Pezizomycetes genomes reveal the molecular basis of ectomycorrhizal truffle lifestyle

To cite this version:

HAL Id: hal-02094494
https://amu.hal.science/hal-02094494
Submitted on 12 Jun 2019

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.

Distributed under a Creative Commons Attribution 4.0 International License
Pezizomycetes genomes reveal the molecular basis of ectomycorrhizal truffle lifestyle

Claude Murat1,2,25*, Thibaut Payen1,2,25, Benjamin Noel2, Alan Kuo3, Emmanuelle Morin4,1, Juan Chen1,4, Annegret Kohler1, Krisztina Krizsán5, Raffaella Balestrini6, Corinne Da Silva2, Barbara Montanini7, Mathieu Hainaut8, Elisabetta Levati7, Kerrie W. Barry3, Beatrice Belfiori9, Nicolas Cichocki1, Alicia Clum3, Rhyan B. Dockter3, Laure Fauchery1, Julie Guy2, Mirco Iotti10, François Le Tacon1, Erika A. Lindquist8, Anna Lipzen3, Fabienne Malagnac11, Antonietta Mello6, Virginie Molinier12,13, Kerrie W. Barry3, Beatrice Belfiori9, Nicolas Cichocki1, Julie Poulain2, Claudia Riccioni9, Andrea Rubini9, Yaron Sitrit14, Richard Spivallo15, Stefanie Traeger16, Mei Wang2, Lucia Žifčáková17, Daniel Wipf12, Alessandra Zambonelli18, Francesco Paolocci9, Minou Nowrousian16, Simone Ottonello3, Petr Baldrian17, Joseph W. Spatafora19, Bernard Henriessat8,20,21, Laszlo G. Nagy8, Jean-Marc Aury2, Patrick Wincker4, Igor V. Grigoriev3,22, Paola Bonfante23 and Francis M. Martin1,2,24*

Tuberaceae is one of the most diverse lineages of symbiotic truffle-forming fungi. To understand the molecular underpinning of the ectomycorrhizal truffle lifestyle, we compared the genomes of Piedmont white truffle (Tuber magnatum), Périgord black truffle (Tuber melanosporum), Burgundy truffle (Tuber aestivum), pig truffle (Choiromyces venosus) and desert truffle (Terfezia boudieri) to saprotrophic Pezizomycetes. Reconstructed gene duplication/loss histories along a time-calibrated phylogeny of Ascomycetes revealed that Tuberaceae-specific traits may be related to a higher gene diversification rate. Genomic features in Tuber species appear to be very similar, with high transposon content, few genes coding lignocellulose-degrading enzymes, a substantial set of lineage-specific fruiting-body-upregulated genes and high expression of genes involved in volatile organic compound metabolism. Developmental and metabolic pathways expressed in ectomycorrhizae and fruiting bodies of T. magnatum and T. melanosporum are unexpectedly very similar, owing to the fact that they diverged ~100 Ma. Volatile organic compounds from pungent truffle odours are not the products of Tuber-specific gene innovations, but rely on the differential expression of an existing gene repertoire. These genomic resources will help to address fundamental questions in the evolution of the truffle lifestyle and the ecology of fungi that have been praised as food delicacies for centuries.

Truffle fungi differentiate into subterranean fruiting bodies bearing spores sequestered in an inconspicuous globose-like mass of hyphae. Truffle-forming species have evolved in nearly every major group of fleshy fungi over 100 times independently within the Ascomycota and Basidiomycota and the majority of evolutionary transitions to a truffle morphology have occurred in lineages establishing ectomycorrhizal mutualistic symbiosis with plants1. This pattern suggests that symbiosis has been a major driver in the evolution of truffle diversity. The evolution of the hypogeous lifestyle across a diversity of truffle lineages also suggests that the transition from epigeous to hypogeous fruiting is driven by strong selection for traits, for example pungent odours, that promote...
animal dispersal. The Tuberaceae family is one of the most diverse lineages of mainly truffle-forming fungi and is presumably one of the earliest diverging clades within the Pezizomycetes (Ascomycota). Truffles, such as the aromatic Périgord black truffle (T. melanosporum Vittad.), Burgundy truffle (T. aestivum Vittad.), Piedmont white truffle (T. boudieri Chatin), and desert truffle (T. pico Tul & C. Tul), have been praised as food delicacies for centuries. Despite their ecological importance and a long history as a gourmet food, there are still many unanswered questions concerning the biology, genetics and ecology of Tuberaceae.

Previously, the genome of T. melanosporum was sequenced and revealed a set of unusual genomic features, including a very high proportion of transposable elements (PCWDEs) and other repeated elements, a substantial gene co-linearity (microsynteny) among Tuberaceae genomes, and saprotrophic black morel (M. importuna) and desert truffle (T. pico) and compared their genomes to Tuberaceae genomes. Predicted protein-coding content ranged from 9,344 for T. aestivum to 17,986 for C. venosus (Fig. 1c and Table 1).

