Diversification slowdown in an evolutionary cul-de-sac

Benoît Perez-Lamarque [1,2 *], Maarja Öpik [3], Odile Maliet [1], Ana C. Afonso Silva [1], Marc-André Selosse [2,4], Florent Martos [2], and Hélène Morlon [1]

[1] *Institut de biologie de l'École normale supérieure (IBENS), École normale supérieure, CNRS, INSERM, Université PSL, 46 rue d'Ulm, 75 005 Paris, France*

[2] *Institut de Systématique, Évolution, Biodiversité (ISYEB), Muséum national d'histoire naturelle, CNRS, Sorbonne Université, EPHE, UA, CP39, 57 rue Cuvier 75 005 Paris, France*

[3] *University of Tartu, 40 Lai Street, 51 005 Tartu, Estonia*

[4] *Department of Plant Taxonomy and Nature Conservation, University of Gdansk, Wita Stwosza 59, 80-308 Gdansk, Poland*

* *corresponding author: benoit.perez@ens.psl.eu; ORCID:* 0000-0001-7112-7197

Abstract (150 words):

Arbuscular mycorrhizal fungi (AMF) are widespread microfungi that provide mineral nutrients to most land plants and form one of the oldest terrestrial symbioses. They have often been referred to as an "evolutionary cul-de-sac" for their limited ecological and species diversity. Here we use a global database of AMF to analyze their diversification dynamics in the past 500 million years (Myr). We find that AMF have low diversification rates. After a diversification peak around 150 Myr ago, they experienced an important diversification slowdown in the last 100 Myr, likely related to a shrinking of their suitable ecological niches. Our results clarify patterns and drivers of diversification in a group of obligate symbionts of major ecological and evolutionary importance. They also highlight a striking example of a diversification slowdown that, instead of reflecting an adaptive radiation as typically assumed, results from a limited ability to colonize new niches in an evolutionary cul-de-sac.

Introduction (500 words):

Arbuscular mycorrhizal fungi (AMF - subphylum Glomeromycotina) are obligate symbionts often referred to as an "evolutionary cul-de-sac, albeit an enormously successful one" [1]. They have limited morphological and species diversities, yet live in the soil and in the roots of >80% of land plants, where they provide mineral resources in exchange for plant-assimilated carbon supply [2]. Present in all terrestrial ecosystems at most latitudes, AMF play key roles in plant protection, nutrient cycles, and ecosystem functions [3]. Fossil evidence and molecular phylogenies suggest that AMF contributed to the emergence of land plants [4–6] and coevolved with their plant partners for more than 400 million years (Myr) [7–9].

Despite the ecological ubiquity and evolutionary importance of AMF, large-scale patterns of their evolutionary history are poorly known. Studies on the diversification history of AMF have been hampered by the difficulty of delineating species, quantifying global scale species richness, and building a robust phylogenetic tree for this group. Indeed, AMF are microscopic soil- and root-dwelling fungi that are poorly differentiated morphologically. Although their classical taxonomy is mostly based on the characters of spores and root colonization [2,10], AMF species delineation has greatly benefited from molecular data [11]. Experts have defined "virtual taxa" (VT) based on 97% similarity of a small region of the 18S small subunit (SSU) rRNA gene and monophyly criteria [12,13]. These delineations are controversial though, and a general, consensual system of AMF classification is still lacking [14]. AMF are also poorly known genetically: the full SSU rRNA gene sequence is known in few species [15], other gene sequences in even fewer [8,16], and complete genomes in very few [17].

The drivers of AMF diversification are unclear. A previous dated phylogenetic tree of AMF VT found that most origination times occurred after the last major continental reconfiguration around 100 Myr ago [18], suggesting that AMF diversification is not linked to vicariant speciations during this geological event. Still, geographical speciation could play an important role in AMF diversification, as these organisms have spores that disperse

59 efficiently over long distances [19–21], which could result in frequent founder-event speciation
60 [22]. Other abiotic factors include habitat: it has for example been suggested that tropical
61 grasslands are diversification hotspots for AMF [23]. Besides abiotic factors, AMF are obligate
62 symbionts and, although current AMF species are relatively generalist [3,24,25], their
63 evolutionary history could be largely influenced by a diffuse coevolution with their host
64 plants [8,26,27]. In the last 400 Myr, land plants have experienced massive radiations [27,28],
65 adapted to various ecosystems [29,30], and evolved associations with different soil
66 microorganisms [31,32]. All these factors could have influenced diversification dynamics in
67 AMF.

68 Here, we used MaarjAM, the largest global database of AMF SSU rRNA gene
69 sequences [12], to reconstruct several thoroughly sampled phylogenetic trees of AMF,
70 considering uncertainty in species delineations and phylogenetic reconstructions. We
71 combined this phylogenetic data with paleoenvironmental data and data of current AMF
72 geographic distributions, ecological traits, interaction with host plants, and genetic
73 diversity to investigate the global patterns and drivers of AMF diversification in the last
74 500 million years.

