



At the nexus of three kingdoms: the genome of the mycorrhizal fungus *Gigaspora margarita* provides insights into plant, endobacterial and fungal interactions

Francesco Venice,¹ Stefano Ghignone ,² Alessandra Salvio di Fossalunga,¹ Joëlle Amselem,³ Mara Novero,¹ Xie Xianan,⁴ Kinga Sędziewska Toro,⁵ Emmanuelle Morin,⁶ Anna Lipzen,^{7,8} Igor V. Grigoriev,⁷ Bernard Henrissat,^{9,10,11} Francis M. Martin⁶ and Paola Bonfante ^{1*}

¹Department of Life Sciences and Systems Biology, University of Turin, Turin, Italy.

²Institute for Sustainable Plant Protection-CNR, Turin Unit, Turin, Italy.

³URGI, INRA, Université Paris-Saclay, Versailles, France.

⁴State Key Laboratory for Conservation and Utilization of Subtropical Agro-Bioresources, Key Laboratory of Innovation and Utilization of Forest Plant Germplasm in Guangdong Province, College of Forestry and Landscape Architecture, South China Agricultural University, Guangzhou, China.

⁵Genetics, Faculty of Biology, Ludwig-Maximilians-University Munich, Planegg-Martinsried, Germany.

⁶Institut National de la Recherche Agronomique (INRA), Laboratory of Excellence Advanced Research on the Biology of Tree and Forest Ecosystems (ARBRE), UMR, 1136, Champenoux, France.

⁷Department of Energy Joint Genome Institute, Walnut Creek, CA, USA.

⁸Department of Plant and Microbial Biology, University of California, Berkeley, Berkeley, CA, 94720, USA.

⁹Architecture et Fonction des Macromolécules Biologiques, CNRS, Aix-Marseille Université, Marseille, 13288, France.

¹⁰Institut National de la Recherche Agronomique, USC1408 Architecture et Fonction des Macromolécules Biologiques, Marseille, F-13288, France.

¹¹Department of Biological Sciences, King Abdulaziz University, Jeddah, 21589, Saudi Arabia.

Summary

As members of the plant microbiota, arbuscular mycorrhizal fungi (AMF, Glomeromycotina) symbiotically colonize plant roots. AMF also possess their own microbiota, hosting some uncultivable endobacteria. Ongoing research has revealed the genetics underlying plant responses to colonization by AMF, but the fungal side of the relationship remains in the dark. Here, we sequenced the genome of *Gigaspora margarita*, a member of the Gigasporaceae in an early diverging group of the Glomeromycotina. In contrast to other AMF, *G. margarita* may host distinct endobacterial populations and possesses the largest fungal genome so far annotated (773.104 Mbp), with more than 64% transposable elements. Other unique traits of the *G. margarita* genome include the expansion of genes for inorganic phosphate metabolism, the presence of genes for production of secondary metabolites and a considerable number of potential horizontal gene transfer events. The sequencing of *G. margarita* genome reveals the importance of its immune system, shedding light on the evolutionary pathways that allowed early diverging fungi to interact with both plants and bacteria.

Introduction

Arbuscular mycorrhizal symbiosis involves 72% of vascular plants and an ancient group of fungi (arbuscular mycorrhizal fungi, AMF) that enhance host fitness by improving mineral nutrition and resistance to biotic and abiotic stresses (Bonfante, 2018). The large number of sequenced plant genomes and availability of mutants allowed many of the mechanisms underlying plant responses to colonization by AMF to be deciphered (MacLean *et al.*, 2017; Lanfranco *et al.*, 2018; Pimprikar and Gutjahr, 2018). By contrast, the biology of the fungal partners remains largely unexplained. AMF are important members of the plant microbiota, present in the rhizosphere and inside roots tissues as obligate symbionts

(Davison *et al.*, 2015). As many other fungi, AMF also possess their own microbiota (Bonfante *et al.*, 2019), often hosting uncultivable endobacteria inside their cytoplasm (Bonfante and Desirò, 2017; Pawlowska *et al.*, 2018). AMF genomes, therefore, offer an unexplored source for investigating the evolution of inter-kingdom interactions among plants, fungi and bacteria.

The genome of the model fungus *Rhizophagus irregularis* was the first AMF to be sequenced (Tisserant *et al.*, 2013; Lin *et al.*, 2014; Chen *et al.*, 2018), offering novel insights into the genetics of a member of the Glomeromycotina, such as the surprising absence of plant cell wall-degrading enzymes and the unexpected presence of genes potentially involved in sexual reproduction. Genome sequences of other AMF followed: *Rhizophagus clarus* (Kobayashi *et al.*, 2018), *Rhizophagus cerebriforme* and *Rhizophagus diaphanus* (Morin *et al.*, 2019), as well as two members of the Diversisporales: *Diversispora epigea* (Sun *et al.*, 2019) and *Gigaspora rosea* (Morin *et al.*, 2019). Irrespective of their phylogenetic position in the AMF tree, these novel genomes consistently confirmed the obligate biotrophic nature of AMF, revealing the loss of genes involved in plant polysaccharide degradation and fatty acid biosynthesis. Most of the sequenced *Rhizophagus* strains, as well as *G. rosea*, do not possess endobacteria (Bonfante and Desirò, 2017), but *D. epigea* and *R. clarus* contain Mycoplasma-Related Endobacteria (MRE) (Naumann *et al.*, 2010; Naito *et al.*, 2015). Here, we present the assembly and annotation of the genome of *Gigaspora margarita*, which belongs to the Gigasporaceae (Krüger *et al.*, 2012), as *G. rosea*, and is a member of an early diverging AMF group, well separated from Glomeraceae. It presents a life cycle with a conspicuous extraradical phase consisting of huge spores, auxiliary cells and extraradical mycelium, as well as an intraradical phase where intercellular hyphae support production of arbuscules inside cortical cells (Fig. S1). Intraradical vesicles are not formed and its presence in natural environments is mostly limited to sandy soils (Stürmer *et al.*, 2018). *Gigaspora margarita* may host *Burkholderia*-related Endobacteria (BRE) and MRE communities (Bianciotto *et al.*, 1996; Desirò *et al.*, 2014), as well as viral sequences (Turina *et al.*, 2018). The isolate used here contains a single population of non-cultivable BRE endobacteria, identified as *Candidatus Glomeribacter gigasporarum* (CaGg), whose genome has been sequenced (Ghignone *et al.*, 2012).

By means of an extended comparative genomic analysis, we reveal that the genome sequence of *G. margarita* isolate BEG34 shares the peculiarities of the other AMF species (conserved core genes, absence of plant cell wall-degrading enzymes, absence of type I fatty acid synthase), but is also remarkably different from all other sequenced fungal genomes in its genome size (predicted to be 831 Mbp by flow cytometry) large number of

transposable elements (TEs); in addition, several horizontal gene transfer (HGT) events have been detected similar to other Diversisporaceae (Sun *et al.*, 2019). Thanks to the availability of OMICs data generated by using fungal lines with and without bacteria (Salvioli *et al.*, 2016; Vannini *et al.*, 2016; Dearth *et al.*, 2018), the genome sequence of *G. margarita* sheds light on the evolutionary signatures of adaptation that have allowed this early diverging fungus to interact with both plants and bacteria since Devonian times (Mondo *et al.*, 2012; Strullu-Derrien *et al.*, 2018). Novel features, i.e. presence of an immune system; secondary metabolite production, and refined tuning of proteins and enzymes that target chitin, offer insights into the biotrophic lifestyle of *G. margarita*. By providing a deeper understanding of inter-kingdom interactions, the genomic information of *G. margarita* allows us to advance hypotheses on the ecological and evolutionary meaning of AMF genomes.

Results and discussion

Separation of the G. margarita nuclear genome sequence from mitochondria and endosymbiont genomes

We sequenced the genome of *G. margarita* BEG34 using one paired-end library (Illumina TruSeq Nano) and two mate-pair (3 and 8 kbp) libraries (Illumina Nextera). Sequencing of *G. margarita* BEG34 produced a first assembly of ~875 Mbp consisting of 6453 scaffolds.

By using BlobTools (Laetsch and Blaxter, 2017) analysis for the visualization and partitioning of metagenome assemblies, scaffolds were visualized as circles of variable size, separated on the basis of their coverage and G/C content, and their taxonomic affiliation was inferred through sequence homology. This analysis allowed us to separate the nuclear component, which is the focus of this study (Fig. 1), from the other compartments containing DNA (mitochondria and CaGg endobacteria, Fig. S2). The mitochondrial reads were re-assembled into a complete circular chromosome (96 998 bp), agreeing with the results of Pelin *et al.* (2012), while the CaGg reads led to an improved assembly (2.07 Mbp) compared with the previous version (Ghignone *et al.*, 2012). The nuclear genome, obtained by reassembling the nuclear reads, consists of 6490 scaffolds and is ~758 Mbp (Table 1), almost six times larger than that of the model species *R. irregularis*, and significantly larger than those of the saprotrophic and phylogenetically related *Mortierella elongata* that has a genome size of around 49 Mbp (Uehling *et al.*, 2017). According to NCBI and JGI Mycocosm (Table S1), the size of the *G. margarita* genome is second only to that of *Austropuccinia psidii* (Pucciniomycotina, Basidiomycota) at 1.2 Gbp, for

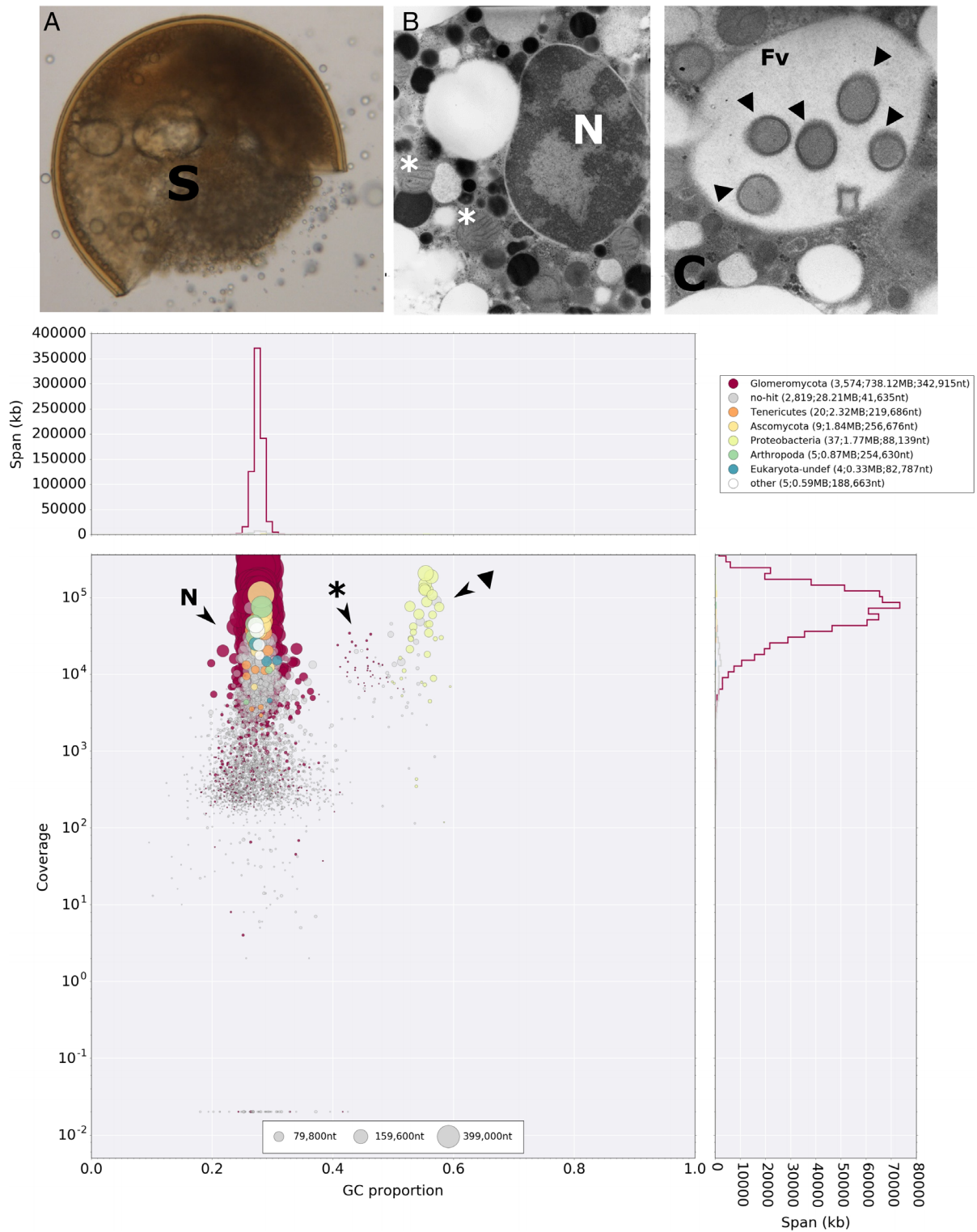


Fig. 1. *Gigaspora margarita* BEG34 as a metaorganism. A.

A *Gigaspora margarita* quiescent spore crushed to allow the spreading of nuclei and cytoplasm.

B. A section of the fungal spore seen under an electron microscope: mitochondria (asterisks) and fungal nuclei (N) are highlighted.

