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

(Article begins on next page)

16 June 2024

1 **At the nexus of three kingdoms: the genome of the mycorrhizal fungus *Gigaspora margarita***

2 **provides insights into plant, endobacterial and fungal interactions**

3 Francesco Venice[1], Stefano Ghignone[2], Alessandra Salvioli di Fossalunga[1], Joëlle Amselem[3], Mara

4 Novero[1], Xie Xianan[4], Kinga Sędzielewska Toro[5], Emmanuelle Morin[6], Anna Lipzen[7,8], Igor V.

5 Grigoriev[7], Bernard Henrissat[9,10,11], Francis M. Martin[6], Paola Bonfante*[1]

6 [1]Department of Life Sciences and Systems Biology, University of Turin, Italy
7 [2] Institute for Sustainable Plant Protection -CNR- Turin Unit, Turin, Italy
8 [3]URGI, INRA, Université Paris-Saclay, Versaille, France
9 [4]State Key Laboratory for Conservation and Utilization of Subtropical Agro-Bioresources, Key
10 Laboratory of Innovation and Utilization of Forest Plant Germplasm in Guangdong Province,
11 College of Forestry and Landscape Architecture, South China Agricultural University, Guangzhou,
12 China
13 [5]Genetics, Faculty of Biology, Ludwig-Maximilians-University Munich, Planegg-Martinsried,
14 Germany
15 [6]Institut National de la Recherche Agronomique (INRA), Laboratory of Excellence Advanced
16 Research on the Biology of Tree and Forest Ecosystems (ARBRE), UMR 1136, Champenoux,
17 France
18 [7]Department of Energy Joint Genome Institute, Walnut Creek, CA, United States
19 [8]Department of Plant and Microbial Biology, University of California Berkeley, Berkeley CA
20 94720
21 [9]Architecture et Fonction des Macromolécules Biologiques, CNRS, Aix⫾Marseille Université,
22 Marseille, 13288 France
23 [10]Institut National de la Recherche Agronomique, USC1408 Architecture et Fonction des
24 Macromolécules Biologiques, Marseille, F-13288, France
25 [11]Department of Biological Sciences, King Abdulaziz University, Jeddah, 21589, Saudi Arabia
26
27
28 *Corresponding author:
29 Paola Bonfante
30 University of Turin
31 Department of Life Sciences and Systems Biology
32 Viale Mattioli, 25 – 10125 Turin, Italy
33 +390116705965
34 paola.bonfante@unito.it
35

36

37

38

39 **Running title:** Fungal Genome Adaptations: nexus of three kingdoms

Originality-Significance Statement

Many exciting studies have recently provided a wealth of data on the arbuscular mycorrhizal (AM) establishment and functioning. However, our knowledge is mostly plant- centric, and the biology of AM fungi is still enigmatic, notwithstanding some isolates have been recently sequenced. Here, we provide the genome sequence of the AM *Gigaspora margarita*, which resulted to be the largest fungal genome so far sequenced with the 64% of transposable elements. By crossing genomics with transcriptomics and experimental evidences, we offer new ideas on the biology of a fungus which is indeed a "holobiont", since it contains endobacterial populations and viral sequences, and constantly interacts with its host plant. We have identified novel elements (from immunity genes- never identified before in AM fungi- to horizontal transfer events) supporting the hypothesis that the presence of the endobacteria - together with the transposable elements- has shaped the fungal genome. Our genomic data open a new window on the evolutionary, environmental, and ecological meaning of an AM fungus which on one hand has a such huge genome and on the other an apparent limited distribution.

