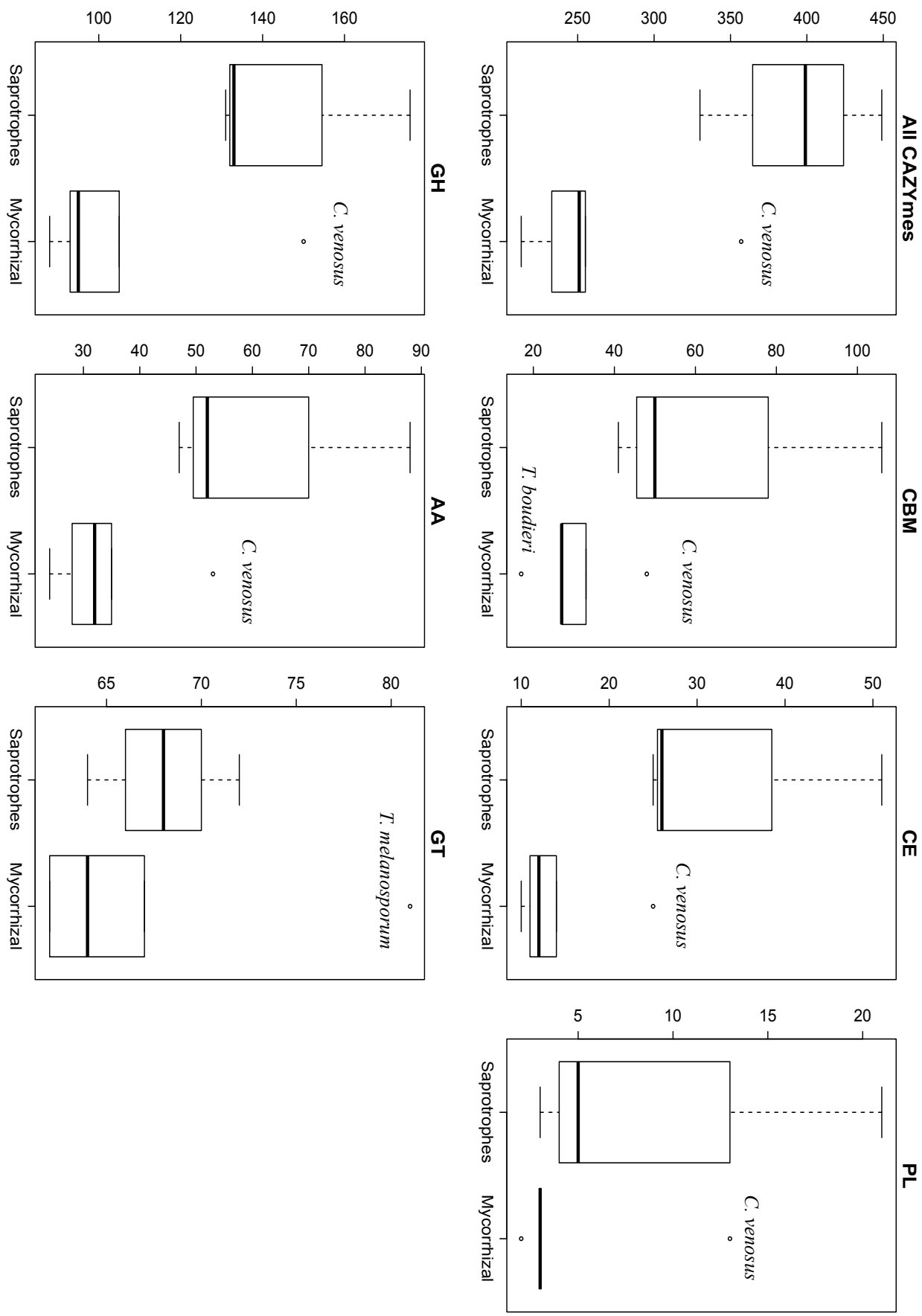


Fig. S20. Number of CAZymes in the different fungal life strategy: mycorrhizal (*C. venosus*, *T. boudieri*, *T. aestivum*, *T. magnatum* and *T. melanosporum*) and saprotroph (*A. immersus*, *M. conica* and *P. confluens*). GH, glycosyl hydrolase; CBM, carbohydrate-binding modules; AA, auxiliary activities ; CE, carbohydrate esterases; GT, glycosyl transferases ; PL, polysaccharide lyases.

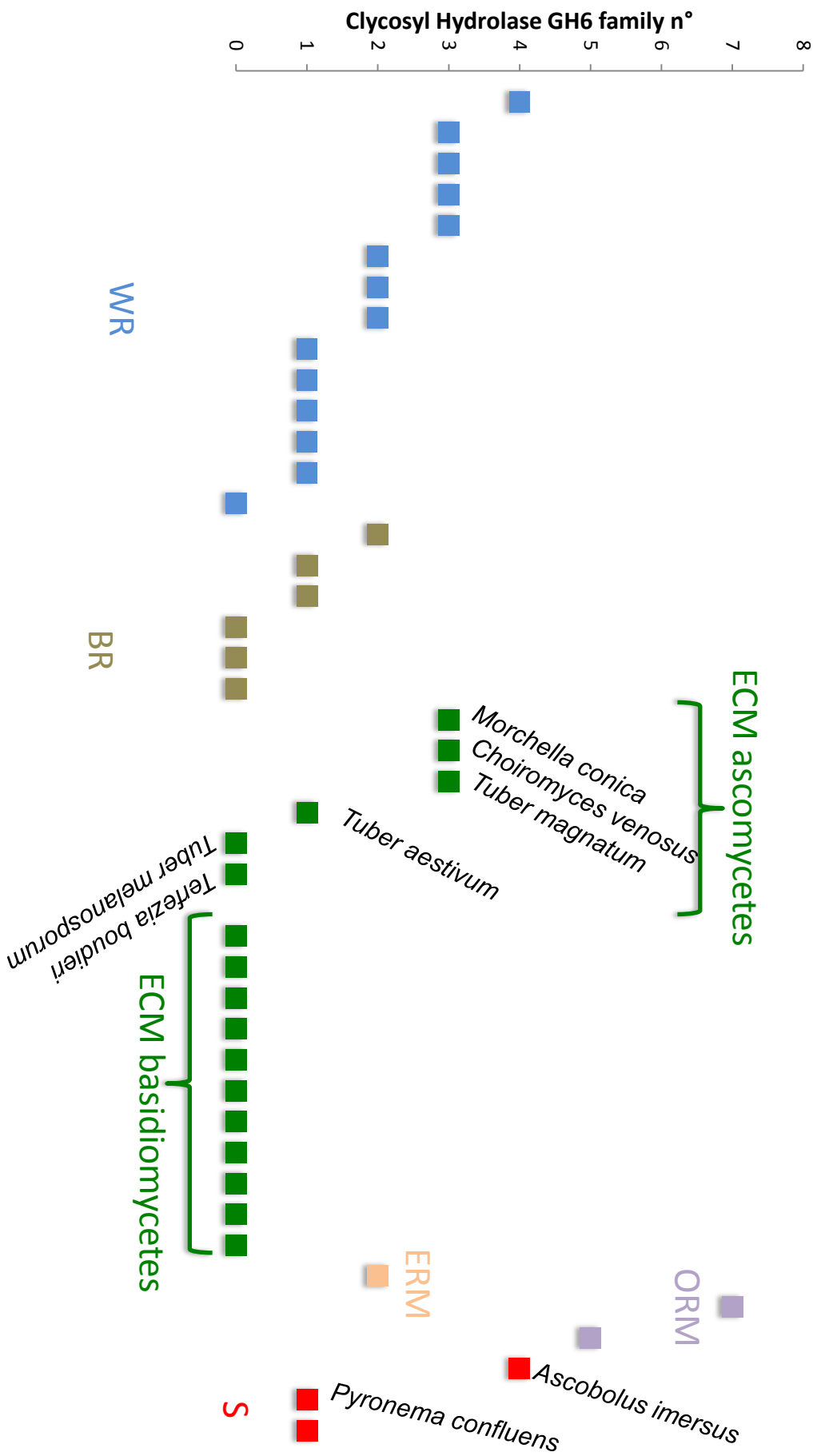


Fig. S21. Number of glycosyl hydrolase 6 (GH6) in white rot (WR), brown rots (BR), ectomycorrhizal (ECM), ericoid mycorrhizal (ERM), orchid mycorrhizal (ORM) and saprotroph (S).

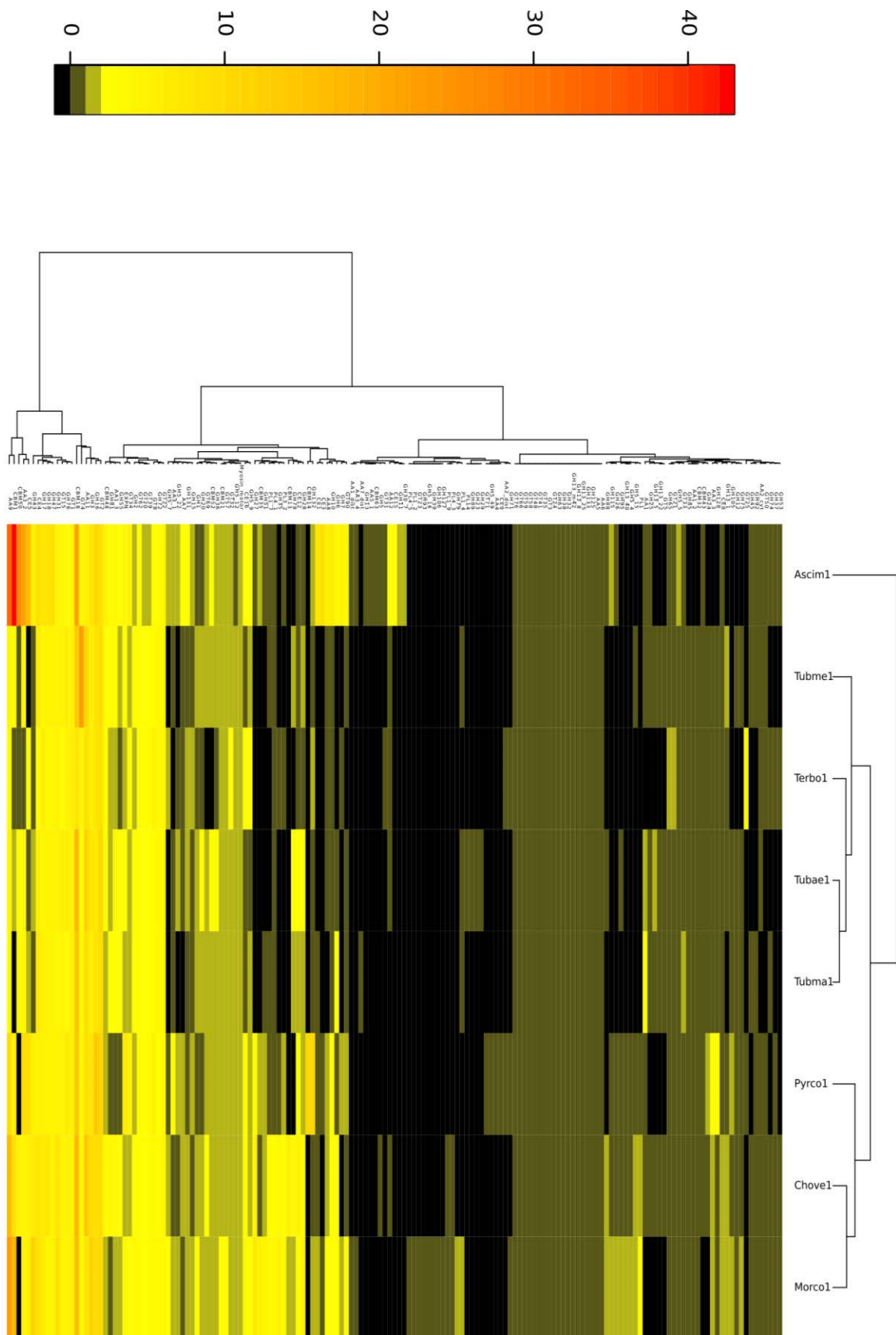


Fig. S22. Hierarchical clustering based on all CAZyme families for the 8 Pezizomycete genomes.

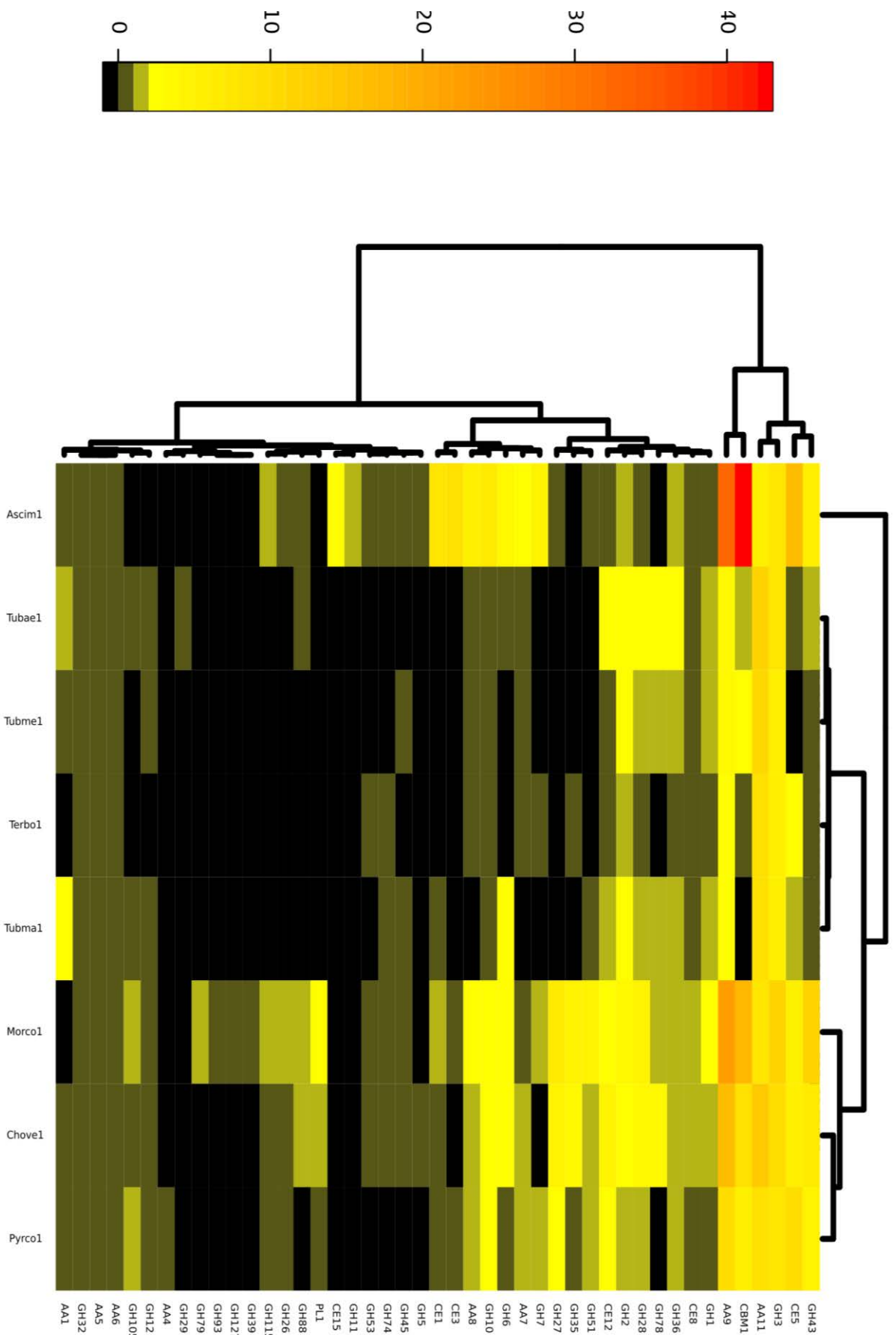


Fig. S23. Hierarchical clustering based on plant cell wall degrading enzymes CAZyme families for the 8 Pezizomycete genomes.

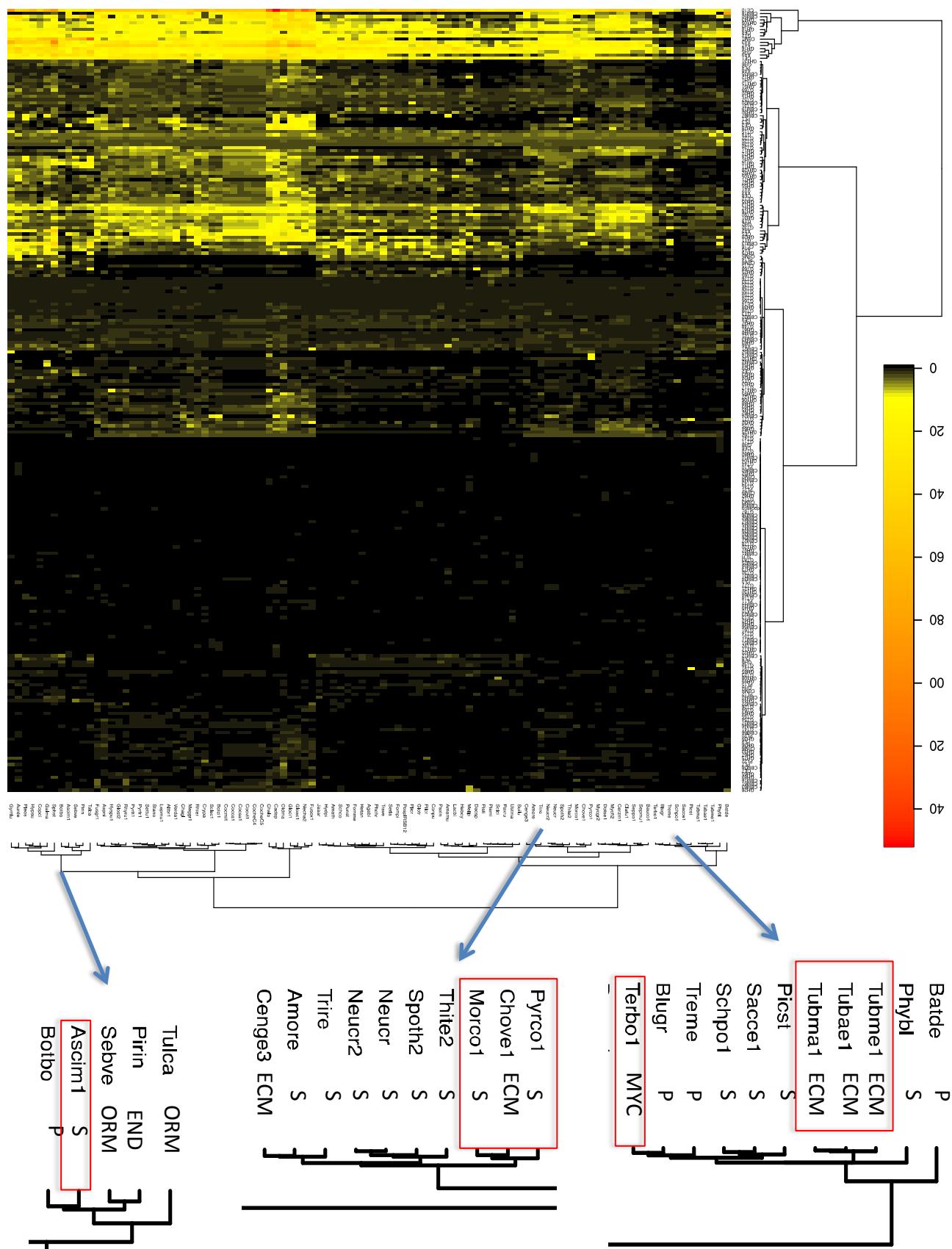


Fig. S24. hierarchical clustering based on all CAZyme families for 103 fungal genomes. The life strategy of the species is indicated: ECM, ectomycorrhizal; END, endophyte; MYC, mycorrhizal; ORM, orchid mycorrhizal; P, parasite; S, saprotroph. The Pezizomycete species are indicated by a red rectangle. Abbreviations for taxon names are defined in supplementary Note. .

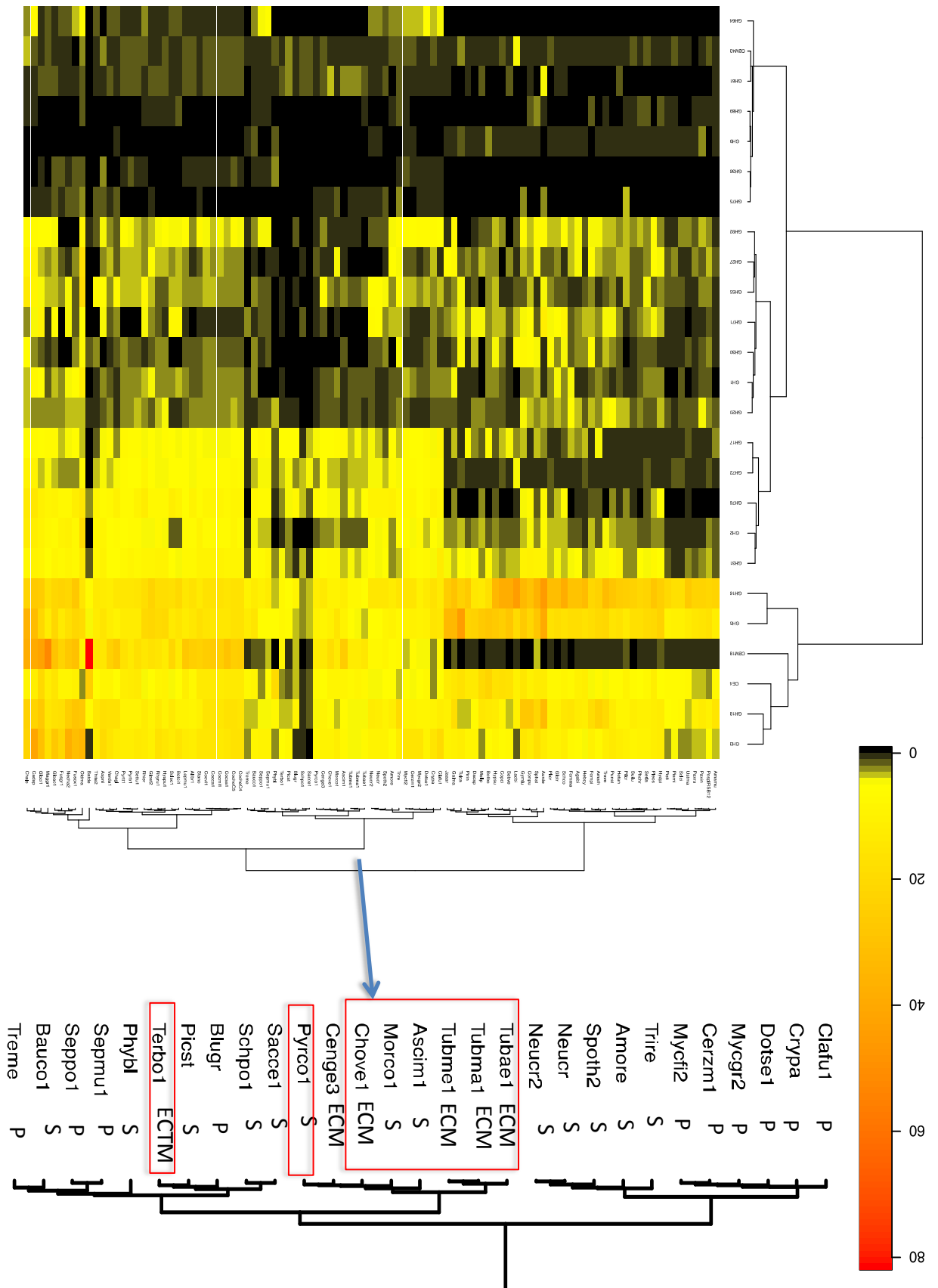


Fig. S25. hierarchical clustering based fungal degrading CAZyme families for 103 fungal genomes. The life strategy of the species is indicated: ECM, ectomycorrhizal; END, endophyte; MYC, mycorrhizal; ORM, orchid mycorrhizal; P, parasite; S, saprotroph. The Pezizomycete species are indicated by a red rectangle. Abbreviations for taxon names are defined in supplementary Note.

5 Etude du polymorphisme génomique de *Tuber melanosporum* par reséquençage d'isolats

Dans ce chapitre je me suis intéressé à l'identification de polymorphismes chez *T. melanosporum* par une approche de génomique comparative intra-spécifique. Ce travail intitulé « **A survey of genome-wide single nucleotide polymorphisms through genome re-sequencing in the Périgord black truffle (*Tuber melanosporum* Vittad.** » a été publié dans le journal « *Molecular Ecology Resources* ».

Le séquençage en 2010 du génome de *T. melanosporum* (Martin et al. 2010) a permis d'avoir un génome de référence. Dans ce travail, les génomes de six isolats ont été reséquencés à une profondeur d'environ 20 X par une approche de séquençage à haut débit Illumina. Afin d'augmenter les chances de trouver du polymorphisme nous avons choisi les isolats dans des populations très contrastantes en terme de climat mais aussi de localisation géographique. Nous avons sélectionné des truffes provenant d'Alsace, de Bourgogne et des Hautes-Alpes en France, de l'Ombrie en Italie et d'Albacete en Espagne.

Nous avons ensuite comparé ces génomes reséquencés au génome de référence nous permettant d'identifier un *core* génome d'environ 110 Mbp. Un total de 442,326 SNPs (*Single Nucleotide Polymorphism*), ou polymorphismes simple base, ont été identifiés tout au long du génome. Notre étude démontre que la truffe a un niveau de polymorphisme significatif et similaire à celui d'autres champignons filamenteux. L'analyse de ces polymorphismes a permis d'identifier des régions ayant un grand nombre de mutations, en particulier les éléments transposables. Au total, 2587 gènes présentent des polymorphismes dans leurs séquences codantes. En se focalisant sur les 119 gènes ayant plus de 5 SNPs dans leurs séquences codantes, nous avons identifié ceux ayant un biais entre les mutations synonymes

(pS) et non synonymes (pN). Dix-huit gènes présentent un ratio pN/pS >1 et 9 ont seulement des mutations non-synonymes suggérant qu'ils sont sous sélection positive. D'autre part, 78 gènes ont un ratio pN/pS <1 suggérant qu'ils sont sous sélection purificatrice. En parallèle, nous avons aussi réalisé un balayage du génome par le calcul de l'indice de Tajima. Cet indice met en évidence des biais dans la distribution des allèles au niveau des SNPs et il permet d'identifier les régions potentiellement sous sélection. Ceci nous a permis d'identifier 36 fenêtres de 10000 pb potentiellement sous sélection positive. En conclusion, nous pouvons dire que des traces de sélection sont présentes dans le génome de *T. melanosporum*, toutefois notre échantillonnage était trop limité pour identifier les facteurs à l'origine de cette sélection. Cette étude a également permis la création d'une ressource de référence de polymorphismes qui pourra être utilisée dans le futur pour caractériser les populations de *T. melanosporum* mais aussi pour rechercher les déterminants génomique de certains traits de vie intéressant liés par exemple aux qualités organoleptiques et à la tolérance aux stress environnementaux.

A survey of genome-wide single nucleotide polymorphisms through genome resequencing in the Périgord black truffle (*Tuber melanosporum* Vittad.)

THIBAUT PAYEN,*†^a CLAUDE MURAT,*†^a ANAÏS GIGANT,*† EMMANUELLE MORIN,*†
STÉPHANE DE MITA*† and FRANCIS MARTIN*†

*INRA, Laboratoire d'Excellence ARBRE, UMR1136 Interactions Arbres-Microorganismes, F-54280 Champenoux, France,

†UMR1136 Interactions Arbres-Microorganismes, Université de Lorraine, Vandoeuvre-lès-Nancy F-54500, France

Abstract

The Périgord black truffle (*Tuber melanosporum* Vittad.), considered a gastronomic delicacy worldwide, is an ectomycorrhizal filamentous fungus that is ecologically important in Mediterranean French, Italian and Spanish woodlands. In this study, we developed a novel resource of single nucleotide polymorphisms (SNPs) for *T. melanosporum* using Illumina high-throughput resequencing. The genome from six *T. melanosporum* geographical accessions was sequenced to a depth of approximately 20×. These geographical accessions were selected from different populations within the northern and southern regions of the geographical species distribution. Approximately 80% of the reads for each of the six resequenced geographical accessions mapped against the reference *T. melanosporum* genome assembly, estimating the core genome size of this organism to be approximately 110 Mbp. A total of 442 326 SNPs corresponding to 3540 SNPs/Mbps were identified as being included in all seven genomes. The SNPs occurred more frequently in repeated sequences (85%), although 4501 SNPs were also identified in the coding regions of 2587 genes. Using the ratio of nonsynonymous mutations per nonsynonymous site (pN) to synonymous mutations per synonymous site (pS) and Tajima's D index scanning the whole genome, we were able to identify genomic regions and genes potentially subjected to positive or purifying selection. The SNPs identified represent a valuable resource for future population genetics and genomics studies.

Keywords: black truffle, genome resequencing, next-generation sequencing, SNP resource

Received 16 July 2014; revision received 9 February 2015; accepted 13 February 2015

Introduction

Single nucleotide polymorphisms (SNPs) have attracted a great deal of interest in the scientific community (Ganal *et al.* 2009). Already widely used for human and plant genetics, due to technological developments and the subsequent reduced costs of high-throughput sequencing technologies, large-scale SNP identification is also available for filamentous fungi (e.g. *Fusarium graminearum*, Cuomo *et al.* 2007; *Coccidioides* spp., Neafsey *et al.* 2010; *Leptographium longiclavatum*, Ojeda *et al.* 2014; *Blumeria graminis*, Wicker *et al.* 2013; *Rhizoctonia solani*, Hane *et al.* 2014). The value of SNPs in comparison with microsatellite markers lies in their abundance throughout the genome, their biallelic nature and their high potential for automation (Brumfield *et al.* 2003). In addition, they are

easy to study, follow relatively robust mutation models and are easily genotyped in large panels. SNPs can be used in population analyses for studying demographic and historical patterns based on large numbers of samples. Compared to microsatellites, which are essentially neutral markers that do not permit a high density, an advantage of surveying SNPs throughout the genome is the higher probability of identifying adaptation signatures. For example, a large-scale SNP analysis uncovered an adaptation signature for temperature in *Neurospora crassa* (Ellison *et al.* 2011).

In *Tuber melanosporum*, the genetic diversity and phylogeography have been investigated with different molecular markers such as randomly amplified polymorphism DNA, microsatellites, SNPs in the internal transcribed spacer (ITS) of ribosomal DNA and inter-simple sequence repeats (ISSR; Bertault *et al.* 1998; Murat *et al.* 2004; Riccioni *et al.* 2008; García-Cunchillos *et al.* 2014). These studies pointed to an important effect of the last glaciations (from 120 000 to 11 000 years ago; Van Andel

Correspondence: Claude Murat, Fax: +33 (0)3 83 39 40 69;
E-mail: claudemurat@nancy.inra.fr

^aThese authors contributed equally to this work.

& Tzedakis 1996) on the truffle population structure. For example, two putative post-glacial recolonization routes were hypothesized using 10 SNPs in ITSs (Murat *et al.* 2004). Through the use of microsatellites and ISSR fingerprinting, it has also been suggested that glacial refuges exist in Italy and Spain (Riccioni *et al.* 2008; García-Cunchillos *et al.* 2014).

Following the publication of the *T. melanosporum* genome sequencing project (Martin *et al.* 2010), this species became one of the model species for studying ectomycorrhizal ascomycetes (Kües & Martin 2011). Ectomycorrhizal fungi are an important group of fungi, because they promote the growth of trees in forests and woodlands by providing the trees with water and nutrients (Smith & Read 2010). Using the *T. melanosporum* genome, highly polymorphic microsatellite makers were developed and used to characterize small-scale spatial genetic diversity in two truffle orchards, identifying a pronounced spatial genetic structure with numerous small-sized genets (Murat *et al.* 2013b). These results suggested that *T. melanosporum* relies heavily on sexual reproduction. Microsatellite-based population genetic analyses have allowed for the investigation of a small proportion of the *T. melanosporum* genome, but with a lower probability of detecting the genomic regions involved in the species' adaptation. Using the whole *T. melanosporum* genome in combination with high-throughput sequencing technologies, large-scale SNP surveys now make it possible to perform an exhaustive investigation of the genomic variation.

The aim of this study was to assess the overall genetic diversity of *T. melanosporum* by identifying and mapping the SNPs in resequenced genomes. The genomes of six geographical accessions of *T. melanosporum* were sequenced using Illumina technology and compared to strain Mel28 as the reference genome (Martin *et al.* 2010). To improve the chances of finding genetic polymorphisms, the resequenced geographical accessions were from samples in different populations within the

northern and southern geographical limits of the species distribution. This SNP resource will be useful for more in-depth investigations of *T. melanosporum* population structure, gene flow and putative ecotype identifications, as well as of selected genes and genomic regions.

Materials and Methods

Sampling and DNA extraction

Tuber melanosporum Vittad. (Ascomycota, Pezizomycotina, Pezizomycetes, Pezizales, Tubercaceae) is native to France, Italy and Spain. Our sampling strategy aimed to cover the natural geographical range of the species as well as the different climates (Mediterranean and continental) where this truffle is produced. The Mel28 isolate (referred to as France-Pro in this study) used for sequencing the reference genome was harvested in southern France (Saint Rémy de Provence, Bouches du Rhône, France; Martin *et al.* 2010). Three *T. melanosporum* were harvested from France (Alps, Burgundy and Alsace), one from Italy (Umbria) and two from Spain (Castilla Leone; Fig. S1; Table 1). Within a few days of harvesting, each ascocarp was shipped to the laboratory, thoroughly washed and the inner section (i.e. gleba) conserved at -20°C pending the DNA extraction.

Total DNA was extracted from 500 mg of gleba using a modified CTAB (cetyl trimethyl ammonium bromide) protocol. After grinding the samples in liquid nitrogen, they were incubated for 30 min at 65°C in 2.5 volumes of buffer A (0.35 M sorbitol; 0.1 M Tris-HCl, pH 9; and 5 mM EDTA, pH 8), 2.5 volumes of buffer B (0.2 M Tris-HCl, pH 9; 50 mM EDTA, pH 8; 2 M NaCl; and 2% CTAB) and 1 volume of buffer C (5% of Sarkosyl; N-lauroylsarcosine sodium salt) in 50-mL Falcon tubes. After the incubation, a 0.33 volume of potassium acetate (5 M) was added, and the tubes were incubated for 30 min on ice to precipitate the polysaccharides. After centrifugation at $5000 \times g$ for 20 min, the supernatant was purified with

Table 1 List of *Tuber melanosporum* samples analysed, their geographical origin and the climate of their sampling area

Sample name	Code in manuscript	Locality	Region	Country	Climate	Sequence origin
091215-1	Spain-1	Sierra de Alcaraz	Castilla Leone	Spain	Mediterranean	This study
091215-4	Spain-2	Sierra de Vianos	Castilla Leone	Spain	Mediterranean	This study
100104-1	France-Bur	Courban	Burgundy	France	Continental	This study
100120-1	France-Als	Rouffach	Alsace	France	Continental	This study
100122-1	Italy	Perugia	Umbria	Italy	Mediterranean	This study
100303-4	France-Alp	Chorges	Provence-Alpes-Côtes d'Azur	France	Alps (> 1000 m altitude)	This study
mel28	France-Pro	St Rémy de Provence	Provence-Alpes-Côtes d'Azur	France	Mediterranean	Martin <i>et al.</i> (2010)

1/10 volume of ammonium acetate (3 M) and 1 volume of chloroform:isoamyl alcohol (24:1) in a Falcon tube and centrifuged at $4000 \times g$ for 10 min. The aqueous phase was transferred to a Nalgene tube (Fisher Scientific, France) and incubated with 100 μ L RNase A (10 mg/mL) for 30 min at 37 °C. The DNA was precipitated with a 1/10 volume of ammonium acetate (3 M) and 1 volume of isopropanol at room temperature for 5 min and centrifuged at $10\,000 \times g$ for 10 min. The pellet was suspended in 2 mL of QBT buffer (0.75 M NaCl; 50 mM MOPS, pH 7.0; 15% isopropanol; and 0.15% Triton X-100) and purified using Genomic-tip 100/G columns (Qiagen Cat# 10243) following the manufacturer instructions with the exception of the QC buffer (1.35 M NaCl; 50 mM MOPS, pH 7.0; 15% isopropanol). The higher NaCl (1.35 M instead of 1 M) concentration in the QC buffer allowed for the exclusion of small DNA fragments from the column. The purified DNA was then concentrated by precipitation with 1/10 volume of ammonium acetate (3 M) and 1 volume of isopropanol at room temperature for 5 min and centrifuged at $10\,000 \times g$ for 10 min. After discarding the supernatant, the pellet was resuspended in 100 μ L of TE buffer and stored at -20 °C.

Riccioni *et al.* (2008) showed that the gleba of *T. melanosporum* is formed by a haploid maternal mycelium. The DNA isolated from each ascocarp by the described protocol is therefore expected to correspond to a haploid mycelium, as no disrupted spores were observed when checked under a microscope (data not shown).

Whole-genome shotgun sequencing and mapping

Each of the six geographical accession DNAs was sequenced in one lane of an Illumina Genome Analyzer

(GAII) at the Beckman Genomics facilities (Brea, CA, USA). Sequencing produced approximately 1.1 Gb of 76-bp single-end reads per sample, and the sequencing depth ranged from 21- to 24-fold (Table 2). The raw data reads can be accessed in the sequence reads archive at the National Center for Biotechnology (NCBI) under the Accession No SRP044130.

The raw reads for each genome were aligned to the France-Pro reference genome available at the Institut National de la Recherche Agronomique (INRA) *Tuber* genome database (<http://mycor.nancy.inra.fr/IMGC/TuberGenome/download.php?select=fast>) using the Burrow-Wheeler Aligner (BWA) software, version 0.7.3a (Li & Durbin 2009); with the exception of the number of mismatches between a read and the reference genome, which was set to two, the *aln/samse* algorithm and default parameters were used. As BWA generates a read mapping quality in phred scaled (MAPQ) according to the read quality, the raw reads were not quality filtered before mapping. To avoid low-quality mapping, only reads mapped with an MAPQ above 25 were considered for analysis with SAMtools (v. 0.1.18; Li *et al.* 2009). This stringent parameter eliminated the reads of low sequencing quality and those mapped at several genomic locations, thereby avoiding problems in the SNP calling due to the higher proportion of repeated sequences such as transposable elements (TE) in the *T. melanosporum* genome (Martin *et al.* 2010). The genomic regions without mapped reads were assigned to the different genomic compartments (i.e. genes, TE and intergenic regions) defined by Martin *et al.* (2010). We considered genes with more than 60% of their sequences without mapped reads as missing.

In this study, we produced an updated gene model repertoire of the *T. melanosporum* genome. Indeed,

Table 2 Illumina sequencing and Burrow-Wheeler Aligner (BWA) mapping statistics

Samples	Read number	Total mapped reads after filtering*		Reads mapping to multiple locations†		Genome reference coverage‡		Number of genes§
		Read Number	%	Number of reads	%	Number of bp	%	
Spain-1	39 275 496	28 744 980	73.19	2 376 408	6.05	113 139 168	91.5	9816
Spain-2	38 003 850	29 780 727	73.62	2 419 174	6.37	113 282 964	91.7	9808
France-Bur	38 921 450	31 145 093	80.02	2 427 651	6.24	114 066 611	92.3	9800
France-Als	34 575 334	26 833 541	77.61	2 073 232	6.00	114 998 792	93.1	9831
Italy	39 184 077	29 780 081	76.00	2 323 320	5.93	114 004 497	92.3	9807
France-Alp	38 597 308	30 651 292	79.41	2 474 195	6.41	113 860 556	92.2	9792

*Excluding low-quality reads and reads mapping to multiple locations.

†These reads were eliminated for the SNP identification.

‡Number of base pairs and the percentage of the reference genome mapped by reads. Excluding the Ns, the reference genome is composed of 123 535 220 bp.

§Number of gene models mapped for each genome. A gene model was considered present if at least 60% of its sequence was covered by reads. The total number of gene models in the reference genome is 9952 (9765 are unique to France-Pro).

transcriptomic analyses have suggested that the 7496 high-confidence protein-coding genes supported by either sequence similarity, the occurrence of Pfam or KOG domains, or oligoarray expression data (Martin *et al.* 2010; available at <http://mycor.nancy.inra.fr/IMGC/TuberGenome/index.html>) omitted several expressed genes (Tisserant *et al.* 2011; A. Kohler, E. Tisserant and F. Martin, unpublished data). Moreover, we could not exclude that gene model families were considered as repeated sequences and excluded from this high-confidence protein-coding gene repertoire. To update the gene model repertoire, we began with the initial 12 826 putative gene models identified by GAZE and discarded those gene models that (i) overlapped with known TEs (Martin *et al.* 2010), (ii) had more than 40% unknown bases (N), (iii) had homology with Repbase (Jurka *et al.* 2005) or (iv) were <20 amino acids in length. The 1315 genes that were manually curated (Martin *et al.* 2010) served as a validation set. Twelve percentage of the *T. melanosporum* genome was covered by uncategorized repeated sequences (so-called no cat) lacking homology with known TE families that could code for either *T. melanosporum*-specific TE or proteins belonging to orphan multigenic families. Gene models overlapping these no cat sequences were retained in the new set of 9952 gene models (Table S1). The expression levels on NimbleGen microarrays and RNAseq for each of the genes in the repertoire were determined for the ectomycorrhizae, free-living mycelium and ascocarps (Martin *et al.* 2010; Tisserant *et al.* 2011). The homology of each gene model searched against the NCBI nr database (September 2013) and UniProt (UniProtKB/Swiss-Prot of September 2013) was computed using BLASTP (v2.2.28+) with an e-value threshold of 10^{-5} (Altschul 1990). Each gene was also analysed for Pfam motifs using the *hmmscan* command of the HMMER package (Eddy 2011).

SNP calling and localization in the genome

SNP calling was performed with two different methods: (i) BWA for the alignment and SAMtools (Li *et al.* 2009) for the SNP calling (referred to as the BWA/SAMtools method) and (ii) the CLC Genomics Workbench version 6.6 (<http://www.clcbio.com>) for both the alignment and calling (referred to as the CLC method).

For the BWA/SAMtools method, a pile-up file (i.e. file describing the mapping results information at each chromosomal position) was created with the SAMtools *mpileup* command using the bam alignment output generated by BWA (see above). The SNP calling was filtered with the *vcfutils* script (available with SAMtools); to be validated, each SNP was required to be supported by at least ten reads, and the root mean square (RMS) of the mapping quality of the SNP position had to be ≥ 25 .

For the CLC method, the reads were mapped using a global alignment with the length fraction set to 1, the similarity fraction set to 0.97 and nonspecific reads ignored. All other parameters were set by default (http://www.clcbio.com/files/usermanuals/CLC_Genomics_Workbench_User_Manual.pdf). The SNPs were called by the quality-based variant detection module ignoring variants in nonspecific regions and using default parameters (i.e. minimum coverage of 10 reads). The maximum expected variation (ploidy) was set to 1, because haploid genomes had previously been sequenced (see above).

For both pipelines (BWA/SAMtools and CLC), we created a file for each sequenced genome with the SNPs localized on the reference genome assembly (France-Pro). The two sets of SNPs identified for each genome were compared, and only the SNPs called by both software methods were considered for further analyses.

All of the SNPs identified by aligning the six geographical accessions against the reference genome were compiled to generate a gff-formatted file available in DRYAD (doi:10.5061/dryad.9gk52). The SNPs were localized according to the new protein-coding gene catalogue defined in this study (see above) and repeated sequences library defined in Martin *et al.* (2010) using python scripts available at the INRA *Tuber* genome portal using the following link (<http://mycor.nancy.inra.fr/IMGC/TuberGenome/download.php?select=anno>).

Polymorphism indices and detecting selection pressure

The level of polymorphism between genomes was assessed by the π index (Nei & Li 1979) calculation in a sliding window of 10 kb throughout the entire reference genome using EGGLIB version 2.1.6 (De Mita & Siol 2012). The π index corresponds to the average number of nucleotide differences per site between two DNA sequences from the sample population. In the sliding windows of 10 kb throughout the entire genome, we also computed the Tajima's D (Tajima 1989) and Watterson Theta (Watterson 1975) values. The ratios of nonsynonymous mutations per nonsynonymous sites (pN) and synonymous mutations per synonymous sites (pS) in gene models were calculated to assess the mutations for deviation from neutral evolution. Positive Tajima's D values are typically attributed to diversifying, balancing or positive selection, whereas negative values are generally attributed to positive or purifying selection (Weedall & Conway 2010); they have an expected normal distribution between -2 and $+2$ for a 95% confidence interval (Tajima 1989; Carlson *et al.* 2005). Therefore, in this study, we considered values $>+2$ or less than -2 as significant. The Tajima's D value was calculated for all (i) of the gene models and (ii) in a sliding window of 10 kb throughout

the whole genome (including the coding and noncoding regions). Only gene models with at least five SNPs in their coding regions were considered for the pN/pS ratio calculation. A comparison between the two indices allowed us to identify candidate gene models subject to positive selection if pN/pS was >1 and purifying selection if pN/pS was <1.

Phylogenetic reconstruction and divergence time among geographical accessions

To construct a phylogeny of the seven *T. melanosporum* samples, the 60 507 SNPs present in the intergenic regions free of selective pressure (excluding repeated sequences and genomic regions with a Tajima's D index above 2) were selected. A maximum-likelihood phylogenetic tree was built using the default parameters of PhyML (Guindon *et al.* 2010) with 100 bootstrap replicates. To investigate the minimal number of SNPs suitable for a population genetic analysis, subsets of 10, 100, 1000, 5000, 10 000, 15 000, 20 000, 25 000, 30 000, 35 000, 40 000 and 50 000 SNPs were randomly selected 100 times among the 60 507 SNPs free of selection, and 100 maximum-likelihood phylogenetic trees were built using the PhyML default parameters. The Robinson–Foulds distance (Robinson & Foulds 1981) was used to measure the distance between each generated phylogenetic tree and the reference tree generated for the whole set of 60 507 SNPs using *RF.dist* in the *phangorn* R library (Schliep 2011). For each subset of SNPs, the number of trees identical to the reference tree was calculated as the number with a Robinson–Foulds distance equal to 0, meaning that the two trees are identical.

The divergence time estimates were performed with the 60 570 SNPs present in the intergenic regions free of selective pressure as described by Wicker *et al.* (2013). For the calculation, we assumed that all of the SNPs were present in regions that had accumulated mutations at the same rate, and we used a rate of 1.3 E^{-8} ($\pm 2.29 \text{ E}^{-9}$) substitutions per site per year as originally proposed by Ma & Bennetzen (2004) for rice. This choice was justified by the common use of this mutation rate for fungi as was done by Wicker *et al.* (2013) to estimate the divergence among isolates of *B. graminis*, another ascomycete fungus. It is also in the range of mutation rates (0.09 E^{-8} to 1.67 E^{-8} substitutions per site per year) proposed by Kasuga *et al.* (2002) for fungi. The number of SNPs in the genomic regions free of selection pressure was used to estimate the time since the most recent common ancestor (MRCA) for all geographical accessions. Using all of the SNPs in the genomic regions free of selective pressure, a Bayesian phylogenetic analysis was conducted with 5 000 000 generations, a sampling for each 1000 generations, and a burn-in value of 1250 with BEAST version

v1.7.5 (Drummond & Rambaut 2007) using the Hasegawa–Kishino–Yano (HKY) DNA substitution model. The estimated age of the MRCA for the tree was used to estimate the different node ages using a relaxed clock model with uncorrelated exponential prior distribution levels. The exponential relaxed clock model was chosen because it had been used previously for Tuberaceae by Bonito *et al.* (2013).

Results and Discussion

Geographical accessions: resequencing and read mappings

Among the approximately 34–39 million reads generated from each genome, 73–80% mapped to a unique position against the reference genome using BWA (Table 2). The raw-read data were deposited in the NCBI sequence reads archive under Accession No SRP044130. The average sequencing quality was good (quality score ≥ 20), as illustrated in Fig. S2, and the low-quality reads were discarded from the mapping for future analyses. The read coverage throughout the genome was continuous with an average depth of approximately 20 \times and did not reveal any important differences between the protein-coding and repeated sequences (Fig. S3). This likely results from the stringent parameters used for the read mapping and the post-processing step in which the reads mapped at several genomic locations were eliminated. Indeed, when multiple mapping was possible, it was shown that an increased density of mapped reads correlates with the location of repeated sequences in *Pyrenophora tritici-repentis* (Manning *et al.* 2013).

Between 91% (Spain-1) and 93% (France-Als) of the France-Pro (Mel28) reference genome was covered by reads (Table 2), and the *T. melanosporum* core genome was estimated at approximately 110 Mbp. The small proportion of the reference genome not covered by reads corresponds in large part to repeated sequences (approximately 80%; Fig. S4). TEs, primarily *gypsy* retrotransposons, were over-represented in the genomic regions not covered (Fig. S4). A total of 187 of the 9952 protein-coding gene models identified in the reference genome were found in the uncovered genomic regions (Table 2), with a maximum of 160 genes found for the France-Alp geographical accession. Among these 187 genes, microarray and RNAseq expression data showed that 73 and 112, respectively, were expressed in at least one tissue (Table S1; Martin *et al.* 2010; Tisserant *et al.* 2011). The regions that were not covered by mapped reads may correspond to (i) regions absent in the resequenced genomes; (ii) regions present in the resequenced genomes, but highly polymorphic, thus preventing proper read mapping; or (iii) regions rich in repeated sequences that prevent

nonambiguous mapping. Unfortunately, the sequencing strategy utilized (i.e. single-end sequencing) was not suitable to address these genomic regions in more detail, because the *de novo* assembly of new resequenced genomes was not possible.

The truffle is a heterothallic species harbouring one of two mating type idiomorphs (i.e. MAT1-1 or MAT1-2) in its haploid genome (Martin *et al.* 2010), and the MAT1-2 idiomorph is present in Scaffold 247 of the France-Pro reference genome (Rubini *et al.* 2011). The resequenced genomes from the geographical accessions Spain-1, France-Als and France-Alp contained reads matching the MAT1-2 idiomorph in the France-Pro reference genome (Fig. S3B), while the genomes of the geographical accessions Spain-2, France-Bur and Italy lack these sequences, suggesting they harbour the MAT1-1 idiomorph. This was confirmed by mapping the Illumina reads against the known sequence of the MAT1-1 idiomorph (data not shown). These results confirmed that either of the two mating type idiomorphs is present in the *T. melanosporum* haploid genome (Rubini *et al.* 2011).

SNP identification

The SNP calling performed with the BWA/SAMtools and CLC Genome Workbench programs produced similar results (Fig. 1), with 93% (442 326 SNPs corresponding to 3540 SNPs/Mbps) of those called by BWA/SAMtools also being called by CLC. As proposed by Zhan *et al.* (2011), only SNPs called by both methods were retained. The gff file with this SNPs resource can be downloaded in DRYAD (doi:10.5061/dryad.9gk52) and in our institution website following this link (<http://mycor.nancy.inra.fr/IMGC/TuberGenome/download.php?select=anno>).