C. CaGg (arrowheads) cells are situated inside the fungal vacuole (Fv). Below, the same components are represented as metagenome scaffolds as a result of the Blobtools analysis, which integrates coverage, G/C content and sequence homology information as a method to identify and separate the different biological entities of the metagenome. Nuclear, mitochondrial and bacterial scaffolds are visualized as circles (blobs) with a diameter corresponding to their size in base pairs. The blobs are distributed on the basis of their G/C content (horizontal axis) and their coverage in terms of genomic reads (vertical axis) and coloured according to their taxonomic affiliation as determined by sequence homology. The higher G/C content in mitochondrial and bacterial genomes is evident from the separation on the horizontal axis. The sequence homology search also confirmed the classification based on the other parameters.

Table 1. Assembly and annotation statistics of the *G. margarita* BEG34 genome.

Flow cytometry genome size	~831 Mbp
Genome size	773.104 Mbp
GC	27.68%
Average scaffold length	~119.122 kbp
N50	326.786 kbp (743scaffolds)
N70	212.988 kbp (1329 scaffolds)
N90	95.586 kbp (2737 scaffolds)
Portion with gaps (Ns)	7.33%
Number of predicted genes	26 603
Number of predicted transcripts	31 568
Benchmarking Universal Single-Copy Orthologs score	98.50%

which no functional annotation is available (McTaggart *et al.*, 2018).

This result was supported by flow cytometry, which revealed a DNA quantity for *G. margarita* corresponding to a genome of about 831 Mbp (Fig. S3). Indeed, previous static cytometry suggested a comparable value (Bianciotto and Bonfante, 1992). Similar to other AMFs, the *G. margarita* genome has a G + C content below 30% (27.68%).

TEs dominate the *G. margarita* genome

The nuclear assembly revealed that repeated sequences cover 64%–86% of the assembled genome sequence, depending on the filtering of highly degenerated copies (Fig. 2, Fig. S4). These repeats can be ascribed to 4264 consensus sequences, representing the putative ancestral TE sequences from which they originated. The

proportion of TEs is very high when compared with the generally low value found in fungal genomes (Castanera *et al.*, 2016), usually ranging from 0% to 25% with the exception of some plant pathogens (Spanu *et al.*, 2010) and ectomycorrhizal species that have undergone massive TE amplifications (Martin *et al.*, 2010). Among the latter, Tuberaceae show a genome expansion that correlates with TEs increase, accounting for 50% of the genomic content (Murat *et al.*, 2018). Similar to the related *G. rosea* (Morin *et al.*, 2019), *G. margarita* confirms this relationship well: the huge genome size is accompanied by the highest proportion of TEs detected so far in AMF; TEs in the model *R. irregularis* (Chen *et al.*, 2018) represent 22%–26% of the total assemblies (from 122 to 138 Mbp).

Overall, 63% of the *G. margarita* TE content (40.6% of genome) remains unclassified. Of the remainder, the largest group of TEs is classified as class I retrotransposons (12% of the genome) (Fig. 2, Fig. S4), where the major classes are LINE (8%) and LTR Gypsy (2.5%) elements. At least one full-length Gypsy element (g2860) was found to be active based on gene expression data (Fig. S5). Since these groups are well represented in *G. rosea* (Morin *et al.*, 2019) and *D. epigaea* (Sun *et al.*, 2019) compared to *Rhizophagus* species, a Diversisporaceae-specific expansion might be hypothesized.

Among the DNA-repeats, at least 15 full-length Toll/interleukin-1 receptors/resistances (TIRs) were expressed, revealing a differential expression across the fungal life stages (Fig. S5, Table S2). Another peculiar feature of *G. margarita* DNA-TEs consists in the presence of repeats with homology to the *sola* class, which is present in plants, bacteria and metazoa, but previously considered to be absent in fungi, except for *R. irregularis* (Gladyshev and Kleckner, 2017). Lastly, *G. margarita*

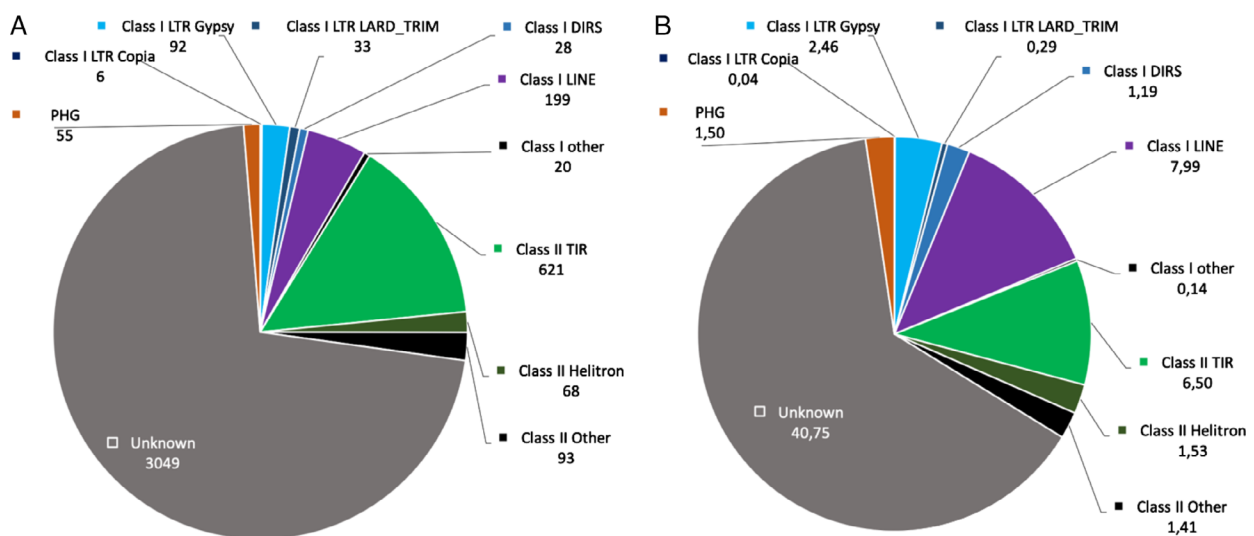


Fig. 2. Distribution of the *G. margarita* TE families according to their main order or superfamily (Gypsy/Copia) in the consensus Library (4/264 consensus; (A) and according to their genome coverage (602/060 copies/ 64% of the genome; (B) PHG stands for 'Potential host genes'.

genome harbours Helitrons, which are characterized by a rolling circle replication mechanism and which may contain genes captured from other organisms. They have been demonstrated to be mediators of HGT in other fungi (HGT, see specific paragraph) (Castanera *et al.*, 2014).

Usually, TE presence is counterbalanced by host genome defences, including fungal repeat-induced point mutation (RIP), methylation induced pre-meiotically (MIP), meiotic silencing of unpaired DNA (MSUD) and quelling (Muszewska *et al.*, 2017). When compared to their ancestor sequences (consensus sequences), the *G. margarita* TE fragments had low frequency of nucleotide transitions (C to T, specifically; Fig. S6a), which are operated by DNA-methylases and are the outcome of RIP activity (Amselem *et al.*, 2015). Furthermore, AT-rich regions could not be statistically detected (Fig. S6b), suggesting that *G. margarita* lacks RIP activity, while the impact of MIP remains to be evaluated. We inferred the absence of MSUD activity as no homology could be found with the associated genes so far characterized in *Neurospora crassa*; by contrast, small RNA-mediated TEs silencing (quelling) cannot be excluded, as *G. margarita* genome encodes for the genes related to this pathway (see the Immune System section), such as Dicer and Argonaute proteins.

It has been suggested that gene-sparse regions in fungal genomes contain small secreted proteins (SSPs), which are often associated with TEs, providing a favourable environment for the diversification of fungal effector repertoire

(Sánchez-Vallet *et al.*, 2018). The *G. margarita* genome is also gene-sparse (Fig. S7a); a further screening revealed a significant spatial association between *G. margarita* candidate SSPs (Table S3) and CRYPTON, TIR and MITE (Table S4, Fig. S7b). In conclusion, a glimpse at the TEs present in *G. margarita* genome reveals many peculiar features: the heterogeneous TEs groups suggest that Diversisporaceae experienced some specific TE bursts; Helitrons could be involved in HGT events; lastly, the TE abundance mirrors the absence of fungal genome defence barriers, confirmed by the expression of some TE typologies.

Genomic features of *Gigaspora margarita*

The gene prediction process combined *ab initio* prediction and hints from physical evidence (see the Methods section). Exon coordinates were determined by genome mapping of the entire set of UniProt Mucoromycota proteins; intron/exon boundaries were defined by spliced alignment of 10 *G. margarita* RNA-seq libraries from different biological conditions (Table S5). Finally, non-exon regions were identified using coordinates of repeated regions. We identified 26 603 coding genes (Table 1). The Program to Assemble Spliced Alignments pipeline refined 11 487 gene models, detecting 4965 transcript isoforms, leading to 31 568 non-redundant transcripts. The Benchmarking Universal Single-Copy Orthologs assessment (Waterhouse *et al.*, 2018) detected 98.5% of the conserved fungal gene

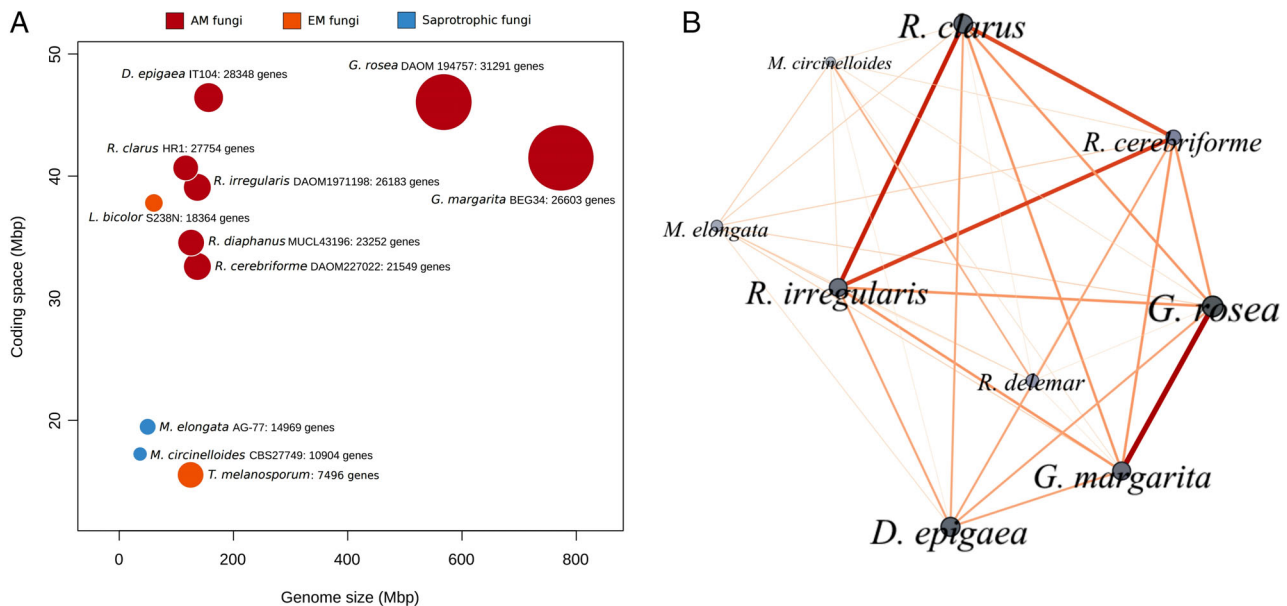


Fig. 3. (A) Comparison of the genomes of *G. margarita* and related fungi. The comparison includes the other sequenced AMF (red circles), free-living Mucoromycota (blue circles), and mycorrhizal fungi from distant clades (orange circles). The size of the circles and their position on the horizontal axis is proportional to the genome sizes and the coding space (the space occupied by protein-coding genes in each genome) is plotted on the vertical axis; (B) a network showing the correlation of *G. margarita* with some of its relatives. In this analysis, the correlation was measured as the number of orthologs between the considered species. The size of the nodes is proportional to the number of proteins in each species and the size/colour of the lines depends on the number of orthologs (the thicker/darker, the higher the number of orthologs).

set as complete sequences, indicating that this gene catalogue is highly comprehensive. The total number of genes found for *G. margarita* was close to that of the other Glomeromycotina so far sequenced, *R. irregularis*, *R. clarus*, *R. cerebriforme*, *R. diaphanus*, *D. epigea* and *G. rosea* (Chen *et al.*, 2018; Kobayashi *et al.*, 2018; Morin *et al.*, 2019; Sun *et al.*, 2019). Figure 3A highlights the exceptionally high genome size of the two *Gigaspora* species, when compared with other AMF, ectomycorrhizal fungi and free-living Mucoromycota. Alongside *G. margarita* and *G. rosea*, the 10 largest fungal genomes (Table S1) include other obligate biotrophs (rusts from Pucciniomycotina), an endophytic fungus and a gut fungus (*Zoophthora radicans* and *Neocallimastix californiae* respectively). Several of these genomes lack annotation, leaving open the question of whether their expansions also implied an increase in gene number. AMF possess more genes than their free-living Mucoromycota relatives; however, plant symbiotic lifestyle and gene number increase do not seem to be strictly related, since ectomycorrhizal truffles (Murat *et al.*, 2018) and obligate plant pathogens (Spanu *et al.*, 2010) possess a reduced gene number. Coding space, defined as the space occupied by the protein-coding genes, is another informative parameter: as expected, among the analysed AMF, *D. epigea* and *G. rosea* possess the largest coding space, in line with their higher gene number (Fig. 3A). A similarity network built up on groups of orthologs (Fig. 3B), and based on gene composition rather than gene number, revealed a considerable distance between AMF genera (*Gigaspora*, *Diversispora* and *Rhizophagus*), while the intra-genus diversity was low. These quantitative parameters suggest that, irrespectively of genome size and comparable gene number, Gigasporaceae and Glomeraceae are characterized by different genes.