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 (Pimprikar and Gutjahr, 2018; MacLean et al., 2017; Lanfranco et al., 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.,

112   2010; Naito et al., 2015). Here, we present the assembly and annotation of the genome of

113   *Gigaspora margarita*, which belongs to the Gigasporaceae (Krüger et al., 2012), as G. rosea, and is

114   a member of an early diverging AMF group, well separated from Glomeraceae It presents a life

115   cycle with a conspicuous extra-radical phase consisting of huge spores, auxiliary cells and extra-

116   radical mycelium, as well as an intra-radical phase where intercellular hyphae support production of

117   arbuscules inside cortical cells (Supporting information Fig. S1). Intra-radical vesicles are not

118   formed and its presence in natural environments is mostly limited to sandy soils (Stürmer et al.,

119   2018). *G. margarita* may host *Burkolderia*- and MRE communities (Bianciotto et al., 1996; Desirò

120   et al., 2014), as well as viral sequences (Turina et al., 2018). The isolate used here contains a single

121   population of non-cultivable endobacteria, identified as *Candidatus* Glomeribacter gigasporarum

122   (*Ca*Gg), whose genome has been sequenced (Ghignone et al., 2012).

123   By means of an extended comparative genomic analysis, we reveal that the genome sequence of *G.*

124   *margarita* isolate BEG34 shares the peculiarities of the other AMF species (conserved core genes,

125   absence of plant cell wall-degrading enzymes, absence of type I fatty acid synthase), but is also

126   remarkably different from all other sequenced fungal genomes in its genome size (predicted to be

127   831 Mbp by flow cytometry) large number of transposable elements (TEs); in addition, several

128   horizontal gene transfer (HGT) events have been detected similar to other Diversisporaceae (Sun et

129   al., 2019). Thanks to the availability of OMICs data generated by using fungal lines with and

130   without bacteria (Salvioli et al., 2016; Vannini et al., 2016; Dearth et al., 2018), the genome

131   sequence of *G. margarita* sheds light on the evolutionary signatures of adaptation that have allowed

132   this early diverging fungus to interact with both plants and bacteria since Devonian times (Strullu-

133   Derrien et al., 2018). Novel features, i.e. presence of an immune system; secondary metabolite

134   production, and refined tuning of proteins and enzymes that target chitin, offer insights into the

135   biotrophic lifestyle of *G. margarita*. By providing a deeper understanding of inter-kingdom

136   interactions, the genomic information of *G. margarita* allows us to advance hypotheses on the

137   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 Kbp and 8 Kbp) libraries (Illumina Nextera). Sequencing of *G. margarita* BEG34 produced a first assembly of ~875 Mbp consisting of 6,453 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 *Ca*Gg endobacteria, Supporting information 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 *Ca*Gg 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 6,490 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 (Supporting information Table S1), the size of the *G. margarita* genome is second only to that of *Austropuccinia psidii* (Pucciniomycotina, Basidiomycota) at 1.2 Gbp, for 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 (Supporting information 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%).