Phylogenetic reconstruction using 2,093 concatenated conserved single-copy protein-coding genes from the eight Pezizomycetes corroborees previously reported topologies (Fig. 1a) with Tuber spp. clustering together and C. venosus being phylogenetically distinct from Tuber spp. within Tuberaceae. We estimated the age of the most recent common ancestor (MRCA) of Pezizomycetes at 470 ± 67 million years (My), in accordance with previous estimations. We also determined that Tuberaceae began diverging around 140 ± 10 Ma in the Early Cretaceous, which corresponds to the time when the Angiosperms evolved.

A comparison of Tuberaceae genomes revealed no evidence of whole genome duplication events (that is, no segmental duplications). Despite the high proportion of scattered transposable elements and other repeated elements, a substantial gene co-linearity (microsynteny) among Tuberaceae genomes was apparent (Supplementary Fig. 1). As expected, there is a high correlation between the number of syntenic regions and the estimated age of divergence between the various clades (Supplementary Fig. 2).

### Invasion of Tuberaceae genomes by transposable elements
Transposable element content ranged from 3.5% in A. immersus to 58% in T. magnatum and T. melanosporum (Table 1 and Supplementary Fig. 3). There was a strong association (Pearson’s, $r = 0.83$, $P = 0.01$) between genome size and the proportion of repeated elements across species (Fig. 1b and Supplementary Fig. 4), confirming that genome expansion in Tuberaceae is mainly driven by transposable element proliferation. Genome size, and consequently proportion of transposable elements, negatively correlates with the size of the gene repertoire (Pearson’s, $r = -0.92$, $P = 0.001$; Supplementary Fig. 3), highlighting the impact of transposable elements on the coding space. Gypsy long-terminal repeat (LTR) retrotransposons are the most frequent known transposable elements, representing up to 46% of the T. magnatum genome (Supplementary Fig. 3). Their age distribution shows that they have been proliferating over the past 5 to 6 Myr for T. melanosporum, T. aestivum and C. venosus, while LTR accumulation in T. magnatum is more ancient and occurred between 6 and 14 Ma (Supplementary Fig. 5). As a result of LTR sequence degeneration, very ancient invasions of

### Genome features and phylogeny of Tuberaceae
Nuclear genomes of the ectomycorrhizal truffles C. venosus, T. aestivum and T. magnatum in Tuberaceae and C. boudieri in Terfeziaceae, as well as of saprotrophic Pezizomycetes M. importuna and A. immersus, were sequenced (Table 1 and Supplementary Information). Genomes ranged from 48 Mb (M. importuna) to 192 Mb (T. magnatum) (Table 1 and Fig. 1b). Tuberaceae genomes are significantly larger than T. boudieri and saprotrophic Pezizomycetes genomes. Predicted protein-coding content ranged from 9,344 for T. aestivum to 17,986 for C. venosus (Fig. 1c and Table 1).

Phylogenetic reconstruction using 2,093 concatenated conserved single-copy protein-coding genes from the eight Pezizomycetes corroborees previously reported topologies (Fig. 1a) with Tuber spp. clustering together and C. venosus being phylogenetically distinct from Tuber spp. within Tuberaceae. We estimated the age of the most recent common ancestor (MRCA) of Pezizomycetes at 470 ± 67 million years (My), in accordance with previous estimations. We also determined that Tuberaceae began diverging around 140 ± 10 Ma in the Early Cretaceous, which corresponds to the time when the Angiosperms evolved.

A comparison of Tuberaceae genomes revealed no evidence of whole genome duplication events (that is, no segmental duplications). Despite the high proportion of scattered transposable elements and other repeated elements, a substantial gene co-linearity (microsynteny) among Tuberaceae genomes was apparent (Supplementary Fig. 1). As expected, there is a high correlation between the number of syntenic regions and the estimated age of divergence between the various clades (Supplementary Fig. 2).

### Invasion of Tuberaceae genomes by transposable elements
Transposable element content ranged from 3.5% in A. immersus to 58% in T. magnatum and T. melanosporum (Table 1 and Supplementary Fig. 3). There was a strong association (Pearson’s, $r = 0.83$, $P = 0.01$) between genome size and the proportion of repeated elements across species (Fig. 1b and Supplementary Fig. 4), confirming that genome expansion in Tuberaceae is mainly driven by transposable element proliferation. Genome size, and consequently proportion of transposable elements, negatively correlates with the size of the gene repertoire (Pearson’s, $r = -0.92$, $P = 0.001$; Supplementary Fig. 3), highlighting the impact of transposable elements on the coding space. Gypsy long-terminal repeat (LTR) retrotransposons are the most frequent known transposable elements, representing up to 46% of the T. magnatum genome (Supplementary Fig. 3). Their age distribution shows that they have been proliferating over the past 5 to 6 Myr for T. melanosporum, T. aestivum and C. venosus, while LTR accumulation in T. magnatum is more ancient and occurred between 6 and 14 Ma (Supplementary Fig. 5). As a result of LTR sequence degeneration, very ancient invasions of

