nature ecology & evolution

SUPPLEMENTARY INFORMATION

https://doi.org/10.1038/s41559-018-0710-4

In the format provided by the authors and unedited.

Pezizomycetes genomes reveal the molecular basis of ectomycorrhizal truffle lifestyle

Claude Murat[1,25]*, Thibaut Payen[1,25], Benjamin Noel[2], Alan Kuo[3], Emmanuelle Morin[1], Juan Chen[1,4], Annegret Kohler[1], Krisztina Krizsán[5], Raffaella Balestrini[6], Corinne Da Silva[2], Barbara Montanini[7], Mathieu Hainaut[8], Elisabetta Levati[7], Kerrie W. Barry[3], Beatrice Belfiori[9], Nicolas Cichocki[1], Alicia Clum[3], Rhyan B. Dockter[3], Laure Fauchery[1], Julie Guy[2], Mirco Iotti[10], François Le Tacon[1], Erika A. Lindquist[3], Anna Lipzen[3], Fabienne Malagnac[11], Antonietta Mello[6], Virginie Molinier[12,13], Shingo Miyauchi[1], Julie Poulain[2], Claudia Riccioni[9], Andrea Rubini[9], Yaron Sitrit[14], Richard Splivallo[15], Stefanie Traeger[16], Mei Wang[3], Lucia Žifčáková[17], Daniel Wipf[12], Alessandra Zambonelli[18], Francesco Paolocci[9], Minou Nowrousian[16], Simone Ottonello[7], Petr Baldrian[17], Joseph W. Spatafora[19], Bernard Henrissat[8,20,21], Laszlo G. Nagy[5], Jean-Marc Aury[2], Patrick Wincker[2], Igor V. Grigoriev[3,22], Paola Bonfante[23] and Francis M. Martin[1,24]*

[1]Institut National de la Recherche Agronomique, Unité Mixte de Recherche 1136 INRA-Université de Lorraine, Interactions Arbres/Microorganismes, Centre INRA-Grand Est-Nancy, Champenoux, France. [2]Commissariat à l'Energie Atomique, Genoscope, Institut de Génomique , Evry, France. [3]US Department of Energy Joint Genome Institute, Walnut Creek, CA, USA. [4]Institute of Medicinal Plant Development, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing, China. [5]Synthetic and Systems Biology Unit, Biological Research Center, Hungarian Academy of Sciences, Szeged, Hungary. [6]National Research Council – Institute for Sustainable Plant Protection, Torino Unit, Torino, Italy. [7]Department of Chemical Life Sciences & Environmental Sustainability, Laboratory of Biochemistry and Molecular Biology, University of Parma, Parma, Italy. [8]Architecture et Fonction des Macromolécules Biologiques, Aix-Marseille Université, Marseille, France. [9]CNR-IBBR, Istituto di Bioscienze e Biorisorse, UOS di Perugia, Perugia, Italy. [10]Department of Life, Health and Environmental Sciences, University of L'Aquila, L'Aquila, Italy. [11]Institute for Integrative Biology of the Cell, CEA, CNRS, Université Paris-Sud, Université Paris-Saclay, Gif-sur-Yvette cedex, France. [12]UMR 1347 Agroécologie AgroSup/INRA/uB, Pôle IPM – ERL CNRS 6300, Dijon, France. [13]UMR 5175 CEFE, CNRS, Université de Montpellier, Université Paul Valéry Montpellier, EPHE, INSERM, Campus CNRS, Montpellier, France. [14]The Jacob Blaustein Institutes for Desert Research, Bergman Campus, Ben-Gurion University of The Negev, Beer-Sheva, Israel. [15]Institute of Molecular Biosciences, Goethe University Frankfurt, Frankfurt am Main, Germany. [16]Lehrstuhl für Allgemeine und Molekulare Botanik, Ruhr-Universität Bochum, Bochum, Germany. [17]Laboratory of Environmental Microbiology, Institute of Microbiology of the CAS, Praha, Czech Republic. [18]Department of Agricultural and Food Sciences, University of Bologna, Bologna, Italy. [19]Department Botany & Plant Pathology, Oregon State University, Corvallis, OR, USA. [20]UMR 7257, Centre National de la Recherche Scientifique, Marseille, France. [21]Department of Biological Sciences, King Abdulaziz University, Jeddah, Saudi Arabia. [22]Department of Plant and Microbial Biology, University of California Berkeley, Berkeley, CA, USA. [23]Department of Life Sciences and Systems Biology, University of Torino, Torino, Italy. [24]Institute of Microbiology, Beijing Forestry University, Beijing, China. [25]These authors contributed equally: Claude Murat, Thibaut Payen. *e-mail: claude.murat@inra.fr; francis.martin@inra.fr

NATECOLEVOL-18044186

Pezizomycetes genomes reveal the molecular basis of ectomycorrhizal truffle lifestyle

by Murat *et al*.

Supplementary Information

Phylogenetic tree. A phylogenetic tree was constructed with 2,338 single-copy protein-coding genes identified by orthoMCL (v 2.0.9)[1]. We aligned the corresponding protein sequence using CLUSTAL omega (v 1.2.1)[2], extracted the conserved blocks from each alignment with Gblocks (v 0.91b)[3] with all default parameters and concatenated all the blocks in one sequence per species. Bootstrap analysis and tree inference were carried out with the RAxML (Randomized Axelerated Maximum Likelihood) program[4] with the LG+G+I+F model and 1000 rounds of bootstrapping. The best model was selected using *ModelTest* function of R package *phangorn*[5]. Maximum Likelihood phylogeny was midpoint-rooted with the *midpoint* function of R package *phangorn*. Smoothing lambda parameter was estimated using *chronopl* function with the cross-validation enabled of R package *ape*[6]. Lambda values ranging from 0 to 100,000 were tested to minimize the value of the cross-validation. Penalized likelihood molecular clock was then realized with the function *chronoPL* of R package *ape*. Tree calibration was realized with lower and higher *Tuberaceae* MRCA (134 and 179 Mya) estimated by Bonito *et al.*[7].

Gene family clustering. Protein sequences from the eight Pezizomycetes were clustered with FastOrtho (v 1.0)[8] which combine BLAST+ (v 2.2.28+)[9], here used with an e-value threshold of $1e^{-5}$, and MCL (v 12.5)[10] used with default parameters.

COMPARE analysis. Evolutionary history of gene families of *Tuberaceae* species and closely related Pezizomycetes was analysed using the COMPARE pipeline[11]. Predicted protein families were clustered based on sequence similarity in 16 Ascomycota species using MCL (inflation parameter, 2), yielding 11,192 protein clusters. Multiple sequence alignments and maximum likelihood gene trees were inferred for each cluster by using PRANK v140603 and RAxML 8.1.2, respectively. For gene tree estimation the WAG model of protein evolution was used. To reconstruct gene duplications and losses, a genome-wide collection of 13,821 gene trees was first reconciled with the species tree using Treefix v1.1.1077. Treefix was run with RAxML site-wise likelihood model, Maximum Parsimony Reconciliation model (MPR) and an alpha threshold of 0.001 to find any alternate gene tree topologies that minimize duplication/loss costs but are statistically equivalent to the ML gene tree. Orthologues

were identified and recoded into a presence/absence matrix, as described in a previous study[11]. We then inferred duplications and losses for each orthogroup along the species tree using Dollo parsimony. Gene trees with less than four proteins were excluded. We also reconstructed the genome size for a given node by summing over gains and losses to the genome size of the most recent common ancestor (MRCA). GO enrichment analysis based on the Fisher exact test with Benjamin–Hochberg correction was performed using PFAM domains mapped to protein clusters and creating GO annotations with pfam2go version 2015/02/14 followed by enrichment analysis at $P < 0.05$.

Comparative genomic analyses and annotation of functional categories

RepeatScout[12] was used to identify *de novo* repetitive DNA in the genome assembly as reported in Peter *et al*.[13]. CAZymes including glycoside hydrolases (GH), glycosyl transferases (GT), polysaccharide lyases (PL), carbohydrate esterases (CE), enzymes that act in conjunction with other CAZymes (Auxiliary activities, AA), carbohydrate-binding modules (CBM) and enzymes distantly related to plant expansins (EXPN), were identified using the CAZy database (www.cazy.org) annotation pipeline[14]. To compare the distribution of genes encoding CAZymes in the various genomes from saprotrophic, symbiotic and pathogenic species, we applied hierarchical clustering of the number of genes for each of the 69 species available at the JGI MycoCosm database (Supplementary Table 15) using the Genesis software[15]. The Euclidian distance was used as the distance metric and a complete linkage clustering was performed. Secreted proteins were identified using a custom pipeline including SignalP v4, WolfPSort, TMHMM, TargetP, and PS-Scan algorithms as reported in Pellegrin *et al*.[16].