A comparison of each of the *T. melanosporum* resequenced genomes to the France-Pro reference genome

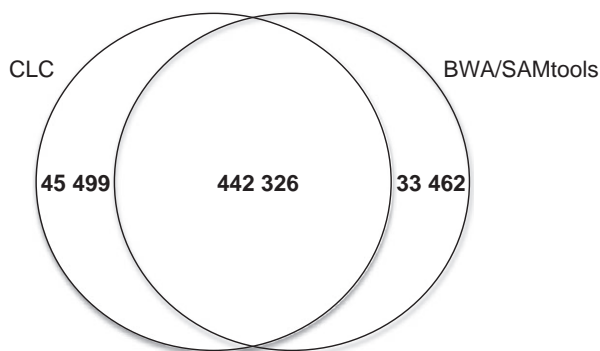


Fig. 1 Venn diagram of the number of single nucleotide polymorphisms called by both the Burrow-Wheeler Aligner/SAMtools and CLC Genomic Workbench methods.

identified between 108 112 and 198 788 SNPs, with the density ranging from 865 to 1591 SNPs/Mbp for the geographical accessions France-Als and Spain-2, respectively (Table 3). According to Fumagalli *et al.* (2013), SNPs identified using high-throughput sequencing technologies should be considered with caution, particularly when the sequencing depth is low (<10×). Here, the SNPs retained for analyses were identified using two different programs, and together with the stringent mapping that limited the multiple mapping of reads and the 10× read depth required for SNP identification, likely limited spurious SNPs (Li *et al.* 2009; Fumagalli *et al.* 2013). However, additional whole-genome sequencing or the targeted sequencing of SNP-rich regions will be conducted to experimentally confirm the existence of these *in silico* SNPs.

The SNP density in filamentous fungi varied from 291 to 14 005 SNPs/Mbp for the *Fusarium graminearum* (Cuomo *et al.* 2007) and *Rhizoctonia solani* (Hane *et al.* 2014) geographical accessions, respectively (Table 4). The differences in the nucleotide polymorphism levels observed could reflect differences in the demographic history of these species (e.g. reduction of polymorphism due to population bottlenecks), evolutionary trends related to the lifestyles of these fungi (e.g. for pathogenic species) as well as their respective ratios between sexual and asexual reproduction. *T. melanosporum*, with 3540 SNPs/Mbp, has a genetic diversity level in the lower range for filamentous fungi (Table 4), although we are aware that the parameters used to call the SNPs and the number of samples differed between the studies.

SNPs are not distributed equally in the genome

The polymorphism index (π ; Nei & Li 1979) was calculated along the genome in 10 Kb sliding windows, which showed some genomic regions to be more polymorphic than others (Fig. 2). Indeed, the SNPs were not distributed equally in the genome, and as expected, they occurred more frequently in repeated sequences than in protein-coding genes (Table 3; Figs. 2 and S5). Most of the SNPs were found in repeated sequences (84.6%) that represented 57.7% of the *T. melanosporum* genome (Martin *et al.* 2010). The SNPs were more frequent in *gypsy* retrotransposons than in DNA transposons (Fig. S5). A bias in the SNP distribution was also observed in the *F. graminearum* fungal genome, where 50% of the SNPs were present in 13% of the genome (Cuomo *et al.* 2007). Large blocks of regions rich in SNPs were also identified in the *B. graminis* genome (Hacquard *et al.* 2013; Wicker *et al.* 2013) and in the poplar leaf rust *Melampsora larici-populina*, in which a large portion of the variants were identified in coding sequences (Persoons *et al.* 2014).

Table 3 The number of single nucleotide polymorphisms (SNPs) and their distribution in the different genomic regions

Samples	Introns		Exons		Untranslated regions (UTRs)		Repeated sequences*		Other genomic regions		Total	
	Number	SNP/Mbp	Number	SNP/Mbp	Number	SNP/Mbp	Number	SNP/Mbp	Number	SNP/Mbp	Number	SNP/Mbp
Spain-1	2727	354	1816	155	102	179	139 824	1961	22 550	670	167 019	1337
Spain-2	3144	409	2050	175	119	209	165 981	2327	27 494	817	198 788	1591
France-Bur	2569	334	1529	131	84	148	110 806	1554	18 295	544	133 283	1067
France-Als	1948	253	1282	110	69	121	89 413	1254	15 400	458	108 112	865
Italy	2288	297	1478	126	91	160	106 57	1494	16 570	492	126 997	1016
France-Alp	1952	254	1332	114	77	135	98 933	1387	15 363	456	117 657	942
Total†	6795	883	4501	385	252	443	374 268	5248	56 510	1679	442 326	3540

*Repeated sequences comprising known transposable elements and uncategorized elements.

†The total number of SNPs excluding redundancy.

Table 4 Polymorphism levels estimated by whole-genome resequencing in filamentous fungi

Species	Phylum	Number of sequenced strains	SNPs/Mbp	Reference
<i>Blumeria graminis</i>	Ascomycota	2	1000	Hacquard <i>et al.</i> (2013)
<i>Coccidioides immitis</i>	Ascomycota	10	5251	Neafsey <i>et al.</i> (2010)
<i>Coccidioides posadasii</i>	Ascomycota	10	9227	Neafsey <i>et al.</i> (2010)
<i>Fusarium graminearum</i>	Ascomycota	2	291	Cuomo <i>et al.</i> (2007)
<i>Leptographium longiclavatum</i>	Ascomycota	71	975	Ojeda <i>et al.</i> (2014)
<i>Neurospora crassa</i>	Ascomycota	48	3375	Ellison <i>et al.</i> (2011)
<i>Tuber melanosporum</i>	Ascomycota	7	3540	This study
<i>Lentinula edodes</i>	Basidiomycota	2	4629	Au <i>et al.</i> (2013)
<i>Melampsora larici-populina</i>	Basidiomycota	15	6051	Persoons <i>et al.</i> (2014)
<i>Puccinia graminis</i>	Basidiomycota	1*	1843	Duplessis <i>et al.</i> (2011)
<i>Puccinia striiformis</i>	Basidiomycota	1*	5980	Cantu <i>et al.</i> (2013)
<i>Rhizoctonia solani</i>	Basidiomycota	2	14005	Hane <i>et al.</i> (2014)
<i>Rhizophagus irregularis</i>	Glomeromycota	6	321	Lin <i>et al.</i> (2014)

*For these species, the SNPs were identified in one dikaryotic strain.

Several mechanisms are known to inactivate transposons in filamentous fungi (Murat *et al.* 2013a), and some such as repeat-induced point mutations (RIPs) introduce mutations in these sequences (Selker *et al.* 1987). While genes involved in RIPs were not identified in the *T. melanosporum* genome (Martin *et al.* 2010), a strong preference for transitions in the CpG dinucleotide was observed by Clutterbuck (2011). Recently, Montanini *et al.* (2014) found that the methylation pattern in *T. melanosporum* selectively targets TEs rather than genes, and their results strongly favour methylation induced premeiotically (MIP) as the process responsible for TE silencing in *T. melanosporum*. Interestingly, MIP can increase the mutation rate of the methylated cytosines, as documented for mammalian DNA (Krickler *et al.* 1992). The SNPs were more frequently found in *gypsy* retrotransposons. Interestingly, these elements colonized the

T. melanosporum genome several millions years ago (Martin *et al.* 2010), and their SNP richness can be explained by their old age, as SNPs in these regions tend to accumulate due to DNA decay (Lisch & Benetzen 2011). The mapping of reads in multiple locations was low (approximately 6%; Table 2), although almost 60% of the *T. melanosporum* genome corresponds to repeated sequences, suggesting that the different TE copies are not conserved.

SNPs in gene models

A total of 903 protein-coding genes presented with more than two SNPs in their untranslated regions (UTRs), introns and/or exons. Among these, 742 had SNPs in their coding regions, including 584 nonsynonymous mutations (Tables S1 and S2). The 20 gene models with the

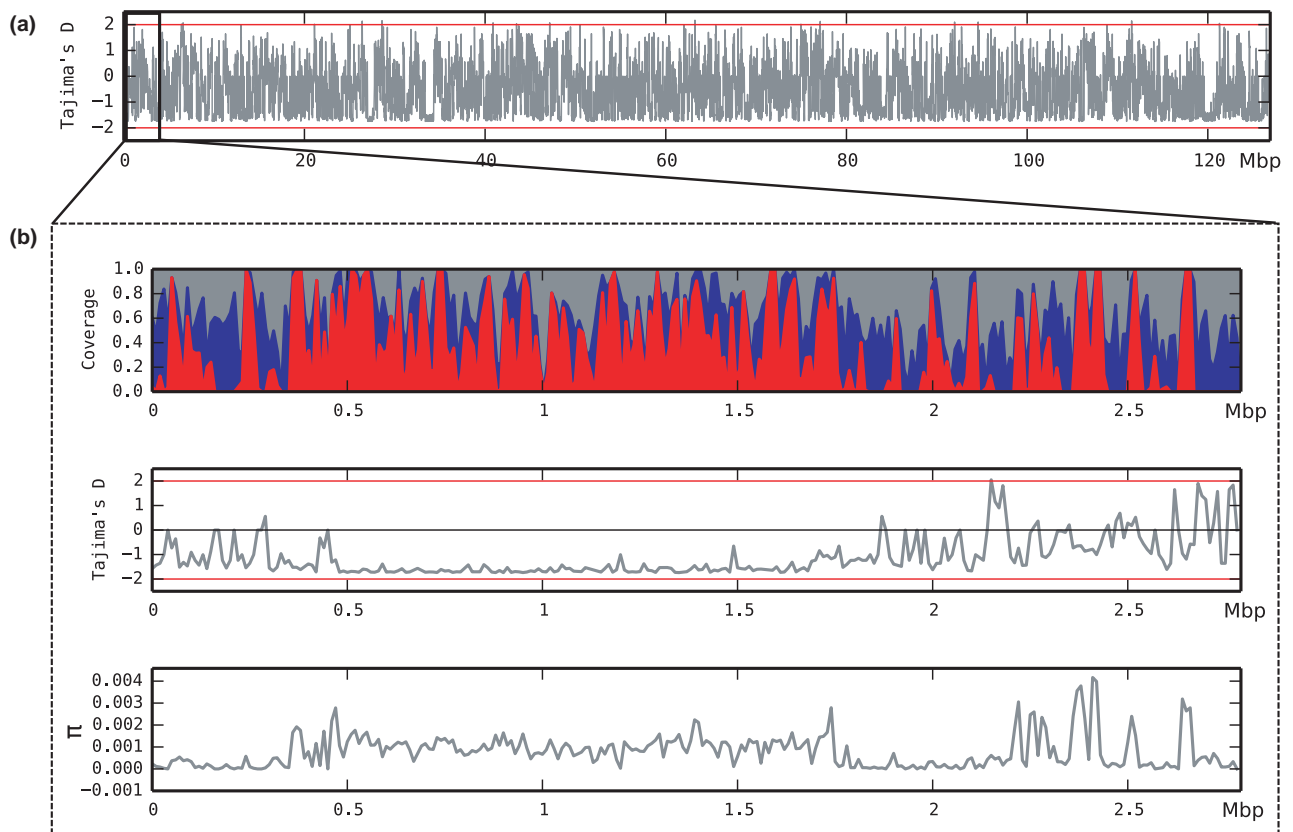


Fig. 2 Genome scans for selection throughout the whole genome and the polymorphism distribution in Scaffold 1. (a) Tajima's D index in 10 Kbp windows throughout the whole genome. The $+2/-2$ threshold limits are indicated. (b) Genomic landscape with the proportion of transposable elements (red), gene models (blue) and other genomic regions (grey) in 10 Kbp windows for Scaffold 1 (top). The Tajima's D indices in the 10 Kbp windows throughout Scaffold 1 are shown (middle). The two windows with a Tajima's D value $>+2$ are indicated with a dotted line. The π index for each nucleotide in the 10 Kbp windows throughout Scaffold 1 is also shown (bottom). The π index corresponds to the average number of nucleotide differences per site between two DNA sequences chosen randomly from the sample population.

highest number of SNPs (>10) in their coding regions are shown in Table S3. Most have sequence similarities in the DNA databases, and four are paralogues coding for the same HET-E-1 protein. Five genes were not expressed in any tissue, although in comparison with the free-living mycelium, four were upregulated in fruiting bodies (coding for an alpha-/beta-glucosidase, an alpha-glucosidase 2, a methylene-tetrahydrofolate reductase 2 and an ankyrin repeat-containing protein), as were three in the ectomycorrhizal root tips (coding for a vegetative incompatibility protein HET-E-1, an alpha/beta-glucosidase and an alpha-glucosidase 2). The same putative alpha-/beta-glucosidase and alpha-glucosidase 2 were upregulated in both tissues (Table S3). For the genes with SNPs, no enrichment of specific metabolic pathways was detected (data not shown).

When compared with the gene contents of the 30 other ectomycorrhizal genomes sequenced under the framework of the Mycorrhizal Genome Initiative

(Martin *et al.* 2011), *T. melanosporum* has a restricted gene content (9952 protein-coding genes). Indeed, the number of gene models per species ranged from 30 282 for *Rhizophagus irregularis* down to 9952 for *T. melanosporum* (http://genome.jgi-psf.org/Mycorrhizal_fungi/Mycorrhizal_fungi.info.html?core=genome&query=%22groups:Mycorrhizal_fungi%22&searchType=Keyword). Thus, *T. melanosporum* genes may have experienced purifying selection at a higher rate in comparison with the species with larger gene repertoires and a higher number of gene families (i.e. functional redundancy). In *T. melanosporum*, 2.6% of the SNPs were found in protein-coding genes and 1% in coding sequences, which was less than found in *B. graminis* and the *Coccidioides* spp. Wicker *et al.* (2013) found between 3.7% and 3.9% of the SNPs in the coding regions for *B. graminis*, and Neafsey *et al.* (2010) identified 33–36% of the SNPs in genes (UTRs, introns and exons) when they compared the genomes

of *C. immitis* and *C. posadasii*. This suggests that the limited gene repertoire of *T. melanosporum* is associated with higher functional constraints and consequently presents a lower rate of genetic variation.

Detecting selection pressure

Two approaches were used to identify selection signatures. First, the rates of nonsynonymous mutations per nonsynonymous site (pN) and synonymous mutations per synonymous site (pS) were calculated for the 119 genes with five or more SNPs. Of those, 18 genes had a pN/pS ratio >1 and 9 had only nonsynonymous mutations (20 expressed in at least one tissue), suggesting they were under positive selection. On the other hand, 78 genes had a pN/pS ratio <1 (68 expressed in at least one tissue), suggesting they were under purifying selection (Tables S4 and S5). The second approach relies on the Tajima's D statistic computed on either the gene models or using a whole-genome sliding window. Here, we considered values >+2 and lower than -2 as significant. When calculated for the gene models, only four had a Tajima's D value >+2, suggesting the existence of a signature for balancing or positive selection (Table S1). These genes, coding for the vegetative incompatibility protein HET-E-1, an NADH-ubiquinone oxidoreductase, a vacuolar protein and a protein kinase, were expressed in the different tissues (Table S1). When the Tajima's D statistic was calculated by scanning the whole genome along a sliding window, 36 genomic regions with a Tajima's D value >+2 were identified (Fig. 2). Thirty-one gene models were present in these 36 genomic regions (Table S6), including the previously identified vegetative incompatibility protein HET-E-1.

The positive Tajima's D values can result not only from balancing selection, but also from population structure and moderately intense bottlenecks (i.e. a reduction in the size of the population; Biswas & Akey 2006). In addition to significant population structure effects (Murat *et al.* 2004; Riccioni *et al.* 2008; García-Cunchillos *et al.* 2014), a population bottleneck due to the last glaciation has also been proposed for *T. melanosporum* (Bertault *et al.* 1998). The six *T. melanosporum* geographical accessions were harvested from different populations, as demonstrated by the phylogeographical analysis (see below). Therefore, we cannot exclude that the high positive Tajima's D values observed resulted from a population bottleneck and/or population structure rather than from balancing selection. These results are preliminary and need to be confirmed by sequencing a larger number of genomes, but they open the way for future investigations of truffle adaptation to environmental stresses.

Phylogeography and divergence time among geographical accessions

The genomic regions putatively free of selection covered 36.6 Mbp, for a total of 60 507 SNPs. This set of SNPs can be downloaded in DRYAD (doi:10.5061/dryad.9gk52) and in our institution website following this link (<http://mycor.nancy.inra.fr/IMGC/TuberGenome/download.php?select=anno>).

The unrooted maximum-likelihood phylogenetic tree clustered together samples according to their geographical origin with a cluster comprised of the northern France samples (France-Als and France-Bur), a cluster grouping samples from south-eastern France (France-Alp and France-Pro) and Italy, and another cluster with the Spanish samples (Fig. S1B). This ability to identify the geographical origin of truffles harvested in natural populations using SNPs is currently being used to design diagnostic SNP arrays for geographical certification. As highlighted by Davey *et al.* (2011), genotyping SNPs across targeted populations is now facilitated by the advent of high-throughput SNP arrays. Indeed, depending on the sample size and the number of SNPs to be analysed, medium- to high-throughput technologies are available such as the competitive allele-specific PCR (KASPar) assay from KBiosciences (Hertfordshire, UK; <http://www.kbioscience.co.uk>) or the Affymetrix Axiom SNP microarrays. The KASPar assay is commonly used for genotyping up to 1000–2000 SNPs, while the Axiom SNP microarrays allow genotyping from 1500 to several million SNPs. As the financial investment for genotyping with KASPar can be three times less expensive than for the Axiom SNP microarray (Charles Poncet, INRA Gentyane Plateform, personal communication), the minimum number of SNPs required for a population genetic analysis was investigated. We found that a minimum of 30 000 SNPs is required to generate all of the maximum-likelihood trees identical to the reference tree produced with the 60 507 SNPs free of selection (Fig. 3). We are thus developing an array based on the 60 507 SNPs for analysing the population genetic structure throughout the natural regions of *T. melanosporum* production.

Using the mutation rate of 1.3×10^{-8} ($\pm 2.29 \times 10^{-9}$) substitutions per site per year (Ma & Bennetzen 2004; Wicker *et al.* 2013), we estimated that the 60 507 mutations had accumulated between 107 703 and 154 763 years ago ($131\,128 \pm 23\,098$ years). These times were used to set the estimated time of the MRCA for the Bayesian phylogenetic reconstruction generated with the 60 507 SNPs free of selection and a relaxed molecular clock. This Bayesian reconstruction clustered the French and the Italian samples together, while the Spanish samples separated earlier (Fig. 4). The Bayesian and maximum-likelihood phylogenies exhibit a single difference in their

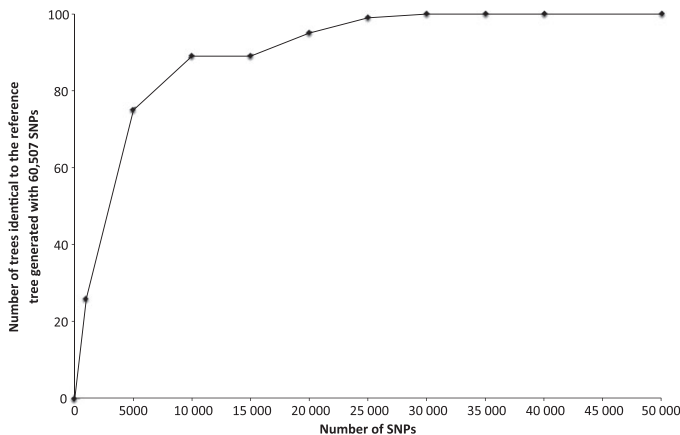


Fig. 3 Bayesian divergence time estimates for the seven truffle geographical accessions. The most recent common ancestor for the tree was calculated using the 60 570 SNPs free of selection and an evolutionary rate of 1.3 E^{-8} ($\pm 2.29 \text{ E}^{-9}$) substitutions per site per year. The maximum clade credibility tree with a posterior probability limit of 0.5 estimated in BEAST is shown with the node placed at the median age. Bare nodes (grey) and values represent the node age for the 95% highest posterior density.

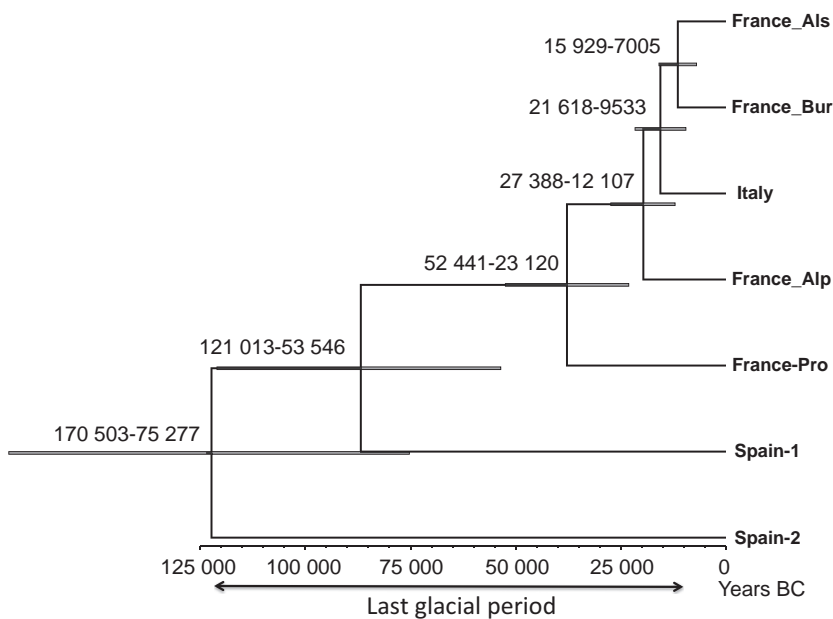


Fig. 4 The percentage of identical pairs of trees based on the number of SNPs. The SNPs were randomly selected 100 times from among the 60 507 SNPs free of selection, and 100 neighbour-joining trees were generated. The dissimilarity among each pair of trees was assessed by the Robinson–Foulds distance. The percentage was calculated as the number of pairs with Robinson–Foulds distances equal to 0 (identical topology) compared to the total number of pairs.

topology: the France-Pro, France-Alp and Italian samples form a monophyletic cluster in the maximum-likelihood phylogeny, but are paraphyletic in the Bayesian phylogeny (Fig. S6). This could be explained not only by the different methods used in the analyses (Bayesian versus maximum-likelihood), but also by the fact that one phylogeny is time-dependent (relaxed molecular clock in Bayesian), while the other is time-independent. That time-dependent and time-independent phylogenies are not always in agreement has been previously discussed (Drummond *et al.* 2006). The phylogenetic signal is also likely to be inconsistent across the genome due to the historical proximity of the samples, which increases the chances of finding different trees with different methods. For outbreeding species, such as *T. melanosporum*, the phylogenetic signal obtained with SNPs could be weakened due to population genetic processes such as

recombination and gene flow. However, both topologies are consistent with the geography: in the Bayesian phylogeny, the French south-eastern samples appear as intermediates between the northern French and the Italian and Spanish samples, while they appear as a separate group in the maximum-likelihood topology (Fig. S6). A further characterization of the overall structure of the *T. melanosporum* populations could be performed using population genetics methods, but they require a larger number of samples to be powerful.

The Bayesian reconstruction suggested that the Spanish samples separated earlier than the French and Italian samples (Fig. 4). Time calibrations and date estimations should be considered with caution, especially for studies without fossil data and incomplete taxon sampling. Thus, we are aware that we cannot use the absolute dates we obtained, but only relative estimates. However, it is

highly probable that the MRCA predates the last glaciation (120 000 to 11 000 years ago; Van Andel & Tzedakis 1996). While our sampling study is not sufficient alone to definitively describe the history of *T. melanosporum* following the last glaciation, this preliminary analysis paves the way for future analyses using the current SNPs resource as we have proposed.

Conclusions

Today, *T. melanosporum* is primarily harvested in truffle orchards in France using tree seedlings that have been inoculated with truffles in greenhouses (Olivier *et al.* 2012). Up to now, the selection of geographically defined sources of truffle inoculum has not been considered for plantations. Interestingly, the Aquitaine regional truffle growers' federation has initiated the production of inoculated plants using truffles sampled from specific natural populations that have appeared to be better adapted to drought or frost (P. Rejou, personal communication). To date, the selection of these truffle populations has been empirical and based only on field observations; this approach could now be validated by genotyping the truffle inocula. Moreover, it is now recognized that truffle aroma has, at least in part, a genetic origin (Martin *et al.* 2010). Investigating a putative phenotype association between SNP markers and traits (such as particular aroma or stress tolerance) can now be contemplated thanks to the current SNPs resource, which is the first step towards a marker-assisted selection of the fungal inocula used by truffle growers.

Acknowledgements

The UMR1136 is supported by a grant overseen by the French National Research Agency (ANR) as part of the Investments for the Future Programme (ANR-11-LABX-0002-01, Lab of Excellence ARBRE). This study benefited from ANR SYSTERRA SYSTRUF (ANR-09-STRA-10). Thibaut Payen's PhD scholarship is cofunded by the Lorraine Region and the European Commission through the EcoFINDERS project (FP7-264465). We would like to thank Francesco Paolucci, Bernard Vonfli, Mario Honrubia, Luc Bernard and Henri Frochet for providing the samples analysed in this study. We also would like to thank Sébastien Duplessis and François Le Tacon for their constructive advice and helpful discussions. Finally, the authors would like to thank the team of American Manuscript Editors for the language and style editing of the manuscript.

References

Altschul S (1990) Basic local alignment search tool. *Journal of Molecular Biology*, **215**, 403–410.

Au CH, Cheung MK, Wong MC *et al.* (2013) Rapid genotyping by low-coverage resequencing to construct genetic linkage maps of fungi: a case study in *Lentinula edodes*. *BMC Research Notes*, **6**, 307.

Bertault G, Raymond M, Berthomieu A *et al.* (1998) Trifling variation in truffles. *Nature*, **394**, 734.

Biswas S, Akey JM (2006) Genomics insights into positive selection. *TRENDS in Genetics*, **22**, 437–446.

Bonito G, Smith ME, Nowak M *et al.* (2013) Historical biogeography and diversification of truffles in the tuberaceae and their newly identified southern hemisphere sister lineage. *PLoS One*, **8**, e52765.

Brumfield RT, Beerli P, Nickerson DA, Edwards SV (2003) The utility of single nucleotide polymorphisms in inferences of population history. *Trends in Ecology & Evolution*, **18**, 249–256.

Cantu D, Segovia V, MacLean D *et al.* (2013) Genome analyses of the wheat yellow (stripe) rust pathogen *Puccinia striiformis f. sp. tritici* reveal polymorphic and haustorial expressed secreted proteins as candidate effectors. *BMC Genomics*, **14**, 270.

Carlson CS, Thomas DJ, Eberle MA *et al.* (2005) Genomic regions exhibiting positive selection identified from dense genotype data. *Genome Research*, **15**, 1553–1565.

Clutterbuck JA (2011) Genomic evidence of repeat-induced point mutation (RIP) in filamentous ascomycetes. *Fungal Genetics and Biology*, **48**, 306–326.

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

Davey JW, Hohenlohe PA, Etter PD *et al.* (2011) Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nature Reviews Genetics*, **12**, 499–510.

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

Drummond AJ, Rambaut A (2007) BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology*, **7**, 214.

Drummond AJ, Ho SY, Phillips MJ *et al.* (2006) Relaxed phylogenetics and dating with confidence. *PLoS biology*, **4**, e88.

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

Eddy SR (2011) Accelerated profile HMM searches. *PLoS Computational Biology*, **7**, e1002195.

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 United States of America*, **108**, 2831–2836.

Fumagalli M, Vieira FG, Korneliusen TS *et al.* (2013) Quantifying population genetic differentiation from next-generation sequencing data. *Genetics*, **195**, 979–992.

Ganal MW, Altmann T, Röder MS (2009) SNP identification in crop plants. *Current opinion in plant biology*, **12**, 211–217.

García-Cunchillos I, Sánchez S, Barriuso JJ, Pérez-Collazos E (2014) Population genetics of the westernmost distribution of the glaciations-surviving black truffle *Tuber melanosporum*. *Mycorrhiza*, **24**, 89–100.

Guindon S, Dufayard JF, Lefort V *et al.* (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Systematic Biology*, **59**, 307–321.

Hacquard S, Kracher B, Maekawa T *et al.* (2013) Mosaic genome structure of the barley powdery mildew pathogen and conservation of transcriptional programs in divergent hosts. *Proceedings of the National Academy of Sciences*, **110**, E2219–E2228.

Hane JK, Anderson JP, Williams AH *et al.* (2014) Genome sequencing and comparative genomics of the broad host-range pathogen *Rhizoctonia solani* AG8. *PLoS Genetics*, **10**, e1004281.

Jurka J, Kapitonov VV, Pavlicek A *et al.* (2005) Repbase Update, a database of eukaryotic repetitive elements. *Cytogenetic and Genome Research*, **110**, 462–467.

Kasuga T, White TJ, Taylor JW (2002) Estimation of nucleotide substitution rates in eurotiomycete fungi. *Molecular Biology and Evolution*, **19**, 2318–2324.

Kricker MC, Drake JW, Radman M (1992) Duplication-targeted DNA methylation and mutagenesis in the evolution of eukaryotic chromosomes. *Proceedings of the National Academy of Sciences*, **89**, 1075–1079.

- Kües U, Martin F (2011) On the road to understanding truffles in the underground. *Fungal genetics and biology*, **48**, 555–560.
- Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics (Oxford, England)*, **25**, 1754–1760.
- Li H, Handsaker B, Wysoker A *et al.* (2009) The sequence alignment/map format and SAMtools. *Bioinformatics (Oxford, England)*, **25**, 2078–2079.
- Lin K, Limpens E, Zhang Z *et al.* (2014) Single nucleus genome sequencing reveals high similarity among nuclei of an endomycorrhizal fungus. *PLoS Genetics*, **10**, e1004078.
- Lisch D, Bennetzen JL (2011) Transposable element origins of epigenetic gene regulation. *Current Opinion in Plant Biology*, **14**, 156–161.
- Ma J, Bennetzen JL (2004) Rapid recent growth and divergence of rice nuclear genomes. *Proceedings of the National Academy of Sciences of the United States of America*, **101**, 12404–12410.
- Manning VA, 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: Genes | Genomes | Genetics*, **3**, 41–63.
- Martin F, Kohler A, Murat C *et al.* (2010) Périgord black truffle genome uncovers evolutionary origins and mechanisms of symbiosis. *Nature*, **464**, 1033–1038.
- Martin F, Cullen D, Hibbett D *et al.* (2011) Sequencing the fungal tree of life. *New Phytologist*, **190**, 818–821.
- Montanini B, Chen PY, Morselli M *et al.* (2014) Non-exhaustive DNA methylation-mediated transposon silencing in the black truffle genome, a complex fungal genome with massive repeat element content. *Genome biology*, **15**, 411.
- Murat C, Díez J, Luis P *et al.* (2004) Polymorphism at the ribosomal DNA ITS and its relation to postglacial re-colonization routes of the Périgord truffle *Tuber melanosporum*. *New Phytologist*, **164**, 401–411.
- Murat C, Payen T, Petitpierre D, Labbé J (2013a) Repeated elements in filamentous fungi with a focus on wood-decay fungi. In: *The Ecological Genomics of Fungi* (ed. Martin F), pp. 21–40. John Wiley & Sons, Inc, Ames, Iowa.
- Murat C, Rubini A, Riccioni C *et al.* (2013b) Fine-scale spatial genetic structure of the black truffle (*Tuber melanosporum*) investigated with neutral microsatellites and functional mating type genes. *New Phytologist*, **199**, 176–187.
- 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.
- Nei M, Li WH (1979) Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proceedings of the National Academy of Sciences*, **76**, 5269–5273.
- Ojeda DI, Dhillon B, Tsui CKM, Hamelin RC (2014) Single-nucleotide polymorphism discovery in *Leptographium longiclavatum*, a mountain pine beetle-associated symbiotic fungus, using whole-genome resequencing. *Molecular Ecology Resources*, **14**, 401–410.
- Olivier JM, Savignac JC, Sourzat P (2012) *Truffe et Trufficulture*. Fanlac, Périgueux.
- Persoons A, Morin E, Delaruelle C *et al.* (2014) Patterns of genomic variation in the poplar rust fungus *Melampsora larici-populina* identify pathogenesis-related factors. *Frontiers in Plant Science*, **5**, 450.
- Riccioni C, Belfiori B, Rubini A *et al.* (2008) *Tuber melanosporum* outcrosses: analysis of the genetic diversity within and among its natural populations under this new scenario. *New Phytologist*, **180**, 466–478.
- Robinson DF, Foulds LR (1981) Comparison of phylogenetic trees. *Mathematical Biosciences*, **53**, 131–147.
- Rubini A, Belfiori B, Riccioni C *et al.* (2011) Isolation and characterization of MAT genes in the symbiotic ascomycete *Tuber melanosporum*. *New Phytologist*, **189**, 710–722.
- Schliep KP (2011) Phangorn: phylogenetic analysis in R. *Bioinformatics*, **27**, 592–593.
- Selker EU, Cambareri EB, Jensen BC, Haack KR (1987) Rearrangement of duplicated DNA in specialized cells of *Neurospora*. *Cell*, **51**, 741–752.
- Smith SE, Read DJ (2010) *Mycorrhizal Symbiosis*. Academic Press, London.
- Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, **123**, 585–595.
- Tisserant E, Da Silva C, Kohler A *et al.* (2011) Deep RNA sequencing improved the structural annotation of the *Tuber melanosporum* transcriptome. *New Phytologist*, **189**, 883–891.
- Van Andel TH, Tzedakis PC (1996) Palaeolithic landscapes of Europe and environs, 150,000–25,000 years ago: an overview. *Quaternary Science Reviews*, **15**, 481–500.
- Watterson GA (1975) On the number of segregating sites in genetical models without recombination. *Theoretical Population Biology*, **7**, 256–276.
- Weedall GD, Conway DJ (2010) Detecting signatures of balancing selection to identify targets of anti-parasite immunity. *Trends in Parasitology*, **26**, 363–369.
- 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.
- Zhan B, Fadista J, Thomsen B *et al.* (2011) Global assessment of genomic variation in cattle by genome resequencing and high-throughput genotyping. *BMC Genomics*, **12**, 557.

F.M. and C.M. designed the project. C.M. extracted the DNA, and C.M., T.P., A.G. and E.M. contributed to the bioinformatics analyses. T.P. and S.D.M. performed the selection analyses. C.M., T.P., S.D.M. and F.M. wrote the manuscript.

Data accessibility

Data sequences: the raw sequence data generated in this study were deposited in the NCBI short reads archive under Accession No SRP044130.

SNPs data: the gff file with all 442 326 SNP resources and the 60 507 SNPs free of selection can be downloaded in DRYAD (doi:10.5061/dryad.9gk52) and in our institution website following this link (<http://mycor.nancy.inra.fr/IMGC/TuberGenome/download.php?select=anno>).

Phylogenetic trees: the nexus and xml files used as input in BEAST as well as the newick files corresponding to the Bayesian and the maximum-likelihood phylogenetic reconstruction are available in DRYAD (doi:10.5061/dryad.9gk52).

Bioinformatic scripts: the python scripts used in this study are available at the INRA *Tuber* genome portal using the following link (<http://mycor.nancy.inra.fr/IMGC/TuberGenome/download.php?select=anno>).

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Fig. S1 (A) Geographical localization of the seven *T. melanosporum* geographic accessions generated with Google Earth (<http://www.google.fr/intl/fr/earth/>).

Fig. S2 Quality of the raw reads for the six re-sequenced genomes.

Fig. S3 Mapping the reads for the six re-sequenced genomes against France-Pro (reference genome) for (A) Scaffold 1 and (B) Scaffold 247.

Fig. S4 Percentage of the different France-Pro genomic regions without mapped reads in the different genomic regions.

Fig. S5 Single nucleotide polymorphisms distribution in the different genomic regions.

Fig. S6 Topology comparison of the Bayesian (left) and maximum-likelihood (right) phylogenies.

Table S1 Summarization of the number of polymorphisms (SNPs) and the different diversity and adaptation indices in the 9952 *T. melanosporum* protein-coding gene models, as well as their expression levels from RNAseq (Tisserant *et al.* 2011) and NimbleGen microarrays in the ectomycorrhizae (ECM), fruiting bodies (FB), and free-living mycelium (FLM; Martin *et al.* 2010).

Table S2 The 742 gene models that have more than two SNPs in their untranslated region (UTR), exon, and intron sequences and at least one SNP in their exon sequence.

Table S3 The 20 protein-coding genes with the highest number of SNPs in their coding regions.

Table S4 The gene models putatively under positive selection (i.e. ratio of non-synonymous to synonymous SNPs (pN/pS) above 1).

Table S5 The gene models putatively under diversifying selection (i.e. ratio of non-synonymous to synonymous SNPs (pN/pS) below 1).

Table S6 The 31 gene models present in the genomic regions exhibiting a Tajima's D statistic >2.

Supplementary figures

Fig. S1 (A) Geographical localization of the seven *T. melanosporum* geographic accessions generated with Google Earth (<http://www.google.fr/intl/fr/earth/>). (B) Unrooted maximum likelihood phylogeny. The bootstrap values are indicated.

A



B

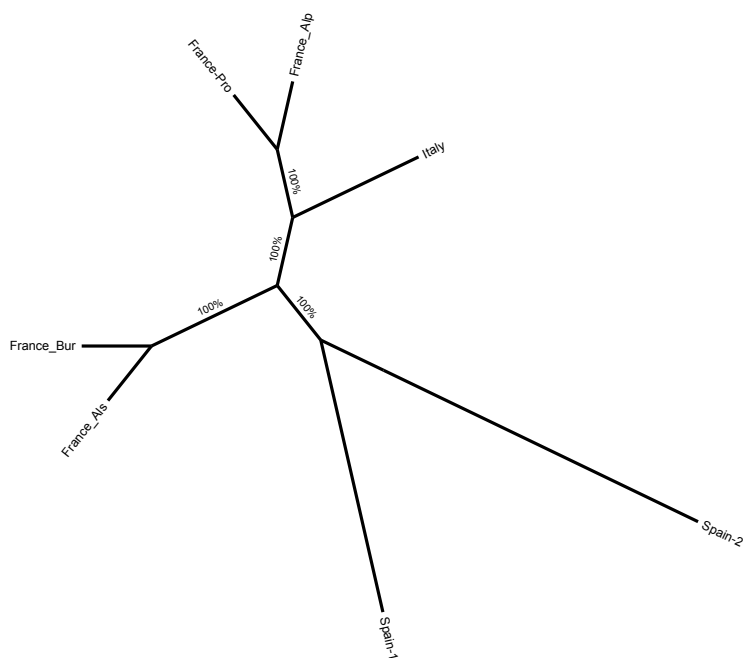


Fig. S2 Quality of the raw reads for the six re-sequenced genomes. Graphs were produced using FastQC (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>).

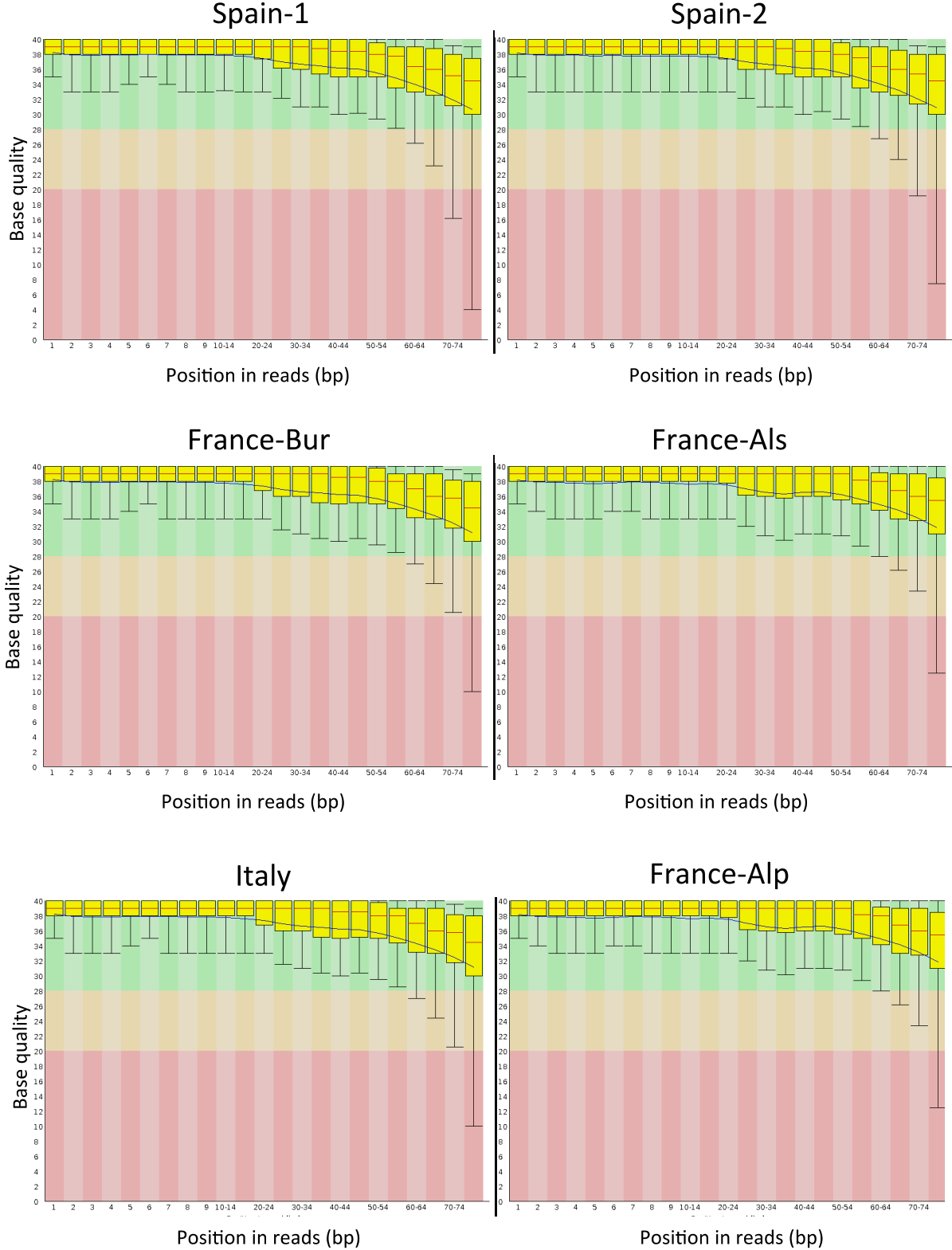


Fig. S3 Mapping the reads for the six re-sequenced genomes against France-Pro (reference genome) for (A) Scaffold 1 and (B) Scaffold 247. The gene models and repeated sequences are indicated. Scaffold 247 harbours the mating type (MAT) idiomorph (Rubini *et al.* 2011).

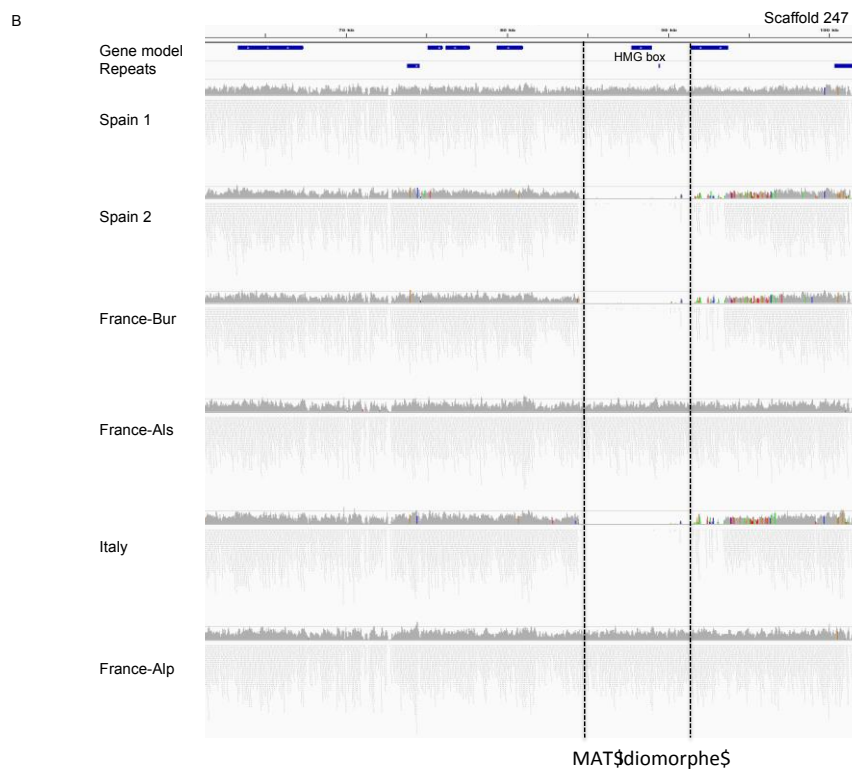
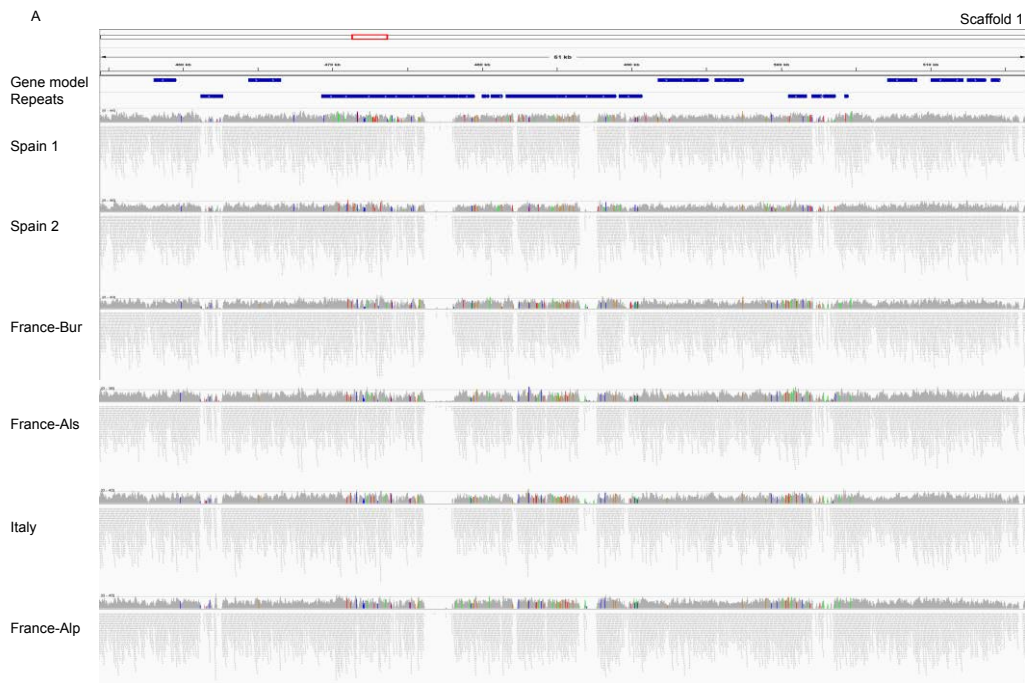


Fig. S4 Percentage of the different France-Pro genomic regions without mapped reads in the different genomic regions.

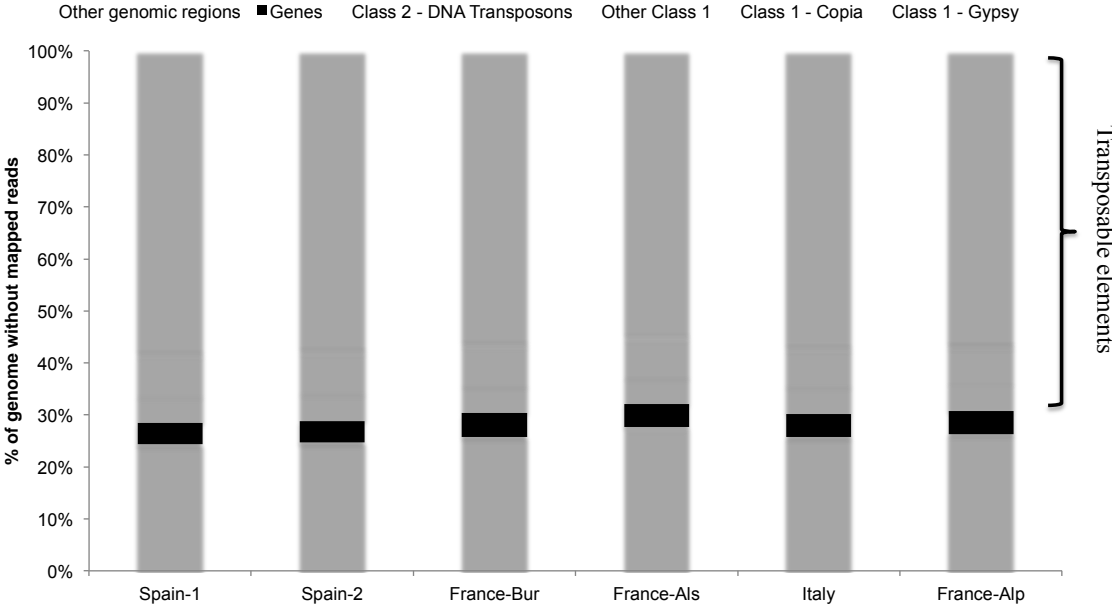


Fig. S5 Single nucleotide polymorphisms distribution in the different genomic regions.