75 ***Results:***

76

77 We began by considering several ways to delineate AMF species. In addition to the
78 VT species proxy, we delineated AMF *de novo* into evolutionary units (EUs) from 36,411
79 sequences of the 520 base pair SSU rRNA gene assembled from MaarjAM [12], using a
80 monophyly criterion and 5 different thresholds of sequence similarity ranging from 97 to
81 99% (see Methods; Fig. 1). The EU97.5 and EU98 delineations (obtained using a threshold
82 of 97.5% and 98% respectively) provided a number of species comparable to the 384
83 currently recognized VT, while the EU97 delineation had much less (182). Conversely, the
84 EU98.5 and EU99 delineations yielded a much larger number of species (1,190 and 2,647,
85 respectively; Supplementary Tables 1 & 2) that was consistent with the number we
86 obtained using coalescent-based species delineation techniques (see Methods, [33,34],
87 Supplementary Tables 2 & 3). This confirms that the actual level of genetic variation within
88 the SSU marker is sufficient to separate AMF species (GMYC LRT: $P<0.05$; [35]), but supports
89 the idea that some VT might lump together several cryptic species that require finer
90 delimitations [14] (Supplementary Note 1). Rarefaction curves as well as Bayesian and Chao2
91 estimates of diversity suggested that more than 90% of the total diversity of AMF is
92 represented in our database regardless of the delineation threshold (Fig. 1b,
93 Supplementary Tables 3, 4, & 5; Supplementary Note 2), which is consistent with the
94 proportion of new species detected in recent studies [36].

95

96 We reconstructed several Bayesian phylogenetic trees of AMF based on VT and EU
97 species delineations (Fig. 2, Supplementary Figs. 1 & 2). Finer delineations resulted in an
98 expected increase in the number of nodes close to the present, but topologies and
99 branching times of internal nodes across our various AMF trees were otherwise rather
100 robust. The branching of the main AMF orders was not resolved, with node supports
101 similar to those of previous studies ([15,18] but see [11]; Supplementary Note 3), confirming that
102 additional genomic evidence is required to reach consensus. Our branching times were
103 overall older than those of Davison *et al.* [18], likely because of a different branching process

prior (see Methods), and suggested that many originations occurred during the period of major continental reconfiguration more than 100 Myr ago. Lineage through time plots (LTTs) indicated a slowdown in the accumulation of new lineages close to the present, even with the finest species delineations (Supplementary Fig. 3).

We estimated lineage-specific speciation rates of AMF with ClaDS, a recently developed Bayesian diversification model accounting for small rate shifts (Maliet *et al.*, 2019). AMF speciation rates ranged from 0.005 to 0.03 events per lineage per million years (Fig. 2; Supplementary Fig. 4), and varied both within and among AMF orders, with Glomerales and Diversisporales having the highest present-day speciation rates (Supplementary Fig. 5). AMF experienced their most rapid diversification between 200 and 100 Myr ago according to estimates of diversification rates through time obtained with ClaDS (Fig. 2; Supplementary Fig. 6), and 150-50 Myr ago according to diversification models with piecewise constant rates (TreePar [38] and CoMET [39], Fig. 2; Supplementary Figs. 7 & 8). This peak of diversification is mainly due to a fast diversification in the Glomeraceae (Fig. 2), and is concomitant with the radiation of flowering plants [26]. It also corresponds to the major continental reconfiguration including the breakdown of Pangea and the formation of new heterogeneous landmasses, which suggests that geographical speciation might have played a more important role than previously supposed [18]. Finally, this period was characterized by a warm and constant climate potentially favorable to AMF diversification. Interestingly, a similar peak of diversification at this period was also found in the Agaricomycetes fungi forming ectomycorrhiza [40].

The fast diversification of AMF around 150 Myr ago was followed by a slowdown in the recent past (Fig. 2; Supplementary Fig. 6), as suggested by the plateauing of the LTTs (Supplementary Fig. 3). A global diminution of the speciation rates through time was supported by ClaDS (Supplementary Fig. 9), TreePar and CoMET analyses (Supplementary Figs. 7 & 8), as well as time-dependent models in RPANDA [41] (Supplementary Fig. 10). Signals of diversification slowdowns sometimes result from

133 methodological artifacts such as incorrect species delineations or under-sampling [42];
134 however the slowdown was robust to all species delineations (Supplementary Figs. 6, 7, 8,
135 & 10), the branching process prior (analyses not shown), phylogenetic uncertainty, and
136 sampling fractions down to 50%, except in the case of ClaDS analyses where the trend
137 disappeared in some EU99 phylogenies and for sampling fractions lower than 70%
138 (Supplementary Figs. 11, 12, & 13).
139
140 Slowdowns in diversification rates close to the present have repeatedly been
141 observed in phylogenetic trees and have often been interpreted as a progressive reduction
142 of the number of available niches as species diversify and accumulate [42,43]. In AMF, this
143 potential effect of the saturation of niche space could be exacerbated by a reduction of the
144 number of niches suitable for AMF, linked to both repetitive breakdowns of their symbiosis
145 with plants and climatic changes. Indeed, in the last 100 Myr, many plant lineages evolved
146 root symbioses with other fungal or bacterial groups or became non-symbiotic [30,31,44], to the
147 point that approximately 20% of the extant plants are not interacting with AMF anymore
148 [3]. Additionally, the cooling of the Earth during the Cenozoic resulted in a reduction of the
149 surface of tropical regions [45,46], which are a reservoir of ecological niches for AMF [18,30,47]. The
150 difficulty of reconstructing past symbiotic associations prevents direct testing of the
151 hypothesis that the emergence of new root symbioses in plants led to a diversification
152 slowdown in AMF. We were, however, able to test the hypothesis that global temperature
153 changes affected diversification rates by fitting various environment-dependent models of
154 diversification [48] (Fig. 3; Supplementary Fig. 14). We found high support for temperature-
155 dependent models compared to time-dependent models for all AMF delineations,
156 sampling fractions, and crown ages (Fig. 3; Supplementary Figs. 15, 16, 17, & 18), with only
157 the exception of some EU99 replicate trees with a sampling fraction of 50% (Supplementary
158 Fig. 18). The support for temperature-dependent models disappeared when the
159 temperature curve was smoothed (Supplementary Fig. 19) or when time-dependent
160 models were simulated (Supplementary Note 4; Supplementary Fig. 20), suggesting that
161 the signal of temperature dependency found in AMF is not artifactually due to a temporal