Crossing genomics with transcriptomics to describe gene expression depending on the fungal life cycle and endobacterial presence

Having predicted the gene repertoire of *G. margarita*, we combined gene information with a large set of transcriptomic data (Table S5). In *G. margarita* native line containing the endobacterium (B+), ~82% of the predicted genes were expressed at the transcript level at least in one of the examined biological conditions: both asymbiotic (spores germinated in H₂O, or in the presence of a synthetic analogue of plant strigolactones, GR24) and symbiotic (intraradical and extraradical mycelium) stages (Fig. S8; Table S6). Around 57% (15,205) of genes were expressed under all biological conditions. Pre-symbiotic and symbiotic stages were characterized by 2431 and 885 unshared genes respectively. Genes encoding enzymes (such as oxidases, carbohydrate-active enzymes and peptidases), transporters and even

SSPs were specifically expressed in these two stages, suggesting a transcriptional shift involved in the development, nutrient acquisition and communication with the host. Most of the expressed symbiotic genes have no known function, as observed in other AMF (Morin *et al.*, 2019). A significant number of genes (444), which were expressed at least in one of the examined conditions in the B+ line, were never expressed in the cured fungal line (B–, without the endobacteria) (Table S6). Similarly, 38 genes were exclusively expressed in the B– line. While most of these B+ or B– specific genes (~63%) are functionally uncharacterized, others encode for enzymes putatively involved in DNA binding, replication and transcription, such as zinc finger domain-containing proteins, far1-related proteins and HMG-box transcription factors, which are indicated as regulators of cryptic sexuality events in AMF (Ropars *et al.*, 2016). A number of genes encoding for peptidases are exclusively expressed in the B+ line, as well as two genes encoding for hydrolases with peptidoglycan as predicted substrate (glycoside hydrolase family 25 proteins).

The following sections will highlight the evolution of gene families and the diversification of relevant and/or novel gene categories in the context of *G. margarita* lifestyle, including biotrophy and endobacterium presence.

Evolution of gene families in Gigasporaceae

We used the Orthofinder (Emms and Kelly, 2015) and CAFE (De Bie *et al.*, 2006) algorithms and a set of fungi to reconstruct the AMF phylogeny and compare close relatives possessing different lifestyles (Table S7). Alongside AMF, we included free-living Mucoromycota, plus two ectomycorrhizal fungi (*Tuber melanosporum* and *Laccaria bicolor*) from distant clades and a saprotroph/opportunistic pathogen (*Aspergillus fumigatus*), considered as outgroup. The pipeline first performed a homology search between the entire proteomes of the selected species; MCL clustering (Enright *et al.*, 2002) was then used to process the homology results, generating 17 341 orthogroups, i.e. gene families gathering orthologs, incorporating ~81% of the 287 540 analysed proteins. Roughly, 84% of *G. margarita* proteins were clustered within these orthogroups. Inter-species distances were calculated through multiple sequence alignment for each orthogroup containing one protein per species (533 single-copy orthogroups). These distances were summarized in a phylogenetic tree, converted into evolutionary time, and finally used by CAFE to detect, for each node and leaf of the tree, orthogroups that underwent accelerated gene gains or losses, interpreted in this context as the effects of evolutionary pressure. The tree was calibrated using the divergence date of 434 MYA between *M. elongata* and *R. irregularis* (Uehling *et al.*, 2017). The

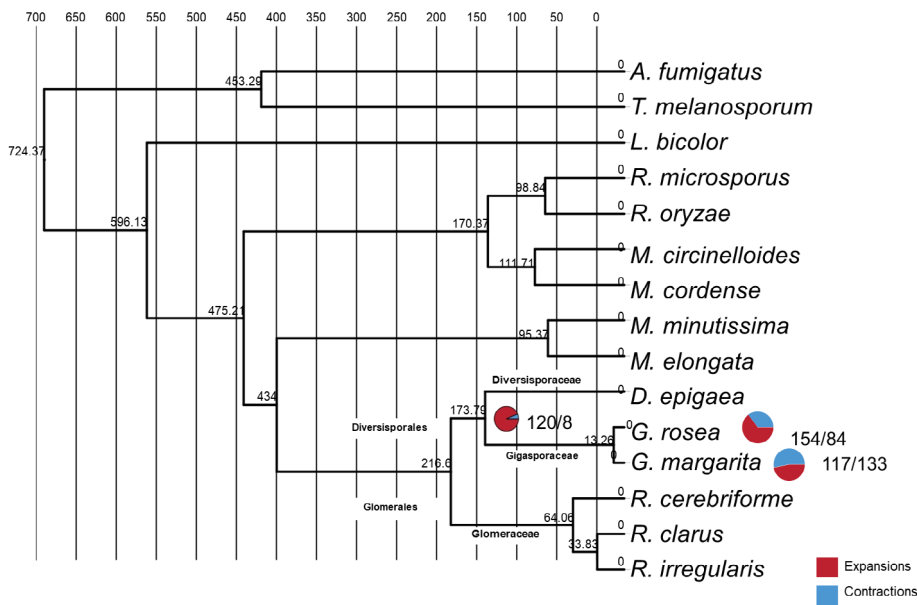


Fig. 4. Genome-scale phylogeny as reconstructed by orthogroups-based data. The ultrametric tree shows the strong phylogenetic relationship between the two *Gigaspora* species, which diverged in recent evolutionary time (the divergence date is shown next to the branches). The divergence between the *Gigaspora* species and the other sequenced AMF groups (Diversisporaceae and Glomerales) is also shown. The leftmost pie chart displays the number of gene families that underwent rapid evolution (red for expansions and blue for contractions) after the divergence between the *Gigaspora* and *Diversispora epigaea* (Diversisporaceae). The pie charts on the right of the tree display the families that rapidly evolved in *G. margarita* or *G. rosea* after the two species diverged.

reconstructed phylogenetic tree (Fig. 4) indicates that *G. margarita* and *G. rosea* diverged very recently, at around 13 MYA. Different from Glomerales, for which all the sequenced fungi belong to the genus *Rhizophagus* within the Glomeraceae, a divergence node (~174 MYA) between Diversisporaceae (*D. epigaea*) and Gigasporaceae (*G. margarita* and *G. rosea*) is present in Diversisporales. Most of the estimated phylogenetic distances resemble those obtained by Chang *et al.* (2019) in the context of a deeper phylogenetic analysis, which was designed for the Endogonales family, also belonging to Mucoromycota. Looking for Gigasporaceae-specific traits, we focused on expansion/contraction events that took place after divergence from the Diversisporaceae. Of the 271 expansion events observed in the *Gigaspora*, 120 originated in *Gigaspora* versus *D. epigaea* differentiation (Table S8a, Table S9 and Fig. S9a-m). Among them, the *Gigaspora* species possess two expanded gene families containing alcohol oxidases which, in brown rot fungi, participate in Fenton chemistry (Guillén *et al.*, 2000; Hernández-Ortega *et al.*, 2012), and contribute to the non-enzymatic degradation of plant cell walls. The analysis also highlighted the presence of a *Gigaspora*-specific feature, i.e. the potential to synthesize secondary metabolites due to the presence of Polyketide synthases, which are normally absent in basal fungi (see also the paragraph on HGT). A rapid gene gain for *Gigaspora* species was particularly dramatic in the families of oligopeptide transporters and of patatins, which can be related to fungal immunity (see the specific paragraph).

After the divergence between *G. rosea* and *G. margarita* (Fig. 4), our pipeline identified 117 families that underwent rapid expansion and 133 that underwent rapid contraction in *G. margarita* (Table S8b, Table S9

and Fig. S9a-m). Gene family evolution in *G. margarita* versus *G. rosea* may have been shaped by the interaction of *G. margarita* with CaG, which—so far—has never been detected in *G. rosea*. Gene loss events in *G. margarita* include proteins related to DNA replication and repair, while among the 117 rapidly expanded families, we found a family of fungal-like Nod-like receptors identified by a central NACHT domain and putatively involved in non-self recognition (discussed later), and MATA-HMG proteins (OG0000316). This latter family of highly diverse transcription factors is widespread among AMF genomes (Morin *et al.*, 2019) and is suggested to regulate cryptic sexuality events in AMF (Ropars *et al.*, 2016). According to our pipeline, MATA-HMG of AMF are clustered in more than one orthogroup, but in OG0000316 *G. margarita* possesses the highest number of proteins; furthermore, two of these genes (g24252 and g7196) are upregulated at the transcript level in germinating spores containing CaGg, when compared to cured spores (Table S9). While it is already demonstrated that the expression of MATA-HMG can be influenced by BRE in their Mucoromycota hosts (Mondo *et al.*, 2017), the putative effects of such regulation remain elusive for *G. margarita*. The expanded family of multicopper oxidases of the AA1 class (OG0000106) gathers two genes (g3919 and g11237) that are strongly upregulated in the B+ line during the mycorrhizal symbiosis, and are also putatively secreted (Table S3). These extracellular oxidases are particularly abundant in plant-interacting Basidiomycota (Kües and Rühl, 2011), where they mediate plant cell wall depolymerization. We speculate that multicopper oxidases might have a role during the interaction of *G. margarita* with the cell wall of its plant host, and hypothesize that the endobacterial presence may have shaped this

symbiotic interplay. Finally, g11471, present in the expanded gene family of immunoreactive mannoproteins (OG0000217), has the highest upregulation in response to the endobacterial presence during the germination phase (Table S9).

In conclusion, the genomes of Gigasporaceae have diversified not only from those of the Glomerales (Mondo *et al.*, 2017) but also from that of *D. epigaea* (Sun *et al.*, 2019). In addition, *G. margarita* shows specific evolution of its gene repertoire when compared with the related *G. rosea*: since the sequenced *G. rosea* isolate does not host bacteria, such events could be related to the presence of the endobacterium CaGg.

As the other AMF, whose obligate biotrophy has been shaped by a lack of specific pathways (Tisserant *et al.*, 2013), *G. margarita* genome does not possess invertase as well as genes involved in fatty acid and thiamine biosynthesis (Supplementary text). The genetic basis for the unculturability in fungi may also involve the lack of other genes, for example those for spermidine and biotin biosynthesis (Ahrendt *et al.*, 2018). This does not seem to be the case of *G. margarita*, since its genome encodes for spermidine and biotin synthases (g1668, g3373 and g11043), all apparently functional on the basis of their consistent expression throughout the life cycle (Table S6). However, a closer inspection revealed that these sequences are all intersected by TEs: the upstream genomic region of both spermidine synthases contains fragments of TIR transposons, while biotin synthase carries the exonic insertion of an unclassified repeat. We conclude that these insertion events did not disrupt gene functionality in *G. margarita*.

Not only phosphate transporter genes: G. margarita has a large number of phosphate-related genes

The iconic function of AMF is characterized by their capacity to take up phosphate from the soil and transfer it to plants through their phosphate transporters (PTs) (Smith and Read, 2008; Ezawa and Saito, 2018). Surprisingly, however, the molecular characterization of fungal PTs is very limited (Harrison and van Buuren, 1995; Xie *et al.*, 2016). Mining the *G. margarita* genome revealed a very rich genetic machinery related to phosphate metabolism, sensing, transport and signalling (Table S10). We first focused on phosphate signal transduction (PHO) genes since, different from metazoa where cells are regularly supplied with phosphate and, as a consequence, lack a PHO network (Lev and Djordjevic, 2018), AMF preferentially thrive in phosphate-depleted conditions and their capacity to successfully interact with their host plants depends strictly on the P content. Under certain conditions, high Pi concentrations block symbiosis establishment (Balzergue *et al.*, 2013; Fiorilli *et al.*, 2013). Comparison of PHO proteins belonging to many fungal groups revealed that some members of the PHO

cascade are absent or poorly represented in the *G. margarita* genome (for instance, SPL2, PHO4 and PHO89 homologues), while other components are enriched (Table S10). *Gigaspora margarita* encodes for at least 11 PT isoforms (Fig. 5, red dots), which are grouped into two distinct lineages. Accordingly, nine Pi transporters, PT (g21463) to PT5 (g26234) and their three paralogs PT8 (g19323) to PT10 (g6532), are closely related to the fungal PHO84 PTs and clustered with Glomeromycotina PTs belonging to the Mucoromycota PHO84-like subfamily. Two other PTs, PT6 (g10792) and PT7 (g26017), which contain SPX domains, are very closely related to the PHO87/90/91 PTs derived from yeast. On the basis of the classification (Fig. 5) and expression (Fig. S10, Table S11) of these *G. margarita* PTs, we speculate that PT1 and PT2, as well as PT4 and PT9, may contribute to Pi uptake from the environment, while PT6-PT7 might be responsible for Pi homeostasis during symbiosis, similar to the SPX-Pi transporters in yeast (Secco *et al.*, 2012). The heatmap drawn on *G. margarita* PTs expression (Fig. S10) revealed that most of the transporters belonging to the Glomeromycotina lineage were expressed in the extraradical phase, confirming the largely acknowledged concept that AMF uptake Pi from the soil. PT (g21463) is also consistently expressed during the symbiotic phase. Pho1 (g17792), encoding PHO1-type Pi transporter, is heavily expressed during the intraradical stage, where it has a predicted role in Pi unloading (or export) from fungus to the plant (Xie *et al.*, 2016). These data might confirm results from laser-dissected cells revealing the unexpected expression of fungal PTs in arbusculated cells (Balestrini *et al.*, 2009). Pi uptake is driven by H⁺-ATPases (Ezawa and Saito, 2018), which are expressed in *G. margarita*, not only at the fungal/substrate interface but also during the symbiotic phase (g4412, g3891; Fig. S10). Once inside the fungus, Pi is polymerized into vacuolar polyphosphate granules maybe thanks to vacuolar transporter chaperons (VTC), which are evenly expressed throughout the *G. margarita* life cycle (Table S11, Fig. S10).