Transposable elements dominate the *G. margarita* genome

The nuclear assembly revealed that repeated sequences cover 64 to 86% of the assembled genome sequence, depending on the filtering of highly degenerated copies (Fig. 2, Supporting information Fig. S4). These repeats can be ascribed to 4,264 consensus sequences, representing the putative ancestral TEs 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 which 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; transposable elements in the model *R. irregularis* (Chen et al., 2018) represent 22 to 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, Supporting information 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 (Supporting information 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 TIRs were expressed, revealing a differential expression across the fungal life stages (Supporting information 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* 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 horizontal gene transfer 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* TEs fragments had low frequency of nucleotide transitions (C to T, specifically; Supporting information 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 (Supporting information 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 (Supporting information Fig. S7a); a further screening revealed a significant spatial association between *G. margarita* candidate SSPs (Supporting information Table S3) and CRYPTON, TIR and MITE (Supporting information Table S4, Fig. S7b). In conclusion, a glimpse at the TEs present in *G. margarita* genome reveals many peculiar features: the heterogenous 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 defense 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 ten *G. margarita* RNA-seq libraries from different biological conditions (Supporting information Table S5). Finally, non-exon regions were identified using coordinates of repeated regions. We identified 26,603 coding genes (Table 1). The PASA pipeline refined 11,487 gene models, detecting 4,965 transcript isoforms, leading to 31,568 non-redundant transcripts. BUSCO assessment (Waterhouse et al., 2018) detected 98.5% of the conserved fungal gene 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). Fig. 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 ten largest fungal genomes (Supporting information 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 analyzed AMF, *D. epigaea* 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*, *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 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 (Supporting information 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 H_2O, or in the presence of a synthetic analogue of plant strigolactones, GR24) and symbiotic (intraradical and extraradical mycelium) stages (Supporting information Fig. S8; Table S6). Around 57% (15,205) of genes were expressed under all biological conditions. Pre-symbiotic and symbiotic stages were characterized by 2,431 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 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) (Supporting information 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 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 (Supporting information 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 analyzed 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 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 and colleagues (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 vs D. epigaea* differentiation (Supporting information Table S8a, Table S9 and Fig. S9 a-m). Among them, the *Gigaspora* species possess two expanded gene families containing alcohol oxidases which, in brown rot fungi, participate in Fenton chemistry (Hernández-Ortega et al., 2012; Guillén et al., 2000), 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 (OPT) 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* (Supporting information Table S8b, Table S9 and Fig. S9 a-m). Gene family evolution in *G. margarita vs G. rosea* may have been shaped by the interaction of *G. margarita* with *Ca*G, 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 up-regulated at the transcript level in germinating spores containing *Ca*Gg, when compared to cured spores (Supporting information Table S9). While it is already demonstrated that the expression of MATA-HMG can be influenced by *Burkholderia*-related endobacteria in their Mucoromycota hosts (Mondo et al., 2017), the putative effects of such regulation remains elusive for *G. margarita*. The expanded family of multicopper oxidases of the AA1 class (OG0000106) gathers two genes (g3919 and g11237) that are strongly up-regulated in the B+ line during the mycorrhizal symbiosis, and are also putatively secreted (Supplemental information Table S3). These extracellular oxidases are particularly abundant in plant-interacting Basidiomycota (Kües and Rühl, 2011), where they mediate plant cell wall depolymerazation. 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 up-regulation in response to the endobacterial presence during the germination phase (Supporting information 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 *Ca*Gg.

As the other AMF, whose obligate biotrophy has been shaped by 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 (Supporting information Table S6). However, a closer inspection revealed that these sequences are all intersected by TEs: the upstream genomic region of both spermidine synthases contain 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 (Supporting information 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-deprived 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 homologs), while other components are enriched (Supporting information Table S10). *G. margarita* encodes for at least 11 PT isoforms (Fig. 5, red dots), which are grouped into two distinct lineages. Accordingly, nine Pi transporters, GigmPT (g21463) to GigmPT5 (g26234) and their three paralogs GigmPT8 (g19323) to GigmPT10 (g6532), are closely related to the fungal PHO84 PTs and clustered with

Glomeromycotina PTs belonging to the Mucoromycota PHO84-like subfamily. Two other PTs, GigmPT6 (g10792) and GigmPT7 (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 (Supporting information Fig. S10, Table S11) of these *G. margarita* PTs, we speculate that GigmPT1 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 (Supporting information Fig. S10) revealed that most of the transporters belonging to the Glomeromycotina lineage were expressed in the extra-radical phase, confirming the largely acknowledged concept that AMF uptake Pi from the soil. GigmPT (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; Supporting information 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 (Supporting information 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 peri-arbuscular and intercellular spaces (Fig. 6, Supporting information Fig. S10). Interestingly, some ATPases and PT transporters are more expressed when the fungus contains its endobacterium (Supporting information Table S11); in particular the HA2 ATPase (g4412) is up-regulated by the presence of