trend. In addition, simulations suggested that heterogeneities in diversification rates do not generate an artifactual support for temperature-dependent models (Supplementary Note 5; Supplementary Fig. 21). Evidence for temperature dependency, however, decreased in some clades closer to the present (see Methods), as small trees tend to be best fit by constant rather than environment-dependent models (Supplementary Fig. 22). Such associations between temperature and diversification rates have been observed before in eukaryotes and have several potential causes [49]. Two prevailing hypotheses are the evolutionary speed hypothesis, stipulating that high temperatures entail high mutation rates and fast speciation [50], and the productivity hypothesis, stating that resources and associated ecological niches are more numerous in warm and productive environments, especially when the tropics are large [51]. The latter hypothesis is particularly relevant in the case of AMF, which have many host plant niches in the tropics and potentially less in temperate regions [52], where a higher proportion of plant species are non-mycorrhizal [53] or associated with ectomycorrhizal fungi [30,40]. Hence, the observed effect of past global temperature could reflect the shrinkage of tropical areas and the associated diminution of the relative proportion of plants relying on AMF.

Among other paleoenvironmental variables we tested (see Methods), few AMF sub-trees in the last 400 or 200 Myr also present significant support for diversification models with a positive dependency on CO_2 concentrations (Supplementary Fig. 22a-b), which reinforces the idea that for the corresponding AMF, benefits retrieved from their interactions with plants could have been amplified by high CO_2 concentrations and fostered diversification [54,55]. Conversely, we only found a significant effect of land plant fossil diversity in very few sub-trees (Supplementary Fig. 22). This indicates that variations in the tempo of AMF diversification did not systematically follow those of land plants, especially in the last 100 Myr when AMF diversification declined while flowering plants radiated, perhaps due to the rise of AMF-free groups such as the species-rich Orchidaceae (Supplementary Fig. 23 [28,56,57]).

191 We found little evidence for species extinction in AMF, including at mass extinction
192 events such as the Cretaceous-Paleogene extinction, as the turnover rate (ratio between
193 extinction and speciation rates) was generally estimated as close to zero with ClaDS
194 (Supplementary Fig. 9b), and the models including extinctions were never selected in
195 RPANDA (Supplementary Fig. 10). Similarly, the extinction rates estimated in piecewise-
196 constant models were not significantly different from 0 (Supplementary Fig. 24). We
197 recognize that these low extinction rate estimates could come from the difficulty of
198 estimating extinction from molecular phylogenies [58], one of the limitations of phylogeny-
199 based diversification analyses (see further discussions in Supplementary Note 6).
200 However, AMF are relatively widespread and generalists, and low extinction rates have
201 been predicted before based on their ecology [59].
202
203 To further investigate the potential factors driving AMF diversification, we assessed
204 the relationship between lineage-specific estimates of present-day speciation rates and
205 characteristics of each AMF species. First, we assessed the relationship between spore size
206 and speciation rate for the few (32) VT that are morphologically characterized [60]. Spore size
207 is often inversely related to dispersal capacity [61], which can either promote diversification
208 by favoring founder speciation events, or limit diversification by increasing gene flow. We
209 did not find a significant correlation (Supplementary Fig. 25), which may be explained
210 either by a weak or absent effect of spore size on speciation or by small sample size. We
211 found no correlation either between spore size and level of endemism (Supplementary Fig.
212 26), suggesting that even AMF with large spores can experience long-distance dispersal
213 [60,62]. Large spores might limit dispersal at smaller (e.g. intra-continental) scales in AMF [20,63],
214 but this does not seem to affect diversification.
215
216 Next, we assessed the relationship between speciation rate and geographic
217 characteristics of AMF species. We found no effect of latitude, regardless of the AMF
218 delineation or the minimum number of sequences per species (MCMCglmm: P>0.05), even
219 though AMF are more frequent and diverse in the tropics (Supplementary Fig. 27, [18,23,52])

220 and their speciation rates are positively affected by global temperature. We did not find
221 any effect of habitat or climatic zone either (Supplementary Fig. 28), nor therefore any
222 evidence that tropical grasslands are diversification hotspots for AMF [23]. Further work,
223 including a more thorough sampling of the distribution of AMF species across latitudes
224 and habitats, would be required to confirm these patterns and to discern whether
225 speciation events are indeed no more frequent in the tropics or, if they are, whether
226 efficient dispersal quickly redistributes the new lineages at different latitudes [23].
227
228 Given our previous results suggesting that temporal changes in the availability of
229 AMF niches influenced the diversification of the group, we tested whether AMF species
230 with larger 'niche width' tend to present higher lineage-specific speciation rates.
231 Macroorganisms often display such a positive relationship between (abiotic and/or biotic)
232 niche width and species diversification [64,65]. We characterized AMF niche width using a set
233 of 10 abiotic and biotic variables recorded for each AMF species and summarized with
234 their two principal component axes (see Methods; Fig. 4; Supplementary Fig. 29): PC1
235 indicates the propensity of a given AMF species to be vastly distributed in various
236 continents, ecosystems and/or associated with many plant species and lineages, and PC2
237 the propensity of a given AMF species to associate with few plant species on many
238 continents (Supplementary Fig. 30). Hence, PC1 reflects AMF niche width, whereas PC2
239 discriminates the width of the abiotic in comparison to the biotic niche (Fig. 4a-b;
240 Supplementary Fig. 31). We note that there are important nutritional and non-nutritional
241 functional aspects of the niche with potential evolutionary significance that we do not (and
242 cannot) account for in our characterization of AMF niche width: some AMF species mainly
243 provide mineral nutrients extracted from the soil, whereas others are specialized in
244 protecting plants from biotic or abiotic stresses, such as pathogens or droughts [66]. We found
245 a positive correlation between PC1 and lineage-specific speciation rates in the majority of
246 the VT and EU99 phylogenetic trees (Fig. 4c-d; Supplementary Fig. 32a). However, these
247 results were no longer significant when controlling for phylogenetic non-independence
248 between AMF species (Supplementary Fig. 32b), likely because a single *Glomeraceae* clade,