On the basis of the genomic and transcriptomic data, we conclude that *G. margarita* possesses a consistent number of functional PTs, confirming its capacity to uptake Pi from the soil; cellular Pi homeostasis is probably maintained through a rather homogenous expression of two SPX-PT transporters (g10792-PT6, g26017-PT7), which could re-uptake Pi from the periarbuscular and intercellular spaces (Fig. 6, Fig. S10). Interestingly, some ATPases and PT transporters are more expressed when the fungus contains its endobacterium (Table S11); in particular, the HA2 ATPase (g4412) is upregulated by the presence of the endobacterium in both germinating spores and intraradical mycelium. Furthermore, the vacuolar VTC1-2 (g24765) is upregulated by the presence of CaGg in germinating spores and GR24-treated spores, suggesting a role in moving the Pi from the fungus to

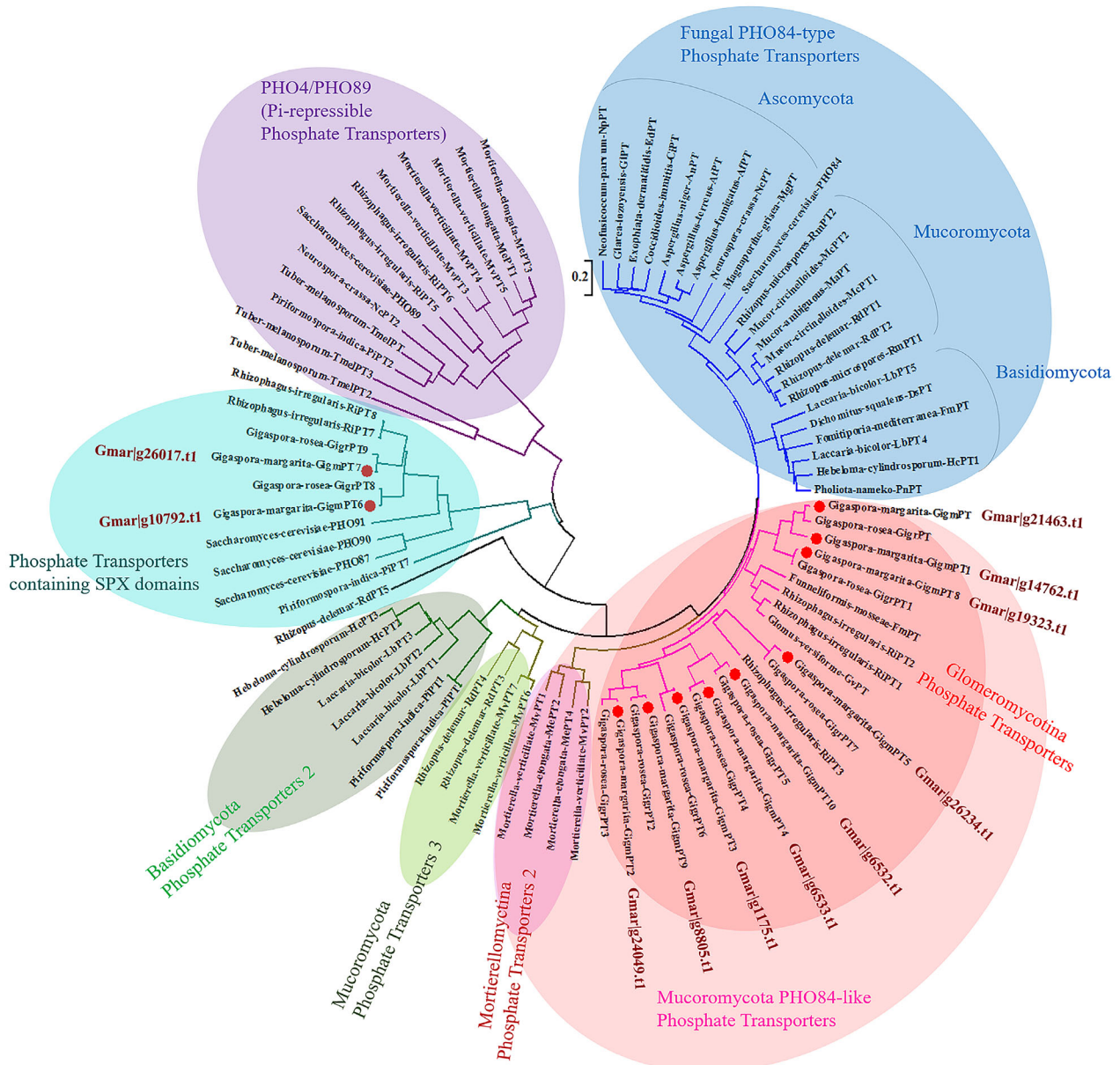


Fig. 5. Evolutionary relationships of phosphate transporters among fungal species. The evolutionary history was inferred using the Neighbour-Joining method 149. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Poisson correction method and are in the units of number of amino acid substitutions per site. The analysis involved 82 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 58 positions in the final data set. Evolutionary analyses were conducted in MEGA7. The marked red circles shown in the figure indicate the multiple isoforms of phosphate transporters from *G. margarita*.

bacterium, which is compartmented inside fungal vacuoles (Fig. S2), and which possesses its own Pi transporter (Ruiz-Lozano and Bonfante, 1999; Ghignone *et al.*, 2012). These results are in agreement with previous data showing that *G. margarita* with its endobacterium has a generally more active metabolism, eventually leading to a higher Pi concentration in the plant host (Salvioli *et al.*, 2016).

Chitin-related genes: molecular tools forming the basis of fungal growth and communication with the plant

Chitin, beta-1,3-glucans, beta-1,6-glucans and mannoproteins are usually listed as the major components of fungal cell walls (Gow *et al.*, 2017). Biochemical analyses are not available for AMF. *In situ* labelling showed chitin in the cell wall in all steps of the AMF life cycle (Bonfante,

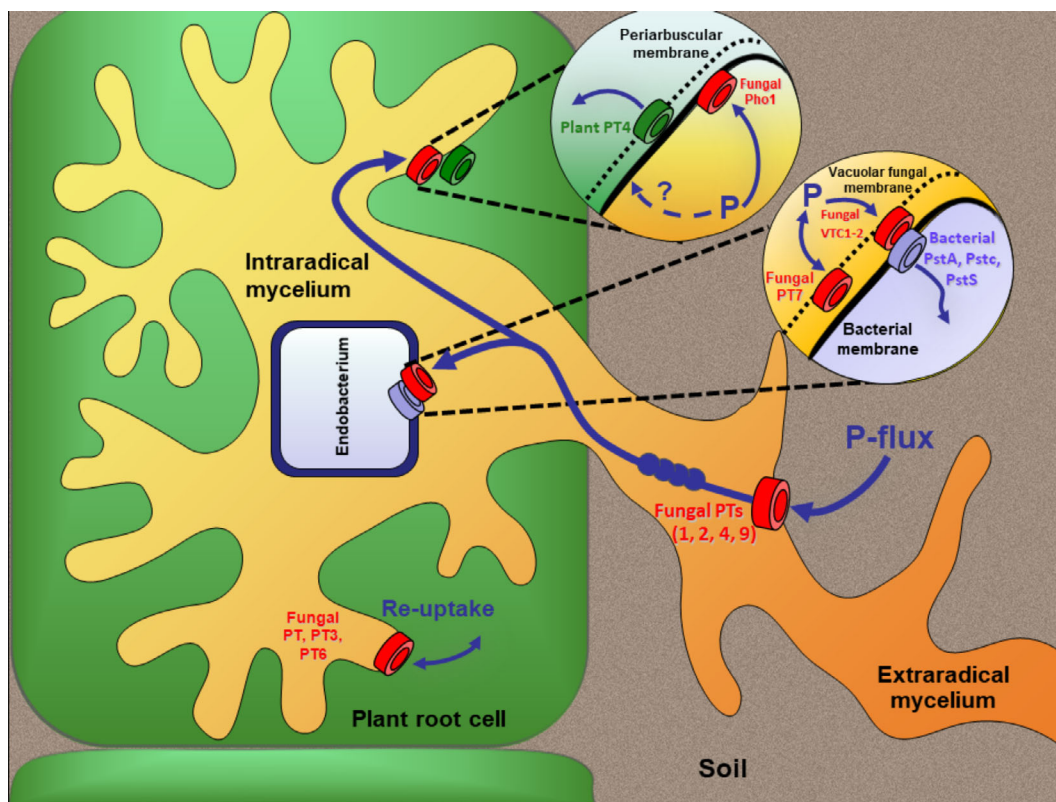


Fig. 6. By crossing genomics and transcriptomic data, the scheme illustrates a hypothetical model of how *G. margarita* handles Pi. PT1-2-4-9, supported by ATPases, may take up phosphate from the soil; the phosphate is aggregated in polyphosphate granules and released at the interfaces with the plant and with the bacterium. The insets reveal the differences between the two contact areas: at the plant interface, the fungus is surrounded by the perifungal membrane, which is continuous with the plant plasma membrane. Here, the fungal Pho1 could mediate Pi export from the intraradical mycelium into the periarbuscular space. Due to the plant PT4 activity, Pi is loaded into plant cells. Pi diffusion may also occur across the perifungal membrane. In addition, PT, PT3 and PT6 could also contribute to Pi re-uptake in the intraradical mycelium under low Pi levels. The bacterium is located inside the fungal vacuole: several fungal vacuolar export chaperones are actively expressed, and the one that mostly seemed to respond to the CaGg presence is VTC1-2. The endobacterium, in turn, has two PTs located at the outer membrane (PstA, PstC), and one located in the periplasmic space (PstS). Furthermore, the bacterium has cytoplasmic components that may respond to intracellular Pi (PstB and PhoU). All the membrane localisations of the fungal P transporters are hypothetical.

2018), while monoclonal antibodies failed to detect beta-1,3-glucans (Lemoine *et al.*, 1995; Ligrone *et al.*, 2007). Indeed, homologues of FSK, the beta-1,3-glucan synthase, were not detected among the protein-coding genes of the *G. margarita* genome, while mannoproteins could be biosynthesized by the numerous enzymes present in the GT15 family (Table S12) and involved in mannosyltransferase activity. The GT2 family of *G. margarita* has 38 members, which is high compared with the other AMF sequenced so far. Among these, 15 show characteristics of chitin synthases: presence of the CON1 region (Liu *et al.*, 2017), and domain organization leading to further divisions and types (Fig. S11). The number of secreted Carbohydrate-Active Enzymes (CAZymes) involved in fungal wall deconstruction is very high in *G. margarita* (Table S3) and probably guarantees the cell wall dynamics described morphologically (Bonfante, 2018); in addition to many 1,2-alpha-mannosidases (GH92), which seem to be specific for Gigasporaceae

(Table S12), many CAZymes are devoted to chitin breakdown. Chitinases are encoded by genes present in the family GH18, which are abundant in Gigasporales genomes.

Chitin can be deacetylated by chitin deacetylases encoded by CE4 family (Fig. 7). Glucosaminidase encoded by GH20 genes might lead to the production of glucosamine residues.

GT2, GH18 and CE4 family members, which are particularly expanded in the *G. margarita* genome (Table S12), seem to be expressed homogeneously during the *G. margarita* life cycle (Fig. S12, Table S13). However, some revealed specific behaviour: g25383 (GH18) was particularly activated in the pre-symbiotic phase and sensitive to strigolactones; expression of others was enhanced during the symbiotic phase, preferentially in the presence of the endobacterium (g19206, g23868); some CE4 members, g7314, g19924 and g19811, had their highest expression in the intraradical phase.