the endobacterium in both germinating spores and intraradical mycelium. Furthermore, the vacuolar VTC1-2 (g24765) is up-regulated by the presence of *Ca*Gg in germinating spores and GR24-treated spores, suggesting a role in moving the Pi from the fungus to bacterium, which is compartmented inside fungal vacuoles (Supporting information Fig. S2), and which possesses its own Pi transporter (Ghignone et al., 2012; Ruiz-Lozano and Bonfante, 1999). These results are in agreement with previous data showing that *G. margarita* with its endobacterium has a general 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* labeling showed chitin in the cell wall in all steps of the AMF life cycle (Bonfante, 2018), while monoclonal antibodies failed to detect beta-1,3-glucans (Lemoine et al., 1995; Ligrone et al. 2007). Indeed, homologs 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 (Supporting information 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 (Supporting information Fig. S11). The number of secreted Carbohydrate Active Enzymes (CAZymes) involved in fungal wall deconstruction is very high in *G. margarita* (Supporting information 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 (Supporting information 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 production of glucosamine residues.

GT2, GH18, and CE4 family members, which are particularly expanded in the *G. margarita* genome (Supporting information Table S12) seem to be expressed homogeneously during the *G. margarita* life cycle (Supporting information 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 intra-radical phase.

Chitin has multiple roles in fungi as a structural and signalling molecule (Pusztahelyi, 2018; Schmitz and Harrison, 2014). This is particularly relevant in AMF, where chitin organization is modulated from the extra-radical to the intra-radical phase (Bonfante, 2018) and chitooligosaccharides (COs) are the main signaling molecules (Zipfel and Oldroyd, 2017). During the pre-symbiotic phase and in the extra-radical 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 defenses, inducing immunity signaling (Pozo and Azcón-Aguilar, 2007; Martinez-Medina et al., 2016; Chialva et al., 2018). Many studies have demonstrated that such plant defenses are limited to the first interaction phase (Giovannetti et al., 2015),

448 questioning whether N-acetyl COs are no longer active during intra-radical colonization. The

449 presence of expressed CE4 members suggests a new hypothesis: the deacetylation process could be

450 more important during the intra-radical phase when changes in molecular organization (loss of

451 acetyl groups) would correspond to loss of chitin fibrillar structure, as seen under the electron

452 microscope. Gow and colleagues (Gow et al., 2017) wrote that deacetylated chitin is not recognized

453 by plant chitin receptors, allowing a deep colonization of plant tissues without any evident rejection.

454 Another alternative hypothesis to explain the stealth colonization of plant tissues by AMF, is based

455 on LysM domain containing proteins. AMF could sequestrate chitin oligosaccharides to elude

456 host's immunity, as already reported for pathogenic Ascomycetes (de Jonge and Thomma, 2009);

457 all the sequenced AMF (except for *G. rosea*) possess one or more of such enzymes, which are

458 gathered in the CBM50 CAZy class (Supporting information Table S12). Lastly, the progressive

459 thinning of the arbuscular cell wall may guarantee passage of SSPs, which act as fungal effectors

460 (Kloppholz et al., 2011). In *G. margarita*, 41 SSPs had absolute expression levels around 10 times

461 greater in the intra-radical mycelium than at all other life stages (Supporting information Fig. S13,

462 Table S14). While no experimental evidence is available of the secretion of these proteins, our

463 results point to the arbuscule as a preferential site for effector production.

464 Combining genomics, transcriptomics and morphological data reveals the finely tuned activity of

465 genes related to the fungal cell wall, and in particular of chitin metabolism, with potential feedback

466 on signaling and defense mechanisms in plant hosts.

467

468 **Between plant and bacterial cells: HGT events**

469 During its life cycle, *G. margarita* lives in intimate contact with both its plant host and its

470 endobacteria (Supporting information Fig. S2). Since this inter-kingdom interaction is expected to

471 have been stable for more than 400 million years (Mondo et al., 2012), we hypothesized that HGT

472 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 (Supporting information 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 (Supporting information 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. Similar sequences belong to a *Gigaspora* specific expanded gene family (Supporting information 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 similarity with bacterial Toll/interleukin-1 receptor/resistance domain-containing proteins (TIR). 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 homologs in MREs, such as the endobacteria of *Dentiscutata heterogama* (Torres-Cortés et al., 2015; Naito et al., 2015) and *D. epigaea* (Sun et al., 2019) but not in *Ca*Gg (Ghignone et al., 2012). *G. 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 *Ca*Gg, and is inactivated upon GR24 treatment (Supporting information Table S16). By contrast, g16267 shows uniform expression across biological conditions. According to Sun and colleagues (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 multilayered; 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.