### Genome features and phylogeny of Tuberaceae
Nuclear genomes of the ectomycorrhizal truffles C. venosus, T. aestivum and T. magnatum in Tuberaceae and C. boudieri in Terfeziaceae, as well as of saprotrophic Pezizomycetes M. importuna and A. immersus, were sequenced (Table 1 and Supplementary Information). Genomes ranged from 48 Mb (M. importuna) to 192 Mb (T. magnatum) (Table 1 and Fig. 1b). Tuberaceae genomes are significantly larger than T. boudieri and saprotrophic

### Table 1 | Genome assembly statistics for the eight Pezizomycetes

<table>
<thead>
<tr>
<th>Species Abbreviation</th>
<th>Tissue of origin</th>
<th>Assembly (Mb)</th>
<th>Scaffolds</th>
<th>N50 scaffold (Mb)</th>
<th>Gap (%)</th>
<th>Transposable element (%)</th>
<th>BUSCO</th>
<th>CEGMA</th>
<th>Sequencing</th>
<th>Annotation</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. immersus</td>
<td>Asci</td>
<td>Mycelium</td>
<td>60</td>
<td>706</td>
<td>0.27</td>
<td>11</td>
<td>3.5</td>
<td>17,877</td>
<td>288 (99%)</td>
<td>238 (96%)</td>
</tr>
<tr>
<td>C. venosus</td>
<td>Chove</td>
<td>Fruiting body</td>
<td>124</td>
<td>1,648</td>
<td>0.31</td>
<td>51</td>
<td>53</td>
<td>17,966</td>
<td>290 (100%)</td>
<td>240 (98%)</td>
</tr>
<tr>
<td>Morchella importuna</td>
<td>Morco</td>
<td>Mycelium</td>
<td>48</td>
<td>540</td>
<td>0.60</td>
<td>21</td>
<td>8.5</td>
<td>11,600</td>
<td>286 (98%)</td>
<td>235 (95%)</td>
</tr>
<tr>
<td>P. confluens</td>
<td>Pyroco</td>
<td>Mycelium</td>
<td>50</td>
<td>1,588</td>
<td>0.14</td>
<td>0.1</td>
<td>14</td>
<td>13,367</td>
<td>288 (99%)</td>
<td>236 (95%)</td>
</tr>
<tr>
<td>Te. boudieri</td>
<td>Terbo</td>
<td>Mycelium</td>
<td>63</td>
<td>516</td>
<td>0.78</td>
<td>5.7</td>
<td>32</td>
<td>10,200</td>
<td>290 (100%)</td>
<td>242 (98%)</td>
</tr>
<tr>
<td>T. aestivum</td>
<td>Tubae</td>
<td>Fruiting body</td>
<td>145</td>
<td>3,244</td>
<td>0.44</td>
<td>9.1</td>
<td>49.5</td>
<td>9,344</td>
<td>277 (95%)</td>
<td>239 (96%)</td>
</tr>
<tr>
<td>T. magnatum</td>
<td>Tubma</td>
<td>Fruiting body</td>
<td>192</td>
<td>1,283</td>
<td>1.81</td>
<td>13.4</td>
<td>58</td>
<td>9,433</td>
<td>286 (98%)</td>
<td>234 (94%)</td>
</tr>
<tr>
<td>T. melanosporum</td>
<td>Tubme</td>
<td>Mycelium</td>
<td>125</td>
<td>398</td>
<td>0.64</td>
<td>1.1</td>
<td>58</td>
<td>10,763</td>
<td>287 (99%)</td>
<td>239 (96%)</td>
</tr>
</tbody>
</table>

BUSCO, benchmarking universal single-copy orthologues; CEGMA, core eukaryotic genes mapping approach.
Species-specific genes exhibit the lowest values for most genetic features, suggesting that they evolve separately from the species-shared genes. The transcriptional RNA-Seq analysis revealed that 34% of the shared genes have a smaller gene size, fewer exons and a lower proportion of expressed genes than the conserved genes, suggesting a differential evolution, that is they might be evolutionarily young genes.

Accelerated rate of genome diversification in Tuberaceae

The new genomes, and more specifically Tuberaceae and Terfeziaceae genomes, offer a unique opportunity to examine the evolution of truffle-related and ectomycorrhiza-related genes, and gene families across the Pezizomycetes clade. Reconstructing genome-wide duplication and contraction events in Pezizomycetes, as well as in a large sample of additional fungal species including representatives of the major subphyla, revealed a considerable heterogeneity in the temporal dynamics of genome diversification between the different clades (Fig. 2). Approximately 1,200 gene duplications are marking the origin of Pezizomycetes, followed by gradual genome contractions in all Pezizomycete lineages. Of note, as few as 84 gene duplications were inferred for the common ancestor of Te. boudieri and A. immersus (Terfeziaceae + Ascobolaceae), whereas considerable species-specific gene diversification was observed in these species, which may be due to the low sampling density in this clade. On the contrary, branches leading to Tuberaceae showed on average 400 to 1,000 gene duplications and 800 to 1,200 gene losses per branch. Notably, we found that the rate of gene duplication within the genus Tuber was 1.7- to 2.2-fold higher than in the other lineages (Supplementary Fig. 8).

