LETTERS

Périgord black truffle genome uncovers evolutionary origins and mechanisms of symbiosis

Francis Martin[1], Annegret Kohler[1], Claude Murat[1], Raffaella Balestrini[2], Pedro M. Coutinho[3], Olivier Jaillon[4–6], Barbara Montanini[7], Emmanuelle Morin[1], Benjamin Noel[4–6], Riccardo Percudani[7], Bettina Porcel[4–6], Andrea Rubini[8], Antonella Amicucci[9], Joelle Amselem[10], Véronique Anthouard[4–6], Sergio Arcioni[8], François Artiguenave[4–6], Jean-Marc Aury[4–6], Paola Ballario[11], Angelo Bolchi[7], Andrea Brenna[11], Annick Brun[1], Marc Buée[1], Brandi Cantarel[3], Gérard Chevalier[12], Arnaud Couloux[4–6], Corinne Da Silva[4–6], France Denoeud[4–6], Sébastien Duplessis[1], Stefano Ghignone[2], Benoît Hilselberger[1,10], Mirco Iotti[13], Benoît Marçais[1], Antonietta Mello[2], Michele Miranda[14], Giovanni Pacioni[15], Hadi Quesneville[10], Claudia Riccioni[8], Roberta Ruotolo[7], Richard Splivallo[16], Vilberto Stocchi[9], Emilie Tisserant[1], Arturo Roberto Viscomi[7], Alessandra Zambonelli[13], Elisa Zampieri[2], Bernard Henrissat[3], Marc-Henri Lebrun[17], Francesco Paolocci[8], Paola Bonfante[2], Simone Ottonello[7] & Patrick Wincker[4–6]

The Périgord black truffle (*Tuber melanosporum* Vittad.) and the Piedmont white truffle dominate today's truffle market[1,2]. The hypogeous fruiting body of *T. melanosporum* is a gastronomic delicacy produced by an ectomycorrhizal symbiont[3] endemic to calcareous soils in southern Europe. The worldwide demand for this truffle has fuelled intense efforts at cultivation. Identification of processes that condition and trigger fruit body and symbiosis formation, ultimately leading to efficient crop production, will be facilitated by a thorough analysis of truffle genomic traits. In the ectomycorrhizal *Laccaria bicolor*, the expansion of gene families may have acted as a 'symbiosis toolbox'[4]. This feature may however reflect evolution of this particular taxon and not a general trait shared by all ectomycorrhizal species[5]. To get a better understanding of the biology and evolution of the ectomycorrhizal symbiosis, we report here the sequence of the haploid genome of *T. melanosporum*, which at ~125 megabases is the largest and most complex fungal genome sequenced so far. This expansion results from a proliferation of transposable elements accounting for ~58% of the genome. In contrast, this genome only contains ~7,500 protein-coding genes with very rare multigene families. It lacks large sets of carbohydrate cleaving enzymes, but a few of them involved in degradation of plant cell walls are induced in symbiotic tissues. The latter feature and the upregulation of genes encoding for lipases and multicopper oxidases suggest that *T. melanosporum* degrades its host cell walls during colonization. Symbiosis induces an increased expression of carbohydrate and amino acid transporters in both *L. bicolor* and *T. melanosporum*, but the comparison of genomic traits in the two ectomycorrhizal fungi showed that genetic predispositions for symbiosis—'the symbiosis toolbox'—evolved along different ways in ascomycetes and basidiomycetes.

The 125-megabase (Mb) genome of *T. melanosporum* is the largest sequenced fungal genome to date[6], but no evidence for whole-genome duplication or large scale dispersed segmental duplications was observed (Supplementary Table 1 and Supplementary Information section 2). The approximately fourfold larger size of the truffle genome compared with other sequenced ascomycetes is accounted for by multi-copy transposable elements (TE) which constitute about 58% of the assembled genome (Fig. 1, Supplementary Figs 5, 6 and 8, Supplementary Information section 3). Estimated insertion times suggest a major wave of retrotransposition at <5 million years ago (Supplementary Fig. 7). TEs are not uniformly spread across the genome, but are clustered in gene-poor or gene-lacking regions (Fig. 1 and Supplementary Fig. 8). The expansion of regions between blocks of protein-coding genes results from an increased density of TEs. The proliferation of TEs within the truffle genome may result from its low effective population size[7] during postglaciation migrations[8] (Supplementary Information section 2.5).

The predicted proteome is in the lower range of sequenced filamentous fungi[6], as only 7,496 protein-coding genes were identified (Supplementary Information section 4). They are mainly located in TE-poor regions and the gene density is heterogeneous when compared with that of other ascomycetes (Fig. 1, Supplementary Figs 8 and 9). Among the predicted proteins, only 3,970, 5,596 and 5,644 showed significant sequence similarity to proteins from *Saccharomyces cerevisiae*, *Neurospora crassa* and *Aspergillus niger*, respectively (Supplementary Fig. 10). This agrees with the predicted ancient separation (>450 Myr ago) of the Pezizomycetes from the other ancestral fungal lineages (Supplementary Fig. 4)[9]. Of the ~5,650 *T. melanosporum* genes that have an orthologue, very few show conservation of neighbouring orthologues (synteny) in at least one of the other species (Supplementary Fig. 11, Supplementary