In silico analysis of VOC biosynthetic pathways

The amino acid sequences of enzymes known to be involved in VOC biosynthetic pathways were used as references for BLAST analyses. The KEGG pathway database (sulfur metabolism pathway 00920; terpenoid backbone biosynthesis pathway 00900; sesquiterpenoid and triterpenoid biosynthesis pathway 00909) along with literature-derived information were used to spot key enzymes in each pathway[17-21]. In particular, *Saccharomyces cerevisiae* or *T. melanosporum* (v. 1.0) predicted polypeptides were retrieved from the KEGG (http://www.genome.jp/kegg/), SGD (the *Saccharomyces* Genome Database, http://www.yeastgenome.org/), or Génoscope (https://www.genoscope.cns.fr/) databases. Reference sequences were used as queries for BLASTP searches to identify homologs in the eight examined Pezizomycetes. Putative orthologues were identified as "best reciprocal hits" and verified by sequence alignment. Multigene families were further investigated and verified by phylogenetic analysis using MEGA v. 6[22].

Synteny analysis

To conduct an analysis of the synteny in the eight genomes, the OMA program[23] was used to generate a list of orthologous genes. *Circos*[24] was used to visualize the synteny based on orthologous genes. We considered a syntenic block only if at least two genes are presents in the same order in the different genomes. We used VISTA available at the JGI MycoCosm *T. magnatum* portal (https://genome.jgi.doe.gov/Tubma1/Tubma1.home.html) to investigate the nucleotide sequence polymophism at the *GH6_1* locus in *Tuberaceae*.

RNA extraction, RNA sequencing and data analysis

For RNA extraction, the free-living mycelium of *T. melanosporum* Mel28 (INRA-Nancy Fungal Collection) was grown as described in Tisserant *et al.*[25] on 1% malt agar (Cristomalt- D, Difal, Villefranche-sur-Saône, France) for five weeks before harvesting. The free-living mycelium of *T. magnatum* was isolated from a fruiting body collected in Città di Castello, Perugia (Italy). It was grown in 100 ml liquid MNM medium supplemented with Bacto Peptone (4 g per L), yeast extract (2 g per L), and malt extract (2 g per L) in the dark at 22-24°C for 30 days. Mycelium was then sampled and snap frozen in liquid nitrogen before RNA extraction. T. *melanosporum* ectomycorrhizae were sampled on 5-month-old hazel seedlings (*Corylus avellana* L.), whereas *T. magnatum* ectomycorrhizae were sampled on 6-month-old inoculated *Q. robur* plantlets. Inoculated hazel plantlets were grown in the greenhouse facilities of the Agritruffe Company (Saint-Maixant, France), while inoculated oak plants were produced by Robin nurseries (St Laurent du Cros, France), following their standard protocols. Five fruiting bodies of *T. melanosporum* of different maturity and from different locations were collected (Supplementary Fig. 13). One *T. magnatum* fruiting body was harvested in Assisi-Valfabbrica (Perugia, Italy) in November 2005 and two in Pietralunga (Perugia, Italy) in November 2016. Total RNA from *T. magnatum* free-living mycelium and fruiting bodies was extracted using the Spectrum Plant Total RNA kit (Sigma-Aldrich), following the protocol A and the on-column DNase Digestion procedure. Total RNA was extracted from ectomycorrhizal root tips of *T. magnatum–Q.robur* using the RNeasy Plant Mini Kit of Qiagen with DNase step and addition of 20 mg per ml polyethylene glycol to the RLC extraction buffer. Three replicates were used for RNA-seq sequencing. Preparation of libraries from total RNA and 2 x 100 bp Illumina HiSeq sequencing (RNA-Seq) was performed at the GET platform (Génopole Toulouse Midi- Pyrénées, Auzeville, France) following their standard protocol. Quality filtered reads were aligned to the *T. magnatum* reference transcripts (https://genome.jgi.doe.gov/Tubma1/Tubma1.home.html) using CLC Genomics Workbench 9 (Qiagen). To identify differentially regulated transcripts in ectomycorrhizae and fruiting bodies compared to free-living mycelium the Baggerly *et al.*'s test implemented in CLC Genomic Workbench and the Benjamini & Hochberg multiple-hypothesis testing to correct for FDR was

applied to the data. If not otherwise indicated, transcripts with ≥5- fold change and a FDR corrected *P*-value <0.05 were used for further analysis. To assess whether symbiosis-regulated transcripts were conserved among Pezizomycetes or lineage-specific, their protein sequences were queried against the protein sequences of the eight Pezizomycetes genomes using BLASTP as described in Kohler *et al.*[26]. The distribution of the symbiosis-regulated transcripts within each cluster were quantified according to their putative function as CAZymes, small secreted proteins (SSPs), with KOG classification or without KOG. The enrichment of SSPs in symbiosis-regulated transcript clusters was assessed by comparing their percentage in symbiosis-regulated transcript clusters to their relative percentage in the total gene repertoire and applying a Fisher's exact test (*P*<0.01 for enrichment). Data were visualized and hierarchically clustered using the euclidean distance metric and the average-linkage clustering method in R (package HeatPlus). For Venn diagrams http://bioinformatics.psb.ugent.be/webtools/Venn/ was used. Microarray data from *T. melanosporum* are described in Martin *et al*.[18] and available at NCBI/GEO as series GSE17529, the *T. magnatum* RNA-Seq data as GEO series [Submitted]. For Supplementary Figure 12, microarray data published by Le Tacon *et al.*[27] from fruiting bodies at different developmental stages and harvested over several years on the truffle orchard at Rollainville, France were used. The *P. confluens* RNA-seq data are available under GEO accession number GSE61274.

Supplementary Results

DNA decay acting on cellobiohydrolase GH6

Secreted cellobiohydrolases (CBH) of glycoside hydrolase families 6 and 7 (GH6 and GH7) act most efficiently on highly ordered microcrystalline cellulose, hydrolyzing either from the reducing or non-reducing end to liberate predominantly cellobiose. Two to five GH7 genes were identified in the saprotrophic Pezizomycetes. In contrast, no such gene was found in *Tuberaceae*, and a single GH7 gene is present in *Te. boudieri* (Supplementary Table 3). No GH6 gene is encoded by *T. melanosporum* genome, whereas a single gene copy was predicted in *T. aestivum* and three gene copies in *T. magnatum* and *C. venosus* (Supplementary Fig. 10A). In *T. magnatum* and *T. aestivum*, a single gene, called GH6_1 (JGI ID# 301672 and GSTUAT00006748001, respectively), is expressed in free-living mycelia and ectomycorrhizae, and its structure is supported by RNA-Seq reads (Supplementary Fig. 10A). Two additional copies found in *T. magnatum*, called GH6_2 (JGI ID# 351473) and GH6_3 (JGI ID# 219130), are probably pseudogenes since they are truncated and not expressed in any of the tissues analyzed (Supplementary Fig. 10A). To investigate the mechanism underlying the loss of GH6_1 in *T. melanosporum,* we used the functional *T. aestivum* GH6_1 gene as a query to search for sequence similarities in *T. melanosporum* genome. We identified small sequence fragments corresponding to *GH6* gene at the expected *GH6* locus in the genome assembly, but 90% of

the *GH6* sequence was missing (Supplementary Fig. 10B). Large syntenic blocks framed the position of the *GH6* locus, indicating that the general local genome organization was not modified as a result of transposon activity (Supplementary Fig. 11). This finding suggests that DNA decay is the inactivation mechanism specifically driving the loss of GH6 in *T. melanosporum*. It remains to be determined whether this inactivation mechanism is involved in the loss of other PCWDEs.