To determine whether there are significant differences between the shared and species-specific genes, we compared their protein length, exon and intron numbers, protein families database (Pfam) and transmembrane domain content, and signal peptide presence (Supplementary Table 2). According to our results, the averages of the above features of the two gene sets are significantly different. Species-specific genes exhibit the lowest values for most genetic features, suggesting that they evolve separately from the species-shared genes. The transcriptional RNA-Seq analysis revealed that 34% (T. melanosporum) to 87% (C. venosus) of species-specific genes have transcripts. Overall, the analysis of genetic features indicates that the two gene sets are noticeably distinct and that species-specific genes have a smaller gene size, fewer exons and a lower proportion of expressed genes than the conserved genes, suggesting a differential evolution, that is they might be evolutionarily young genes.

Core, dispensable and species-specific genes

To compare gene repertoires encoded by sequenced Pezizomycetes and identify species-specific gene families that might contribute to trait diversification, we clustered the predicted proteins to infer orthologous gene groups (orthogroups), including core genes (that is, occurring in the eight species), dispensable genes (that is, found in at least two species) and species-specific genes (that is, unique to a taxa). To define species-specific genes in a broader context, we then searched for orthologous sequences in >500 published fungal genomes. While the set of core genes is in the range of 2,443 to 2,511, the repertoire of species-specific genes is much lower in Tuber species compared to the other taxa (Supplementary Fig. 7). Species-specific genes, which are also referred to as taxonomically cations in the lineage leading to C. venosus, followed by gradual genome contractions in all Pezizomycete lineages. Of note, as few as 84 gene duplications were inferred for the common ancestor of Te. boudieri and A. immersus (Terfeziaceae + Ascobolaceae), whereas considerable species-specific gene diversification was observed in these species, which may be due to the low sampling density in this clade. On the contrary, branches leading to Tuberaceae showed on average 400 to 1,000 gene duplications and 800 to 1,200 gene losses per branch. Notably, we found that the rate of gene duplication within the genus Tuber was 1.7- to 2.2-fold higher than in the other lineages (Supplementary Fig. 8). In addition to Tuber species, we inferred >7,000 gene duplications in the lineage leading to C. venosus, although most of the duplicated genes are of unknown function (Supplementary Fig. 7). This latter lineage shows the highest rate of genome diversification among Pezizomycetes (~10-fold faster than that estimated for the other clades), although further sampling of species will be necessary to understand how duplications happened along this lineage.

Altogether, Tuberaceae has a significantly accelerated evolution of its gene repertoire (analysis of variance (ANOVA), P = 0.015) compared to other ascomycete species in this dataset, possibly reflecting an increased rate of genomic evolution underlying traits specific to either symbiosis and/or hypogeous fruiting-body formation.
ANOVA within the entire hypogeous Ascomycetes dataset (that is, Tuberaceae + Terfeziaceae) also indicated a significantly higher rate of gene duplication compared to non-hypogeous ascomycetes ($P = 0.014$). A similar, albeit slightly weaker trend was observed for the gene loss rates in the Tuberaceae, which is significantly higher than in the other clades ($P = 0.032$), indicating once again an accelerated gene turnover in these particular species (but not in the hypogeous Ascomycetes in general, $P = 0.061$). Of note, we also inferred a significant number of terminal duplication events in Tuber species (1,000 to 1,200 per branch, yielding orphan genes) (Supplementary Fig. 8), suggesting a possible role of species-specific adaptive changes in the evolution of the genus.

To gain insights into the unique features of Tuberaceae and Tuber genomes, we compared the frequencies of Gene Ontology annotations of genes that showed duplications in these clades. Although most of the duplicated genes specific to Tuberaceae had no functional annotation, several functions were found to be overrepresented compared to other Pezizomycetes in particular genes involved in DNA binding, transposition, RNA metabolism and biosynthesis of aromatic and organic cyclic compounds (Supplementary Fig. 9). These compounds contribute to the pungent flavours released by truffle fruiting bodies.

**Tuberaceae have a restricted set of PCWDEs**

In sequenced Pezizomycetes, the total number of secreted carbohydrate-active enzymes involved in plant polysaccharide degradation ranged from 49 for the symbiotic Te. boudieri to 203 for the coprophilic A. immersus (Fig. 3). As expected, based on their saprotrophic lifestyle, genomes of A. immersus, M. importuna and P. confluens encode a large repertoire of PCWDEs, including lignin--, cellulose-, hemicellulose- and pectin-degrading enzymes, as well as multiple carbohydrate-binding modules, indicating the potential to degrade a large set of substrates found in decaying soil organic matter and plant debris (for example, herbivore dung) (Supplementary Table 3). In contrast, these PCWDE genes are lacking, or occur at a very low level, in Tuber spp. and Te. boudieri (Supplementary Table 3).

