IRIS AperTO

UNIVERSITÀ DEGLI STUDI DI TORINO

AperTO - Archivio Istituzionale Open Access dell'Università di Torino

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

This is the author's manuscript

Original Citation:

Availability:

This version is available http://hdl.handle.net/2318/1728284 since 2020-02-18T17:16:37Z

Published version:

DOI:10.1111/1462-2920.14827

Terms of use:

Open Access

Anyone can freely access the full text of works made available as "Open Access". Works made available under a Creative Commons license can be used according to the terms and conditions of said license. Use of all other works requires consent of the right holder (author or publisher) if not exempted from copyright protection by the applicable law.

(Article begins on next page)

24 May 2024

1 **At the nexus of three kingdoms: the genome of the mycorrhizal fungus *Gigaspora margarita* provides insights into plant, endobacterial and fungal interactions**
2

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

40 **Originality-Significance Statement**

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

54

55

56

57

58

59

60

61

62

63

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.

87 **Introduction**

88 Arbuscular mycorrhizal symbiosis involves 72% of vascular plants and an ancient group of fungi
89 (arbuscular mycorrhizal fungi, AMF) that enhance host fitness by improving mineral nutrition and
90 resistance to biotic and abiotic stresses (Bonfante, 2018). The large number of sequenced plant
91 genomes and availability of mutants allowed many of the mechanisms underlying plant responses to
92 colonization by AMF to be deciphered (Pimprikar and Gutjahr, 2018; MacLean et al., 2017;
93 Lanfranco et al., 2018). By contrast, the biology of the fungal partners remains largely unexplained.
94 AMF are important members of the plant microbiota, present in the rhizosphere and inside roots
95 tissues as obligate symbionts (Davison et al., 2015). As many other fungi, AMF also possess their
96 own microbiota (Bonfante et al, 2019), often hosting uncultivable endobacteria inside their
97 cytoplasm (Bonfante and Desirò, 2017, Pawlowska et al., 2018). AMF genomes therefore offer an
98 unexplored source for investigating the evolution of inter-kingdom interactions among plants, fungi
99 and bacteria.

100 The genome of the model fungus *Rhizophagus irregularis* was the first AMF to be sequenced
101 (Tisserant et al., 2013; Lin et al., 2014; Chen et al., 2018), offering novel insights into the genetics
102 of a member of the Glomeromycotina, such as the surprising absence of plant cell wall-degrading
103 enzymes and the unexpected presence of genes potentially involved in sexual reproduction.
104 Genome sequences of other AMF followed: *Rhizophagus clarus* (Kobayashi et al., 2018),
105 *Rhizophagus cerebriforme* and *Rhizophagus diaphanus* (Morin et al., 2019), as well as two
106 members of the Diversisporales: *Diversispora epigea* (Sun et al., 2019) and *Gigaspora rosea*
107 (Morin et al., 2019). Irrespective of their phylogenetic position in the AMF tree, these novel
108 genomes consistently confirmed the obligate biotrophic nature of AMF, revealing the loss of genes
109 involved in plant polysaccharide degradation and fatty acid biosynthesis. Most of the sequenced
110 *Rhizophagus* strains, as well as *G. rosea,* do not possess endobacteria (Bonfante and Desirò, 2017),
111 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.

138 **Results and Discussion**

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

141 We sequenced the genome of *G. margarita* BEG34 using one paired-end library (Illumina TruSeq
142 Nano) and two mate-pair (3 Kbp and 8 Kbp) libraries (Illumina Nextera). Sequencing of *G.*
143 *margarita* BEG34 produced a first assembly of ~875 Mbp consisting of 6,453 scaffolds.
144 By using BlobTools (Laetsch and Blaxter, 2017) analysis for the visualization and partitioning of
145 metagenome assemblies, scaffolds were visualized as circles of variable size, separated on the basis
146 of their coverage and G/C content, and their taxonomic affiliation was inferred through sequence
147 homology. This analysis allowed us to separate the nuclear component, which is the focus of this
148 study (Fig. 1), from the other compartments containing DNA (mitochondria and *Ca*Gg
149 endobacteria, Supporting information Fig. S2). The mitochondrial reads were re-assembled into a
150 complete circular chromosome (96,998 bp), agreeing with the results of Pelin et al. (2012), while
151 the *Ca*Gg reads led to an improved assembly (2.07 Mbp) compared with the previous version
152 (Ghignone et al., 2012). The nuclear genome, obtained by reassembling the nuclear reads, consists
153 of 6,490 scaffolds and is ~758 Mbp (Table 1), almost six times larger than that of the model species
154 *R. irregularis*, and significantly larger than those of the saprotrophic and phylogenetically related
155 *Mortierella elongata* that has a genome size of around 49 Mbp (Uehling et al., 2017). According to
156 NCBI and JGI Mycocosm (Supporting information Table S1), the size of the *G. margarita* genome
157 is second only to that of *Austropuccinia psidii* (Pucciniomycotina, Basidiomycota) at 1.2 Gbp, for
158 which no functional annotation is available (McTaggart et al., 2018).
159 This result was supported by flow cytometry, which revealed a DNA quantity for *G. margarita*
160 corresponding to a genome of about 831 Mbp (Supporting information Fig. S3). Indeed, previous
161 static cytometry suggested a comparable value (Bianciotto and Bonfante, 1992). Similar to other
162 AMFs, the *G. margarita* genome has a G+C content below 30% (27.68%).

163 **Transposable elements dominate the *G. margarita* genome**

164 The nuclear assembly revealed that repeated sequences cover 64 to 86% of the assembled genome 165 sequence, depending on the filtering of highly degenerated copies (Fig. 2, Supporting information 166 Fig. S4). These repeats can be ascribed to 4,264 consensus sequences, representing the putative 167 ancestral TEs sequences from which they originated. The proportion of TEs is very high when 168 compared with the generally low value found in fungal genomes (Castanera et al., 2016), usually 169 ranging from 0 to 25% with the exception of some plant pathogens (Spanu et al., 2010) and 170 ectomycorrhizal species that have undergone massive TE amplifications (Martin et al., 2010). 171 Among the latter, Tuberaceae show a genome expansion which correlates with TEs increase, 172 accounting for 50% of the genomic content (Murat et al., 2018). Similar to the related *G. rosea* 173 (Morin et al, 2019), *G. margarita* confirms this relationship well: the huge genome size is 174 accompanied by the highest proportion of TEs detected so far in AMF; transposable elements in the 175 model *R. irregularis* (Chen et al., 2018) represent 22 to 26% of the total assemblies (from 122 to 176 138 Mbp).

177 Overall, 63% of the *G. margarita* TE content (40.6% of genome) remains unclassified. Of the 178 remainder, the largest group of TEs is classified as class I retrotransposons (12% of the genome) 179 (Fig. 2, Supporting information Fig. S4), where the major classes are LINE (8%) and LTR Gypsy 180 (2.5%) elements. At least one full length Gypsy element (g2860) was found to be active based on 181 gene expression data (Supporting information Fig. S5). Since these groups are well represented in 182 *G. rosea* (Morin et al., 2019) and *D. epigaea* (Sun et al., 2019) compared to *Rhizophagus* species, a 183 Diversisporaceae-specific expansion might be hypothesized.