All Tuberaceae *are heterothallic*

The sequenced strains of *T. aestivum* and *T. magnatum* only encode the *MAT1-2-1* gene, while *C. venosus* only harboured the *MAT1-1-1* gene (Supplementary Table S11). This indicates heterothallism as a common reproductive mode in *Tuberaceae*. Not only the already sequenced genome of *T. melanosporum,* but also single strains of *T. indicum* and *T. borchii* harbor either the *MAT1-1-1* or the *MAT1-2-1* genes[18, 28-30]. Among other sequenced Pezizomycetes, both *MAT1-1-1* and *MAT1-2-1* genes were found in *P. confluens,* as shown by Traeger and colleagues[31], whereas only *MAT1-1-1* gene was found in *A. immersus*. No *MAT* gene was found in *Te. boudieri* when either *MAT1-1-1* or *MAT1-2-1* genes from other Pezizomycetes were used as BLAST queries (Supplementary Table S8). A low level of sequence synteny was found around the *MAT* locus of these fungi. Homologs of the *RSM22* and *COX13* genes were identified next to the *MAT* locus of *T. magnatum*, *T. aestivum* and *T. melanosporum,* but not in *C. venosus*, *M. importuna*, *P. confluens* and *A. immersus,* whereas homologs of *T. melanosporum* gene model GSTUMT200009628001 were found next to the *MAT* locus of all *Tuberaceae* along with *M. importuna*.

References

1. Li, L., Stoeckert, C. J. & Roos, D. S. OrthoMCL: Identification of Ortholog Groups for Eukaryotic Genomes. *Genome Res.* **13**, 2178–2189 (2003).
2. Sievers, F. et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7**, 539 (2011).
3. Castresana, J. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol. Biol. Evol.* **17**, 540–552 (2000).
4. Stamatakis, A. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22,** 2688–2690 (2006).
5. Schliep, K. P. phangorn: phylogenetic analysis in R. Bioinformatics 27, 592–593 (2011).
6. Paradis, E., Claude, J. & Strimmer, K. APE: Analyses of Phylogenetics and Evolution in R language. *Bioinformatics* **20**, 289–290 (2004).
7. Bonito, G. et al. Historical biogeography and diversification of truffles in the Tuberaceae and their newly identified southern hemisphere sister lineage. *PLoS One* **8**, e52765 (2013).

8. Wattam, A. R. et al. PATRIC, the bacterial bioinformatics database and analysis resource. *Nucleic Acids Res.* **42**, D581–D591 (2014).

9. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).

10. Enright, A. J., Van Dongen, S. & Ouzounis, C. A. An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Res.* **30**, 1575–1584 (2002).

11. Nagy, L. G. *et al.* Latent homology and convergent regulatory evolution underlies the repeated emergence of yeasts. *Nat. Commun.* **5,** 4471 (2014).

12. Price, A. L., Jones, N. C. & Pevzner, P. A. De novo identification of repeat families in large genomes. *Bioinformatics* **21,** i351–i358 (2005).

13. Peter, M. *et al.* Ectomycorrhizal ecology is imprinted in the genome of the dominant symbiotic fungus *Cenococcum geophilum*. *Nat. Commun.* **7,** 12662 (2016).

14. Cantarel, B. L. *et al.* The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. *Nucleic Acids Res.* **37,** D233–D238 (2009).

15. Sturn, A., Quackenbush, J. & Trajanoski, Z. Genesis: cluster analysis of microarray data. *Bioinformatics* **18,** 207–208 (2002)

16. Pellegrin, C., Morin, E., Martin, F. M. & Veneault-Fourrey, C. Comparative Analysis of Secretomes from Ectomycorrhizal Fungi with an Emphasis on Small-Secreted Proteins. *Front. Microbiol.* **6,** (2015).

17. Hazelwood, L. A., Daran, J.-M., Maris, A. J. A. van, Pronk, J. T. & Dickinson, J. R. The Ehrlich pathway for fusel alcohol production: a century of research on *Saccharomyces cerevisiae* Metabolism. *Appl. Environ. Microbiol.* **74,** 2259–2266 (2008).

18. Martin, F. *et al.* Périgord black truffle genome uncovers evolutionary origins and mechanisms of symbiosis. *Nature* **464,** 1033–1038 (2010).

19. Heshof, R. *et al.* A novel class of fungal lipoxygenases. *Appl. Microbiol. Biotechnol.* **98,** 1261–1270 (2014).

20. Saerens Sofie M. G., Delvaux Freddy R., Verstrepen Kevin J. & Thevelein Johan M. Production and biological function of volatile esters in *Saccharomyces cerevisiae*. *Microb. Biotechnol.* **3,** 165–177 (2009).

21. Combet, E., Henderson, J., Eastwood, D. C. & Burton, K. S. Eight-carbon volatiles in mushrooms and fungi: properties, analysis, and biosynthesis. *Mycoscience* **47,** 317– 326 (2006).

22. Tamura, K., Stecher, G., Peterson, D., Filipski, A. & Kumar, S. MEGA6: Molecular evolutionary genetics analysis version 6.0. *Mol. Biol. Evol.* **30,** 2725–2729 (2013).

23. Dessimoz, C. *et al.* OMA, A comprehensive, automated project for the identification of

orthologs from complete genome data: Introduction and first achievements. in *Comparative Genomics* 61–72 (Springer, Berlin, Heidelberg, 2005). doi:10.1007/11554714_6

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

25. Tisserant E. *et al.* Deep RNA sequencing improved the structural annotation of the *Tuber melanosporum* transcriptome. *New Phytol.* **189,** 883–891(2011).

26. Kohler, A. *et al.* Convergent losses of decay mechanisms and rapid turnover of symbiosis genes in mycorrhizal mutualists. *Nat. Genet.* **47,** 410–415 (2015).

27. Le Tacon, F. *et al.* Study of nitrogen and carbon transfer from soil organic matter to *Tuber melanosporum* mycorrhizas and ascocarps using ^{15}N and ^{13}C soil labelling and whole-genome oligoarrays. *Plant Soil* **395,** 351–373 (2015).

28. Rubini, A. *et al.* Isolation and characterization of *MAT* genes in the symbiotic ascomycete *Tuber melanosporum*. *New Phytol.* **189,** 710–722 (2011).

29. Belfiori, B., Riccioni, C., Paolocci, F. & Rubini, A. Mating type locus of Chinese black truffles reveals heterothallism and the presence of cryptic species within the *T. indicum* species complex. *PLoS One* (2013).

30. Belfiori, B., Riccioni, C., Paolocci, F. & Rubini, A. Characterization of the reproductive mode and life cycle of the whitish truffle *T. borchii*. *Mycorrhiza* **26,** 515– 527 (2016).

31. Traeger, S. *et al.* The genome and development-dependent transcriptomes of *Pyronema confluens*: a window into fungal evolution. *PLoS Genet.* **9,** e1003820 (2013).

Legends of Supplementary tables as presented in the SupplementaryTables.xlsx file

Supplementary Table 1. Conservation of genes involved in RIP-like genome defense in the sequenced Pezizomycetes.

Supplementary Table 2. Structural features of total genes, species-specific genes and shared (non-specific) genes in the sequenced Pezizomycetes.

Supplementary Table 3. Distribution of carbohydrate-active enzymes (CAZymes) in the sequenced Pezizomycetes.

Supplementary Table 4. Upregulated genes (>5-fold) in *T. magnatum* ectomycorrhizae (ECM) compared to free-living mycelium (FLM). Upregulated genes in fruiting-body (FB) are also showed (see Supplementary Table S9). Clusters referred to Fig. 4A.

Supplementary Table 5. Downregulated genes (>5-fold) in *T. magnatum* ectomycorrhizae (ECM) compared to free-living mycelium (FLM). Downregulated genes in fruiting-body (FB) are also showed. ND, not detected in ECM.

Supplementary Table 6. Most highly upregulated transcripts in ectomycorrhizae (ECM) of both *T. melanosporum-C. avellana* and *T. magnatum-Q. robur* associations compared to free-living mycelium (FLM).

Supplementary Table 7. Number of genes and percentage of KOG categories up-regulated *in T. magnatum* (Tubma) and *T. melanosporum* (Tubme) ectomycorrhiza (ECM) and fruiting bodies (FB), compared to free-living mycelium (FLM). For Tubme oligoarray data was used, whereas for Tubma RNA-Seq data was used.

Supplementary Table 8. Presence of mating-type genes in the sequenced Pezizomycetes. The scaffold containing the *MAT* locus is indicated. ND, not detected

Supplementary Table 9. Upregulated genes (>5-fold) in *T. magnatum* fruiting body (FB) compared to free-living mycelium (FLM). Upregulated genes in ectomycorrhizae (ECM) are also showed (see Supplementary Table S4).