The set of PCWDEs in truffles is as low as the repertoire reported from previously sequenced ectomycorrhizal basidiomycetes \(^{16}\) (Fig. 3). This corroborates a convergent decrease in plant cell wall degrading capacity of ectomycorrhizal ascomycetous and basidiomycetous symbionts\(^{16,17}\), resulting in a limited ability to decompose soil (plant) organic matter and host plant cell wall materials. DNA decay is probably the inactivation mechanism specifically driving the loss of GH6 in T. melanosporum (Supplementary Material, Supplementary Figs. 10 and 11).

Of note, genes coding for pectate lyase (PL1), rhamnogalacturonan endolyase (PL4) and pectin acetylesterase (CE12) are present in ectomycorrhizal Pezizomycetes, whereas they have been lost in ectomycorrhizal basidiomycetes, indicating that the former symbionts may use these enzymes for the colonization of the middle lamella of host roots. Intriguingly, the PCWDE repertoire of C. venosus is closer to saprotrophic taxa (for example P. confluens) than to ectomycorrhizal Tuber spp. Its substantial PCWDE repertoire suggests that C. venosus mycelium can also acquire carbon directly from soil organic matter, although the natural abundance \(^{15}N/^{13}C\) ratio of its fruiting body, a signature of carbon sources used for elaborating fruiting-body tissues, is typical of...
ectomycorrhizal symbionts, that is host photoassimilates are the major source of carbon.

On the other hand, Tuberaceae has a richer set of copper-dependent lytic polysaccharide monooxygenases involved in chitin (AA11) degradation (Supplementary Table 3), which suggests their ability to partly acquire nitrogen and carbon from the soil fungal necromass.

Shared symbiosis genes in truffle ectomycorrhizae

We found that 5,229 (75%) of the 6,952 orthologous T. magnatum/T. melanosporum genes were expressed in T. magnatum–Quercus robur and T. melanosporum–Corylus avellana ectomycorrhizae (Supplementary Fig. 12A), indicating that developmental and metabolic pathways associated with symbiosis in these truffles are very similar, despite their interaction with different host species. However, among orthologous genes, 218 and 253 displayed a species-specific expression in either T. melanosporum or T. magnatum, respectively (Supplementary Fig. 12A). Differential gene expression analysis of T. magnatum–Quercus robur ectomycorrhizae identified 514 upregulated and 928 downregulated genes, relative to free-living mycelium (fold-change > 5, false discovery rate (FDR), P < 0.05) (Supplementary Tables 4 and 5). Highly upregulated genes of known function include oxidoreductases (thioredoxin, cupredoxin, laccase, cytochrome P450), transcription factors, RNA metabolism-related proteins, and membrane transporters (carbohydrates, ammonium, amino acids, peptides) (Supplementary Table 4). As previously observed for T. melanosporum, root colonization by T. magnatum triggers the induction of a limited set of PCWDEs. Genes coding for
GH5 endoglucanases, GH28 polygalacturonase, CE12 carbohydrate esterase, GH43 arabinosidase and GH12 xyloglucan hydrolase are amongst the most upregulated genes (Supplementary Table 4), supporting their involvement in hemicellulose degradation during root apoplast colonization.

Most *T. magnatum* symbiosis-upregulated genes are conserved in other Pezizomycetes, including saprotrophic species (66%, clusters V and VI in Fig. 4a). These genes code for key metabolic and cellular functions (for example, membrane transporters, signalling proteins, oxidases, Zn-finger transcriptional factors) that are related to the symbiotic metabolism (for example, increased nutrient fluxes between symbionts) and symbiosis-related changes in hyphal development (Supplementary Table 4). Approximately one-third of symbiosis-induced genes are homologous to genes only present in Tuberaceae (clusters I, II and III). This large set of mycorrhiza-induced genes conserved in Tuberaceae indicates that similar ancient gene networks
are involved in development and functioning of ectomycorrhizae in Tuberaceae species (for example, Supplementary Table 7). This is in agreement with the fact that the ancestral ecology for Tuberaceae is presumed to be ectomycorrhizal because all of the extant species possess this ecology, and the MRCA of the family was probably ectomycorrhizal with angiosperm hosts at the end of the Jurassic period (156 Ma). A low proportion (15%) of T. magnatum symbiosis-upregulated genes are similarly induced in T. melanosporum–Corylus avellana ectomycorrhizae (Supplementary Fig. 12B, Supplementary Table 6). They are involved in core metabolic pathways required for an efficient symbiosis, such as carbohydrate, oligopeptide, amino acid and lipid transporters, or root colonization (for example, GH5 endoglucanase), but most of them (62–69%) have no known function (Supplementary Tables 6 and 7).

Shared fruiting-body-related genes in truffle

Although the vegetative mycelia of several Tuber species can be grown on agar medium and ectomycorrhizae produced by inoculated trees under controlled environmental conditions, truffle fruiting cannot be triggered in the laboratory. This is further complicated due to the heterothallic mating system of Tuber species (Supplementary materials, Supplementary Table 8). Therefore, fruiting bodies of T. magnatum were sampled in truffle grounds, their transcriptome analyzed by RNA sequencing and compared to published T. melanosporum transcript profiling. We found that 5,173 genes (74%) of orthologous T. magnatum/T. melanosporum genes were also expressed in T. magnatum fruiting bodies (Supplementary Fig. 12C). Most of these genes (5,010) are expressed in both truffles, indicating that developmental and metabolic pathways associated with the reproductive (sexual) stage of these two truffles are very similar. This is surprising, given that these species diverged ~100 Ma and belong to two separate clades.