G. 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 allorecognition 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 2 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) (Supporting information 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 up-regulated upon treatment with the plant signal analogue GR24 (Supporting information 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* (Supporting information 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 (Supporting information 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). *G. 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. *G. margarita* and *G. rosea* share an enriched set of CAZymes, which explains 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 extra-radical 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 defenses. The absence of beta 1-3-glucans in the wall of the intra-radical 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 *Ca*Gg 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 Nod-like-receptors 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 phosphate transporters 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. *G. 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 amongst the most nitrogen- and phosphorous-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 which is commonly associated with long-term survival strategies (Aguilar-Trigueros 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

G. margarita strain BEG34 with (B+) and without (B-) the *Ca*Gg endobacterium was used for this study. The spores were obtained as described in Salvioli et al., 2016 (19) (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 cm^2 of young *S. lycopersicum* leaf tissues were processed as described in Sędzielewska et al., 2011 (Sędzielewska et al., 2011) and measured using a FACSAria (BD Biosciences, San Josè, 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 Kbp and 8 Kbp) were constructed in 2 X 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/; last accessed 2019-02-28), 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 (Supporting information 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 (Hoff et al., 2016; Stanke et al., 2008; Stanke et al., 2006; Lomsadze et al., 2014). 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 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 Transposable Elements (TE) 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 consensuses, with at least one full-length fragment (FLF), 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 explanations on TEs discovery are 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.

Acknowledgments

680 analysis, and to Dr J. Mach for the critical reading of the manuscript.

681

682

683

686

687

688

689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726

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(3):379–89.

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(12): 1417.

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(2): 81–92.

Balzergue, C., Chabaud, M., Barker, D.G., Bécard, G., and Rochange, 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., and Bonfante, P. (1992) Quantification of the nuclear DNA content of two arbuscular mycorrhizal fungi. Mycol Res 96(12): 1071–6.

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(8): 3005–10.

Bolger, A.M., Lohse, M., and Usadel, B. (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30(15): 2114–20.

Bonfante, P. (2018) The future has roots in the past: the ideas and scientists that shaped mycorrhizal research. New Phytol 220(4): 982–95.

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(8): 1727–35.

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

Bonfante, P., Balestrini, R., and Mendgen, K. (1994) Storage and secretion processes in the spore of *Gigaspora margarita* Becker & Hall as revealed by high-pressure freezing and freeze substitution. New Phytol 128(1): 93–101.

Bonfante, P., Venice, F., and Lanfranco, L. (2019). The Mycobiota: Fungi take their place between Plants and Bacteria. Current Opinion in Microbiology, under revision.

Bushnell, B. (2014) BBMap: A Fast, Accurate, Splice-Aware Aligner. Lawrence Berkeley National Lab. (LBNL), Berkeley, CA (United States). Report No.: LBNL-7065E. Available from: https://www.osti.gov/biblio/1241166-bbmap-fast-accurate-splice-aware-aligner

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(4): 1440–6.

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(6): 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(1): 511–25.

Chen, E.C.H., Morin, E., Beaudet, D., Noel, J., Yildirir, G., Ndikumana, S., *et al.* (2018) High intraspecific genome diversity in the model arbuscular mycorrhizal symbiont *Rhizophagus irregularis*. New Phytol 220(4): 1161–71.

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(4): 1296–308.

Coombes, B.K. (2009) Type III secretion systems in symbiotic adaptation of pathogenic and non-pathogenic bacteria. Trends Microbiol 17(3): 89–94.

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(6251): 970–3.