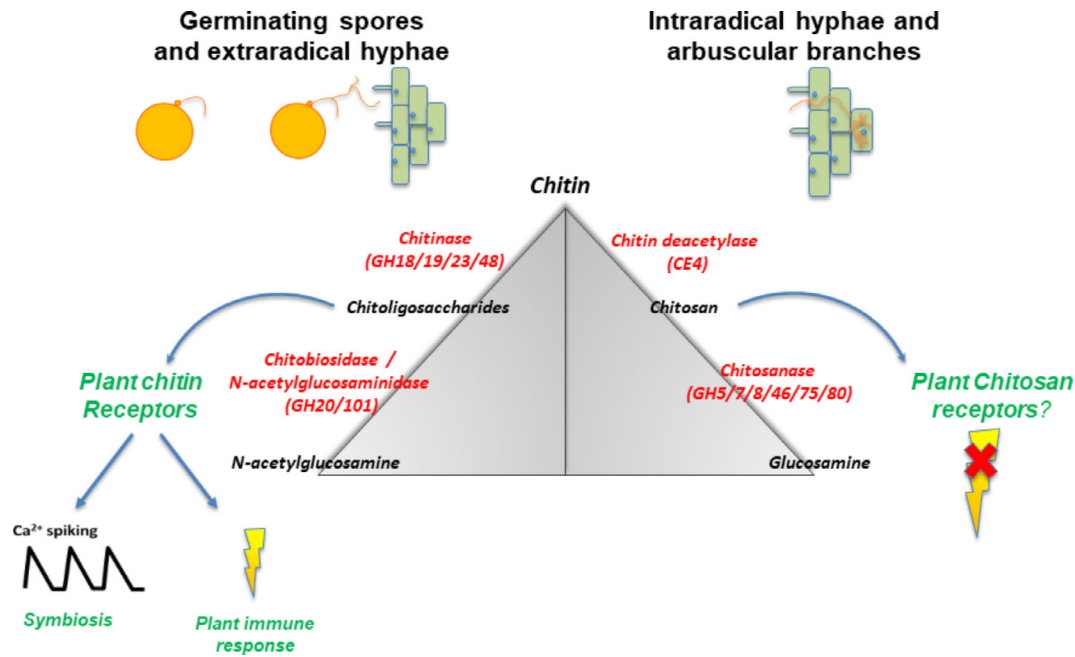


Fig. 7. *Gigaspora margarita* has two pathways for chitin breakdown. On one side, the activity of chitinases leads to the formation of chitoligosaccharides, which are recognized by plant chitin receptors.

Chitin has multiple roles in fungi as a structural and signalling molecule (Schmitz and Harrison, 2014; Pusztahelyi, 2018). This is particularly relevant in AMF, where chitin organization is modulated from the extraradical to the intraradical phase (Bonfante, 2018) and chitoligosaccharides (COs) are the main signalling molecules (Zipfel and Oldroyd, 2017). During the pre-symbiotic phase and in the extraradical symbiotic mycelium, *G. margarita* produces COs, which may or not be decorated by lipid chains. This is enabled by the GH18 family, supported by the action of CBM14, which represent the chitin-binding domains of chitinases. *N*-acetylated COs are recognized by plant chitin receptors, (Miyata *et al.*, 2014; Zhang *et al.*, 2015; Carotenuto *et al.*, 2017), activating a conserved downstream symbiotic signal transduction pathway. Here, one of the first signatures is a calcium spiking response, which is activated by *N*-acetylated COs and enhanced by treatment with the synthetic strigolactone GR24 (Genre *et al.*, 2013). However, chitin not only acts as a signal for symbiosis but also for pathogenicity (Sánchez-Vallet *et al.*, 2015). *N*-Acetyl COs released by AMF activate host plant defences, inducing immunity signalling (Pozo and Azcón-Aguilar, 2007; Martínez-Medina *et al.*, 2016; Chialva *et al.*, 2018). Many studies have demonstrated that such plant defences are limited to the first interaction phase (Giovannetti *et al.*, 2015), questioning whether *N*-acetyl COs are no longer active during intraradical colonization. The presence of expressed CE4 members suggests a

new hypothesis: the deacetylation process could be more important during the intraradical phase when changes in molecular organization (loss of acetyl groups) would correspond to loss of chitin fibrillar structure, as seen under the electron microscope. Gow *et al.* (2017) wrote that deacetylated chitin is not recognized by plant chitin receptors, allowing a deep colonization of plant tissues without any evident rejection. Another alternative hypothesis to explain the stealth colonization of plant tissues by AMF is based on LysM domain-containing proteins. AMF could sequester chitin oligosaccharides to elude host's immunity, as already reported for pathogenic Ascomycetes (de Jonge and Thomma, 2009); all the sequenced AMF (except for *G. rosea*) possess one or more of such enzymes, which are gathered in the CBM50 CAZy class (Table S12). Lastly, the progressive thinning of the arbuscular cell wall may guarantee the passage of SSPs, which act as fungal effectors (Kloppholz *et al.*, 2011). In *G. margarita*, 41 SSPs had absolute expression levels around 10 times greater in the intraradical mycelium than at all other life stages (Fig. S13, Table S14). While no experimental evidence is available of the secretion of these proteins, our results point to the arbuscule as a preferential site for effector production.

Combining genomics, transcriptomics and morphological data reveals the finely tuned activity of genes related to the fungal cell wall, and in particular of chitin metabolism, with potential feedback on signalling and defence mechanisms in plant hosts.

Between plant and bacterial cells: HGT events

During its life cycle, *G. margarita* lives in intimate contact with both its plant host and its endobacteria (Fig. S2). Since this inter-kingdom interaction is expected to have been stable for more than 400 million years (Mondo *et al.*, 2012), we hypothesized that HGT events might have occurred. The predicted gene models were analysed using a pipeline for HGT discovery (Li *et al.*, 2018). Following this method, two candidate lists were generated, representing sequences putatively transferred to *G. margarita* from plants and bacteria respectively. Among genes possibly derived from plants (Table S15), we found two sequences related to fucosyltransferases. The BLASTX top hits for these sequences included *Chara braunii*, a green alga that is considered the ancestor of AMF-hosting green plants (Delaux *et al.*, 2015), as well as *Spizellomyces punctatus*, a basally branching chytrid fungus in phylum Chytridiomycota.

Of the genes putatively transferred from bacteria (Table S15), the ones possessing the highest alien index were non-ribosomal peptide synthetases-polyketide synthases (NRPS-PKS). Since these sequences also possess a high alien index when considering a potential plant origin, they might confirm an extensive HGT from bacteria to a wide variety of eukaryotes (Lawrence *et al.*, 2011). NRPS-PKSs are involved in the biosynthesis of secondary metabolites including antibiotics, toxins and siderophores: as an example, rhizoxin, a pathogenic molecule synthesized by *Burkholderia rhizoxinica*, the endosymbiont of *Rhizopus microsporus*, is a NRPS-PKSproduct (Lackner *et al.*, 2011). Similar sequences belong to a *Gigaspora* specific expanded gene family (Table S8a) but are absent in *R. irregularis* genome (Tisserant *et al.*, 2013). These findings, together with the limited potential of early diverging fungi to biosynthesize secondary metabolites (Voigt *et al.*, 2016), suggest that the Gigasporaceae lineage might have acquired such a peculiar ability via HGT from bacteria.

Two gene models (g16267 and g8176) share similarities with bacterial TIR domain-containing proteins. This domain is a protein-protein interaction domain widely distributed in animals, plants and bacteria but considered absent in fungi (Ve *et al.*, 2015). In plants and animals, TIRs play roles in innate immunity, while in bacteria some of them interfere with the innate immune pathways of the host (Ve *et al.*, 2015). The TIR-like *G. margarita* sequences found homologues in MREs, such as the endobacteria of *Dentiscutata heterogama* (Naito *et al.*, 2015; Torres-Cortés *et al.*, 2015) and *D. epigaea* (Sun *et al.*, 2019) but not in CaGg (Ghignone *et al.*, 2012). *Gigaspora margarita* BEG34 does not host MREs; furthermore, TIR-like *G. margarita* sequences were also expressed in the cured fungal line, ruling them out as

bacterial contaminants. In *G. margarita*, transcription of g8176 is strongly increased in germinating spores in the presence of CaGg, and is inactivated upon GR24 treatment (Table S16). By contrast, g16267 shows uniform expression across biological conditions. According to Sun *et al.* (2019), we suggest that such sequences originated from MREs, which were probably present in the common ancestors of Glomeromycotina and Mucoromycotina (Bonfante and Desirò, 2017). Finally, as Helitrons-mediated HGT has been reported in fungi (Castanera *et al.*, 2014), we searched for HGT candidates whose genomic locations overlap those of Helitrons: we found two similar sequences encoding S1 peptidases (g18889 and g8238) and an amidohydrolase, both having putative bacterial origin, along with an α -mannosidase (g17903) with high homology with both bacteria and plants.

In conclusion, HGT events from plants to the *G. margarita* genome appear to be limited and represent ancient events related to the algal origin of land plants (Delaux *et al.*, 2015) and to the interactions with basal Chytridiomycota fungi, which may feed on plants (Berbee *et al.*, 2017). HGT events from bacteria to AMF are multi-layered; on the one hand, they reflect the interactions occurring on the surface of AMF, which—in the soil—is regularly colonized by multiple bacterial communities (Agnolucci *et al.*, 2015) with different metabolic capacities. On the other, these HGT events reflect the ancient presence of endobacteria, which are hosted in the *G. margarita* cytoplasm, making genetic exchange between bacteria and their fungal hosts easier.

Gigaspora margarita genome and the fungal immune system

Since *G. margarita* interacts with viruses, bacteria and plants, we mined its genome to identify genetic determinants potentially involved in the fungal immune system. Nod-like-receptors (NLRs) are the main actors of the immune system in plants and animals but are also present in fungi, in a number ranging from 0 to 200 (Dyrka *et al.*, 2014). They are involved in allrecognition processes, mediating multiple processes from heterokaryon incompatibility to restriction of mycovirus transfer along the hyphae, as well as xenorecognition, working during pathogen attack or symbiosis establishment (Heller *et al.*, 2018). These proteins harbour a nucleotide-binding domain of the NACHT or AAA family, flanked by an N-terminal recognition domain and C-terminal repeats for protein-protein interactions. In addition to the two TIR-like sequences (HGT events), we found 18 proteins in *G. margarita* where both the central domain and the terminal repeats were present, most of them belonging to an expanded gene family; the central portion consisted of a NACHT domain in 17 cases, while an

NB-ARC domain was found in one case (g12905) (Table S16, Fig. S14). NB-ARCs are normally found in plant and animals NLR, where they are associated with resistance and cell death respectively (Van der Biezen and Jones, 1998). The C-terminal of the identified proteins was always composed of pentapeptide repeats. By contrast, their N-terminal recognition domain could not be functionally annotated, probably due to the lack of information in public databases, or to the modular structure of these proteins, where such modules could be present in separate proteins (Uehling *et al.*, 2017). The N-terminal domains associated with fungal NLRs can possess enzymatic activity such as lipase (patatins) and peptidase (subtilisin-related) activity: in *G. margarita*, the expression of four patatins seemed to be sensitive to the presence of the endobacterium, while three subtilisins were upregulated upon treatment with the plant signal analogue GR24 (Table S16), suggesting their involvement in plant or bacterial interaction. These patatins belong to expanded protein families in the *Gigaspora* genus (OG0000175), while subtilisins (OG0000164) are expanded exclusively in *G. margarita*, when compared to *G. rosea* (Table S8b).

Finally, *G. margarita* possesses several genes that may be involved in defence against viruses, including PIWI domain-containing proteins, a DICER-like protein and three argonaute-binding proteins, all related to RNA-mediated gene silencing (Fig. S14).

In conclusion, the genome of *G. margarita* reveals novel genetic determinants that might form the basis of the fungal immune system, which must activate non-self recognition events throughout the life cycle.

Conclusions

The genome sequencing of *G. margarita* BEG34 revealed key genomic traits shedding new light on the biology of the Glomeromycotina (Spatafora *et al.*, 2016). *Gigaspora margarita* possesses the largest fungal genome annotated so far. Its genome expansion is mostly driven by a TE explosion, also characterizing many other mycorrhizal fungi, from truffles with their low gene number (Murat *et al.*, 2018) to the related *G. rosea* (Morin *et al.*, 2019). Diversification time of Gigasporaceae has been placed at around 170 MYA (Davison *et al.*, 2015), being congruent with phylogenetic trees reconstructed using genome data, which suggest that divergence of Gigasporaceae and Glomeraceae from the last common ancestor was around 300 MYA (Sun *et al.*, 2019). These different speciation times suggest diverse evolution pathways among AMF and provide strong support for the many unique features reported in the *G. margarita* genome.

Similar to other AMF investigated, *G. margarita* is a strictly biotrophic fungus (Bonfante and Genre, 2010), meaning that it cannot uptake complex sugars from the environment and

cannot biosynthesize fatty acids, being therefore dependent on its host plants. *Gigaspora margarita* and *G. rosea* share an enriched set of CAZymes, which explain the capacity of *Gigaspora* to colonize different plants, maybe using a combination of non-enzymatic and enzymatic mechanisms. The high number of chitinase-encoding genes supports the dynamics of the fungal cell wall, which thins progressively moving from the extraradical phase to the thinner arbusculated branches. The discovery of highly regulated chitin deacetylases during the symbiotic phase suggests that deacetylated chitin might offer a good tool to escape from the plant immune system, avoiding the activation of plant defences. The absence of beta 1-3-glucans in the wall of the intraradical hyphae might reflect the same strategy, similar to that described for the pathogenic fungus *Colletotrichum* (Oliveira-Garcia and Deising, 2013). By contrast, *G. margarita* has had stable interactions with its endobacterium CaGg for a very long time (Mondo *et al.*, 2012). The intimate contact that *G. margarita* establishes with its host plants, its endobacterium and fungal-associated bacteria has provided the physical possibility of HGT events. Indeed, mining the genome revealed a unique transfer event from an algal ancestor towards *G. margarita*; many transfer events from a range of soil bacteria; and an interesting HGT event, involving TIRs, probably transferred from Mollicutes-related endobacteria detected in many AMF (Sun *et al.*, 2019) and some African strains of *G. margarita* (Desirò *et al.*, 2014). The presence of these TIRs together with a number of NLRs allowed us to identify the first core genes of the immune system present in AMF, which probably allows these fungi to safely interact with plants, bacteria and viruses.