[1]INRA, UMR 1136, INRA-Nancy Université, Interactions Arbres/Microorganismes, 54280 Champenoux, France. [2]Istituto per la Protezione delle Piante del CNR, sez. di Torino and Dipartimento di Biologia Vegetale, Università degli Studi di Torino, Viale Mattioli, 25, 10125 Torino, Italy. [3]Architecture et Fonction des Macromolécules Biologiques, UMR 6098 CNRS-Universités Aix-Marseille I & II, 13288 Marseille, France. [4]CEA, IG, Genoscope, 2 rue Gaston Crémieux CP5702, F-91057 Evry, France. [5]CNRS, UMR 8030, 2 rue Gaston Crémieux, CP5706, F-91057 Evry, France. [6]Université d'Evry, F-91057 Evry, France. [7]Dipartimento di Biochimica e Biologia Molecolare, Università degli Studi di Parma, Viale G.P. Usberti 23/A, 43100 Parma, Italy. [8]CNR-IGV Istituto di Genetica Vegetale, Unità Organizzativa di Supporto di Perugia, via Madonna Alta, 130, 06128 Perugia, Italy. [9]Dipartimento di Scienze Biomolecolari, Università degli Studi di Urbino, Via Saffi 2 - 61029 Urbino (PU), Italy. [10]INRA, Unité de Recherche Génomique Info, Route de Saint-Cyr, 78000 Versailles, France. [11]Dipartimento di Genetica e Biologia Molecolare & IBPM (CNR), Università La Sapienza, Roma, Piazzale, A. Moro 5, 00185 Roma, Italy. [12]INRA, UMR Amélioration et Santé des Plantes, INRA-Université Blaise Pascal, INRA – Clermont-Theix, 63122 Saint-Genes-Champanelle, France. [13]Dipartimento di Protezione e Valorizzazione Agroalimentare, Università degli Studi di Bologna, 40 126 Bologna, Italy. [14]Dipartimento di Biologia di Base ed Applicata, [15]Dipartimento di Scienze Ambientali, Università degli Studi dell'Aquila, Via Vetoio Coppito 1 - 67100 L'Aquila, Italy. [16]University of Goettingen, Molecular Phytopathology and Mycotoxin Research, Grisebachstrasse 6, D-37077 Goettingen, Germany. [17]INRA, UMR BIOGER-CPP, INRA-Grignon, av Lucien Brétignières - 78850 Thiverval Grignon, France.

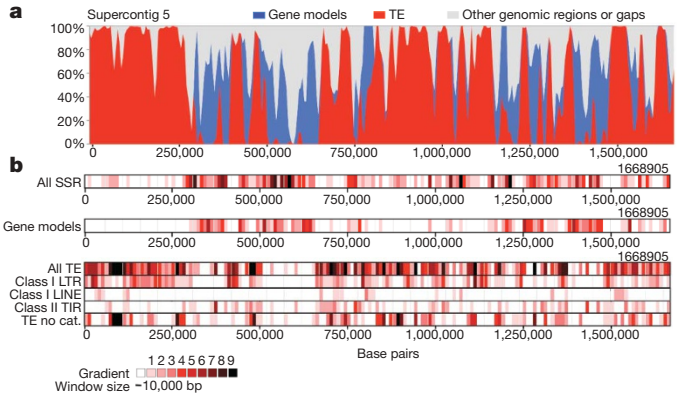

Figure 1 | Genomic landscape of *T. melanosporum*. **a**, The area chart quantifies the distribution of transposable elements (TE) and protein-coding genes (Gene models) along supercontig 5. The *y* axis represents the percentage of base pairs corresponding to TE (red), genes (blue), and other regions and gaps (white) in 10,000-bp sliding windows. **b**, Heat maps display the distribution of selected elements, including simple sequence repeats (SSR), gene models, all TE, long terminal repeat retrotransposons (class I LTR), long interspersed elements (class I LINE), terminal inverted repeats (class II TIR), and unknown TE classes (TE no cat.). Abundance of TE, protein-coding genes and other sequences is represented by a colour scale from 0 (white) to ≥9 occurrences (black) per 10 kbp window.

Information section 5.2). The *T. melanosporum* genome shows a structural organization strikingly different from other sequenced ascomycetes; the largest syntenic region (with *Coccidioides immitis*) only contains 99 genes with 39 orthologues (Supplementary Fig. 12). TE proliferation probably facilitated genome rearrangements. Some regions of meso-synteny were however detected, suggesting that *T. melanosporum* could be used for assessing the genome organization of ancestral ascomycete clades.

Expression of most predicted genes was detected in free-living mycelium, ectomycorrhizal (ECM) root tips and/or fruiting body by custom oligoarrays, expressed-sequence-tag pyrosequencing and Illumina RNA-Seq (Supplementary Information sections 2.4 and 8, Supplementary Table 2, Supplementary Fig. 26). Only a low proportion of transcripts (7.6%) is differentially expressed (fold-ratio ≥4.0, $P < 0.05$) in either ectomycorrhiza or fruiting body by comparison to free-living mycelium (Table 1, Supplementary Table 4). Only 61 transcripts unique to ectomycorrhiza, fruiting body or free-living mycelium were detected (Supplementary Table 5). A few transcripts coding for a H-type lectin, an arabinogalactan protein, a LysM-domain containing protein, major facilitator superfamily (MFS) transporters, laccase/tyrosinase, a lipase and polysaccharide-degrading enzymes are strikingly enriched (>1,000-fold) in symbiotic tissues (Table 1). They may play a role in adhesion to host cells, detoxication of plant defence metabolites, nutrient exchange, and colonization of root apoplast through the deconstruction of cell walls.