249 which included the abundant and widespread morphospecies *Rhizophagus irregularis* and
250 *Glomus clarum,* has both the highest speciation rates among AMF (Fig. 2a-b) and among the
251 largest niche widths (Supplementary Fig. 33).
252
253 Finally, as a first attempt at connecting AMF macroevolutionary diversification to
254 microevolutionary processes, we measured intraspecific genetic diversities across AMF
255 species. Genetic diversity measures the intraspecific variability on which selection can act,
256 potentially leading to species diversification. In AMF, intraspecific variability is an
257 important source of functional diversity [67,68]. We computed intraspecific genetic diversities
258 using Tajima's $\theta\pi$ estimator [69] for all species represented by at least 10, 15, or 20 sequences
259 (Supplementary Fig. 34). AMF genetic diversity was not correlated with spore size (PGLS:
260 $P>0.05$), latitude, habitat or climatic zone (MCMCglmm: $P>0.05$); it was however
261 significantly and positively correlated with niche width (PC1) for all AMF delineations and
262 minimal numbers of sequences per species considered, and in particular with abiotic
263 aspects of the niche (PC2) in many cases (Fig. 4e-h; Supplementary Fig. 32). Hence,
264 geographically widespread AMF species tend to be more genetically diverse, as previously
265 suggested by population genomics [68], but do not necessarily speciate faster. Along with a
266 decoupling between genetic diversity and lineage-specific speciation rate (Supplementary
267 Fig. 32), this suggests that the accumulation of genetic diversity among distant
268 subpopulations is not enough to spur AMF speciation.

Discussion (2 paragraphs)

Our findings that AMF have low speciation and extinction rates primarily constrained by the availability of suitable niches reinforce the vision of AMF as an "evolutionary cul-de-sac" [1]. We found speciation rates for AMF an order of magnitude lower than rates typically found for macro-eukaryotes [37,70], including plants [27], or other mycorrhizal fungi such as Agaricomycetes [40]. Low speciation rates in AMF may be linked to their (debated) asexual reproduction [71], to their strong dispersal capacity that homogenizes populations globally [68], or to the fact that they are non-specific obligate symbionts [72]. Regardless of the proximal cause, and contrary to Agaricomycetes for example, which present a large diversity of species, morphologies and ecologies, the niche space exploited by AMF in plant roots is rather restricted and has been conserved for more than 400 Myr.

Several of our results suggest a preponderant role of the global availability of AMF niches in modulating their diversification. We found a peak of diversification concomitant with several important factors, including continental reconfigurations, Angiosperms radiation, and warm climates, such that disentangling the role of these various factors is not an easy task. Continental reconfigurations may have played a role, yet the absence of correlation between speciation rates and spore size or geographic spread suggests that this role is limited. Geographic spread increases intraspecific genetic variability, but this does not result in speciation, potentially because long-distance dispersal in AMF counteracts the effect of continental separation on reducing gene flow. We found little support for the influence of plant fossil diversity on AMF diversification, particularly in the past 100 Myr. Still, the concordance of the peak of AMF diversification with the radiation of the Angiosperms is noteworthy, in particular in Glomeraceae that frequently interact with present-day Angiosperms [15]. The co-diversification with the plants might have been an important driver from the emergence of land plants until the Mesozoic [8,59], but less so after this period, as many plant lineages evolved different mycorrhizal interactions [31], blurring

global co-diversification patterns (Supplementary Fig. 23). Finally, we found substantial support for an effect of climate, which we attribute to its effect on the size of tropical areas and the associated relative proportion of plants relying on AMF. Hence, we hypothesize that a conjunction of the emergence of plant lineages not associated with AMF, of the reduction of tropical areas induced by climate cooling, and of the difficulty AMF have colonizing new niches led to a diversification slowdown in the past 100 Myr. Diversification slowdowns have often been interpreted as the signal of adaptive radiation [42,73], that is clades that experienced a rapid accumulation of morphological, ecological, and species diversity[74]. AMF provide a striking example of a clade with slow morphological, ecological and species diversification that features a pattern of diversification slowdown. In AMF, and potentially in many other species groups [49], the slowdown reflects a climate-driven reduction of the global availability of niches.

Material & methods (no words limit):

Virtual taxa phylogenetic reconstruction:

We built a phylogenetic tree of the 384 virtual taxa (VT, defined based on at least 97% sequence similarity and a criterion of monophyly [12]) from the corresponding representative sequences available in the MaarjAM database [12] updated in June 2019. We used the full 18S SSU rRNA gene sequences from Rimington *et al.* [15] to better align the VT sequences using MAFFT [75] (--add option). The 18S SSU rRNA gene is generally preferred to the universal fungal barcode (the ITS region) for characterizing AMF diversity [76]; in addition, the ITS has a higher substitution rate which makes it more sensitive to potential saturation, which could be problematic for investigating AMF evolutionary history from phylogenetic data over the last 500 Myr. We selected the 520 base pair central barcode of the VT aligned sequences and used BEAST2 [77] for the phylogenetic reconstruction. We recognize that phylogenies constructed from single and short sequences are limited, and we hope that studies like ours will foster efforts to obtain better genetic data with the aim of reconstructing comprehensive phylogenies of better quality. In the current state of data acquisition, short metabarcoding sequences provide for most microbial groups, including AMF, the only current possibility to analyze their diversification dynamics [18,78,79].