184 Among the DNA-repeats, at least 15 full-length TIRs were expressed, revealing a differential 185 expression across the fungal life stages (Supporting information Fig. S5, Table S2). Another 186 peculiar feature of *G. margarita* DNA-TEs consists in the presence of repeats with homology to the 187 *sola* class, which is present in plants, bacteria and metazoa, but previously considered to be absent 188 in fungi, except for *R. irregularis* (Gladyshev and Kleckner, 2017). Lastly, *G. margarita* genome

189 harbours Helitrons, which are characterized by a rolling circle replication mechanism and which
190 may contain genes captured from other organisms. They have been demonstrated to be mediators of
191 horizontal gene transfer in other fungi (HGT, see specific paragraph) (Castanera et al., 2014).
192 Usually, TE presence is counterbalanced by host genome defences, including fungal repeat-induced
193 point mutation (RIP), methylation induced pre-meiotically (MIP), meiotic silencing of unpaired
194 DNA (MSUD) and quelling (Muszewska et al., 2017). When compared to their ancestor sequences
195 (consensus sequences), the *G. margarita* TEs fragments had low frequency of nucleotide transitions
196 (C to T, specifically; Supporting information Fig. S6a), which are operated by DNA-methylases and
197 are the outcome of RIP activity (Amselem et al., 2015). Furthermore, AT-rich regions could not be
198 statistically detected (Supporting information Fig. S6b), suggesting that *G. margarita* lacks RIP
199 activity, while the impact of MIP remains to be evaluated. We inferred the absence of MSUD
200 activity as no homology could be found with the associated genes so far characterized in
201 *Neurospora crassa*; by contrast, small RNA-mediated TEs silencing (quelling) cannot be excluded,
202 as *G. margarita* genome encodes for the genes related to this pathway (see the Immune System
203 section), such as Dicer and Argonaute proteins.
204 It has been suggested that gene-sparse regions in fungal genomes contain small secreted proteins
205 (SSPs) which are often associated with TEs, providing a favourable environment for the
206 diversification of fungal effector repertoire (Sánchez-Vallet et al., 2018). The *G. margarita* genome
207 is also gene-sparse (Supporting information Fig. S7a); a further screening revealed a significant
208 spatial association between *G. margarita* candidate SSPs (Supporting information Table S3) and
209 CRYPTON, TIR and MITE (Supporting information Table S4, Fig. S7b). In conclusion, a glimpse
210 at the TEs present in *G. margarita* genome reveals many peculiar features: the heterogenous TEs
211 groups suggest that Diversisporaceae experienced some specific TE bursts; Helitrons could be
212 involved in HGT events; lastly, the TE abundance mirrors the absence of fungal genome defense
213 barriers, confirmed by the expression of some TE typologies.
214

215 **Genomic features of *Gigaspora margarita***

216 The gene prediction process combined *ab initio* prediction and hints from physical evidence (see the
217 Methods section). Exon coordinates were determined by genome mapping of the entire set of
218 UniProt Mucoromycota proteins; intron/exon boundaries were defined by spliced alignment of ten
219 *G. margarita* RNA-seq libraries from different biological conditions (Supporting information Table
220 S5). Finally, non-exon regions were identified using coordinates of repeated regions. We identified
221 26,603 coding genes (Table 1). The PASA pipeline refined 11,487 gene models, detecting 4,965
222 transcript isoforms, leading to 31,568 non-redundant transcripts. BUSCO assessment (Waterhouse
223 et al., 2018) detected 98.5% of the conserved fungal gene set as complete sequences, indicating that
224 this gene catalogue is highly comprehensive. The total number of genes found for *G. margarita* was
225 close to that of the other Glomeromycotina so far sequenced, *R. irregularis*, *R. clarus*, *R.*
226 *cerebriforme*, *R. diaphanus D. epigea* and *G. rosea* (Chen et al., 2018; Kobayashi et al., 2018;
227 Morin et al., 2019; Sun et al., 2019). Fig. 3a highlights the exceptionally high genome size of the
228 two *Gigaspora* species, when compared with other AMF, ectomycorrhizal fungi and free-living
229 Mucoromycota. Alongside *G. margarita* and *G. rosea*, the ten largest fungal genomes (Supporting
230 information Table S1) include other obligate biotrophs (rusts from Pucciniomycotina), an
231 endophytic fungus and a gut fungus (*Zoophthora radicans* and *Neocallimastix californiae*,
232 respectively). Several of these genomes lack annotation, leaving open the question of whether their
233 expansions also implied an increase in gene number. AMF possess more genes than their free-living
234 Mucoromycota relatives; however, plant symbiotic lifestyle and gene number increase do not seem
235 to be strictly related, since ectomycorrhizal truffles (Murat et al., 2018) and obligate plant
236 pathogens (Spanu et al., 2010) possess a reduced gene number. Coding space, defined as the space
237 occupied by the protein-coding genes, is another informative parameter: as expected, among the
238 analyzed AMF, *D. epigaea* and *G. rosea* possess the largest coding space, in line with their higher
239 gene number (Fig. 3a). A similarity network built up on groups of orthologs (Fig. 3b), and based on
240 gene composition rather than gene number, revealed a considerable distance between AMF genera

241 (*Gigaspora*, *Diversispora*, *Rhizophagus*), while the intra-genus diversity was low. These
242 quantitative parameters suggest that, irrespectively of genome size and comparable gene number,
243 Gigasporaceae and Glomeraceae are characterized by different genes.
244

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

247 Having predicted the gene repertoire of *G. margarita*, we combined gene information with a large
248 set of transcriptomic data (Supporting information Table S5). In *G. margarita* native line containing
249 the endobacterium (B+), ~82% of the predicted genes were expressed at the transcript level at least
250 in one of the examined biological conditions: both asymbiotic (spores germinated in H_2O, or in the
251 presence of a synthetic analogue of plant strigolactones, GR24) and symbiotic (intraradical and
252 extraradical mycelium) stages (Supporting information Fig. S8; Table S6). Around 57% (15,205) of
253 genes were expressed under all biological conditions. Pre-symbiotic and symbiotic stages were
254 characterized by 2,431 and 885 unshared genes, respectively. Genes encoding enzymes (such as
255 oxidases, carbohydrate-active enzymes and peptidases), transporters and even SSPs were
256 specifically expressed in these two stages, suggesting a transcriptional shift involved in
257 development, nutrient acquisition and communication with the host. Most of the expressed
258 symbiotic genes have no known function, as observed in other AMF (Morin et al., 2019). A
259 significant number of genes (444) which were expressed at least in one of the examined conditions
260 in the B+ line, were never expressed in the cured fungal line (B-, without the endobacteria)
261 (Supporting information Table S6). Similarly, 38 genes were exclusively expressed in the B- line.
262 While most of these B+ or B- specific genes (~63%) are functionally uncharacterized, others
263 encode for enzymes putatively involved in DNA binding, replication and transcription, such as zinc
264 finger domain-containing proteins, far1-related proteins and HMG-box transcription factors, which
265 are indicated as regulators of cryptic sexuality events in AMF (Ropars et al., 2016). A number of
266 genes encoding for peptidases are exclusively expressed in the B+ line, as well as two genes

267 encoding for hydrolases with peptidoglycan as predicted substrate (glycoside hydrolase family 25
268 proteins).
269 The following sections will highlight the evolution of gene families and diversification of relevant
270 and /or novel gene categories in the context of *G. margarita* lifestyle, including biotrophy and
271 endobacterium presence.
272

273 **Evolution of gene families in Gigasporaceae**

274 We used the Orthofinder (Emms and Kelly, 2015) and CAFE (De Bie et al., 2006) algorithms and a
275 set of fungi to reconstruct the AMF phylogeny and compare close relatives possessing different
276 lifestyles (Supporting information Table S7). Alongside AMF, we included free-living
277 Mucoromycota, plus two ectomycorrhizal fungi (*Tuber melanosporum* and *Laccaria bicolor*) from
278 distant clades and a saprotroph/opportunistic pathogen (*Aspergillus fumigatus*), considered as
279 outgroup. The pipeline first performed a homology search between the entire proteomes of the
280 selected species; MCL clustering (Enright et al., 2002) was then used to process the homology
281 results, generating 17,341 orthogroups, i.e. gene families gathering orthologs, incorporating ~81%
282 of the 287,540 analyzed proteins. Roughly 84% of *G. margarita* proteins were clustered within
283 these orthogroups. Inter-species distances were calculated through multiple sequence alignment for
284 each orthogroup containing one protein per species (533 single-copy orthogroups). These distances
285 were summarized in a phylogenetic tree, converted into evolutionary time, and finally used by
286 CAFE to detect, for each node and leaf of the tree, orthogroups that underwent accelerated gene
287 gains or losses, interpreted in this context as the effects of evolutionary pressure. The tree was
288 calibrated using the divergence date of 434 MYA between *M. elongata* and *R. irregularis* (Uehling
289 et al., 2017). The reconstructed phylogenetic tree (Fig. 4) indicates that *G. margarita* and *G. rosea*
290 diverged very recently, at around 13 MYA Different from Glomerales, for which all the sequenced
291 fungi belong to the genus *Rhizophagus* within the Glomeraceae, a divergence node (~174 MYA)
292 between Diversisporaceae (*D. epigaea)* and Gigasporaceae (*G. margarita* and *G. rosea*) is present