Supplementary Table 10. Downregulated genes (>5-fold) in *T. magnatum* fruiting body (FB) compared to free-living mycelium (FLM). Downregulated genes in ectomycorrhiza (ECM) are also showed.

Supplementary Table 11. Upregulated genes in fruiting body (FB) of both *T. melanosporum* and

T. magnatum by comparison to free-living mycelium (FLM).

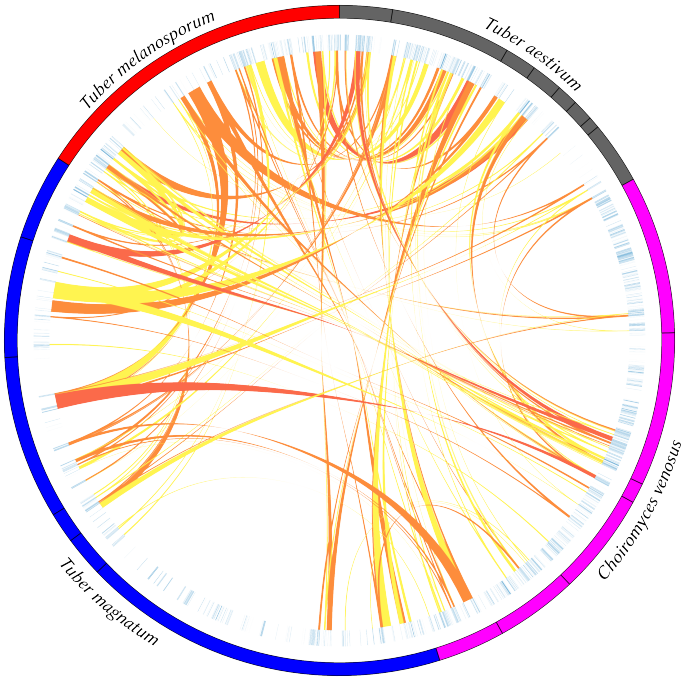
Supplementary Table 12. Presence and sequence similarity of fruiting body-upregulated genes from *T. magnatum* in genomes of other sequenced Pezizomycetes (linked to Fig. 4B)

Supplementary Table 13. Relative expression of genes implicated in sexual reproduction in fruiting bodies of *T. melanosporum*, *T. magnatum* and *T. aestivum*. Orthologs of genes involved in mating and meiosis in yeast are listed on the right panel. For *T. melanosporum* (Tubmel), Tubme1, fruiting bodies of *T. melanosporum* harvested on different sites at different dates. Tubme2, immature fruiting bodies of *T. melanosporum* harvested at the Rollainville truffle orchard during Winter 2010-2011; Tubme3, mature fruiting bodies of *T. melanosporum* harvested at the Rollainville truffle orchard during Winter 2010-2011; Tubme4, mature fruiting bodies of *T. melanosporum* harvested at the Rollainville truffle orchard during Winter 2013-2014; Tubmag, mature fruiting bodies of *T. magnatum* harvested in Piedmont (Italy); Tubae, a fruiting body of *T. aestivum* harvested in Northeastern France. See also Supplemental Fig. S13.

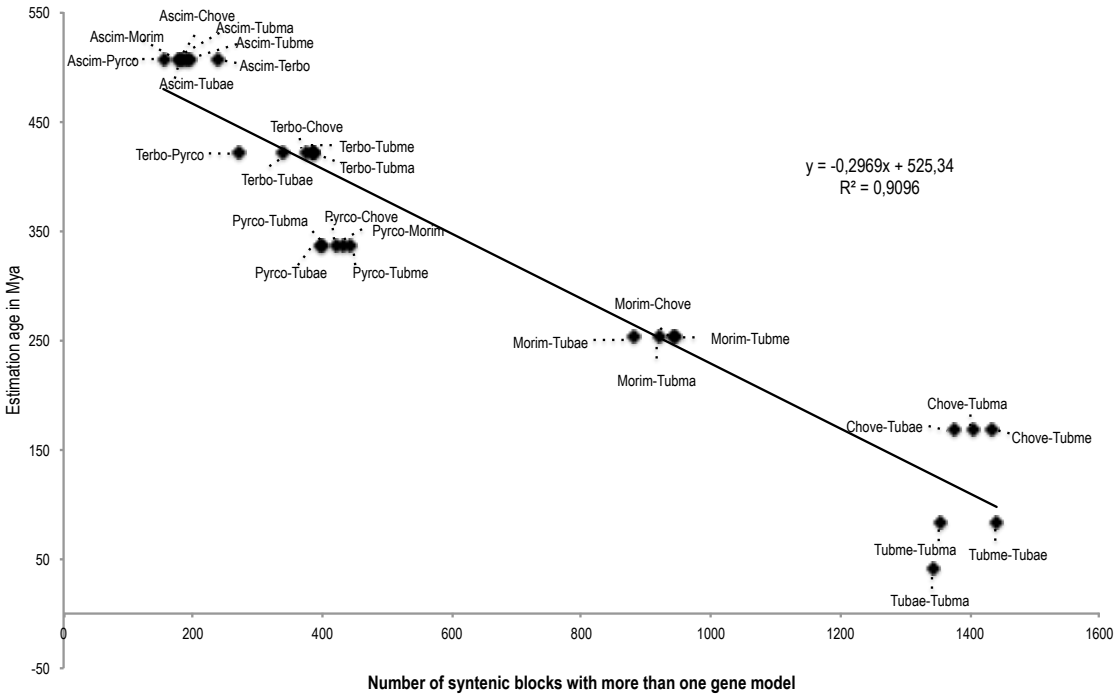
Supplementary Table 14. Identification and expression of the enzymes involved in the biosynthesis of volatile organic compounds (VOC) in the fruiting body of the sequenced Pezizomycetes. For Pyrco the gene number corresponds to those of Mycocosm (DOE-JGI).

Supplementary Table 15. Fungal species, their lifestyle and abbreviations of the genome assemblies used in this study.