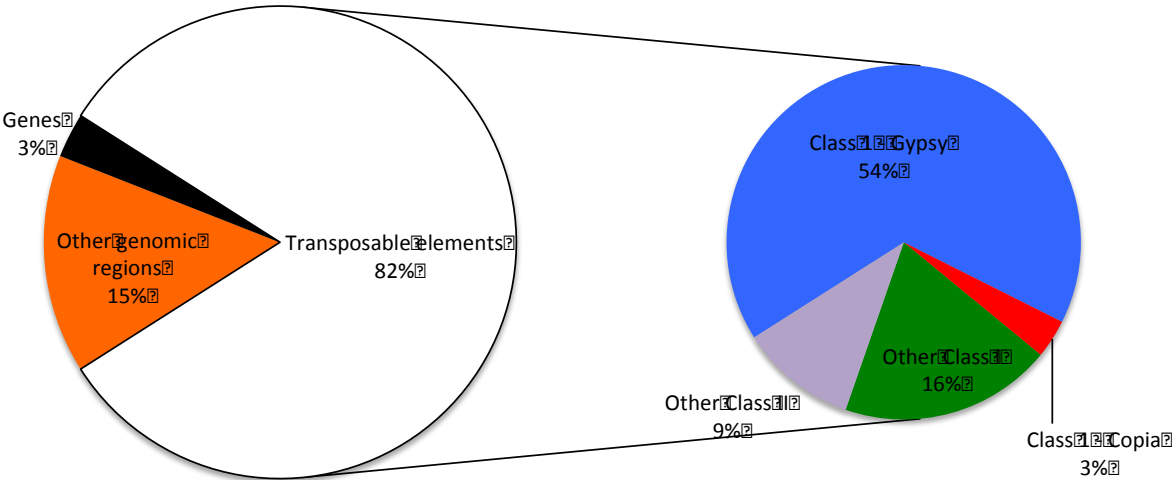
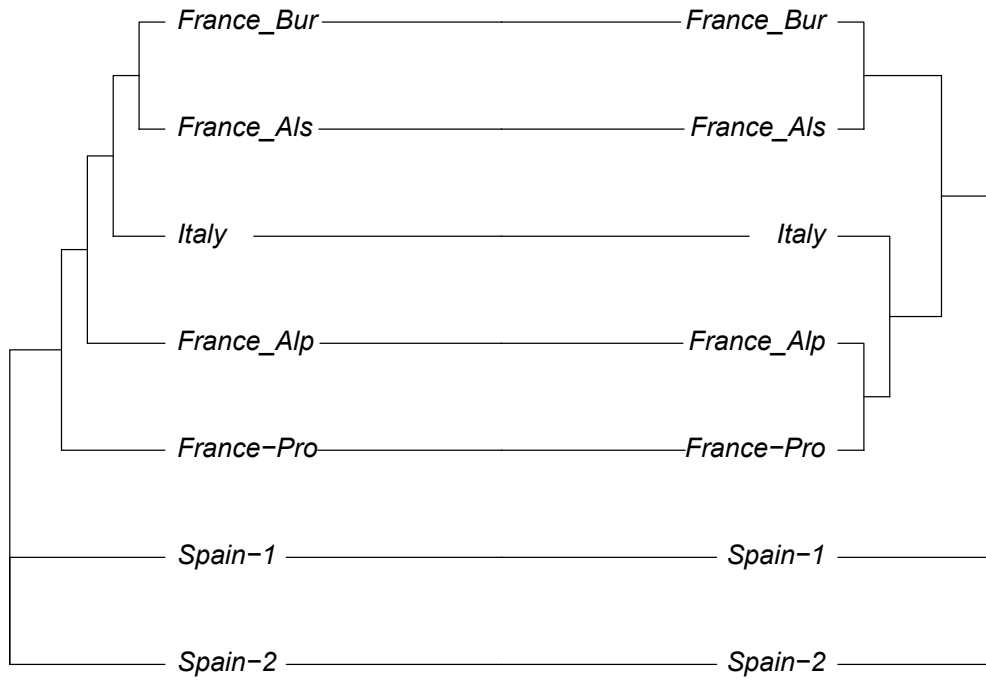


Fig. S6 Topology comparison of the Bayesian (left) and maximum-likelihood (right) phylogenies. The figure was generated using the *cophyloplot* function of the *ape* R package.



6 Reconstruction de l'histoire évolutive des rétrotransposons *Gypsy* chez *Tuber melanosporum*

Dans ce dernier chapitre j'ai focalisé mon attention sur les rétrotransposons *gypsy* de *T. melanosporum*. Cette étude a été compilée dans un article en préparation intitulé « **Reconstructing the evolutionary history of Gypsy retrotransposons in the black truffle of Périgord (*Tuber melanosporum* Vittad.)** » qui va être soumis au journal *Mobile DNA*.

Pourquoi s'intéresser aux rétrotransposons *gypsy* de *T. melanosporum* ? Comme cela a déjà été montré précédemment (voir chapitre 3 et 4), le génome de *T. melanosporum* est constitué à 58% d'éléments transposables dont 30 % sont des rétrotransposons *gypsy*. Les rétrotransposons *gypsy* appartiennent à la classe I des éléments transposables ce qui signifie qu'ils sont transcrits en ARN pour transposer. Ce mécanisme complexe a pour effet de multiplier le nombre de copie de façon assez rapide et permet donc une invasion du génome si les rétrotransposons *gypsy* ne sont pas inactivés. Malgré leur prédominance chez *T. melanosporum* leur impact sur le génome, leur évolution et leur diversité était mal connu au début de ma thèse. En effet, il existe de nombreuses familles de rétrotransposons *gypsy* puisque 35 familles ont été identifiées chez les animaux, plantes et champignons. L'objectif de ce travail a été de connaître la diversité des rétrotransposons *gypsy* chez *T. melanosporum* et ensuite de reconstruire leur histoire démographique.

En analysant les reverse-transcriptase, une protéine hautement conservée dans les *Gypsy*, il a été possible de mettre en évidence 6 familles de rétrotransposons *gypsy*. Parmi celles-ci, deux familles représentent 14 et 13% du génome. La première d'entre elles *Tmt1* avait déjà été décrite précédemment, la seconde en terme d'abondance, *Tmt6* est quant à elle très semblable à *Tcn2* qui n'avait jamais été identifié chez un Ascomycète. Les autres familles, qui sont présentes de façon moins importante sont pour certaines des familles nouvellement décrites et

pour l'une d'entre elle, *Tmt2*, une famille décrite uniquement chez les Arthropodes. Dans un second temps, les copies complètes ont pu être datées grâce à l'étude des séquences de reconnaissance terminales (LTR). De plus, les données de reséquençage (voir Chapitre 4) ont été utilisées pour étudier les points d'insertion des éléments identifiés comme étant en copies complètes. Nous avons trouvé que 62 des copies complètes de *Tmt6* ne sont pas insérées aux mêmes points dans tous les isolats de truffes. Parmi ces 62 copies, deux ne sont présentes que chez le génome de référence et en plus il s'agit de copies récentes (peu de polymorphismes entre les deux LTR) suggérant que ces éléments ont transposé après la séparation des isolats. Ce constat semble indiquer que certains éléments de la famille *Tmt6* étaient encore actifs relativement récemment. Malgré ce constat, dans les tissus pour lesquels nous possédons des données d'expression (fructification, mycelium et mycorhizes) nous n'avons pas été en mesure de détecter d'expression pour les protéines portées par les rétrotransposons *gypsy*.

Afin de mieux connaître la dynamique de colonisation du génome de *T. melanosporum* par ces éléments, leur histoire démographique a été étudiée. Cette analyse a montré que les taux de transposition étaient plus importants dans le passé, que *Tmt6* a probablement eu deux vagues de colonisation et que ces éléments ont tendance à s'agréger dans le génome. D'autre part, il apparaît que le système de transposition soit de type « copie modèle aléatoire », où après la création d'une nouvelle copie les deux copies jumelles ont autant de chance de produire un événement de transposition.

En conclusion, cette étude a permis d'appréhender la diversité des rétrotransposons *gypsy* dans le génome de *T. melanosporum*. Si la colonisation du génome de cette truffe par ces éléments est ancienne, certains éléments semblent avoir transposé récemment suggérant que les rétrotransposons *gypsy* peuvent continuer de jouer un rôle dans l'évolution du génome de *T. melanosporum*.

1 **Reconstructing the evolutionary history of Gypsy retrotransposons in the black truffle**
2 **of Périgord (*Tuber melanosporum* Vittad.)**

3

4

5 Payen Thibaut^{1,2*}, Murat Claude^{1,2,3*}, Martin Francis^{1,2}

6

7 ¹INRA, Laboratoire d'Excellence ARBRE, UMR1136 Interactions Arbres-Microorganismes,
8 F-54280 Champenoux, France

9 ²Université de Lorraine, UMR1136 Interactions Arbres-Microorganismes, F-54500
10 Vandoeuvre-lès-Nancy, France

11

12

13 ³Corresponding author:

14 Claude Murat

15 UMR1136 INRA/Université de Lorraine, "Interactions Arbres/Micro-organismes",

16 Laboratoire d'Excellence ARBRE, INRA Nancy, 54280 Champenoux, France

17 tel: +33 (0)3 83 39 41 27; fax: +33 (0)3 83 39 40 69; email: claude.murat@nancy.inra.fr

18

19 *These authors contributed equally as first authors

20

21 **Keywords:** chromovirus, gypsy-like retrotransposons, transposable elements, truffle,
22 transposition dynamic

23

24

25 **Abstract**

26 Truffles are ascomycete fungi belonging to genus *Tuber* and forming ectomycorrhizal
27 associations with trees and shrubs. Transposable elements (TE) are the most frequent repeated
28 sequences but their impact on the genome evolution is unknown in the genome of the black
29 Périgord truffle (*Tuber melanosporum*) black truffle where they constitute more than 50% of
30 the assembled genome sequence. The *T. melanosporum* genome is particularly rich in gypsy
31 retrotransposons. This study investigates the diversity of gypsy retrotransposons in this
32 species and their evolutionary history by analysing the reference genome and six resequenced
33 genomes of different geographic accession. Using reverse transcriptase sequences, six
34 different gypsy retrotransposons clades called *Tmt1* to *Tmt6* were identified. *Tmt1*, *Tmt2* and
35 *Tmt6* corresponded to known clades *Tmt1*, *Circe* and *Tcn2*, respectively. *Circe* gypsy
36 retrotransposon were identified till now only in arthropods and *Tcn2* was not yet identified in
37 Ascomycetes. *Tmt1* and *Tmt6* are the two clades that widely invaded the *T. melanosporum*
38 genome since they are the most abundant TEs representing 14 and 13 %, respectively, of this
39 genome. The *Tmt6* proliferation was more recent with 226 full-length copies identified and 62
40 that have not the same insertion points in seven *T. melanosporum* geographic accessions
41 resequenced genomes. However, transcriptional expression was not detected in these *Tmt6*
42 elements. All the *Tmts* have not a random distribution in the genome and their mode of
43 transposition is more likely random suggesting that each new copy can transpose. This study
44 allows having a better view of the gypsy retrotransposons diversity and dynamic in *T.*
45 *melanosporum*. If the major invasion waves are old some elements seems to have transposed
46 recently suggesting that retrotransposons could continue to play an important role in truffle
47 evolution.

48 **Introduction**

49 Repeated sequences are ubiquitous in all prokaryotic and eukaryotic genomes. Transposable
50 elements (TE) are the most frequent repeated sequences and they can constitute a large
51 proportion of fungal genomes. TE proliferation was proposed to explain the large genome size
52 found for Pucciniales species averaging 380.2 Mbp, while the average for Basidiomycota was
53 49.9 Mbp and was 37.7 Mbp for all fungi (Murat, Payen, et al. 2013; Tavares et al. 2014).
54 TEs are the most dynamic part the genomes and they are considered a major force in the
55 genomes evolution (Hua-Van et al. 2011). They can lead to chromosomal rearrangements,
56 speciation and evolution to stress (Capy et al. 2000; Kehrer-Sawatzki and Cooper 2006;
57 Kazazian 2004). TEs are also known to have different effect on genes, such as the inhibition
58 of their expression by insertions in the promoting sequences, the loss or gain of genes by TE
59 insertion and excision and the gain of a function by the exaptation of one TE gene by the host
60 organism (Kidwell and Lisch 2000). TEs are classified in retrotransposons (RNA
61 intermediate) and DNA transposons. Retrotransposons are composed by long terminal repeats
62 (LTRs), DIRS, Penelope-like, and short interspersed nuclear elements (SINEs) (Wicker et al.
63 2007). LTR retrotransposons are divided into several superfamilies: Copia (Pseudoviridae),
64 Gypsy (Metaviridae), Bel-Pao, Retrovirus (Retroviridae), and ERV (Wicker et al. 2007).
65 Gypsy retrotransposons populate many eukaryotic genomes, such as oomycetes (Judelson
66 2002), slime molds (Glöckner et al. 2001), fungi (Neuvéglise et al. 2002), plants (Suoniemi,
67 Tanskanen, and Schulman 1998), and animals (Bae et al. 2001). They have been classified
68 into three genera on the basis of an envelope gene (*env*) that is present (Errantivirus) or absent
69 (Metavirus) and the presence of a chromodomain (Chromovirus) (Malik and Eickbush 1999;
70 Marín and Lloréns 2000). According to Llorens and colleagues (2011), thirty-five clades of
71 gypsy retrotransposons have been identified in animals, fungi and plants. In fungi,
72 chromodomain-containing LTR retrotransposons or chromoviruses are the most frequent with
73 twenty clades identified (O. Novikova, Smyshlyaev, and Blinov 2010). Chromoviruses are
74 characterized by the presence of an additional domain - the chromodomain (CHD). The
75 evolutionary history and dynamics of chromoviruses has been explained by strict vertical
76 transmission in plants, while in fungi and metazoan it seems more complex (Gorinsek,
77 Gubensek, and Kordis 2004). Although chromoviruses have been described in different
78 fungal genomes (O. Novikova, Smyshlyaev, and Blinov 2010) their evolutionary history in a
79 single species is still largely unknown.

80 Thanks to the sequencing of its genome, the Black Truffle of Périgord (*Tuber melanosporum*)
81 became in the last years a model species for ectomycorrhizal Ascomycetes (Martin et al.
82 2010; Kües and Martin 2011). Truffles are ascomycetes fungi belonging to genus *Tuber* and
83 forming ectomycorrhizal associations with trees and shrubs. Truffles are also famous for their
84 hypogeous fruiting bodies, called ascocarps, particularly appreciated for their organoleptic
85 qualities. *T. melanosporum* is endemic of Mediterranean forest, but thanks to controlled
86 mycorrhization processes it has been introduced in many countries worldwide (Murat 2014).
87 With 125 Mbp the *T. melanosporum* genome is particularly large for a fungal species (Martin
88 et al. 2010). As for several lineages of filamentous plant pathogens (Raffaele and Kamoun
89 2012), *T. melanosporum* genome has been shaped by repeat-driven expansion of repeated
90 sequences corresponding to almost 60 % of its genome (Martin et al. 2010; Murat, Payen, et
91 al. 2013). While gypsy retrotransposons are the most frequent TE in *T. melanosporum*, their
92 diversity and evolutionary history are unknown.

93 In 2008, Riccioni and colleagues characterized a gypsy retrotransposon clade in
94 *T. melanosporum*, called *Tmt1*. *Tmt1* was *T. melanosporum* species specific and silent
95 (Riccioni et al. 2008). Similarly, a genomic suppressive subtractive hybridization of
96 *T. melanosporum* against *T. borchii* and *T. indicum* experiment identified several clones
97 specific to *T. melanosporum* corresponding to *Tmt1* (Murat et al. 2011). In the
98 *T. melanosporum* genome, 271 putative full-length gypsy retrotransposons have been
99 identified (Martin et al. 2010). Using the solo-LTR sequences and the substitution rate of 1.3
100 $\times 10^{-8}$ mutations per site per year (Ma and Bennetzen 2004), it was suggested that
101 *T. melanosporum* underwent at least two bursts of gypsy retrotransposons amplifications. A
102 burst taking place more than 10 Mya was followed by a gradual increased starting 5.5 Mya to
103 about 2.5 Mya (Martin et al. 2010). However, some full-length elements seem to have a more
104 recent insertion (see supplementary figure 7 in Martin et al. 2010). By the resequencing of six
105 new *T. melanosporum* geographic accessions genome, we showed that gypsy retrotransposons
106 generated genome diversity since 54 % of the single nucleotide polymorphisms were
107 identified in them (Payen et al. 2015). The aim of this study was to take advantage of the
108 genomic resources available for *T. melanosporum* to investigate the diversity of the gypsy
109 retrotransposons in this species and their evolutionary history. The number of gypsy clades
110 was addressed by isolating the reverse transcriptase sequence in *T. melanosporum* gypsy
111 retrotransposons sequences. The evolutionary history was addressed by 1) the birth-death

112 models that have been developed to infer speciation and extinction rates in phylogenies (Yule
113 1925; Kendall 1948; Nee, May, and Harvey 1994); 2) the tree imbalance (Le Rouzic, Payen,
114 and Hua-Van 2012)) and 3) the investigation of the full length gypsy insertion point
115 differences in the genome of six *T. melanosporum* geographic accessions recently
116 resequenced (Payen et al. 2015).

117

118 **Materials and Methods**

119 *T. melanosporum* repeated sequences library generation

120 REPEATSCOUT (Price, Jones, and Pevzner 2005) was used to identify *de novo* repetitive
121 DNA sequences in the *T. melanosporum* Mel28 genome version 1.0 (Martin et al. 2010). The
122 default parameters (with $l = 15$) were used. REPEATSCOUT library was then filtered as
123 follows: (1) all the sequences less than 100 bp in size were discarded; (2) repeats having less
124 than ten copies in the genome were removed (as they may correspond to protein-coding gene
125 families); (3) repeats having significant hits to known proteins in UNIPROT (UniProt
126 Consortium 2008) other than proteins known as belonging to TEs were removed. The
127 consensus sequences remaining were manually annotated by a TBLASTX search (Altschul
128 1990) against REPBASE (version 16.03; Jurka et al. 2005). The 271 putative full-length
129 gypsy retrotransposons identified using LTR_STRUC (McCarthy and McDonald 2003) in
130 Martin et al. (2010) were also included in this study. The number of occurrences and the
131 percent of genome coverage were assessed by masking the genome assemblies using
132 REPEATMASKER version open-4.0.5 (Smit, Hubley, and Green 1996) with the consensus
133 sequences coming from the REPEATSCOUT and the LTR_STRUC pipelines. The
134 REPEATSCOUT consensus sequences and the putative full-length gypsy elements are
135 available in DRYAD ().

136

137 *Isolation of reverse transcriptase and gypsy retrotransposons clades identification*

138 In the consensus sequences coming from the REPEATSCOUT and the LTR_STRUC
139 pipelines described above only those having the corresponding canonical protein sequences
140 with gypsy retrotransposons were selected. The PFAM profile number 00078 corresponding
141 to reverse transcriptase (RT) was used with RPSBLAST (version 2.2.28+) to isolate RT
142 sequences in the selected sequences. Only sequences more than 80 amino acid were kept for

143 further analyses. The resulting sequences were then used in a TBLASTN (version 2.2.28+)
144 against REPBASE to confirm they corresponded to gypsy retrotransposons.
145 The resulting protein sequences were then clustered using USEARCH version 7 (Edgar 2010)
146 with the 80-80-80 rule described by Wicker and colleagues (2007). This rule state that
147 sequences having 80% of their sequences with more than 80% of similarity for a fragment of
148 at least 80 bp long belong to the same clade. For each cluster generated by USEARCH the
149 consensus sequence corresponding to the centroid sequence was selected.
150 In order to check if the resulting consensus sequences correspond to know clades, RT of
151 known gypsy retrotransposons clades were downloaded from the gypsy database (Llorens et
152 al. 2011) and Novinska et al. (2010). A global alignment with all the consensus and reference
153 sequences was realized using CLUSTAL OMEGA (Sievers et al. 2011) with default
154 parameters. A neighbor joining reconstruction was realized using QUICKTREE (Howe,
155 Bateman, and Durbin 2002) with default parameters and 10,000 bootstraps replication. The
156 resulting tree was collapsed using python library *ete2* (Huerta-Cepas, Dopazo, and Gabaldón
157 2010) for the nodes with less than 40% bootstrap value. The tree was manually edited using
158 FIGTREE 1.4.2 (<http://tree.bio.ed.ac.uk/software/figtree/>). Each *T. melanosporum* RT cluster
159 was assigned to the closest known gypsy retrotransposon clade if the bootstrap value was
160 above 50 %. We decided to name each *T. melanosporum* gypsy retrotransposon clade as
161 proposed in Riccioni and colleagues (2008): TmtX (T: Tuber; m: melanosporum; t:
162 transposon; X: number of the clade). A subsequent BLASTN search of all REPEATSCOUT
163 and LTR_STRUC sequences (see above) against themselves was realized to assign the
164 sequences missing a RT to the different clades.
165 The percentage of genome coverage for each clade was calculated using REPEATMASKER
166 as described above.
167 For all the putative full-length gypsy retrotransposons the identification of the presence of a
168 dUTPase, integrase, protease and chromodomain has been realized by TBLASTN of the DNA
169 sequences of the elements against the respective protein sequences downloaded from gypsy
170 database (Llorens et al. 2011). A PFAM search for known motifs was also realized
171 (<http://pfam.xfam.org/search>).
172
173 *Maximum likelihood phylogeny of gypsy retrotransposon clades*

174 For this analysis all the copies for each *Tmts* families longer than 5,000 bp were isolated. In
175 order to have a reliable alignment and considering the low level of nucleotide conservation
176 among copies (Data not shown), a clustering was realized using USEARCH version 8.0.1517
177 (<http://www.drive5.com/usearch/manual/>) with the *cluster_fast* command and id 0.7
178 parameter specifying the minimum sequence identity of a hit to be clustered. The cluster with
179 the highest number of sequences was selected for further analyses. All the sequences for the
180 cluster with the highest number of sequences for *Tmt1*, *Tmt2*, *Tmt3* and *Tmt6* were aligned
181 using MAFFT v7.164b (Kato and Standley 2013). Due to the low number of sequences for
182 *Tmt4* and *Tmt5* we decided to exclude these two clades in the further analyses. Maximum
183 likelihood (ML) phylogenies were realized with R (version 3.1.1) using packages *ape* version
184 3.1-4 (Paradis, Claude, and Strimmer 2004) and *phangorn* version 1.99-7 (Schliep 2011). The
185 model of substitution GTR (general time reversible) + G (gamma) + I (invariable) model of
186 substitution was selected according to the function *modelTest()* from the *phangorn*.
187 Ultrametric tree were calculated from the ML trees using the *pathd8* method (Britton et al.
188 2007).

189

190 *Birth-Death model*

191 A birth-death model described the stochastic branching process in which branches can split or
192 disappear in the course of time. In phylogenies, branching events correspond to speciation and
193 disappearance of branch to death. For TE, the branching events are considered as
194 transposition and extinct branch as deletion events (Le Rouzic, Payen, and Hua-Van 2012).
195 The “pure birth” model or Yule process was tested using the function *yule()* from the R
196 package *ape*. The “birth-death” model was tested by the function *bd()* from the R package
197 *laser* version 2.4-1 (Rabosky 2007).

198

199 *Tree balance*

200 Using the ML phylogenetic trees produced for the gypsy clades (see above) it is possible to
201 investigate the tree balance (or imbalance) meaning for a perfectly balanced tree all branches
202 duplicate once and for unbalanced tree all duplications happens in the same branch. For TE,
203 balanced trees occurred when all copies can duplicate at the same rate corresponding to
204 random template model. In this model the phylogeny consisted of numerous clades (Clough et
205 al. 1996). For unbalanced tree one copy in the genome is able to duplicate corresponding to a

206 “master copy”, i.e. the strict master model. This model consisted as a single uninterrupted
207 lineage (Clough et al. 1996). The quantification of TE phylogeny balance leads therefore to
208 meaningful insights on transposition history (Le Rouzic, Payen, and Hua-Van 2012).
209 The tree balance was assessed by β -index. The ML phylogeny were midpoint rooted with the
210 function *midpoint()* from R package *Phangorn*. The rooted trees were then transformed by
211 the function *as.treeshape()* from R package *apTreeshape* version 1.4-5 (Bortolussi et al.
212 2006). The β -index were computed using function *maxlik.betasplit()* from R package
213 *apTreeshape* with 500 bootstrap. For balanced random trees there are two traditional models
214 proposed: the “Proportional to Distinguishable Arrangement” (PDA) model and the “Equal
215 Rate Markov” (ERM) model corresponding to trees generated by Yule process. Trees
216 generated by ERM model have a β -index of 0 while trees generated by PDA model have a β -
217 index of -1.5 (Le Rouzic, Payen, and Hua-Van 2012). The interpretation of the β -index can be
218 done as follow: imbalanced trees ($-2 < \beta < -1.5$), random trees ($-1.5 < \beta < 0$) and trees too
219 perfectly balanced to be random ($0 < \beta < \infty$).

220

221 *Lineage through time and skyline plots*

222 The lineages through time (LTT) analyses were realized using the ultrametric trees (see
223 above) and the function *ltt.plot()* from the *ape* R package. The simulation of the Yule model
224 was realized using the function *pbtree()* from the R package *phytools* version 0.4-31 (Revell
225 2012) with the following parameters: the birth rate (b) found with function *yule()* from the R
226 package *ape* as described above, the extinction rate was set to 0 (d=0), one tree was simulated
227 (nsim=1), the tree was simulated with continuous time (type = continuous) and the scale was
228 set to 1 (scale = 1). Skyline plots display the effective population size along a time axis given
229 a graphical representation of population demographic history (Ho and Shapiro 2011). Skyline
230 plots have already been used to assess HIV-1 populations (Holmes 2004) and mosquito’s
231 transposons demography (Struchiner et al. 2009). Generalized skyline plots (Strimmer and
232 Pybus 2001) were draw from the ultrametric trees and *ape* R package. First, the coalescent
233 intervals information were extracted with the function *coalescent.intervals()*. The function
234 *find.skyline.epsilon()* was used to estimate epsilon parameters that was used in the function
235 *skyline()*.

236

237 *Evaluation of Tmts insertions in T. melanosporum genomic sequences*

238 The distribution of *Tmts* in the *T. melanosporum* genome was assessed using the coefficient
239 of aggregation described by Hopkins (Hopkins and Skellam 1954) corresponding to the ratio
240 of the square of the mean distance between a random point and its nearest neighbouring
241 individual (P) to the square of the mean distance between an individual and its nearest
242 neighbouring individual (I), i.e. $A = \Sigma P^2 / \Sigma I^2$. For random distribution the coefficient is equal
243 to one, greater than one for aggregation distribution, and less than one for regular distribution.
244 For each *Tmt* the total number of copies found by REPEATMASKER (see above) was
245 considered and a file in bed format (<http://genome.ucsc.edu/FAQ/FAQformat.html#format1>)
246 was generated. For each *Tmt* a bed file with random intervals was generated with *random*
247 function from BEDTOOLS version 2.20.1 (Quinlan and Hall 2010). The number of generated
248 intervals was the same than those of each *Tmt* and the length of intervals was set to 1. The bed
249 files were merged to eliminate overlapping using *merge* function from BEDTOOLS with
250 option *-d 5*, i.e. maximum distance of 5 bp between features to be merged. The interdistances
251 P and I for each *Tmt* were calculated with *closest* function from BEDTOOLS. The deviation
252 of A from one was tested for significance by calculating the parameter $x = A/(1+A)$ that tends
253 to normality as n (i.e. the number of pairs of observation) increases. If n is greater than 50, (x-
254 0.50) may be treated as a normal deviate with zero mean and standard deviation $1/(2\sqrt{(2n+1)})$.
255 For n less than 50 a specific diagram given by Hopkins can be used (Hopkins and Skellam
256 1954). Kolmogorov-Smirnov test was used to test if (x-0.50) deviate from normal distribution
257 using *ks.test()* function from R package *stats* version 3.1.1.

258

259 Investigation of the gypsy retrotransposon insertion sites among seven *T. melanosporum*
260 geographic accessions

261 The full-length gypsy retrotransposons elements dynamic in *T. melanosporum* was also
262 investigated using the genomic data of six *T. melanosporum* geographic accession recently
263 resequenced available in the NCBI sequence reads archive under accession SRP044130
264 (Payen et al. 2015). For each full-length gypsy retrotransposon elements the first 50 bp of 5'
265 and 3' LTR as well as the first 50 bp of the non-LTR neighbour sequence was extracted (Fig.
266 S1). This leads for each element to have two sequences of 100bp corresponding to the 5' and
267 3' insertion sequence. Then, the illumina reads for each geographic accession were mapped
268 against these sequences using ERNE-MAP version 1.3 (Vezi et al. 2012) allowing only two
269 SNPs per reads and a single mapping for each reads. Using "Integrative Genome Viewer"

270 (Robinson et al. 2011) the alignments were manually visualized to identify those with no
271 reads overlapping the LTR and non-LTR genomic region (Fig. S1).

272

273 **Results**

274

275 *Number of gypsy clades in T. melanosporum genomes*

276 Among the 1,199 consensus sequences produced by REPEATSCOUT, 59 were identified as
277 gypsy LTR due to their homology in REPBASE. LTR_STRUCT identified 271 putative full-
278 length gypsy LTR (Martin et al. 2010). Using the 59 REPEATSCOUT consensus sequences
279 and the 271 putative full-length gypsy LTR, REPEATMASKER identified 37 % of the *T.*
280 *melanosporum* genome assembly corresponding to gypsy LTR. A total of 173 RT sequences
281 were isolated corresponding to 158 and 15 RT from full-length gypsy LTR and
282 REPEATSCOUT consensus sequences, respectively (Table S1). These RT amino acid
283 sequences were grouped in 13 clusters. The neighbour-joining reconstruction allowed
284 identifying six clades for the *T. melanosporum* gypsy retrotransposons RT (Fig. 1 and Table
285 1). Only the REPEATSCOUT element number 7 presented two RT belonging to two different
286 clades (*Tmt1* and *Tmt2*; data not showed). We decided to separate this element in two for
287 further analysis.

288 One clade corresponded to the previously described *Tmt1* gypsy (Riccioni et al. 2008). The
289 five other clades were named *Tmt2* to *Tmt6*. *Tmt2* and *Tmt6* grouped with *Circe* and *Tcn2*,
290 respectively. *Tmt3*, *Tmt4* and *Tmt5* did not grouped with known RT retrotransposons clades
291 (Fig. 1).

292 *Tmt1* and *Tmt6* were the most frequent clades in the *T. melanosporum* genomes
293 corresponding to 14 and 13% of the genome (Table 1). Other clades were less frequent with 7,
294 2, 0.05 and 0.04% for clades *Tmt3*, *Tmt2*, *Tmt4* and *Tmt5*, respectively. Among the 271 full-
295 length gypsy LTR, 224, 26, 16 and 4 elements were identified using their RT as *Tmt6*, *Tmt1*,
296 *Tmt3*, and *Tmt5* (Table 1). Full-length elements were not identified for *Tmt2* and *Tmt4*.

297 The putative structure of the six clades is presented in Fig 2. *Tmt1* is composed of four
298 protein-coding genes (protease, dUTPase, RT and integrase). A chromodomain was also
299 identified in some copies. *Tmt2* is composed of three protein-coding genes (exo-endo-
300 phosphatase, RT and RNaseH). Three protein-coding genes (RT, dUTPase and integrase) and
301 a chromodomain were identified in *Tmt3*. For *Tmt4*, three protein-coding genes (RT, dUTPase

302 and integrase) and two domains (Zf-CCHC and chromodomain) were identified. *Tmt5* is
303 composed of three protein-coding genes (protease, RT and integrase) and a chromodomain.
304 Finally, *Tmt6* is composed of three protein-coding gene (protease, RT and integrase) and two
305 domains (Zf-CCHC and chromodomain). As presented in supplementary table 1 the presence
306 of all these protein-coding genes and domains was not conserved in the putative full-length
307 elements.

308 The RNAseq data (Tisserant et al. 2011) indicated that the level of expression of full-length
309 gypsy LTRs is low or null at the exception of one *Tmt1* element (8656_RT42_B24_L24_236;
310 Table S1). However, this element did not harbour known retrotransposons protein coding
311 genes, instead it harbour a non-transposon protein suggesting this is 1) a misidentification of a
312 full-length gypsy LTR or 2) an elements that capture a protein coding gene.

313

314 *Historical demographic evolution: birth-death models, lineage through time, skyline plots*
315 *and tree imbalance*

316 In order to investigate their evolutionary history, 1,135, 256, 255, 7, 7 and 1,363 sequences
317 longer than 5,000 bp were isolated for *Tmt1*, *Tmt2*, *Tmt3*, *Tmt4*, *Tmt5* and *Tmt6*, respectively.
318 *Tmt4* and *Tmt5* were not included in this analysis due to the low number of sequences
319 isolated.

320 Due to the structural polymorphisms observed in putative full-length elements (Table S1) and
321 the low level of sequence conservation among these sequences (data not shown), for each
322 *Tmts* a previous clustering was realized. The cluster with highest number of reads (320, 115,
323 108 and 517 sequences for *Tmt1*, *Tmt2*, *Tmt3* and *Tmt6*, respectively) was selected to
324 generated a maximum likelihood (ML) reconstruction. We first applied a “pure birth” model
325 (i.e. constant transposition rate and no deletion). The estimated diversification rate ranged
326 from 1.61 to 3.62 for *Tmt3* and *Tmt6*, respectively (Table 2). Using the “birth-death” model
327 (allowing both transposition and deletion) the estimated diversification rate (i.e. similar to
328 transposition rate of the transposons) ranged from 1.26 to 3.56 for *Tmt3* and *Tmt6*,
329 respectively, although no significant deletions (i.e. disappearance rate of the transposons)
330 were detected. The ultrametric trees suggested changes in the molecular evolution of the
331 different branches (i.e. differences in branch length; Fig. 3A). The resulting LTT plots suggest
332 departure from simple Yule model (Fig. 3B). The curves of the four *Tmts* are above the pure
333 birth simulation model suggesting the past rate of transposition was higher than the current

334 one. This is in agreement with the insertional time estimation based on the genetic distance
335 between both solo-LTR of each element (Fig. 4) and considering that only one full-length
336 element has no polymorphism in its LTRs (number 6239_RT32_B24_L22_98) (Martin et al.
337 2010). Figure 3C presents the different profiles under which the effective population size
338 (Emerson, Paradis, and Thébaud 2001) unfolds in time. *Tmt1* and *Tmt2* exemplify
339 amplification trajectories that rise explosively at first, and continuously increasing thereafter.
340 However, *Tmt1* exhibits a recent decreasing in effective population size. *Tmt4* and *Tmt6* also
341 exemplify amplification trajectories, follow by an increase that is interrupt to increase again
342 suggesting different proliferation waves.

343

344 The quantification of TE phylogeny balance leads to meaningful insights on transposition
345 history. Indeed, for unbalanced trees, one copy in the genome is able to duplicate
346 corresponding to a “master copy” and for balanced trees, all copies can duplicate at the same
347 rate corresponding. The quantification of tree balance was estimated by the β -index that gave
348 the following results: $\beta_{Tmt1} = -1.07$ (-1.21, -0.92), $\beta_{Tmt2} = -1.07$ (-1.31, -0.80), $\beta_{Tmt3} = -0.85$ (-
349 1.14, -0.39) and $\beta_{Tmt6} = -1.17$ (-1.27, -1.08). All β estimates are consistent with random trees.
350 The estimates were more precise for *Tmt1* and *Tmt6* (lower support interval) confirming that
351 these analyses are more robust with larger trees (Le Rouzic, Payen, and Hua-Van 2012).

352

353 *Tmts are not distributed randomly in the T. melanosporum genome*

354 The coefficient of aggregation described by Hopkins (1954) showed that *Tmt1*, *Tmt2*, *Tmt3*,
355 *Tmt4* and *Tmt6* deviate from random distribution in the *T. melanosporum* genome (Fig 5).
356 *Tmt1*, *Tmt3*, *Tmt4* and *Tmt6* elements are highly aggregated although *Tmt2* have a regular
357 distribution with important inter-element distances compared to other *Tmts* (Fig. 5 and 6).
358 Due to the low number of *Tmt5* elements it was not possible to calculate the coefficient of
359 aggregation since random interval were positioned in different scaffold. But interestingly
360 several *Tmt5* elements are present in the same scaffold suggesting that *Tmt5* also has a
361 tendency of aggregation (Fig. 5).

362

363 *Dynamic of full-length elements insertion in seven T. melanosporum geographic accessions*

364 We have also investigated the insertion points of the full-length gypsy elements in seven
365 *T. melanosporum* geographic accession genomes (Payen et al. 2015). Among the 271 putative

366 full length gypsy identified in the Mel28 genome (Martin et al. 2010), 204 were identified at
367 the same localization in the resequenced geographic accessions since sequence reads
368 overlapping the LTR and contiguous non-LTR junction sequences were found (Fig S1). The
369 lack of sequence reads overlapping the LTR and contiguous non-LTR junction sequences for
370 at least one genome was found for 67 elements (Table S2), two belong to *Tmt1*, three to *Tmt3*
371 and 62 to *Tmt6* (Table S2). Two *Tmt6* elements (6239_RT_B24_L22_98 and
372 17999_RT32_B24_L22_276) were not found at the same localization in all the resequenced
373 geographic accession genomes (Table S2). Interestingly, the insertion of these elements was
374 estimated as recent in Mel28 (0 and 135,428 years ago; Table S2) when both solo-LTRs were
375 compared (Martin et al. 2010). These LTRs likely inserted at this current location in Mel28
376 after the separation of the analysed geographic accessions suggesting that these LTRs were
377 actives recently.

378

379 **Discussion**

380

381 *T. melanosporum* genome hosted six classes of gypsy LTR

382 The *T. melanosporum* genome was invaded by gypsy retrotransposons several millions years
383 ago (Martin et al. 2010). In this study, we used the RT sequences to identify the different
384 gypsy LTR clades present in this genome. RT is the most highly conserved of the
385 retrotransposon proteins and it is commonly used to classify retrotransposons (Havecker, Gao,
386 and Voytas 2004). Our analysis demonstrated that six different clades of gypsy
387 retrotransposons are presents in the *T. melanosporum* genome. Their organisation in term of
388 protein-coding genes and conserved domains is different.

389 *Tmt1* and *Tmt6* expanded in the *T. melanosporum* genome since they corresponded to more
390 than 10 % of its genome. *Tmt1* was the first gypsy retrotransposon identified for *T.*
391 *melanosporum* (Riccioni et al. 2008). This clade was not identified by southern blot in other
392 *Tuber* species and genomic suppression subtractive hybridization (Murat et al. 2011)
393 suggesting that *Tmt1* is *T. melanosporum* specific. *Tmt1* presented 26 full-length copies but
394 their insertion are ancient (Fig. 4; Table S1). This was confirmed by a recent decrease in the
395 effective population size (Fig. 3C). The phylogenetic position of *Tmt1* suggested it belongs to
396 chromovirus (Fig. 1 and Riccioni et al. 2008) although Riccioni et al. (2008) did not detect a
397 chromodomain. In our study, we detected one fragmented copy of *Tmt1* in scaffold 67

398 (positions 266819 to 271029) that presented a chromodomain (Fig. S2). Using genomic data
399 of six resequenced *T. melanosporum* geographic accessions (Payen et al. 2015), two elements
400 did not present overlapping reads in their insertion point (Table S2). This could be explained
401 by insertion of these elements in the Mel28 reference genome after its separation with
402 geographic accessions, the removal of these elements in resequenced geographic accessions
403 or an accumulation of polymorphisms avoiding the mapping of reads. Since this is only in one
404 of both LTR insertion point it is more probably due to polymorphism after the insertion.
405 These results suggested that the *T. melanosporum* genome invasion by *Tmt1* is old and
406 probably this clade is not yet active as showed by transcriptomic analyses (Table S1 and
407 Riccioni et al. 2008).

408
409 *Tmt1* harboured a dUTPase between the protease and the RT (Fig. 2; Riccioni et al. 2008)
410 while *Tmt3* presented a dUTPase located between the RT and integrase (Fig. 2). dUTPase is
411 an essential factor in DNA metabolism and repair (Barabás et al. 2003). Retroviral dUTPase
412 genes are located in non-primate lentiviruses and betaretroviruses at different genomic
413 locations (McGeoch 1990; Baldo and McClure 1999). Lentiviral dUTPase genes are between
414 RT and integrase genes (Turelli et al. 1997) suggesting *Tmt3* has a similar organization than
415 lentivirus. The organization of the *Tmt1* with dUTPase between protease and RT is similar to
416 those of *Pc_Metavir 4* and *Pc_Metavir 5* gypsy retrotransposons elements from
417 *Phaenoerochaete chrysosporum* (O. S. Novikova and Blinov 2008) and a *MarY1* element
418 from *Postia placenta* (*PosPlaTy3-5*; Novikova et al. 2010). However, the dUTPase
419 phylogeny realized by Riccioni et al. (2008) and Novikova et al. (2010) did not allow
420 identifying the origin of the *Tmt1* dUTPase.

421
422 *Tmt6* is the second most frequent gypsy retrotransposons and it is related to *Tcn2* (Goodwin
423 and Poulter 2001). *Tcn2* has been described in *Cryptococcus neoformans* and was thought to
424 be specific of basidiomycetes (Novikova, Smyshlyaev, and Blinov 2010; Murat et al. 2013).
425 The identification of this clade in *T. melanosporum* demonstrated that *Tcn2*-like
426 retrotransposons are also present in ascomycetes. Future investigations are needed to know if
427 *Tcn2*-like retrotransposons are found only in *T. melanosporum* or also in other truffle
428 genomes. Sixty-two *Tmt6* full-length elements seem to have different insertion points in at
429 least one geographic accession. Interestingly, two recent elements, belonging to *Tmt6*, were

430 present only in the Mel28 genome suggesting that for those elements an insertion in Mel28
431 genome may have occurred after the separation of the geographic accession. This result
432 suggests that *Tmt6* activity occur recently (i.e. few thousand years ago) and that some
433 retrotransposons could also continue to drive truffle evolution. This is confirmed by the recent
434 observed exponential increase in its effective population size (Fig. 3C). Unfortunately, the
435 sequencing strategy used was not suitable to address more in detail transposon diversity since
436 *de novo* assembly was not possible (Payen et al. 2015).

437

438 At the exception of *Tmt2*, all the other clades belong to chromovirus. Non chromovirus gypsy
439 retrotransposons are not frequent in fungi; indeed an analysis of 31 fungal genomes RT
440 identified only one non chromovirus clade (i.e. *Cigr*) present in one fungal species,
441 *Melampsora larici-populina* (Murat, Payen, et al. 2013). *Tmt2* is related to *Circe*
442 retrotransposon that has been described in *Drosophila melanogaster* (Losada et al. 1999).
443 *Circe* belongs to clade *Oswaldo* according to gypsy database (Llorens et al. 2011) and until
444 now gypsy retrotransposons of the *Oswaldo* clade have been described only in arthropods. It
445 has been suggested that gypsy retrotransposons originated early after the first eukaryotes and
446 reached their current wide dispersal by a combination of vertical and horizontal transmission
447 (Kumar and Bennetzen 1999). Recently, El Baidouri and colleagues (2014) identified
448 numerous horizontal transfer of LTR retrotransposons between plants species, suggesting that
449 plants can frequently exchange genetic material through horizontal transfers and that this
450 mechanism may be important in TE-driven genome evolution. The truffle aroma is known to
451 attract insects such as *Leiodes cinnamomea* and *Suillia gigantea* (Hochberg et al. 2003). This
452 last species is famous as an alternative to dogs for truffle hunting since it uses truffle
453 ascocarps to lay eggs. The acquisition of gypsy retrotransposon clades in fungi by horizontal
454 transfer has been proposed for *Tcn1* gypsy retrotransposons from non-seed plant to fungi (O.
455 Novikova, Smyshlyayev, and Blinov 2010). Due to this close relation between truffle and
456 arthropods, we may speculate that *Circe*-like gypsy retrotransposons could have been
457 transferred to *T. melanosporum* by horizontal transfer from arthropods. The sequencing of
458 new truffle species genomes (Payen et al. 2015) as well as the availability of other
459 Pezizomycete genomes will allow gaining more information on the gypsy diversity as well as
460 the transversal versus horizontal acquisition of gypsy retrotransposon in this fungal class.

461

462 *Historical dynamic of Tmts in T. melanosporum genome*

463 Phylogenies realized with copies of *Tmt1*, *Tmt2*, *Tmt3* and *Tmt6* longer than 5,000 bp
464 produced random trees suggesting most copies are able to transpose and to generate new
465 branches. This process is widespread for TE phylogenies (Perdue and Nuzhdin 2000;
466 Cordaux, Hedges, and Batzer 2004; Le Rouzic, Payen, and Hua-Van 2012). The *Tmts*
467 elements did not presented a random distribution in the *T. melanosporum* genome. Although a
468 random insertion was recently found for the DNA transposon *Tn7* (Green et al. 2012), many
469 transposons have preferential insertion sites. For example, the retrotransposon *Tos17* from
470 rice presented hotspots of insertions (Miyao et al. 2003) and *RIRE-7*, a gypsy retrotransposon
471 of rice is preferentially inserted in centromeric regions (Kumekawa et al. 2001). Gypsy LTRs
472 of fungi were proposed to preferentially insert in heterochromatin whereas copia LTR have
473 been reported to integrate into transcriptionally active regions of the genome (Muszewska,
474 Hoffman-Sommer, and Grynberg 2011).

475 The estimated rate of transposition was similar for *Tmt1*, *Tmt2* and *Tmt3* although it is about
476 three times higher for *Tmt6*. The demographic dynamic of these clades is therefore different.
477 Since the curves in LTT for the four *Tmts* are above the pure birth simulation model, the past
478 rate of transposition was probably higher than the current one. This result supports an old
479 colonization of the *T. melanosporum* genome by gypsy LTRs (Martin et al. 2010) although
480 evidence of recent transposition has been found for *Tmt6* clade suggesting that this clade was
481 active recently. However, we did not detected significant level of transcription in putative
482 full-length element. This could be due to an inactivation of these sequences by methylation
483 induced premeiotically (MIP). Indeed, the level of methylation was high in *T. melanosporum*
484 gypsy LTR sequences and it has been shown that the level of expression is negatively
485 correlated with the level of methylation (Montanini et al. 2014).

486
487 To conclude, gypsy retrotransposons contributed greatly to shape the *T. melanosporum*
488 genome. They also generated genetic diversity with a higher rate of polymorphisms in these
489 elements among geographic isolates (Payen et al. 2015). Transposons are known to be
490 activated during sexual reproduction and also during environmental stresses (Daboussi and
491 Capy 2003). *T. melanosporum* was demonstrated to invest mainly in sexual reproduction
492 compared to vegetative spread (Murat, Rubini, et al. 2013) and climate changes affect this
493 species (Büntgen et al. 2012; Le Tacon et al. 2014). It cannot be excluded that active gypsy

494 clade (e.g. *Tmt6*) could continue to colonize and invade the *T. melanosporum* genome in the
495 future. The population genomic studies currently started in our laboratory will allow
496 addressing the impact of such TE activity in this species.

497

498 **Acknowledgements**

499 The UMR1136 is supported by a grant overseen by the French National Research Agency
500 (ANR) as part of the "Investissements d'Avenir" program (ANR-11-LABX-0002-01, Lab of
501 Excellence ARBRE). This study benefited from ANR SYSTERRA SYSTRUF (ANR-09-
502 STRA-10). The PhD fellow of Thibaut Payen is co-funded by Lorraine Region and the
503 European Commission through the EcoFINDERS project (FP7-264465).

504

505 **References**

506

507 Altschul, S. 1990. "Basic Local Alignment Search Tool." *Journal of Molecular Biology* 215 (3): 403–10.
508 doi:10.1006/jmbi.1990.9999.

509 Bae, Young-An, Seo-Yun Moon, Yoon Kong, Seung-Yull Cho, and Mun-Gan Rhyu. 2001. "CsRn1, a Novel
510 Active Retrotransposon in a Parasitic Trematode, Clonorchis Sinensis, Discloses a New Phylogenetic
511 Clade of Ty3/gypsy-like LTR Retrotransposons." *Molecular Biology and Evolution* 18 (8): 1474–83.

512 Baldo, Angela M., and Marcella A. McClure. 1999. "Evolution and Horizontal Transfer of dUTPase-Encoding
513 Genes in Viruses and Their Hosts." *Journal of Virology* 73 (9): 7710–21.

514 Barabás, Orsolya, Michaela Rumlová, Anna Erdei, Veronika Pongrácz, Iva Pichová, and Beáta G. Vértessy.
515 2003. "dUTPase and Nucleocapsid Polypeptides of the Mason-Pfizer Monkey Virus Form a Fusion
516 Protein in the Virion with Homotrimeric Organization and Low Catalytic Efficiency." *Journal of*
517 *Biological Chemistry* 278 (40): 38803–12. doi:10.1074/jbc.M306967200.

518 Bortolussi, Nicolas, Eric Durand, Michael Blum, and Olivier François. 2006. "apTreeshape: Statistical Analysis
519 of Phylogenetic Tree Shape." *Bioinformatics* 22 (3): 363–64. doi:10.1093/bioinformatics/bti798.

520 Britton, Tom, Cajsja Lisa Anderson, David Jacquet, Samuel Lundqvist, and Kåre Bremer. 2007. "Estimating
521 Divergence Times in Large Phylogenetic Trees." *Systematic Biology* 56 (5): 741–52.
522 doi:10.1080/10635150701613783.

523 Büntgen, Ulf, Simon Egli, J. Julio Camarero, Erich M. Fischer, Ulrich Stobbe, Håvard Kausrud, Willy Tegel,
524 Ludger Sproll, and Nils C. Stenseth. 2012. "Drought-Induced Decline in Mediterranean Truffle
525 Harvest." *Nature Climate Change* 2 (12): 827–29. doi:10.1038/nclimate1733.

526 Capy, P, G Gasperi, C Biéumont, and C Bazin. 2000. "Stress and Transposable Elements: Co-Evolution or Useful
527 Parasites?" *Heredity* 85 (Pt 2) (March): 101–6.

- 528 Clough, James E., James A. Foster, Michael Barnett, and Holly A. Wichman. 1996. "Computer Simulation of
529 Transposable Element Evolution: Random Template and Strict Master Models." *Journal of Molecular*
530 *Evolution* 42 (1): 52–58. doi:10.1007/BF00163211.
- 531 Cordaux, Richard, Dale J. Hedges, and Mark A. Batzer. 2004. "Retrotransposition of Alu Elements: How Many
532 Sources?" *TRENDS in Genetics* 20 (10): 464–67.
- 533 Daboussi, Marie-Josée, and Pierre Capy. 2003. "Transposable Elements in Filamentous Fungi." *Annual Review*
534 *of Microbiology* 57 (January): 275–99. doi:10.1146/annurev.micro.57.030502.091029.
- 535 Edgar, Robert C. 2010. "Search and Clustering Orders of Magnitude Faster than BLAST." *Bioinformatics* 26
536 (19): 2460–61. doi:10.1093/bioinformatics/btq461.
- 537 El Baidouri, Moaine, Marie-Christine Carpentier, Richard Cooke, Dongying Gao, Eric Lasserre, Christel Llauro,
538 Marie Mirouze, Nathalie Picault, Scott A. Jackson, and Olivier Panaud. 2014. "Widespread and
539 Frequent Horizontal Transfers of Transposable Elements in Plants." *Genome Research* 24 (5): 831–38.
- 540 Emerson, Brent C., Emmanuel Paradis, and Christophe Thébaud. 2001. "Revealing the Demographic Histories
541 of Species Using DNA Sequences." *Trends in Ecology & Evolution* 16 (12): 707–16.
542 doi:10.1016/S0169-5347(01)02305-9.
- 543 Glöckner, Gernot, Karol Szafranski, Thomas Winckler, Theodor Dinger, Michael A. Quail, Edward Cox,
544 Ludwig Eichinger, Angelika Anna Noegel, and André Rosenthal. 2001. "The Complex Repeats of
545 Dictyostelium Discoideum." *Genome Research* 11 (4): 585–94. doi:10.1101/gr.162201.
- 546 Goodwin, Timothy JD, and Russell Poulter. 2001. "The Diversity of Retrotransposons in the Yeast
547 *Cryptococcus Neoformans*." *Yeast* 18 (9): 865–80.
- 548 Gorinsek, Benjamin, Franc Gubensek, and Dusan Kordis. 2004. "Evolutionary Genomics of Chromoviruses in
549 Eukaryotes." *Molecular Biology and Evolution* 21 (5): 781–98. doi:10.1093/molbev/msh057.
- 550 Green, Brian, Christiane Bouchier, Cécile Fairhead, Nancy L. Craig, and Brendan P. Cormack. 2012. "Insertion
551 Site Preference of Mu, Tn5, and Tn7 Transposons." *Mobile DNA* 3 (1): 3.
- 552 Havecker, Ericka R., Xiang Gao, and Daniel F. Voytas. 2004. "The Diversity of LTR Retrotransposons."
553 *Genome Biology* 5 (6): 225. doi:10.1186/gb-2004-5-6-225.
- 554 Ho, Simon Y. W., and Beth Shapiro. 2011. "Skyline-Plot Methods for Estimating Demographic History from
555 Nucleotide Sequences." *Molecular Ecology Resources* 11 (3): 423–34. doi:10.1111/j.1755-
556 0998.2011.02988.x.
- 557 Hochberg, Michael E., Guillaume Bertault, Karine Poitrineau, and Arne Janssen. 2003. "Olfactory Orientation of
558 the Truffle Beetle, *Leiodes Cinnamomea*." *Entomologia Experimentalis et Applicata* 109 (2): 147–53.
- 559 Holmes, Edward C. 2004. "The Phylogeography of Human Viruses." *Molecular Ecology* 13 (4): 745–56.
560 doi:10.1046/j.1365-294X.2003.02051.x.
- 561 Hopkins, Brian, and J. G. Skellam. 1954. "A New Method for Determining the Type of Distribution of Plant
562 Individuals." *Annals of Botany* 18 (2): 213–27.
- 563 Howe, Kevin, Alex Bateman, and Richard Durbin. 2002. "QuickTree: Building Huge Neighbour-Joining Trees
564 of Protein Sequences." *Bioinformatics* 18 (11): 1546–47.