293 in Diversisporales. Most of the estimated phylogenetic distances resemble those obtained by Chang
294 and colleagues (Chang et al., 2019) in the context of a deeper phylogenetic analysis, which was
295 designed for the Endogonales family, also belonging to Mucoromycota. Looking for
296 Gigasporaceae-specific traits, we focused on expansion/contraction events that took place after
297 divergence from the Diversisporaceae. Of the 271 expansion events observed in the *Gigaspora*, 120
298 originated in *Gigaspora vs D. epigaea* differentiation (Supporting information Table S8a, Table S9
299 and Fig. S9 a-m). Among them, the *Gigaspora* species possess two expanded gene families
300 containing alcohol oxidases which, in brown rot fungi, participate in Fenton chemistry (Hernández-
301 Ortega et al., 2012; Guillén et al., 2000), and contribute to the non-enzymatic degradation of plant
302 cell walls. The analysis also highlighted the presence of a *Gigaspora*-specific feature, i.e. the
303 potential to synthesize secondary metabolites due to the presence of Polyketide synthases, which are
304 normally absent in basal fungi (see also the paragraph on HGT). A rapid gene gain for *Gigaspora*
305 species was particularly dramatic in the families of oligopeptide (OPT) transporters and of patatins,
306 which can be related to fungal immunity (see the specific paragraph).
307 After the divergence between G. rosea and G. margarita (Fig. 4), our pipeline identified 117
308 families that underwent rapid expansion and 133 that underwent rapid contraction in *G. margarita*
309 (Supporting information Table S8b, Table S9 and Fig. S9 a-m). Gene family evolution in *G.*
310 *margarita vs G. rosea* may have been shaped by the interaction of *G. margarita* with *Ca*G, which -
311 so far- has never been detected in *G. rosea.* Gene loss events in *G. margarita* include proteins
312 related to DNA replication and repair, while among the 117 rapidly expanded families, we found a
313 family of fungal-like Nod-like receptors identified by a central NACHT domain and putatively
314 involved in non-self recognition (discussed later), and MATA-HMG proteins (OG0000316). This
315 latter family of highly diverse transcription factors is widespread among AMF genomes (Morin et
316 al., 2019) and is suggested to regulate cryptic sexuality events in AMF (Ropars et al., 2016).
317 According to our pipeline, MATA-HMG of AMF are clustered in more than one orthogroup, but in
318 OG0000316 *G. margarita* possesses the highest number of proteins; furthermore, two of these

319 genes (g24252 and g7196) are up-regulated at the transcript level in germinating spores containing
320 *Ca*Gg, when compared to cured spores (Supporting information Table S9). While it is already
321 demonstrated that the expression of MATA-HMG can be influenced by *Burkholderia*-related
322 endobacteria in their Mucoromycota hosts (Mondo et al., 2017), the putative effects of such
323 regulation remains elusive for *G. margarita*. The expanded family of multicopper oxidases of the
324 AA1 class (OG0000106) gathers two genes (g3919 and g11237) that are strongly up-regulated in
325 the B+ line during the mycorrhizal symbiosis, and are also putatively secreted (Supplemental
326 information Table S3). These extracellular oxidases are particularly abundant in plant-interacting
327 Basidiomycota (Kües and Rühl, 2011), where they mediate plant cell wall depolymerazation. We
328 speculate that multicopper oxidases might have a role during the interaction of *G. margarita* with
329 the cell wall of its plant host, and hypothesize that the endobacterial presence may have shaped this
330 symbiotic interplay. Finally g11471, present in the expanded gene family of immunoreactive
331 mannoproteins (OG0000217) has the highest up-regulation in response to the endobacterial presence
332 during the germination phase (Supporting information Table S9).
333
334 In conclusion, the genomes of Gigasporaceae have diversified not only from those of the
335 Glomerales (Mondo et al., 2017), but also from that of *D. epigaea* (Sun et al., 2019). In addition, *G.*
336 *margarita* shows specific evolution of its gene repertoire when compared with the related *G. rosea*:
337 since the sequenced *G. rosea* isolate does not host bacteria, such events could be related to the
338 presence of the endobacterium *Ca*Gg.
339
340 As the other AMF, whose obligate biotrophy has been shaped by lack of specific pathways
341 (Tisserant et al., 2013), *G. margarita* genome does not possess invertase as well as genes involved
342 in fatty acid and thiamine biosynthesis (Supplementary text). The genetic basis for the
343 unculturability in fungi may also involve the lack of other genes, for example those for spermidine
344 and biotin biosynthesis (Ahrendt et al., 2018). This does not seem to be the case of *G. margarita*,

345 since its genome encodes for spermidine and biotin synthases (g1668, g3373 and g11043), all
346 apparently functional on the basis of their consistent expression throughout the life cycle
347 (Supporting information Table S6). However, a closer inspection revealed that these sequences are
348 all intersected by TEs: the upstream genomic region of both spermidine synthases contain
349 fragments of TIR transposons, while biotin synthase carries the exonic insertion of an unclassified
350 repeat. We conclude that these insertion events did not disrupt gene functionality in *G. margarita*.
351
352 **Not only phosphate transporter genes: *G. margarita* has a large number of phosphate-related**
353 **genes**
354 The iconic function of AMF is characterized by their capacity to take up phosphate from the soil
355 and transfer it to plants through their phosphate transporters (PTs) (Smith and Read, 2008; Ezawa
356 and Saito, 2018). Surprisingly, however, the molecular characterization of fungal PTs is very
357 limited (Harrison and van Buuren, 1995; Xie et al., 2016). Mining the *G. margarita* genome
358 revealed a very rich genetic machinery related to phosphate metabolism, sensing, transport and
359 signalling (Supporting information Table S10). We first focused on phosphate signal transduction
360 (PHO) genes since, different from metazoa where cells are regularly supplied with phosphate and,
361 as a consequence, lack a PHO network (Lev and Djordjevic, 2018), AMF preferentially thrive in
362 phosphate-deprived conditions and their capacity to successfully interact with their host plants
363 depends strictly on the P content. Under certain conditions, high Pi concentrations block symbiosis
364 establishment (Balzergue et al., 2013; Fiorilli et al., 2013). Comparison of PHO proteins belonging
365 to many fungal groups revealed that some members of the PHO cascade are absent or poorly
366 represented in the *G. margarita* genome (for instance, SPL2, PHO4 and PHO89 homologs), while
367 other components are enriched (Supporting information Table S10). *G. margarita* encodes for at
368 least 11 PT isoforms (Fig. 5, red dots), which are grouped into two distinct lineages. Accordingly,
369 nine Pi transporters, GigmPT (g21463) to GigmPT5 (g26234) and their three paralogs GigmPT8
370 (g19323) to GigmPT10 (g6532), are closely related to the fungal PHO84 PTs and clustered with

371 Glomeromycotina PTs belonging to the Mucoromycota PHO84-like subfamily. Two other PTs,
372 GigmPT6 (g10792) and GigmPT7 (g26017), which contain SPX domains, are very closely related
373 to the PHO87/90/91 PTs derived from yeast. On the basis of the classification (Fig. 5) and
374 expression (Supporting information Fig. S10, Table S11) of these *G. margarita* PTs, we speculate
375 that GigmPT1 and PT2, as well as PT4 and PT9, may contribute to Pi uptake from the environment,
376 while PT6-PT7 might be responsible for Pi homeostasis during symbiosis, similar to the SPX-Pi
377 transporters in yeast (Secco et al., 2012). The heatmap drawn on *G. margarita* PTs expression
378 (Supporting information Fig. S10) revealed that most of the transporters belonging to the
379 Glomeromycotina lineage were expressed in the extra-radical phase, confirming the largely
380 acknowledged concept that AMF uptake Pi from the soil. GigmPT (g21463) is also consistently
381 expressed during the symbiotic phase. Pho1 (g17792), encoding PHO1-type Pi transporter, is
382 heavily expressed during the intraradical stage, where it has a predicted role in Pi unloading (or
383 export) from fungus to the plant (Xie et al., 2016). These data might confirm results from laser-
384 dissected cells revealing the unexpected expression of fungal PTs in arbusculated cells (Balestrini et
385 al., 2009). Pi uptake is driven by H^+-ATPases (Ezawa and Saito, 2018), which are expressed in *G.*
386 *margarita*, not only at the fungal/substrate interface, but also during the symbiotic phase (g4412,
387 g3891; Supporting information Fig. S10). Once inside the fungus, Pi is polymerized into vacuolar
388 polyphosphate granules maybe thanks to vacuolar transporter chaperons (VTC), which are evenly
389 expressed throughout the *G. margarita* life cycle (Supporting information Table S11, Fig. S10).
390 On the basis of the genomic and transcriptomic data, we conclude that *G. margarita* possesses a
391 consistent number of functional PTs, confirming its capacity to uptake Pi from the soil; cellular Pi
392 homeostasis is probably maintained through a rather homogenous expression of two SPX-PT
393 transporters (g10792-PT6, g26017-PT7), which could re-uptake Pi from the peri-arbuscular and
394 intercellular spaces (Fig. 6, Supporting information Fig. S10). Interestingly, some ATPases and PT
395 transporters are more expressed when the fungus contains its endobacterium (Supporting
396 information Table S11); in particular the HA2 ATPase (g4412) is up-regulated by the presence of