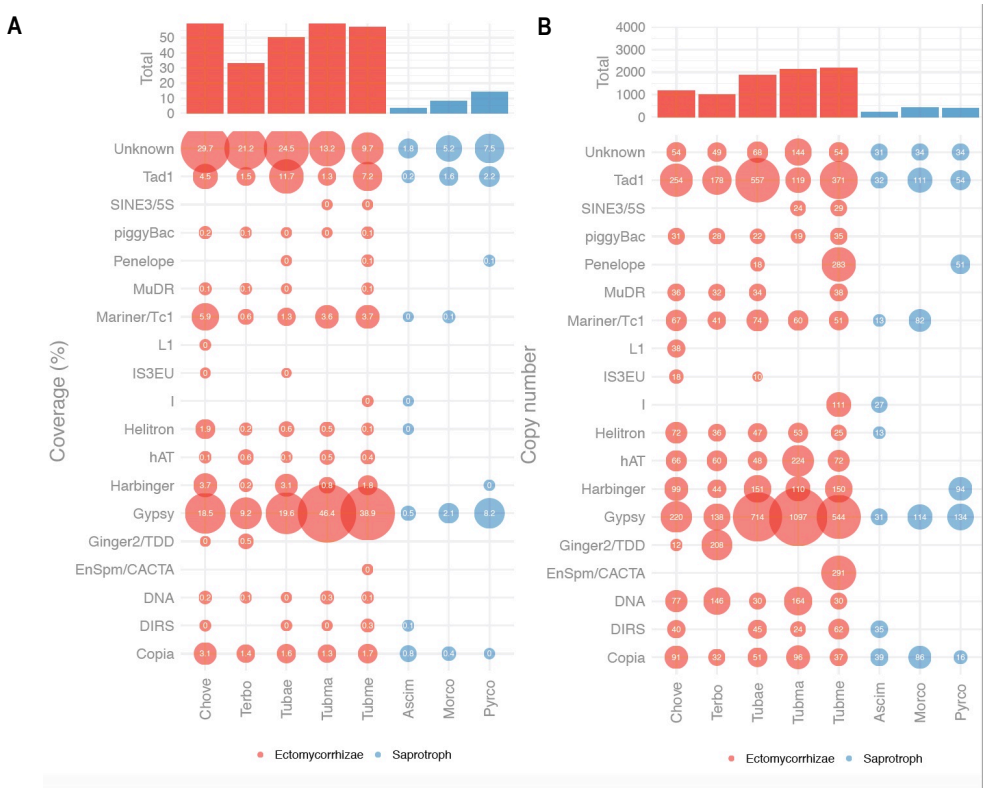
Supplementary figures



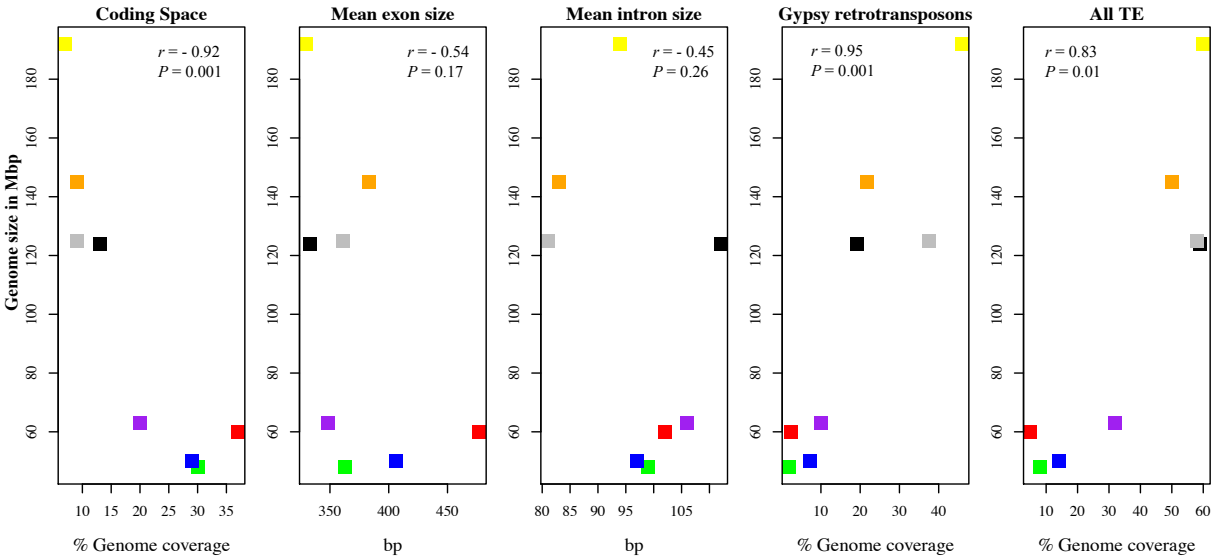
Supplementary Figure 1. Genome synteny in *Tuberaceae*. Circos plot showing syntenic regions in four *Tuberaceae* species. One syntenic block is considered only if at least two orthologous genes are presents in the same order in at least two genomes. The yellow, orange and red links correspond to syntenic blocks composed by <5, 5-10 and >10 genes, respectively.



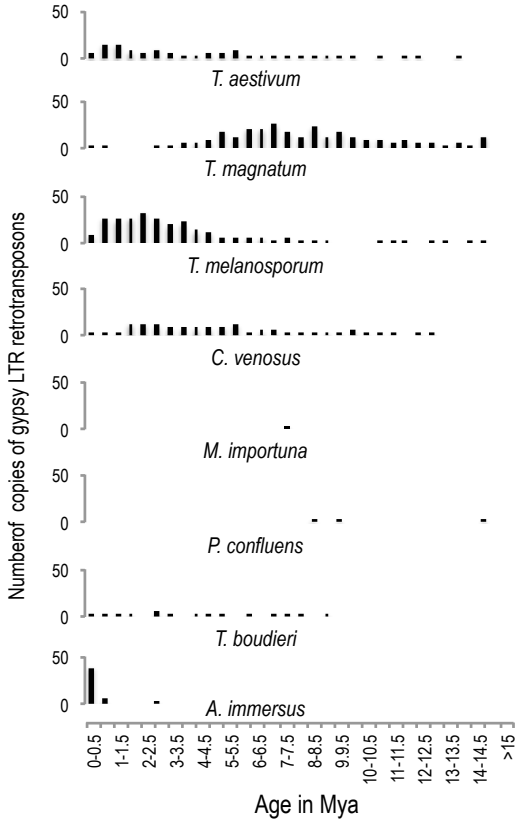
Supplementary Figure 2. Genome synteny in Pezizomycetes. A high correlation (Pearson's, r = -0.96, P = 1.3e-15) between the number of syntenic blocks and the estimation age for each pair of Pezizomycetes genomes. Up to 1,441 syntenic blocks are conserved between *T. melanosporum* and *T. aestivum*, whereas only 153 syntenic blocks are shared by *A. immersus* and *P. confluens*. Abbreviations, *Ascobolus immersus* (Ascmi1), *Terfezia boudieri* (Terbo), *Pyronema confluens* (Pyrco), *Morchella importuna* (Morim), *Choiromyces venosus* (Chove), *T. melanosporum* (Tubme), *T. magnatum* (Tubma) and *T. aestivum* (Tubae).



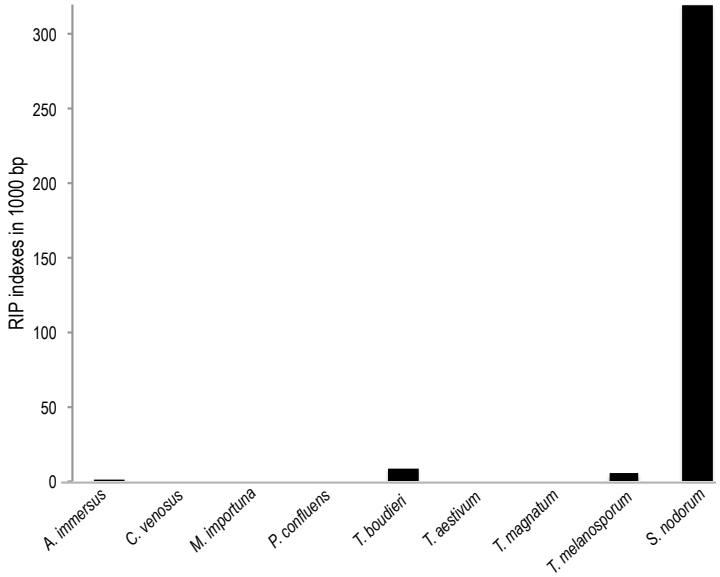
Supplementary Figure 3. The genome coverage (**a**) and copy number (**b**) of transposable elements in the sequenced Pezizomycetes. The numbers in bubbles indicate the coverage/copy number of transposable element families. The total coverage/copy number are presented on top. Chove, *Choiromyces venosus*; Terbo, *Terfezia boudieri*; Tubae, *Tuber aestivum*; Tubma, *Tuber magnatum*; Ascim, *Ascobolus immersus*; Morco; *Morchella importuna*; Pyrco, *Pyronema confluens*.