A total of 489 upregulated and 1,073 downregulated genes (fold-change >5, FDR <0.05) were identified in T. magnatum fruiting bodies by comparison to free-living mycelium (Fig. 4b and Supplementary Tables 9, 10 and 12). Among upregulated genes, 121 (25%) are well conserved and share a sequence homology with sequenced Pezizomycetes (cluster VI); 163 (33%) have homologs in other Tuberaceae only (clusters I, III, IV and V). Among the most highly upregulated, shared genes, oxidoreductases (for example alcohol oxidase, cytochrome P450), membrane transporters and enzymes of sulfur metabolism (sulfate adenyltransferase, adenylylsulfate kinase) were found (Supplementary Tables 7 and 9).

As many as 178 genes (36%) are specific to T. magnatum (cluster VII). Up to 90% of the latter have no known function or conserved protein domain (Supplementary Table 12). Some of these species-specific genes may partly explain T. magnatum morphological and metabolic idiosyncrasies, although specificities in fruiting-body morphologies and truffle metabolism can also be driven by expression changes in a few conserved genes.

Expression of genes involved in the sexual reproduction machinery (that is, mating and meiosis) was assessed in several immature and mature T. melanosporum and T. magnatum fruiting bodies sampled in truffle orchards over several years. Despite a substantial transcript profile variability, probably reflecting the unavoidable heterogeneity inherent to environmental samples (Supplementary Fig. 13, Supplementary Table 13), the expression levels of these genes were overall remarkably similar between species and among samples.

Aroma biosynthesis in truffles

Truffle aroma comprises a complex mixture of VOCs. By combining in silico analysis of biosynthetic pathways with transcriptome profiling, we found that genomes of the Tuberaceae, but also those from the other sequenced Pezizomycetes, encode genes of biosynthetic pathways (for example sulfate reduction, amino acid and fatty acid catabolism) leading to truffle VOCs (Supplementary Table 14). However, specific gene expression patterns were characterized in several taxa by a closer comparison of their transcript profiles. Expression of genes coding for enzymes and transporters involved in specific (often rate-limiting) steps of sulfur-amino acid metabolism, such as sulfate reduction, the homocysteine-methionine cycle and VOC synthesis from methionine, is particularly sustained in Tuber species (Fig. 5), in line with the fact that sulfur volatiles are key constituents of truffle VOCs and play a crucial role in their sexual reproduction, that is sulfur-related volatile organic compounds (S-VOCs) released by truffles are attractive to rodents and truffle flies, which disperse their spores.

C. venosus shows a low expression of almost all of the S-VOC-related genes. This finding suggests that the unique flavour of the pig truffle may originate from a different set of pathways. In addition, the expression patterns of genes involved in the Ehrlich pathway, leading to alcohol and aldehyde VOCs, appear to be quite distinctive in most Tuber species (Supplementary Table 14). For instance, the higher expression levels of genes coding for branched-chain and aromatic amino acid transferases probably explain the preferential occurrence of specific VOCs (for example veratrole, anisole, 3-methyl-1-butanal) in T. melanosporum compared to the other truffles. Unexpectedly, specific VOC compositions of individual truffle species may thus be largely explained by the differential expression of selected subsets of metabolic genes, while variation in gene content and/or gene copy number appear to play a relatively minor role in Tuber aroma formation.

The analysis of the expression profiles of genes involved in fatty acid synthesis revealed significant similarities between Tuber species and M. importuna, yet failed to explain the quite unique volatile ester profile of T. melanosporum (Supplementary Table 14). Similarly, the production of 2,4-dithiapentane by T. magnatum could not be traced down to a specific gene repertoire, corroborating recent results showing that VOCs may actually be produced by the community of bacteria, yeasts and moulds specifically associated with different truffle species. Future challenges include the experimental assessment of the catalytic properties of specific enzymes involved in VOC biosynthesis and the elucidation of the role played by fruiting-body-associated microorganisms in aroma formation.

Conclusions

In Pezizomycetes, the transition from saprotrophic nutrition mode and epigeous fruiting body to the ectomycorrhizal lifestyle and truffle habit occurred at least twice in the MRCAs of Tuberaceae and Terfeziaceae. From a genomic point of view, Tuber species appear quite similar overall, with a strikingly high abundance of transposons, low protein-coding gene repertoires, restricted sets of PCWDEs, highly expressed genes involved in VOC synthesis, and conserved genes with no known function, preferentially expressed in fruiting bodies, and presumably associated with their formation. Some of the observed differences at the genome and transcriptome levels have probably contributed to the evolution of their specific phenotypic features. Our findings also suggest that the complex cocktail of Tuber VOCs is not the result of specific gene innovations, but mostly rely on the differential expression of existing gene repertoires. The present Tuberaceae genomes have provided significant insight into our understanding of the evolution of complex developmental and morphological traits. It also led to the identification of a number of genes specifically related to ectomycorrhizae and truffle development. A detailed understanding of the ontogenetic and metabolic roles played by these genes will require an in-depth functional characterization of their protein products.
**Methods**