565 Hua-Van, Aurélie, Arnaud Le Rouzic, Thibaud S Boutin, Jonathan Filée, and Pierre Capy. 2011. “The Struggle
566 for Life of the Genome’s Selfish Architects.” *Biology Direct* 6 (1): 19–19. doi:10.1186/1745-6150-6-
567 19.

568 Huerta-Cepas, Jaime, Joaquín Dopazo, and Toni Gabaldón. 2010. “ETE: A Python Environment for Tree
569 Exploration.” *BMC Bioinformatics* 11 (1): 24. doi:10.1186/1471-2105-11-24.

570 Judelson, Howard S. 2002. “Sequence Variation and Genomic Amplification of a Family of Gypsy-like
571 Elements in the Oomycete Genus *Phytophthora*.” *Molecular Biology and Evolution* 19 (8): 1313–22.

572 Jurka, J., V.V. Kapitonov, A. Pavlicek, P. Klonowski, O. Kohany, and J. Walichiewicz. 2005. “Rebase Update,
573 a Database of Eukaryotic Repetitive Elements.” *Cytogenetic and Genome Research* 110 (1-4): 462–67.
574 doi:10.1159/000084979.

575 Katoh, Kazutaka, and Daron M. Standley. 2013. “MAFFT Multiple Sequence Alignment Software Version 7:
576 Improvements in Performance and Usability.” *Molecular Biology and Evolution* 30 (4): 772–80.
577 doi:10.1093/molbev/mst010.

578 Kazazian, Haig H. 2004. “Mobile Elements: Drivers of Genome Evolution.” *Science* 303 (5664): 1626–32.
579 doi:10.1126/science.1089670.

580 Kehrer-Sawatzki, Hildegard, and David N. Cooper. 2006. “Structural Divergence between the Human and
581 Chimpanzee Genomes.” *Human Genetics* 120 (6): 759–78. doi:10.1007/s00439-006-0270-6.

582 Kendall, David G. 1948. “On Some Modes of Population Growth Leading to RA Fisher’s Logarithmic Series
583 Distribution.” *Biometrika*, 6–15.

584 Kidwell, Mg, and Dr Lisch. 2000. “Transposable Elements and Host Genome Evolution.” *Trends in Ecology &
585 Evolution* 15 (3): 95–99.

586 Kües, Ursula, and Francis Martin. 2011. “On the Road to Understanding Truffles in the Underground.” *Fungal
587 Genetics and Biology : FG & B* 48 (6): 555–60. doi:10.1016/j.fgb.2011.02.002.

588 Kumar, Amar, and Jeffrey L. Bennetzen. 1999. “Plant Retrotransposons.” *Annual Review of Genetics* 33 (1):
589 479–532.

590 Kumekawa, N., N. Ohmido, K. Fukui, E. Ohtsubo, and H. Ohtsubo. 2001. “A New Gypsy-Type
591 Retrotransposon, RIRE7: Preferential Insertion into the Tandem Repeat Sequence TrsD in
592 Pericentromeric Heterochromatin Regions of Rice Chromosomes.” *Molecular Genetics and Genomics*
593 265 (3): 480–88.

594 Le Rouzic, Arnaud, Thibaut Payen, and Aurélie Hua-Van. 2012. “Reconstructing the Evolutionary History of
595 Transposable Elements.” *Genome Biology and Evolution* 5 (class I): 77–86. doi:10.1093/gbe/evs130.

596 Le Tacon, François, Benoît Marçais, Michel Courvoisier, Claude Murat, Pierre Montpied, and Michel Becker.
597 2014. “Climatic Variations Explain Annual Fluctuations in French Périgord Black Truffle Wholesale
598 Markets but Do Not Explain the Decrease in Black Truffle Production over the Last 48 Years.”
599 *Mycorrhiza* 24 (1): 115–25. doi:10.1007/s00572-014-0568-5.

600 Llorens, Carlos, Ricardo Futami, Laura Covelli, Laura Domínguez-Escribá, Jose M. Viu, Daniel Tamarit, Jose
601 Aguilar-Rodríguez, et al. 2011. “The Gypsy Database (GyDB) of Mobile Genetic Elements: Release
602 2.0.” *Nucleic Acids Research* 39 (suppl 1): D70–D74. doi:10.1093/nar/gkq1061.

- 603 Losada, Ana, José P. Abad, Marta Agudo, and Alfredo Villasante. 1999. "The Analysis of Circe, an LTR
604 Retrotransposon of *Drosophila Melanogaster*, Suggests That an Insertion of Non-LTR Retrotransposons
605 into LTR Elements Can Create Chimeric Retroelements." *Molecular Biology and Evolution* 16 (10):
606 1341–46.
- 607 Ma, Jianxin, and Jeffrey L. Bennetzen. 2004. "Rapid Recent Growth and Divergence of Rice Nuclear Genomes."
608 *Proceedings of the National Academy of Sciences of the United States of America* 101 (34): 12404–10.
609 doi:10.1073/pnas.0403715101.
- 610 Malik, Harmit S., and Thomas H. Eickbush. 1999. "Modular Evolution of the Integrase Domain in the
611 Ty3/Gypsy Class of LTR Retrotransposons." *Journal of Virology* 73 (6): 5186–90.
- 612 Marín, Ignacio, and Carlos Lloréns. 2000. "Ty3/Gypsy Retrotransposons: Description of New Arabidopsis
613 Thaliana Elements and Evolutionary Perspectives Derived from Comparative Genomic Data." *Molecular Biology and Evolution* 17 (7): 1040–49.
- 615 Martin, Francis, Annegret Kohler, Claude Murat, Raffaella Balestrini, Pedro M Coutinho, Olivier Jaillon,
616 Barbara Montanini, et al. 2010. "Périgord Black Truffle Genome Uncovers Evolutionary Origins and
617 Mechanisms of Symbiosis." *Nature* 464 (7291): 1033–38. doi:10.1038/nature08867.
- 618 McCarthy, E. M., and J. F. McDonald. 2003. "LTR_STRUC: A Novel Search and Identification Program for
619 LTR Retrotransposons." *Bioinformatics* 19 (3): 362–67. doi:10.1093/bioinformatics/btf878.
- 620 McGeoch, Duncan J. 1990. "Protein Sequence Comparisons Show That the 'pseudoproteases' Encoded by
621 Poxviruses and Certain Retroviruses Belong to the Deoxyuridine Triphosphatase Family." *Nucleic
622 Acids Research* 18 (14): 4105–10.
- 623 Miyao, Akio, Katsuyuki Tanaka, Kazumasa Murata, Hiromichi Sawaki, Shin Takeda, Kiyomi Abe, Yoriko
624 Shinozuka, Katsura Onosato, and Hirohiko Hirochika. 2003. "Target Site Specificity of the Tos17
625 Retrotransposon Shows a Preference for Insertion within Genes and against Insertion in
626 Retrotransposon-Rich Regions of the Genome." *The Plant Cell Online* 15 (8): 1771–80.
- 627 Montanini, Barbara, Pao-Yang Chen, Marco Morselli, Artur Jaroszewicz, David Lopez, Francis Martin, Simone
628 Ottonello, and Matteo Pellegrini. 2014. "The Black Truffle Methylome: Non-Exhaustive DNA
629 Methylation-Mediated Transposon Silencing in a Complex Fungal Genome with Massive Repeat
630 Element Content." *Genome Biology* 15 (7): 411. doi:10.1186/PREACCEPT-1025065114129803.
- 631 Murat, Claude. 2014. "Forty Years of Inoculating Seedlings with Truffle Fungi: Past and Future Perspectives."
632 *Mycorrhiza* 25 (1): 77–81. doi:10.1007/s00572-014-0593-4.
- 633 Murat, Claude, Thibaut Payen, Denis Petitpierre, and Jessy Labbé. 2013. "Repeated Elements in Filamentous
634 Fungi with a Focus on Wood-Decay Fungi." In *The Ecological Genomics of Fungi*, edited by Francis
635 Martin, 21–40. John Wiley & Sons, Inc.
636 <http://onlinelibrary.wiley.com/doi/10.1002/9781118735893.ch2/summary>.
- 637 Murat, Claude, Andrea Rubini, Claudia Riccioni, Herminia De la Varga, Emila Akroume, Beatrice Belfiori,
638 Marco Guaragno, et al. 2013. "Fine-Scale Spatial Genetic Structure of the Black Truffle (*Tuber
639 Melanosporum*) Investigated with Neutral Microsatellites and Functional Mating Type Genes." *New
640 Phytologist* 199 (1): 176–87. doi:10.1111/nph.12264.

641 Murat, Claude, Elisa Zampieri, Marta Vallino, Stefania Daghino, Silvia Perotto, and Paola Bonfante. 2011.
642 “Genomic Suppression Subtractive Hybridization as a Tool to Identify Differences in Mycorrhizal
643 Fungal Genomes.” *FEMS Microbiology Letters* 318 (2): 115–22. doi:10.1111/j.1574-
644 6968.2011.02248.x.

645 Muszewska, Anna, Marta Hoffman-Sommer, and Marcin Grynberg. 2011. “LTR Retrotransposons in Fungi.”
646 *PLoS One* 6 (12): e29425.

647 Nee, Sean, Robert M. May, and Paul H. Harvey. 1994. “The Reconstructed Evolutionary Process.”
648 *Philosophical Transactions of the Royal Society B: Biological Sciences* 344 (1309): 305–11.

649 Neuvéglise, Cécile, Horst Feldmann, Elisabeth Bon, Claude Gaillardin, and and Serge Casaregola. 2002.
650 “Genomic Evolution of the Long Terminal Repeat Retrotransposons in Hemiascomycetous Yeasts.”
651 *Genome Research* 12 (6): 930–43. doi:10.1101/gr.219202.

652 Novikova, O. S., and A. G. Blinov. 2008. “dUTPase-Containing Metaviridae LTR Retrotransposons from the
653 Genome of Phanerochaete Chrysosporium (Fungi: Basidiomycota).” In *Doklady Biochemistry and*
654 *Biophysics*, 420:146–49. Springer. <http://www.springerlink.com/index/78U10J8102728364.pdf>.

655 Novikova, Olga, Georgiy Smyshlyaev, and Alexander Blinov. 2010. “Evolutionary Genomics Revealed
656 Interkingdom Distribution of Tcn1-like Chromodomain-Containing Gypsy LTR Retrotransposons
657 among Fungi and Plants.” *BMC Genomics* 11 (January): 231–231. doi:10.1186/1471-2164-11-231.

658 Paradis, Emmanuel, Julien Claude, and Korbinian Strimmer. 2004. “APE: Analyses of Phylogenetics and
659 Evolution in R Language.” *Bioinformatics* 20 (2): 289–90. doi:10.1093/bioinformatics/btg412.

660 Payen, Thibaut, Claude Murat, Anaïs Gigant, Emmanuelle Morin, Stéphane De Mita, and Francis Martin. 2015.
661 “A Survey of Genome-Wide Single Nucleotide Polymorphisms through Genome Re-Sequencing in the
662 Périgord Black Truffle (*Tuber Melanosporum* Vittad.).” *Molecular Ecology Resources*, February, n/a–
663 n/a. doi:10.1111/1755-0998.12391.

664 Perdue, Sarah, and Sergey V. Nuzhdin. 2000. “Master Copy Is Not Responsible for the High Rate of Copia
665 Transposition in *Drosophila*.” *Molecular Biology and Evolution* 17 (6): 984–86.

666 Price, Alkes L, Neil C Jones, and Pavel A Pevzner. 2005. “De Novo Identification of Repeat Families in Large
667 Genomes.” *Bioinformatics* 21 Suppl 1 (June): i351–8. doi:10.1093/bioinformatics/bti1018.

668 Quinlan, Aaron R., and Ira M. Hall. 2010. “BEDTools: A Flexible Suite of Utilities for Comparing Genomic
669 Features.” *Bioinformatics* 26 (6): 841–42. doi:10.1093/bioinformatics/btq033.

670 Rabosky, Daniel L. 2007. “LASER: A Maximum Likelihood Toolkit for Detecting Temporal Shifts in
671 Diversification Rates From Molecular Phylogenies.” *Evolutionary Bioinformatics Online* 2 (February):
672 247–50.

673 Raffaele, Sylvain, and Sophien Kamoun. 2012. “Genome Evolution in Filamentous Plant Pathogens: Why
674 Bigger Can Be Better.” *Nature Reviews. Microbiology* 10 (6): 417–30. doi:10.1038/nrmicro2790.

675 Revell, Liam J. 2012. “Phytools: An R Package for Phylogenetic Comparative Biology (and Other Things).”
676 *Methods in Ecology and Evolution* 3 (2): 217–23. doi:10.1111/j.2041-210X.2011.00169.x.

677 Riccioni, Claudia, Andrea Rubini, Beatrice Belfiori, Valentina Passeri, Francesco Paolocci, and Sergio Arcioni.
678 2008. “Tmt1: The First LTR-Retrotransposon from a Tuber Spp.” *Current Genetics* 53 (1): 23–34.
679 doi:10.1007/s00294-007-0155-9.

- 680 Robinson, J.T., H. Thorvaldsdóttir, W. Winckler, M. Guttman, E.S. Lander, G. Getz, and J.P. Mesirov. 2011.
681 “Integrative Genomics Viewer.” *Nature Biotechnology* 29 (1): 24–26. doi:10.1038/nbt0111-24.
- 682 Schliep, Klaus Peter. 2011. “Phangorn: Phylogenetic Analysis in R.” *Bioinformatics* 27 (4): 592–93.
683 doi:10.1093/bioinformatics/btq706.
- 684 Sievers, Fabian, Andreas Wilm, David Dineen, Toby J Gibson, Kevin Karplus, Weizhong Li, Rodrigo Lopez, et
685 al. 2011. “Fast, Scalable Generation of High-Quality Protein Multiple Sequence Alignments Using
686 Clustal Omega.” *Molecular Systems Biology* 7 (539). doi:10.1038/msb.2011.75.
- 687 Smit, Arian FA, Robert Hubley, and Phil Green. 1996. *RepeatMasker Open-3.0*.
- 688 Strimmer, Korbinian, and Oliver G. Pybus. 2001. “Exploring the Demographic History of DNA Sequences
689 Using the Generalized Skyline Plot.” *Molecular Biology and Evolution* 18 (12): 2298–2305.
- 690 Struchiner, Claudio J., Eduardo Massad, Zhijian Tu, and José M. C. Ribeiro. 2009. “The Tempo and Mode of
691 Evolution of Transposable Elements as Revealed by Molecular Phylogenies Reconstructed from
692 Mosquito Genomes.” *Evolution* 63 (12): 3136–46. doi:10.1111/j.1558-5646.2009.00788.x.
- 693 Suoniemi, Annu, Jaakko Tanskanen, and Alan H. Schulman. 1998. “Gypsy-like Retrotransposons Are
694 Widespread in the Plant Kingdom.” *The Plant Journal* 13 (5): 699–705. doi:10.1046/j.1365-
695 313X.1998.00071.x.
- 696 Tavares, Sílvia, Ana Paula Ramos, Ana Sofia Pires, Helena G. Azinheira, Patrícia Caldeirinha, Tobias Link, Rita
697 Abranches, et al. 2014. “Genome Size Analyses of Pucciniales Reveal the Largest Fungal Genomes.”
698 *Frontiers in Plant Science* 5 (August). doi:10.3389/fpls.2014.00422.
- 699 Tisserant, E., C. Da Silva, A. Kohler, E. Morin, P. Wincker, and F. Martin. 2011. “Deep RNA Sequencing
700 Improved the Structural Annotation of the Tuber *Melanosporium* Transcriptome.” *New Phytologist* 189
701 (3): 883–91. doi:10.1111/j.1469-8137.2010.03597.x.
- 702 Turelli, Priscilla, Francois Guiguen, Jean-Francois Mornex, Robert Vigne, and Gilles Quérat. 1997. “dUTPase-
703 Minus Caprine Arthritis-Encephalitis Virus Is Attenuated for Pathogenesis and Accumulates G-to-A
704 Substitutions.” *Journal of Virology* 71 (6): 4522–30.
- 705 UniProt Consortium. 2008. “The Universal Protein Resource (UniProt).” *Nucleic Acids Research* 36 (suppl 1):
706 D190–D195. doi:10.1093/nar/gkm895.
- 707 Vezzi, Francesco, Cristian Del Fabbro, Alexandru I. Tomescu, and Alberto Policriti. 2012. “rNA: A Fast and
708 Accurate Short Reads Numerical Aligner.” *Bioinformatics* 28 (1): 123–24.
709 doi:10.1093/bioinformatics/btr617.
- 710 Wicker, Thomas, François Sabot, Aurélie Hua-Van, Jeffrey L Bennetzen, Pierre Capy, Boulos Chalhoub,
711 Andrew Flavell, et al. 2007. “A Unified Classification System for Eukaryotic Transposable Elements.”
712 *Nature Reviews. Genetics* 8 (12): 973–82. doi:10.1038/nrg2165.
- 713 Yule, G. Udny. 1925. “A Mathematical Theory of Evolution, Based on the Conclusions of Dr. JC Willis, FRS.”
714 *Philosophical Transactions of the Royal Society of London. Series B, Containing Papers of a Biological*
715 *Character*, 21–87.

716

717 **Author’s contribution**

718 FM and CM designed the project. TP, have done the bioinformatic analyses. CM, TP and FM
719 wrote the paper.

720

721 **Figure legend**

722 Fig.1 Neighbour-joining phylogeny realized with reverse transcriptase isolated in *T.*
723 *melanosporum* (Tmel) and coming from gypsy databases. The bootstrap value above 40% are
724 shown.

725 Fig. 2 Structural organization of gypsy retrotransposons clades found in *T. melanosporum*.
726 The different motifs and genes were identified with pfam search (<http://pfam.xfam.org/search>)
727 Abbreviation: LTR: long terminal repeat; PBS: putative primer-binding site; Zf-CCHC: zinc
728 finger; Exo-endo-phos: exonuclease-endonuclease-phosphatase; PRO: proteinase; RT: reverse
729 transcriptase; RH: ribonuclease H; INT: integrase; CHR: chromodomain; PPT: polypurine
730 tract.

731

732 Fig. 3 Maximum likelihood reconstructed phylogenies and lineage through time for three *Tmt*
733 clades. A. Ultrametric trees were obtained through the “pathd8” algorithm (see Materials and
734 Methods). B. Continue line indicated lineage through time based on ultrametric trees and
735 dashed line illustrates the expectation for a “pure birth” model (constant transposition, no
736 deletions).

737

738 Fig. 4 Time estimation of full-lengths elements for *Tmt1*, *Tmt3* and *Tmt6* clades. Kimura-2-
739 parameter genetic distance was calculated using solo LTRs of each elements according to
740 Martin et al. (2010).

741

742 Fig. 5 Distribution of the inter-element distances in genomic window sizes. For each elements
743 the coefficient of aggregation A (Hopkins, 1954) is indicated as well as the parameter $x =$
744 $A/(1+A)$ that tends to normality as n (i.e. the number of pairs of observation) increases. For
745 random distribution the coefficient A is equal to one, greater than one for aggregation
746 distribution, and less than one for regular distribution. For all the *Tmts* the random
747 distribution was excluded with p-value < 0.0001.

748

749 Fig. 6 Boxplots representing the inter-element distances for the five *Tmts*.

Fig. 1

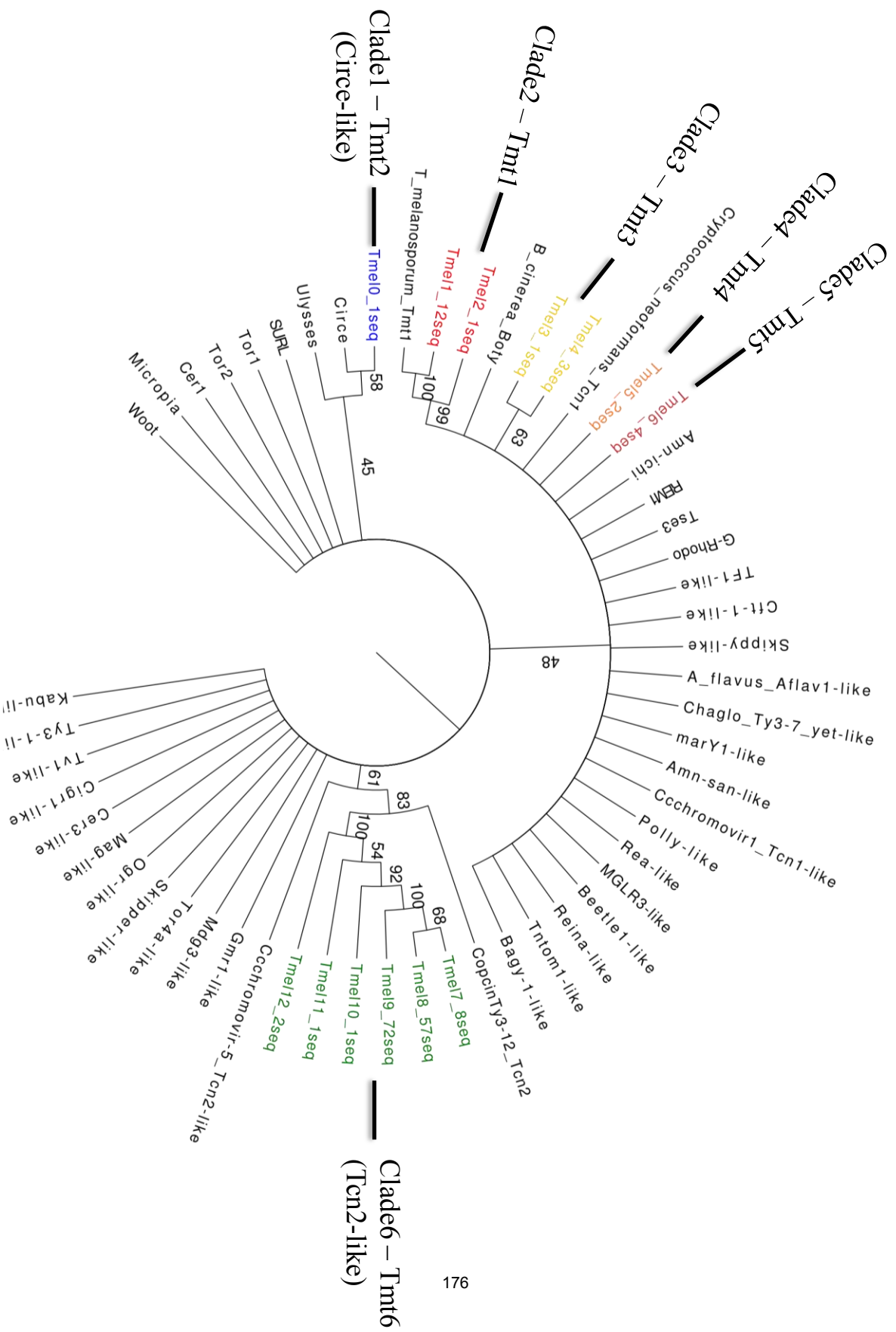


Fig. 2

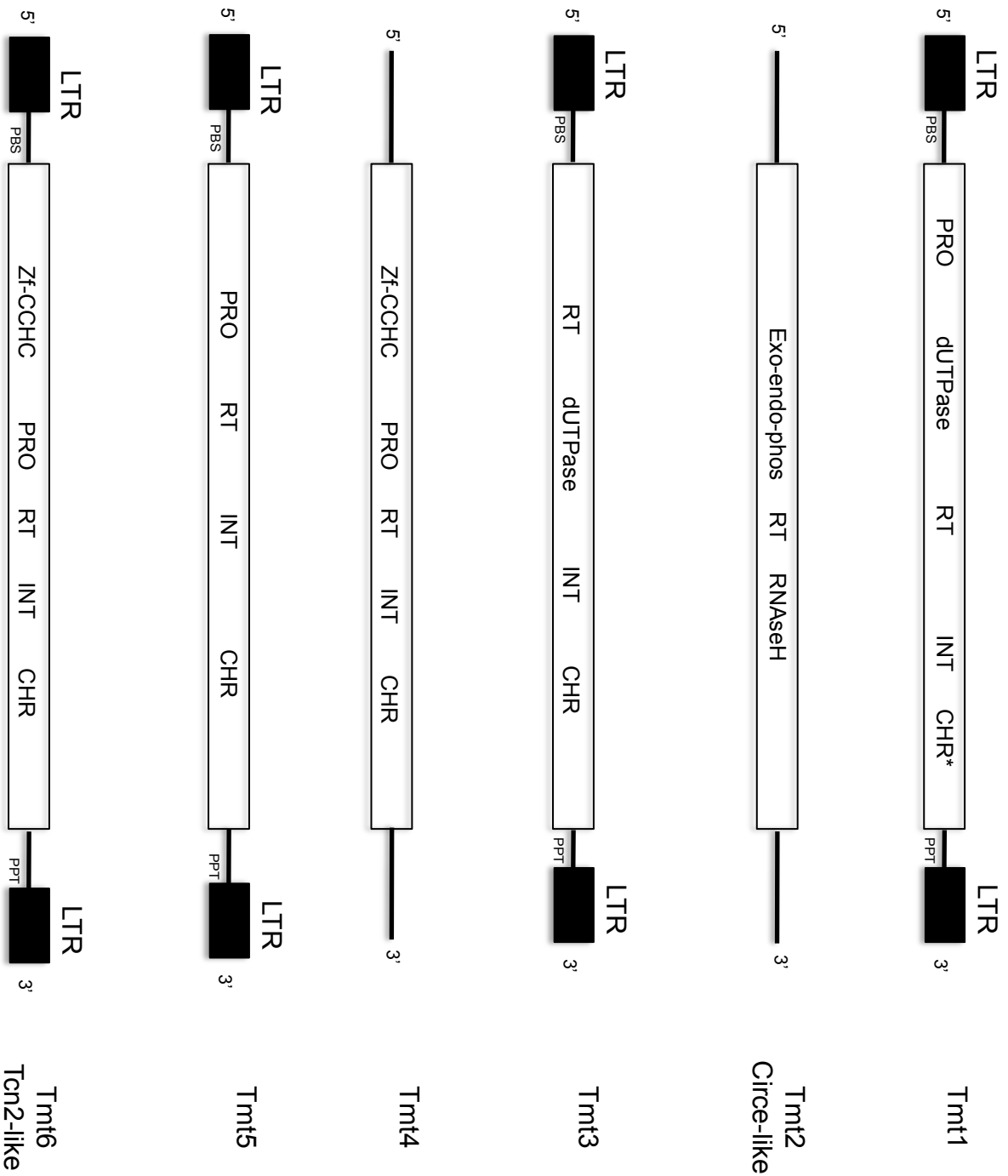
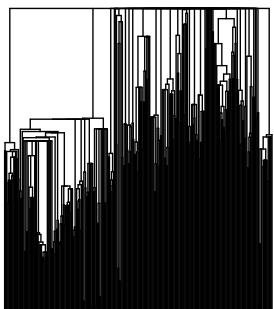
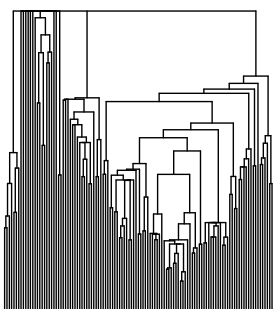


Fig. 3

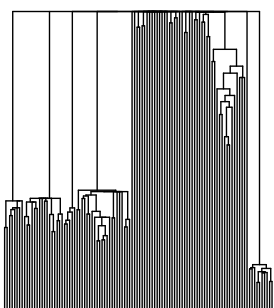
A



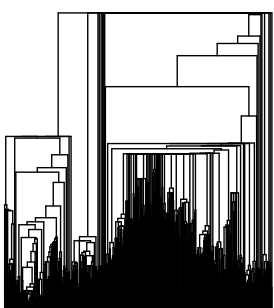
Tm1



Tm2

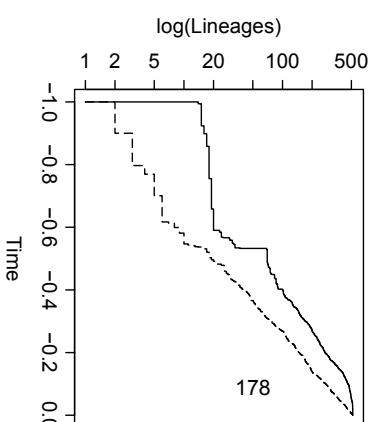
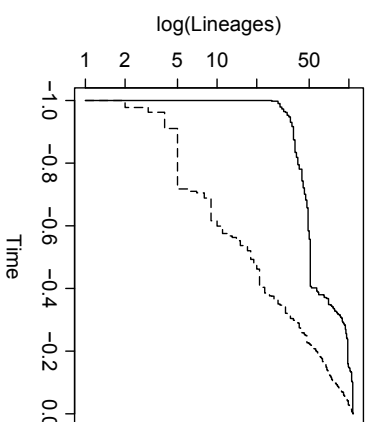
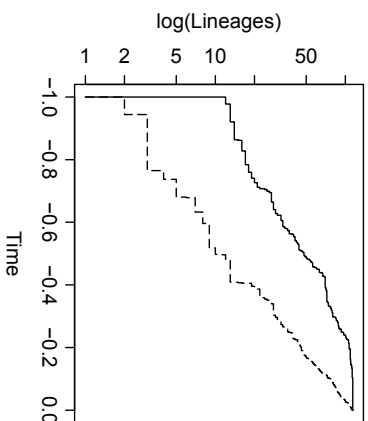
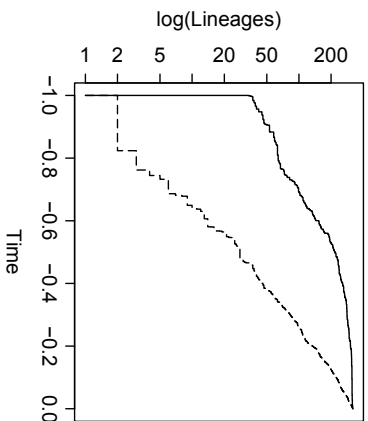


Tm3



Tm6

B



C

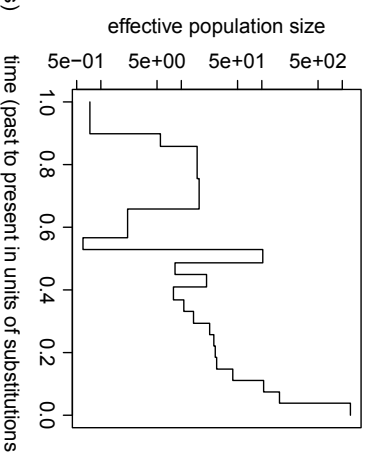
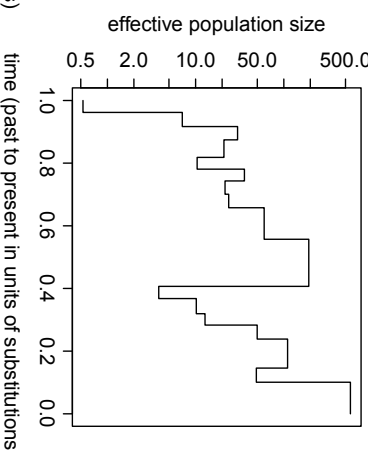
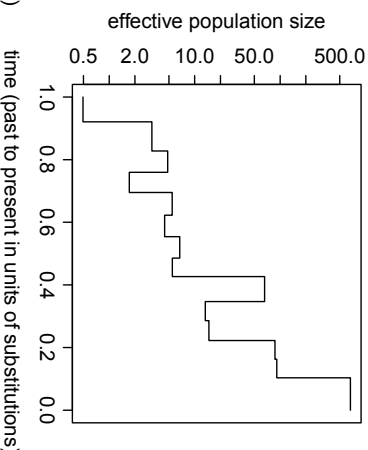
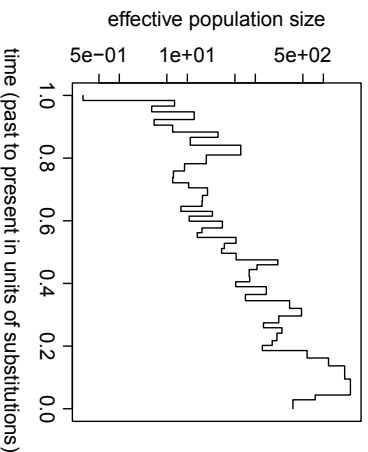


Fig. 4

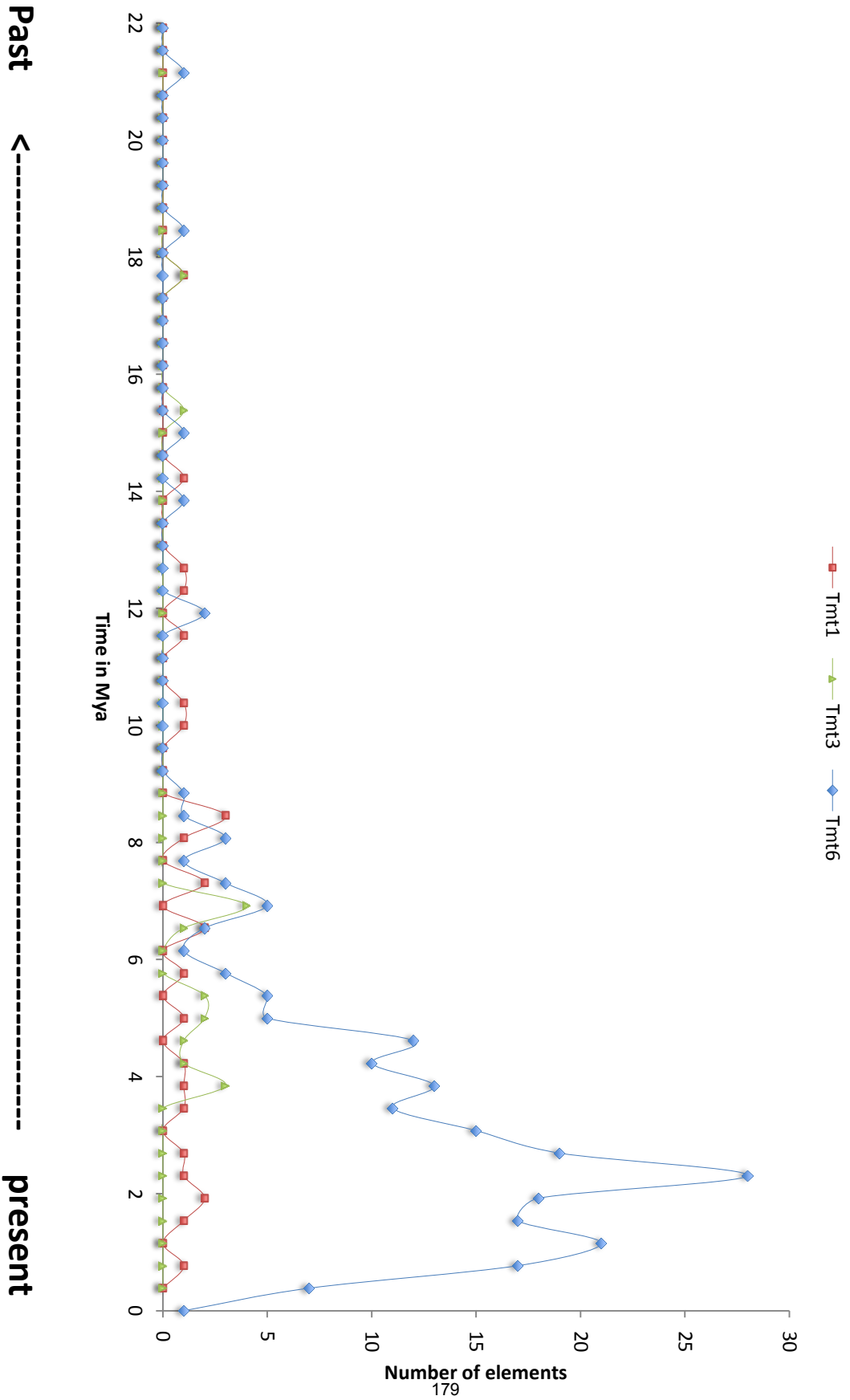
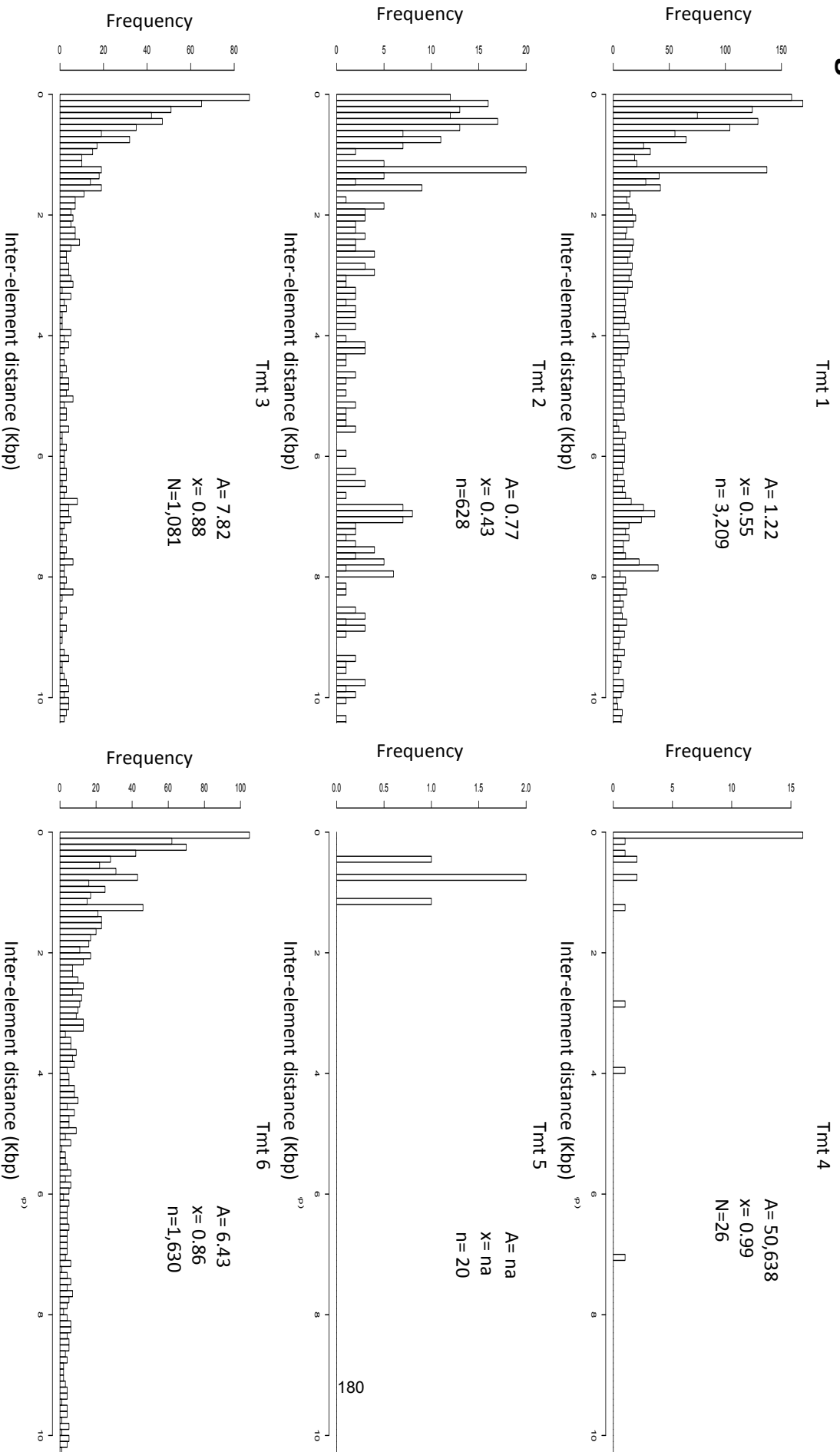
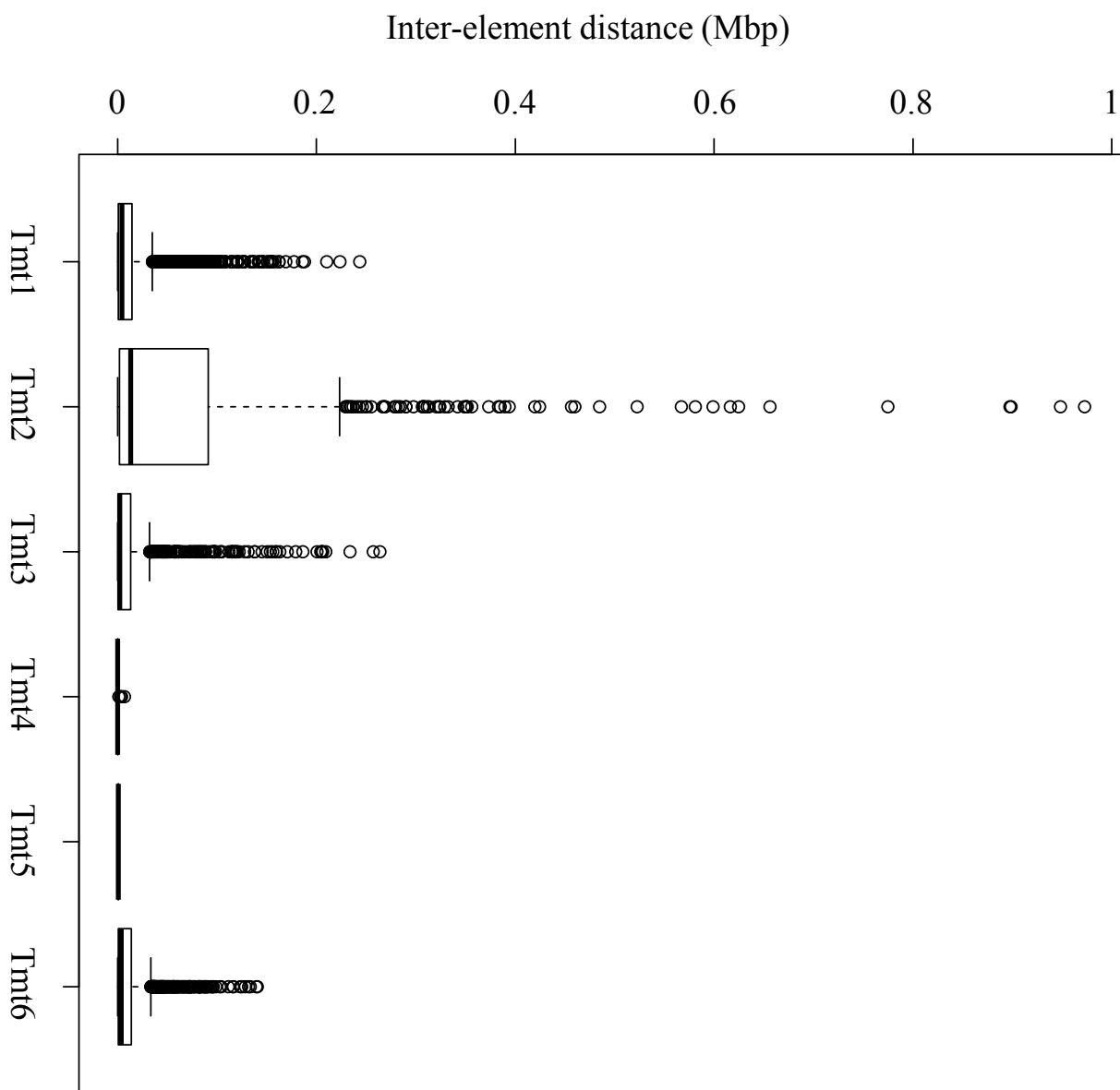


Fig. 5





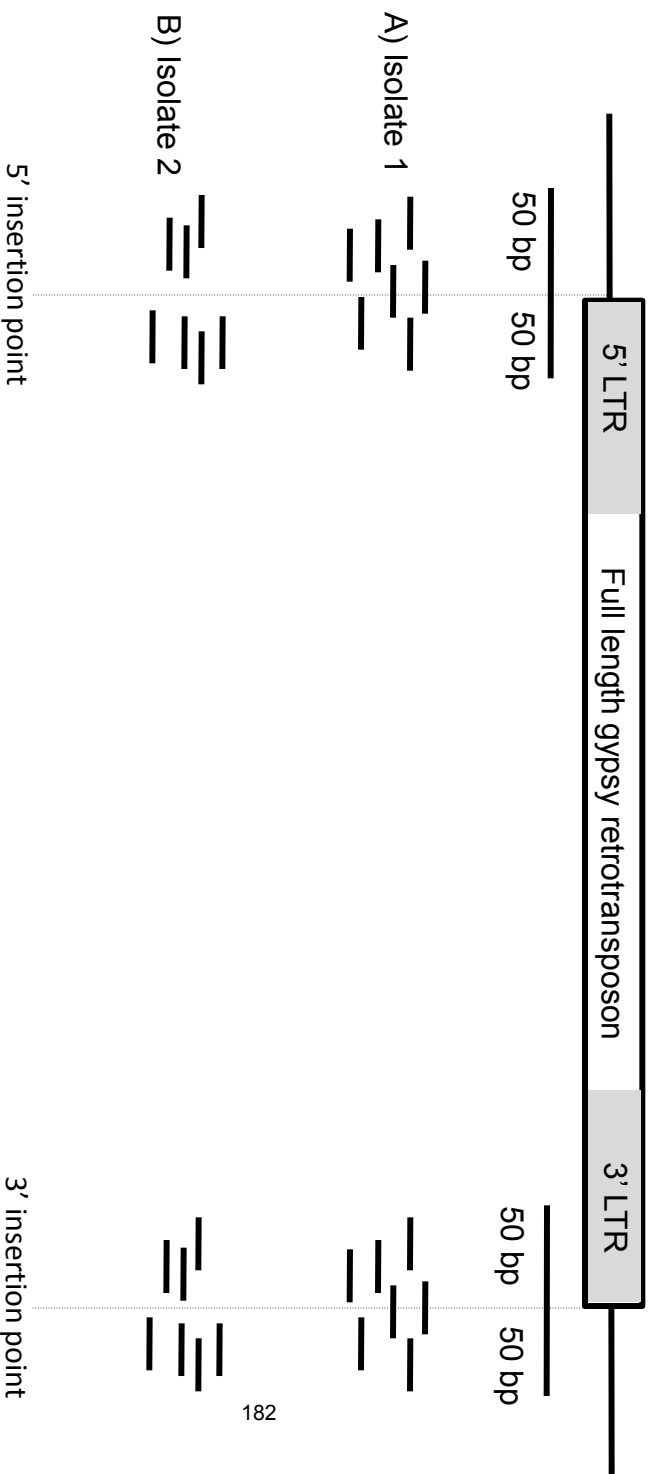
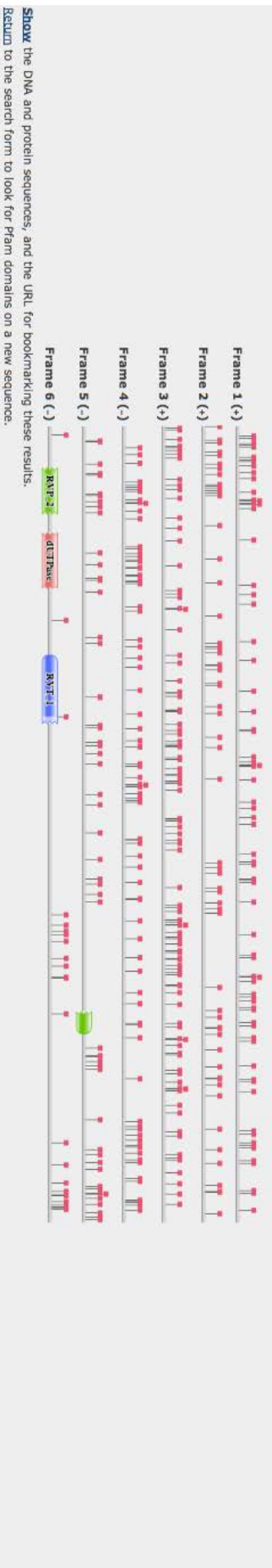


Fig. S1. Strategy used to investigate the insertion point of full length retrotransposons gypsy. For each elements 50 bp of the LTR (long terminal repeat) and 50 bp of the genomic sequence outside the LTR was selected for both 5' and 3'. Reads for the six resequenced genomes were mapped against the 100 bp obtained sequences. Two examples are presented: A) for isolate 1 the full length element has the same insertion point although B) for isolate 2 the full length element has not the same insertion point since reads overlapping insertion point are lacking.



Show the DNA and protein sequences, and the URL for bookmarking these results.
 Return to the search form to look for Pfam domains on a new sequence.

Significant Pfam-A Matches

Show or hide all alignments. Toggle between amino-acid and DNA sequence coordinates.