A process that is crucial to the success of ECM interactions is the mutualistic exchange of nutrients between the microsymbiont and its host plant. A comparison with other fungi revealed that the total number of predicted transporters is lower in *T. melanosporum* (381 members) compared with *L. bicolor* (491 members) as well as with saprotrophic and pathogenic ascomycetes (481–781 members) (Supplementary Table 26). However, 64 predicted membrane transporters showed an upregulated expression in truffle ectomycorrhizas, suggesting increased fluxes of carbohydrates, oligopeptides, amino acids and polyamines at the symbiotic interface (Supplementary Table 27). PFAM classification of fungal genes induced in symbiotic tissues of either *L. bicolor* or *T. melanosporum* ECM root tips revealed strikingly divergent fungal symbiotic proteomes (Supplementary Fig. 15). However, the PFAM categories corresponding to the MFS transporters (PFAM00083), aquaporin-related major intrinsic proteins (PFAM00230) and amino acid permeases (PFAM000324) were among the most strongly overrepresented in genes that were transcriptionally upregulated in both *L. bicolor* and *T. melanosporum* ectomycorrhizas.

Orthologous genes (that is, reciprocal best hits, BLASTP e-value $\leq 10^{-5}$) significantly induced in the symbiosis represent only 1.5% and

Table 1 | The most highly upregulated transcripts in *T. melanosporum*/*Corylus avellana* ECM root tips

SEQ_ID	ECM level	FB level	FLM level	ECM/FLM ratio	FB/FLM ratio	Definition	Size (AA)	Location	TMD
GSTUMT00012772001	13,262	731	1	**13,262**	731	H-type lectin	279	C	0
GSTUMT00012792001	11,423	250	1	**11,423**	250	Fasciclin-like arabinogalactan protein	414	S	0
GSTUMT00012437001	19,542	180	2	**9,796**	90	Lipase/esterase	346	S	0
GSTUMT00009894001	18,205	2	2	**7,721**	1	Cytochrome P450	396	M	0
GSTUMT00008973001	10,866	2	2	**6,213**	1	Endoglucanase GH5	342	C	0
GSTUMT00008992001	19,305	9	4	**4,538**	2	Laccase	586	S	0
GSTUMT00006890001	5,436	3,787	2	**2,636**	1,836	Sporulation-induced protein	481	C	0
GSTUMT00010076001	2,063	1	1	**2,063**	1	Tyrosinase	604	C	0
GSTUMT00003538001	3,588	2,818	2	**2,050**	1,610	FAD oxidoreductase	564	C	0
GSTUMT00012780001	6,542	44	4	**1,745**	12	LysM domain protein	87	S	0
GSTUMT00009016001	6,737	3,257	5	**1,432**	692	DUF1479-domain protein	379	C	0
GSTUMT00005760001	12,966	989	9	**1,401**	107	Major facilitator superfamily (MFS) permease*	142	S	2
GSTUMT00007927001	1,073	593	1	**1,073**	593	Hypothetical protein	306	C	0
GSTUMT00008954001	19,493	35	18	**1,066**	2	MFS permease	496	C	10
GSTUMT00000763001	902	2	1	**902**	2	Cytochrome P450	397	M	0
GSTUMT00002130001	1,534	1	2	**833**	1	β-Glucan synthesis-associated protein SKN1	575	M	1
GSTUMT00006579001	14,241	8,701	22	**662**	404	DUF1479-domain protein	426	M	0
GSTUMT00010279001	7,848	1,926	12	**651**	160	DUF2235-domain protein	403	C	0
GSTUMT00000499001	49,524	634	77	**644**	8	Phosphatidylserine decarboxylase	316	C	0
GSTUMT00012667001	1,151	7	2	**575**	3	Hypothetical protein	883	C	2
GSTUMT00009500001	2,529	2,846	6	403	453	*Tuber*-specific protein	86	M	0

Upregulation in ectomycorrhizal root tips and fruiting body is assessed by comparing transcript profiles to those from free-living mycelium. Ectomycorrhizal root tips were sampled from five-month-old common hazel (*Corylus avellana* L.) plantlets. Values are the means of seven, four and five biological replicates for free-living mycelium (FLM), ectomycorrhizal root tips (ECM) and fruiting body (FB), respectively. Based on the statistical analysis, a gene was considered significantly upregulated if it met two criteria: (1) *t*-test *P* value <0.05 (ArrayStar, DNASTAR); (2) mycorrhiza or fruiting body versus free-living mycelium fold change ≥4; 571 genes (7.6% of the total gene repertoire) showed an upregulated expression. The highest signal intensity value observed on these arrays was 65,189 arbitrary units. Signals below the cut-off values (100 arbitrary units) were assigned a signal intensity value of 1. Bold font emphasises upregulation in symbiosis. AA, amino acid; C, predicted as cytosolic; S, predicted as secreted; M, predicted mitochondrial protein; TMD, no. of transmembrane domain.
* Truncated sequence. See Supplementary Information section 8 for details.

Table 2 | Orthologous symbiosis upregulated genes of *L. bicolor* and *T. melanosporum*

Tuber gene models	*Lb* gene models (protein ID)	Percentage identity	Putative function	*Tm*- Hazel ECM/FLM ratio	*Lb*-poplar ECM/FLM ratio	*Lb*-Douglas fir ECM/FLM ratio
GSTUMT00008973001	303005	61	Cellulase CBM1-GH5	**6,213**	4	6
GSTUMT00008992001	229839	31	Laccase	**4,538**	8	19
GSTUMT00002130001	239682	38	β-Glucan synthesis-associated protein	**833**	6	4
GSTUMT00005583001	314483	45	MFS permease	**55**	6	8
GSTUMT00005380001	306393	39	DUF1996	**54**	12	22
GSTUMT00005614001	249535	33	Sugar (and other) transporter	**35**	5	8
GSTUMT00010098001	300838	41	Amino acid permease	**33**	18	13
GSTUMT00012011001	247019	36	GH28 pectinase	**29**	5	6
GSTUMT00001268001	308894	35	Hypothetical protein	**23**	5	5
GSTUMT00009221001	187852	37	MFS permease	**7**	8	6
GSTUMT00003668001	256686	53	Sugar (and other) transporter	**6**	12,548	16,648
GSTUMT00000861001	186401	45	Sulphate transporter	**6**	6	5
GSTUMT00010105001	162486	43	Hypothetical protein	**5**	7	5
GSTUMT00007281001	152257	40	Amino oxidase	**5**	5	4