397 the endobacterium in both germinating spores and intraradical mycelium. Furthermore, the vacuolar
398 VTC1-2 (g24765) is up-regulated by the presence of *Ca*Gg in germinating spores and GR24-treated
399 spores, suggesting a role in moving the Pi from the fungus to bacterium, which is compartmented
400 inside fungal vacuoles (Supporting information Fig. S2), and which possesses its own Pi transporter
401 (Ghignone et al., 2012; Ruiz-Lozano and Bonfante, 1999). These results are in agreement with
402 previous data showing that *G. margarita* with its endobacterium has a general more active
403 metabolism, eventually leading to a higher Pi concentration in the plant host (Salvioli et al., 2016).
404

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

407 Chitin, beta-1,3-glucans, beta-1,6-glucans and mannoproteins are usually listed as the major
408 components of fungal cell walls (Gow et al., 2017). Biochemical analyses are not available for
409 AMF. *In situ* labeling showed chitin in the cell wall in all steps of the AMF life cycle (Bonfante,
410 2018), while monoclonal antibodies failed to detect beta-1,3-glucans (Lemoine et al., 1995; Ligrone
411 et al. 2007). Indeed, homologs of FSK, the beta-1,3-glucan synthase, were not detected among the
412 protein-coding genes of the *G. margarita* genome, while mannoproteins could be biosynthesized by
413 the numerous enzymes present in the GT15 family (Supporting information Table S12) and
414 involved in mannosyltransferase activity. The GT2 family of *G. margarita* has 38 members, which
415 is high compared with the other AMF sequenced so far. Among these, 15 show characteristics of
416 chitin synthases: presence of the CON1 region (Liu et al., 2017), and domain organization leading
417 to further divisions and types (Supporting information Fig. S11). The number of secreted
418 Carbohydrate Active Enzymes (CAZymes) involved in fungal wall deconstruction is very high in
419 *G. margarita* (Supporting information Table S3) and probably guarantees the cell wall dynamics
420 described morphologically (Bonfante, 2018); in addition to many 1,2-alpha-mannosidases (GH92),
421 which seem to be specific for Gigasporaceae (Supporting information Table S12), many CAZymes

422 are devoted to chitin breakdown. Chitinases are encoded by genes present in the family GH18,
423 which are abundant in Gigasporales genomes.
424 Chitin can be deacetylated by chitin deacetylases encoded by CE4 family (Fig. 7). Glucosaminidase
425 encoded by GH20 genes might lead to production of glucosamine residues.
426 GT2, GH18, and CE4 family members, which are particularly expanded in the *G. margarita*
427 genome (Supporting information Table S12) seem to be expressed homogeneously during the *G.*
428 *margarita* life cycle (Supporting information Fig. S12, Table S13). However, some revealed
429 specific behaviour: g25383 (GH18) was particularly activated in the pre-symbiotic phase and
430 sensitive to strigolactones; expression of others was enhanced during the symbiotic phase,
431 preferentially in the presence of the endobacterium (g19206, g23868); some CE4 members, g7314,
432 g19924 and g19811, had their highest expression in the intra-radical phase.
433 Chitin has multiple roles in fungi as a structural and signalling molecule (Pusztahelyi, 2018;
434 Schmitz and Harrison, 2014). This is particularly relevant in AMF, where chitin organization is
435 modulated from the extra-radical to the intra-radical phase (Bonfante, 2018) and
436 chitooligosaccharides (COs) are the main signaling molecules (Zipfel and Oldroyd, 2017). During
437 the pre-symbiotic phase and in the extra-radical symbiotic mycelium, *G. margarita* produces COs,
438 which may or not be decorated by lipid chains. This is enabled by the GH18 family, supported by
439 the action of CBM14, which represent the chitin-binding domains of chitinases. N-Acetylated COs
440 are recognized by plant chitin receptors, (Miyata et al., 2014; Zhang et al., 2015; Carotenuto et al.,
441 2017), activating a conserved downstream symbiotic signal transduction pathway. Here, one of the
442 first signatures is a calcium spiking response, which is activated by N-acetylated COs and enhanced
443 by treatment with the synthetic strigolactone GR24 (Genre et al., 2013). However, chitin not only
444 acts as a signal for symbiosis, but also for pathogenicity (Sánchez-Vallet et al., 2015). N-Acetyl
445 COs released by AMF activate host plant defenses, inducing immunity signaling (Pozo and Azcón-
446 Aguilar, 2007; Martinez-Medina et al., 2016; Chialva et al., 2018). Many studies have demonstrated
447 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

473 discovery (Li et al., 2018). Following this method, two candidate lists were generated, representing
474 sequences putatively transferred to *G. margarita* from plants and bacteria, respectively. Among
475 genes possibly derived from plants (Supporting information Table S15), we found two sequences
476 related to fucosyltransferases. The BLASTX top hits for these sequences included *Chara braunii*, *a*
477 green alga that is considered the ancestor of AMF-hosting green plants (Delaux et al., 2015), as well
478 as *Spizellomyces punctatus*, a basally branching chytrid fungus in phylum Chytridiomycota.
479 Of the genes putatively transferred from bacteria (Supporting information Table S15), the ones
480 possessing the highest alien index were non-ribosomal peptide synthetases-polyketide synthases
481 (NRPS-PKS). Since these sequences also possess a high alien index when considering a potential
482 plant origin, they might confirm an extensive HGT from bacteria to a wide variety of eukaryotes
483 (Lawrence et al., 2011). NRPS-PKSs are involved in the biosynthesis of secondary metabolites
484 including antibiotics, toxins and siderophores. Similar sequences belong to a *Gigaspora* specific
485 expanded gene family (Supporting information Table S8a), but are absent in *R. irregularis* genome
486 (Tisserant et al., 2013). These findings, together with the limited potential of early diverging fungi
487 to biosynthesize secondary metabolites (Voigt et al., 2016), suggest that the Gigasporaceae lineage
488 might have acquired such a peculiar ability via HGT from bacteria.
489 Two gene models (g16267 and g8176) share similarity with bacterial Toll/interleukin-1
490 receptor/resistance domain-containing proteins (TIR). This domain is a protein-protein interaction
491 domain widely distributed in animals, plants and bacteria but considered absent in fungi (Ve et al.,
492 2015). In plants and animals, TIRs play roles in innate immunity, while in bacteria some of them
493 interfere with the innate immune pathways of the host (Ve et al., 2015). The TIR-like *G.margarita*
494 sequences found homologs in MREs, such as the endobacteria of *Dentiscutata heterogama* (Torres-
495 Cortés et al., 2015; Naito et al., 2015) and *D. epigaea* (Sun et al., 2019) but not in *Ca*Gg (Ghignone
496 et al., 2012). *G. margarita* BEG34 does not host MREs; furthermore TIR-like *G.margarita*
497 sequences were also expressed in the cured fungal line, ruling them out as bacterial contaminants.
498 In *G. margarita*, transcription of g8176 is strongly increased in germinating spores in the presence