Frame (Gene)	Family	Description	Entry type	Clan	Envelope		Alignment		HMM	HMM	Bit	E-value	Predicted active sites	Show/hide alignment
					Start	End	Start	End	From	To	score			
5 (-)	Chromo	Chromo (Chromatin Organisation Modifier [™])	Domain	CL0049	1373	1424	1376	1424	4	55	21.7	0.00011	n/a	Show
6 (-)	RVP_2	Retroviral aspartyl protease	Domain	CL0129	98	205	105	203	36	131	27.1	3.1e-06	n/a	Show
6 (-)	dUTPase	dUTPase	Domain	CL0153	250	379	251	378	2	128	103.0	7.4e-30	n/a	Show
6 (-)	RNT1_1	Reverse transcriptase (RNA-dependent DNA [™])	Family	CL0022	536	696	536	681	1	197	60.5	1.4e-16	n/a	Show

Fig. S2. PFAM motifs of one copy of Tmt1 (scaffold 67 in positions 266819 to 271029) that present a chromodomain

Table S1: Name of the full-length and repeatscout gypsy retrotransposons with their clade attribution and if a RT sequence has been identified inside

Name full-length elements	Clade	Reverse transcriptase	dUTPase	Internal proteins				Number of illumina reads		
				Integrase	Chromodomaine	AP	RNAseH	ECM	FB	FLM
10000_PT_B24_L22_265	Tmt1	No	Yes	Yes	No	Yes	Yes	0	2	2
10119_RT32_B24_L22_206	Tmt1	No	Yes	Yes	No	Yes	Yes	0	0	1
11679_RT32_B24_L24_201	Tmt1	Yes	Yes	Yes	No	Yes	Yes	0	2	1
11679_RT6_B24_L24_134	Tmt1	Yes	Yes	Yes	No	Yes	Yes	0	0	0
12985_RT32_B24_L24_174	Tmt1	Yes	Yes	Yes	No	Yes	Yes	0	1	0
15000_RT32_B24_L24_79	Tmt1	No	Yes	Yes	No	Yes	Yes	0	0	0
16199_RT32_B24_L26_16	Tmt1	Yes	Yes	Yes	No	Yes	Yes	0	1	0
16559_RT6_B24_L24_197	Tmt1	Yes	Yes	Yes	No	Yes	Yes	0	0	0
18799_RT32_B21_L22_116	Tmt1	Yes	Yes	Yes	No	Yes	Yes	0	5	0
4130_B24_L28_189	Tmt1	No	No	No	No	Yes	No	0	3	0
4788_B24_L28_32	Tmt1	No	No	No	No	No	No	0	0	0
5132_PT_B24_L22_105	Tmt1	Yes	Yes	Yes	No	Yes	Yes	0	0	0
5850_RT32_272	Tmt1	Yes	Yes	Yes	No	Yes	Yes	0	2	5
6984_PT_B24_L24_28	Tmt1	No	Yes	Yes	No	Yes	Yes	0	0	1
6987_RT32_B24_L22_52	Tmt1	Yes	Yes	Yes	No	Yes	Yes	0	0	0
7128_RT43_B21_L22_278	Tmt1	No	Yes	Yes	No	No	Yes	0	3	5
7299_RT33_B24_L28_41	Tmt1	No	No	Yes	No	No	Yes	2	0	8
7479_PT_B24_L22_126	Tmt1	No	Yes	Yes	No	Yes	Yes	0	0	1
7944_RT32_B24_L22_122	Tmt1	No	Yes	Yes	No	Yes	Yes	0	13	6
8656_RT42_B24_L24_236	Tmt1	No	No	No	No	No	No	203	311	199
8658_PT_B24_L24_50	Tmt1	No	Yes	Yes	No	Yes	Yes	0	4	3
8755_RT32_B24_L22_115	Tmt1	No	Yes	Yes	No	Yes	Yes	0	0	0
9052_RT32_B24_L24_251	Tmt1	No	Yes	Yes	No	Yes	Yes	0	0	0
9722_RT5_B21_L22_161	Tmt1	No	Yes	Yes	No	Yes	Yes	0	0	1
9954_RT32_B24_L26_21	Tmt1	No	Yes	Yes	No	Yes	Yes	1	3	1
11040_RT6_B21_L24_26	Tmt3	No	Yes	Yes	No	No	Yes	0	0	1
12599_RT38_B21_L5_86	Tmt3	No	Yes	Yes	No	No	Yes	0	3	2
12599_RT39_B21_L24_241	Tmt3	No	Yes	Yes	No	No	Yes	0	0	1
12599_RT4_B19_L5_1	Tmt3	No	Yes	Yes	No	No	Yes	0	4	0
14199_RT6_B24_L22_245	Tmt3	No	Yes	Yes	No	Yes	Yes	0	0	2
16469_RT39_B21_L5_127	Tmt3	No	Yes	Yes	No	No	Yes	0	0	0
3291_PT_B21_L5_171	Tmt3	No	Yes	Yes	No	No	Yes	3	5	2
4328_RT34_B21_L22_59	Tmt3	No	No	Yes	No	No	Yes	0	0	0
6929_RT6_B21_L24_39	Tmt3	No	Yes	Yes	No	No	Yes	0	2	1
7656_RT6_B21_L5_196	Tmt3	No	Yes	Yes	No	No	Yes	0	3	2
8148_RT5_B20_L22_2	Tmt3	Yes	Yes	Yes	Yes	No	Yes	0	0	0
8405_RT6_B22_L24_221	Tmt3	No	Yes	Yes	No	Yes	Yes	0	0	0
8802_RT6_B24_L22_31	Tmt3	No	Yes	Yes	No	No	Yes	0	0	1
8955_RT6_B24_L22_136	Tmt3	No	Yes	Yes	No	No	Yes	0	9	6
9149_RT34_B21_L24_60	Tmt3	No	No	Yes	No	No	Yes	3	5	2
9306_RT36_B24_L22_74	Tmt3	No	No	Yes	No	No	Yes	0	2	0
14649_RT6_B21_L24_121	Tmt5	Yes	No	Yes	Yes	Yes	Yes	0	2	1
14649_RT6_B26_L28_23	Tmt5	Yes	No	Yes	Yes	Yes	Yes	0	2	1
16649_RT6_B24_L22_194	Tmt5	No	No	Yes	Yes	Yes	Yes	0	1	0
16649_RT6_B24_L22_55	Tmt5	Yes	No	Yes	No	Yes	Yes	0	1	0
10000_PT_B24_L22_114	Tmt6	Yes	No	Yes	Yes	No	Yes	0	0	0
10000_PT_B24_L22_120	Tmt6	No	No	Yes	No	No	Yes	0	1	0
10000_PT_B24_L22_124	Tmt6	Yes	No	Yes	Yes	No	Yes	0	0	1
10000_PT_B24_L22_140	Tmt6	Yes	No	Yes	No	No	Yes	0	1	0
10000_PT_B24_L22_167	Tmt6	Yes	No	Yes	No	No	Yes	0	0	0
10000_PT_B24_L22_198	Tmt6	Yes	No	Yes	No	No	Yes	0	1	1
10000_PT_B24_L22_246	Tmt6	No	No	Yes	No	No	Yes	0	1	0
10000_PT_B24_L22_288	Tmt6	No	No	Yes	No	No	Yes	0	0	0
10000_PT_B24_L22_293	Tmt6	Yes	No	Yes	Yes	No	Yes	0	0	0
10000_PT_B24_L22_294	Tmt6	Yes	No	Yes	Yes	No	Yes	0	0	0
10000_PT_B24_L22_300	Tmt6	Yes	No	Yes	No	No	Yes	0	0	0
10119_RT32_B24_L22_183	Tmt6	Yes	No	Yes	No	No	Yes	0	0	0
10187_RT41_B24_L22_186	Tmt6	Yes	Yes	Yes	Yes	Yes	Yes	0	3	1
10223_RT32_B24_L28_107	Tmt6	No	No	Yes	No	No	Yes	2	0	0
10361_RT5_B24_L22_242	Tmt6	No	No	Yes	Yes	No	Yes	0	0	0
10544_RT32_B24_L22_230	Tmt6	Yes	No	Yes	Yes	No	Yes	0	0	0
10995_RT32_B24_L24_220	Tmt6	Yes	Yes	Yes	No	Yes	Yes	0	0	0
11119_RT32_B24_L22_91	Tmt6	Yes	No	Yes	Yes	No	Yes	0	0	0
11339_RT32_B24_L22_254	Tmt6	Yes	Yes	Yes	No	No	Yes	0	1	2
11513_RT32_B24_L22_123	Tmt6	No	No	Yes	No	No	Yes	0	2	0
11513_RT32_B24_L22_187	Tmt6	Yes	No	Yes	Yes	No	Yes	0	0	0
11859_RT32_B24_L24_249	Tmt6	No	Yes	Yes	No	Yes	Yes	0	1	1
11985_RT32_B24_L22_292	Tmt6	Yes	No	Yes	No	No	Yes	0	0	0
12120_RT32_B21_L22_11	Tmt6	Yes	No	Yes	Yes	No	Yes	0	1	0
12120_RT32_B21_L22_12	Tmt6	Yes	No	Yes	No	No	Yes	0	0	1
12120_RT32_B21_L22_7	Tmt6	Yes	No	Yes	No	No	Yes	0	0	0
12120_RT32_B24_L22_118	Tmt6	Yes	No	Yes	Yes	No	Yes	0	0	2
12120_RT32_B24_L22_150	Tmt6	Yes	No	Yes	Yes	No	Yes	0	1	0
12120_RT32_B24_L22_151	Tmt6	Yes	No	Yes	No	No	Yes	0	3	1
12120_RT32_B24_L22_162	Tmt6	Yes	No	Yes	No	No	Yes	0	1	0
12120_RT32_B24_L22_170	Tmt6	No	No	Yes	No	No	Yes	0	0	0
12120_RT32_B24_L22_172	Tmt6	Yes	No	Yes	No	No	Yes	0	0	0
12120_RT32_B24_L22_177	Tmt6	Yes	No	Yes	Yes	No	Yes	0	1	1
12120_RT32_B24_L22_199	Tmt6	Yes	No	Yes	No	No	Yes	0	0	0
12120_RT32_B24_L22_202	Tmt6	Yes	No	Yes	No	No	Yes	0	0	1
12120_RT32_B24_L22_211	Tmt6	Yes	No	Yes	No	No	Yes	0	0	0
12120_RT32_B24_L22_212	Tmt6	Yes	No	Yes	No	No	Yes	0	0	0
12120_RT32_B24_L22_218	Tmt6	No	No	Yes	Yes	No	Yes	0	0	0
12120_RT32_B24_L22_219	Tmt6	No	No	Yes	No	No	Yes	0	0	0
12120_RT32_B24_L22_227	Tmt6	Yes	No	Yes	Yes	No	Yes	0	0	0
12120_RT32_B24_L22_255	Tmt6	Yes	No	Yes	No	No	Yes	0	0	0
12120_RT32_B24_L22_282	Tmt6	Yes	No	Yes	No	No	Yes	0	0	1
12120_RT32_B24_L22_291	Tmt6	No	No	Yes	No	No	Yes	0	0	0
12120_RT32_B24_L22_30	Tmt6	Yes	No	Yes	Yes	No	Yes	0	0	0
12120_RT32_B24_L22_53	Tmt6	No	No	Yes	No	No	Yes	0	1	0
12120_RT32_B24_L22_71	Tmt6	No	No	Yes	No	No	Yes	0	0	1
12120_RT32_B24_L22_76	Tmt6	Yes	No	Yes	No	No	Yes	0	0	0
12120_RT32_B24_L22_94	Tmt6	Yes	No	Yes	Yes	No	Yes	0	0	0
12120_RT5_B24_L22_240	Tmt6	No	No	Yes	No	No	Yes	0	0	0
12187_RT32_B24_L22_207	Tmt6	No	No	Yes	Yes	No	Yes	0	0	1
12224_RT32_B24_L22_289	Tmt6	No	No	Yes	No	No	Yes	0	0	0
12750_RT32_B24_L22_33	Tmt6	No	No	Yes	Yes	No	Yes	0	0	0

12750_RT32_B24_L22_37	Tmt6	No	No	Yes	No	No	Yes	0	0	0
13485_RT32_B24_L22_164	Tmt6	Yes	No	Yes	No	No	Yes	0	0	0
13679_RT6_B21_L24_5	Tmt6	No	No	Yes	No	No	Yes	0	0	0
14983_RT32_B24_L22_250	Tmt6	No	No	Yes	Yes	No	Yes	0	3	1
14999_RT32_B24_L22_48	Tmt6	Yes	No	Yes	No	No	Yes	0	0	0
15389_RT32_B24_L22_259	Tmt6	No	No	Yes	No	No	Yes	0	3	0
15432_RT28_B24_L22_225	Tmt6	Yes	No	Yes	Yes	No	Yes	0	0	0
15799_RT32_B24_L22_147	Tmt6	Yes	No	Yes	Yes	No	Yes	0	2	1
15999_RT32_B24_L22_100	Tmt6	No	No	Yes	No	No	Yes	0	0	0
16199_RT32_B21_L22_15	Tmt6	No	No	Yes	No	No	Yes	0	0	0
16199_RT32_B24_L22_104	Tmt6	Yes	No	Yes	Yes	No	Yes	0	0	0
16199_RT32_B24_L22_301	Tmt6	Yes	No	Yes	No	No	Yes	0	0	0
16199_RT32_B24_L22_92	Tmt6	No	No	Yes	No	No	Yes	0	1	0
16399_RT32_B24_L22_58	Tmt6	Yes	No	Yes	No	No	Yes	0	1	0
16999_RT32_B24_L22_78	Tmt6	Yes	No	Yes	No	No	Yes	0	0	0
17000_RT32_B24_L22_84	Tmt6	Yes	No	Yes	No	No	Yes	0	0	0
17099_RT32_B24_L22_156	Tmt6	No	No	Yes	No	No	Yes	0	0	0
17099_RT32_B24_L22_180	Tmt6	No	No	Yes	No	No	Yes	0	2	1
17099_RT32_B24_L22_243	Tmt6	Yes	No	Yes	No	No	Yes	0	0	0
17099_RT32_B24_L22_262	Tmt6	No	No	Yes	Yes	No	Yes	0	0	0
17099_RT32_B24_L22_68	Tmt6	Yes	No	Yes	Yes	No	Yes	0	0	0
17194_RT32_B24_L22_66	Tmt6	Yes	No	Yes	Yes	No	Yes	0	0	0
17999_RT28_B24_L22_281	Tmt6	Yes	No	Yes	No	No	Yes	0	0	0
17999_RT32_B21_L22_13	Tmt6	Yes	No	Yes	No	No	Yes	0	0	0
17999_RT32_B21_L22_17	Tmt6	No	No	Yes	Yes	No	Yes	0	1	0
17999_RT32_B21_L22_19	Tmt6	Yes	No	Yes	No	No	Yes	0	0	0
17999_RT32_B21_L22_20	Tmt6	Yes	No	Yes	No	No	Yes	0	0	0
17999_RT32_B21_L22_8	Tmt6	Yes	No	Yes	No	No	Yes	0	0	0
17999_RT32_B24_L22_101	Tmt6	No	No	Yes	Yes	No	Yes	0	0	0
17999_RT32_B24_L22_110	Tmt6	Yes	No	Yes	Yes	No	Yes	0	0	0
17999_RT32_B24_L22_111	Tmt6	Yes	No	Yes	No	No	Yes	0	1	0
17999_RT32_B24_L22_128	Tmt6	No	No	Yes	No	No	Yes	0	1	0
17999_RT32_B24_L22_129	Tmt6	Yes	No	Yes	No	No	Yes	0	0	0
17999_RT32_B24_L22_130	Tmt6	No	No	Yes	No	No	Yes	0	0	0
17999_RT32_B24_L22_132	Tmt6	Yes	No	Yes	Yes	No	Yes	0	0	0
17999_RT32_B24_L22_135	Tmt6	Yes	No	Yes	Yes	No	Yes	0	0	0
17999_RT32_B24_L22_137	Tmt6	No	No	Yes	No	No	Yes	0	3	1
17999_RT32_B24_L22_141	Tmt6	Yes	No	Yes	No	No	Yes	0	0	1
17999_RT32_B24_L22_144	Tmt6	Yes	No	Yes	No	No	Yes	0	1	0
17999_RT32_B24_L22_146	Tmt6	Yes	No	Yes	No	No	Yes	0	3	1
17999_RT32_B24_L22_148	Tmt6	Yes	No	Yes	No	No	Yes	0	2	1
17999_RT32_B24_L22_153	Tmt6	Yes	No	Yes	Yes	No	Yes	0	1	1
17999_RT32_B24_L22_154	Tmt6	Yes	No	Yes	No	No	Yes	0	0	0
17999_RT32_B24_L22_159	Tmt6	No	No	Yes	No	No	Yes	0	0	0
17999_RT32_B24_L22_163	Tmt6	No	No	Yes	No	No	Yes	0	0	1
17999_RT32_B24_L22_165	Tmt6	No	No	Yes	No	No	Yes	0	0	0
17999_RT32_B24_L22_181	Tmt6	Yes	No	Yes	Yes	No	Yes	0	0	0
17999_RT32_B24_L22_185	Tmt6	Yes	No	Yes	No	No	Yes	0	0	0
17999_RT32_B24_L22_200	Tmt6	Yes	No	Yes	Yes	No	Yes	0	0	0
17999_RT32_B24_L22_205	Tmt6	Yes	No	Yes	Yes	No	Yes	0	0	0
17999_RT32_B24_L22_209	Tmt6	Yes	No	Yes	Yes	No	Yes	0	0	0
17999_RT32_B24_L22_213	Tmt6	Yes	No	Yes	No	No	Yes	0	0	0
17999_RT32_B24_L22_22	Tmt6	Yes	No	Yes	No	No	Yes	0	0	0
17999_RT32_B24_L22_222	Tmt6	Yes	No	Yes	Yes	No	Yes	0	0	1
17999_RT32_B24_L22_223	Tmt6	No	No	Yes	No	No	Yes	0	0	2
17999_RT32_B24_L22_226	Tmt6	Yes	No	Yes	No	No	Yes	0	0	0
17999_RT32_B24_L22_229	Tmt6	No	No	Yes	No	No	Yes	0	0	0
17999_RT32_B24_L22_235	Tmt6	No	No	Yes	No	No	Yes	0	0	0
17999_RT32_B24_L22_238	Tmt6	Yes	No	Yes	No	No	Yes	0	0	0
17999_RT32_B24_L22_24	Tmt6	Yes	No	Yes	No	No	Yes	0	0	0
17999_RT32_B24_L22_247	Tmt6	Yes	No	Yes	No	No	Yes	0	2	0
17999_RT32_B24_L22_248	Tmt6	No	No	Yes	Yes	No	Yes	0	0	0
17999_RT32_B24_L22_252	Tmt6	Yes	No	Yes	No	No	Yes	0	0	0
17999_RT32_B24_L22_253	Tmt6	Yes	No	Yes	No	No	Yes	0	0	0
17999_RT32_B24_L22_256	Tmt6	Yes	No	Yes	Yes	No	Yes	0	0	0
17999_RT32_B24_L22_260	Tmt6	Yes	No	Yes	Yes	No	Yes	0	0	0
17999_RT32_B24_L22_261	Tmt6	Yes	No	Yes	Yes	No	Yes	0	0	0
17999_RT32_B24_L22_268	Tmt6	Yes	No	Yes	Yes	No	Yes	0	0	0
17999_RT32_B24_L22_27	Tmt6	Yes	No	Yes	Yes	No	Yes	0	0	0
17999_RT32_B24_L22_273	Tmt6	Yes	No	Yes	No	No	Yes	0	0	0
17999_RT32_B24_L22_274	Tmt6	Yes	No	Yes	Yes	No	Yes	0	0	0
17999_RT32_B24_L22_275	Tmt6	Yes	No	Yes	Yes	No	Yes	0	0	1
17999_RT32_B24_L22_276	Tmt6	Yes	No	Yes	Yes	No	Yes	0	0	0
17999_RT32_B24_L22_277	Tmt6	Yes	No	Yes	Yes	No	Yes	0	1	1
17999_RT32_B24_L22_283	Tmt6	Yes	No	Yes	No	No	Yes	0	0	0
17999_RT32_B24_L22_284	Tmt6	Yes	No	Yes	No	No	Yes	0	0	0
17999_RT32_B24_L22_29	Tmt6	Yes	No	Yes	Yes	No	Yes	0	0	0
17999_RT32_B24_L22_290	Tmt6	Yes	No	Yes	Yes	No	Yes	0	0	0
17999_RT32_B24_L22_295	Tmt6	No	No	Yes	Yes	No	Yes	0	0	0
17999_RT32_B24_L22_297	Tmt6	Yes	No	Yes	No	No	Yes	0	0	0
17999_RT32_B24_L22_298	Tmt6	Yes	No	Yes	Yes	No	Yes	0	0	1
17999_RT32_B24_L22_303	Tmt6	No	No	Yes	No	No	Yes	0	0	0
17999_RT32_B24_L22_304	Tmt6	Yes	No	Yes	No	No	Yes	0	0	0
17999_RT32_B24_L22_36	Tmt6	No	No	Yes	No	No	Yes	0	2	0
17999_RT32_B24_L22_40	Tmt6	No	No	Yes	Yes	No	Yes	0	0	0
17999_RT32_B24_L22_45	Tmt6	No	No	Yes	No	No	Yes	0	0	0
17999_RT32_B24_L22_46	Tmt6	No	No	Yes	Yes	No	Yes	0	0	1
17999_RT32_B24_L22_51	Tmt6	No	No	Yes	No	No	Yes	0	0	0
17999_RT32_B24_L22_54	Tmt6	No	No	Yes	No	No	Yes	0	1	1
17999_RT32_B24_L22_62	Tmt6	Yes	No	Yes	No	No	Yes	0	0	0
17999_RT32_B24_L22_63	Tmt6	Yes	No	Yes	No	No	Yes	0	0	0
17999_RT32_B24_L22_65	Tmt6	No	No	Yes	No	No	Yes	0	0	0
17999_RT32_B24_L22_72	Tmt6	Yes	No	Yes	No	No	Yes	0	0	0
17999_RT32_B24_L22_75	Tmt6	Yes	No	Yes	No	No	Yes	0	0	0
17999_RT32_B24_L22_77	Tmt6	No	No	Yes	Yes	No	Yes	0	0	0
17999_RT32_B24_L22_80	Tmt6	Yes	No	Yes	Yes	No	Yes	0	0	0
17999_RT32_B24_L22_83	Tmt6	No	No	Yes	No	No	Yes	0	0	0
17999_RT32_B24_L22_85	Tmt6	No	No	Yes	No	No	Yes	0	0	0
17999_RT32_B24_L22_89	Tmt6	No	No	Yes	Yes	No	Yes	0	0	0
17999_RT32_B24_L22_93	Tmt6	No	No	Yes	Yes	No	Yes	1	0	1
17999_RT32_B24_L22_95	Tmt6	Yes	No	Yes	No	No	Yes	0	0	0

17999_RT32_B24_L22_99	Tmt6	Yes	No	Yes	No	No	Yes	0	0	0
17999_RT40_B24_L22_175	Tmt6	Yes	No	Yes	No	No	Yes	0	0	0
17999_RT6_B21_L22_3	Tmt6	Yes	No	Yes	No	No	Yes	0	0	0
3059_B24_L22_69	Tmt6	Yes	No	Yes	No	No	Yes	0	0	0
3112_B24_L23_149	Tmt6	No	No	Yes	No	No	No	0	2	2
3732_PT_B24_L28_179	Tmt6	No	No	Yes	No	No	Yes	0	0	0
4120_B24_L22_270	Tmt6	Yes	No	Yes	No	No	Yes	0	0	0
4684_RT32_6	Tmt6	Yes	No	Yes	No	No	Yes	0	0	0
4760_PT_B24_L22_184	Tmt6	No	No	Yes	No	No	Yes	4	6	3
5080_PT_B24_L22_157	Tmt6	No	No	Yes	No	No	Yes	0	0	0
6194_PT_B24_L28_176	Tmt6	Yes	No	Yes	No	No	Yes	0	2	1
6239_RT32_B24_L22_178	Tmt6	Yes	No	Yes	No	No	Yes	0	0	0
6239_RT32_B24_L22_214	Tmt6	Yes	No	Yes	No	No	Yes	0	1	0
6239_RT32_B24_L22_239	Tmt6	Yes	No	Yes	No	No	Yes	0	0	0
6239_RT32_B24_L22_57	Tmt6	Yes	No	Yes	Yes	No	Yes	0	0	0
6239_RT32_B24_L22_98	Tmt6	Yes	No	Yes	Yes	No	Yes	0	0	0
6274_RT32_B24_L22_232	Tmt6	Yes	No	Yes	No	No	Yes	0	0	0
6305_RT32_B24_L22_97	Tmt6	Yes	No	Yes	No	No	Yes	0	0	0
6358_RT32_B24_L22_125	Tmt6	No	No	Yes	No	No	Yes	0	0	1
6480_RT32_B24_L28_106	Tmt6	Yes	No	Yes	No	No	Yes	0	0	1
6523_RT32_B24_L22_35	Tmt6	No	No	Yes	No	No	Yes	0	0	0
6665_RT32_B24_L22_42	Tmt6	No	No	Yes	Yes	No	Yes	0	0	0
6665_RT32_B24_L22_43	Tmt6	No	No	Yes	No	No	Yes	0	0	0
6704_RT32_B24_L28_234	Tmt6	No	No	Yes	No	No	Yes	0	1	0
6800_RT32_B24_L22_224	Tmt6	Yes	No	Yes	Yes	No	Yes	0	0	0
6909_RT32_B24_L22_113	Tmt6	Yes	No	Yes	No	No	Yes	0	1	0
6925_RT32_B22_L23_10	Tmt6	Yes	No	Yes	Yes	No	Yes	0	0	0
6946_RT32_B24_L22_188	Tmt6	No	No	Yes	No	No	Yes	0	0	0
7060_PT_B24_L22_231	Tmt6	No	No	No	No	No	No	0	0	0
7110_RT32_B23_L25_14	Tmt6	Yes	Yes	Yes	Yes	Yes	Yes	0	0	0
7127_RT6_B21_L23_4	Tmt6	No	No	Yes	Yes	No	Yes	0	0	0
7199_RT32_B24_L22_117	Tmt6	Yes	No	Yes	Yes	No	Yes	0	1	2
7199_RT32_B24_L22_142	Tmt6	Yes	No	Yes	No	No	Yes	0	4	0
7272_RT32_B24_L22_131	Tmt6	No	No	Yes	Yes	No	Yes	0	0	0
7272_RT32_B24_L22_210	Tmt6	Yes	No	Yes	No	No	Yes	0	0	0
7332_RT32_B25_L27_18	Tmt6	Yes	Yes	Yes	No	No	Yes	0	2	2
7403_RT32_B24_L22_145	Tmt6	No	No	Yes	No	No	Yes	0	0	0
7456_RT32_B24_L28_143	Tmt6	No	No	Yes	Yes	No	Yes	0	0	1
7695_RT32_B24_L22_88	Tmt6	Yes	No	Yes	No	No	Yes	0	0	0
7886_B24_L22_215	Tmt6	No	No	Yes	No	No	Yes	0	0	0
7910_RT32_B24_L22_269	Tmt6	Yes	No	Yes	No	No	Yes	0	2	0
7948_RT32_B24_L28_139	Tmt6	Yes	No	Yes	Yes	No	Yes	0	0	0
7990_RT32_B24_L22_190	Tmt6	Yes	No	Yes	No	No	Yes	0	0	0
8039_RT32_B24_L22_38	Tmt6	No	No	Yes	No	No	Yes	0	0	2
8041_RT32_B24_L28_296	Tmt6	No	No	Yes	No	No	Yes	0	0	0
8099_B24_L22_90	Tmt6	No	No	Yes	No	No	Yes	0	0	0
8160_RT32_B24_L22_233	Tmt6	Yes	No	Yes	Yes	No	Yes	0	0	0
8160_RT32_B24_L22_244	Tmt6	Yes	No	Yes	Yes	No	Yes	0	0	0
8160_RT32_B24_L22_25	Tmt6	No	No	Yes	No	No	Yes	0	0	0
8160_RT32_B24_L22_285	Tmt6	Yes	No	Yes	No	No	Yes	0	0	0
8160_RT32_B24_L22_56	Tmt6	Yes	No	Yes	No	No	Yes	0	0	0
8446_PT_B24_L22_287	Tmt6	No	No	Yes	No	No	Yes	4	13	4
8786_RT32_B24_L22_102	Tmt6	Yes	No	Yes	No	No	Yes	0	0	1
8909_RT32_B24_L28_258	Tmt6	Yes	No	Yes	Yes	No	Yes	0	0	0
8934_RT32_B24_L22_228	Tmt6	Yes	No	Yes	No	No	Yes	0	0	0
8999_B24_L22_158	Tmt6	Yes	No	Yes	No	No	Yes	0	0	0
8999_B24_L22_263	Tmt6	No	No	Yes	Yes	No	Yes	0	0	0
8999_B24_L22_70	Tmt6	Yes	No	Yes	No	No	Yes	0	0	0
8999_B24_L22_87	Tmt6	Yes	No	Yes	Yes	No	Yes	0	0	0
9122_PT_B21_L22_9	Tmt6	Yes	No	Yes	No	No	Yes	0	0	0
9150_RT32_B24_L24_108	Tmt6	Yes	No	No	No	No	Yes	0	0	1
9549_PT_B24_L22_96	Tmt6	Yes	No	Yes	No	No	Yes	0	0	1
9900_RT32_B24_L22_112	Tmt6	Yes	No	Yes	No	No	Yes	0	2	1
9900_RT32_B24_L22_152	Tmt6	Yes	No	Yes	No	No	Yes	0	0	0
9900_RT32_B24_L22_266	Tmt6	No	No	Yes	No	No	Yes	0	0	0
9900_RT32_B24_L22_267	Tmt6	Yes	No	Yes	No	No	Yes	0	1	0
9900_RT32_B24_L22_271	Tmt6	No	No	Yes	No	No	Yes	0	1	1
9900_RT32_B24_L22_279	Tmt6	Yes	No	Yes	Yes	No	Yes	0	0	0
9900_RT32_B24_L22_299	Tmt6	No	No	Yes	No	No	Yes	0	0	0
9900_RT40_B24_L22_217	Tmt6	No	No	Yes	No	No	Yes	0	0	0
9936_RT32_B24_L22_119	Tmt6	Yes	No	Yes	No	No	Yes	0	0	0

Name repeatscout elements	Clade	RT identification
R=0	Tmt6	Yes
R=1	Tmt1	Yes
R=10	Tmt3	Yes
R=106	Tmt1	No
R=1070	Tmt6	No
R=11	Tmt3	Yes
R=110	Tmt3	Yes
R=120	Tmt3	No
R=139	Tmt3	No
R=1394	Tmt3	No
R=14	Tmt6	Yes
R=1400	Tmt6	No
R=1420	Tmt1	Yes
R=1457	Tmt6	No
R=1522	Tmt1	No
R=1586	Tmt6	No
R=1693	Tmt1	No
R=1751	Tmt3	No
R=2	Tmt1	Yes
R=218	Tmt3	No

R=221	Tmt3	No
R=3	Tmt3	No
R=308	Tmt6	No
R=364	Tmt6	No
R=379	Tmt3	No
R=4	Tmt1	Yes
R=42	Tmt3	No
R=432	Tmt3	No
R=434	Tmt5	Yes
R=473	Tmt1	Yes
R=5	Tmt3	No
R=546	Tmt6	No
R=576	Tmt3	No
R=608	Tmt3	No
R=624	Tmt3	No
R=711	Tmt3	No
R=734	Tmt3	No
R=762	Tmt4	Yes
R=778	Tmt4	Yes
R=7a	Tmt1	Yes
R=7b	Tmt2	Yes
R=873	Tmt6	No

Table S1: Table summarizing the mapping of reads against LTR and non LTR of full length gypsy retrotransposons corresponding to insertion point. Only the elements for those at least one insertion point is lacking are showed. A 0 indicated that reads did not overlapped although a 1 indicated that at least one reads overlapped the LTR and non LTR sequence. The insertional time estimated in Martin et al (2010) is indicated as well as for each element both 5' and 3' LTR.

Name element	Estimated insertional time	Clade	Spain-1		Spain-2		France-Bur		France-Als		Italy		France-Alp	
			5	3	5	3	5	3	5	3	5	3	5	3
6239_RT32_B24_L22_98	0	Tmt6	0	0	0	0	0	0	0	0	0	0	0	0
17999_RT32_B24_L22_29	122880	Tmt6	1	1	0	0	1	1	1	1	1	1	1	1
17999_RT32_B24_L22_276	135428	Tmt6	0	0	0	0	0	0	0	0	0	0	0	0
17999_RT32_B24_L22_80	257692	Tmt6	0	0	1	1	1	1	0	0	1	1	1	1
9900_RT32_B24_L22_279	308923	Tmt6	1	1	1	1	0	0	1	1	1	1	1	1
17999_RT32_B24_L22_200	309554	Tmt6	1	1	0	0	1	1	1	1	0	0	0	0
17999_RT32_B24_L22_63	389815	Tmt6	0	0	1	1	1	1	1	1	1	1	1	1
17999_RT32_B24_L22_205	516262	Tmt6	1	1	0	0	0	0	0	0	0	0	0	0
17999_RT32_B24_L22_283	519046	Tmt6	1	1	1	1	1	1	1	1	1	1	1	0
12120_RT32_B24_L22_118	553400	Tmt6	1	1	0	0	1	1	1	1	0	1	0	0
17999_RT32_B21_L22_19	553400	Tmt6	0	1	0	0	1	1	1	1	1	1	1	1
17999_RT32_B24_L22_238	616615	Tmt6	1	1	1	1	1	1	0	0	0	1	1	1
17999_RT32_B24_L22_213	642092	Tmt6	0	0	0	0	0	0	0	0	1	1	1	1
12120_RT32_B24_L22_218	951338	Tmt6	0	0	0	0	0	0	1	1	0	0	1	1
17999_RT32_B24_L22_153	951631	Tmt6	1	1	1	0	1	1	1	1	1	1	1	1
8909_RT32_B24_L28_258	977831	Tmt6	1	1	1	1	1	1	1	1	0	1	1	1
12120_RT32_B24_L22_212	1075846	Tmt6	1	1	0	0	0	0	0	0	0	0	0	0
10544_RT32_B24_L22_230	1076923	Tmt6	1	0	1	0	1	0	1	0	1	0	1	0
12120_RT32_B24_L22_150	1096662	Tmt6	1	1	1	1	0	0	0	0	1	1	1	1
8160_RT32_B24_L22_285	1131215	Tmt6	1	0	1	1	1	1	1	1	1	1	1	1
17999_RT32_B24_L22_261	1143846	Tmt6	1	1	1	1	1	1	1	1	0	0	1	1
17999_RT32_B24_L22_99	1225862	Tmt6	1	1	1	1	1	1	1	1	1	1	0	0
17999_RT28_B24_L22_281	1290646	Tmt6	1	1	1	1	1	1	1	1	1	1	0	0
6665_RT32_B24_L22_42	1390169	Tmt6	0	0	1	1	1	1	1	1	1	1	1	1
8160_RT32_B24_L22_233	1434462	Tmt6	0	0	0	0	1	1	1	1	1	1	1	1
17999_RT32_B24_L22_235	1441185	Tmt6	1	1	0	1	1	1	1	1	1	1	1	1
17999_RT32_B24_L22_85	1636615	Tmt6	1	0	1	1	1	1	0	1	1	1	1	1
17999_RT32_B24_L22_110	1639846	Tmt6	0	0	0	1	0	0	1	1	1	1	1	1
12120_RT32_B24_L22_219	1804154	Tmt6	0	0	0	0	0	0	0	0	0	0	1	1
17999_RT32_B24_L22_304	1853538	Tmt6	1	1	1	1	1	0	1	1	1	0	0	1
9122_PT_B21_L22_9	1897538	Tmt6	1	0	1	1	1	1	1	1	1	1	1	1
17999_RT32_B21_L22_20	1955692	Tmt6	0	0	1	0	1	1	1	1	1	1	1	1
17999_RT32_B24_L22_45	2174462	Tmt6	0	0	1	1	1	1	1	1	1	1	1	1
17999_RT32_B24_L22_141	2218923	Tmt6	0	1	1	1	1	1	1	1	1	1	1	1
8999_B24_L22_263	2306308	Tmt6	1	1	1	0	0	0	1	1	1	1	1	1
8999_B24_L22_87	2426154	Tmt6	1	1	1	1	1	0	1	0	1	0	1	1
8999_B24_L22_70	2471385	Tmt6	1	0	1	0	1	0	1	1	1	0	1	1
3059_B24_L22_69	2496462	Tmt6	1	1	1	0	1	1	1	1	1	0	1	1
17999_RT32_B24_L22_165	2532462	Tmt6	1	1	1	1	1	0	1	1	1	0	1	1
12120_RT32_B24_L22_170	2625846	Tmt6	1	1	1	1	1	1	1	1	0	1	0	1
17999_RT32_B24_L22_159	2637385	Tmt6	1	1	1	1	1	1	1	1	1	1	0	0
6909_RT32_B24_L22_113	2736923	Tmt6	0	0	0	0	0	0	1	1	1	1	1	1
17099_RT32_B24_L22_262	2777077	Tmt6	1	1	1	1	1	1	0	1	0	1	1	1
17999_RT32_B24_L22_260	3065231	Tmt6	1	1	1	0	1	1	1	1	1	0	1	1
8099_B24_L22_90	3116308	Tmt6	0	0	1	1	1	1	1	1	1	1	1	1
12750_RT32_B24_L22_33	3320462	Tmt6	0	1	0	1	1	1	1	1	1	1	1	1
10000_PT_B24_L22_198	3391385	Tmt6	1	1	1	1	0	1	1	1	1	1	1	1
9149_RT34_B21_L24_60	3427231	Tmt3	1	0	0	1	1	0	1	1	1	1	1	1
12120_RT32_B24_L22_53	3503692	Tmt6	1	1	1	1	1	1	1	1	1	1	0	0
17000_RT32_B24_L22_84	3717538	Tmt6	0	0	1	1	1	1	0	0	1	1	1	1
12985_RT32_B24_L24_174	3831385	Tmt1	1	0	1	0	1	0	1	1	1	0	1	0
12120_RT32_B24_L22_76	3861077	Tmt6	1	1	1	1	1	0	1	0	1	0	1	1
9900_RT32_B24_L22_266	3936308	Tmt6	1	1	1	1	0	1	0	1	1	1	1	1
13485_RT32_B24_L22_164	3954000	Tmt6	1	1	1	1	1	0	1	1	1	0	1	1
17099_RT32_B24_L22_180	3988615	Tmt6	1	0	1	0	1	0	1	0	1	1	1	1
17999_RT32_B24_L22_290	4165077	Tmt6	1	0	1	1	1	1	1	1	1	1	1	1
6665_RT32_B24_L22_43	4281231	Tmt6	1	1	1	0	1	1	1	1	0	0	1	1
17999_RT32_B24_L22_93	4313385	Tmt6	0	0	1	1	0	0	1	1	0	0	0	0
12120_RT32_B24_L22_282	4424154	Tmt6	1	1	1	1	1	1	1	1	1	1	1	0
4328_RT34_B21_L22_59	4569692	Tmt3	1	0	1	1	0	0	1	1	0	1	1	1
17999_RT32_B24_L22_83	4573846	Tmt6	0	1	1	1	1	1	0	1	0	1	1	1
17999_RT32_B24_L22_253	4967385	Tmt6	1	1	1	1	1	1	1	1	0	1	1	1
4120_B24_L22_270	5137692	Tmt6	1	1	1	0	1	1	1	1	1	1	1	1
12599_RT38_B21_L5_86	6308154	Tmt3	1	1	1	1	1	1	1	1	0	1	1	1
11119_RT32_B24_L22_91	6493538	Tmt6	1	1	1	1	1	0	1	1	1	1	1	1
11513_RT32_B24_L22_123	6993077	Tmt6	1	1	0	0	1	1	1	1	0	0	0	0
10000_PT_B24_L22_265	7380615	Tmt1	1	1	1	1	1	1	0	1	1	1	1	1

7 Discussion générale et perspectives

Les truffes sont connues et étudiées depuis plusieurs centaines d'années. Toutefois, les connaissances acquises sur leur cycle biologique et leur écologie sont restées limitées jusque dans les années 1990. Depuis une vingtaine d'années, les études portant sur leur diversité génétique, leur écologie et leur biologie se sont multipliées grâce à l'essor des approches moléculaires et l'activité concertée d'un réseau scientifique franco-italien ayant abouti au séquençage du génome de la truffe noire du Périgord. Les progrès de la biologie moléculaire ont permis d'améliorer nos connaissances sur la taxonomie de plusieurs espèces, mais aussi de mettre au point des diagnostics moléculaires utilisés par les services des fraudes. Il est maintenant possible d'identifier rapidement la plupart des truffes européennes et asiatiques à partir d'échantillons environnementaux ou d'aliments (Douet et al. 2004; Mabru et al. 2004). De même, les données concernant la diversité génétique et la structure génétique des populations de truffes se sont multipliées (Bertault et al. 1998; Bertault et al. 2001; Murat et al. 2004; Murat et al. 2013; Riccioni et al. 2008; García-Cunchillos et al. 2013). Depuis le séquençage du génome de la truffe noire du Périgord, *Tuber melanosporum*, en 2010 (Martin et al. 2010) de nouveaux outils moléculaires sont maintenant disponibles. Grâce à l'amélioration des techniques de séquençage à haut débit, des projets de génomique comparative abordant le polymorphisme inter- et intraspécifique de différentes espèces de truffes d'intérêt économique sont désormais entrepris.

Mon projet de thèse s'inscrit dans le cadre de ces projets de génomique comparative. Les objectifs de mon travail de thèse étaient d'utiliser les ressources génomiques que nous avons générées pour mieux connaître l'évolution des champignons mycorhiziens ascomycètes en prenant les truffes comme modèles d'études. Ce travail complète les travaux de génomique comparative des champignons ectomycorhiziens basidiomycètes de l'équipe qui viennent

d'être publiés dans *Nature genetics* (Kohler et al. 2015). Pour mener à bien ce travail, j'ai développé une approche de génomique comparative au niveau interspécifique et intraspécifique. Les principaux résultats de mon travail sont :

1. Une meilleure connaissance de l'évolution des génomes au sein des *Tuberaceae* qui ont subi une expansion remarquable de leur taille due à la prolifération d'éléments transposables et pour les *Tuber* une réduction de leur répertoire de gènes principalement ceux impliqués dans la dégradation de la matière organique végétale (PCWDE). Au contraire, *Choiromyces venosus* se caractérise par une expansion principalement de gènes orphelins (Chapitre 3).
2. La mise en évidence d'une convergence évolutive entre champignons ECM basidiomycètes et ascomycètes. En effet, le mode de vie symbiotique a entraîné (ou a été favorisé par) la diminution considérable de l'arsenal enzymatique de dégradation de la lignocellulose. Du fait de cette érosion génomique remarquable, les champignons ECM, dont les truffes, ont perdu une partie de leurs capacités saprotrophiques (Chapitre 3).
3. En comparant les génomes de sept truffes noires (*T. melanosporum*), nous avons mis en évidence un niveau de diversité génétique intraspécifique similaire (3540 SNPs/Mpb) à celui d'autres champignons filamenteux (Chapitre 4).
4. Nous avons identifié plus de 400 000 mutations nucléotidiques ponctuelles de type « *single nucleotide polymorphism* » (SNP) pouvant être utilisées pour étudier la structure géographique des populations de cette truffe (Chapitre 4).
5. Il existe des traces d'adaptation génétique dans le génome codant de *T. melanosporum*, ouvrant la voie à une possible sélection de variants génétiques

possédant des caractéristiques phénotypiques et organoleptiques spécifiques (Chapitre 4).

6. Six familles de rétrotransposons de type gypsy sont présentes dans le génome de *T. melanosporum* et dont deux l'ont massivement colonisé représentant plus de 20 % de celui-ci. L'invasion du génome par ces rétrotransposons est ancienne (> 1 million d'années) et le taux de transposition était plus important dans le passé. Par contre, une famille (*Tmt6*) semble avoir été active récemment (entre 50 et 20 mille ans). Enfin, leur mécanisme de transposition semble aléatoire, c'est à dire que chaque copie peut transposer. La prolifération des rétrotransposons de type gypsy à partir d'une séquence ancestrale unique n'est pas supportée par les analyses que nous avons effectuées (Chapitre 5).

7.1 Evolution des génomes au sein des Tuberaceae : comment expliquer la grande taille de leurs génomes ?

Les génomes des Tuberaceae étudiés (*C. venosus*, *T. aestivum*, *T. magnatum* et *T. melanosporum*) se caractérisent par leur grande taille qui est comprise entre 125 Mpb et 192 Mpb (Chapitre 3). Cette taille est remarquable dans la mesure où la plupart des génomes des champignons filamenteux ont des génomes de taille comprise entre 20 Mpb et 60 Mpb (Raffaele and Kamoun 2012). Toutefois, un certain nombre de champignons pathogènes possèdent des génomes supérieurs à 100 Mpb comme *Blumeria graminis*, *Golovinomyces orontii* et *Erysiphe pisi* (Raffaele and Kamoun 2012). Dans le cas de *B. graminis*, la taille de son génome est due à un excès de séquences répétées (~ 64 %), mais par contre il possède peu de gènes codant des protéines (5884). Raffaele et Kamoun (2012) suggèrent qu'un fort contenu en transposons constitue un avantage pour les oomycètes et champignons pathogènes car ces génomes sont plus plastiques. Ils peuvent évoluer plus rapidement, les transposons

favorisant la néo-fonctionnalisation des gènes. L'activation et la prolifération de ces transposons ont comme conséquence d'augmenter la taille du génome. Il est intéressant de noter que les génomes de ces pathogènes présentent des caractéristiques très proches de ceux des truffes. Mais comment expliquer ce caractère et surtout l'expansion du génome due à une colonisation par des éléments transposables ? En 2009, Zeh et collaborateurs (Zeh, Zeh, and Ishida 2009) ont établi une théorie appelée « *epi-transposon equilibrium* » proposant que l'évolution est guidée par des épisodes rapides de colonisation des génomes par des transposons (voir Chapitre 1). Elle suggère que les transposons sont contrôlés par des mécanismes réversibles. A la suite d'événements de stress comme un changement d'hôte par exemple, ces mécanismes de contrôle peuvent être inactivés ce qui va permettre aux transposons de rapidement coloniser le génome. Leur prolifération provoquerait des mutations par insertion, des néo-fonctionnalisations par duplication et des changements dans la régulation de certains gènes. Rapidement le contrôle de cette invasion se met en place afin de protéger le génome colonisé. Chez les Pézizomycètes, nous n'avons pas mis en évidence des traces de *repeated induced point mutations* (RIP) qui est un mécanisme irréversible induisant des mutations ponctuelles pour contrôler les séquences répétées (Chapitre 3). Une étude récente du méthylome de *T. melanosporum* met en évidence un taux important de méthylation dans les séquences de transposons suggérant des mécanismes de défense comme le MIP (*Methylation Induced Premeiotically* ; Montanini et al. 2014). En se basant sur cette théorie de l'« *epi-transposon equilibrium* » deux modèles d'évolution des génomes de Tuberales peuvent être proposés (Fig. 1 et 2). Ces deux modèles postulent l'existence d'un ancêtre commun aux Pézizomycètes vivant il y a environ 450 millions d'années. Cet ancêtre aurait eu un génome de taille réduite (20-60 Mpb), des transposons efficacement contrôlés par des mécanismes réversibles de type MIP et probablement, un contenu en CAZymes important. En effet, il a été récemment suggéré que l'ancêtre des ECM Basidiomycètes était probablement

une pourriture blanche riche en CAZymes et peroxydases (Kohler et al. 2015). L'ancêtre des Tuberales (*Tuber* spp et *Choiromyces*) aurait subi une augmentation de la taille de son génome suite à la prolifération d'éléments transposables. Sur la base de la théorie « *epi-transposon equilibrium* » une première hypothèse postule que le stress engendré par un changement d'hôte ou bien l'apparition d'un premier hôte possible, c'est à dire un arbre, aux alentours de 180 millions d'années aurait induit l'inactivation des transposons. La deuxième hypothèse postule l'arrêt de la régulation des transposons suite à un stress inconnu avant la mise en place de l'interaction avec les arbres. La colonisation du génome par les transposons aurait entraîné une augmentation de la plasticité de l'espèce et donc sa meilleure adaptation à de nouvelles niches, lui permettant d'interagir avec les arbres. Il est à l'heure actuelle impossible de discriminer entre ces deux théories mais il semble que l'expansion du génome de l'ancêtre des Tuberales a débuté aux alentours de 180 millions d'années, époque à laquelle les forêts des ancêtres des Pinaceae, premiers arbres ECM, couvraient la planète (Rothwell et al. 2012). Aux alentours de 100 millions d'années, l'ancêtre des *Tuber* a vu son répertoire de gènes diminuer, avec en particulier une perte massive des PCWDE. En revanche, *C. venosus* a connu une évolution différente de celle des *Tuber* et a conservé un large répertoire de gènes (~18000) dont de nombreux gènes orphelins. Le carbone utilisé pour la croissance de ces ascomycètes semble d'origine symbiotique sur la base de l'abondance naturelle $^{13}\text{C}/^{15}\text{N}$ (Hobbie, Weber, and Trappe 2001) comme cela a été aussi démontré récemment pour *T. melanosporum* (Le Tacon et al. 2013). *C. venosus* a cependant un contenu en PCWDE plus proche d'une espèce saprotrophe comme le montrent les figures S21 et S22 du chapitre 4. Peu d'études existent sur cette espèce ; il me semble pertinent de mieux connaître son écologie et son mode d'interaction avec les arbres à travers, par exemple, une analyse de transcriptomique de ces mycorhizes.

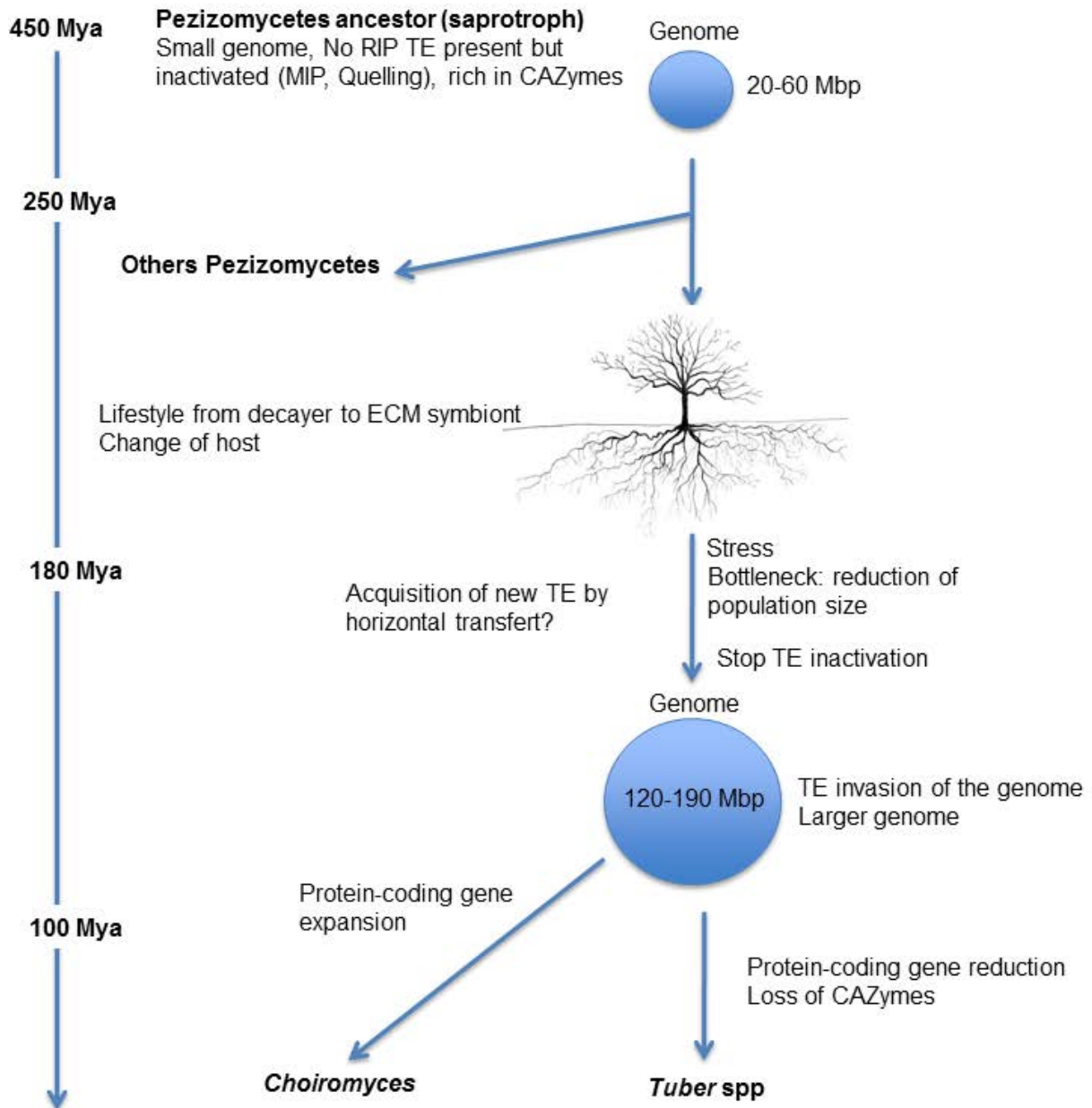


Figure 1 : Première hypothèse expliquant l'évolution des génomes de Tuberaeae (RIP : Repeat Induced Point Mutation ; MIP : Methylation Induced Premeiotically ; TE : Transposable elements ; CAZymes : Carbohydrate-Active enZymes).