Prior to phylogenetic analysis, we used ModelFinder [80] to select the best substitution model and performed nested sampling analyses (NS [81] in BEAST2) to select the most accurate models among strict, relaxed log-normal, or exponential molecular clocks and coalescent, pure-birth, or birth-death branching process priors. As a result, we generated an input file using BEAUti with the following parameters: a GTR model with 4 classes of rates and invariant sites, a pure birth prior, and relaxed log-normal clock (Supplementary Table 6). The selection of a pure birth prior over a birth death prior by the NS analyses suggested that there is low support for extinction, and this differed from the phylogenetic analyses of Davison *et al.* (2015) who selected a coalescent prior (which tends to push the nodes close to the present compared to a pure birth prior).

339

340 In order to efficiently explore tree space and reduce computation time we constrained the
341 monophyly of the orders Diversisporales, Glomerales, and Paraglomerales +
342 Archaeosporales, based on preliminary analyses (not shown) and previous phylogenetic
343 reconstructions of the Glomeromycotina subphylum [15,18]. We ran BEAST2 to generate a
344 posterior distribution of ultrametric trees sampled using Markov chain Monte Carlo
345 (MCMC) with 4 independent chains each composed of 200,000,000 steps sampled every
346 20,000 generations. The convergences of the 4 chains were checked using Tracer [82]. We used
347 LogCombiner to merge the results setting a 25% burn-in and TreeAnnotator to obtain the
348 maximum clade credibility tree with median branch lengths, hereafter referred to as the
349 VT consensus tree. We also selected 12 trees equally spaced in the 4 independent chains (3
350 trees per chain) to account for phylogenetic uncertainty in the subsequent diversification
351 analyses, hereafter referred to as the VT replicate trees. Phylogenetic trees were visualized
352 using FigTree (v1.5) and potential negative branch lengths in the consensus tree were set
353 equal to 0 using R [83].

354

355 Finally, we converted relative branch lengths to absolute time by setting the crown root
356 age at 505 Myr [18], which is coherent with fossil data and previous dated molecular
357 phylogenies [8,9].

358

359 **Evolutionary units delineation and phylogenetic reconstruction:**

360 We gathered 41,989 AMF sequences from the MaarjAM database accessed in June 2019,
361 from which we filtered 36,411 sequences corresponding to an approximately 520 base pair
362 barcode region of the 18S SSU rRNA gene, mostly amplified by the primer pair NS31 -
363 AML2 [84,85] (dataset 1, Supplementary Table 7). In MaarjAM, most of these fungal sequences
364 are assigned to a Glomeromycotina order (Diversisporales, Glomerales, Paraglomerales or
365 Archaeosporales), a Glomeromycotina family, and are mapped to one of the 384 VT. We
366 performed *de novo* automated delineations into evolutionary units (EUs); we used a

367 criterium of monophyly, as incorporating phylogeny into AMF molecular delineations
368 seems to give more accurate units [35].

369 To do so, we first built the phylogenetic tree of all the sequences from dataset 1: we
370 aligned the 36,411 sequences using MAFFT with the VT representative sequences as a
371 backbone, filtered out gaps present in more than 50% of the sequences using TrimAl [86],
372 removed sequences having more than 25% of gaps, and selected the 520 base pair barcode
373 region. We obtained a total of 34,205 sequences reduced to 27,728 after removing identical
374 sequences using the GetHaplo function from the R-package *sidier* [87]. We used FastTree [88] to
375 build an uncalibrated tree of all AMF sequences constrained with the VT consensus tree
376 (option -*constraints*), that we rooted using midpoint rooting in R.

377 Next, we wrote an R algorithm, *delineate_phylotypes* available in the R-package
378 RPANDA [89], which traverses a tree from the root to the tips, at every node computes the
379 average similarity of all sequences descending from the node, and collapses the sequences
380 into a single EU if their sequence dissimilarity is lower than a given threshold [90]. To
381 compute average pairwise similarity among sequences, we used the formula for estimating
382 nucleotide diversity of Ferretti *et al.* (2012) [91] which has the advantage of accounting for
383 gaps in the alignment (Supp. Material 1). We ran this algorithm on the tree of AMF
384 sequences for 97, 97.5, 98, 98.5, and 99% average sequence similarity thresholds, resulting
385 in the delineation of EUs denoted EU97, EU97.5, EU98, EU98.5, and EU99.

386 Finally, we performed Bayesian phylogenetic reconstructions of the EUs. We
387 removed singletons (EUs represented by a unique sequence) as they mostly corresponded
388 to sequences with many gaps and/or badly placed tips in the phylogenetic tree
389 (Supplementary Figs. 35 & 36; Supplementary Note 2). We assigned to each EU a
390 representative sequence corresponding to its longest available sequence. We performed
391 the same BEAST2 Bayesian phylogenetic reconstructions as the one used to build the VT
392 trees, resulting in a dated consensus tree and 12 replicate trees for each delineation
393 threshold.