Lastly, the high number of PTs first provides a rationale for the experimental evidence of Pi transporters expressed during the symbiotic mycorrhizal phase but also sheds a different light on the biology of AMF, usually considered as biofertilizers. *Gigaspora margarita* has a staggeringly high genome size: as for plant genomes (Pellicer *et al.*, 2018), we suggest that this peculiar trait has shaped its evolution. Plants with small genomes seem to be more widespread than those with large genomes, which persist only under conditions where selective pressures are more relaxed (Pellicer *et al.*, 2018). From a mechanistic point of view, nucleic acids are among the most nitrogen- and phosphorus-rich molecules of the cell, so under limiting nutrient N and P conditions, we predict that species with large genomes, which are more demanding and costly to build and maintain than species with small genomes, would be less competitive. This reasoning fits well with *G. margarita* BEG34, which is commonly isolated from sand dunes (Stürmer *et al.*, 2018) but poorly represented in soils originating from many other diverse environments (Davison *et al.*, 2015; Davison *et al.*, 2018). On the other hand, its spores are among the largest produced by AMF, a feature that is commonly associated with long-term survival

strategies (Carlos *et al.*, 2018). In conclusion, the genome description of *G. margarita* revealed novel and unique features, showing its weaknesses (a huge genome may be a limiting factor in terms of ecological success), but also its biological strengths through the capacity to communicate with plants and endobacteria, acting therefore as a hub of inter-kingdom interactions.

Materials and methods

AMF culture, flow cytometry and gDNA extraction

Gigaspora margarita strain BEG34 with (B+) and without (B-) the *CaGg* endobacterium was used for this study. The spores were obtained as described in Salvioli *et al.*, 2016 (see Supplementary Text).

The genome size of *G. margarita* was first determined by flow cytometry using *Solanum lycopersicum* as an internal reference. For each set of isolated nuclei, 100 sterilized B+ spores and 0.5 cm² of young *S. lycopersicum* leaf tissues were processed as described in Sędziewska *et al.* (2011) and measured using a FACSAria (BD Biosciences, San José, CA, USA) with a 488 nm laser. The absolute DNA amounts of the samples were calculated based on the values of the G1 peak means. At least 10 000 nuclei per sample were analysed.

The gDNA was extracted from batches of 100 spores using the CTAB method. Sample purity was assessed at Nanodrop1000 (Thermo Scientific, Wilmington, NC, USA). Only samples with a 260/230–260/280 > 1.8 were kept and quantified with Qubit (Broad Range Kit, Thermo-Fisher). In total, 15 µg of DNA, extracted from 16 batches of spores (~1600 spores), was sent to the sequencing facility.

DNA sequencing and genome assembly

The genome of *G. margarita* was sequenced using Illumina platform. One fragment Paired-End library (PE; Illumina TruSeq Nano) and two long Mate-Pair libraries (MP; Illumina Nextera; inserts sizes 3 and 8 kbp) were constructed in 2 × 150 bp format using one lane of Illumina HiSeq3000 by GeT-PlaGe GenoToul (Castanet-Tolosan, France). All libraries have been filtered for adapter sequences and low quality reads, using combinations of trim_galore v. 0.4.1 (http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/), trimmomatic v. 0.36 (Bolger *et al.*, 2014) and 1, cutadapt v. 1.2.1, (Martin, 2011). Genome assembly was performed by Bison SeqTech ApS (Frederiksberg, Denmark) with ALLPATH-LG v.3 (Gnerre *et al.*, 2011). More details on assembly, quality check and refinements are provided in Supplementary Text.

RNA-seq experiments

The single-end libraries used to reconstruct the *G. margarita* transcriptome (Salvioli *et al.*, 2016) were used in this study along with new libraries obtained from *G. margarita* B+ symbiotic extraradical mycelium (JGI proposal ID 1450), and from *Lotus japonicus* B+ and the B–mycorrhizal roots (Table S5). RNA for mycorrhizal roots was extracted with NucleoSpin RNA Plant and Fungi Kit (Macherey-Nagel) and checked at BioAnalyzer 2100 (Agilent). Samples with RIN < 7.0 were discarded. Sequencing was performed by Macrogen (South Korea).

All the libraries used in this study were pre-processed with BBmap (Bushnell, 2014) for contaminants removal and quality trimming. Each library was mapped to the set of transcripts with salmon v. 0.10.2 (Patro *et al.*, 2017). Differential gene expression was calculated as detailed in Supplementary Text.

Gene prediction and annotation

Protein-coding genes were predicted using the BRAKER2 pipeline (Stanke *et al.*, 2006; Stanke *et al.*, 2008; Lomsadze *et al.*, 2014; Hoff *et al.*, 2016). The training process was performed with the aid of the GOU13/14/15, BUNYW/X/Y and BUOAC/G/H RNA-seq libraries (Supporting Information 5), the UNIPROT Mucoromycota proteins, and the coordinates of repeated sequences. The annotation process is detailed in Supplementary Text.

The predicted gene models were functionally annotated with a blastx search against the nr protein database. Blast2GO v.4.1 (Götz *et al.*, 2008) was used for the attribution of the best blast hit, Gene Ontology Terms and E.C. numbers.

Identification and analysis of repetitive elements

The REPET v. 2.5 pipeline (Flutre *et al.*, 2011) was used for the detection, classification (TEdenovo) and annotation (TEannot) of TEs and other repetitive DNA sequences. We used the same analytic strategy already used elsewhere (Plomion *et al.*, 2018). The library of classified consensus sequences provided by the TEdenovo pipeline was used to annotate the TE copies with the TEannot pipeline. After manual curation, we obtained a final set of TEs consensus, with at least one full-length fragment, representing the 63.8% of the genome. OcculterCut v. 1.1 (Testa *et al.*, 2016) was used to detect the eventually occurring GC-bias in the *G. margarita* genome. Ripcal v.2.0 (Hane and Oliver, 2008) was used to detect the occurrence of RIP in *G. margarita*. A more extensive explanation on TEs discovery is reported in Supplementary Text.

Additional analyses

All the other downstream analyses performed on the *G. margarita* genome (including gene family evolution, detection of putatively horizontally transferred genes, classification chitin synthases, identification of secreted proteins and CAZymes, etc...) are reported in Supplementary Text.

Acknowledgements

The authors express their thanks to Dr. J. Uheling for the useful suggestions on fungal immunity. Transcriptome sequencing for the JGI Community Science Project 1450 'Comparative genomics of early-diverging terrestrial fungi and their bacterial endosymbionts' (Principal Investigator Dr. Greg Bonito) were conducted by the DOE's Joint Genome Institute, a DOE Office of Science User Facility, and supported by the Office of Science of the US Department of Energy under Contract DE-AC02-05CH11231. FM is funded by the Laboratory of Excellence ARBRE (ANR-11-LABX-0002-01), Region Lorraine, European Regional Development Fund and Beijing Advanced Innovation Center for Tree Breeding by Molecular Design, Beijing Forest University. Research was funded by local grants (60%) to Dr. Paola Bonfante. The authors are grateful to Dr. Jörg Fuchs (IPK, Gatersleben) for providing FC facilities and help with analysis, and to Dr. J. Mach for the critical reading of the manuscript.

References

Agnolucci, M., Battini, F., Cristani, C., and Giovannetti, M. (2015) Diverse bacterial communities are recruited on spores of different arbuscular mycorrhizal fungal isolates. *Biol Fertil Soils* **51**: 379–389.

Ahrendt, S.R., Quandt, C.A., Ciobanu, D., Clum, A., Salamov, A., Andreopoulos, B., *et al.* (2018) Leveraging single-cell genomics to expand the fungal tree of life. *Nat Microbiol* **3**: 1417–1428.

Amselem, J., Lebrun, M.H., and Quesneville, H. (2015) Whole genome comparative analysis of transposable elements provides new insight into mechanisms of their inactivation in fungal genomes. *BMC Genomics* **16**: 141.

Balestrini, R., Gómez-Ariza, J., Klink, V.P., and Bonfante, P. (2009) Application of laser microdissection to plant pathogenic and symbiotic interactions. *J Plant Interact* **4**: 81–92.

Balzerque, C., Chabaud, M., Barker, D.G., Bécard, G., and Rochage, S.F. (2013) High phosphate reduces host ability to develop arbuscular mycorrhizal symbiosis without affecting root calcium spiking responses to the fungus. *Front Plant Sci* **4**: 426.

Berbee, M.L., James, T.Y., and Strullu-Derrien, C. (2017) Early diverging fungi: diversity and impact at the dawn of terrestrial life. *Annu Rev Microbiol* **71**: 41–60.

Bianciotto, V., Bandi, C., Minerdi, D., Sironi, M., Tichy, H.V., and Bonfante, P. (1996) An obligately endosymbiotic

mycorrhizal fungus itself harbors obligately intracellular bacteria. *Appl Environ Microbiol* **62**: 3005–3010.

Bianciotto, V., and Bonfante, P. (1992) Quantification of the nuclear DNA content of two arbuscular mycorrhizal fungi. *Mycol Res* **96**: 1071–1076.

Bolger, A.M., Lohse, M., and Usadel, B. (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**: 2114–2120.

Bonfante, P. (2018) The future has roots in the past: the ideas and scientists that shaped mycorrhizal research. *New Phytol* **220**: 982–995.

Bonfante, P., and Desirò, A. (2017) Who lives in a fungus? The diversity, origins and functions of fungal endobacteria living in Mucoromycota. *ISME J* **11**: 1727–1735.

Bonfante, P., and Genre, A. (2010) Mechanisms underlying beneficial plant–fungus interactions in mycorrhizal symbiosis. *Nat Commun* **1**: 48.

Bonfante, P., Venice, F., and Lanfranco, L. (2019) The Mycobiota: fungi take their place between plants and bacteria. *Curr Opin Microbiol*. <https://doi.org/10.1016/j.mib.2019.08.004>

Bushnell, B. (2014) *BBMap: A Fast, Accurate, Splice-Aware Aligner*. Berkeley, CA (United States): Lawrence Berkeley National Lab. (LBNL) Report No.: LBNL-7065E. Available from: <https://jgi.doe.gov/data-and-tools/bbtools/bb-tools-user-guide/installation-guide/>

Carlos, A., Aguilar-Trigueros, C.A., Hempel, S., Powell, J.R., Cornwell, W.K., and Rillig, M.C. (2018) Bridging reproductive and microbial ecology: a case study in arbuscular mycorrhizal fungi. *ISME J* **13**: 873–884.

Carotenuto, G., Chabaud, M., Miyata, K., Capozzi, M., Takeda, N., Kaku, H., *et al.* (2017) The rice LysM receptor-like kinase OsCERK1 is required for the perception of short-chain chitin oligomers in arbuscular mycorrhizal signaling. *New Phytol* **214**: 1440–1446.

Castanera, R., López-Varas, L., Borgognone, A., LaButti, K., Lapidus, A., Schmutz, J., *et al.* (2016) Transposable elements *versus* the fungal genome: impact on whole-genome architecture and transcriptional profiles. *PLoS Genet* **12**: e1006108.

Castanera, R., Perez, G., Lopez, L., Sancho, R., Santoyo, F., Alfaro, M., *et al.* (2014) Highly expressed captured genes and cross-kingdom domains present in Helitrons create novel diversity in *Pleurotus ostreatus* and other fungi. *BMC Genomics* **15**: 1071.

Chang, Y., Desirò, A., Na, H., Sandor, L., Lipzen, A., Clum, A., *et al.* (2019) Phylogenomics of Endogonaceae and evolution of mycorrhizas within Mucoromycota. *New Phytol* **222**: 511–525.

Chen, E.C.H., Morin, E., Beaudet, D., Noel, J., Yildirim, G., Ndikumana, S., *et al.* (2018) High intraspecific genome diversity in the model arbuscular mycorrhizal symbiont *Rhizophagus irregularis*. *New Phytol* **220**: 1161–1171.

Chialva, M., Salvioli di Fossalunga, A., Daghino, S., Ghignone, S., Bagnaresi, P., Chiapello, M., *et al.* (2018) Native soils with their microbiotas elicit a state of alert in tomato plants. *New Phytol* **220**: 1296–1308.

Davison, J., Moora, M., Öpik, M., Adholeya, A., Ainsaar, L., Bâ, A., *et al.* (2015) Global assessment of arbuscular mycorrhizal fungus diversity reveals very low endemism. *Science* **349**: 970–973.