499 of *Ca*Gg, and is inactivated upon GR24 treatment (Supporting information Table S16). By contrast,
500 g16267 shows uniform expression across biological conditions. According to Sun and colleagues
501 (Sun et al., 2019), we suggest that such sequences originated from MREs, which were probably
502 present in the common ancestors of Glomeromycotina and Mucoromycotina (Bonfante and Desirò,
503 2017). Finally, as Helitrons-mediated HGT has been reported in fungi (Castanera et al., 2014), we
504 searched for HGT candidates whose genomic locations overlap those of Helitrons: we found two
505 similar sequences encoding S1 peptidases (g18889 and g8238) and an amidohydrolase, both having
506 putative bacterial origin, along with an α-mannosidase (g17903) with high homology with both
507 bacteria and plants.
508 In conclusion, HGT events from plants to the *G. margarita* genome appear to be limited and
509 represent ancient events related to the algal origin of land plants (Delaux et al., 2015) and to the
510 interactions with basal Chytridiomycota fungi, which may feed on plants (Berbee et al., 2017). HGT
511 events from bacteria to AMF are multilayered; on the one hand, they reflect the interactions
512 occurring on the surface of AMF, which – in the soil – is regularly colonized by multiple bacterial
513 communities (Agnolucci et al., 2015) with different metabolic capacities. On the other, these HGT
514 events reflect the ancient presence of endobacteria, which are hosted in the *G. margarita* cytoplasm,
515 making genetic exchange between bacteria and their fungal hosts easier.
516

517 ***G. margarita* genome and the fungal immune system**

518 Since *G. margarita* interacts with viruses, bacteria and plants, we mined its genome to identify
519 genetic determinants potentially involved in the fungal immune system. Nod-like-receptors (NLRs)
520 are the main actors of the immune system in plants and animals, but are also present in fungi, in a
521 number ranging from 0 to 200 (Dyrka et al., 2014). They are involved in allorecognition processes,
522 mediating multiple processes from heterokaryon incompatibility to restriction of mycovirus transfer
523 along the hyphae, as well as xenorecognition, working during pathogen attack or symbiosis
524 establishment (Heller et al., 2018). These proteins harbour a nucleotide-binding domain of the

525 NACHT or AAA family, flanked by an N-terminal recognition domain and C-terminal repeats for
526 protein-protein interactions. In addition to the 2 TIR-like sequences (HGT events), we found 18
527 proteins in *G. margarita* where both the central domain and the terminal repeats were present, most
528 of them belonging to an expanded gene family; the central portion consisted of a NACHT domain
529 in 17 cases, while an NB-ARC domain was found in one case (g12905) (Supporting information
530 Table S16, Fig. S14). NB-ARCs are normally found in plant and animals NLR, where they are
531 associated with resistance and cell death, respectively (Van der Biezen and Jones, 1998). The C-
532 terminal of the identified proteins was always composed of pentapeptide repeats. By contrast, their
533 N-terminal recognition domain could not be functionally annotated, probably due to the lack of
534 information in public databases, or to the modular structure of these proteins, where such modules
535 could be present in separate proteins (Uehling et al., 2017). The N-terminal domains associated with
536 fungal NLRs can possess enzymatic activity such as lipase (patatins) and peptidase (subtilisin-
537 related) activity: in *G. margarita*, the expression of four patatins seemed to be sensitive to the
538 presence of the endobacterium, while three subtilisins were up-regulated upon treatment with the
539 plant signal analogue GR24 (Supporting information Table S16), suggesting their involvement in
540 plant or bacterial interaction. These patatins belong to expanded protein families in the *Gigaspora*
541 genus (OG0000175), while subtilisins (OG0000164) are expanded exclusively in *G. margarita*,
542 when compared to *G. rosea* (Supporting information Table S8b).
543 Finally, *G. margarita* possesses several genes that may be involved in defence against viruses,
544 including PIWI domain-containing proteins, a DICER-like protein and three argonaute-binding
545 proteins, all related to RNA-mediated gene silencing (Supporting information Fig. S14).
546 In conclusion, the genome of *G. margarita* reveals novel genetic determinants that might form the
547 basis of the fungal immune system, which must activate non-self recognition events throughout the
548 life cycle.
549

550

551 **Conclusions**

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

562 Similar to other AMF investigated, *G. margarita* is a strictly biotrophic fungus (Bonfante and
563 Genre, 2010), meaning that it cannot uptake complex sugars from the environment and cannot
564 biosynthesize fatty acids, being therefore dependent on its host plants. *G. margarita* and *G. rosea*
565 share an enriched set of CAZymes, which explains the capacity of *Gigaspora* to colonize different
566 plants, maybe using a combination of non-enzymatic and enzymatic mechanisms. The high number
567 of chitinase-encoding genes supports the dynamics of the fungal cell wall, which thins progressively
568 moving from the extra-radical phase to the thinner arbusculated branches. The discovery of highly
569 regulated chitin deacetylases during the symbiotic phase suggests that deacetylated chitin might
570 offer a good tool to escape from the plant immune system, avoiding the activation of plant defenses.
571 The absence of beta 1-3-glucans in the wall of the intra-radical hyphae might reflect the same
572 strategy, similar to that described for the pathogenic fungus *Colletotrichum* (Oliveira-Garcia and
573 Deising, 2013). By contrast, *G. margarita* has had stable interactions with its endobacterium *Ca*Gg
574 for a very long time (Mondo et al., 2012). The intimate contact that *G. margarita* establishes with
575 its host plants, its endobacterium and fungal-associated bacteria has provided the physical
576 possibility of HGT events. Indeed, mining the genome revealed a unique transfer event from an

577 algal ancestor towards *G. margarita*; many transfer events from a range of soil bacteria; and an
578 interesting HGT event, involving TIRs, probably transferred from Mollicutes-related endobacteria
579 detected in many AMF (Sun et al., 2019) and some African strains of *G. margarita* (Desirò et al.,
580 2014). The presence of these TIRs together with a number of Nod-like-receptors allowed us to
581 identify the first core genes of the immune system present in AMF, which probably allows these
582 fungi to safely interact with plants, bacteria and viruses.
583 Lastly, the high number of phosphate transporters first provides a rationale for the experimental
584 evidence of Pi transporters expressed during the symbiotic mycorrhizal phase, but also sheds a
585 different light on the biology of AMF, usually considered as biofertilizers. *G. margarita* has a
586 staggeringly high genome size: as for plant genomes (Pellicer et al., 2018), we suggest that this
587 peculiar trait has shaped its evolution. Plants with small genomes seem to be more widespread than
588 those with large genomes, which persist only under conditions where selective pressures are more
589 relaxed (Pellicer et al., 2018). From a mechanistic point of view, nucleic acids are amongst the most
590 nitrogen- and phosphorous-rich molecules of the cell, so under limiting nutrient N and P conditions,
591 we predict that species with large genomes, which are more demanding and costly to build and
592 maintain than species with small genomes, would be less competitive. This reasoning fits well with
593 *G. margarita* BEG34, which is commonly isolated from sand dunes (Stürmer et al., 2018), but
594 poorly represented in soils originating from many other diverse environments (Davison et al., 2015;
595 Davison et al. 2018). On the other hand, its spores are among the largest produced by AMF, a
596 feature which is commonly associated with long-term survival strategies (Aguilar-Trigueros et al.,
597 2018). In conclusion, the genome description of *G. margarita* revealed novel and unique features,
598 showing its weaknesses (a huge genome may be a limiting factor in terms of ecological success),
599 but also its biological strengths through the capacity to communicate with plants and endobacteria,
600 acting therefore as a hub of inter-kingdom interactions.
601
602

603 **Materials and methods**

604 **AMF culture, Flow cytometry, and gDNA extraction**

605 *G. margarita* strain BEG34 with (B+) and without (B-) the *Ca*Gg endobacterium was used for this
606 study. The spores were obtained as described in Salvioli et al., 2016 (19) (see Supplementary Text).
607 The genome size of *G. margarita* was first determined by flow cytometry using *Solanum*
608 *lycopersicum* as an internal reference. For each set of isolated nuclei, 100 sterilized B+ spores and
609 0.5 cm^2 of young *S. lycopersicum* leaf tissues were processed as described in Sędzielewska et al.,
610 2011 (Sędzielewska et al., 2011) and measured using a FACSAria (BD Biosciences, San Josè, CA,
611 USA) with a 488 nm laser. The absolute DNA amounts of the samples were calculated based on the
612 values of the G1 peak means. At least 10,000 nuclei per sample were analysed.
613 The gDNA was extracted from batches of 100 spores using the CTAB method. Sample purity was
614 assessed at Nanodrop1000 (Thermo Scientific, Wilmington, NC, USA). Only samples with a
615 260/230 - 260/280 > 1.8 were kept, and quantified with Qubit (Broad Range Kit, Thermo-Fisher).
616 In total, 15 µg of DNA, extracted from 16 batches of spores (~1600 spores), was sent to the
617 sequencing facility.