394

395 **Coalescent-based species delineation analyses:**

396 To compare the number of AMF species obtained using VT and EUs to an estimate that
397 does not depend on an arbitrary similarity threshold, we used GMYC models [33,34]. GMYC
398 estimates the time *t* in a reconstructed calibrated tree that separates species diversification
399 (Yule process – before *t*) and intraspecific differentiation (coalescent process – after *t*).
400 Because GMYC is too computationally intensive to be run on large clades (such as the
401 family Glomeraceae), we separated the AMF sequences according to family and performed
402 the analyses on smaller Glomeromycotina clades: (i) the family Claroideoglomeraceae, (ii)
403 the order Diversisporales composed of the families Acaulosporaceae, Diversisporaceae,
404 Pacisporaceae, and Gigasporaceae; and (iii) a "basal" clade composed of the families
405 Ambisporaceae, Archaeosporaceae, Paraglomeraceae, and Geosiphonaceae. For each of
406 these clades we aligned the sequences using MAFFT, filtered the alignment using TrimAl,
407 and reconstructed calibrated trees using BEAST2, following the same procedure as above.
408 Finally, we ran GMYC analyses (splits R-package [92]) on each of these trees, evaluated the
409 support of the threshold models compared to null models (in which all tips are assumed
410 to be different species) using likelihood ratio tests (LRT), and compared the number of
411 GMYC-delineated species to the number of VT and EUs.
412
413 **Estimating global AMF species diversity:**
414 In order to evaluate how thoroughly sampled our species-level AMF phylogenetic trees
415 are and to account for missing species in our diversification analyses, we estimated the
416 total number of Glomeromycotina VT and EUs and the corresponding sampling fractions
417 using two different approaches. First, we constructed rarefaction curves of the number of
418 VT or EUs as a function of the number of sequences and estimated the total number of
419 units using Chao2 (index *specpool* function, vegan R-package). Second, we used the
420 Bayesian Diversity Estimation Software (BDES [93]), which estimates the total number of
421 species by extrapolating sampled taxa abundance distributions. For every delineation, we
422 generated global-scale taxa abundance distributions and assumed that they followed either
423 a log-normal, log-Student, inverse Gaussian, or Sichel distribution [79]. For each tested
424 distribution, we ran three independent MCMC chains. Each chain included 250,000 steps

425 including a burn-in period of 100,000 steps and was complemented by non-informative
426 prior distributions from trial MCMC runs. The estimated total number of units was given
427 by the median value of the last 150,000 steps in the three MCMC chains. We selected the
428 most likely distributions among the 4 tested according to the lowest deviance information
429 criterion (DIC = -2 log(likelihood) + the number of parameters).
430
431 **Diversification analyses:**
432 All diversification analyses were performed for each delineation (VT or EUs) on the
433 consensus tree and on the 12 independent replicate trees to account for phylogenetic
434 uncertainty. We accounted for missing species in all these diversification analyses by
435 imputing sampling fractions, computed as the number of VT or EUs represented in a given
436 phylogeny divided by the corresponding BDES estimates of global AMF diversity
437 estimated from Sichel distributions (see Results, Supplementary Table 4). As the analyses
438 can be highly dependent on diversity estimates, we checked the robustness of the results
439 by replicating all diversification analyses using lower sampling fractions of 50%, 60%, 70%,
440 and 80%.
441
442 We estimated lineage-specific diversification rates using ClaDS, a Bayesian diversification
443 model that accounts for rate heterogeneity by modeling small rate shifts at speciation
444 events [37]. At each speciation event, the two daughter lineages inherit new speciation rates
445 sampled from a log-normal distribution with an expected value equal to $\log[\alpha \times \lambda]$ (where
446 λ represents the parental speciation rate) and a standard deviation σ. We considered the
447 model with constant turnover ε (i.e. constant ratio between extinction rates and speciation
448 rates; *ClaDS2*) and ran a newly developed ClaDS algorithm based on data augmentation
449 techniques that is much more efficient computationally than the original MCMC algorithm
450 and which estimates mean rates through time (https://github.com/OdileMaliet/
451 ClaDS_Julia). We ran ClaDS2 with 3 independent chains, checked the convergence of the
452 chains using a Gelman-Rubin diagnostic criterion [94], and recorded lineage-specific
453 speciation rates. We also extracted from the chains the maximum *a posteriori* (MAP)

454 average speciation rate across all fungal lineages (including extinct and unsampled
455 'augmented' lineages) back in time every 5 Myr. We also looked at the estimated
456 hyperparameters (α, σ, ε) and the value $m = \alpha \times \exp(\sigma^2/2)$, which represents the general
457 trend of the rate through time [37].
458
459 In order to confirm the temporal variation in rates found using ClaDS, we applied TreePar
460 [38], another diversification approach that does not consider rate variation across lineages,
461 but models temporal shifts in diversification rates affecting all lineages simultaneously. We
462 searched for up to ten shifts in diversification rates at every 2-million-year interval in each
463 phylogenetic tree. We estimated the number of temporal shifts in AMF diversification rates
464 using maximum likelihood inferences and likelihood ratio tests. We also used its
465 equivalent piecewise-constant model in a Bayesian framework (CoMET [39]) using the R-
466 package TESS [95]. We chose the Bayesian priors according to maximum likelihood estimates
467 from TreePar, we disallowed mass extinction events, and we ran the MCMC chains until
468 convergence (minimum effective sample size values of 500).
469
470 We also fitted a series of time-dependent and environment-dependent birth-death
471 diversification models using RPANDA [89]. For the time-dependent models, we used the
472 *fit_bd* function and considered models with constant or exponential variation of speciation
473 rates through time and null or constant extinction rates. The models with constant rates
474 correspond to the null hypothesis of clock-like speciation. The exponential variation was
475 chosen here as a convenient functional form that ensures positive rates; it can also be
476 viewed as an approximation of diversity dependence, a process that has often been
477 invoked during the radiation of clades [96]. For the environment-dependent models, we used
478 the *fit_env* function and considered an exponential dependency of the speciation rates with
479 the environmental variable (env), i.e. speciation rate = b * exp(a * env), where a and b are
480 two parameters estimated by maximum likelihood. The exponential variation was again
481 chosen here as a convenient functional form that ensures positive rates, and corresponds
482 to a simple linear regression of log-transformed rates. Environment curves were smoothed