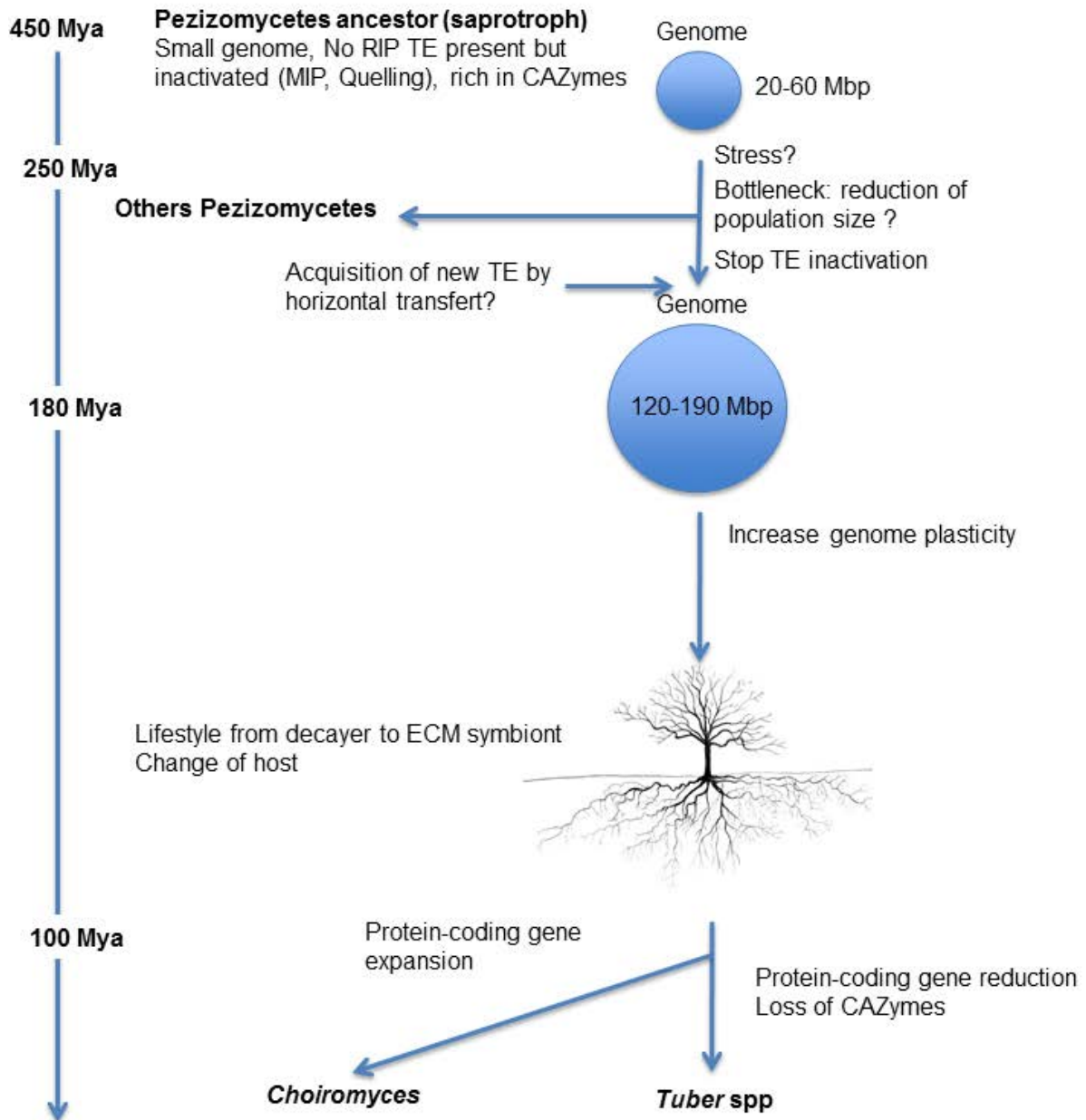


Figure 2 : Seconde hypothèse expliquant l'évolution des génomes de Tuberaeae.

Afin de mieux comprendre l'évolution des truffes, d'autres génomes sont en cours de séquençage (voir Chapitre 2) comme celui de *T. borchii* au JGI, *T. lyonii* à l'université de Duke (US), *T. brumale* et *T. indicum* toutes au sein de notre laboratoire.

Les différences trouvées entre *C. venosus* et les *Tuber* sont surprenantes. Il me semble pertinent de compléter notre analyse comparative en étudiant des génomes d'autres espèces de Tuberales comme, par exemple, des *Dingleya*, *Reddellomyces*, *Labyrinthomyces* ou *Nothojafnea* (Bonito et al. 2013). De même, le séquençage d'espèces proches et intermédiaires entre les Tuberales et Morchella comme des espèces appartenant aux Morchellaceae (ex. *Verpa* et *Disciotis*) et Discinaceae (ex. *Gymnohydnotrya*, *Leucangium*, *Hydnotrya* *Gyromitra* ou *Discina*) permettrait de mieux appréhender l'évolution des Pézizomycètes (Hansen et al. 2013; Bonito et al. 2013). Une autre possibilité serait de se concentrer sur les champignons mycorhiziens du genre *Wilcoxina* afin d'étudier les points communs entre les champignons mycorhiziens chez les Pézizomycètes.

Avec l'amélioration constante des techniques de séquençage à haut débit et la réduction de leurs coûts, le séquençage de nouveaux génomes est maintenant plus aisé. En tenant compte des résultats obtenus pour *T. magnatum* (voir Chapitre 4), la stratégie de séquençage que nous avons adopté repose sur la technique Illumina avec une couverture de séquençage de 50 à 100 X via la construction de trois banques d'ADN : une banque « *paired end* » avec des inserts de petite taille (environ 200 pb) et deux banques « *mate paired* » avec des inserts de 3 et 8 Kpb. Cette stratégie de séquençage permet l'utilisation de l'assembleur génomique ALLPATH-LG développé par le Broad Institute (Gnerre et al. 2010). Les résultats préliminaires obtenus sur *T. brumale* et *T. indicum* nous confirment que cette stratégie permet d'obtenir un génome de bonne qualité pour un coût réduit.

7.2 Convergences et divergences évolutives entre les champignons ectomycorhiziens ascomycètes et basidiomycètes

Dans une étude de génomique comparative conduite par le consortium *Mycorrhizal Genomics Initiative*, coordonné par l'Inra, Kohler et collaborateurs (2015) ont analysé les processus évolutifs ayant conduit à l'apparition des champignons mycorhiziens il y a plus de 200 millions d'années. Treize génomes de champignons ECM basidiomycètes ont été séquencés et analysés avec ceux de *Laccaria bicolor* (Martin et al. 2008), *T. melanosporum* (Martin et al. 2010), *Rhizophagus irregularis*, un champignons endomycorhiziens (Tisserant et al. 2013) et 33 génomes de champignons saprotrophes. Cette étude met en évidence une corrélation remarquable entre le contenu en PCWDE et la stratégie de vie des champignons (pourriture blanche, pourriture brune ou symbiote ectomycorhizien). D'autre part, l'apparition de la symbiose ECM dans plusieurs des lignées fongiques est confirmée suggérant une convergence évolutive de ces champignons appartenant à des lignées couvrant plus de 180 millions d'années. Parmi ces convergences dans les répertoires de gènes, on observe une diminution du contenu en PCWDE, certaines étant rares voire absentes comme les enzymes des familles glycosyl hydrolases 6 et 7 (GH6 et GH7). Ces enzymes sont des cellobiohydrolases se distinguant par leurs mécanismes d'actions pour dégrader la cellulose et générer le cellobiose à partir de l'extrémité non réductrice pour la GH6 (<http://www.cazy.org/GH6.html>) et de l'extrémité réductrice pour la GH7 (<http://www.cazy.org/GH7.html>). Cette absence de PCWDE est caractéristique de *T. melanosporum*, le seul ascomycète ECM étudié jusqu'à maintenant (Martin et al. 2010). Dans le chapitre 3, nous avons montré que les trois espèces de truffes (*T. aestivum*, *T. magnatum* et *T. melanosporum*) ont un contenu en CAZymes proche d'un autre champignon ECM ascomycète (*Cenococcum geophilum*) et de plusieurs champignons ECM basidiomycètes comme *Scleroderma citrinum*, *Paxillus involus*, *Paxillus rubicundulus*, *Amanita muscaria* et *Piloderma croceum*. La convergence évolutive est donc

confirmée d'autant plus que la GH7 n'a pas été identifié dans ces génomes. Toutefois, *T. aestivum* a conservé une copie de GH6 et *T. magnatum* trois copies. En conclusion, nous confirmons que les trois truffes se caractérisent par une perte massive de PCWDE, également observée chez les ECM basidiomycètes.

Les champignons ECM, quels que soient leurs origines phylogénétiques, forment une structure qui pénètre dans les racines de la plante (réseau de Hartig) et doivent donc envoyer des signaux à la plante afin de minimiser l'induction des réactions de défense. Le fonctionnement des ECM consiste à un échange réciproque entre la plante qui fournit des sucres simples et le champignon qui fournit des minéraux. Ces échanges sont vraisemblablement régis par des un échange de signaux entre les deux partenaires impliquant des protéines sécrétées. Parmi les protéines sécrétées les MiSSP (*Mycorrhizal induced Small Secreted Protein*) sont de petite taille et sont fortement induites dans l'ECM lorsque *L. bicolor* interagit avec sa plante hôte (Martin et al. 2008; Plett et al. 2011; Plett et al. 2014). Chez *L. bicolor*, l'interaction entre MiSSP7 et JAZ6, le co-récepteur de l'acide jasmonique du peuplier est indispensable à l'établissement de la symbiose ECM (Plett et al. 2014). A la différence de *L. bicolor*, aucune MiSSP n'a été identifiée parmi les protéines les plus fortement surexprimées chez *T. melanosporum* (Martin et al. 2010). Toutefois, 19 MiSSP sont induites chez *T. melanosporum* mais à des niveaux d'induction bien plus faible que ceux de *L. bicolor* (Chapitre 3). Aucune de celles-ci n'est spécifique à *T. melanosporum* ce qui est une différence avec ce qui a été trouvé chez les ECM basidiomycètes (Kohler et al. 2015). Dans notre étude, nous avons trois espèces du même genre alors que pour la plupart des ECM basidiomycètes une seule espèce par genre est présente. Toutefois, un certain nombre de MiSSP, dont MiSSP7, sont spécifiques à *L. bicolor* alors qu'il a été comparé à *L. amethystina*, suggérant que l'évolution des MiSSP chez les basidiomycètes est peut-être plus rapide.

Malheureusement, nous ne disposons du transcriptome des ascocarpes, du mycélium et des ECM que pour *T. melanosporum*. Pour *T. aestivum*, uniquement un séquençage RNAseq d'ECM et d'ascocarpes a été effectué et pour *T. magnatum* uniquement un séquençage utilisant la technologie 454 est disponible. Il faut toutefois rappeler que la culture *in vitro* des truffes est très difficile puisque les mycéliums poussent très lentement. Il est donc difficile d'obtenir le matériel pour les extractions d'ARN. Ceci est surtout vrai pour *T. magnatum* qui pendant longtemps a été récalcitrante à toute culture *in vitro* et à la synthèse d'ECM. Toutefois, grâce à la collaboration des Pépinières Robin, des ECM de *T. aestivum*, *T. magnatum* et *T. melanosporum* sur *Quercus robur* ont été récemment récoltées. Le choix de *Q. robur* est dû au fait que le génome de cette espèce de chêne est en cours de finalisation ce qui permettra d'analyser simultanément les transcriptomes des champignons et de la plante.

7.3 Les ressources génomiques vont-elles impacter la gestion des truffières ?

Il existe de nombreux reportages à la télévision sur les truffes. Ces reportages se focalisent toujours sur la gastronomie et le folklore et donnent souvent une image erronée et passéiste de la filière trufficole. En effet, il s'agit d'une filière professionnelle dynamique tournée vers l'innovation. A titre d'exemple, à partir du séquençage du génome de la truffe noire en 2010, un brevet international a été déposé en 2012 sur un kit de diagnostics des types sexuels (sexage). Ce dernier a fait l'objet d'une licence de savoir-faire avec une société de biotechnologie. Cette innovation devrait faciliter l'étude des facteurs influant la production de truffes. D'autre part, des marqueurs microsatellites très polymorphes ont été identifiés chez *T. melanosporum* (Murat et al. 2011) mais également chez *T. aestivum* (Molinier et al. 2013). Toutefois, les marqueurs microsatellites neutres ne permettent pas d'étudier les traces de

sélection et l'adaptation des espèces à leur environnement. Dans le chapitre 4, nous avons identifié un jeu de SNPs (*Single Nucleotide Polymorphism*) par la comparaison de sept génomes de *T. melanosporum*. Ces outils moléculaires faciliteront le développement de plusieurs études. Tout d'abord, nous avons montré que leur utilisation pour en phylogénie permet de discriminer les échantillons en fonction de leur origine géographique. Si les SNPs permettent de discriminer l'origine géographique des truffes ou tout au moins les truffes d'origine française par rapport à celles provenant d'Espagne ou d'Italie cela pourrait aider la fédération française des trufficulteurs (FFT) dans son souhait d'une appellation « truffe de France ». Cette demande est justifiée par un afflux de plus en plus important de truffes d'autres pays Européens entraînant une baisse des prix sur les marchés de gros français.

Parmi les techniques culturales de plus en plus populaires chez les trufficulteurs on note les apports de spores annuels en périphérie des arbres mycorhizés. Toutefois, nous ne savons pas si ces apports de spores ont un réel rôle dans promotion de la production des ascocarpes. Les outils mis au point dans cette thèse, tel que les SNPs pour *T. melanosporum*, la caractérisation des gènes de compatibilité sexuelle, les outils bioinformatiques pour comparer les génomes vont nous permettre de déterminer si les spores apportées contribuent à la formation des ascocarpes. Pour cela, nous caractériserons les mélanges de spores à utiliser comme inoculum et nous pourrons ainsi les tracer *in situ* au niveau des ECM et des ascocarpes récoltés dans la truffière traitée.

Depuis plusieurs millénaires, l'agriculteur sélectionne les semences, ou bien les animaux, en fonction des caractères qu'il souhaite maintenir et favoriser. Les truffes sont cultivées depuis le 19^{ème} siècle, mais à l'époque les trufficulteurs plantés des arbres uniquement en comptant sur l'inoculum naturel de truffes. Grâce à l'INRA, il est possible depuis 1972, d'inoculer les arbres en conditions contrôlées pour ensuite réaliser des vergers truffiers. Négligée jusqu'à

aujourd'hui, la sélection classique a débuté chez certains trufficulteurs. En effet, l'Union régionale des trufficulteurs d'Aquitaine fait produire des plants inoculés avec des truffes qu'elle considère comme tolérantes à différents stress comme le gel ou la sécheresse. Il est encore impossible de dire s'il existe des caractères génétiques d'adaptation des truffes aux stress environnementaux mais étant donné que nous avons trouvé une diversité génétique significative pour *T. melanosporum* et des traces de sélection permettent d'émettre l'hypothèse que cette truffe pourrait s'adapter à son environnement. Les SNPs sont utilisés pour identifier des régions génomiques intéressantes pour des traits quantitatifs (QTL : *quantitative traits loci*). Les SNPs que nous avons identifiés pourraient donc être utilisés pour identifier des QTLs chez *T. melanosporum*. Toutefois, il faut être réaliste et les croisements entre mycélium de truffes n'étant pas encore maîtrisés il sera difficile d'identifier des QTL. Par contre, une approche de génomique des populations par séquençage d'échantillons multiples de populations différentes, et représentatives, ayant des traits différents peut être envisagée. Parmi les traits qui pourraient être étudiés, l'adaptation à la sécheresse édaphique semble l'un des plus pertinents. Le lien entre le climat et la production de truffes est connu depuis longtemps puisque Chatin (1869) et De Bosredon (1887) le suggéraient déjà. Les études récentes publiées dans des revues internationales (Büntgen et al. 2012; Le Tacon et al. 2014) suggèrent un lien entre le climat (voire le changement climatique) et la production de truffes. Parmi les facteurs importants le bilan hydrique et surtout P-ETP (précipitations moins évapotranspiration) semblent déterminants (Le Tacon et al. 2014). Dans le contexte des changements climatiques, les modèles prévoient des changements importants. Pour compenser ce déficit hydrique l'irrigation est une solution qui a donné des résultats intéressants (Martin and Danoy 1910; Le Tacon et al. 1982). Une autre possibilité serait d'utiliser des truffes adaptées aux conditions de sécheresse, c'est à dire des génotypes capables de fructifier même lorsqu'il existe un déficit hydrique. Il est impossible de dire si de tels génotypes existent mais

en tout cas les ressources génomiques maintenant disponibles permettent de développer de telles recherches.

7.4 Conclusion

Le travail réalisé dans le cadre de ma thèse a permis d'améliorer les connaissances sur un groupe de champignons peu étudiés : les Pézizomycètes. En effet, au début de cette thèse seul le génome de *Tuber melanosporum* était séquencé. Cette étude a permis de mettre en avant une partie des convergences et des différences évolutives entre huit espèces de cette classe de champignons ayant connu une diversification il y a 500 millions d'années.

En utilisant les résultats d'autres études réalisées en parallèle (Kohler et al. 2015), j'ai pu comparer l'évolution de la symbiose mycorhizienne chez les Ascomycètes (truffes) et les Basidiomycètes. Ainsi, j'ai mis en évidence :

- 1) une convergence évolutive entre les truffes et les Basidiomycètes ectomycorhiziens avec une réduction importante des enzymes dégradant la paroi cellulosique des plantes ;
- 2) une divergence évolutive avec l'absence chez les truffes de MiSSP spécifique de *T. melanosporum* et le maintien de la GH6 qui n'existe plus chez les basidiomycètes alors qu'elle est présente chez *C. venosus*, *T. aestivum* et *T. magnatum*.

En parallèle de l'analyse de génomique comparative au sein de la classe des Pézizomycètes, j'ai aussi réalisé une analyse comparative de sept génomes de truffes noires (*T. melanosporum*). J'ai ainsi développé une ressource de plus de 400 000 mutations nucléotidiques ponctuelles de type « *single nucleotide polymorphism* » (SNP) montrant que le niveau de diversité génétique chez *T. melanosporum* est similaire à celui d'autres champignons filamenteux. Enfin, cette étude a permis d'identifier des traces d'adaptation chez la truffe noire.

Mon travail a confirmé que les éléments transposables ont joué un rôle important au cours de l'évolution des génomes chez les Pézizomycètes, L'étude de la dynamique de ces éléments dans le génome de *Tuber melanosporum* a mis en évidence l'existence de plusieurs familles de rétrotransposons *Gypsy* couvrant 30% du génome et pour l'une d'entre elles ayant été active après la spéciation.

Mon travail de thèse a des retombées et des perspectives aussi bien en terme de biologique que de méthodologie. Concernant les perspectives biologiques je peux mettre en avant :

- 1) La ressource de SNPs que j'ai développé sera utilisée pour étudier la structure géographique des populations de *T. melanosporum*. En effet, ces marqueurs pourraient permettre d'identifier l'origine géographique des truffes ce qui est fortement demandé par la filière trufficole dans le but de créer des appellations d'origine géographique.
- 2) Dans ma thèse j'ai identifié des gènes d'intérêt qui sont surexprimés chez *Tuber melanosporum* dans les tissus liés à la fructification et à la symbiose et cela permettra, quand les données seront disponibles de regarder les profils d'expressions des homologues détectées lors de cette thèse chez les autres Pézizomycètes.
- 3) Il va sans dire que la bioinformatique et l'analyse *in silico* de génomes en général et de gènes en particulier, se base sur l'annotation fonctionnelle de ceux-ci. Les annotations sont pour la plupart générés par similarité de séquence et les gènes dit « orphelins » ou n'ayant des homologues que dans des espèces proches annotées de la même façon se retrouvent sans annotation. La seule solution pour contourner ce problème reste la biologie expérimentale. Les difficultés rencontrées dans l'étude en laboratoire de certains champignons comme les truffes peuvent ainsi être contournées en utilisant des espèces modèles comme *Ascobolus immersus*.

- 4) Les mécanismes de défense de ces génomes contre les séquences répétées sont encore mal connus, il sera donc pertinent d'étudier en détail les enzymes impliqués dans ces mécanismes de défenses ainsi que le méthylome de plusieurs espèces comme cela a déjà été initié pour *T. melanosporum* (Montanini et al. 2014).
- 5) Enfin les truffes étant des organismes particulièrement reconnus pour leurs arômes, il me semble important de réaliser une analyse fine des voies métaboliques impliquées dans la production des composés volatiles. De même, il faudrait étudier les voies de biosynthèse des mycotoxines, des allergènes et des toxines pour voir si les champignons comestibles comme les truffes et les morilles présentent des différences avec les champignons saprotrophes.

Mon travail de thèse concernant la génomique comparative chez les Pézizomycètes s'insère dans un consortium international « *Pezizomycete Pan-Genome consortium* ». Les analyses fines des mécanismes de défense contre les séquences répétées et celles portant sur les composés volatiles sont en cours au sein de ce consortium.

D'un point de vue méthodologique, ma thèse m'a permis de développer des outils d'analyses des SNPs dans les gènes codants se basant sur des travaux d'autres personnes du laboratoire (De Mita and Siol 2012). Ce pipeline bioinformatique va pouvoir être utilisé dans des projets de reséquençage qui ont pour certain déjà été initié au sein du laboratoire chez *T. aestivum* et *T. melanosporum* mais aussi chez *Pisolithus tinctorius* et *Tricholoma matsutake*. D'ailleurs, j'ai utilisé ces outils chez la rouille du peuplier *Melanpsora larici-populina* (Persoons et al. 2014; voir annexe).

7.5 Bibliographie

- Bertault, G., M. Raymond, A. Berthomieu, G. Callot, and D. Fernandez. 1998. "Trifling Variation in Truffles." *Nature* 394 (6695): 734–734. doi:10.1038/29428.
- Bertault, Guillaume, François Rousset, Diana Fernandez, Arnaud Berthomieu, Michael E. Hochberg, Gabriel Callot, and Michel Raymond. 2001. "Population Genetics and Dynamics of the Black Truffle in a Man-Made Truffle Field." *Heredity* 86 (4): 451–58. doi:10.1046/j.1365-2540.2001.00855.x.
- Bonito, Gregory, Matthew E Smith, Michael Nowak, Rosanne a Healy, Gonzalo Guevara, Efren Cázares, Akihiko Kinoshita, et al. 2013. "Historical Biogeography and Diversification of Truffles in the Tuberales and Their Newly Identified Southern Hemisphere Sister Lineage." *PloS One* 8 (1): e52765–e52765. doi:10.1371/journal.pone.0052765.
- Büntgen, Ulf, Simon Egli, J. Julio Camarero, Erich M. Fischer, Ulrich Stobbe, Håvard Kauserud, Willy Tegel, Ludger Sproll, and Nils C. Stenseth. 2012. "Drought-Induced Decline in Mediterranean Truffle Harvest." *Nature Climate Change* 2 (12): 827–29. doi:10.1038/nclimate1733.
- Chatin, Adolphe. 1869. *La Truffe. Étude des conditions générales de la production truffière*. AURORÆ LIBRI, Éditeur.
- De Bosredon, Alexandre. 1887. *Manuel Du Trufficulteur: Exposé Complet de La Méthode Pratique Pour L'entretien et La Création Des Truffières, Suivi de La Description Des Principales Variétés de Truffes et de L'histoire Gastronomique et Commerciale de Ce Tubercule*. Imprimerie E. Laporte.
- De Mita, Stéphane, and Mathieu Siol. 2012. "EggLib: Processing, Analysis and Simulation Tools for Population Genetics and Genomics." *BMC Genetics* 13 (1): 27–27. doi:10.1186/1471-2156-13-27.
- Douet, J. P., M. Castroviejo, D. Mabru, G. Chevalier, C. Dupré, F. Bergougnoux, J. M. Ricard, and B. Médina. 2004. "Rapid Molecular Typing of Tuber *Melanosporum*, T. *Brumale* and T. *Indicum* from Tree Seedlings and Canned Truffles." *Analytical and Bioanalytical Chemistry* 379 (4): 668–73. doi:10.1007/s00216-004-2643-9.
- García-Cunchillos, Iván, Sergio Sánchez, Juan José Barriuso, and Ernesto Pérez-Collazos. 2013. "Population Genetics of the Westernmost Distribution of the Glaciations-Surviving Black Truffle Tuber *Melanosporum*." *Mycorrhiza*, November, 1–12. doi:10.1007/s00572-013-0540-9.
- Gnerre, S., I. MacCallum, D. Przybylski, F. J. Ribeiro, J. N. Burton, B. J. Walker, T. Sharpe, et al. 2010. "High-Quality Draft Assemblies of Mammalian Genomes from Massively Parallel Sequence Data." *Proceedings of the National Academy of Sciences* 108 (4): 1513–18. doi:10.1073/pnas.1017351108.
- Hansen, Karen, Brian A. Perry, Andrew W. Dranginis, and Donald H. Pfister. 2013. "A Phylogeny of the Highly Diverse Cup-Fungus Family Pyronemataceae (Pezizomycetes, Ascomycota) Clarifies Relationships and Evolution of Selected Life History Traits." *Molecular Phylogenetics and Evolution* 67 (2): 311–35. doi:10.1016/j.ympev.2013.01.014.
- Hobbie, Erik A., Nancy S. Weber, and James M. Trappe. 2001. "Mycorrhizal vs Saprotrophic Status of Fungi: The Isotopic Evidence." *New Phytologist* 150 (3): 601–10. doi:10.1046/j.1469-8137.2001.00134.x.
- Kohler, Annegret, Alan Kuo, Laszlo G. Nagy, Emmanuelle Morin, Kerrie W. Barry, Francois Buscot, Björn Canbäck, et al. 2015. "Convergent Losses of Decay Mechanisms and Rapid Turnover of Symbiosis Genes in Mycorrhizal Mutualists." *Nature Genetics* advance online publication. doi:10.1038/ng.3223.
- Le Tacon, F., J. Delmas, R. Gleyze, and D. Bouchard. 1982. "Influence Du Regime Hydrique Du Sol et de La Fertilisation Sur La Fructification de La Truffe Noire Du Perigord (Tuber *Melanosporum* Vitt.) Dans Le Sud-Est de La France [Vaucluse; Irrigation, Paillage]." *Acta Oecologia Oecologia Applicata* 3.
- Le Tacon, François, Benoît Marçais, Michel Courvoisier, Claude Murat, Pierre Montpied, and Michel Becker. 2014. "Climatic Variations Explain Annual Fluctuations in French Périgord Black Truffle Wholesale Markets but Do Not Explain the Decrease in Black Truffle Production over the Last 48 Years." *Mycorrhiza* 24 (1): 115–25. doi:10.1007/s00572-014-0568-5.
- Le Tacon, François, Bernd Zeller, Caroline Plain, Christian Hossann, Claude Bréchet, and Christophe Robin. 2013. "Carbon Transfer from the Host to Tuber *Melanosporum* Mycorrhizas and Ascocarps Followed Using a ¹³C Pulse-Labeling Technique." *PLoS ONE* 8 (5): e64626. doi:10.1371/journal.pone.0064626.
- Mabru, D, J. P Douet, A Mouton, C Dupré, J. M Ricard, B Médina, M Castroviejo, and G Chevalier. 2004. "PCR-RFLP Using a SNP on the Mitochondrial Lsu-rDNA as an Easy Method to Differentiate Tuber *Melanosporum* (Perigord Truffle) and Other Truffle Species in Cans." *International Journal of Food Microbiology* 94 (1): 33–42. doi:10.1016/j.ijfoodmicro.2003.12.016.
- Martin, Francis, A. Aerts, D. Ahrén, A. Brun, E. G. J. Danchin, F. Duchaussoy, J. Gibon, et al. 2008. "The Genome of *Laccaria bicolor* Provides Insights into Mycorrhizal Symbiosis." *Nature* 452 (7183): 88–92. doi:10.1038/nature06556.

- Martin, Francis, Annegret Kohler, Claude Murat, Raffaella Balestrini, Pedro M Coutinho, Olivier Jaillon, Barbara Montanini, et al. 2010. "Périgord Black Truffle Genome Uncovers Evolutionary Origins and Mechanisms of Symbiosis." *Nature* 464 (7291): 1033–38. doi:10.1038/nature08867.
- Martin, Joseph, and Henry Danoy. 1910. *La Truffe noire et les truffières rationnelles dans le département de Vaucluse*. F. Seguin.
- Molinier, Virginie, Claude Murat, Emmanuelle Morin, Armelle Gollotte, Daniel Wipf, and Francis Martin. 2013. "First Identification of Polymorphic Microsatellite Markers in the Burgundy Truffle, *Tuber aestivum* (Tuberaceae)." *Applications in Plant Sciences* 1 (2): 1200220. doi:10.3732/apps.1200220.
- Montanini, Barbara, Pao-Yang Chen, Marco Morselli, Artur Jaroszewicz, David Lopez, Francis Martin, Simone Ottonello, and Matteo Pellegrini. 2014. "The Black Truffle Methylome: Non-Exhaustive DNA Methylation-Mediated Transposon Silencing in a Complex Fungal Genome with Massive Repeat Element Content." *Genome Biology* 15 (7): 411. doi:10.1186/PREACCEPT-1025065114129803.
- Murat, C, C Riccioni, B Belfiori, N Cichocki, J Labbé, E Morin, E Tisserant, F Paolucci, A Rubini, and F Martin. 2011. "Distribution and Localization of Microsatellites in the Périgord Black Truffle Genome and Identification of New Molecular Markers." *Fungal Genetics and Biology* 48 (6): 592–601. doi:10.1016/j.fgb.2010.10.007.
- Murat, Claude, Jesús Díez, Patricia Luis, Christine Delaruelle, Chantal Dupré, Gérard Chevalier, Paola Bonfante, and Francis Martin. 2004. "Polymorphism at the Ribosomal DNA ITS and Its Relation to Postglacial Re-Colonization Routes of the Périgord Truffle *Tuber melanosporum*." *New Phytologist* 164 (2): 401–11. doi:10.1111/j.1469-8137.2004.01189.x.
- Murat, Claude, Andrea Rubini, Claudia Riccioni, Herminia De la Varga, Emila Akroume, Beatrice Belfiori, Marco Guaragno, et al. 2013. "Fine-Scale Spatial Genetic Structure of the Black Truffle (*Tuber melanosporum*) Investigated with Neutral Microsatellites and Functional Mating Type Genes." *New Phytologist* 199 (1): 176–87. doi:10.1111/nph.12264.
- Persoons, Antoine, Emmanuelle Morin, Christine Delaruelle, Thibaut Payen, Fabien Halkett, Pascal Frey, Stéphane De Mita, and Sébastien Duplessis. 2014. "Patterns of Genomic Variation in the Poplar Rust Fungus *Melampsora larici-populina* Identify Pathogenesis-Related Factors." *Plant-Microbe Interaction* 5: 450. doi:10.3389/fpls.2014.00450.
- Plett, Jonathan M., Yohann Daguere, Sebastian Wittulsky, Alice Vayssières, Aurelie Deveau, Sarah J. Melton, Annegret Kohler, et al. 2014. "Effector MiSSP7 of the Mutualistic Fungus *Laccaria bicolor* Stabilizes the Populus JAZ6 Protein and Represses Jasmonic Acid (JA) Responsive Genes." *Proceedings of the National Academy of Sciences* 111 (22): 8299–8304. doi:10.1073/pnas.1322671111.
- Plett, Jonathan M., Minna Kempainen, Shiv D. Kale, Annegret Kohler, Valérie Legué, Annick Brun, Brett M. Tyler, Alejandro G. Pardo, and Francis Martin. 2011. "A Secreted Effector Protein of *Laccaria bicolor* Is Required for Symbiosis Development." *Current Biology: CB* 21 (14): 1197–1203. doi:10.1016/j.cub.2011.05.033.
- Raffaele, Sylvain, and Sophien Kamoun. 2012. "Genome Evolution in Filamentous Plant Pathogens: Why Bigger Can Be Better." *Nature Reviews. Microbiology* 10 (6): 417–30. doi:10.1038/nrmicro2790.
- Riccioni, Claudia, Beatrice Belfiori, Andrea Rubini, Valentina Passeri, Sergio Arcioni, and Francesco Paolucci. 2008. "Tuber *Melanosporum* Outcrosses: Analysis of the Genetic Diversity within and among Its Natural Populations under This New Scenario." *New Phytologist* 180 (2): 466–78. doi:10.1111/j.1469-8137.2008.02560.x.
- Rothwell, Gar W., Gene Mapes, Ruth A. Stockey, and Jason Hilton. 2012. "The Seed Cone *Eathiestrobus* Gen. Nov.: Fossil Evidence for a Jurassic Origin of Pinaceae." *American Journal of Botany* 99 (4): 708–20. doi:10.3732/ajb.1100595.
- Tisserant, E., M. Malbreil, A. Kuo, A. Kohler, A. Symeonidi, R. Balestrini, P. Charron, et al. 2013. "Genome of an Arbuscular Mycorrhizal Fungus Provides Insight into the Oldest Plant Symbiosis." *Proceedings of the National Academy of Sciences*, November. doi:10.1073/pnas.1313452110.
- Zeh, David W., Jeanne A. Zeh, and Yoichi Ishida. 2009. "Transposable Elements and an Epigenetic Basis for Punctuated Equilibria." *BioEssays* 31 (7): 715–26. doi:10.1002/bies.200900026.

Liste des publications

- Le Rouzic, Arnaud, **Thibaut Payen**, and Aurélie Hua-Van. 2012. “Reconstructing the Evolutionary History of Transposable Elements.” *Genome Biology and Evolution* 5 (class I): 77–86. doi:10.1093/gbe/evs130.
- Kamoun, Choumouss, **Thibaut Payen**, Aurélie Hua-Van, and Jonathan Filée. 2013. “Improving Prokaryotic Transposable Elements Identification Using a Combination of de Novo and Profile HMM Methods.” *BMC Genomics* 14 (1): 700. doi:10.1186/1471-2164-14-700.
- Murat, Claude, **Thibaut Payen**, Denis Petitpierre, and Jessy Labbé. 2013. “Repeated Elements in Filamentous Fungi with a Focus on Wood-Decay Fungi.” In *The Ecological Genomics of Fungi*, edited by Francis Martin, 21–40. John Wiley & Sons, Inc. (Chapitre 2)
- Labbé, Jessy, Jessie Uehling, **Thibaut Payen**, and Jonathan Plett. 2014. “Fungal Biology: Compiling Genomes and Exploiting Them.” *New Phytologist* 203 (2): 359–61. doi:10.1111/nph.12891. (Annexe 1)
- Payen, Thibaut**, Claude Murat, and Gregory Bonito. 2014. “Chapter Seven - Truffle Phylogenomics: New Insights into Truffle Evolution and Truffle Life Cycle.” In *Advances in Botanical Research*, edited by Francis M. Martin, Volume 70:211–34. Fungi. Academic Press. (Chapitre 3)
- Persoons, Antoine, Emmanuelle Morin, Christine Delaruelle, **Thibaut Payen**, Fabien Halkett, Pascal Frey, Stéphane De Mita, and Sébastien Duplessis. 2014. “Patterns of Genomic Variation in the Poplar Rust Fungus *Melampsora Larici-Populina* Identify Pathogenesis-Related Factors.” *Plant-Microbe Interaction* 5: 450. doi:10.3389/fpls.2014.00450. (Annexe 2)
- Payen, Thibaut**, Claude Murat, Anaïs Gigant, Emmanuelle Morin, Stéphane De Mita, and Francis Martin. 2015. “A Survey of Genome-Wide Single Nucleotide Polymorphisms through Genome Re-Sequencing in the Périgord Black Truffle (*Tuber Melanosporum Vittad.*)” *Molecular Ecology Resources*, February, n/a–n/a. doi:10.1111/1755-0998.12391. (Chapitre 5)

Liste de participation à des congrès internationaux

- Journées Ouvertes Biologie Informatique et Mathématiques, 2012, Rennes, France (présentation d'un poster)
- Journées Francophones des Mycorhizes, 2012, Nancy, France (Présentation orale)
- First International Congress on Trufficulture, Teruel, Espagne (Présentation orale)
- Comparative Genomics of Eukaryotic Microorganisms: Complexity Patterns in Eukaryotic Genomes, 2013, Sant Feliu de Guixols, Espagne (Présentation d'un poster)
- Third Annual Consortium Meeting of EcoFINDERS, 2014, Manchester, Royaume-Uni (Présentation d'un poster)
- 12th European Conference on Fungal Genetics, 2014, Séville, Espagne (Présentation d'un poster)
- First Pezizomycete pan-genomic workshop, 2015, Nancy, France (Présentation orale)

Meetings

Fungal biology: compiling genomes and exploiting them

12th European Conference on Fungal Genetics, Seville, Spain, March 2014

The 12th European Conference on Fungal Genetics (ECFG12) took place in March 2014. Every 2 years a European country welcomes this meeting, which is held in coordination with the Fungal Genetics Conferences that take place every 2 years in Asilomar (USA). This year over 700 participants from 40 countries gathered in Seville (Spain) to exchange ideas on the central theme of fungal genetics and general fungal biology including molecular and cell biology, genomics, evolution and biotechnology. The conference included plenary sessions in honor of three prominent fungal geneticists (Charles Yanofsky, John Clutterbuck and Claudio Scazzocchio), concurrent sessions such as the tremendous 'Fungal genomes: now what?', and six satellite meetings including the 4th Mycorrhizal Genomics Initiative Workshop (MGIW4). In the latter meeting, coordinated by Francis Martin (INRA, Nancy, France), approx. 30 junior and senior scientists discussed progress made on exploring the genome diversity of mycorrhizal fungi and debated future directions on how to use the current data sets to bridge mycorrhizal genomics, metagenomics and forest ecology. In this meeting report, we focus on the engaging discussions surrounding fungal genomes and their utilization.

'... the evolution of effectors is a key aspect to both speciation within symbiotic fungi as well as a determinant in host plant specificity.'

Fungal genomics moving forward

The last 10 years have seen the cost of sequencing complete genomes decrease at an incredible speed. This has led to an increase in the number of genomes sequenced in all the fungal tree of life as well as a wide variety of plant genomes. The increase in sequencing has permitted us to study the evolution of organisms on a genomic scale. A number of talks during the conference discussed the importance of transposable elements (TEs) that are present in almost all species of fungi. These TEs represent an especially large

percentage of genomic space in fungi that interact with plants. Thierry Rouxel (INRA, Nancy, France) showed the link between speciation in the *Leptosphaeria* complex and the expansion of TE families. For example in the *Leptosphaeria* complex, one species associated with oilseed rape has experienced a recent and massive burst of movement by a few TE families. The alterations caused by these TEs took place in discrete regions of the genome leading to shuffling of the genomic landscape and the appearance of genes specific to the species, such as effectors useful for the interactions with a particular plant (Rouxel *et al.*, 2011). Other presentations showed the importance of TEs in affecting genome organization. For example, in *Amanita* different species appear to have been invaded by different TE families (Veneault-Fourrey & Martin, 2011).

There were also many engaging oral presentations on the potential applications of the genomic data now available from multiple sources. These data enable inferences of how the genomic content and structure of fungi can impact interactions with, and response to, their hosts and environments. One particularly successful strategy has been using the newly available genes from phylogenomic studies to ask questions about how closely related fungal lineages developed different ecological strategies. Joey Spatafora (Oregon State University, Corvallis, OR, USA) discussed work on homologous nonribosomal peptide synthetase (NRPS) clusters in the Hypocreales. His team found that duplications, rearrangements, and fusions of conserved adenylation domains in derived members of close pathogenic relatives mirror host switches between plants, fungi, and insects. This work raises questions about the mechanisms involved in genome expansion, contraction, and rearrangement. It also raises the question concerning which of these processes most strongly influence shifts in host interactions and other changes in life strategies (Veneault-Fourrey & Martin, 2011). Francis Martin summarized analyses of several mycorrhizal fungal genomes from the Mycorrhizal Genomics Initiative (<http://mycor.nancy.inra.fr/IMGC/Mycogenomes/>), which is an ongoing multi-institution joint program with the JGI focused on sequencing mycorrhizal fungal species (Martin *et al.*, 2011). Large emerging patterns include losses in lignocellulolytic decomposition capabilities in several groups of ectomycorrhizal fungi compared to their saprotrophic ancestors. Interestingly, ericoid and orchid mycorrhiza retain the potential to decay most components of secondary plant cell walls. As more fungal genomes are sequenced through this, and other, initiatives it will be interesting to evaluate commonalities of genomic content and structure in addition to determining which aspects of fungal biology underwent convergent evolution in order to produce the mycorrhizal habit in phylogenetically distinct fungal lineages. Hanna Johannesson (Uppsala University, Sweden) and colleagues presented data comparing rates and locations of selection and recombination in nearly 100 *Neurospora tetrasperma* genomes. These data suggest population

level sweeps of entire mating-type chromosomes driven by introgression and ultimately fixations are detectable in some lineages (Nygren *et al.*, 2011). It has been demonstrated that variants in multi-allelic genes can selectively sweep through populations of many diverse organisms, and evidence for larger scale selective sweeps has recently emerged in several model organisms. The frequency of such sweeps and implications for the evolutionary trajectories of fungal groups with this capability are an exciting avenue for future research. Michael Brunner (Heidelberg Biochemistry Center, Germany) spoke about dynamics of light induced transcriptional re-profiling in *Neurospora*. He presented data suggesting that chromatin re-modification via interaction of GATA type transcription factor SUB1 with a putative acetyltransferase may facilitate access of the transcription factor White Collar Complex (WCC) to promoters of genes involved in the response. The epigenetic control of functional genes has been demonstrated in several other fungal groups, and in some instances appears to play an important role in symbiotic relationship establishment and maintenance (Soyer *et al.*, 2014). Future investigations into the extent of epigenetic control of fungal response to stimuli from the environment and other organisms will be informative into the question of how fungi regulate and strategically use their genomic contents. Jason Stajich (University of California, Riverside, CA, USA) discussed recent work on the evolution of multi-cellularity in fungi by comparing several genomes in lineages of Ascomycetes and Basidiomycetes that convergently gained multi-cellularity. Comparative analyses of genome content and organization revealed increases in copy number of signaling, perception, and cell wall gene families among others are likely involved in gain of multi-cellularity. The future sequencing of many genomes that falls along a spectrum of physical complexity will allow a more thorough analysis of the question, does complex trait gain most often result from genome content gain and loss, organization, or higher-level control or some combination of the three?

Deciphering the interactive toolbox one effector at a time

The conference was a stirring update on the progress in our understanding of effector biology in both pathogenic and mutualistic fungi. Effectors are small secreted fungal proteins that are used during the symbiotic interaction with plants to alter the physical or chemical environment of plant tissues to favor colonization. Sophien Kamoun (The Sainsbury Laboratory, Norwich, UK) took us back into the evolutionary past of the *Phytophthora* genus with his group's recent work with the EpiC1 effector (Dong *et al.*, 2014). They have shown that this effector evolved specific activities after a host jump and, through phylogenetic analysis, were able to predict the ancestral version of this effector and demonstrate how its activity had changed. This work has demonstrated that the evolution of effectors is a key aspect to both speciation within symbiotic fungi as well as a determinant in host plant specificity. This work, leaves the intriguing question: how many effectors are required before a fungus can colonize a new host?

Within the effector community one aspect of their biology has hampered the *in silico* annotation of their functions: very rarely do effectors encode conserved domains nor do they exhibit close homology with other proteins at the nucleotide level. Despite these differences, as more effectors are characterized, the more we find that they have similar biological activities or consequences. Alga Zuccaro's group (Max Planck Institute, Marburg, Germany) working with the effectors of *Piriformospora indica* is tackling just this problem. By taking effectors with no known homologues and by doing the tricky work of crystallization, her group is determining the three-dimensional (3D) structure of these effectors. And the results have been worth the effort. New data coming from Alga's group is now showing that there are conserved motifs between fungal effectors and other known proteins – even proteins from bacteria! These results would indicate that the next challenge of this field is to look to the arduous task of understanding the biology of our favorite effectors in 3D.

Back in the pathogenic camp, Shigeyuki Tanaka (Max Planck Institute, Marburg, Germany) gave a great talk concerning Tin2, one of the virulence promoting effectors of *Ustilago* (Tanaka *et al.*, 2014). In an interesting twist, this effector targets, and promotes, the production of anthocyanins in the host plant. While, at first, this seems like a rather wandering path toward assuring virulence, the metabolic reasoning behind this effect is brilliant. In promoting anthocyanin synthesis in infected leaves, the fungus is forcing the host plant to divert maximal levels of the precursor p-Coumaric acid toward a benign end-point and away from the lignin pathway. In depleting the host of lignin, *Ustilago* is able to form infection structures and hijack plant cellular functioning, a circumstance that would not happen if the host were able to lignify the infected tissues. This case study is very pertinent in highlighting the fact that effectors may not directly target problematic host pathways that would hurt their chances of colonization. Rather they may act upon a parallel pathway to indirectly aid the take-over of plant functioning. Thus, the only way to fully understand these complex problems is to have an intimate knowledge of the host biology as well as the fungal symbiont.

Typically in our areas of specialization, we choose to focus on the effectors of either pathogenic fungi or mutualistic fungi with very little crossover. With so many effectors to characterize this is a rational approach. If, however, we are to understand the full biological relevance of these intriguing proteins we must begin to cross disciplines and begin to work on a more unified vision of effector biology. Natalia Requena's presentation addressed just this point. In her group (Karlsruhe Institute of Technology, Karlsruhe, Germany) they are beginning to work on both pathogens and mutualists to begin the process of identifying common host proteins targeted by effectors from both lifestyles. While this work is still in its infancy, we await with great anticipation this next step in the study of these proteins. Will effectors from both lifestyles affect the same host proteins? If such host 'hub' proteins exist, will mutualistic and pathogenic effectors control their target in the same way or in opposing fashions? We anticipate the answer to these questions and so many more in 2 years in Paris at the 13th ECFG.

Acknowledgements

Many thanks to Krista Plett and Francis Martin for helpful comments on the manuscript. Thanks also to all who attended and gave freely of their data, thoughts and opinions. The authors acknowledge the Genomic Science Program (project 'Plant-Microbe Interactions'), US Department of Energy, Office of Science, Biological and Environmental Research, for supporting the authors' participation to this meeting, under the contract DE-AC05-00OR22725.

Jessy Labbé^{1*}, Jessie Uehling², Thibaut Payen³ and Jonathan Plett⁴

¹Biosciences Division, Oak Ridge National Laboratory, PO Box 2008, Oak Ridge, TN 37831-6422, USA;

²Biology Department, Duke University, Durham, NC 27708, USA;

³UMR 1136, INRA/Lorraine University, 'Tree-Microbe Interactions', Lab of Excellence ARBRE, INRA, Nancy 54280, Champenoux, France;

⁴Hawkesbury Institute for the Environment, University of Western Sydney, Richmond, NSW, Australia

(*Author for correspondence: tel +1 865 576 3478; email labbejj@ornl.gov)

References

- Dong S, Stam R, Cano LM, Song J, Sklenar J, Yoshida K, Bozkurt TO, Olivia R, Liu Z, Tian M *et al.* 2014. Effector specialization in a lineage of the Irish potato famine pathogen. *Science* 343: 552–555.
- Martin F, Cullen D, Hibbett D, Pisabarro A, Spatafora JW, Baker SE, Grigoriou IV. 2011. Sequencing the fungal tree of life. *New Phytologist* 190: 818–821.
- Nygren K, Strandberg R, Wallberg A, Nabholz B, Gustafsson T, García D, Cano J, Guarro J, Johannesson H. 2011. A comprehensive phylogeny of *Neurospora* reveals a link between reproductive mode and molecular evolution in fungi. *Molecular Phylogenetics and Evolution* 59: 649–663.
- Rouxel T, Grandaubert J, Hane JK, Hoede C, van de Wouw AP, Couloux A, Dominguez V, Anthouard V, Bally P, Bourras S *et al.* 2011. Effector diversification within compartments of the *Leptosphaeria maculans* genome affected by Repeat-Induced Point mutations. *Nature Communications* 2: 202.
- Soyer JL, El Ghalid M, Glaser N, Ollivier B, Linglin J, Grandaubert J, Balesdent MH, Connolly LR, Freitag M, Rouxel T, Fuda I. 2014. Epigenetic control of effector gene expression in the plant pathogenic fungus *Leptosphaeria maculans*. *PLoS Genetics* 10: e1004227.
- Tanaka S, Brefort T, Neidig N, Djamei A, Kahnt J, Vermerris W, Koenig S, Feussner K, Feussner I, Kahmann R. 2014. A secreted *Ustilago maydis* effector promotes virulence by targeting anthocyanin biosynthesis in maize. *eLife* 3: e01355.
- Veneault-Fourrey C, Martin F. 2011. Mutualistic interactions on a knife-edge between saprotrophy and pathogenesis. *Current Opinion in Plant Biology* 14: 444–450.

Key words: effectors, fungal evolution, fungal genetics and genomics, mycorrhizal fungi, pathogenic and mutualistic interactions, saprotrophs, symbiosis.