Davison, J., Moora, M., Öpik, M., Ainsaar, L., Ducousso, 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(9): 2211–24.

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(10): 1269–71.

de Jonge, R., Thomma, B.P.H.J. (2009) Fungal LysM effectors: extinguishers of host immunity? Trends Microbiol 17(4):151–7.

Dearth, S.P., Castro, H.F., Venice, F., Tague, E.D., Novero, M., Bonfante, P., *et al.* (2018) Metabolome changes are induced in the arbuscular mycorrhizal fungus *Gigaspora margarita* by germination and by its bacterial endosymbiont. Mycorrhiza 28(5–6): 421–33.

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 USA 112(43): 13390–5.

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. (2): 257–70.

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(12): 3137–58.

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(7): 1575–84.

Ezawa, T., and Saito, K. (2018) How do arbuscular mycorrhizal fungi handle phosphate? New insight into fine-tuning of phosphate metabolism. New Phytol 220(4): 1116–21.

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(5): 1267–77.

Flutre, T., Duprat, E., Feuillet, C., and Quesneville, H. (2011) Considering transposable element diversification in de novo annotation approaches. PLoS One 2011 6(1): 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 Ca_2^+ spiking in *Medicago truncatula* roots and their production is enhanced by strigolactone. New Phytol 198(1): 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(1): 136–45.

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(3): 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 USA 108(4): 1513–8.

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(10): 3420–35.

Gow, N.A.R., Latge, J.P., and Munro, C.A. (2017) The fungal cell wall: structure, biosynthesis, and function. Microbiol Spectr 5(3).

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(1): 142–7.

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(1): 478.

Harrison, M.J., and van Buuren, M.L. (1995) A phosphate transporter from the mycorrhizal fungus *Glomus versiforme*. Nature 378(6557): 626–9.

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 USA 115(10): 2292–301.

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(4): 1395–410.

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(5): 767–9.

Kamanova, J., Sun, H., Lara-Tejero, M., and Galán, J.E. (2016) The salmonella effector protein SopA modulates innate immune responses by targeting TRIM E3 ligase family members. PLoS Pat 12(4): e1005552.

Kloppholz, S., Kuhn, H., and Requena, N. (2011) A secreted fungal effector of *Glomus intraradices* promotes symbiotic biotrophy. Curr Biol 21(14): 1204–9.

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 Schüßler, A. (2012), Phylogenetic reference data for systematics and phylotaxonomy of arbuscular mycorrhizal fungi from phylum to species level. New Phytologist, 193: 970-984.

Kües, U. and Rühl, M. (2011) Multiple Multi-Copper Oxidase Gene Families in Basidiomycetes – What for? Curr Genomics 12(2): 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(1): 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(4): 1031–46.

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(11): e28231.

Lemoine, M.C., Gollotte, A., and Gianinazzi-Pearson, V. (1995) Localization of β(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(1): 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(6): 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(11):1756–77.

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(1): 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(1): 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(15): e119.

MacLean, A.M., Bravo, A., and Harrison, M.J. (2017) Plant Signaling and Metabolic Pathways Enabling Arbuscular Mycorrhizal Symbiosis. Plant Cell. 29(10): 2319–35.

Maculins, T., Fiskin, E., Bhogaraju, S., and Dikic, I. (2016) Bacteria-host relationship: ubiquitin ligases as weapons of invasion. Cell Res 26(4): 499–510.

Martin, F., Kohler, A., Murat, C., Balestrini, R., Coutinho, P.M., Jaillon, O., *et al.* (2010) Périgord black truffle genome uncovers evolutionary origins and mechanisms of symbiosis. Nature 464(7291):1033–8.

Martin, M. (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal 17(1): 10–2.

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(10): 818–22.

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(3): 729–34.

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(11): 1864–72.

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 Comm 8(1): 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(8): 2564–76.

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(3): 1584-1598

919 Murat, C., Kuo, A., Barry, K.W., Clum, A., Dockter, R.B., Fauchery, L., *et al.* (2018) Draft genome
920 sequence of *Tuber borchii* Vittad., a whitish edible truffle. Genome Announc 6(25).