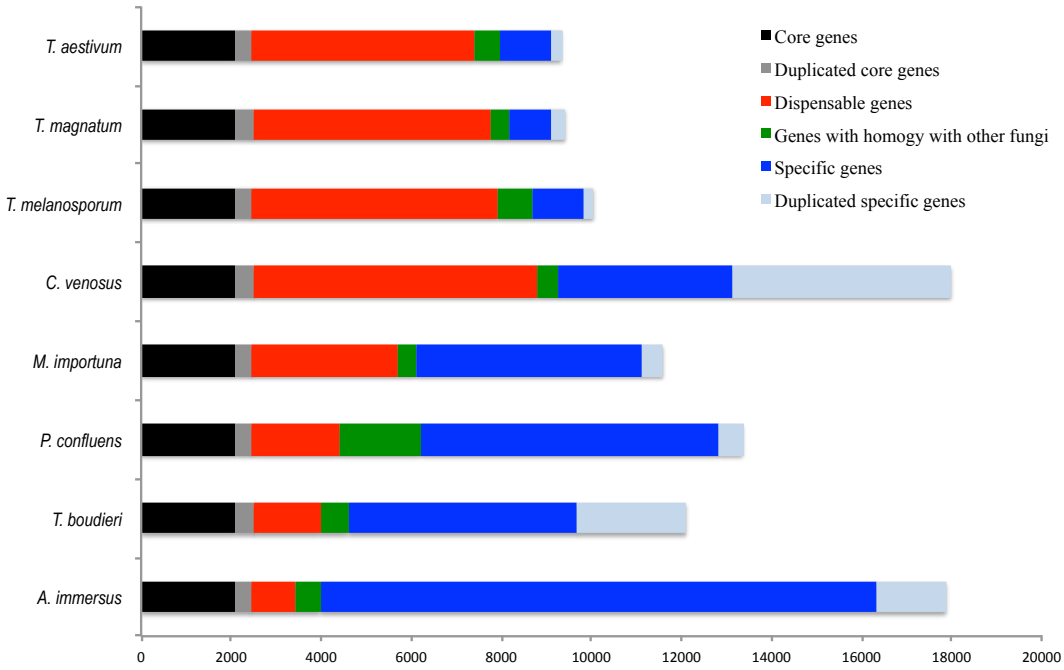
Supplementary Figure 4. The relationship between gene numbers, exon size, intron size, *Gypsy* LTR and total TE content (in Mb) and genome assembly size (in Mb) in *Tuberaceae* and other Pezizomycetes. The significance of the correlation was tested by Spearman test in R using *cov.test*() function of stats package. Each species are coded by a color: *Ascobolus immersus* (Ascmi1) in red, *Terfezia boudieri* (Terbo) in purple, *Pyronema confluens* (Pyrco) in blue, *Morchella importuna* (Morim) in green, *Choiromyces venosus* (Chove) in black, *Tuber melanosporum* (Tubme) in grey, *T. magnatum* (Tubma) in yellow and *T. aestivum* (Tubae) in orange.



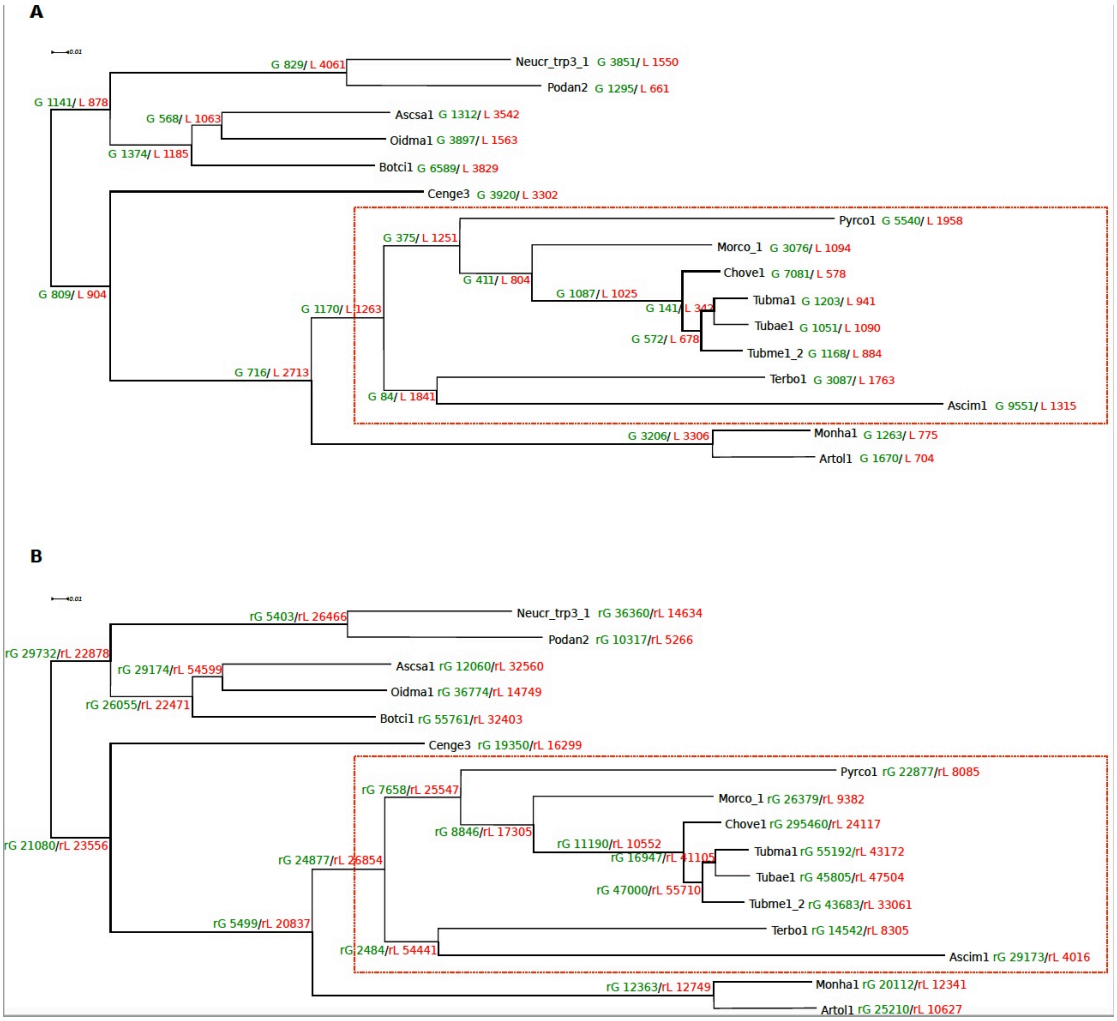
Supplementary Figure 5. Age distribution of G*ypsy* LTR in the eight selected Pezizomycetes over the last 15 millions years.



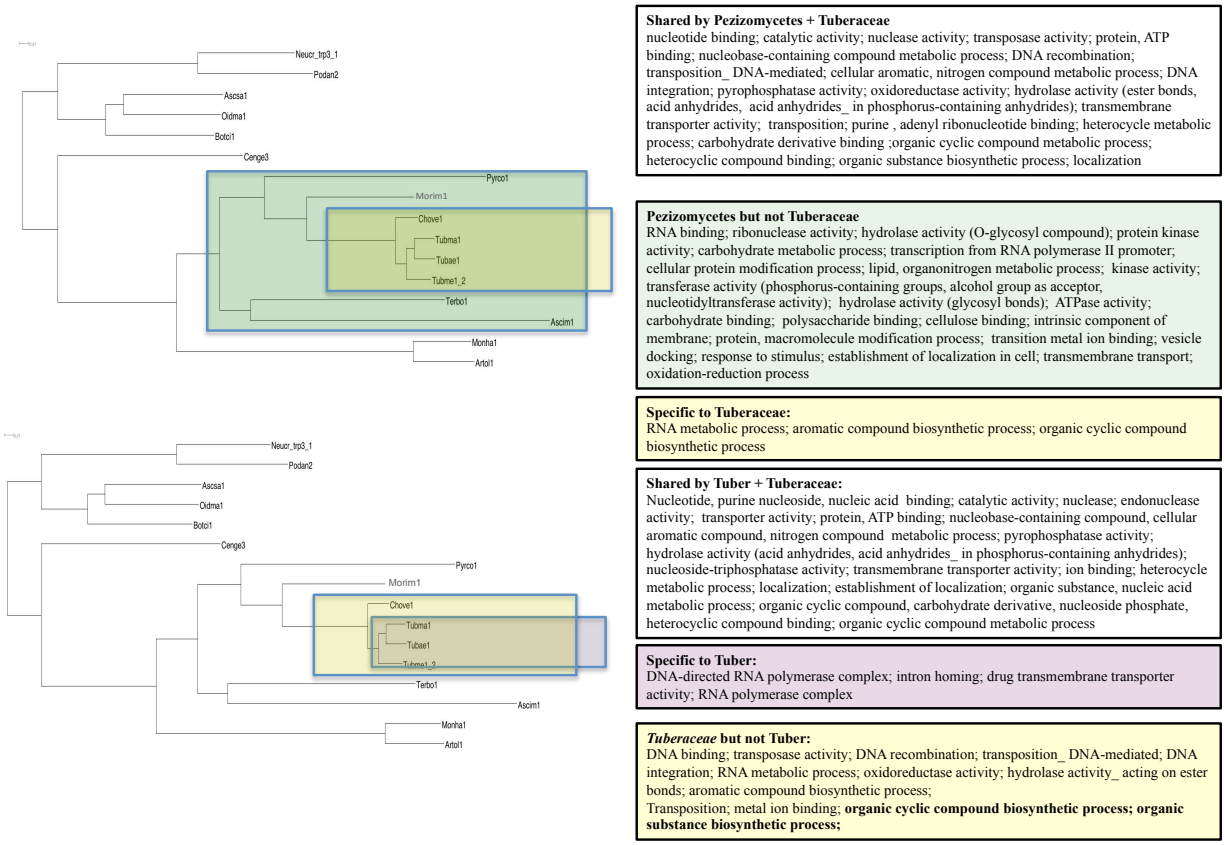
Supplementary Figure 6. Estimation of RIP (Repeated Induced Points) indexes in 1000 bp windows. The number of windows with both TpA/ApT > 2 and (CpA+TpG)/(ApC+GpT) < 0.7 are indicated. *Stagonospora nodorum* was included in this analysis since this species has a RIP defense mechanism.