483 using the function *smooth.spline* (R-package stats). We used the corrected Akaike
484 information criterion (AICc) to select the best-fit models, considering that a difference of 2
485 in AICc indicates that the model with the lowest AICc is better. We tested the influence of
486 temperature on the complete AMF phylogenetic trees, using estimates of past global
487 temperature relative to the average temperature of the period 1960-1990 from Royer *et al.*
488 (2004). As these temporal analyses can be sensitive to the root age calibration (fixed at 505
489 Myr [18]), we replicated them using the youngest (437 Myr) and oldest (530 Myr) crown age
490 estimates from Lutzoni *et al.* (2018). We checked that environment-dependent models are
491 not artifactually selected using simulated trees under the best-supported time-dependent
492 models (Supplementary Note 4) or simulated trees with heterogeneity in rates across
493 lineages (Supplementary Note 5). Following Clavel & Morlon [98] and Condamine *et al.* [49],
494 we tested whether the support for temperature-dependent models was due to a global
495 temporal trend or to details in the temperature curve by performing our model comparison
496 using increasingly smoothed temperature curves, using cubic splines with a range of
497 degrees of freedom (df) from 20 to 2. If support for temperature-dependent models
498 disappears when the temperature curve is smoothed, this means that support for
499 temperature dependency is not linked only to a global temporal trend.
500
501 Starting from 400 Myr ago, we also tested the influence of pCO_2 [99] and of land plant fossil
502 diversity, as these environmental data are not available for more ancient times. For these
503 analyses we sliced the phylogenies at 400 and 200 Myr ago, and applied the diversification
504 models to the sliced sub-trees larger than 50 tips. Estimates of land plant diversity were
505 obtained using all available Embryophyta fossils from the Paleobiology database
506 (https://paleobiodb.org) and using the shareholder quorum subsampling method [100]
507 implemented in the pipeline http://fossilworks.org [101] (Supplementary Note 7).
508
509 **Estimating genetic diversity:**
510 Among dataset 1, given that some interactions were characterized in disturbed ecosystems
511 (e.g. agricultural systems, urban areas, etc.) or sampled directly from soil, we only retained

512 the sequences of AMF that occurred in natural ecosystems and interacted with a plant
513 identified at the species level [12] (dataset 2, Supplementary Table 7). This filtering led to 351
514 fungal VT interacting with a total of 490 plant species [24]. For each VT or EU "species"
515 containing at least 10 sequences in dataset 2, we computed genetic diversity using Tajima's
516 estimator ($\theta\pi$) [69], which corresponds to the mean pairwise differences between sequences,
517 while accounting for the nucleotide gaps frequent in such metabarcoding datasets
518 (Supplementary Materials 1) [91]. By using a minimal number of 10 sequences, the estimates
519 of genetic diversity were either not significantly or only slightly correlated with the
520 number of sampled sequences per species (Supplementary Fig. 34). This criterion
521 considerably reduced the number of EU99 units represented in the analyses, as many EU99
522 had less than 10 sequences (Supplementary Table 8).
523
524 **Correlation with biotic and abiotic niche width:**
525 For each VT or EU species, we reported the number of continents where it has been
526 sampled, as well as the number of ecosystems, climatic zones, realms, habitats, and biomes
527 defined in MaarjAM [12]. We simultaneously reported the number of plant partners that have
528 been recorded (degree of the AMF species) and computed, for each fungal species, (i) the
529 phylogenetic diversity of these interacting partners (indicating whether a fungus is
530 interacting with closely related plant species or not; R-package *picante* [102]) and (ii) its
531 centrality in the plant-fungus bipartite network (we reported both the closeness and the
532 betweenness centrality; R-package bipartite [103]). The measure of centrality in an interaction
533 network indicates whether a given species is isolated and interacting with few specialist
534 partners (reciprocal specialization – restricted niche) versus part of the interaction core
535 (well-connected generalists interacting with generalists – large niche). For the plants, we
536 used the Plant List (http://www.theplantlist.org) to update some taxonomic assignations
537 following the Angiosperm Phylogeny Group (APG) III system. The land plant
538 phylogenetic tree of the 490 species was obtained by pruning the synthesis phylogeny [27]
539 using Phylomatic (http://phylodiversity.net/phylomatic/) and manually grafting the 37
540 missing taxa onto the Phylomatic tree as polytomies based on the current literature [24].

As most of these 10 scaled variables are highly correlated, we performed a principal component analysis (PCA), reported the contributions of the different variables to the first axes, and extracted the first two components for each fungal species that are represented by at least 10 sequences (in dataset 2). We selected the first two axes based on parallel analyses [104] using the function *parallel_analysis* from the R-package GeometricMorphometricsMix [105]. PC1 and PC2 explained more than 75% of the variance of the 10 variables (Supplementary Fig. 29). Log-transforming the count variables such as the number of plant species associated with an AMF species did not significantly change the PCA and the downstream analyses. We tested whether these PCA coordinates reflecting the biotic and abiotic characteristics of fungal species were correlated with the genetic diversities and the present-day speciation rates using both linear mixed-models (not accounting for AMF phylogeny) or MCMCglmm models [106]. For each MCMCglmm, we assumed a Gaussian residual distribution, included the fungal phylogenetic tree as a random effect, and ran the MCMC chains for 1,300,000 iterations with a burn-in of 300,000 and a thinning interval of 500. We tested that these results were robust to the minimal number of sequences per species used to compute the genetic diversity and the PCA, either 10, 15 or 20 sequences.