921 Muszewska, A., Steczkiewicz, K., Stepniewska-Dziubinska, M., and Ginalski, K. (2017) Cut-and-
922 paste transposons in fungi with diverse lifestyles. Genome Biol Evol 9(12): 3463-3477.

923 Naito, M., Morton, J.B., and Pawlowska, T.E. (2015) Minimal genomes of mycoplasma-related
924 endobacteria are plastic and contain host-derived genes for sustained life within Glomeromycota.
925 Proc Natl Acad Sci USA 112(25): 7791–6.

926 Naumann, M., Schüssler, A., and Bonfante, P. (2010) The obligate endobacteria of arbuscular
927 mycorrhizal fungi are ancient heritable components related to the Mollicutes. ISME J. 4(7): 862–71.

928 Oliveira-Garcia, E., and Deising, H.B. (2013) Infection structure-specific expression of β-1,3-
929 glucan synthase is essential for pathogenicity of *Colletotrichum graminicola* and evasion of β-
930 glucan-triggered immunity in maize. Plant Cell 25(6): 2356–78.

931 Patro, R., Duggal, G., Love, M.I., Irizarry, R.A., Kingsford, C. (2017) Salmon provides fast and
932 bias-aware quantification of transcript expression. Nature Methods 417–9.

933 Pawlowska, T.E., Gaspar, M.L., Lastovetsky, O.A., Mondo, S.J., Real-Ramirez, I., Shakya, E., *et al.*
934 (2018) Biology of fungi and their bacterial endosymbionts. Annu Rev Phytopathol 56(1): 289–309.

935 Pelin, A., Pombert, J.F., Salvioli, A., Bonen, L., Bonfante, P., and Corradi, N. (2012) The
936 mitochondrial genome of the arbuscular mycorrhizal fungus *Gigaspora margarita* reveals two
937 unsuspected trans-splicing events of group I introns. New Phytol 194(3): 836–45.

938 Pellicer, J., Hidalgo, O., Dodsworth, S., and Leitch, I.J. (2018) Genome size diversity and its impact
939 on the evolution of land plants. Genes (Basel) 9(2).

940 Pimprikar, P., and Gutjahr, C. (2018) Transcriptional Regulation of Arbuscular Mycorrhiza
941 Development. Plant Cell Physiol 59(4): 673–90.

942 Plomion, C., Aury, J.M., Amselem, J., Leroy, T., Murat, F., Duplessis, S., *et al.* (2018) Oak genome
943 reveals facets of long lifespan. Nature Plants 4(7): 440.

944 Pozo, M.J., and Azcón-Aguilar, C. (2007) Unraveling mycorrhiza-induced resistance. Curr Opin
945 Plant Biol 10(4): 393–8.

946 Pusztahelyi, T. (2018) Chitin and chitin-related compounds in plant-fungal interactions. Mycology
947 9(3): 189–201.

948 Ropars, J., Toro, K.S., Noel, J., Pelin, A., Charron, P., Farinelli, L., *et al.* (2016) Evidence for the
949 sexual origin of heterokaryosis in arbuscular mycorrhizal fungi. Nat Microbiol 1(6): 16033.

950 Ruiz-Lozano, J.M., and Bonfante, P. (1999) Identification of a putative P-transporter operon in the
951 genome of a Burkholderia strain living inside the arbuscular mycorrhizal fungus *Gigaspora*
952 *margarita*. J Bacteriol 181(13): 4106–9.

953 Salvioli, A., Ghignone, S., Novero, M., Navazio, L., Venice, F., Bagnaresi, P., *et al.* (2016)
954 Symbiosis with an endobacterium increases the fitness of a mycorrhizal fungus, raising its
955 bioenergetic potential. ISME J 10(1): 130–44.