618

619 **DNA sequencing and genome assembly**

620 The genome of *G. margarita* was sequenced using Illumina platform. One fragment Paired-End
621 library (PE; Illumina TruSeq Nano) and two long Mate-Pair libraries (MP; Illumina Nextera; inserts
622 sizes 3 Kbp and 8 Kbp) were constructed in 2 X 150 bp format using one lane of Illumina
623 HiSeq3000 by GeT-PlaGe GenoToul (Castanet Tolosan, France). All libraries have been filtered for
624 adapter sequences and low-quality reads, using combinations of trim_galore v.0.4.1
625 (http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/; last accessed 2019-02-28),
626 trimmomatic v.0.36 (Bolger et al., 2014) and 1, cutadapt v.1.2.1, (Martin, 2011). Genome assembly
627 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

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. 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.

Competing interest: the authors declare there are no competing financial interests in relation to the work described.

727 **References**

728 Agnolucci, M., Battini, F., Cristani, C., and Giovannetti, M. (2015) Diverse bacterial communities
729 are recruited on spores of different arbuscular mycorrhizal fungal isolates. Biol Fertil Soils
730 51(3):379–89.

731 Ahrendt, S.R., Quandt, C.A., Ciobanu, D., Clum, A., Salamov, A., Andreopoulos, B., *et al.* (2018)
732 Leveraging single-cell genomics to expand the fungal tree of life. Nat Microbiol 3(12): 1417.

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

736 Balestrini, R., Gómez-Ariza, J., Klink, V.P., and Bonfante, P. (2009) Application of laser
737 microdissection to plant pathogenic and symbiotic interactions. J Plant Interact 4(2): 81–92.

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

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

743 Bianciotto, V., and Bonfante, P. (1992) Quantification of the nuclear DNA content of two
744 arbuscular mycorrhizal fungi. Mycol Res 96(12): 1071–6.

745 Bianciotto, V., Bandi, C., Minerdi, D., Sironi, M., Tichy, H.V., and Bonfante, P. (1996) An
746 obligately endosymbiotic mycorrhizal fungus itself harbors obligately intracellular bacteria. Appl
747 Environ Microbiol 62(8): 3005–10.

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

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

752 Bonfante, P., and Desirò, A. (2017) Who lives in a fungus? The diversity, origins and functions of
753 fungal endobacteria living in Mucoromycota. ISME J. 11(8): 1727–35.

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

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

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

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

764 Carlos A. Aguilar-Trigueros C.A., Hempel, S., Powell, J.R., Cornwell, W.K., and Rillig, M.C.
765 (2018) Bridging reproductive and microbial ecology: a case study in arbuscular mycorrhizal fungi.
766 ISME J 13: 873–884

767 Carotenuto, G., Chabaud, M., Miyata, K., Capozzi, M., Takeda, N., Kaku, H., *et al.* (2017) The rice
768 LysM receptor-like kinase OsCERK1 is required for the perception of short-chain chitin oligomers
769 in arbuscular mycorrhizal signaling. New Phytol 214(4): 1440–6.

770 Castanera, R., López-Varas, L., Borgognone, A., LaButti, K., Lapidus, A., Schmutz, J., *et al.* (2016)
771 Transposable elements *versus* the fungal genome: impact on whole-genome architecture and
772 transcriptional profiles. PLOS Genet 12(6): e1006108.

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

776 Chang, Y., Desirò, A., Na, H., Sandor, L., Lipzen, A., Clum, A., *et al.* (2019) Phylogenomics of
777 Endogonaceae and evolution of mycorrhizas within Mucoromycota. New Phytol 222(1): 511–25.

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

781 Chialva, M., Salvioli di Fossalunga, A., Daghino, S., Ghignone, S., Bagnaresi, P., Chiapello, M., *et*
782 *al.* (2018) Native soils with their microbiotas elicit a state of alert in tomato plants. New Phytol
783 220(4): 1296–308.

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

786 Davison, J., Moora, M., Öpik, M., Adholeya, A., Ainsaar, L., Bâ A, *et al.* (2015) Global assessment
787 of arbuscular mycorrhizal fungus diversity reveals very low endemism. Science 349(6251): 970–3.

788 Davison, J., Moora, M., Öpik, M., Ainsaar, L., Ducousso, M., Hiiesalu, I., *et al.* (2018) Microbial
789 island biogeography: isolation shapes the life history characteristics but not diversity of root-
790 symbiotic fungal communities. ISME J 12(9): 2211–24.

791 De Bie, T., Cristianini, N., Demuth, J.P., and Hahn, M.W. (2006) CAFE: a computational tool for
792 the study of gene family evolution. Bioinformatics 22(10): 1269–71.

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

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

798 Delaux, P.M., Radhakrishnan, G.V., Jayaraman, D., Cheema, J., Malbreil, M., Volkening, J.D., *et*
799 *al.* (2015) Algal ancestor of land plants was preadapted for symbiosis. Proc Natl Acad Sci USA
800 112(43): 13390–5.

801 Desirò, A., Salvioli, A., Ngonkeu, E.L., Mondo, S.J., Epis, S., Faccio, A., *et al.* (2014) Detection of
802 a novel intracellular microbiome hosted in arbuscular mycorrhizal fungi. ISME J. (2): 257–70.

803 Dyrka, W., Lamacchia, M., Durrens, P., Kobe, B., Daskalov, A., Paoletti, M., *et al.* (2014) Diversity
804 and variability of NOD-Like receptors in fungi. Genome Biol Evol 6(12): 3137–58.

805 Emms, D,M,, and Kelly, S. (2015) OrthoFinder: solving fundamental biases in whole genome
806 comparisons dramatically improves orthogroup inference accuracy. Genome Biol 16: 157.

807 Enright, A.J., Van Dongen, S., and Ouzounis, C.A. (2002) An efficient algorithm for large-scale
808 detection of protein families. Nucleic Acids Res 30(7): 1575–84.

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

811 Fiorilli, V., Lanfranco, L., and Bonfante, P. (2013) The expression of *GintPT*, the phosphate
812 transporter of *Rhizophagus irregularis*, depends on the symbiotic status and phosphate availability.
813 Planta 237(5): 1267–77.

814 Flutre, T., Duprat, E., Feuillet, C., and Quesneville, H. (2011) Considering transposable element
815 diversification in de novo annotation approaches. PLoS One 2011 6(1): e16526.

816 Genre, A., Chabaud, M., Balzergue, C., Puech-Pagès, V., Novero, M., Rey, T., *et al.* (2013) Short-
817 chain chitin oligomers from arbuscular mycorrhizal fungi trigger nuclear Ca_2^+ spiking in *Medicago*
818 *truncatula* roots and their production is enhanced by strigolactone. New Phytol 198(1): 190–202.

819 Ghignone, S., Salvioli, A., Anca, I., Lumini, E., Ortu, G., Petiti, L., *et al.* (2012) The genome of the
820 obligate endobacterium of an AM fungus reveals an interphylum network of nutritional interactions.
821 ISME J 6(1): 136–45.

822 Giovannetti, M., Mari, A., Novero, M., and Bonfante, P. (2015) Early *Lotus japonicus* root
823 transcriptomic responses to symbiotic and pathogenic fungal exudates. Front Plant Sci 6: 480.

824 Gladyshev, E., and Kleckner, N. (2017) Recombination-independent recognition of DNA homology
825 for repeat-induced point mutation. Curr Genet 63(3): 389–400.

826 Gnerre, S., MacCallum, I., Przybylski, D., Ribeiro, F.J., Burton, J.N., Walker, B.J., *et al.* (2011)
827 High-quality draft assemblies of mammalian genomes from massively parallel sequence data. Proc
828 Natl Acad Sci USA 108(4): 1513–8.