**Strains and fungal material used for genome sequencing.** Genomic DNA used for sequencing *T. aestivum*, *T. magnatum* and *C. venosus* genomes were extracted from fruiting bodies (ascomata) sampled in truffle grounds at Montdoré (Haute-Saône, France), Montemagno (Piedmont, Italy) and Charmes State Forest (Vosges, France), respectively. Cultures of free-living mycelium were used for *M. importuna* respectively. Cultures of free-living mycelium were used for *A. immersus* and *Te. boudieri*. Strain RN42 (CBS 140666) of *A. immersus* was grown as described in Rhouim et al.²². Strain S1 (ATCC MYA-4762) of *Te. boudieri* was grown on cellophane sheet placed on a solid Fontana medium in the dark at 25°C for 1 month²³.

**Genome sequencing, assembly and annotation.** *T. melanoporum* and *P. confluens* genomes have previously been published²¹. Genomes of *A. immersus*, *M. importuna*, *Te. boudieri* and *C. venosus* were sequenced by the US Department of Energy Joint Genome Institute (JGI) (Table 1) using a combination of Illumina fragment (270 base pair (bp) insert size) and 4 kb long mate-pair (LMP) libraries, and assembled using ALLPATHS-LG.²⁴ *T. aestivum* genome was sequenced by the Genoscope (Institut de Génomique, Evry, France) using 454 and Illumina sequencing technologies, and assembled using Newbler (Roche). Final consensus was polished as described in Aury et al.²⁵.

*Genomes sequenced by the Genoscope (Institut de Génomique, Evry, France) using 454 and Illumina sequencing technologies, and assembled using Newbler (Roche). Final consensus was polished as described in Aury et al.²⁵.*
Articles


8. O’Donnell, K., Cigelnik, E., Weber, N. S. & Trappe, J. M. Phylogenetic sequencing and annotation at the Genoscope. T.P., B.N., A. Kuo, J.C., A. Kohler, K.K., Morisset, C., P.B., L.N., B.H., P.B., S.O., M.N., F.P. and J.S. planned and managed the research. I.V.G. support by the US Department of Energy Joint Genome Institute, a DOE Office of Science User Facility, was performed by the Department of Energy (DOE) Joint Genome Institute, a DOE Office of Science User Facility, with support by the US Department of Energy (DOE) under contract no. DE-AC02-05CH11231. The Genoscope is partly supported by France Génomique (grant no. ANR-10-IBHS-09). Computer facilities at INRA-Nancy are funded by the Region Lorraine Research Council and the European Commission (European Regional Development Fund). Financial support by the University of Parma (local funding, P.I. programme) to B.M. and S.O. and by the Interuniversity Consortium for Biotechnologies to S.O., and by Piedmont Region to P.B. is also gratefully acknowledged. L.Z. and P.Ba. were supported by the Czech Science Foundation (grant no.16–08916S). M.N. was supported by the German Research Council (grant no. NA04077–1) and would like to thank U. Kuck for support at the Department of General and Molecular Botany. We would like to thank the Agriturtle and Robin Nurseries companies for providing trees inoculated with *T. melanosporum* and *T. magnatum*, respectively. We are grateful to G. Chevalier, V. Gavazza and J.-F. Maurice for harvesting *T. aestival* and *T. magnatum* and *C. venosus* fruiting bodies. We also thank F. Sillo and E. Zampieri for obtaining and analysing the initial RNA-seq dataset from *T. magnatum* fruiting body.

Acknowledgements

The research was funded by the French National Research Agency through the projects SYSTRUF (grant nos. ANR-14-C006-0020 and ANR-09-STRAT-10) and SYMWOOD (Laboratory of Excellence Advanced Research on the Biology of Tree and Forest Ecosystems, grant no. ANR-11-LABX 0002). The work conducted by the US Department of Energy Joint Genome Institute, a DOE Office of Science User Facility, within the framework of the 1000 Fungal Genome project, was supported by the Office of Science of the US Department of Energy under contract no. DE-AC02-05CH11231. The Genoscope is partly supported by France Génomique (grant no. ANR-10-INBS-09). Computer facilities at INRA-Nancy are funded by the Region Lorraine Research Council and the European Commission (European Regional Development Fund). Financial support by the University of Parma (local funding, P.I. programme) to B.M. and S.O. and by the Interuniversity Consortium for Biotechnologies to S.O., and by Piedmont Region to P.B. is also gratefully acknowledged. L.Z. and P.Ba. were supported by the Czech Science Foundation (grant no.16–08916S). M.N. was supported by the German Research Council (grant no. NA04077–1) and would like to thank U. Kuck for support at the Department of General and Molecular Botany. We would like to thank the Agriturtle and Robin Nurseries companies for providing trees inoculated with *T. melanosporum* and *T. magnatum*, respectively. We are grateful to G. Chevalier, V. Gavazza and J.-F. Maurice for harvesting *T. aestival* and *T. magnatum* and *C. venosus* fruiting bodies. We also thank F. Sillo and E. Zampieri for obtaining and analysing the initial RNA-seq dataset from *T. magnatum* fruiting body.