Table shows genes upregulated in ectomycorrhizas (ECM) compared to free-living mycelia (FLM). ECM of *T. melanosporum* were sampled on roots of five-month-old common hazel (*Corylus avellana* L.) plantlets, whereas ECM of *L. bicolor* were sampled on roots of nine-month-old Douglas fir (*Pseudotsuga menziesii*) seedlings or three-month old poplar (*P. trichocarpa*) plantlets[4]. Orthologous genes have been identified by BLASTP (reciprocal best hits, e-value $\leq 10^{-5}$). Transcript profiling was performed on FLM and ECM root tips. Values are the means of seven and four biological replicates, respectively. Based on the statistical analysis, a gene was considered significantly upregulated if it met two criteria: (1) *t*-test *P* value <0.05 (ArrayStar, DNASTAR); (2) ECM versus FLM fold change ≥4. Bold font emphasises upregulation in *Tuber* symbiosis. *Lb*, *L. bicolor*; *Tm*, *T. melanosporum*; MFS, major facilitator superfamily. See Supplementary Information section 8 for details.

4.1% of the ectomycorrhiza-upregulated genes in both *T. melanosporum* and *L. bicolor*, respectively. Most of these rare transcripts code for membrane transporters involved in sugar, amino acid or sulphate uptake (Table 2). This transcriptome trait appears to be a hallmark of the mycorrhizal symbiosis. The resulting increased nutrient flux probably explains the beneficial effect of the symbionts on the growth of their host seedlings (Supplementary Information section 1 and Supplementary Fig. 3). Other overrepresented PFAM categories displayed different patterns in the two symbionts. None of the effector-like small secreted MiSSP proteins specifically expressed in *L. bicolor* ectomycorrhizas[4] were detected among ectomycorrhiza-regulated *T. melanosporum* transcripts.

One of the most striking characteristics of the *T. melanosporum* genome is the almost complete absence of highly similar gene pairs. Of the predicted 7,496 protein-coding genes, only seven pairs share >90% amino-acid identity in their coding sequence, whereas 30 pairs share >80% identity (Supplementary Information section 5.3, Fig. 2). This feature was also observed in the ascomycetous saprotroph *N. crassa*[10]. In striking contrast to the ECM *L. bicolor*[4], multigene families in *T. melanosporum* are limited in number and comprise only 19% of the predicted proteome; most families have only two members (Supplementary Fig. 13). The rate of gene family gain is much lower than the rate of gene loss and among the 11,234 gene families found in ascomycetes, 5,695 appear to be missing in *T. melanosporum* (Supplementary Information section 5.4, Fig. 2). This compact gene coding space may reflect the genome organization of an ascomycete common ancestor, as the Pezizomycetes clade is the earliest diverging lineage within the Pezizomycotina (Supplementary Fig. 4). By comparison to other ascomycetes, gene families predicted to encode metabolite transporters (for example, amino acid and sugar permeases) and secondary metabolism enzymes (such as polyketide synthases and cytochrome P450s) are much smaller. Only 465 genes encoded by expanding gene families of *L. bicolor* are also found in the *T. melanosporum* genome (BLASTP, e-value $\leq 10^{-5}$) and 154 orthologues are shared between expanding gene families of both symbionts. None of them is differentially expressed in ectomycorrhizas. Differences in gene family expansion, in particular dynamic repertoires of genes encoding symbiosis-regulated effector-like proteins and sugar-cleaving enzymes (see below), are probably responsible for different symbiotic traits between *T. melanosporum* and *L. bicolor*, such as altered host specificity. The compact genome of *T. melanosporum* might be a product of selection for specialization; this is because genome expansion, as observed in *L. bicolor*, is probably driven by selection on the symbiont to exploit a diversity of encountered substrates provided by multiple potential hosts and by their diverse soils[4,5].

The volatiles released by truffles are attractive to rodents and truffle flies[11], which disperse their spores, but also to humans who consider this elusive mushroom a delicacy. *T. melanosporum* is the first sequenced fungus producing highly flavoured hypogeous fruiting bodies (Supplementary Information section 6.4, Fig. 3). Genomic signatures of the long-standing (>2,000-year-old) reputation of the