- Davison, J., Moora, M., Öpik, M., Ainsaar, L., Ducouso, M., Hiiesalu, I., et al. (2018) Microbial Island biogeography: isolation shapes the life history characteristics but not diversity of root-symbiotic fungal communities. *ISME J* **12**: 2211–2224.
- De Bie, T., Cristianini, N., Demuth, J.P., and Hahn, M.W. (2006) CAFE: a computational tool for the study of gene family evolution. *Bioinformatics* **22**: 1269–1271.
- de Jonge, R., and Thomma, B.P.H.J. (2009) Fungal LysM effectors: extinguishers of host immunity? *Trends Microbiol* **17**: 151–157.
- Dearth, S.P., Castro, H.F., Venice, F., Tague, E.D., Novero, M., Bonfante, P., and Campagna, S.R. (2018) Metabolome changes are induced in the arbuscular mycorrhizal fungus *Gigaspora margarita* by germination and by its bacterial endosymbiont. *Mycorrhiza* **28**: 421–433.
- Delaux, P.M., Radhakrishnan, G.V., Jayaraman, D., Cheema, J., Malbreil, M., Volkening, J.D., et al. (2015) Algal ancestor of land plants was preadapted for symbiosis. *Proc Natl Acad Sci U S A* **112**: 13390–13395.
- Desirò, A., Salvioli, A., Ngonkeu, E.L., Mondo, S.J., Epis, S., Faccio, A., et al. (2014) Detection of a novel intracellular microbiome hosted in arbuscular mycorrhizal fungi. *ISME J* **8**: 257–270.
- Dyrka, W., Lamacchia, M., Durrens, P., Kobe, B., Daskalov, A., Paoletti, M., et al. (2014) Diversity and variability of NOD-like receptors in fungi. *Genome Biol Evol* **6**: 3137–3158.
- Emms, D.M., and Kelly, S. (2015) OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol* **16**: 157.
- Enright, A.J., Van Dongen, S., and Ouzounis, C.A. (2002) An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Res* **30**: 1575–1584.
- Ezawa, T., and Saito, K. (2018) How do arbuscular mycorrhizal fungi handle phosphate? New insight into fine-tuning of phosphate metabolism. *New Phytol* **220**: 1116–1121.
- Fiorilli, V., Lanfranco, L., and Bonfante, P. (2013) The expression of *GintPT*, the phosphate transporter of *Rhizophagus irregularis*, depends on the symbiotic status and phosphate availability. *Planta* **237**: 1267–1277.
- Flutre, T., Duprat, E., Feuillet, C., and Quesneville, H. (2011) Considering transposable element diversification in de novo annotation approaches. *PLoS One* **6**: e16526.
- Genre, A., Chabaud, M., Balzergue, C., Puech-Pagès, V., Novero, M., Rey, T., et al. (2013) Short-chain chitin oligomers from arbuscular mycorrhizal fungi trigger nuclear Ca₂⁺ spiking in *Medicago truncatula* roots and their production is enhanced by strigolactone. *New Phytol* **198**: 190–202.
- Ghignone, S., Salvioli, A., Anca, I., Lumini, E., Ortu, G., Petiti, L., et al. (2012) The genome of the obligate endobacterium of an AM fungus reveals an interphylum network of nutritional interactions. *ISME J* **6**: 136–145.
- Giovannetti, M., Mari, A., Novero, M., and Bonfante, P. (2015) Early *Lotus japonicus* root transcriptomic responses to symbiotic and pathogenic fungal exudates. *Front Plant Sci* **6**: 480.
- Gladyshev, E., and Kleckner, N. (2017) Recombination-independent recognition of DNA homology for repeat-induced point mutation. *Curr Genet* **63**: 389–400.
- Gnerre, S., MacCallum, I., Przybylski, D., Ribeiro, F.J., Burton, J.N., Walker, B.J., et al. (2011) High-quality draft assemblies of mammalian genomes from massively parallel sequence data. *Proc Natl Acad Sci U S A* **108**: 1513–1518.
- Götz, S., García-Gómez, J.M., Terol, J., Williams, T.D., Nagaraj, S.H., Nueda, M.J., et al. (2008) High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Res* **36**: 3420–3435.
- Gow, N.A.R., Latge, J.P., and Munro, C.A. (2017) The fungal cell wall: structure, biosynthesis, and function. *Microbiol Spectr* **5**. <https://doi.org/10.1128/microbiolspec.FUNK-0035-2016>
- Guillén, F., Gómez-Toribio, V., Martínez, M.J., and Martínez, A.T. (2000) Production of hydroxyl radical by the synergistic action of fungal laccase and aryl alcohol oxidase. *Arch Biochem Biophys* **383**: 142–147.
- Hane, J.K., and Oliver, R.P. (2008) RIPCAL: a tool for alignment-based analysis of repeat-induced point mutations in fungal genomic sequences. *BMC Bioinformatics* **9**: 478.
- Harrison, M.J., and van Buuren, M.L. (1995) A phosphate transporter from the mycorrhizal fungus *Glomus versiforme*. *Nature* **378**: 626–629.
- Heller, J., Clavé, C., Gladieux, P., Saupe, S.J., and Glass, N.L. (2018) NLR surveillance of essential SEC-9 SNARE proteins induces programmed cell death upon allorecognition in filamentous fungi. *Proc Natl Acad Sci U S A* **115**: 2292–2301.
- Hernández-Ortega, A., Ferreira, P., and Martínez, A.T. (2012) Fungal aryl-alcohol oxidase: a peroxide-producing flavoenzyme involved in lignin degradation. *Appl Microbiol Biotechnol* **93**: 1395–1410.
- Hoff, K.J., Lange, S., Lomsadze, A., Borodovsky, M., and Stanke, M. (2016) BRAKER1: unsupervised RNA-Seq-based genome annotation with GeneMark-ET and AUGUSTUS. *Bioinformatics* **32**: 767–769.
- Kloppholz, S., Kuhn, H., and Requena, N. (2011) A secreted fungal effector of *Glomus intraradices* promotes symbiotic biotrophy. *Curr Biol* **21**: 1204–1209.
- Kobayashi, Y., Maeda, T., Yamaguchi, K., Kameoka, H., Tanaka, S., Ezawa, T., et al. (2018) The genome of *Rhizophagus clarus* HR1 reveals a common genetic basis for auxotrophy among arbuscular mycorrhizal fungi. *BMC Genomics* **19**: 465.
- Krüger, M., Krüger, C., Walker, C., Stockinger, H., and Schübler, A. (2012) Phylogenetic reference data for systematics and phylotaxonomy of arbuscular mycorrhizal fungi from phylum to species level. *New Phytol* **193**: 970–984.
- Kües, U., and Rühl, M. (2011) Multiple multi-copper oxidase gene families in basidiomycetes – what for? *Curr Genomics* **12**: 72–94.
- Lackner, G., Moebius, N., Partida-Martinez, L.P., Boland, S., and Hertweck, C. (2011) Evolution of an endofungal lifestyle: deductions from the *Burkholderia rhizoxinica* genome. *BMC Genomics* **12**: 210.
- Laetsch, D.R., and Blaxter, M.L. (2017) BlobTools: interrogation of genome assemblies. *F1000Research* **6**: 1287.
- Lanfranco, L., Fiorilli, V., and Gutjahr, C. (2018) Partner communication and role of nutrients in the arbuscular mycorrhizal symbiosis. *New Phytol* **220**: 1031–1046.

- Lawrence, D.P., Kroken, S., Pryor, B.M., and Arnold, A.E. (2011) Interkingdom gene transfer of a hybrid NPS/PKS from bacteria to filamentous ascomycota. *PLoS One* **6**: e28231.
- Lemoine, M.C., Gollotte, A., and Gianinazzi-Pearson, V. (1995) Localization of $\beta(1-3)$ glucan in walls of the endomycorrhizal fungi *Glomus mosseae* (Nicol. and Gerd.) Gerd. and Trappe and *Acaulospora laevis* Gerd. and Trappe during colonization of host roots. *New Phytol* **129**: 97–105.
- Lev, S., and Djordjevic, J.T. (2018) Why is a functional PHO pathway required by fungal pathogens to disseminate within a phosphate-rich host: a paradox explained by alkaline pH-simulated nutrient deprivation and expanded PHO pathway function. *PLoS Pathog* **14**: e1007021.
- Li, M., Zhao, J., Tang, N., Sun, H., and Huang, J. (2018) Horizontal gene transfer from bacteria and plants to the arbuscular mycorrhizal fungus *Rhizophagus irregularis*. *Front Plant Sci* **9**: 701.
- Ligrone, R., Carafa, A., Lumini, E., Bianciotto, V., Bonfante, P., and Duckett, J.G. (2007) Glomeromycotean associations in liverworts: a molecular, cellular, and taxonomic analysis. *Am J Bot* **94**: 1756–1777.
- Lin, K., Limpens, E., Zhang, Z., Ivanov, S., Saunders, D.G.O., Mu, D., et al. (2014) Single nucleus genome sequencing reveals high similarity among nuclei of an endomycorrhizal fungus. *PLoS Genet* **10**: e1004078.
- Liu, Z., Gay, L.M., Tuveng, T.R., Agger, J.W., Westereng, B., Mathiesen, G., et al. (2017) Structure and function of a broad-specificity chitin deacetylase from *Aspergillus nidulans* FGSC A4. *Sci Rep* **7**: 1746.
- Lomsadze, A., Burns, P.D., and Borodovsky, M. (2014) Integration of mapped RNA-Seq reads into automatic training of eukaryotic gene finding algorithm. *Nucleic Acids Res* **42**: e119.
- MacLean, A.M., Bravo, A., and Harrison, M.J. (2017) Plant Signaling and metabolic pathways enabling arbuscular mycorrhizal Symbiosis. *Plant Cell* **29**: 2319–2335.
- Martin, F., Kohler, A., Murat, C., Balestrini, R., Coutinho, P. M., Jaillon, O., et al. (2010) Périgord black truffle genome uncovers evolutionary origins and mechanisms of symbiosis. *Nature* **464**: 1033–1038.
- Martin, M. (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnetJ* **17**: 10–12.
- Martinez-Medina, A., Flors, V., Heil, M., Mauch-Mani, B., Pieterse, C.M.J., Pozo, M.J., et al. (2016) Recognizing plant defense priming. *Trends Plant Sci* **21**: 818–822.
- McTaggart, A.R., Shuey, L.S., Granados, G.M., du Plessis, E., Fraser, S., Barnes, I., et al. (2018) Evidence that *Austropuccinia psidii* may complete its sexual life cycle on Myrtaceae. *Plant Pathol* **67**: 729–734.
- Miyata, K., Kozaki, T., Kouzai, Y., Ozawa, K., Ishii, K., Asamizu, E., et al. (2014) The bifunctional plant receptor, OsCERK1, regulates both chitin-triggered immunity and arbuscular mycorrhizal symbiosis in rice. *Plant Cell Physiol* **55**: 1864–1872.
- Mondo, S.J., Lastovetsky, O.A., Gaspar, M.L., Schwardt, N.H., Barber, C.C., Riley, R., et al. (2017) Bacterial endosymbionts influence host sexuality and reveal reproductive genes of early divergent fungi. *Nat Commun* **8**: 1843.
- Mondo, S.J., Toomer, K.H., Morton, J.B., Lekberg, Y., and Pawlowska, T.E. (2012) Evolutionary stability in a 400-million-year-old heritable facultative mutualism. *Evolution* **66**: 2564–2576.
- Morin, E., Miyauchi, S., San Clemente, H., Chen, E.C.H., Pelin, A., de la Providencia, I., et al. (2019) Comparative genomics of *Rhizophagus irregularis*, *R. cerebriforme*, *R. diaphanus* and *Gigaspora rosea* highlights specific genetic features in Glomeromycotina. *New Phytol* **222**: 1584–1598.
- Murat, C., Kuo, A., Barry, K.W., Clum, A., Dockter, R.B., Fauchery, L., et al. (2018) Draft genome sequence of *Tuber borchii* Vittad., a whitish edible truffle. *Genome Announc* **6**: 1956–1965.
- Muszevska, A., Steczkiewicz, K., Stepniewska-Dziubinska, M., and Ginalski, K. (2017) Cut-and-paste transposons in fungi with diverse lifestyles. *Genome Biol Evol* **9**: 3463–3477.
- Naito, M., Morton, J.B., and Pawlowska, T.E. (2015) Minimal genomes of mycoplasma-related endobacteria are plastic and contain host-derived genes for sustained life within Glomeromycota. *Proc Natl Acad Sci U S A* **112**: 7791–7796.
- Naumann, M., Schüssler, A., and Bonfante, P. (2010) The obligate endobacteria of arbuscular mycorrhizal fungi are ancient heritable components related to the Mollicutes. *ISME J* **4**: 862–871.
- Oliveira-Garcia, E., and Deising, H.B. (2013) Infection structure-specific expression of β -1,3-glucan synthase is essential for pathogenicity of *Colletotrichum graminicola* and evasion of β -glucan-triggered immunity in maize. *Plant Cell* **25**: 2356–2378.
- Patro, R., Duggal, G., Love, M.I., Irizarry, R.A., and Kingsford, C. (2017) Salmon provides fast and bias-aware quantification of transcript expression. *Nat Methods* **14**: 417–419.
- Pawlowska, T.E., Gaspar, M.L., Lastovetsky, O.A., Mondo, S.J., Real-Ramirez, I., Shakya, E., and Bonfante, P. (2018) Biology of fungi and their bacterial endosymbionts. *Annu Rev Phytopathol* **56**: 289–309.
- Pelin, A., Pombert, J.F., Salvioli, A., Bonen, L., Bonfante, P., and Corradi, N. (2012) The mitochondrial genome of the arbuscular mycorrhizal fungus *Gigaspora margarita* reveals two unsuspected trans-splicing events of group I introns. *New Phytol* **194**: 836–845.
- Pellicer, J., Hidalgo, O., Dodsworth, S., and Leitch, I.J. (2018) Genome size diversity and its impact on the evolution of land plants. *Genes (Basel)* **9**: 88–102.
- Pimprikar, P., and Gutjahr, C. (2018) Transcriptional regulation of arbuscular mycorrhiza development. *Plant Cell Physiol* **59**: 673–690.
- Plomion, C., Aury, J.M., Amselem, J., Leroy, T., Murat, F., Duplessis, S., et al. (2018) Oak genome reveals facets of long lifespan. *Nat Plants* **4**: 440–452.
- Pozo, M.J., and Azcón-Aguilar, C. (2007) Unraveling mycorrhiza-induced resistance. *Curr Opin Plant Biol* **10**: 393–398.
- Pusztahelyi, T. (2018) Chitin and chitin-related compounds in plant-fungal interactions. *Mycology* **9**: 189–201.