Author contributions

Competing interests
The authors declare no competing interests.

Additional information
Supplementary Information is available for this paper at https://doi.org/10.1038/s41559-018-0710-4.
Reprints and permissions information is available at www.nature.com/reprints.
Correspondence and requests for materials should be addressed to C.M. or E.M.M.
Publisher’s note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.
Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. figure legend, table legend, main text, or Methods section).

<table>
<thead>
<tr>
<th>n/a</th>
<th>Confirmed</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>✗ The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement</td>
</tr>
<tr>
<td></td>
<td>✗ An indication of whether measurements were taken from distinct samples or whether the same sample was measured repeatedly</td>
</tr>
<tr>
<td></td>
<td>✗ Only common tests should be described solely by name; describe more complex techniques in the Methods section.</td>
</tr>
<tr>
<td></td>
<td>✗ A description of all covariates tested</td>
</tr>
<tr>
<td></td>
<td>✗ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons</td>
</tr>
<tr>
<td></td>
<td>✗ A full description of the statistics including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)</td>
</tr>
<tr>
<td></td>
<td>✗ For null hypothesis testing, the test statistic (e.g. F, t, r) with confidence intervals, effect sizes, degrees of freedom and P value noted</td>
</tr>
<tr>
<td></td>
<td>✗ Give P values as exact values whenever suitable.</td>
</tr>
<tr>
<td></td>
<td>✗ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings</td>
</tr>
<tr>
<td></td>
<td>✗ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes</td>
</tr>
<tr>
<td></td>
<td>✗ Estimates of effect sizes (e.g. Cohen’s d, Pearson’s r), indicating how they were calculated</td>
</tr>
<tr>
<td></td>
<td>✗ Clearly defined error bars</td>
</tr>
<tr>
<td></td>
<td>✗ State explicitly what error bars represent (e.g. SD, SE, CI)</td>
</tr>
</tbody>
</table>

Our web collection on statistics for biologists may be useful.

Software and code

Policy information about availability of computer code

Data collection

NA

Data analysis

The software used in this paper is extensive. It is detailed in the methods and supplementary methods sections (with version parameters and run parameters)

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Genome assemblies and annotations are available at the JGI MycoCosm and Genoscope portals, and NCBI as detailed in the paper.
Field-specific reporting

Please select the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

- Life sciences
- Behavioural & social sciences
- Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/authors/policies/ReportingSummary-flat.pdf

Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

<table>
<thead>
<tr>
<th>Study description</th>
<th>Comparative genomics of full nuclear genomes and comparative transcriptomics of field-collected truffles</th>
</tr>
</thead>
<tbody>
<tr>
<td>Research sample</td>
<td>Tuber aestivum fruit body, T. magnatum fruit body, Choiromyces venosus fruit body, Ascobolus immersus mycelium, Terfezia boudieri mycelium, Morchella conica mycelium</td>
</tr>
<tr>
<td>Sampling strategy</td>
<td>not applicable</td>
</tr>
<tr>
<td>Data collection</td>
<td>The genomic DNA used for sequencing T. aestivum, T. magnatum and C. venosus genomes were extracted from fruiting bodies (ascomata) sampled in truffle ground Montdoré (Haute-Saône, France), Montemagno (Piedmont, Italy) and Charmes State Forest (Grand-Est, France), respectively.</td>
</tr>
<tr>
<td>Timing and spatial scale</td>
<td>Samples were collected from truffle grounds in France and Italy. Sample sizes were set based on the availability of samples because truffles cannot be produced under controlled conditions. Our results and discussion are all framed within the parameters of statistical certainty.</td>
</tr>
<tr>
<td>Data exclusions</td>
<td>not applicable</td>
</tr>
<tr>
<td>Reproducibility</td>
<td>For transcriptomic analyses 3 replicates were realized</td>
</tr>
<tr>
<td>Randomization</td>
<td>not applicable</td>
</tr>
<tr>
<td>Blinding</td>
<td>not applicable</td>
</tr>
<tr>
<td>Did the study involve field work?</td>
<td>Yes</td>
</tr>
</tbody>
</table>

Reporting for specific materials, systems and methods

Materials & experimental systems

- n/a Involved in the study
- Unique biological materials
- Antibodies
- Eukaryotic cell lines
- Palaeontology
- Animals and other organisms
- Human research participants

Methods

- n/a Involved in the study
- ChIP-seq
- Flow cytometry
- MRI-based neuroimaging

Unique biological materials

Policy information about availability of materials

Obtaining unique materials

Fruiting bodies of truffles are unique materials collected in truffle grounds and produced in environmental settings. All the sample tissues were used for DNA or RNA extraction.