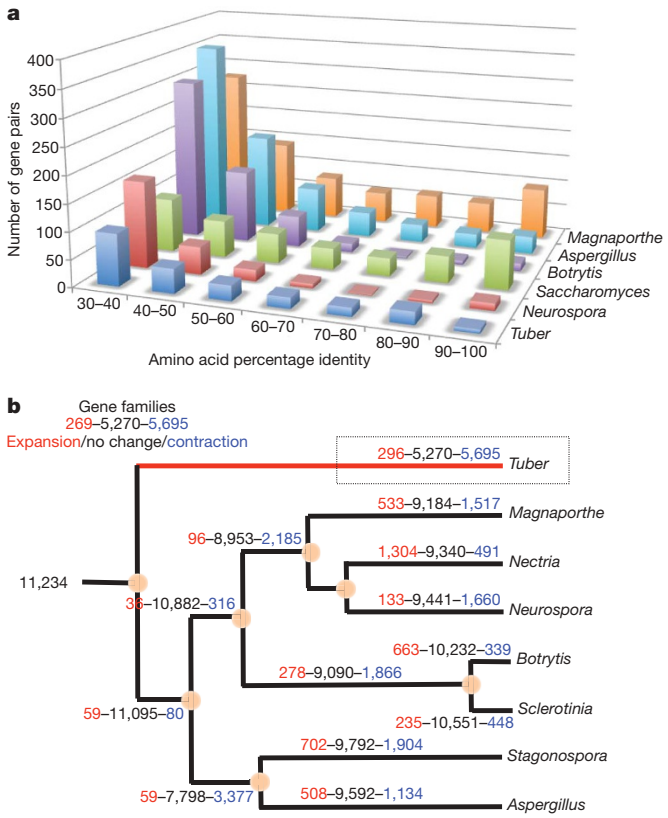

Figure 2 | Genome redundancy in the truffle genome. a, The percentage of amino-acid identity of the top-scoring self-matches for protein-coding genes in *T. melanosporum, Saccharomyces cerevisiae, Aspergillus nidulans, Neurospora crassa, Magnaporthe grisea,* and *Botrytis cinerea.* For each fungus, the protein-coding regions for each gene were compared with those of every other gene in the same genome using BLASTX. **b**, The figure represents the total number of protein families in each species or node. 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.

black truffle as a gastronomic delicacy are its extremely low allergenic potential (Supplementary Fig. 18), coupled with the lack of key mycotoxin biosynthetic enzymes (Supplementary Information section 6.2, Supplementary Table 14), and the preferential overexpression of various flavour-related enzymes in the fruiting body (Supplementary Figs 19–21). Among the latter are specific subsets of sulphur assimilation and S-amino acid interconversion enzymes. These include cystathionine lyases known to promote the side-formation of methyl sulphide volatiles abundant in truffles[12] as well as various enzymes involved in amino acid degradation through the Ehrlich pathway which are giving rise to known truffle volatiles and flavours, for example, 2-methyl-1-butanal (Fig. 3, Supplementary Information section 6.4, Supplementary Figs 20 and 21). Also notable, given the subterranean habitat of this fungus, is the presence of various putative light-sensing components (Supplementary Information section 6.6), which might be involved in light avoidance mechanisms and/or in the control of seasonal developmental variations, especially those related to fruiting body formation and sexual reproduction.

The analysis of genes implicated in the mating process, including pheromone response, meiosis and fruiting body development showed that most sex-related components identified in other ascomycetes are also present in *T. melanosporum* (Supplementary Table 11). Sexual reproduction in ascomycete filamentous fungi is partly controlled by two different mating-type (*MAT*) genes that establish sexual compatibility[13]: one *MAT* gene codes for a protein with an α-box domain, whereas the other encodes a high mobility group (HMG) DNA binding protein (Supplementary Information section 6.5). It was widely believed that *T. melanosporum* was a homothallic or even an exclusively selfing species[14]. The sequenced Mel28 strain contains the HMG locus, and the opposite linked *MAT*α locus was identified in another natural isolate (Supplementary Fig. 22), confirming recent hints that *T. melanosporum* is heterothallic and thus an obligate outcrossing species[15]. This result has major implications for truffle cultivation, which will be improved by the use of host plants harbouring truffle strains of opposite mating types. In most ascomycetes, the genomic regions flanking the *MAT* locus show an extended conservation[13], but there is no synteny of the *MAT* loci between *T. melanosporum* and other sequenced fungi (Supplementary Fig. 23).

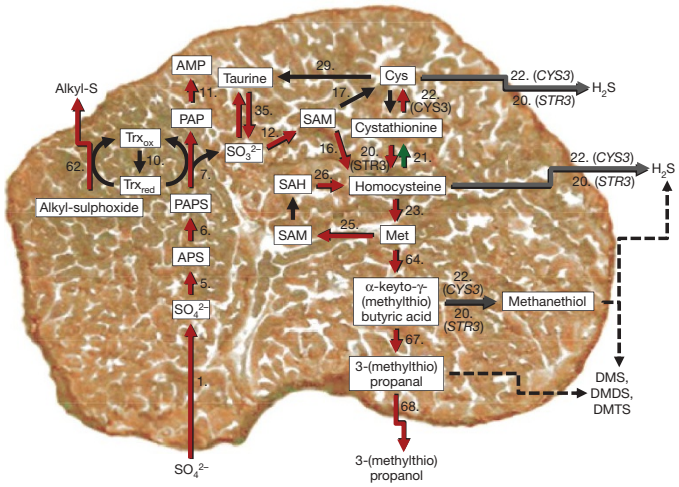

Figure 3 | Outline of sulphur metabolism in *T. melanosporum* fruiting body. Numbers identify enzymes and gene models as specified in Supplementary Table 15. Reactions (arrows) catalysed by enzymes whose mRNAs are upregulated in fruiting bodies are shown in red; mRNAs downregulated by at least twofold in fruiting bodies, or whose expression levels differ by less than twofold compared to free-living mycelia and ectomycorrhizas are represented by green and black arrows, respectively. Off-pathway cystathionine-γ-lyase (no. 22)- and cystathionine β-lyase (no. 20)-supported reactions, and spontaneous (non-enzymatic) breakdown reactions are indicated by grey and dashed arrows, respectively. APS, adenosine phosphosulphate; PAPS, phosphoadenosine phosphosulphate; PAP, phosphoadenosine phosphate; Trx, thioredoxin; SAM, S-adenosylmethionine; SAH, S-adenosylhomocysteine; DMS, dimethylsulphide; DMDS, dimethyl-disulphide; DMTS, dimethyl-trisulphide.