Sánchez-Vallet, A., Fouché, S., Fudal, I., Hartmann, F.E., Soyer, J.L., Tellier, A., *et al.* (2018) The genome biology of effector gene evolution in filamentous plant pathogens. Annu Rev Phytopathol 56(1): 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(2): 171–183.

Schmitz, A.M., and Harrison, M.J. (2014) Signaling events during initiation of arbuscular mycorrhizal symbiosis. J Integr Plant Biol 56(3): 250–61.

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(4): 289–95.

Sędzielewska, K.A., Fuchs, J., Temsch, E.M., Baronian, K., Watzke, R., and Kunze, G. (2011) Estimation of the *Glomus intraradices* nuclear DNA content. New Phytol 192(4): 794–7.

Smith, S., and Read, D. (2008) Mycorrhizal Symbiosis - 3rd Edition, Academic Press, London.

Spanu, P.D., Abbott, J.C., Amselem, J., Burgis, T.A., Soanes, D.M., Stüber, K., *et al.* (2010) Genome expansion and gene loss in powdery mildew fungi reveal tradeoffs in extreme parasitism. Science 330(6010): 1543–6.

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(5): 1028–46.

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(5): 637–44.

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.

Stenmark, H. (2009) Rab GTPases as coordinators of vesicle traffic. Nat Rev Mol Cell Biol 10(8): 513–25.

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(7): 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(3): 1556–73.

Tang, F., Wang, B., Li, N., Wu, Y., Jia, J., Suo, T., *et al.* (2011) RNF185, a novel mitochondrial ubiquitin E3 ligase, regulates autophagy through interaction with BNIP1. PLoS ONE 6(9): e24367.

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(6): 2044–64.

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. USA 110(50): 20117–22.

Torres-Cortés, G., Ghignone, S., Bonfante, P., and Schüßler, A. (2015) Mosaic genome of endobacteria in arbuscular mycorrhizal fungi: transkingdom gene transfer in an ancient mycoplasma-fungus association. Proc Natl Acad Sci USA 112(25): 7785–90.

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(6): 2012–25.

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 Mycoavidus cysteinexigens. Environ Microbiol 19(8): 2964–83.

Van der Biezen, E.A., and Jones, J.D. (1998) Plant disease-resistance proteins and the gene-for-gene concept. Trends Biochem Sci 23(12): 454–6.

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(1): 265–75.

Ve, T., Williams, S.J., and Kobe, B. (2015) Structure and function of Toll/interleukin-1 receptor/resistance protein (TIR) domains. Apoptosis 20(2): 250–61.

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, editor. Biochemistry and Molecular Biology. Springer International Publishing; p. 361–85.

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(3): 543–8.

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(12): 1583–608.

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(2): 258–67.

Zipfel, C., and Oldroyd, G.E.D. (2017) Plant signalling in symbiosis and immunity. Nature 543(7645): 328–36.

Figure Legends

Figure 1: *G. margarita* BEG34 as a metaorganism A) a *G. 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) *Ca*Gg (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 colored 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.

Figure 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".

Figure 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 depend on the number of orthologs (the thicker/darker, the higher the number of orthologs).

Figure 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 are 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.

Figure 5: evolutionary relationships of phosphate transporters among fungal species. the evolutionary history was inferred using the Neighbor-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 was a total of 58 positions in the final dataset. Evolutionary analyses were conducted in MEGA7. The marked red circles shown in the figure indicate the multiple isoforms of phosphate transporters from *G. margarita.*

Figure 6: By crossing genomics and transcriptomic data the scheme illustrates an 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 *Ca*Gg 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 (PtsS). Furthermore, the bacterium has cytoplasmic components which may respond to intracellular Pi (PstB and PhoU). All the membrane localisations of the fungal P transporters are hypothetical.

Figure 7: *G. 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.

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

In addition to these files we provide:

16 supplemental tables (individual excel workbooks)

16 supplemental figures (embedded as separate pages in a single PDF file)

1 Supplementary text file (a word document containing supplementary Materials and Methods and Results)