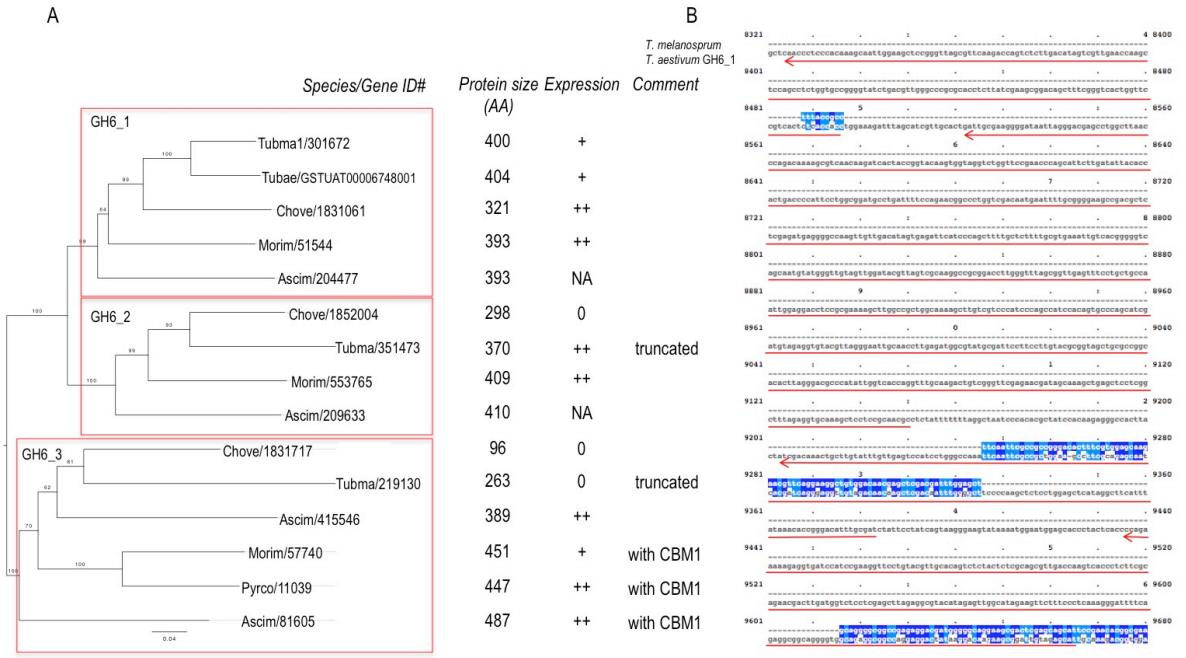
Supplementary Fig. 7. Gene conservation in the eight sequenced Pezizomycete species. Bar graphs represent the sets of core (black and grey), dispensable (red), and species-specific (blues) genes as well as those with homology with genes of other fungal species (green) according to a blast query against JGI Mycocosm database (https://genome.jgi.doe.gov/programs/fungi/index.jsf).



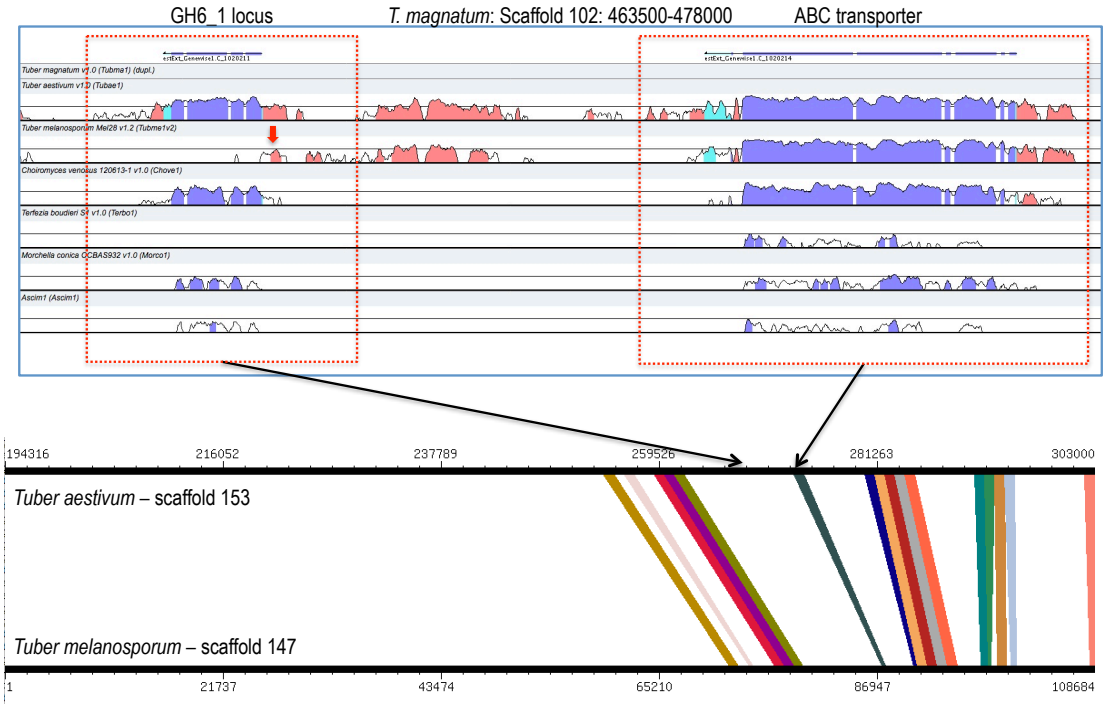
Supplementary Figure 8. Rate of genome diversification in Ascomycota. (**a**) Gene duplication (G)/loss (L) histories along a phylogeny of 16 fungal species of Pezizomycetes (red box) and other Ascomycota. Green numbers correspond to duplication and red numbers to loss. (**b**) Gene duplication (rG) and loss rates (rL). Species tree was inferred from 1,028 single-copy genes using partitioned maximum likelihood (RAxML) (alignment length, 329,332 amino acids). Abbreviations: Artol1, *Arthrobotrys oligospora*; Ascim1, *Ascobolus immersus*; Ascsa1, *Ascocoryne sarcoides*; Botci1, *Botrytis cinerea*; Chove1, *Choiromyces venosus*; Cenge3, *Cenococcum geophilum*; Monha1, *Monacrosporium haptotylum*; Neucr_trp3_1, *Neurospora crassa*; Podan2, *Podospora anserina*; Oidma1, *Oidiodendron maius*; Morim1, *Morchella importuna*; Pyrco1, *Pyronema confluens*; Tubae1, *Tuber aestivum*; Tubma1, *Tuber magnatum*; Tubme1_2 , *Tuber melanosporum*; Terbo1, *Terfezia boudieri*.