To determine whether *T. melanosporum* sugar-cleaving capabilities resemble those of other fungi, we have undertaken a comparison of the glycoside hydrolase (GH) and polysaccharide lyase (PL) repertoires[16] of 18 completely sequenced fungi (Fig. 4). As expected for a symbiotic fungus living in the root apoplast, *T. melanosporum* has a relatively small number of GH-encoding genes (91 members; Supplementary Tables 23 and 24); much fewer than phytopathogens (for example, *Magnaporthe grisea*, *Fusarium graminearum*) and saprotrophs (for example, *N. crassa*, *Podospora anserina*). The *T. melanosporum* GH repertoire bears some similarity with that of *L. bicolor*[4], especially a reduced spectrum of enzymes targeting the plant cell wall compared to saprobes, culminating in both fungi with the absence of cellulases from families GH6 and GH7. There are however significant differences in the spectrum of enzymes present in these two symbiotic fungi. For instance, *T. melanosporum* has hemicellulases from families GH10 and GH43, whereas *L. bicolor* has none. Similarly, *T. melanosporum* has a family GH45 cellulase that is absent from the *L. bicolor* genome. Other differences include different strategies to cleave pectin: whereas *L. bicolor* utilizes six hydrolytic GH28 pectinases, *T. melanosporum* has only two, but these are complemented by three pectin lyases and a pectin methylesterase that are missing in *L. bicolor*. Both fungi have a set of proteins, few in number, bearing cellulose-binding domains, but differences appear here too: the single cellulose-binding CBM1 motif of *L. bicolor* is appended to a GH5 endoglucanase, whereas *T. melanosporum* has two CBM1 motifs attached to a GH61 enzyme and to a protein of unknown function. GH61 enzymes have been reported to display weak cellulolytic activity[17].

An extended comparison with other sequenced fungi (Fig. 4) shows that *T. melanosporum* clusters neither with *L. bicolor* nor with saprotrophic ascomycetes, most probably because of its limited overall number of GHs and PLs that make it closer to yeasts and fungi that do not interact with plants, but rather with animals (*Cryptococcus neoformans*, *Malassezia globosa*). Differences between the enzyme repertoires of *T. melanosporum* and *L. bicolor* suggest differences in the mode of interaction of the two symbionts with their respective host plants. A striking difference is the presence of an invertase gene in *T. melanosporum*, whereas *L. bicolor* has none and is therefore completely dependent on its host for its provision of glucose[5]. In contrast, *T. melanosporum* could access and hydrolyse the plant-derived sucrose. This would suggest that although both fungi develop symbiotic relationships with plants, *T. melanosporum* is probably less dependent than *L. bicolor*. The overall pattern of induction of genes coding for enzymes acting on polysaccharides is similar in both *L. bicolor* and *T. melanosporum* symbiotic transcriptomes, although a larger number of carbohydrate-cleaving enzyme transcripts are upregulated for some families—for example, GH16 (β-1,6-glucanases), GH18 (chitinases) and GT20 (α,α-trehalose-phosphate synthase) in *L. bicolor* (Supplementary Table 25 and Supplementary Fig. 24). Intriguingly, a GH5 cellulase and a GH28 pectinase are among the rare transcripts that are highly upregulated in both *L. bicolor* and *T. melanosporum* ectomycorrhizas, suggesting that they play a key role in the symbiosis. On the other hand, the β-glucan synthesis-associated protein present in both ectomycorrhizas is involved in fungal cell wall remodelling[16] and may play a role in the alteration of cell wall surface during infection to conceal the hyphae from the host.

The ability to establish ECM symbioses is a widespread characteristic of various ascomycetes and basidiomycetes[3]. The truffle genome reveals features of an ancestral fungal lineage that diverged from other lineages >450 Myr ago[9]. Despite their similar symbiotic structures and similar beneficial effects on plant growth, the ascomycete *T. melanosporum* and the basidiomycete *L. bicolor* encode strikingly