- Ropars, J., Toro, K.S., Noel, J., Pelin, A., Charron, P., Farinelli, L., et al. (2016) Evidence for the sexual origin of heterokaryosis in arbuscular mycorrhizal fungi. *Nat Microbiol* **1**: 16033.
- Ruiz-Lozano, J.M., and Bonfante, P. (1999) Identification of a putative P-transporter operon in the genome of a Burkholderia strain living inside the arbuscular mycorrhizal fungus *Gigaspora margarita*. *J Bacteriol* **181**: 4106–4109.
- Salvioli, A., Ghignone, S., Novero, M., Navazio, L., Venice, F., Bagnaresi, P., and Bonfante, P. (2016) Symbiosis with an endobacterium increases the fitness of a mycorrhizal fungus, raising its bioenergetic potential. *ISME J* **10**: 130–144.
- Sánchez-Vallet, A., Fouché, S., Fudal, I., Hartmann, F.E., Soyer, J.L., Tellier, A., and Croll, D. (2018) The genome biology of effector gene evolution in filamentous plant pathogens. *Annu Rev Phytopathol* **56**: 21–40.
- Sánchez-Vallet, A., Mesters, J.R., and Thomma, B.P.H.J. (2015) The battle for chitin recognition in plant-microbe interactions. *FEMS Microbiol Rev* **39**: 171–183.
- Schmitz, A.M., and Harrison, M.J. (2014) Signaling events during initiation of arbuscular mycorrhizal symbiosis. *J Integr Plant Biol* **56**: 250–261.
- Secco, D., Wang, C., Shou, H., and Whelan, J. (2012) Phosphate homeostasis in the yeast *Saccharomyces cerevisiae*, the key role of the SPX domain-containing proteins. *FEBS Lett* **586**: 289–295.
- Sędziewska, K.A., Fuchs, J., Tensch, E.M., Baronian, K., Watzke, R., and Kunze, G. (2011) Estimation of the *Glomus intraradices* nuclear DNA content. *New Phytol* **192**: 794–797.
- Smith, S., and Read, D. (2008) *Mycorrhizal Symbiosis*, 3rd ed. London: Academic Press.
- Spanu, P.D., Abbott, J.C., Amselem, J., Burgis, T.A., Soanes, D.M., Stüber, K., et al. (2010) Genome expansion and gene loss in powdery mildew fungi reveal tradeoffs in extreme parasitism. *Science* **330**: 1543–1546.
- Spatafora, J.W., Chang, Y., Benny, G.L., Lazarus, K., Smith, M.E., Berbee, M.L., et al. (2016) A phylum-level phylogenetic classification of zygomycete fungi based on genome-scale data. *Mycologia* **108**: 1028–1046.
- Stanke, M., Diekhans, M., Baertsch, R., and Haussler, D. (2008) Using native and syntenically mapped cDNA alignments to improve de novo gene finding. *Bioinformatics* **24**: 637–644.
- Stanke, M., Schöffmann, O., Morgenstern, B., and Waack, S. (2006) Gene prediction in eukaryotes with a generalized hidden Markov model that uses hints from external sources. *BMC Bioinformatics* **7**: 62.
- Strullu-Derrien, C., Selosse, M., Kenrick, P., and Martin, F. M. (2018) The origin and evolution of mycorrhizal symbioses: from palaeomycology to phylogenomics. *New Phytol* **220**: 1012–1030.
- Stürmer, S.L., Bever, J.D., and Morton, J.B. (2018) Biogeography of arbuscular mycorrhizal fungi (Glomeromycota): a phylogenetic perspective on species distribution patterns. *Mycorrhiza* **28**: 587–603.
- Sun, X., Chen, W., Ivanov, S., MacLean, A.M., Wight, H., Ramaraj, T., et al. (2019) Genome and evolution of the arbuscular mycorrhizal fungus *Diversispora epigaea* (formerly *Glomus versiforme*) and its bacterial endosymbionts. *New Phytol* **221**: 1556–1573.
- Testa, A.C., Oliver, R.P., and Hane, J.K. (2016) OcculterCut: a comprehensive survey of AT-rich regions in fungal genomes. *Genome Biol Evol* **8**: 2044–2064.
- Tisserant, E., Malbreil, M., Kuo, A., Kohler, A., Symeonidi, A., Balestrini, R., et al. (2013) Genome of an arbuscular mycorrhizal fungus provides insight into the oldest plant symbiosis. *Proc Natl Acad Sci U S A* **110**: 20117–20122.
- Torres-Cortés, G., Ghignone, S., Bonfante, P., and Schübler, A. (2015) Mosaic genome of endobacteria in arbuscular mycorrhizal fungi: transkingdom gene transfer in an ancient mycoplasma-fungus association. *Proc Natl Acad Sci U S A* **112**: 7785–7790.
- Turina, M., Ghignone, S., Astolfi, N., Silvestri, A., Bonfante, P., and Lanfranco, L. (2018) The virome of the arbuscular mycorrhizal fungus *Gigaspora margarita* reveals the first report of DNA fragments corresponding to replicating non-retroviral RNA viruses in fungi. *Environ Microbiol* **20**: 2012–2025.
- Uehling, J., Gryganskyi, A., Hameed, K., Tschaplinski, T., Misztal, P.K., Wu, S., et al. (2017) Comparative genomics of *Mortierella elongata* and its bacterial endosymbiont *Mycosporium cysteinexigens*. *Environ Microbiol* **19**: 2964–2983.
- Van der Biezen, E.A., and Jones, J.D. (1998) Plant disease-resistance proteins and the gene-for-gene concept. *Trends Biochem Sci* **23**: 454–456.
- Vannini, C., Carpentieri, A., Salvioli, A., Novero, M., Marsoni, M., Testa, L., et al. (2016) An interdomain network: the endobacterium of a mycorrhizal fungus promotes antioxidative responses in both fungal and plant hosts. *New Phytol* **211**: 265–275.
- Ve, T., Williams, S.J., and Kobe, B. (2015) Structure and function of toll/interleukin-1 receptor/resistance protein (TIR) domains. *Apoptosis* **20**: 250–261.
- Voigt, K., Wolf, T., Ochsenreiter, K., Nagy, G., Kaerger, K., Shelest, E., et al. (2016) 15 genetic and metabolic aspects of primary and secondary metabolism of the Zygomycetes. In Hoffmeister, D. (ed). *Biochemistry and Molecular Biology*. Switzerland: Springer International Publishing, pp. 361–385.
- Waterhouse, R.M., Seppey, M., Simão, F.A., Manni, M., Ioannidis, P., Klioutchnikov, G., et al. (2018) BUSCO applications from quality assessments to gene prediction and phylogenomics. *Mol Biol Evol* **35**: 543–548.
- Xie, X., Lin, H., Peng, X., Xu, C., Sun, Z., Jiang, K., et al. (2016) Arbuscular mycorrhizal symbiosis requires a phosphate transceptor in the *Gigaspora margarita* fungal symbiont. *Mol Plant* **9**: 1583–1608.
- Zhang, X., Dong, W., Sun, J., Feng, F., Deng, Y., He, Z., et al. (2015) The receptor kinase CERK1 has dual functions in symbiosis and immunity signalling. *Plant J* **81**: 258–267.
- Zipfel, C., and Oldroyd, G.E.D. (2017) Plant signalling in symbiosis and immunity. *Nature* **543**: 328–336.

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

Supplementary Table S1: List of the largest fungal genome sequences deposited in the NCBI and JGI databases (updated November 2018).

Supplementary Table S2: differentially expressed genes identified as putative active transposases. The differential gene expression was calculated with DESeq2 and log2 normalized. The padj threshold was set to 0.05, and no filtering was applied to the log2 fold changes. B+/B- germinating spores: spores with/without the CaGg endobacterium/ grown for 10 days in H2O; B+/B- SL treatment: spores with/without the CaGg endobacterium/ pre-germinated for 3 days in H2O/ and then grown for 7 days in presence of the synthetic strigolactone GR24; B+ extraradical mycelium: symbiotic mycelium of the line containing CaGg/ sampled from the root surface of mycorrhized Lotus japonicus roots; B+/B- symbiotic mycelium: intraradical and extraradical mycelium sampled from *L. japonicus* roots colonized by the line with/without CaGg.

Supplementary Table S3: *G. margarita* secreted proteins prediction

Supplementary Table S4: Results of permutation tests for the association of Transposon families with candidate Small Secreted Proteins.

Supplementary Table S5: RNAseq libraries generated from different life stages used in this study. B+ identifies the line of *G. margarita* containing CaGg; the cured line, lacking the endobacterium, is defined as B-.

Supplementary Table S6: expression of *G. margarita* genes, and classification of genes expressed only in specific life stages or depending on the presence of CaGg (i.e. with a normalized expression value >0.1).

Supplementary Table S7: Organisms used for the analysis of gene family evolution. Saprotrophic fungi from the Mucoromycotina and Mortierellomycotina were included in the analysis due to their phylogenetic relationship with *G. margarita*, as well as two ectomycorrhizal fungi from distant clades and a pathogenic fungus.

Supplementary Table S8: biologically relevant protein families that underwent rapid evolution in a) the Gigaspora clade and in b) *G. margarita* when compared to *G. rosea*. Family-wide functional annotations are reported, as well as the number of proteins from a specific organism contained in each family. Family-wide functional annotations were obtained for each family from the annotation of its protein members.

Supplementary Table S9: differential expression of genes belonging to *Gigaspora* spp. and *G. margarita* rapidly evolved gene families. The differential gene expression was calculated with DESeq2 and log2 normalized. The padj threshold was set to 0.05, and no filtering was applied to the log2 fold changes. B+/B- germinating spores: spores with/without the CaGg endobacterium/ grown for 10 days in H2O; B+/B- SL treatment: spores

with/without the CaGg endobacterium/ pre-germinated for 3 days in H2O/ and then grown for 7 days in presence of the synthetic strigolactone GR24; B+ extraradical mycelium: symbiotic mycelium of the line containing CaGg/ sampled from the root surface of mycorrhized Lotus japonicus roots; B+/B- symbiotic mycelium: intraradical and extraradical mycelium sampled from *L. japonicus* roots colonized by the line with/without CaGg. A gene coding for an immunoreactive mannoprotein (g11471), belonging to OG0000217, is the *G. margarita* gene that mostly respond to the presence of CaGg throughout all the gene expression dataset presented in the current work.

Supplementary Table S10: *G. margarita* genes involved in phosphate sensing, transport, and metabolism.

Supplementary Table S11: differential expression of phosphate-related genes. The differential gene expression was calculated with DESeq2 and log2 normalized. The padj threshold was set to 0.05, and no filtering was applied to the log2 fold changes. B+/B- germinating spores: spores with/without the CaGg endobacterium/ grown for 10 days in H2O; B+/B- SL treatment: spores with/without the CaGg endobacterium/ pre-germinated for 3 days in H2O/ and then grown for 7 days in presence of the synthetic strigolactone GR24; B+ extraradical mycelium: symbiotic mycelium of the line containing CaGg/ sampled from the root surface of mycorrhized Lotus japonicus roots; B+/B- symbiotic mycelium: intraradical and extraradical mycelium sampled from *L. japonicus* roots colonized by the line with/without CaGg.

Supplementary Table S12: Distribution of genes coding for carbohydrate-active enzymes (CAZymes) in the sequenced AMF.

Supplementary Table S13: absolute expression of *G. margarita* chitin related genes shown in Fig. S13+

Supplementary Table S14: absolute expression of *G. margarita* SSPs shown in Fig. S13

Supplementary Table S15: *G. margarita* genes with a putative plant/bacterial origin.

Supplementary Table S16: differential expression of immunity-related *G. margarita* genes. The differential gene expression was calculated with DESeq2 and log2 normalized. The padj threshold was set to 0.05, and no filtering was applied to the log2 fold changes. B+/B- germinating spores: spores with/without the CaGg endobacterium/ grown for 10 days in H2O; B+/B- SL treatment: spores with/without the CaGg endobacterium/ pre-germinated for 3 days in H2O/ and then grown for 7 days in presence of the synthetic strigolactone GR24; B+ extraradical mycelium: symbiotic mycelium of the line containing CaGg/ sampled from the root surface of mycorrhized Lotus japonicus roots; B+/B- symbiotic mycelium: intraradical and extraradical mycelium sampled from *L. japonicus* roots colonized by the line with/without CaGg.

Supplementary Table S16: *G. margarita* possess several transporters which may support sugar uptake in the environment. By contrast, *G. margarita* depends on host-derived lipids, since it does not possess a cytosolic FASII,

while it possesses all the tools to further transform the acquired lipids.

Figure S1. Supplementary Figures

Appendix. S1: Supplementary text.