New Phytologist

Symposia for 2014

www.newphytologist.org/symposia

33rd New Phytologist Symposium

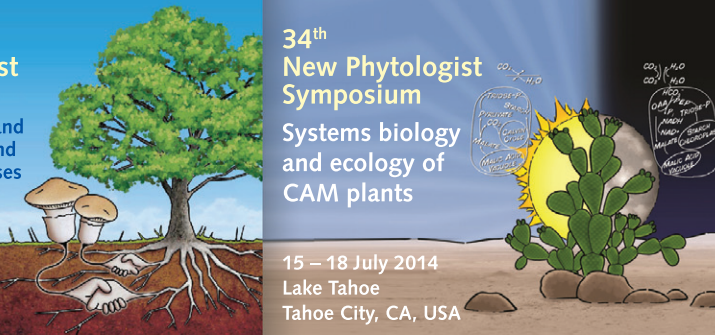
Networks of Power and Influence: ecology and evolution of symbioses between plants and mycorrhizal fungi


14 – 16 May 2014
Agroscope, Zurich, Switzerland

34th New Phytologist Symposium

Systems biology and ecology of CAM plants

15 – 18 July 2014
Lake Tahoe
Tahoe City, CA, USA





A special event for researchers in the early stages of their careers

New Phytologist
next generation scientists

29 – 30 July 2014
John Innes Conference Centre, Norwich UK



Patterns of genomic variation in the poplar rust fungus *Melampsora larici-populina* identify pathogenesis-related factors

Antoine Persoons^{1,2}, Emmanuelle Morin^{1,2}, Christine Delaruelle^{1,2}, Thibaut Payen^{1,2}, Fabien Halkett^{1,2}, Pascal Frey^{1,2}, Stéphane De Mita^{1,2} and Sébastien Duplessis^{1,2*}

¹ Institut National de la Recherche Agronomique, Unité Mixte de Recherche 1136 Institut National de la Recherche Agronomique/Université de Lorraine Interactions Arbres/Microorganismes, Champenoux, France

² Université de Lorraine, Unité Mixte de Recherche 1136 Institut National de la Recherche Agronomique/Université de Lorraine Interactions Arbres/Microorganismes, Vandoeuvre-lès-Nancy Cedex, France

Edited by:

Francine Govers, Wageningen University, Netherlands

Reviewed by:

John P. Rathjen, The Australian National University, Australia
Dario Cantu, University of California, Davis, USA

*Correspondence:

Sébastien Duplessis, INRA, Unité Mixte de Recherche 1136
INRA/Université de Lorraine Interactions Arbres/Microorganismes,
54280 Champenoux, France
e-mail: duplessis@nancy.inra.fr

Melampsora larici-populina is a fungal pathogen responsible for foliar rust disease on poplar trees, which causes damage to forest plantations worldwide, particularly in Northern Europe. The reference genome of the isolate 98AG31 was previously sequenced using a whole genome shotgun strategy, revealing a large genome of 101 megabases containing 16,399 predicted genes, which included secreted protein genes representing poplar rust candidate effectors. In the present study, the genomes of 15 isolates collected over the past 20 years throughout the French territory, representing distinct virulence profiles, were characterized by massively parallel sequencing to assess genetic variation in the poplar rust fungus. Comparison to the reference genome revealed striking structural variations. Analysis of coverage and sequencing depth identified large missing regions between isolates related to the mating type loci. More than 611,824 single-nucleotide polymorphism (SNP) positions were uncovered overall, indicating a remarkable level of polymorphism. Based on the accumulation of non-synonymous substitutions in coding sequences and the relative frequencies of synonymous and non-synonymous polymorphisms (i.e., P_N/P_S), we identify candidate genes that may be involved in fungal pathogenesis. Correlation between non-synonymous SNPs in genes encoding secreted proteins (SPs) and pathotypes of the studied isolates revealed candidate genes potentially related to virulences 1, 6, and 8 of the poplar rust fungus.

Keywords: effector, virulence, Pucciniales, obligate biotroph, genomics, polymorphism

INTRODUCTION

Worldwide, *Melampsora* spp. (Basidiomycota, Pucciniales) are the most devastating pathogens of poplars (Steenackers et al., 1996), and *Melampsora larici-populina* is a major threat in European poplar plantations (Pinon and Frey, 2005). The poplar rust fungus has a complex life cycle with five different types of spores that develop on two distinct host plants: *Populus*, on which it performs several asexual reproduction cycles during summer and autumn, and *Larix* spp., on which it performs a single sexual reproduction cycle once a year in spring. Poplars are particularly susceptible to *M. larici-populina* mostly because of their intensive monoclonal cultivation over several decades (Gérard et al., 2006). Until now eight qualitative resistances ($R1$ to $R8$) have been deployed in plantations and each has been overcome by *M. larici-populina*. The most damaging resistance breakdown occurred in 1994 when the resistance $R7$ was overcome and led to the invasion of France by virulent 7 *M. larici-populina* isolates (Xhaard et al., 2011). In accordance with the gene-for-gene relationship (Flor, 1971), *M. larici-populina* isolates which successfully infect resistant poplar possess the corresponding virulence factors (i.e., $vir1$ to $vir8$) determined at an avirulence

locus. Up to now, none of the poplar R genes, nor the poplar rust virulence genes have been characterized (Hacquard et al., 2011).

Pathogenicity factors, i.e., effectors, contribute to the success of pathogen infection. Their recognition by cytoplasmic plant R receptors leads to a rapid and strong defense reaction through specific signaling cascades and expression of defense-related genes that stop pathogen growth, notably through the expression of a localized hypersensitive response at infection site (Dodds and Rathjen, 2010; Win et al., 2012). Most effectors described to date in rust fungi correspond to avirulence factors such as AvrL567, AvrP4, AvrP123, and AvrM of the flax rust fungus *Melampsora lini* (Ravensdale et al., 2011) and PGTAUSPE-10-1, a candidate AvrSr22 factor of the wheat stem rust *Puccinia graminis* f. sp. *tritici* (Upadhyaya et al., 2014), but their role in pathogenesis remain unknown. Another effector, the Rust Transferred Protein 1 (RTP1) from the bean rust fungus *Uromyces fabae*, forms fibrils in the extrahaustorial matrix and is transferred from haustoria into infected host cells, and may have protease inhibitory function (Kemen et al., 2005, 2013; Pretsch et al., 2013). So far, only a handful of fungal candidate effectors have been fully characterized

(Stergiopoulos and de Wit, 2009; Tyler and Rouxel, 2012; Giraldo and Valent, 2013). Fungal effectors share several features, which are not exclusive, i.e., most have a N-terminal secretion signal, enrichment in cysteine residues and a lack of functional homology in databases and present a small size. Such features have been widely used to determine sets of candidate effectors in the predicted proteome of fungal pathogens for which a reference genome has been sequenced (Lowe and Howlett, 2012; Duplessis et al., 2014a).

Host immunity escape by pathogens is frequently mediated by deletion or mutations in effector genes, which often show elevated levels of non-synonymous polymorphism as a result of their antagonistic co-evolution with the host (Stukenbrock and McDonald, 2009). The relative abundance of non-synonymous and synonymous polymorphisms (P_N and P_S) measures the direct effect of positive selection that tends to remove deleterious non-synonymous variants in coding sequences. When considered at the interspecific level, the rates of non-synonymous and synonymous substitutions (termed dN and dS, respectively) can be assessed to contrast patterns of variation between species (Stukenbrock and Bataillon, 2012). Such approaches have been applied at the genome scale to detect sets of candidate effectors in oomycetes and fungi (Raffaele and Kamoun, 2012; Cantu et al., 2013; Stergiopoulos et al., 2013; Stukenbrock, 2013). Evidence of positive selection was reported in avirulence genes of rust fungi at the intraspecific (AvrL567, Dodds et al., 2004; AvrP4 and AvrP123, Barrett et al., 2009) or interspecific levels (AvrP4, Van der Merwe et al., 2009). Genome-scale approaches were also used with sets of candidate effectors at the intraspecific level in *Puccinia striiformis* f. sp. *tritici* (Cantu et al., 2013) or by considering clusters of paralogous genes (CPG) in the genome of *M. larici-populina* (Hacquard et al., 2012).

Genomics is becoming a method of choice to identify new candidate effectors, particularly in obligate biotrophs where functional approaches are impeded. Only a handful of rust fungi genomes are available (Cantu et al., 2011, 2013; Duplessis et al., 2011a; Zheng et al., 2013; Nemri et al., 2014). In these, repertoires of candidate effectors corresponding to small secreted proteins (SSPs) have been defined (Hacquard et al., 2012; Saunders et al., 2012; Cantu et al., 2013; Zheng et al., 2013; Nemri et al., 2014). The poplar-poplar rust pathosystem is a model in forest pathology because it is one of the few pathosystems for which both the host and pathogen genomes are available (Tuskan et al., 2006; Duplessis et al., 2011a). *M. larici-populina* has a remarkably large diploid genome of 101 Mb enriched in repetitive and transposable elements (TE), a common feature of rust fungi genomes. There is a striking number of 16,399 predicted genes in the poplar rust genome, another feature shared with other rust fungi (Duplessis et al., 2014b). Among genes encoding secreted proteins (SPs), a set of 1184 SSP genes showing typical features of pathogen effectors was uncovered; most of these are cysteine-rich, belong to multigene families and are lineage specific (Duplessis et al., 2011a; Hacquard et al., 2012). In order to prioritize functional analysis of such candidates, other features were searched including specific expression during the interaction with the poplar host (Duplessis et al., 2011b), presence of conserved motifs in proteins, and gene families exhibiting evidences of positive selection by considering

a classification into CPG (Joly et al., 2010; Hacquard et al., 2012). Another way to identify promising effectors is to study gene polymorphism at the intraspecific or interspecific level, as has been performed in *M. lini* (Ravensdale et al., 2011).

In the present study, we report on the genome sequencing of 15 *M. larici-populina* isolates and their comparison to the reference genome of isolate 98AG31 (Duplessis et al., 2011a) in order to identify patterns of genomic variations that may relate to fungal pathogenesis. Genes that accumulate intraspecific polymorphism in their coding sequence as well as in their non-coding upstream regions were scrutinized, thus providing a new filter to prioritize candidate effectors of interest.

MATERIALS AND METHODS

FUNGAL MATERIAL

Isolates were selected in a laboratory collection (Frey P., INRA Nancy, Champenoux, France) in order to maximize historical and geographical repartitions and virulence profiles (Table 1). Phenotypes of all isolates (i.e., combination of virulences) were confirmed in triplicate on eight poplar cultivars each carrying a single resistance (R1 to R8) to *M. larici-populina* (Table 1) and on the universal clone 'Robusta', as a positive control. To ensure their purity and to avoid potential clones within the selected isolates, genotyping was performed using 25 microsatellite markers (Xhaard et al., 2011). Urediniospores of each isolates were multiplied on 'Robusta' detached leaves to obtain enough material for genomic DNA isolation.

DNA ISOLATION

A total of 100–300 mg of urediniospores were used for DNA isolation using a CTAB method. Spores were crushed using a Retsch Tissue Lyser (Qiagen, Courtaboeuf, France) at a frequency of 30 Hz for 1 min. Broken spores were resuspended in CTAB buffer (Tris 0.1 M, NaCl 1.43 M, EDTA 0.02 M, CTAB 0.02 M) and heated at 65°C for 30 min. The suspension was subjected to centrifugation at 8000 rpm at room temperature for 5 min to pellet spore debris. Supernatant was gently mixed with an equal volume of phenol:chloroform:isoamyl alcohol (50:48:2; Euromedex, Souffelweyersheim, France) and centrifuged at 8000 rpm at room temperature for 10 min. The aqueous phase was recovered, gently mixed with an equal volume of chloroform and centrifuged at 8000 rpm at room temperature for 10 min. The aqueous phase was subjected to RNA digestion with RNaseA at 10 μ M (Fermentas, Saint-Remy-lès-chevreuses, France) at 37°C for 30 min. A final extraction with an equal volume of chloroform was realized followed by centrifugation at 8000 rpm at room temperature for 10 min. The recovered aqueous phase was then subjected to isopropanol (0.75 of final volume) precipitation, followed by centrifugation at 14,000 rpm at 4°C for 30 min. DNA pellet was washed twice with 70%, then absolute ethanol, each followed by centrifugation at 14,000 rpm at 4°C for 10 min. The DNA pellet was finally dried under a hood for 20 min and resuspended in 1X Tris EDTA. Quality and quantity of recovered high molecular weight DNA was assessed by electrophoresis on agarose gel, by spectrophotometry (Nanodrop, Saint-Remy-lès-Chevreuse, France) and with the QuBit (Life Technologie, Villebon-sur-Yvette, France) fluorometric quantitation system.

Table 1 | Summary of *Melampsora larici-populina* isolates.

Isolate	Year	Location	Latitude, Longitude	Host	Pathotype
93ID6	1993	Champenoux (NE France)	N 48° 45' 02", E 06° 20' 20"	<i>P. x euramericana</i> 'I45-51'	3-4
02Y5	2002	Charrey-sur-Saône (NE France)	N 47° 05' 18", E 05° 09' 11"	<i>P. x euramericana</i> 'Robusta'	2-3-4-7-8
09BS12	2009	Mirabeau (SE France)	N 43° 41' 29", E 05° 40' 21"	<i>P. nigra</i>	4-6
94ZZ15	1994	Saulchoy (N France)	N 50° 21', E 01° 50'	<i>P. x euramericana</i> 'Luisa Avanzo'	3-4-5-7
94ZZ20	1994	Nogent-sur-Vernisson (Central France)	N 47° 50', E 02° 45'	<i>P. x interamericana</i> 'Boelare'	3-4-7
08EA47	2008	Prelles (SE France)	N 44° 51' 00", E 06° 34' 47"	<i>P. nigra</i>	2-4
95XD10	1995	Rogécourt (N France)	N 49° 39', E 03° 25'	<i>P. x euramericana</i> 'Flevo'	3-4-5-7
08EA20	2008	Prelles (SE France)	N 44° 51' 00", E 06° 34' 47"	<i>P. nigra</i>	4
08EA77	2008	Prelles (SE France)	N 44° 51' 00", E 06° 34' 47"	<i>P. nigra</i>	4-6
97CF1	1997	Champenoux (NE France)	N 48° 45' 02", E 06° 20' 20"	<i>P. x interamericana</i> 'Hoogvorst'	3-4-7
08KE26	2008	Mirabeau (SE France)	N 43° 41' 29", E 05° 40' 21"	<i>P. nigra</i>	4
9683B13	1996	Orléans (Central France)	N 47° 49' 39", E 01° 54' 40"	<i>P. x interamericana</i> '83B13'	1-3-4-5-6-7
98AG31	1998	Moy-de-l'Aisne (N France)	N 49° 45', E 03° 21'	<i>P. x interamericana</i> 'Beaupré'	3-4-7
93JE3	1993	Champenoux (NE France)	N 48° 45' 02", E 06° 20' 20"	<i>P. x euramericana</i> 'Blanc du Poitou'	2-4
98AR1	1998	Geraardsbergen (Flanders, Belgium)	N 50° 45', E 03° 52'	<i>P. x interamericana</i> 'B71085/A1'	1-3-4-5-7-8

Isolate name, year, and location of sampling are indicated. Host indicates the poplar species/cultivar on which the isolate was sampled. The pathotype profile (combination of virulences) was confirmed in triplicate by inoculation on a differential set of poplar cultivars carrying the eight known resistances to *M. larici-populina*.

GENOME RE-SEQUENCING

For all isolates, except 98AR1, genomic DNA libraries were prepared using TruSeq DNA sample preparation kit (v3) followed by paired-end 100 nt massively parallel sequencing on Illumina HiSeq2000 by Integragen (Evry, France). Briefly, 3 µg of each genomic DNA were fragmented by sonication and purified to yield fragments of 400–500 nt. Paired-end adapter oligonucleotides from Illumina were ligated on repaired A-tailed DNA fragments, then purified and enriched by PCR cycles. Each library was quantified by qPCR and sequenced on Illumina HiSeq2000 platform as paired-end 100 nt reads. Image analysis and base calling were performed using Illumina Real Time Analysis (RTA 1.13.48.0) pipeline with default parameters. Isolate 98AR1 genomic DNA was sequenced by a single read strategy of 75 bases on Illumina Genome Analyzer II (Beckman Coulter Genomics, Grenoble, France).

FILTERING AND MAPPING OF SHORT READS

Adapter and quality filtering was carried out using CLC Genomics Workbench 6.5 (CLC bio, QIAgen, Aarhus, Denmark). For each batch of reads, 3 and 10 low quality terminal nucleotides were trimmed at the 5' and 3' ends, respectively. FASTQ files of trimmed sequences were used to proceed with mapping onto the 98AG31 reference genome available at the Joint Genome Institute (JGI; <http://genome.jgi.doe.gov/programs/fungi/index.jsf>; Duplessis et al., 2011a). The 462 scaffolds composing the reference genome were uploaded in CLC Genomics Workbench and the annotation was superimposed onto the scaffolds using the annotation plugin. The following parameters were applied for mapping: masking mode = no masking; mismatch cost = 2; insertion cost = 3; deletion cost = 3; length fraction = 1.0; similarity fraction = 0.95; global alignment = no; auto-detect paired distances = yes; non-specific match handling = map randomly. Sequencing data and assemblies were deposited at the National Center for Biotechnology Information (NCBI) and

the Short Reads Archive (Bioproject PRJNA251864 study SRA accession SRP042998). Coverage and sequencing depth values were extracted from the CLC stand-alone read mapping files and were further used to compare scaffolds of resequenced isolates. Sequencing depth and coverage on each scaffold were visually inspected using the CLC read tracks functions used for further detection of structural variants.

SCAFFOLD DEPTH ANALYSIS AND VARIANTS DETECTION

Cross-comparison of average coverage and sequencing depth onto the 462 reference scaffolds was performed within and between isolates based on the CLC Genomics Workbench mapping outputs to detect the potential presence/absence of regions and the sequencing coverage or depth bias. In the case of missing regions or coverage bias, read mapping profiles and distribution of genes and TEs on the scaffolds were inspected manually. In these manual inspections, regions with high concentrations of ambiguous mappings were excluded from consideration, because of the possibility of artifactually divergent coverage. In parallel, the coverage analysis tool implemented in CLC Genomics Workbench (version 7.0) was used to detect regions within scaffolds showing significantly unexpected low or high coverage relative to the reference genome, according to a Poisson distribution of observed coverage in mapping positions (p -value threshold = 0.0001 and minimum length of the coverage region of 100 bp). Search for SP genes in the low-coverage regions was performed using an in-house Python script. Notably, this script was limited to detection of genes which laid entirely inside the corresponding region.

Single Nucleotide Variants (SNVs, i.e., Single Nucleotide Polymorphisms, SNPs), Multiple Nucleotide Variants (MNVs, i.e., successive SNVs), and small Insertion/Deletion variants (i.e., InDels) were detected in the genome of each isolate based on mapping outputs using the quality-based variant detection option of CLC Genomics Workbench (version 6.5.1). This option

considers minimum quality levels and minimum coverage of bases where the variant is detected and in surrounding bases. The following parameters were considered: neighborhood radius = 5; maximum gap and mismatch count = 2; minimum neighborhood quality = 15; minimum central quality = 20; ignore non-specific matches = yes; ignore broken pairs = yes; minimum coverage = 10; minimum variant frequency = 35%; maximum expected alleles = 2; advanced = no; require presence in both forward and reverse reads = no; ignore variants in non-specific regions = no; genetic code = standard. Variant tables were generated for all isolates. Selection of synonymous and non-synonymous polymorphism in genes and variants in 1 Kb upstream regions of genes was performed using in-house Python scripts.

SEQUENCE ANALYSIS

Gene and protein sequences and Gene Ontology (GO) and Eukaryotic Orthologous Group (KOG) functional annotations were retrieved from the *M. larici-populina* genome sequence on the MycoCosm website at the JGI (<http://genome.jgi.doe.gov/programs/fungi/index.jsf>). Homology searches were carried out using the Blastp algorithm (Altschul et al., 1997) against the non-redundant database at the NCBI (March 2014). AvrP4 sequences from Van der Merwe et al. (2009) and Barrett et al. (2009) were retrieved from the NCBI and used for multiple alignments with members of the CPG5464 gene family previously identified in the *M. larici-populina* genome (Hacquard et al., 2012). Alignment with variants of the CPG5464 gene family retrieved in the *M. larici-populina* isolates was conducted using the program ClustalW (Thompson et al., 2002) and gaps were manually inserted to strictly align sites reported under positive selection in the above-mentioned articles, before generating conservation profiles on the WebLogo server (Crooks et al., 2004).

KOG ENRICHMENT ANALYSIS

KOG (Tatusov et al., 2003) annotation of each *M. larici-populina* gene was retrieved from the JGI genome website. Each gene was classified according to the KOG functional classification using custom Perl scripts. Over-represented KOG categories in a selected gene set were calculated relative to the global gene distribution in the genome. Fisher's exact test was used to determine significant differences in the distribution of genes by KOG categories between the selected gene set and all genes ($p < 0.05$).

P_N/P_S ANALYSIS

For each gene, an alignment was generated with a custom Python script based on the reference genome and gene annotations (gff files from the *M. larici-populina* JGI website) taking into account the SNP variants generated by CLC Genomics Workbench. Alignments interrupted by an early stop codon were excluded from the computation of synonymous and non-synonymous polymorphisms. Polymorphism index was computed for each gene using Egglip version 2.1.6 (De Mita and Siol, 2012). This Python library computes from an alignment the number of synonymous or non-synonymous sites either polymorphic or non-polymorphic. P_N/P_S is computed as the ratio of the number of synonymous over non-synonymous polymorphisms corrected

by the number of synonymous and non-synonymous sites, respectively.

RESULTS

SEQUENCING EFFICIENCY

Genomes of 15 *M. larici-populina* isolates, including the 98AG31 reference isolate, were sequenced at a targeted depth of ~40X. A total of 64 billion bases were generated, corresponding to 2.5–6.2 billion reads per genome. After filtering, the average read length was 84.4 nt. A number of length and similarity parameters were tested for mapping reads onto the reference genome. Loose default parameters tended to generate multiple mappings in repetitive sequences including large gene families, impinging on further call of variants in a given isolate (data not shown). Stringent parameters were retained (i.e., total length of the sequence showing a minimum of 95% similarity) for optimal mapping and subsequent variant calling. On average, 78% of the reads aligned to the 462 scaffolds of the reference genome (63–90%), and only one isolate had a lower percentage of mapped reads (isolate 9683B13, 40%). Examination of 1000 randomly selected unmapped reads from genome 9683B13 showed contamination with bacterial sequences (68%; >30% *Pseudomonas* sp. and >10% *Stenotrophomonas maltophilia*, data not shown), so these sequences were discarded. Overall, this led to a sequencing depth average of 32X per genome (22X–46X; **Table 2**). Overall coverage was between 90.7 and 96.3% for the 15 isolates. For all genomes sequenced with paired-end reads (that is, all except 98AR1), the number of broken paired reads was relatively moderate (<11% and average of 9%).

COVERAGE AND SEQUENCING DEPTH ANALYSIS

Cross-comparison of mapping outputs identified a bias of average coverage and sequencing depth among the 462 reference scaffolds within and between isolates. For instance, several scaffolds systematically showed very high (>100X) or low (<1X) depths in all sequenced isolates, and others showed discrepancies for a given scaffold between different isolates. Such situations were manually inspected and led to the survey of 151 scaffolds (representing about 10% of the genome sequence) for which the mapping depth profile and the presence of genes along the scaffolds were recorded (Supporting Table 1). Notably, scaffold 484 showed a systematic high depth >1000X. Four mitochondrial scaffolds were previously identified and removed from the poplar rust genome assembly (Duplessis et al., 2011a). Mapping of Illumina reads from the 15 isolates onto these four scaffolds showed much higher depth than the average observed for other scaffolds (178X–1211X, data not shown). Inspection of scaffold 484 indicated that it is most likely a portion of the mitochondrial genome. Indeed, this 5.4 Kb scaffold bears two genes showing high homology to two mitochondrial genes (ATP synthase F0 subunit and NADH dehydrogenase subunit).

For other scaffolds with systematic high coverage and sequencing depth biases, major differences are explained by missing regions in one or several isolates. Such scaffolds were marked by no mapping support for the entire scaffold, or for some regions of the scaffold at the same positions in a given subset of isolates (i.e., probable large deletions or highly variable loci). For

Table 2 | General mapping information for the 15 *Melampsora larici-populina* isolates.

Isolate	Total reads number	Mapped reads	% Mapped reads	Broken pairs	Average read length	Sequencing depth
93ID6	3,594,455,577	2,656,764,147	73.9	226,296,523	84.4	26.3
02Y5	3,691,995,994	3,218,997,193	87.2	269,105,383	85.4	31.8
09BS12	6,230,429,688	4,717,557,005	75.7	479,213,815	84.2	46.6
94ZZ15	3,653,741,644	3,290,238,877	90.1	278,395,986	85.3	32.5
94ZZ20	3,387,309,786	3,045,158,939	89.9	253,470,401	85.2	30.1
08EA47	4,659,300,813	3,460,505,640	74.3	352,258,523	83.3	34.2
95XD10	4,701,407,950	3,993,529,488	84.9	396,812,163	83.7	39.5
08EA20	4,829,802,826	3,034,419,164	62.8	290,972,918	83.2	30.0
08EA77	4,259,571,919	3,840,082,037	90.2	340,127,111	84.7	38.0
97CF1	3,570,560,916	3,083,826,749	86.4	270,564,864	84.5	30.5
08KE26	5,407,393,523	4,626,803,739	85.6	434,085,871	85.0	45.8
9683B13	6,378,404,736	2,537,206,558	39.8	223,243,679	83.1	25.1
98AG31	2,779,485,081	2,529,716,573	91.0	218,294,868	85.2	25.0
93JE3	4,310,048,066	2,796,054,256	64.9	258,175,513	84.1	27.6
98AR1	2,562,143,464	2,227,530,892	86.9	na	76.0	22.0

Illumina reads of each genome were mapped onto the 98AG31 JGI reference genome. na, not applicable.

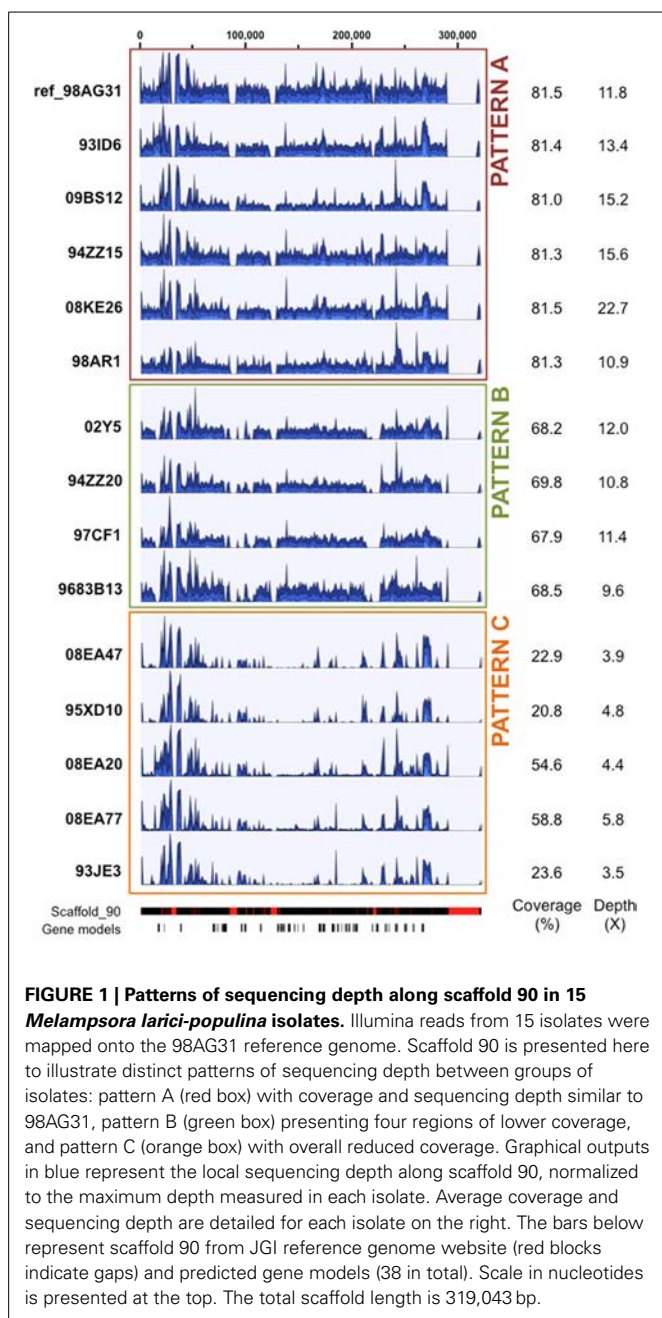
instance, the 319 Kb scaffold 90 showed either a similar depth along the scaffold in reference isolate 98AG31 and five other isolates (pattern A; **Figure 1**), or the absence of regions at the same positions for two patterns, each grouping different isolates (patterns B and C; **Figure 1**). Pattern C exhibited an overall low sequencing depth ranging from 3.5X to 5.8X, that mostly corresponds to repetitive elements regions marked by peaks of high depth similar to those present in patterns A and B. This indicates that the missing regions were not related to sequencing depth (**Figure 1**). For pattern C with the longest missing regions, a total of 38 genes were not supported by reads, including 4 pheromone genes related to mating type in the poplar rust fungus. Despite a generally similar profile and sequencing depths within pattern C, isolates 08EA20 and 08EA77 showed a higher coverage (54.6 and 58.8%, respectively) than the other three isolates (20.8, 22.9 and 23.6%). This is explained by a light and continuous depth in the central region of the scaffold that was totally absent in the other isolates (**Figure 1**). In isolate 08EA20 two genes located at 16–17 Kb (hypothetical protein) and 22–23 Kb (chitinase) were present. In isolate 08EA77, only the chitinase encoding gene was present, whereas these two genes were missing in the other three isolates of pattern C. Assembling unmapped reads from isolates exhibiting pattern C onto the 38 missing genes using loose similarity parameters retrieved only highly divergent and/or partial sequences (data not shown). Because of the presence of pheromone genes on scaffold 90, we looked at previously described mating type loci in the *M. larici-populina* genome (Duplessis et al., 2011a). A missing region containing a pheromone gene and a STE3 pheromone receptor gene was also observed in scaffold 172 for the isolates with pattern C. This prompted us to examine the homeodomain locus, composed of the genes HD1 and HD2. The five isolates that exhibited missing regions in scaffolds 90 and 172 also presented a missing region at the homeodomain locus in the scaffold 35. Using the homeodomain loci and pheromone/receptor loci genes as baits, divergent alleles were identified for *M. larici-populina* HD1,

HD2 and some pheromone genes in the unmapped reads of these isolates (data not shown).

A total of 212 genes lie in the missing regions of the surveyed scaffolds, including 12 SP genes in 7 scaffolds (Supporting Table 1). We therefore conducted a systematic analysis of regions of 100 bp or more showing coverage differences using the CLC coverage analysis tool, in order to detect possible deletions or amplifications. In total, 18,564–81,325 regions with significantly high/low coverage differences relative to the 98AG31 reference genome were identified in the 14 isolates (Supporting Table 2). Search for SP genes within these regions revealed that between 12 (9683B13) and 59 (95XD10) SP genes are in low coverage regions, indicating a possible deletion compared to isolate 98AG31. However, we could not find any correlation between a probable SP gene deletion and the pathotypes of the isolates, i.e., the absence of a SP gene explaining virulences 1, 2, 5, 6, or 8 (98AG31 reference isolate being virulence 3, 4, 7).

POLYMORPHISM AND INSERTION/DELETION DETECTION

In order to assess polymorphism in the 15 isolates, variants (SNVs/SNPs, MNVs, and InDels) were recorded using the CLC Genomics Workbench program. The 98AG31 reference genome had been sequenced at a 6.9X sequencing depth from dikaryotic urediniospores by Sanger sequencing, following a whole-genome shotgun strategy. Therefore, the 462 scaffolds represent a chimeric version of the genome combining the two haplotypes (Duplessis et al., 2011a). Resequencing by Illumina at a sequencing depth of 25X identified a total of 93,189 variants including 86,877 SNPs, 1741 MNVs, 2945 insertions and 1626 deletions in isolate 98AG31 (representing 96,099 bases; **Table 3**), which is in close range with the 88,083 SNPs recorded by Sanger sequencing. However, only 40,001 SNPs from the initial assembly were confirmed, highlighting differences due to the sequencing approaches. An average of 163,477 variants (including 152,936 SNPs) representing 168,708 bases was found in the 14 other isolates mapped onto the reference genome, representing a larger number of polymorphic sites



at the inter-individual level (0.17% of the genome; 1.51 SNPs/Kb). When the 15 genomes were considered together, 11,683 SNPs were conserved, whereas in total 611,824 unique SNPs were found. The variant caller implemented in CLC allowed the determination of the zygosity of nucleotides at the polymorphic sites. The heterozygosity rate was 0.45–0.55 in 12 isolates, whereas it was lower in 09BS12 and 08KE26 (0.35 and 0.37, respectively) and higher in 98AG31 (0.85). The latter is as expected, as it was the reference genome to which reads were mapped (Table 3). For all genomes, the ratio of transition over transversion mutations was 2.31 ± 0.11 (Table 4), which is similar in range to previous observations in rust fungi (Cantu et al., 2013). Individually, all isolates except the reference 98AG31 showed similar numbers of

SNPs, MNVs, and InDels (Table 3), indicating a homogeneous polymorphism rate at the intraspecific level. Polymorphic sites residing within coding DNA sequences (CDS) were more closely scrutinized and represented 20% of the SNPs, 17% of the MNVs, and 5% of deletions, and 5% of insertions in InDels. These proportions were rather similar in the different isolates (Table 4). In total, more SNPs were present in exons than in introns (average $30,077 \pm 3893$ SD and $14,982 \pm 1871$ SD, respectively; Table 4), but when exon and intron size were accounted for, introns tended to accumulate more SNPs than the coding sequences (data not shown).

HIGHLY VARIABLE GENES

Synonymous and non-synonymous polymorphisms within the 15 isolates were inspected in the gene complement of *M. larici-populina*, considering only SNPs that were represented in most of the observed variants (90%). Both homozygous and heterozygous SNPs were considered. For cross-comparison of SNPs between isolates, non-redundant SNPs (i.e., nucleotides in the reference isolate presenting polymorphism in at least one other isolate) were considered. Overall, a very large portion of the genes (89%) was marked at least by one SNP, and 5332 and 10 genes exhibited more than 10 and 100 SNPs, respectively (Supporting Table 3). A total of 1089 genes in the 15 isolates had more than 10 non-synonymous SNPs in CDS, the maximum number being 66 (proteinID 66139). Table 5 presents the top 30 genes with the highest number of non-synonymous SNPs over the 15 genomes, with 20.5 SNPs/Kb and 11.8 non-synonymous SNPs/Kb on average. Homology searches by Blastp against the NCBI nr protein database indicated a putative function or presence of a conserved domain for nine of the genes, six of which are associated with predicted nuclear activity. In total, 14 genes had GO and/or KOG annotations, and the majority encode predicted proteins of unknown function. A functional KOG analysis of the 4142 genes exhibiting ≥ 5 non-synonymous SNPs revealed significant enrichment for gene categories related to chromatin structure and dynamics; cell cycle control, cell division and chromosome partitioning; nuclear structure; defense mechanisms and extracellular structures (Figure 2). SNPs were also inspected in the 1 Kb upstream regions of CDS, where they may impact transcription. Most genes also had at least one polymorphic site in their 1 Kb upstream regions (89%) and 2554 genes each had more than 10 SNPs in these regions (Supporting Table 3). Half of the 30 genes with the highest number of SNPs had an annotation in various cellular categories including two SSP genes, the other half corresponded to genes encoding predicted proteins of unknown function (Supporting Table 4).

HIGHLY VARIABLE SECRETED PROTEIN ENCODING GENES

A set of 1184 SSP-encoding genes representing candidate poplar rust effectors was previously reported (Hacquard et al., 2012). Because larger effectors were also described (e.g., flax rust AvrM; Ravensdale et al., 2011), we decided to place a particular focus on secreted protein encoding genes as possible candidate effectors (i.e., a total of 2050 SPs identified by automatic annotation, including the 1184 SSPs). We further distinguish SSPs from SPs as SSP genes were manually annotated in the *M. larici-populina*

Table 3 | Genomic variants identified in 15 *Melampsora larici-populina* isolates by mapping onto the 98AG31 JGI reference genome.

Isolate	Zygoty		Variant types		Total			
	Homozygous	Heterozygous	Deletion	Insertion	MNVs	SNVs	Variants	Nucleotides
93ID6	84,849	88,855	3534	4198	3302	162,670	173,704	179,274
02Y5	76,511	95,418	3514	4399	3348	160,668	171,929	177,658
09BS12	91,934	54,500	3170	4020	2835	136,409	146,434	151,298
94ZZ15	84,155	82,478	3485	4287	3160	155,701	166,633	172,001
94ZZ20	80,851	80,541	3385	4085	3002	150,920	161,392	166,613
08EA47	85,423	75,527	3435	4158	3026	150,331	160,950	166,117
95XD10	68,735	87,520	2909	3554	2903	146,889	156,255	160,886
08EA20	90,268	91,000	3723	4354	3469	169,722	181,268	187,146
08EA77	89,765	83,569	3599	4275	3222	162,238	173,334	178,887
97CF1	75,954	76,585	3061	3902	2958	142,618	152,539	157,525
08KE26	102,244	55,022	3578	4268	3100	146,320	157,266	162,670
9683B13	70,974	82,208	3004	3708	2866	143,604	153,182	157,967
98AG31	14,219	78,970	1626	2945	1741	86,877	93,189	96,099
93JE3	91,933	75,793	3277	3938	3182	157,329	167,726	172,951
98AR1	77,267	88,799	3315	3932	3130	155,689	166,066	170,921

MNV, Multiple Nucleotide Variant; SNV, Single Nucleotide Variant (i.e., Single Nucleotide Polymorphism).

Table 4 | Analysis of polymorphism in 15 *Melampsora larici-populina* isolates.

Isolate	SNPs				% Polymorphism in CDS				
	Tr/Tv	SNPs in exon	SNPs in intron	SNPs intergenic	Non-synonymous SNP	Deletion	Insertion	MNV	SNV
93ID6	2.30	33,428	16,489	112,753	15,950	5.0	5.7	18.3	20.5
02Y5	2.30	32,904	16,325	111,439	15,553	5.5	5.4	15.9	20.5
09BS12	2.34	26,086	13,365	96,958	12,905	5.2	5.6	16.8	19.1
94ZZ15	2.29	31,938	16,056	107,707	15,252	6.2	5.7	17.2	20.5
94ZZ20	2.29	31,035	15,461	104,424	14,859	5.2	5.2	17.7	20.6
08EA47	2.30	29,848	14,986	105,497	14,493	5.3	5.5	17.3	19.9
95XD10	2.42	29,932	14,817	101,940	15,950	4.5	4.9	14.8	20.4
08EA20	2.30	35,069	17,230	117,423	16,911	5.3	5.4	18.0	20.7
08EA77	2.35	32,152	16,383	113,703	15,653	5.4	5.4	17.1	19.8
97CF1	2.33	29,566	14,649	98,403	14,218	5.7	6.1	17.9	20.7
08KE26	2.36	27,137	13,886	105,297	13,862	5.3	5.6	16.5	18.5
9683B13	2.33	29,776	14,719	99,109	14,442	5.1	5.5	17.7	20.7
98AG31	2.27	18,749	9335	58,793	8825	6.6	5.8	19.9	21.6
93JE3	2.36	32,155	15,684	109,490	15,651	4.9	5.4	18.0	20.4
98AR1	2.21	31,389	15,352	108,948	14,441	4.7	5.2	17.0	20.2

CDS, Coding DNA sequence. Tr/Tv, rate of transition to transversion; MNV, Multiple Nucleotide Variants; SNV/SNP, Single Nucleotide Variant/Polymorphism.

genome (Hacquard et al., 2012). Overall, a very large portion of the SP genes (89%) was marked by at least one SNP and 586 exhibited 10 SNPs or more (Supporting Table 5). A total of 386 and 119 genes had more than 5 and 10 non-synonymous SNPs, respectively (maximum = 45 non-synonymous SNPs; proteinID 66458). **Table 6** presents the top 30 SP genes with the highest numbers of non-synonymous SNPs/Kb, of which 24 are SSP genes. Only six SPs showed homology to other fungal proteins, including an *M. lini* avirulence factor AvrP4, a metallopeptidase, and a pleckstrin homology-like domain involved in binding to interacting protein partners. Rates of synonymous (P_S) and non-synonymous (P_N) substitutions were calculated for all genes with

the EggLib package (Supporting Table 3) and SP genes were more particularly scrutinized. The P_N/P_S rate could be measured for 14,052 genes, while 1073 genes had a mutation generating a stop codon in the sequence and were excluded. P_N/P_S showed similar distributions between SP genes and other genes (**Figure 3**) and the highest P_N/P_S (4.9) was found for a gene encoding a hypothetical protein (ProteinID_70080; Supporting Table 3). In SP genes, the highest P_N/P_S was 2.47 and corresponds to a SSP of 200 amino acids with three homologs in *Puccinia graminis* f. sp. *tritici* and no conserved domain (ProteinID_124304; Supporting Table 5). The average P_N/P_S observed in SP genes (0.20) was lower than for other genes (0.25). A total of 68 SP genes showed a $P_N/P_S >$

Table 5 | Top 30 genes accumulating non-synonymous (NS) Single Nucleotide Polymorphism (SNP).

JGI Protein ID ^a	Protein length	Transcript length	SNP	NS	Annotation	GO ID ^a	KOG ID ^a
66139	5273	15819	227	66	AAA+ ATPase	0003677	1808
84101	1325	3975	95	57	Hypothetical protein	No hit	No hit
93626	1737	5211	82	54	Hypothetical protein	No hit	No hit
62079	1821	5463	73	47	Hypothetical protein, telomere-length maintenance and DNA damage repair domain	0001584	No hit
106057	2195	6585	136	45	Hypothetical protein, NAM-like protein C-terminal domain	No hit	No hit
92944	1135	3405	71	45	Hypothetical protein, DNA breaking-rejoining enzymes, C-terminal catalytic domain	No hit	No hit
95670	893	2679	87	45	Hypothetical protein	No hit	1187
66458	929	2787	55	45	Hypothetical protein	No hit	1245
70222	1542	4626	73	44	DEAD-like helicase superfamily	No hit	0351
101154	1470	4410	79	44	Hypothetical protein	No hit	No hit
114610	948	2844	91	41	Hypothetical protein	No hit	No hit
85441	1256	3768	56	40	Hypothetical protein	No hit	0714
92226	1393	4179	59	38	Hypothetical protein	No hit	No hit
67208	1203	3609	76	37	Hypothetical protein	No hit	1015
108793	931	2793	54	37	Hypothetical protein	No hit	No hit
96388	1344	4032	63	36	Hypothetical protein	No hit	No hit
108574	2851	8553	114	35	Hypothetical protein, down-regulated in metastasis domain	No hit	No hit
91870	1131	3393	72	35	Hypothetical protein, alpha kinase domain family	0004674	3614
118268	1649	4947	108	34	Hypothetical protein, sister-chromatid cohesion C-terminus domain	0006520	No hit
68278	1507	4521	54	34	Hypothetical protein	No hit	4475
65221	568	1704	44	34	Hypothetical protein	No hit	No hit
91258	771	2313	51	33	Hypothetical protein, GCM transcription factor family motif	No hit	2992
88323	575	1725	55	33	Hypothetical protein	No hit	No hit
60895	698	2094	58	33	Hypothetical protein	No hit	2992
84177	639	1917	57	33	Hypothetical protein	No hit	No hit
92190	551	1653	52	33	Hypothetical protein	0006306	No hit
101664	1102	3306	63	32	Hypothetical protein	No hit	No hit
95815	1486	4458	46	32	Hypothetical protein	No hit	1245
107058	720	2160	51	32	Hypothetical protein	No hit	No hit
64441	1107	3321	45	31	Hypothetical protein	No hit	No hit

^a Protein ID number, Eukaryotic Orthologous Group (KOG) and Gene Ontology (GO) annotations were retrieved from the 98AG31 reference genome at the Joint Genome Institute MycoCosm website (<http://genome.jgi.doe.gov/programs/fungi/index.jsf>).

1, whereas 668 had a $P_N/P_S > 1$ in other genes (Supporting Table 6). Among the 30 genes with the highest numbers of non-synonymous SNPs, nine have a $P_N/P_S > 1$ (Table 6). These genes represent particularly interesting candidates that could have evolved under the selection pressure exerted by the interaction with the host plant. No enrichment in KOG functional annotation was detected for the 736 genes presenting a $P_N/P_S > 1$.

In the panel of 15 *M. larici-populina* isolates, only two of the eight virulences described in the poplar rust fungus presented a balanced frequency: virulence 3 with six avirulent isolates and

nine virulent isolates and virulence 7 with seven avirulent isolates and eight virulent isolates (Table 1). SP genes presenting conserved non-synonymous SNPs in avirulent isolates and not in virulent isolates (including the reference genome 98AG31 which carries virulences 3 and 7) could be strong candidates, however none of the SP genes presented such a pattern for virulence 3 and 7, suggesting that events other than non-synonymous substitutions in coding sequence may explain the emergence of the virulences 3 and 7. Four SP genes (Protein IDs 89167, 91014, 105154, and 123753) presented non-synonymous SNPs in isolates

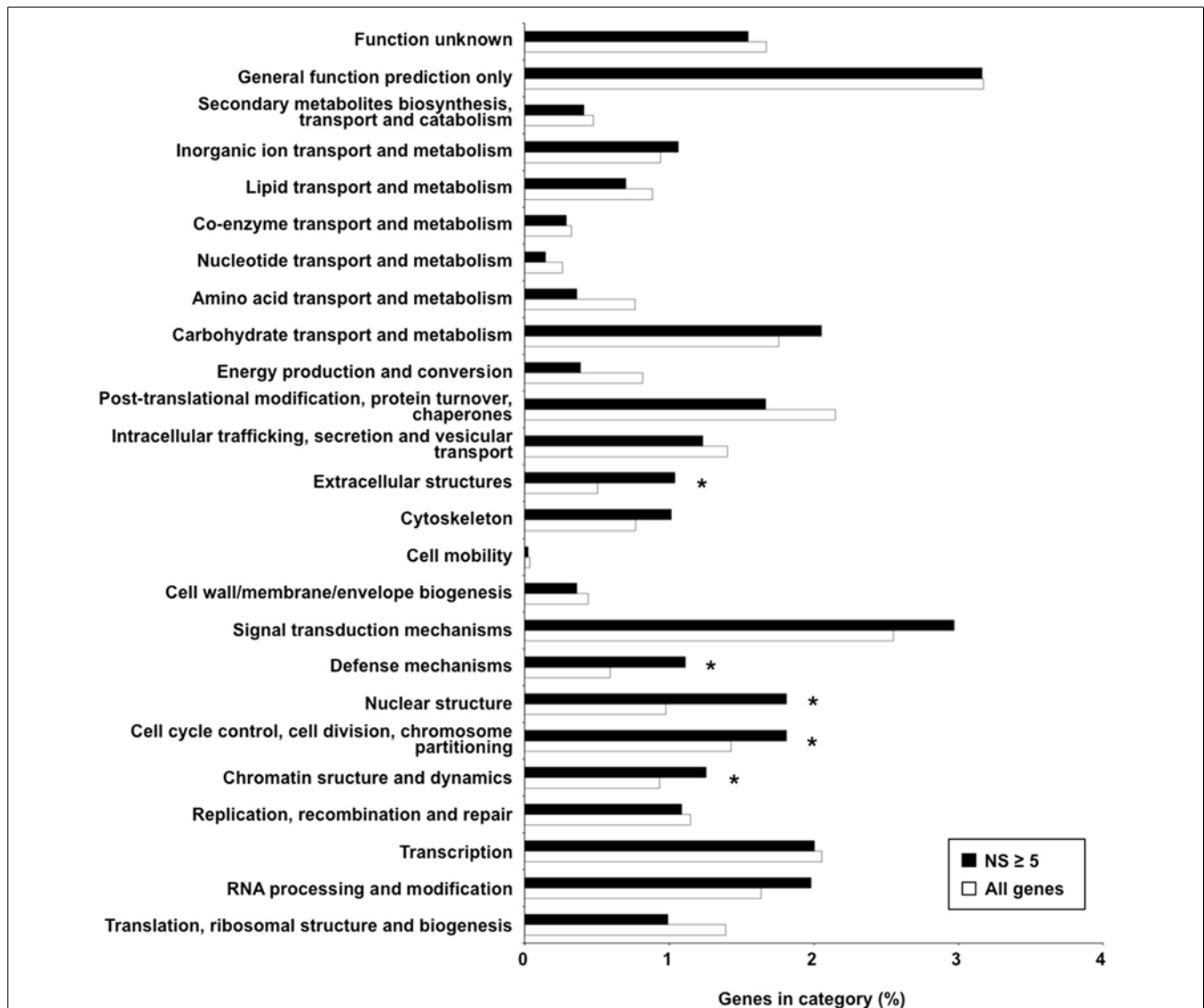


FIGURE 2 | Functional categories over-represented among genes exhibiting five non-synonymous polymorphisms or more. Percentages of genes falling in the different KOG categories among genes exhibiting five non-synonymous polymorphisms or more (NS \geq 5) relative to the global gene distribution are shown. Black and white bars correspond to selected NS \geq 5

genes and all genes, respectively. The category “No hits” corresponding to genes with no KOG annotation (~75% in both sets) is not represented on the graph to facilitate visualization of other categories. Significantly over-represented KOG categories are indicated by asterisks (Fisher’s exact test, $p < 0.05$).

98AR1 and 02Y5 which bear the virulence 8, whereas these were absent from the other 13 avirulent isolates, suggesting these genes could be candidate effectors for virulence 8. One SP gene (Protein ID 104703) presented non-synonymous SNPs in isolates 98AR1 and 9683B13 that were absent from the other isolates, indicating that this gene could be a candidate related to virulence 1. One SP gene (Protein ID 108857) presented non-synonymous SNPs in isolates 08EA77, 9683B13, and 09BS12, whereas they were absent from the 12 other isolates, suggesting also that this gene could be a candidate for virulence 6. No correlation was found between mutations in SP genes and other virulences. Similarly, none of the genes interrupted by stop codons correlated with the pathotypes of the 15 isolates.

M. larici-populina SSP genes showing homology to *M. lini* Avr genes *AvrL567*, *AvrP123*, and *AvrP4* do not exhibit important accumulation of non-synonymous SNPs (Supporting Table 5). Interestingly, the polymorphic sites identified for the *M. lini* *AvrL567* homolog in the poplar rust genome correspond to those that were previously identified by PCR-cloning in a panel of 32 *M. larici-populina* isolates (Hacquard et al., 2012), which included isolate 98AR1, validating the SNPs found in this candidate. Evidence of positive selection were previously recorded for *AvrP4* genes at the intraspecific level in *M. lini* (Barrett et al., 2009) and at the interspecific level in the Melampsoraceae family (Van der Merwe et al., 2009), as well as in a cluster of paralogous genes encoding *AvrP4*-homologs (multigene family CPG5464;

Table 6 | Top 30 genes encoding secreted proteins accumulating non-synonymous SNPs/Kb.