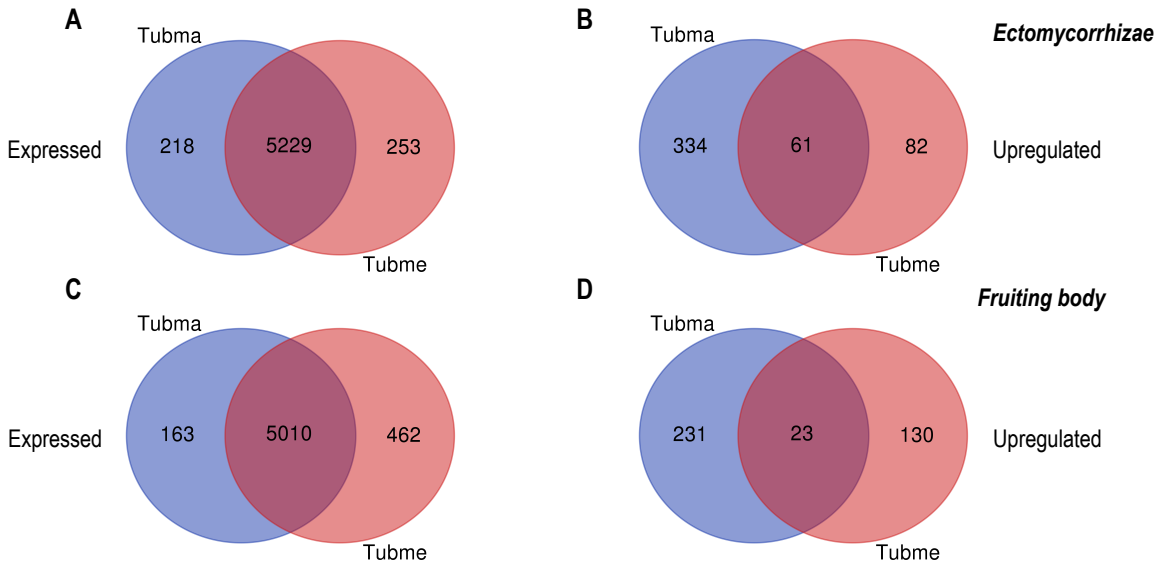
Supplementary Figure 9. Gene diversification along a phylogeny of 16 fungal species, including the eight sequenced Pezizomycetes. Biochemical functions based on Gene Ontology categories enriched are listed for Pezizomycetes, Pezizomycetes but not *Tuberaceae* and *Tuberaceae* only. See Supplementary Figure 8 for species abbreviations.



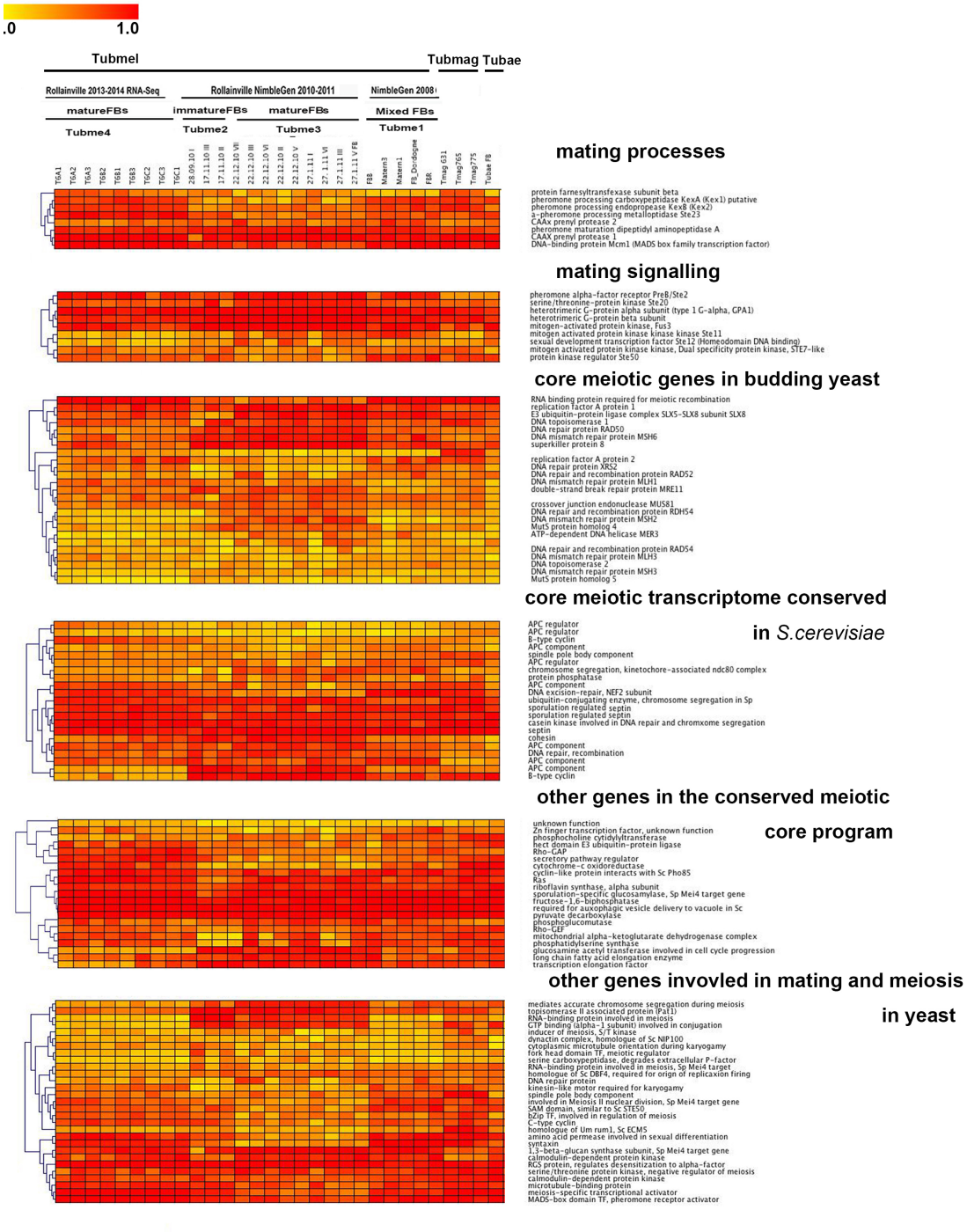
Supplementary Figure 10. Evolution of the GH6 cellobiohydrolase genes in *Tuberaceae*. (A) Neighbour-joining tree of GH6_1, GH6_2, and GH6_3 cellobiohydrolases in the sequenced Pezizomycetes. Right panel shows the protein size in amino acids and the approximate level of expression in at least one tissue (- no expression, + low expression and ++ high expression). (B) Alignment of the nucleotide sequences of the *GH6_1* locus in *Tuber aestivum* and *T. melanosporum*. The red arrows underline exons in the *T. aestivum* sequence. NA, not available.



Supplementary Figure 11. Comparison of the genomic region containing *T. aestivum* GH6 cellobiohydrolase gene to syntenic regions in *T. melanosporum, C. venosus, Te. boudieri, M. importuna (*formerly *M. conica) and A. immersus (Ascim1)* using the VISTA program available at *T. magnatum* JGI genome portal. The red arrow indicates the GH6 gene remnants in *T. melanosporum.*



Supplementary Figure 12. Venn diagrams showing shared and species-specific gene expression (**A, C**) and up-regulation (**B, D**) in *T. magnatum* (Tubma) and *T. melanosporum* (Tubme) ectomycorrhizae (**A, B**) or fruiting bodies (**C, D**), compared to their respective free-living mycelium (FLM). For the comparison, 6,952 orthologous genes were identified by Best Reverse BlastP Hit. From this set, 5,809 were called in *T. melanosporum* array datasets used for this analysis. A gene was considered as expressed when its rpkm value was >5 or its array expression value >100 [5]. A gene was considered as up-regulated when its fold change was >5 (FDR *p*-value>0.05).



Supplementary Figure 13. Expression of genes implicated in sexual reproduction in fruiting bodies of *Tuber melanosporum, T. magnatum* and *T. aestivum*. Orthologs of genes involved in mating and meiosis in yeast are listed on the right panel. For *T. melanosporum* (Tubmel), Tubme1, fruiting bodies of *T. melanosporum* harvested on different sites at different dates. Tubme2, immature fruiting bodies of *T. melanosporum* harvested at the Rollainville truffle orchard during Winter 2010-2011; Tubme3, mature fruiting bodies of *T. melanosporum* harvested at the Rollainville truffle orchard during Winter 2010-2011; Tubme4, mature fruiting bodies of *T. melanosporum* harvested at the

Rollainville truffle orchard during Winter 2013-2014; Tubmag, mature fruiting bodies of *T. magnatum* harvested in Piedmont (Italy); Tubae, a fruiting body of *T. aestivum* harvested in Grand-Est region (France). Color scale of the heat map from no expression (= 0) in yellow to highest expression in red (= 1). Data were visualized and clustered using R (package HeatPlus). The hierarchical clustering was done by using a binary distance metric and Ward clustering method. See also Supplementary Table S13.