829 Götz, S., García-Gómez, J.M., Terol, J., Williams, T.D., Nagaraj, S.H., Nueda, M.J., *et al.* (2008)
830 High-throughput functional annotation and data mining with the Blast2GO suite. Nucleic Acids Res
831 36(10): 3420–35.

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

834 Guillén, F., Gómez-Toribio, V., Martínez, M.J., and Martínez, A.T. (2000) Production of hydroxyl
835 radical by the synergistic action of fungal laccase and aryl alcohol oxidase. Arch Biochem Biophys
836 383(1): 142–7.

837 Hane, J.K., and Oliver, R.P. (2008) RIPCAL: a tool for alignment-based analysis of repeat-induced
838 point mutations in fungal genomic sequences. BMC Bioinformatics 9(1): 478.

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

841 Heller, J., Clavé, C., Gladieux, P., Saupe, S.J., and Glass, N.L. (2018) NLR surveillance of essential
842 SEC-9 SNARE proteins induces programmed cell death upon allorecognition in filamentous fungi.
843 Proc Natl Acad Sci USA 115(10): 2292–301.

844 Hernández-Ortega, A., Ferreira, P., and Martínez, A.T. (2012) Fungal aryl-alcohol oxidase: a
845 peroxide-producing flavoenzyme involved in lignin degradation. Appl Microbiol Biotechnol 93(4):
846 1395–410.

847 Hoff, K.J., Lange, S., Lomsadze, A., Borodovsky, M., and Stanke, M. (2016) BRAKER1:
848 unsupervised RNA-Seq-based genome annotation with GeneMark-ET and AUGUSTUS.
849 Bioinformatics 32(5): 767–9.

850 Kamanova, J., Sun, H., Lara-Tejero, M., and Galán, J.E. (2016) The salmonella effector protein
851 SopA modulates innate immune responses by targeting TRIM E3 ligase family members. PLoS Pat
852 12(4): e1005552.

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

855 Kobayashi, Y., Maeda, T., Yamaguchi, K., Kameoka, H., Tanaka, S., Ezawa, T., *et al.* (2018) The
856 genome of *Rhizophagus clarus* HR1 reveals a common genetic basis for auxotrophy among
857 arbuscular mycorrhizal fungi. BMC Genomics 19: 465

858 Krüger, M., Krüger, C., Walker, C., Stockinger, H. and Schüßler, A. (2012), Phylogenetic reference
859 data for systematics and phylotaxonomy of arbuscular mycorrhizal fungi from phylum to species
860 level. New Phytologist, 193: 970-984.

861 Kües, U. and Rühl, M. (2011) Multiple Multi-Copper Oxidase Gene Families in Basidiomycetes –
862 What for? Curr Genomics 12(2): 72–94.

863 Lackner, G., Moebius, N., Partida-Martinez, L.P., Boland, S., and Hertweck, C. (2011) Evolution of
864 an endofungal lifestyle: deductions from the *Burkholderia rhizoxinica* genome. BMC Genomics
865 12(1): 210.

866 Laetsch, D.R., and Blaxter, M.L. (2017) BlobTools: Interrogation of genome assemblies.
867 F1000Research. 6: 1287.

868 Lanfranco, L., Fiorilli, V., and Gutjahr, C. (2018) Partner communication and role of nutrients in
869 the arbuscular mycorrhizal symbiosis. New Phytol 220(4): 1031–46.

870 Lawrence, D.P., Kroken, S., Pryor, B.M., and Arnold, A.E. (2011) Interkingdom gene transfer of a
871 hybrid NPS/PKS from bacteria to filamentous ascomycota. PLoS One 6(11): e28231.

872 Lemoine, M.C., Gollotte, A., and Gianinazzi-Pearson, V. (1995) Localization of β(1-3) glucan in
873 walls of the endomycorrhizal fungi *Glomus mosseae* (Nicol. and Gerd.) Gerd. and Trappe and
874 *Acaulospora laevis* Gerd. and Trappe during colonization of host roots. New Phytol 129(1): 97–
875 105.

876 Lev, S., and Djordjevic, J.T. (2018) Why is a functional PHO pathway required by fungal
877 pathogens to disseminate within a phosphate-rich host: A paradox explained by alkaline pH-
878 simulated nutrient deprivation and expanded PHO pathway function. PLoS Pathog 14(6):
879 e1007021.

880 Li, M., Zhao, J., Tang, N., Sun, H., and Huang, J. (2018) Horizontal gene transfer from bacteria and
881 plants to the arbuscular mycorrhizal fungus *Rhizophagus irregularis*. Front Plant Sci 9:701.

882 Ligrone, R., Carafa, A., Lumini, E., Bianciotto, V., Bonfante, P., and Duckett, J.G. (2007)
883 Glomeromycotean associations in liverworts: a molecular, cellular, and taxonomic analysis. Am J
884 Bot 94(11):1756–77.

885 Lin, K., Limpens, E., Zhang, Z., Ivanov, S., Saunders, D.G.O., Mu, D., *et al.* (2014) Single nucleus
886 genome sequencing reveals high similarity among nuclei of an endomycorrhizal fungus. PLoS
887 Genet 10(1): e1004078.

888 Liu, Z., Gay, L.M., Tuveng, T.R., Agger, J.W., Westereng, B., Mathiesen, G., *et al.* (2017)
889 Structure and function of a broad-specificity chitin deacetylase from *Aspergillus nidulans* FGSC
890 A4. Sci Rep 7(1): 1746.

891 Lomsadze, A., Burns, P.D., and Borodovsky, M. (2014) Integration of mapped RNA-Seq reads into
892 automatic training of eukaryotic gene finding algorithm. Nucleic Acids Res 42(15): e119.

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

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

897 Martin, F., Kohler, A., Murat, C., Balestrini, R., Coutinho, P.M., Jaillon, O., *et al.* (2010) Périgord
898 black truffle genome uncovers evolutionary origins and mechanisms of symbiosis. Nature
899 464(7291):1033–8.

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

902 Martinez-Medina, A., Flors, V., Heil, M., Mauch-Mani, B., Pieterse, C.M.J., Pozo, M.J., *et al.*
903 (2016) Recognizing plant defense priming. Trends Plant Sci 21(10): 818–22.

904 McTaggart, A.R., Shuey, L.S., Granados, G.M., du Plessis, E., Fraser, S., Barnes, I., *et al*. (2018)
905 Evidence that *Austropuccinia psidii* may complete its sexual life cycle on Myrtaceae. Plant Pathol
906 67(3): 729–34.

907 Miyata, K., Kozaki, T., Kouzai, Y., Ozawa, K., Ishii, K., Asamizu, E., *et al.* (2014) The bifunctional
908 plant receptor, OsCERK1, regulates both chitin-triggered immunity and arbuscular mycorrhizal
909 symbiosis in rice. Plant Cell Physiol 55(11): 1864–72.

910 Mondo, S.J., Lastovetsky, O.A., Gaspar, M.L., Schwardt, N.H., Barber, C.C., Riley, R., *et al.*
911 (2017) Bacterial endosymbionts influence host sexuality and reveal reproductive genes of early
912 divergent fungi. Nat Comm 8(1): 1843.

913 Mondo, S.J., Toomer, K.H., Morton, J.B., Lekberg, Y., and Pawlowska, T.E. (2012) Evolutionary
914 stability in a 400-Million-year-old heritable facultative mutualism. Evolution. 66(8): 2564–76.

915 Morin, E., Miyauchi, S., San Clemente, H., Chen, E.C.H., Pelin, A., de la Providencia, I., *et al.*
916 (2019) Comparative genomics of *Rhizophagus irregularis*, *R. cerebriforme*, *R. diaphanus* and
917 *Gigaspora rosea* highlights specific genetic features in Glomeromycotina. New Phytol 222(3):
918 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.

956 Sánchez-Vallet, A., Fouché, S., Fudal, I., Hartmann, F.E., Soyer, J.L., Tellier, A., *et al.* (2018) The
957 genome biology of effector gene evolution in filamentous plant pathogens. Annu Rev Phytopathol
958 56(1): 21–40.

959 Sánchez-Vallet, A., Mesters, J.R., and Thomma, B.P.H.J. (2015) The battle for chitin recognition in
960 plant-microbe interactions. FEMS Microbiol Rev 39(2): 171–183.