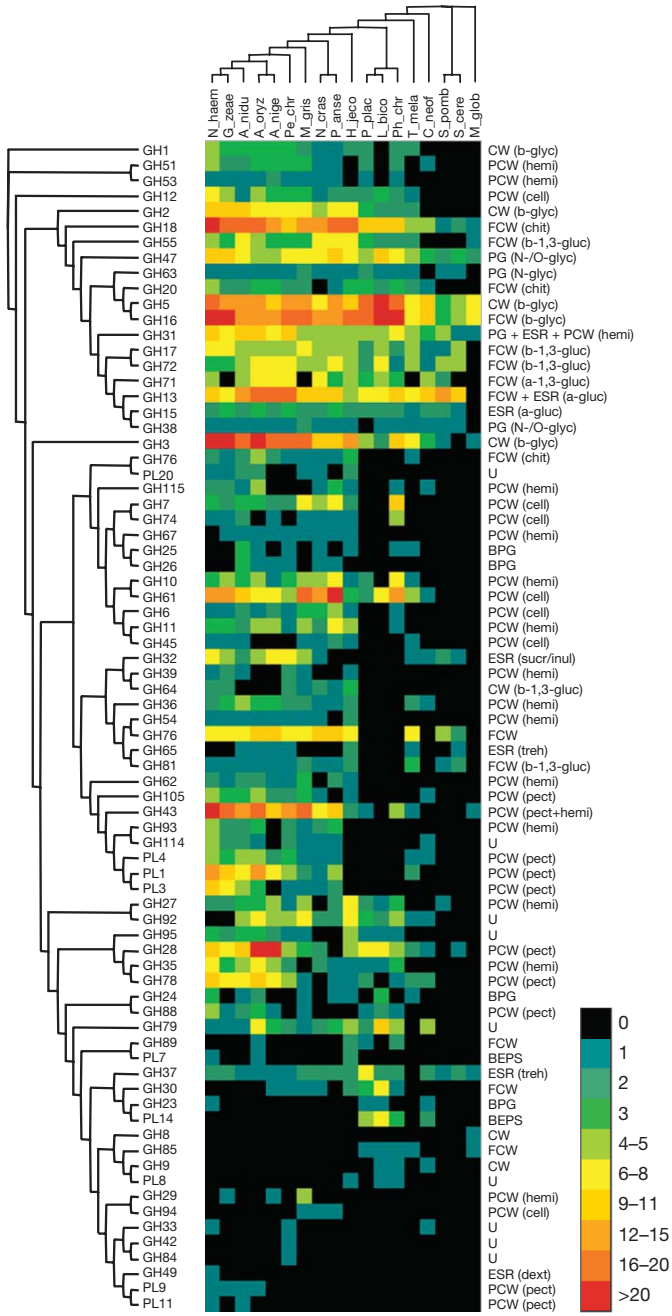

Figure 4 | Double clustering of the carbohydrate-cleaving families from representative fungal genomes. Top tree: the fungi named are *Aspergillus nidulans* (A_nidu), *Aspergillus niger* (A_nige), *Aspergillus oryzae* (A_oryz), *Cryptococcus neoformans* (C_neof), *Gibberella zeae* (G_zeae), *Hypocrea jecorina* (H_jeco), *Laccaria bicolor* (L_bico), *Magnaporthe grisea* (M_gris), *Malassezia globosa* (M_glob), *Nectria haematococca* (N_haem), *Neurospora crassa* (N_cras), *Penicillium chrysogenum* (Pe_chr), *Phanerochaete chrysosporium* (Ph_chr), *Podospora anserina* (P_anse), *Postia placenta* (P_plac), *Saccharomyces cerevisiae* (S_cere), *Schizosaccharomyces pombe* (S_pomb), and *Tuber melanosporum* (T_mela). Left tree: the enzyme families are represented by their class (GH, glycoside hydrolase; PL, polysaccharide lyase) and family number according to the carbohydrate-active enzyme database[16]. Right side: known substrate of CAZy families (most common forms in brackets): BPG, bacterial peptidoglycan; BEPS, bacterial exopolysaccharides; CW, cell wall; ESR, energy storage and recovery; FCW, fungal cell wall; PCW, plant cell wall; PG, protein glycosylation; U, undetermined; a-gluc, α-glucans (including starch/glycogen); b-glyc, β-glycans; b-1,3-gluc, β-1,3-glucan; cell, cellulose; chit, chitin/chitosan; dext, dextran; hemi, hemicelluloses; inul, inulin; N-glyc, N-glycans; N-/O-glyc, N- / O-glycans; pect, pectin; sucr, sucrose; and tre, trehalose. Abundance of the different enzymes within a family is represented by a colour scale from 0 (black) to ≥20 occurrences (red) per species.

different proteomes—compact with very few multigene families, versus large with many expanded multigene families—and symbiosis-regulated genes. Effector-like proteins, such as the *L. bicolor* ECM-induced SSP MiSSP7 (ref. 4), are not expressed in *T. melanosporum* ectomycorrhizas. On the basis of our results, the ECM symbiosis appears as an ancient innovation that developed several times during the course of Mycota evolution using different 'molecular toolkits'[18]. Sequencing of the *T. melanosporum* genome has provided unprecedented insights into the molecular bases of symbiosis, sex and fruiting in a most popular representative of the only lifestyle not yet addressed by Ascomycota genomics[19]. This sequencing will be a major step in moving truffle research into the realm of ecosystem science, and a deeper understanding of the genome of the Périgord black truffle is expected to have substantial social and cultural impact.

METHODS SUMMARY

A whole-genome shotgun strategy was adopted for sequencing and assembling the *T. melanosporum* genome (Supplementary Information section 2). All genomic DNA was obtained from the homokaryotic haploid strain Mel28. All data were generated by paired-end sequencing of cloned 3 kb and 10 kb inserts using Sanger technology. The pool of data available for the assembly consisted of 1,262,177 reads, with ~1,250 Mb of sequence. The data were assembled using the ARACHNE assembler. The 4,464 contigs (N50 = 62 kb) were assembled in 398 supercontigs (N50 = 638 kb) corresponding to 124.946 Mb of sequence. The main genome scaffolds were at a depth of 10. Assemblies and annotations are available at INRA (http://mycor.nancy.inra.fr/IMGC/TuberGenome/) and Genoscope (www.genoscope.cns.fr/tuber).

The GAZE pipeline selected a best representative gene model for each locus on the basis of expressed-sequence-tag support and similarity to known proteins from other organisms, and predicted 7,496 protein-coding gene models (Supplementary Information section 4). All predicted genes were annotated using Gene Ontology and KEGG pathways. Protein domains were predicted using InterProScan. Gene families were built from proteins using Tribe-MCL.

Single dye labelling of cDNAs, hybridization procedures, data acquisition, background correction and normalization of custom-exon expression arrays were performed at the NimbleGen facilities following their standard protocol. A Student *t*-test with false discovery rate was applied to the data using the ARRAYSTAR software (DNASTAR). Transcripts with a significant *P* value (<0.05) and ≥4-fold change in transcript level were considered as differentially expressed in ECM root tips or fruiting body.

Received 20 August 2009; accepted 28 January 2010.
Published online 28 March 2010.