Protein ID ^a	Protein length	Transcript length	SNP	NS	NS/Kb	Annotation	KOG ID ^a	Go ID ^a
124497	77	231	5	5	21.6	hypothetical secreted protein of 8 kDa	No hit	No hit
124050	151	453	13	9	19.9	hypothetical secreted protein of 17 kDa	No hit	No hit
124361	88	264	5	5	18.9	hypothetical secreted protein of 9 kDa	No hit	No hit
109910	230	690	17	13	18.8	hypothetical secreted protein	No hit	No hit
123541	75	225	6	4	17.8	hypothetical secreted protein of 8 kDa	No hit	No hit
123852	135	405	55	7	17.3	hypothetical secreted protein of 15 kDa	No hit	No hit
104907	117	351	6	6	17.1	hypothetical secreted protein	1245	No hit
123868	139	417	15	7	16.8	hypothetical secreted protein of 15 kDa	No hit	No hit
66458	929	2787	55	45	16.1	hypothetical secreted protein	No hit	No hit
103402	151	453	15	7	15.5	hypothetical secreted protein	No hit	No hit
101262	131	393	18	6	15.3	hypothetical secreted protein	No hit	No hit
124304	200	600	10	9	15.0	hypothetical secreted protein of 22 kDa	No hit	No hit
107425	268	804	28	12	14.9	hypothetical secreted protein	No hit	No hit
124511	67	201	3	3	14.9	hypothetical secreted protein of 7 kDa	No hit	No hit
124264	90	270	5	4	14.8	hypothetical secreted protein of 10 kDa, <i>Melampsora lini</i> AvrP4 homolog	No hit	9055
107508	720	2160	51	32	14.8	hypothetical secreted protein	No hit	No hit
124351	92	276	7	4	14.5	hypothetical secreted protein of 10 kDa	No hit	No hit
95362	301	903	18	13	14.4	hypothetical secreted protein	No hit	No hit
64885	188	564	23	8	14.2	hypothetical secreted protein of 21 kDa	No hit	No hit
58423	142	426	10	6	14.1	hypothetical secreted protein of 14 kDa	No hit	No hit
124524	71	213	3	3	14.1	hypothetical secreted protein of 8 kDa	No hit	No hit
63656	315	945	22	13	13.8	hypothetical secreted protein	No hit	No hit
70838	97	291	9	4	13.7	hypothetical secreted protein of 10 kDa	No hit	No hit
123559	146	438	10	6	13.7	hypothetical secreted protein of 16 kDa	No hit	No hit
61241	392	1176	39	16	13.6	hypothetical secreted protein, PLECKSTRIN homology domain	No hit	No hit
68348	247	741	18	10	13.5	hypothetical secreted protein	No hit	No hit
123552	150	450	12	6	13.3	hypothetical secreted protein of 17 kDa	No hit	No hit
124134	125	375	14	5	13.3	hypothetical secreted protein of 14 kDa	No hit	No hit
108793	931	2793	54	37	13.2	hypothetical secreted protein	No hit	No hit
36743	179	537	8	7	13.0	hypothetical secreted protein of 21 kDa, peptidase M, neutral zinc metallopeptidase	No hit	8237

^aProtein ID number, Eukaryotic Orthologous Group (KOG) and Gene Ontology (GO) annotations were retrieved from the 98AG31 reference genome at the Joint Genome Institute MycoCosm website (<http://genome.jgi.doe.gov/programs/fungi/index.jsf>).

Hacquard et al., 2012). The 13 members of the CPG5464 family in *M. larici-populina* were more closely examined in the 15 isolates (Figure 4). The 13 members of the family were rather conserved and only four had non-synonymous SNPs between isolates (CPG5464_124256, CPG5464_124262, CPG5464_124264, CPG5464_124266). In total, substitutions were noted at four different positions, two within the signal peptide and two after the conserved K/R and E/D regions. None of these substitutions corresponded to positions previously shown under positive selection at the intraspecific or interspecific level (Figure 4). Notably, CPG5464_124564, which includes three different substitution sites in three isolates, presented a P_N/P_S value of 1 and was among the SP genes exhibiting the highest numbers of SNPs/Kb (Table 6, Supporting Table 5). Among the eight homologs of *M. lini* AvrM genes, one showed 15 non-synonymous SNPs (ProteinID_124207; Supporting Table 3).

Three *Uromyces fabae* RTP1 homologs have been described in *M. larici-populina* (Hacquard et al., 2012). Only one RTP1 homolog (ProteinID_123932; Supporting Table 3) that consists of a fusion between a *M. lini* HESP-327 homolog and an *U. fabae* RTP1 homolog exhibited an important number of non-synonymous SNPs (7, of which 5 reside in the C-terminal RTP1 region). No substitution occurred at the positions of the four conserved cysteine residues under purifying selection identified by Pretsch et al. (2013).

DISCUSSION

The sequencing of the *M. larici-populina* genome has opened new avenues for the study of effector genes in a model pathosystem composed of a perennial plant and an obligate biotrophic rust fungus (Duplessis et al., 2011a; Hacquard et al., 2011). A set of 1184 candidate poplar rust effectors were identified on the

basis of a combination of typical features of effectors reported in other fungal pathogens, including an initial arbitrary size filter to focus on small proteins of less than 300 amino acids (Hacquard et al., 2012). Because rust fungi effectors such as the *M. lini* AvrM avirulence factor (Ravensdale et al., 2011) can be larger,

all predicted SPs were subsequently considered in the search for candidate effectors. Complementary information such as transcript profiling during host infection or the pathogen life cycle can help in reducing the set of genes likely to be *bona fide* effectors (Duplessis et al., 2011b; Hacquard et al., 2013a). Another filter commonly used to identify candidate effectors in plant pathogens is the detection of positive selection in virulence genes, indicative of the evolutionary pressure exerted by the plant-pathogen co-evolution (Alfano, 2009; Stergiopoulos and de Wit, 2009). Events such as non-synonymous substitutions, gene gain, gene loss or differential regulation of gene expression can affect avirulence genes and generate new virulences in plant pathogens; comparative genomics using new generation sequencing technologies have uncovered such types of events (Raffaele and Kamoun, 2012; Spanu, 2012). In the present study, we applied Illumina sequencing by synthesis to explore the genetic diversity of *M. larici-populina*, focusing on 15 isolates collected on poplar trees in the wild or in experimental poplar nurseries in the past 21 years in France, and with a wide range of virulence profiles. The main goal here is to provide another level of information about *M. larici-populina* genes in order to guide selection of pathogenesis-related genes, including effectors, for future functional analyses. The mapping of Illumina reads onto the 98AG31 reference genome helped in the detection of variations such as SNPs and InDels. To date, only a few reports explored genetic diversity at the genome scale in rust fungi using Illumina technology, but they provide ground for comparison within the Pucciniales order (Duplessis et al., 2014b).

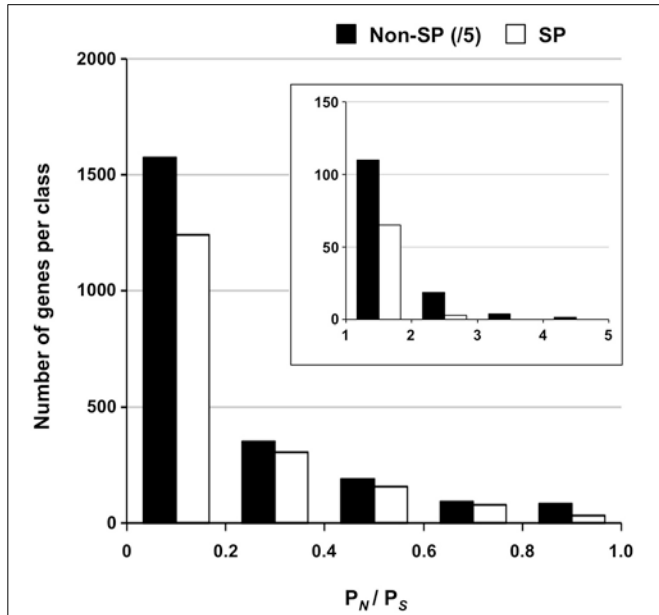


FIGURE 3 | Distribution of P_N/P_S for SP and non-SP genes. Ratios of non-synonymous to synonymous polymorphisms (P_N/P_S) between 0 and 1 are shown for SP genes and non-SP genes. The insert shows distribution of genes with a $P_N/P_S > 1$. Numbers of non-SP genes were divided by 5 for representation. Note the different scale for y-axes in figure and insert.

RESEQUENCING *M. LARICI-POPULINA* GENOMES REVEALS STRUCTURAL VARIATIONS

Reads were mapped onto the 98AG31 reference genome with good overall coverage and sequencing depth. Although there was

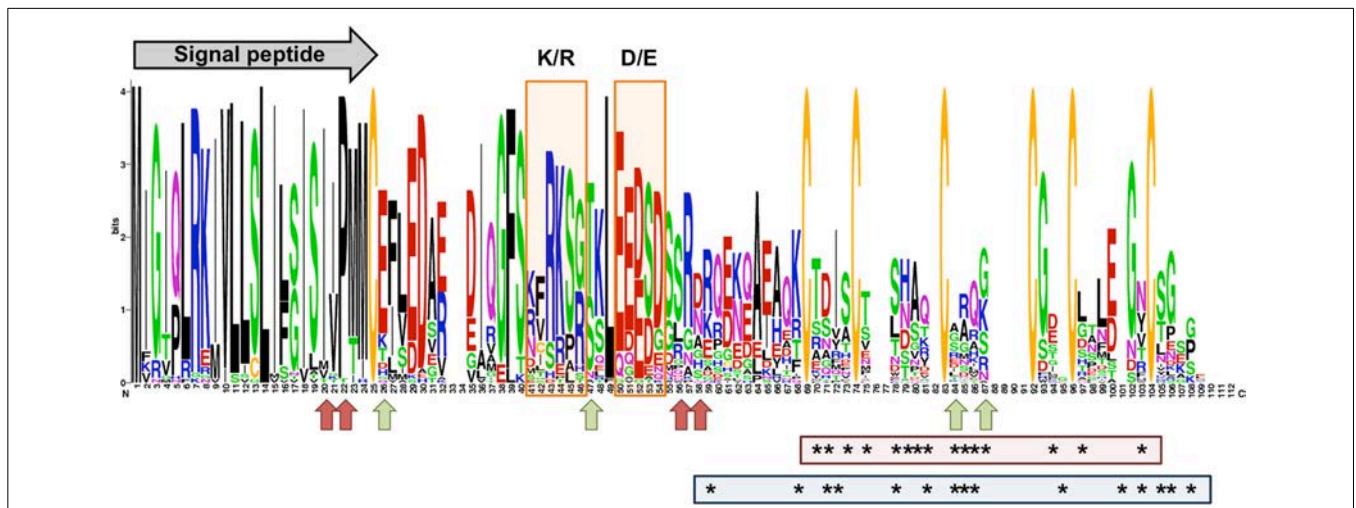


FIGURE 4 | Conservation protein profile of the *M. larici-populina* CPG5464 family and AvrP4 homologs in Melampsoraceae. The profile was designed using WebLogo with 40 sequences corresponding to the 12 members in the CPG5464 family (Hacquard et al., 2012), six variants deduced from the 15 genomes sequenced in this study, 22 AvrP4 homologs sequenced from 9 *Melampsora* spp. (Van der Merwe et al., 2009) and 16 *Melampsora lini* AvrP4 variants. The predicted signal peptide

and K/R and D/E rich regions previously shown in Hacquard et al. (2012) are depicted on the profile. Green arrows point to sites under selection in Barrett et al. (2009). Red arrows point to sites of substitution observed in *M. larici-populina* variants. Asterisks in the red box indicate amino acids under positive selection in Van der Merwe et al. (2009) and asterisks in the blue box indicate amino acids under positive selection in Hacquard et al. (2012).

a narrow range in the average coverage by isolate, discrepancies were observed for given scaffolds. Particularly, the small scaffold 484 presented a strikingly high sequencing depth. Two genes encoding an ATP synthase F0 subunit and a NADH dehydrogenase subunit presenting strong similarity with resident genes of the soybean rust *Phakopsora pachyrhizi* mitochondrial genome (Stone et al., 2010) are present on this scaffold. Thus, our analysis identifies a new mitochondrial scaffold that will help in refining the genome assembly. Detailed examination of scaffolds that presented divergent coverage and sequencing depth between isolates revealed on some occasions rather large missing gene-containing regions compared to the reference genome. Although still unresolved, the poplar rust fungus seems to possess a tetrapolar mating system, as for many other basidiomycetes (Duplessis et al., 2011a). In this system, two unlinked loci govern the sexual cycle, and both loci should differ to complete mating (Fraser et al., 2007). Three distinct patterns of conserved missing regions were observed between isolates of unrelated pathotypes collected on different years at different locations (see **Table 1** for collection details). Scaffold 90 showed the most striking differences, where missing regions encompass a total of 38 genes, including four pheromone genes that were previously annotated in mating type loci of *M. larici-populina* (Duplessis et al., 2011a). Other mating type loci (i.e., the pheromone/receptor and the homeodomain loci) are also missing in these isolates suggesting that their mating type loci are highly divergent. Despite the quality of the reference genome assembly, the organization of the mating type loci is still not resolved (Duplessis et al., 2011a). This study will provide support to further explore and resolve the organization and composition of the poplar rust fungus mating loci. Other missing regions unrelated to the mating loci suggest that the poplar rusts possess a great genomic variability. In *M. oryzae*, 1.68 Mb (of a total of 38 Mb) were missing in isolate Ina168 resequenced by 454-pyrosequencing compared to the 70-15 reference genome (Yoshida et al., 2009). This has led to the discovery of many missing SSP genes including known avirulence genes between the two *M. oryzae* isolates (Yoshida et al., 2009). In *M. larici-populina*, none of the missing regions contained large numbers of SP genes (only 12 in total). By performing a wider coverage analysis in the 15 isolates, up to 59 SP genes were found in low coverage regions, representing possible deletions. However, no such deletion correlates with the poplar rust virulences. In *P. striiformis* f. sp. *tritici*, less than 1.3% of the secretome (15 SP genes) was absent between the most divergent sequenced isolates (Cantu et al., 2013), which indicates that the same set of SP genes occurs at the intraspecific level in these rust fungi.

M. LARICI-POPULINA GENOMES SHOW REMARKABLE LEVELS OF POLYMORPHISM

The reference genome 98AG31 was included in the panel of 15 isolates. This genome was previously characterized by Sanger sequencing, which provided an adequate assembly into 462 scaffolds (considering the large size of 101 Mb and a large content in TE, i.e., 45%), however at a rather low sequencing depth of 6.9X (Duplessis et al., 2011a). A total of 88,083 SNPs were previously identified in the reference genome by mapping back Sanger sequencing reads onto the assembled reference genome, with a

loose criterion considering a minimum of four reads at a given position (Duplessis et al., 2011a). Illumina sequencing identified a total of 93,189 variants including 86,877 SNPs, of which only 40,001 confirmed SNPs found in the initial assembly. This finding strengthens the support for the use of resequencing at a greater depth to confidently assess SNPs. The total number of SNPs we report is slightly lower than the one found in *P. graminis* f. sp. *tritici* (129,172; Duplessis et al., 2011a). It differs, too, to the numbers reported in *P. striiformis* f. sp. *tritici*, with 81,001–108,785 depending on the isolate considered in Zheng et al. (2013) and more than 350,000 with important variations between isolates in Cantu et al. (2013). The large variation in SNPs in these studies could be explained by the wide variation in geographical origin of the isolates and the varying rates of occurrence of sexual reproduction at these sites. Population analyses of the poplar rust fungus with neutral markers indicate that the fungus frequently undergoes sexual recombination resulting in regular gene flow within natural population (Gérard et al., 2006; Barrès et al., 2008; Xhaard et al., 2011). Overall, these findings indicate a great genetic diversity in rust fungi that possess a complex life cycle with a sexual reproduction stage achieved on an alternate host (Duplessis et al., 2014b).

Because of the high TE content and the large size of the poplar rust genome, together with putatively large differences between isolates (as previously reported in *P. striiformis* f. sp. *tritici*), we did not expect *de novo* assembly to be optimal for analysis of the 14 isolates sequenced for the first time in this study. Indeed, *de novo* assembly generated large numbers of scaffolds (i.e., >30,000, data not shown). Instead, Illumina reads from the 14 isolates were directly mapped onto the 98AG31 reference genome for variants detection, similar as in Zheng et al. (2013). In *M. larici-populina*, an average of 148,532 SNPs per isolate were uncovered, which is slightly higher than in *P. striiformis* f. sp. *tritici* according to Zheng et al. (2013). The proportions of heterozygous SNPs in the two isolates 08KE26 and 09BS12 (35 and 37%, respectively), might reflect their assignment to an asexual group as described by the poplar rust population genetic analysis of Xhaard et al. (2011). A much higher proportion of heterozygous SNPs were found between *P. striiformis* f. sp. *tritici* isolates: 82–84% in Zheng et al. (2013) and 87–99% in Cantu et al. (2013). The observed differences between the two studies may reflect differences in the sequencing and analysis process used (Duplessis et al., 2014b), or could be related to a different reproduction regime, as *P. striiformis* f. sp. *tritici* is mostly asexual which fosters individual heterozygosity (Balloux et al., 2003; Halkett et al., 2005). It would be interesting to compare this with the genetic diversity in rust fungi such as *P. pachyrhizi* or *H. vastatrix* with no known sexual reproduction to date (Rodrigues et al., 1975; Goellner et al., 2010). InDel variants were also inspected and ranged from 4571 to 8077 in the 15 *M. larici-populina* isolates, which is slightly larger than in *P. striiformis* f. sp. *tritici* where 1863 on average were reported (Zheng et al., 2013), but smaller than in the yeast *Saccharomyces* sp. (Liti et al., 2009). A substantial level of polymorphism is noted in *M. larici-populina* at the intraspecific level (~6 SNPs/Kb), which is in close accordance with those reported in the shiitake mushroom *Lentinula edodes* (4.6 SNPs/Kb, Au et al., 2013) or in the wheat stripe rust fungus

P. striiformis f. sp. *tritici* (Cantu et al., 2013). It is slightly larger than in plant pathogenic ascomycetes such as *Pyrenophora tritici-repentis* (1.9 SNPs/Kb, Manning et al., 2013), *Blumeria graminis* (less than 2 SNPs/Kb; Hacquard et al., 2013b; Wicker et al., 2013) or *Leptosphaeria maculans* (0.5 SNPs/Kb; Zander et al., 2013) but much lower than in the yeast *S. cerevisiae* (59.8 SNPs/Kb; Liti et al., 2009) or in the plant pathogen *Rhizoctonia solani* (~15 SNPs/Kb; Hane et al., 2014). The observed differences in the levels of polymorphism could reflect evolutionary trends related to the lifestyle of these fungi. Rust fungi, exhibit a remarkable level of polymorphism, providing ground for detection of loci that may underlie the co-evolution with their associated hosts and/or their unique life cycle, which is marked by the formation of five spore types and infection of two alternate hosts (Duplessis et al., 2014b).

PATTERNS OF GENETIC VARIATIONS IN POPLAR RUST GENES UNCOVER CANDIDATE PATHOGENESIS-RELATED GENES

A large part of the variants was identified in coding sequences, similar to *P. striiformis* f. sp. *tritici* (Cantu et al., 2013; Zheng et al., 2013). In total, 89 and 74% of the 16,399 *M. larici-populina* genes were marked by at least one SNP, or one non-synonymous SNP, respectively, in one of the isolates. Such valuable information provides ground for detailed analysis of the functions that may be under selection in the poplar rust genome, particularly those evolving under the pressure of the host plant. P_N/P_S values can be informative to the detection of positive selection and the understanding of how fungi adapt to their environment (Stukenbrock and Bataillon, 2012). We examined the genes showing a $P_N/P_S > 1$ with a particular focus on candidate effectors. Strikingly, whereas other comparative genomic studies have revealed candidate effector genes under positive selection (Cooke et al., 2012; Wicker et al., 2013), we did not detect any enrichment in SP genes exhibiting a high P_N/P_S compared to all genes in the poplar rust genome. However, 68 SP genes in total showed a $P_N/P_S > 1$ and are priority candidates. Other genes falling in this category may be related to pathogenesis-related functions, but no particular enrichment in functional annotation could be detected. However, the missing regions in *M. larici-populina* isolates contain many genes encoding small proteins (i.e., less than 300 amino acids) with no predicted signal peptide. In the obligate biotroph *B. graminis*, selection analysis carried out between formae speciales identified candidate effectors with no predicted signal peptide that share other common evolutionary features with annotated effectors (Wicker et al., 2013). A total of 262 *M. larici-populina* genes encoding small proteins were found with a $P_N/P_S > 1$ (Supporting Table 6). Such small protein encoding genes are also found among *in planta* highly expressed genes of *M. larici-populina* (Duplessis et al., 2011b). Although no unconventional secretory system is known so far in rust fungi, it would be tempting to consider such proteins in future analysis as possible candidate effectors. We therefore examined the genes presenting a large proportion of non-synonymous substitutions in their sequence and detected enrichment in KOG categories related to nuclear structure and function. Interestingly, genomes of rust fungi contain significantly expanded gene families encoding helicases that may play an important role in DNA repair and maintenance, and nucleic acid and zinc-finger proteins

corresponding to putative transcription factors (Duplessis et al., 2011a; Zheng et al., 2013). DNA repair systems can have a dramatic impact on genomic diversity (Seidl and Thomma, 2014) and their possible role in the evolution of the poplar rust genome is still to be determined.

In our study, variations occurring in upstream sequence of genes were also inspected, on the grounds that they may relate to regulation of expression. In total, 16% of the genes had more than 10 SNPs in their 1 Kb upstream region. Detailed transcriptome-driven analyses of conserved cis-acting regulatory elements in *P. infestans* have revealed motifs underlying specific expression of pathogenesis-related genes (Seidl et al., 2012; Roy et al., 2013a,b). The transcriptome analysis of poplar leaf infection by *M. larici-populina* has shown conserved patterns of coordinated expression of several sets of SSP genes along a time course experiment (Duplessis et al., 2011a). Several other transcriptomic studies have confirmed this trend for SP genes in rust fungi (Fernandez et al., 2012; Cantu et al., 2013; Tremblay et al., 2013; Bruce et al., 2014; Duplessis et al., 2014b). A better knowledge of cis-acting regulatory elements in the genome of *M. larici-populina* is needed to further explore the impact of mutations in upstream gene regions. Other molecular mechanisms may control regulation of expression profiles, as recently exemplified in the oilseed rape ascomycete pathogen *L. maculans* (Soyer et al., 2014). Particularly of note, a significant enrichment in genes falling in the chromatin structure and dynamics KOG category was found in genes accumulating non-synonymous SNPs, and it remains to be explored whether such a control of the chromatin structure could relate to the control of gene expression in rust fungi.

A major goal of the present study was to uncover the presence of polymorphic effectors within a set of predefined candidates that may reflect specific adaptation to the host plant in the classical scheme of the plant-pathogen arms race. A similar approach conducted in *P. striiformis* f. sp. *tritici* identified five polymorphic candidate effectors by comparing two isolates presenting distinct pathotypes (Cantu et al., 2013). Another study identified such possible avirulence genes among secreted protein transcripts showing patterns of non-synonymous mutations between different *Puccinia triticina* isolates (Bruce et al., 2014). In the panel of *M. larici-populina* isolates, virulences 1, 6, and 8 presented correlations with the presence of non-synonymous SNPs in one, one and four genes of virulent isolates compared to avirulent isolates, respectively. Such genes could be candidates underlying virulences 1, 6, and 8. No such correlation was observed for the other virulences carried by the poplar rust isolates, indicating that other events than non-synonymous substitutions in coding sequences may explain their emergence.

Sequence polymorphism has been reported in several avirulence genes of the flax rust *M. lini* (Catanzariti et al., 2006; Dodds et al., 2006; Barrett et al., 2009; Van der Merwe et al., 2009; Ravensdale et al., 2011). Homologs of flax rust avirulence genes retrieved in the *M. larici-populina* genome did not exhibit high P_N/P_S or excess of non-synonymous substitutions in the 15 isolates, except in a very few cases. Interestingly, non-synonymous substitutions observed in the CPG5464 family homologous to *M. lini* AvrP4 did not match sites previously shown under selection in *M. lini* at the intraspecific level (Barrett

et al., 2009), in Melampsoraceae at the interspecific level (Van der Merwe et al., 2009) or between members of the paralogous gene cluster CPG5464 of *M. larici-populina* (Hacquard et al., 2012). Members of this gene family are rather conserved within the Melampsoraceae, suggesting that AvrP4/CPG5464 could play an important role as an effector during the interaction with the relative host plants. A high diversity is observed at both the intraspecific and interspecific level highlighting the probable interplay with the different host plants, but to date, such an interaction in a gene-for-gene manner has only been demonstrated for the flax rust fungus (Ravensdale et al., 2011). At least one *M. larici-populina* homolog of the *M. lini* AvrM gene shows a high level of polymorphic sites (e.g., in isolate 98AR1, 30 SNPs of which 15 are non-synonymous), similar to those reported in *M. lini* (Catanzariti et al., 2006; Ravensdale et al., 2011). Some of these mutations are particularly important for the direct interaction with the corresponding *M* resistance gene in flax (Catanzariti et al., 2010; Ve et al., 2013). It will be particularly interesting to further study the potential role of AvrM homologs in the poplar-poplar rust fungus interaction.

FUTURE STEPS IN POPLAR RUST GENOMICS

Genomics is a powerful approach to identify pathogenesis-related candidates, as the present study illustrates. From the perspective of population biology, it is well-known that structure and demography can affect all loci equally. To identify loci under selection, a population genomics approach is required to take into account demographic history. A population genomics study is ongoing in collaboration with the JGI to identify loci related to virulence 7. As large portions of the genome were missing in different *M. larici-populina* isolates, it might be required to study presence/absence at a larger scale using *de novo* assembled genomes. Many mechanisms can underlie genome evolution (Raffaele and Kamoun, 2012; Seidl and Thomma, 2014) and a better knowledge of the structural rearrangements occurring in the poplar rust genome will help to determine their impact on virulence evolution. In this regard, we have initiated the genome sequencing of an avirulent 7 isolate by combining paired-end and mate-pair Illumina sequencing to compare with the virulent 7 reference genome. Together, these genomic analyses will foster functional studies by pinpointing numerous sites of sequence variation, i.e., positions that may have important implications at the structural level for the function of effectors.

AUTHOR CONTRIBUTIONS

Sébastien Duplessis and Pascal Frey designed research; Antoine Persoons, Sébastien Duplessis, Christine Delaruelle, and Pascal Frey performed research; Antoine Persoons, Sébastien Duplessis, Emmanuelle Morin, Stéphane De Mita, and Thibaut Payen analyzed data; Antoine Persoons and Sébastien Duplessis drafted the manuscript and, Antoine Persoons, Sébastien Duplessis, Pascal Frey, Fabien Halkett, Thibaut Payen, and Stéphane De Mita wrote the paper.

ACKNOWLEDGMENTS

We warmly thank Katherine Hayden for comments on the manuscript. We would like to acknowledge the help of Bénédicte

Fabre (INRA Nancy) in the production of poplar plants in greenhouses and of *M. larici-populina* urediniospores. We also thank our colleagues Claude Murat and Francis Martin at INRA Nancy for fruitful discussions during the course of the study. This work was supported by the French National Research Agency through the Laboratory of Excellence ARBRE (ANR-12-LABXARBRE-01), the Young Scientist Grant POPRUST to Sébastien Duplessis (ANR-2010-JCJC-1709-01) and the GANDALF project (ANR-12-ADAP0009) and by the Région Lorraine (Researcher Award to Sébastien Duplessis). Antoine Persoons is supported by a Doctoral Scholarship from the Institut National de la Recherche Agronomique and the Region Lorraine. We thank the Joint Genome Institute for the access to the *M. larici-populina* genome sequence.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://www.frontiersin.org/journal/10.3389/fpls.2014.00450/abstract>

REFERENCES

- Alfano, J. R. (2009). Roadmap for future research on plant pathogen effectors. *Mol. Plant Pathol.* 10, 805–813. doi: 10.1111/j.1364-3703.2009.00588.x
- Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., et al. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25, 3389–3402.
- Au, C. H., Cheung, M. K., Wong, M. C., Chu, A. K., Law, P. T., and Kwan, H. S. (2013). Rapid genotyping by low-coverage resequencing to construct genetic linkage maps of fungi: a case study in *Leptinula edodes*. *BMC Res. Notes* 6:307. doi: 10.1186/1756-0500-6-307
- Ballou, F., Lehmann, L., and De Meeus, T. (2003). The population genetics of clonal and partially clonal diploids. *Genetics* 164, 1635–1644.
- Barrès, B., Halkett, F., Dutech, C., Andrieux, A., Pinon, J., and Frey, P. (2008). Genetic structure of the poplar rust fungus *Melampsora larici-populina*: evidence for isolation by distance in Europe and recent founder effects overseas. *Infect. Genet. Evol.* 8, 577–587. doi: 10.1016/j.meegid.2008.04.005
- Barrett, L. G., Thrall, P. H., Dodds, P. N., van der Merwe, M., Linde, C. C., Lawrence, G. J., et al. (2009). Diversity and evolution of effector loci in natural populations of the plant pathogen *Melampsora lini*. *Mol. Biol. Evol.* 26, 2499–2513. doi: 10.1093/molbev/msp166
- Bruce, M., Neugebauer, K. A., Joly, D. L., Migeon, P., Cuomo, C. A., Wang, S., et al. (2014). Using transcription of six *Puccinia triticina* races to identify the effective secretome during infection of wheat. *Front. Plant Sci.* 4:520. doi: 10.3389/fpls.2013.00520
- Cantu, D., Govindarajulu, M., Kozik, A., Wang, M., Chen, X., Kojima, K. K., et al. (2011). Next generation sequencing provides rapid access to the genome of *Puccinia striiformis* f. sp. *tritici*, the causal agent of wheat stripe rust. *PLoS ONE* 6:e24230. doi: 10.1371/journal.pone.0024230
- Cantu, D., Segovia, V., MacLean, D., Bayles, R., Chen, X., Kamoun, S., et al. (2013). Genome analyses of the wheat yellow (stripe) rust pathogen *Puccinia striiformis* f. sp. *tritici* reveal polymorphic and haustorial expressed secreted proteins as candidate effectors. *BMC Genomics* 14:270. doi: 10.1186/1471-2164-14-270
- Catanzariti, A.-M., Dodds, P. N., Lawrence, G. J., Ayliffe, M. A., and Ellis, J. G. (2006). Haustorially expressed secreted proteins from flax rust are highly enriched for avirulence elicitors. *Plant Cell* 18, 243–256. doi: 10.1105/tpc.105.035980
- Catanzariti, A.-M., Dodds, P. N., Ve, T., Kobe, B., Ellis, J. G., and Staskawicz, B. J. (2010). The AvrM effector from flax rust has a structured C-terminal domain and interacts directly with the M resistance protein. *Mol. Plant Microbe Interact.* 23, 49–57. doi: 10.1094/MPMI-23-1-0049
- Cooke, D. E., Cano, L. M., Raffaele, S., Bain, R. A., Cooke, L. R., Etherington, G. J., et al. (2012). Genome analyses of an aggressive and invasive lineage of the Irish potato famine pathogen. *PLoS Pathog.* 8:e1002940. doi: 10.1371/journal.ppat.1002940

- Crooks, G. E., Hon, G., Chandonia, J. M., and Brenner, S. E., (2004). WebLogo: a sequence logo generator. *Genome Res.* 14, 1188–1190. doi: 10.1101/gr.849004
- De Mita, S., and Siol, M. (2012). EggLib: processing, analysis and simulation tools for population genetics and genomics. *BMC Genet.* 11, 13–27. doi: 10.1186/1471-2156-13-27
- Dodds, P. N., Lawrence, G. J., Catanzariti, A.-M., Ayliffe, M. A., and Ellis, J. G. (2004). The *Melampsora lini* AvrL567 avirulence genes are expressed in haustoria and their products are recognized inside plant cells. *Plant Cell.* 16, 755–768. doi: 10.1105/tpc.020040
- Dodds, P. N., Lawrence, G. J., Catanzariti, A.-M., Teh, T., Wang, C. I., Ayliffe, M. A., et al. (2006). Direct protein interaction underlies gene-for-gene specificity and coevolution of the flax resistance genes and flax rust avirulence genes. *Proc. Natl. Acad. Sci. U.S.A.* 103, 8888–8893. doi: 10.1073/pnas.0602577103
- Dodds, P. N., and Rathjen, J. P. (2010). Plant immunity: towards an integrated view of plant-pathogen interactions. *Nat. Rev. Genet.* 11, 539–548. doi: 10.1038/nrg2812
- Duplessis, S., Bakkeren, G., and Hamelin, R. (2014b). Advancing knowledge on biology of rust fungi through genomics. *Adv. Bot. Res.* 70, 173–209. doi: 10.1016/B978-0-12-397940-7.00006-9
- Duplessis, S., Cuomo, C. A., Lin, Y. C., Aerts, A., Tisserant, E., Veneault-Fourrey, C., et al. (2011a). Obligate biotrophy features unraveled by the genomic analysis of rust fungi. *Proc. Natl. Acad. Sci. U.S.A.* 108, 9166–9171. doi: 10.1073/pnas.1019315108
- Duplessis, S., Hacquard, S., Delaruelle, C., Tisserant, E., Frey, P., Martin, F., et al. (2011b). *Melampsora larici-populina* transcript profiling during germination and timecourse infection of poplar leaves reveals dynamic expression patterns associated with virulence and biotrophy. *Mol. Plant Microbe Interact.* 24, 808–818. doi: 10.1094/MPMI-01-11-0006
- Duplessis, S., Spanu, P. D., and Schirawski, J. (2014a). “Biotrophic fungi (powdery mildews, Rusts and Smuts),” in *Ecological Genomics of the Fungi. Plant-Interacting Fungi Section*, ed F. Martin (Oxford: Wiley-Blackwell), 149–168.
- Fernandez, D., Tisserant, E., Talhinhas, P., Azinheira, H., Vieira, A., Petitot, A. S., et al. (2012). 454-pyrosequencing of *Coffea arabica* leaves infected by the rust fungus *Hemileia vastatrix* reveals in planta-expressed pathogen-secreted proteins and plant functions in a late compatible plant-rust interaction. *Mol. Plant Pathol.* 13, 17–37. doi: 10.1111/j.1364-3703.2011.00723
- Flor, H. H. (1971). Current status of the gene-for-gene concept. *Annu. Rev. Phytopathol.* 9, 275–296.
- Fraser, J. A., Hsueh, Y. P., Findley, K. M., and Heiman, J. (2007). “Evolution of the mating type locus: the basidiomycetes,” in *Sex in Fungi: Molecular Determination and Evolutionary Implications*, eds J. Heitman, J. W. Kronstad, J. W. Taylor, and L. A. Casselton (Washington, DC: ASM Press), 19–34.
- Gérard, P. R., Husson, C., Pinon, J., and Frey, P. (2006). Comparison of genetic and virulence diversity of *Melampsora larici-populina* populations on wild and cultivated poplar and influence of the alternate host. *Phytopathology* 96, 1027–1036. doi: 10.1094/PHYTO-96-1027
- Giraldo, M. C., and Valent, B. (2013). Filamentous plant pathogen effectors in action. *Nat. Rev. Microbiol.* 11, 800–814. doi: 10.1038/nrmicro3119
- Goellner, K., Loehrer, M., Langenbach, C., Conrath, U. W. E., Koch, E., and Schaffrath, U. (2010). *Phakopsora pachyrhizi*, the causal agent of Asian soybean rust. *Mol. Plant Pathol.* 11, 169–177. doi: 10.1111/j.1364-3703.2009.00589.x
- Hacquard, S., Delaruelle, C., Frey, P., Tisserant, E., Kohler, A., and Duplessis, S. (2013a). Transcriptome analysis of poplar rust telia reveals overwintering adaptation and tightly coordinated karyogamy and meiosis processes. *Front. Plant Sci.* 4:456. doi: 10.3389/fpls.2013.00456
- Hacquard, S., Joly, D. L., Lin, Y. C., Tisserant, E., Feau, N., Delaruelle, C., et al. (2012). A comprehensive analysis of genes encoding small secreted proteins identifies candidate effectors in *Melampsora larici-populina* (Poplar Leaf Rust). *Mol. Plant Microbe Interact.* 25, 279–293. doi: 10.1094/MPMI-09-11-0238
- Hacquard, S., Kracher, B., Maekawa, T., Vernaldi, S., Schulze-Lefert, P., and Ver Loren van Themaat, E. (2013b). Mosaic genome structure of the barley powdery mildew pathogen and conservation of transcriptional programs in divergent hosts. *Proc. Natl. Acad. Sci. U.S.A.* 110, E2219–E2228. doi: 10.1073/pnas.1306807110
- Hacquard, S., Petre, B., Frey, P., Hecker, A., Rouhier, N., and Duplessis, S. (2011). The poplar-poplar rust interaction: insights from genomics and transcriptomics. *J. Pathog.* 2011, 716041. doi: 10.4061/2011/716041
- Halkett, F., Simon, J.-C., and Balloux, F. (2005). Tackling the population genetics of clonal and partially clonal organisms. *Trends Ecol. Evol.* 20, 194–201. doi: 10.1016/j.tree.2005.01.001
- Hane, J. K., Anderson, J. P., Williams, A. H., Sperschneider, J., and Singh, K. B. (2014). Genome sequencing and comparative genomics of the broad host-range pathogen *Rhizoctonia solani* AG8. *PLoS Genet.* 10:e1004281. doi: 10.1371/journal.pgen.1004281
- Joly, D. L., Feau, N., Tanguay, P., and Hamelin, R. C. (2010). Comparative analysis of secreted protein evolution using expressed sequence tags from four poplar leaf rusts (*Melampsora* spp.). *BMC Genomics* 11:422. doi: 10.1186/1471-2164-11-422
- Kemen, E., Kemen, A. C., Rafiqi, M., Hempel, U., Mendgen, K., Hahn, M., et al. (2005). Identification of a protein from rust fungi transferred from haustoria into infected plant cells. *Mol. Plant Microbe Interact.* 18, 1130–1139. doi: 10.1094/MPMI-18-1130
- Kemen, E., Kemen, A., Ehlers, A., Voegele, R., and Mendgen, K. (2013). A novel structural effector from rust fungi is capable of fibril formation. *Plant J.* 75, 767–780. doi: 10.1111/tpj.12237
- Liti, G., Carter, D. M., Moses, A. M., Warringer, J., Parts, L., James, S. A., et al. (2009). Population genomics of domestic and wild yeasts. *Nature* 458, 337–341. doi: 10.1038/nature07743
- Lowe, R. G., and Howlett, B. J. (2012). Indifferent, affectionate, or deceitful: lifestyles and secretomes of fungi. *PLoS Pathog.* 8:e1002515. doi: 10.1371/journal.ppat.1002515
- Manning, V. A., Pandelova, I., Dhillon, B., Wilhelm, L. J., Goodwin, S. B., Berlin, A. M., 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. doi: 10.1534/g3.112.004044
- Nemri, A., Saunders, D. G. O., Anderson, C., Upadhyaya, N. M., Win, J., Lawrence, G. J., et al. (2014). The genome sequence and effector complement of the flax rust pathogen *Melampsora lini*. *Front. Plant Sci.* 5:98. doi: 10.3389/fpls.2014.00098
- Pinon, J., and Frey, P. (2005). “Interactions between poplar clones and *Melampsora* populations and their implications for breeding for durable resistance,” in *Rust Diseases of Willow and Poplar*, eds M. H. Pei and A. R. McCracken (Wallingford: CAB International), 139–154.
- Pretsch, K., Kemen, A., Kemen, E., Geiger, M., Mendgen, K., and Voegele, R. (2013). The rust transferred proteins—a new family of effector proteins exhibiting protease inhibitor function. *Mol. Plant Pathol.* 14, 96–107. doi: 10.1111/j.1364-3703.2012.00832
- Raffaele, S., and Kamoun, S. (2012). Genome evolution in filamentous plant pathogens: why bigger can be better. *Nat. Rev. Microbiol.* 10, 417–430. doi: 10.1038/nrmicro2790
- Ravensdale, M., Nemri, A., Thrall, P. H., Ellis, J. G., and Dodds, P. N. (2011). Co-evolutionary interactions between host resistance and pathogen effector genes in flax rust disease. *Mol. Plant Pathol.* 12, 93–102. doi: 10.1111/j.1364-3703.2010.00657
- Rodrigues, C. J. Jr., Bettencourt, A. J., and Rijo, L. (1975). Races of the pathogen and resistance to coffee rust. *Annu. Rev. Phytopathol.* 13, 49–70.
- Roy, S., Kagda, M., and Judelson, H. S. (2013a). Genome-wide prediction and functional validation of promoter motifs regulating gene expression in spore and infection stages of *Phytophthora infestans*. *PLoS Pathog.* 9:e1003182. doi: 10.1371/journal.ppat.1003182
- Roy, S., Poidevin, L., Jiang, T., and Judelson, H. S. (2013b). Novel core promoter elements in the oomycete pathogen *Phytophthora infestans* and their influence on expression detected by genome-wide analysis. *BMC Genomics* 14:106. doi: 10.1186/1471-2164-14-106
- Saunders, D. G. O., Win, J., Cano, L. M., Szabo, L. J., Kamoun, S., and Raffaele, S. (2012). Using hierarchical clustering of secreted protein families to classify and rank candidate effectors of rust fungi. *PLoS ONE* 7:e29847. doi: 10.1371/journal.pone.0029847
- Seidl, M. F., and Thomma, B. P. H. J. (2014). Sex or no sex: evolutionary adaptation occurs regardless. *Bioessays* 36, 335–345. doi: 10.1002/bies.201300155
- Seidl, M. F., Wang, R.-P., Van den Ackerveken, G., Govers, F., and Snel, B. (2012). Bioinformatic inference of specific and general transcription factor binding sites in the plant pathogen *Phytophthora infestans*. *PLoS ONE* 7:e51295. doi: 10.1371/journal.pone.0051295

- Soyer, J. L., El Ghalid, M., Glaser, N., Ollivier, B., Linglin, J., Grandaubert, J., et al. (2014). Epigenetic control of effector gene expression in the plant pathogenic fungus *Leptosphaeria maculans*. *PLoS Genet.* 10:e1004227. doi: 10.1371/journal.pgen.1004227
- Spanu, P. D. (2012). The genomics of obligate (and nonobligate) biotrophs. *Annu. Rev. Phytopathol.* 50, 91–109. doi: 10.1146/annurev-phyto-081211-173024
- Steenackers, J., Steenackers, M., Steenackers, V., and Stevens, M. (1996). Poplar diseases, consequences on growth and wood quality. *Biomass Bioenerg.* 10, 267–274.
- Stergiopoulos, I., Cordovez, V., Okmen, B., Beenen, H. G., Kema, G. H., and de Wit, P. J. (2013). Positive selection and intragenic recombination contribute to high allelic diversity in effector genes of *Mycosphaerella fijiensis*, causal agent of the black leaf streak disease of banana. *Mol. Plant Pathol.* 15, 447–460. doi: 10.1111/mpp.12104
- Stergiopoulos, I., and de Wit, P. J. (2009). Fungal effector proteins. *Annu. Rev. Phytopathol.* 47, 233–263. doi: 10.1146/annurev.phyto.112408.132637
- Stone, C. L., Buitrago, M. L., Boore, J. L., and Frederick, R. D. (2010). Analysis of the complete mitochondrial genome sequences of the soybean rust pathogens *Phakopsora pachyrhizi* and *P. meibomia*. *Mycologia* 102, 887–897. doi: 10.3852/09-198
- Stukenbrock, E. H. (2013). Evolution, selection and isolation: a genomic view of speciation in fungal plant pathogens. *New Phytol.* 199, 895–907. doi: 10.1111/nph.12374
- Stukenbrock, E. H., and Bataillon, T. (2012). A population genomics perspective on the emergence and adaptation of new plant pathogens in agro-ecosystems. *PLoS Pathog.* 8:e1002893. doi: 10.1371/journal.ppat.1002893
- Stukenbrock, E. H., and McDonald, B. A. (2009). Population genetics of fungal and oomycete effectors involved in gene-for-gene interactions. *Mol. Plant Microbe Interact.* 22, 371–380. doi: 10.1094/MPMI-22-4-0371
- Tatusov, R. L., Fedorova, N. D., Jackson, J. D., Jacobs, A. R., Kiryutin, B., Koonin, E. V., et al. (2003). The COG database: an updated version includes eukaryotes. *BMC Bioinform.* 4:41. doi: 10.1186/1471-2105-4-41
- Thompson, J. D., Gibson, T. J., and Higgins, D. G. (2002). Multiple sequence alignment using ClustalW and ClustalX. *Curr. Protoc. Bioinform. Chapter 2:Unit 2.3*. doi: 10.1002/0471250953.bi0203s00
- Tremblay, A., Hosseini, P., Li, S., Alkharouf, N. W., and Matthews, B. F. (2013). Analysis of *Phakopsora pachyrhizi* transcript abundance in critical pathways at four time-points during infection of a susceptible soybean cultivar using deep sequencing. *BMC Genomics* 14:614. doi: 10.1186/1471-2164-14-614
- Tuskan, G. A., Difazio, S., Jansson, S., Bohlmann, J., Grigoriev, I., Hellsten, U., et al. (2006). The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). *Science* 313, 1596–1604. doi: 10.1126/science.1128691
- Tyler, B. M., and Rouxel, T. (2012). “Effectors of fungi and oomycetes: their virulence and avirulence functions and translocation from pathogen to host cells,” in *Molecular Plant Immunity*, ed G. Sessa (Oxford: Wiley-Blackwell), 123–167. doi: 10.1002/9781118481431.ch7
- Upadhyaya, N. M., Mago, R., Staskawicz, B. J., Ayliffe, M. A., Ellis, J. G., and Dodds, P. N. (2014). A bacterial type III secretion assay for delivery of fungal effector proteins into wheat. *Mol. Plant Microbe Interact.* 27, 255–264. doi: 10.1094/MPMI-07-13-0187-FI
- Van der Merwe, M. M., Kinnear, M. W., Barrett, L. G., Dodds, P. N., Ericson, L., Thrall, P. H., et al. (2009). Positive selection in AvrP4 avirulence gene homologues across the genus *Melampsora*. *Proc. Biol. Sci.* 276, 2913–2922. doi: 10.1098/rspb.2009.0328
- Ve, T., Williams, S. J., Catanzariti, A.-M., Rafiqi, M., Rahman, M., Ellis, J. G., et al. (2013). Structures of the flax-rust effector AvrM reveal insights into the molecular basis of plant-cell entry and effector-triggered immunity. *Proc. Natl. Acad. Sci. U.S.A.* 11, 17594–17599. doi: 10.1073/pnas.1307614110
- Wicker, T., Oberhaensli, S., Parlange, F., Buchmann, J. P., Shatalina, M., Roffler, S., et al. (2013). The wheat powdery mildew genome shows the unique evolution of an obligate biotroph. *Nat. Genet.* 45, 1092–1096. doi: 10.1038/ng.2704
- Win, J., Chaparro-Garcia, A., Belhaj, K., Saunders, D. G., Yoshida, K., Dong, S., et al. (2012). Effector biology of plant-associated organisms: concepts and perspectives. *Cold Spring Harb. Symp. Quant. Biol.* 77, 235–247. doi: 10.1101/sqb.2012.77.015933
- Xhaard, C., Fabre, B., Andrieux, A., Gladieux, P., Barrès, B., Frey, P., et al. (2011). The genetic structure of the plant pathogenic fungus *Melampsora larici-populina* on its wild host is extensively impacted by host domestication. *Mol. Ecol.* 20, 2739–2755. doi: 10.1111/j.1365-294X.2011.05138
- Yoshida, K., Saitoh, H., Fujisawa, S., Kanzaki, H., Matsumura, H., Yoshida, K., et al. (2009). Association genetics reveals three novel avirulence genes from the rice blast fungal pathogen *Magnaporthe oryzae*. *Plant Cell* 21, 1573–1591. doi: 10.1105/tpc.109.066324
- Zander, M., Patel, D. A., Van de Wouw, A., Lai, K., Lorenc, M. T., Campbell, E., et al. (2013). Identifying genetic diversity of avirulence genes in *Leptosphaeria maculans* using whole genome sequencing. *Funct. Integr. Genomics* 13, 295–308. doi: 10.1007/s10142-013-0324-5
- Zheng, W., Huang, L., Huang, J., Wang, X., Chen, X., Zhao, J., et al. (2013). High genome heterozygosity and endemic genetic recombination in the wheat stripe rust fungus. *Nat. Commun.* 4, 2673. doi: 10.1038/ncomms3673

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

Received: 30 May 2014; accepted: 20 August 2014; published online: 15 September 2014.

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

This article was submitted to *Plant-Microbe Interaction*, a section of the journal *Frontiers in Plant Science*.

Copyright © 2014 Persoons, Morin, Delaruelle, Payen, Halkett, Frey, De Mita and Duplessis. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Résumé

Les truffes sont des champignons ectomycorhiziens du genre *Tuber*, au sein des Pézizomycètes, vivant en symbiose avec de nombreux arbres et arbustes. Parmi les Pézizomycètes se retrouvent des espèces saprotrophes, pathogènes et symbiotiques ainsi que des champignons très connus comme les truffes et les morilles. Malgré leur intérêt, la structure et l'évolution des génomes des champignons de la classe des Pézizomycètes est encore mal connue. Les objectifs de ma thèse étaient de caractériser la structure et l'évolution des génomes de Truffes et d'autres Pézizomycètes. Une analyse de génomique comparative entre huit génomes de Pézizomycètes, dont trois de truffes, a montré que l'expansion de la taille des génomes des truffes, due principalement aux rétrotransposons gypsy, a probablement eu lieu chez l'ancêtre commun des Tuberaceae il y a environ 150 millions d'années. Ensuite, une réduction de leur contenu en gènes, et principalement en enzymes dégradant la paroi végétale, a eu lieu chez l'ancêtre des *Tuber* il y a environ 100 millions d'années. Des convergences et des divergences évolutives ont été mise en évidence entre les champignons ectomycorhiziens basidiomycètes et les truffes. Une analyse de génomique comparative au sein de l'espèce *T. melanosporum* a quant à elle permis de caractériser une ressource de plus de 400000 polymorphismes et de mettre en évidence des traces de sélection. Enfin une analyse plus détaillée des rétrotransposons gypsy présents dans le génome de *T. melanosporum* a montré que leur colonisation du génome est ancienne. Certains éléments semblent toutefois avoir transposé récemment suggérant que les rétrotransposons gypsy peuvent continuer de jouer un rôle dans l'évolution du génome de *T. melanosporum*.

Mots clefs : *Tuber melanosporum*, truffes, Pézizomycètes, bioinformatique, génomique comparative

Abstract

Truffles are ectomycorrhizal (ECM) fungi, belonging to *Tuber* genus in the Pezizomycete Class, that form symbiotic associations with numerous trees and shrubs. The Pezizomycetes constitute an early diverging lineage of Ascomycota composed of saprophytic, mycorrhizal and pathogenic species. Some Pezizomycetes, such as truffles and morels, are widely recognized by scientist and lay-person alike. Despite their importance, the genomic structure and evolution within the Pezizomycetes is largely unknown. The aims of my thesis were to characterize the structure and the evolution of the truffle and other Pezizomycete genomes. A comparative analysis performed using eight Pezizomycete fungi, among them three truffles, showed that the genome of truffles has evolved by a size expansion, mainly due to gypsy retrotransposons, in the common ancestor of Tuberaceae about 150 Mya. Then, a loss of the number of gene models, such as plant cell wall degrading enzymes occurred in the common ancestor of *Tuber* spp about 100 Mya. This study provides new insights into the evolution of the truffles, and ECM symbiosis in general, and highlights cases of divergence and convergence between Basidiomycota and truffle symbiotic species. A comparative analysis in *T. melanosporum* allowed characterizing a resource of more than 400,000 single nucleotide polymorphisms. This population genomic analysis allowed identifying candidate genomic regions with trace of selection. Finally, a more detailed characterization of *T. melanosporum* gypsy retrotransposons showed that the major invasive waves are old, but that more recently some elements have moved suggesting that retrotransposons continue to play an important role in the evolution of the truffle genome.

Keywords: *Tuber melanosporum*, truffles, Pezizomycete, bioinformatic, comparative genomic