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

963 Secco, D., Wang, C., Shou, H., and Whelan, J. (2012) Phosphate homeostasis in the yeast
964 *Saccharomyces cerevisiae*, the key role of the SPX domain-containing proteins. FEBS Lett 586(4):
965 289–95.

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

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

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

972 Spatafora, J.W., Chang, Y., Benny, G.L., Lazarus, K., Smith, M.E., Berbee, M.L., *et al.* (2016) A
973 phylum-level phylogenetic classification of zygomycete fungi based on genome-scale data.
974 Mycologia 108(5): 1028–46.

975 Stanke, M., Diekhans, M., Baertsch, R., and Haussler, D. (2008) Using native and syntenically
976 mapped cDNA alignments to improve de novo gene finding. Bioinformatics 24(5): 637–44.

977 Stanke, M., Schöffmann, O., Morgenstern, B., and Waack, S. (2006) Gene prediction in eukaryotes
978 with a generalized hidden Markov model that uses hints from external sources. BMC
979 Bioinformatics 7: 62.

980 Stenmark, H. (2009) Rab GTPases as coordinators of vesicle traffic. Nat Rev Mol Cell Biol 10(8):
981 513–25.

982 Strullu-Derrien, C., Selosse, M., Kenrick, P. and Martin, F. M. (2018), The origin and evolution of mycorrhizal
983 symbioses: from palaeomycology to phylogenomics. New Phytol, 220: 1012-1030.

984 Stürmer, S.L., Bever, J.D., and Morton, J.B. (2018) Biogeography of arbuscular mycorrhizal fungi
985 (Glomeromycota): a phylogenetic perspective on species distribution patterns. Mycorrhiza 28(7):
986 587–603.

987 Sun, X., Chen, W., Ivanov, S., MacLean, A.M., Wight, H., Ramaraj, T., *et al.* (2019) Genome and
988 evolution of the arbuscular mycorrhizal fungus *Diversispora epigaea* (formerly *Glomus versiforme*)
989 and its bacterial endosymbionts. New Phytol 221(3): 1556–73.

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

992 Testa, A.C., Oliver, R.P., and Hane, J.K. (2016) OcculterCut: A Comprehensive Survey of AT-Rich
993 Regions in Fungal Genomes. Genome Biol Evol 8(6): 2044–64.

994 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.

997 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.

1000 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.

1004 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.

1007 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.

1009 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.

1012 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.

1014 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.

1017 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.

1020 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.

1023 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.

1025 Zipfel, C., and Oldroyd, G.E.D. (2017) Plant signalling in symbiosis and immunity. Nature 543(7645): 328–36.

1027

1028

1029

1030

1031 **Figure Legends**

1032

1033 **Figure 1**: *G. margarita* BEG34 as a metaorganism A) a *G. margarita* quiescent spore crushed to
1034 allow the spreading of nuclei and cytoplasm; B) a section of the fungal spore seen under an electron
1035 microscope: mitochondria (asterisks) and fungal nuclei (N) are highlighted; C) *Ca*Gg (arrowheads)
1036 cells are situated inside the fungal vacuole (Fv). Below, the same components are represented as
1037 metagenome scaffolds as a result of the Blobtools analysis, which integrates coverage, G/C content
1038 and sequence homology information as a method to identify and separate the different biological
1039 entities of the metagenome. Nuclear, mitochondrial and bacterial scaffolds are visualized as circles
1040 (blobs) with a diameter corresponding to their size in base pairs. The blobs are distributed on the
1041 basis of their G/C content (horizontal axis) and their coverage in terms of genomic reads (vertical
1042 axis), and colored according to their taxonomic affiliation as determined by sequence homology.
1043 The higher G/C content in mitochondrial and bacterial genomes is evident from the separation on
1044 the horizontal axis. The sequence homology search also confirmed the classification based on the
1045 other parameters.

1046

1047 **Figure 2***:* Distribution of the *G. margarita* TE families according to their main order or superfamily
1048 (Gypsy/Copia) in the consensus Library (4/264 consensus; A) and according to their genome
1049 coverage (602/060 copies/ 64% of the genome; B). PHG stands for "Potential host genes".

1050

1051 **Figure 3:** a) Comparison of the genomes of *G. margarita* and related fungi. The comparison
1052 includes the other sequenced AMF (red circles), free living Mucoromycota (blue circles), and
1053 mycorrhizal fungi from distant clades (orange circles). The size of the circles and their position on
1054 the horizontal axis is proportional to the genome sizes and the coding space (the space occupied by
1055 protein coding genes in each genome) is plotted on the vertical axis; b) a network showing the
1056 correlation of *G. margarita* with some of its relatives. In this analysis the correlation was measured

1057 as the number of orthologs between the considered species. The size of the nodes is proportional to
1058 the number of proteins in each species and the size/colour of the lines depend on the number of
1059 orthologs (the thicker/darker, the higher the number of orthologs).

1060

1061 **Figure 4:** genome-scale phylogeny as reconstructed by orthogroups-based data. The ultrametric tree
1062 shows the strong phylogenetic relationship between the two *Gigaspora* species, which diverged in
1063 recent evolutionary time (the divergence date are shown next to the branches). The divergence
1064 between the *Gigaspora* species and the other sequenced AMF groups (Diversisporaceae and
1065 Glomerales) is also shown. The leftmost pie chart displays the number of gene families that
1066 underwent rapid evolution (red for expansions and blue for contractions) after the divergence
1067 between the *Gigaspora* and *Diversispora epigaea* (Diversisporaceae). The pie charts on the right of
1068 the tree display the families that rapidly evolved in *G. margarita* or *G. rosea* after the two species
1069 diverged.

1070

1071 **Figure 5:** evolutionary relationships of phosphate transporters among fungal species. the
1072 evolutionary history was inferred using the Neighbor-Joining method 149. The tree is drawn to
1073 scale, with branch lengths in the same units as those of the evolutionary distances used to infer the
1074 phylogenetic tree. The evolutionary distances were computed using the Poisson correction method
1075 and are in the units of number of amino acid substitutions per site. The analysis involved 82 amino
1076 acid sequences. All positions containing gaps and missing data were eliminated. There was a total
1077 of 58 positions in the final dataset. Evolutionary analyses were conducted in MEGA7. The marked
1078 red circles shown in the figure indicate the multiple isoforms of phosphate transporters from *G.*
1079 *margarita.*

1080

1081 **Figure 6:** By crossing genomics and transcriptomic data the scheme illustrates an hypothetical
1082 model of how *G. margarita* handles Pi. PT1-2-4-9, supported by ATPases, may take up phosphate

1083 from the soil; the phosphate is aggregated in polyphosphate granules and released at the interfaces
1084 with the plant and with the bacterium. The insets reveal the differences between the two contact
1085 areas: at the plant interface, the fungus is surrounded by the perifungal membrane, which is
1086 continuous with the plant plasma membrane. Here the fungal Pho1 could mediate Pi export from the
1087 intraradical mycelium into the periarbuscular space. Due to the plant PT4 activity, Pi is loaded into
1088 plant cells. Pi diffusion may also occur across the perifungal membrane. In addition, PT, PT3 and
1089 PT6 could also contribute to Pi re-uptake in the intraradical mycelium under low Pi levels. The
1090 bacterium is located inside the fungal vacuole: several fungal vacuolar export chaperones are
1091 actively expressed, and the one that mostly seemed to respond to the *Ca*Gg presence is VTC1-2.
1092 The endobacterium in turn has two PTs located at the outer membrane (PstA, PstC), and one located
1093 in the periplasmic space (PtsS). Furthermore, the bacterium has cytoplasmic components which
1094 may respond to intracellular Pi (PstB and PhoU). All the membrane localisations of the fungal P
1095 transporters are hypothetical.
1096
1097 **Figure 7:** *G. margarita* has two pathways for chitin breakdown. On one side, the activity of
1098 chitinases leads to the formation of chitoligosaccharides, which are recognized by plant chitin
1099 receptors.
1100
1101 **Table 1**: assembly and annotation statistics of the *G. margarita* BEG34 genome.
1102
1103 In addition to these files we provide:
1104 16 supplemental tables (individual excel workbooks)
1105 16 supplemental figures (embedded as separate pages in a single PDF file)
1106 1 Supplementary text file (a word document containing supplementary Materials and Methods and
1107 Results)