1. Mello, A., Murat, C. & Bonfante, P. Truffles: much more than a prized and local fungal delicacy. *FEMS Microbiol. Lett.* **260,** 1–8 (2006).
2. Hall, I. R., Brown, G. T. & Zambonelli, A. *Taming the Truffle. The History, Lore, and Science of the Ultimate Mushroom* (Timber Press, Portland, 2007).
3. Smith, S. E. & Read, D. J. *Mycorrhizal Symbiosis* 3rd edn (Academic, 2008).
4. Martin, F. *et al.* The genome of *Laccaria bicolor* provides insights into mycorrhizal symbiosis. *Nature* **452,** 88–92 (2008).
5. Martin, F. & Selosse, M. A. The *Laccaria* genome: a symbiont blueprint decoded. *New Phytol.* **180,** 296–310 (2008).
6. Galagan, J. E., Henn, M. R., Ma, L. J., Cuomo, C. A. & Birren, B. Genomics of the fungal kingdom: insights into eukaryotic biology. *Genome Res.* **15,** 1620–1631 (2005).
7. Lynch, M. & Conery, J. S. The origins of genomic complexity. *Science* **302,** 1401–1404 (2003).
8. Murat, C. *et al.* Polymorphism at the ribosomal DNA ITS and its relation to postglacial re-colonization routes of the Perigord truffle *Tuber melanosporum*. *New Phytol.* **164,** 401–411 (2004).
9. Taylor, J. W. & Berbee, M. L. Dating divergences in the fungal tree of life: review and new analyses. *Mycologia* **98,** 838–849 (2006).
10. Galagan, J. E. & Selker, E. U. RIP: the evolutionary cost of genome defense. *Trends Genet.* **20,** 417–423 (2004).
11. Maser, C., Claridge, A. W. & Trappe, J. M. *Trees, Truffles and Beasts* (Rutgers Univ. Press, 2008).
12. Liu, M., Nauta, A., Francke, C. & Siezen, R. J. Comparative genomics of enzymes in flavor-forming pathways from amino acids in lactic acid bacteria. *Appl. Environ. Microbiol.* **74,** 4590–4600 (2008).
13. Fraser, J. A. & Heitman, J. Evolution of fungal sex chromosomes. *Mol. Microbiol.* **51,** 299–306 (2004).
14. Bertault, G., Raymond, M., Berthomieu, A., Callot, G. & Fernandez, D. Trifling variation in truffles. *Nature* **394,** 734 (1998).
15. Riccioni, C. *et al. Tuber melanosporum* outcrosses: analysis of the genetic diversity within and among its natural populations under this new scenario. *New Phytol.* **180,** 466–478 (2008).

16. Cantarel, B. L. *et al.* The Carbohydrate-Active EnZymes database (CAZy): an expert resource for glycogenomics. *Nucleic Acids Res.* **37,** D233–D238 (2009).
17. Karkehabadi, S. *et al.* The first structure of a glycoside hydrolase family 61 member, Cel61B from *Hypocrea jecorina*, at 1.6 Å resolution. *J. Mol. Biol.* **383,** 144–154 (2008).
18. Hibbett, D. S. & Matheny, P. B. Relative ages of ectomycorrhizal mushrooms and their plant hosts. *BMC Biol.* **7,** 13 (2009).
19. Soanes, D. M. *et al.* Comparative genome analysis of filamentous fungi reveals gene family expansions associated with fungal pathogenesis. *PLoS One* **3,** e2300 (2008).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank the late L. Riousset and C. Dupré for providing the Mel28 isolate, and acknowledge F. Le Tacon and J. Weissenbach for continuous support. The genome sequencing of *T. melanosporum* was funded by the Génoscope, Institut de Génomique, CEA, and Agence Nationale de la Recherche (ANR). Genome annotation and transcriptome analysis were supported by INRA, the European FP6 Network of Excellence EVOLTREE, Région Lorraine, the ANR FungEffector project, Fondazione Cariparma, Compagnia di San Paolo-Torino, the Italian Ministry of Education, University and Research (MIUR), Regione Umbria and Instituto Pasteur Fondazione Cenci Bolognetti. We thank D. Hibbett and J. Heitman for comments on an early draft of the manuscript, and J. Plett for a critical reading of the paper.

Author Contributions B.H., M.-H.L., F.P., P. Bonfante, S.O. and P.W. contributed equally to this work as senior authors; A.K., C.M., R.B., P.M.C., O.J., B.M., E.M., B.N., R.P., B.P. and A.R. contributed equally to this work as second authors. F.M. initiated the project and coordinated the genome annotation, data analysis and manuscript preparation; P.W. coordinated the sequencing and automated annotation at Genoscope. F.M. and S.O. wrote the manuscript with input from P. Bonfante. R.B., P.M.C., B.H., O.J., A.K., M.H.L., B.M., E.M., C.M., B.N., F.P., R.P., A.R. and P.W. also made substantial contributions (listed in alphabetical order). All other authors are members of the genome sequencing consortium and contributed annotation, analyses or data throughout the project, and are listed in alphabetical order. We also thank A. Bonfigli, M. Buffalini, S. Colafarina, T. Flutre, S. Kamal, P. Ceccaroli, C. Roux, R. Saltarelli, S. von Pall di Tolna and O. Zarivi for their assistance in annotation.

Author Information Genome assemblies together with predicted gene models and annotations have been deposited at DNA Data Bank of Japan/European Molecular Biology Laboratory/GenBank under the project accession numbers CABJ01000001–CABJ01004455 (whole genome shotgun sequencing data) and FN429986–FN430383 (scaffolds and annotations). The complete expression dataset is available as series (accession number GSE17529) at the Gene Expression Omnibus at NCBI. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to F.M. (fmartin@nancy.inra.fr).