Testing the latitudinal effect:

To test the effect of latitude on speciation rates and genetic diversity, we associated each species with its set of latitudes and used similar MCMCglmm with an additional random effect corresponding to the fungal species. To consider the inhomogeneous sampling along the latitudinal gradient, we re-ran the model on jackknifed datasets (we re-sampled 1,000 interactions per slice of latitude of twenty degrees). Similarly, we tested the effect of climatic zone and ecosystems on the diversification rates.

Testing the effect of spore size:

569 As some VT containing sequences of morphologically characterized AMF isolates, we used
570 measures of their average spore length [60] to test whether spore size influenced AMF
571 speciation rates, by using a phylogenetic generalized least square regression (PGLS) to take
572 into account phylogenetic relatedness among VT. We carefully checked the normality of
573 the residuals of the models.
574
575 All statistical models were replicated on the different phylogenetic trees (consensus or
576 replicates) for a given delineation and we reported p-values (*P*) corresponding to two-
577 sided tests.

578 **References:**

579

580 1. Malloch, D. M. The evolution of mycorrhizae. *Can. J. Plant. Path.* **9**, 398–402 (1987).

581 2. Smith, S. E. & Read, D. J. *Mycorrhizal Symbiosis*. *Mycorrhizal Symbiosis* (Elsevier,
582 2008). doi:10.1016/B978-0-12-370526-6.X5001-6

583 3. van der Heijden, M. G. A. A., Martin, F. M., Selosse, M.-A. & Sanders, I. R.
584 Mycorrhizal ecology and evolution: the past, the present, and the future. *New*
585 *Phytol.* **205**, 1406–1423 (2015).

586 4. Selosse, M.-A. & Le Tacon, F. The land flora: a phototroph-fungus partnership?
587 *Trends Ecol. Evol.* **13**, 15–20 (1998).

588 5. Field, K. J., Pressel, S., Duckett, J. G., Rimington, W. R. & Bidartondo, M. I.
589 Symbiotic options for the conquest of land. *Trends Ecol. Evol.* **30**, 477–486 (2015).

590 6. Feijen, F. A. A., Vos, R. A., Nuytinck, J. & Merckx, V. S. F. T. Evolutionary
591 dynamics of mycorrhizal symbiosis in land plant diversification. *Sci. Rep.* **8**, 10698
592 (2018).

593 7. Simon, L., Bousquet, J., Lévesque, R. C. & Lalonde, M. Origin and diversification of
594 endomycorrhizal fungi and coincidence with vascular land plants. *Nature* **363**, 67–
595 69 (1993).

596 8. Lutzoni, F. *et al.* Contemporaneous radiations of fungi and plants linked to
597 symbiosis. *Nat. Commun.* **9**, 1–11 (2018).

598 9. Strullu-Derrien, C., Selosse, M. A., Kenrick, P. & Martin, F. M. The origin and
599 evolution of mycorrhizal symbioses: from palaeomycology to phylogenomics. *New*
600 *Phytol.* **220**, 1012–1030 (2018).

601 10. Stürmer, S. L. A history of the taxonomy and systematics of arbuscular mycorrhizal
602 fungi belonging to the phylum Glomeromycota. *Mycorrhiza* **22**, 247–258 (2012).

603 11. Krüger, M., Krüger, C., Walker, C., Stockinger, H. & Schüßler, A. Phylogenetic
604 reference data for systematics and phylotaxonomy of arbuscular mycorrhizal fungi
605 from phylum to species level. *New Phytol.* **193**, 970–984 (2012).

606 12. Öpik, M. *et al.* The online database MaarjAM reveals global and ecosystemic

607 distribution patterns in arbuscular mycorrhizal fungi (Glomeromycota). *New*
608 *Phytol.* **188**, 223–241 (2010).
609 13. Öpik, M., Davison, J., Moora, M. & Zobel, M. DNA-based detection and
610 identification of Glomeromycota: the virtual taxonomy of environmental
611 sequences. *Botany* **92**, 135–147 (2014).
612 14. Bruns, T. D., Corradi, N., Redecker, D., Taylor, J. W. & Öpik, M. Glomeromycotina:
613 what is a species and why should we care? *New Phytol.* **220**, 963–967 (2018).
614 15. Rimington, W. R. *et al.* Ancient plants with ancient fungi: liverworts associate with
615 early-diverging arbuscular mycorrhizal fungi. *Proc. R. Soc. B Biol. Sci.* **285**, 20181600
616 (2018).
617 16. James, T. Y. *et al.* Reconstructing the early evolution of Fungi using a six-gene
618 phylogeny. *Nature* **443**, 818–822 (2006).
619 17. Venice, F. *et al.* At the nexus of three kingdoms: the genome of the mycorrhizal
620 fungus Gigaspora margarita provides insights into plant, endobacterial and fungal
621 interactions. *Environ. Microbiol.* **22**, 122–141 (2020).
622 18. Davison, J. *et al.* Global assessment of arbuscular mycorrhizal fungus diversity
623 reveals very low endemism. *Science (80-.).* **349**, 970–973 (2015).
624 19. Egan, C., Li, D.-W. & Klironomos, J. Detection of arbuscular mycorrhizal fungal
625 spores in the air across different biomes and ecoregions. *Fungal Ecol.* **12**, 26–31
626 (2014).
627 20. Bueno, C. G. & Moora, M. How do arbuscular mycorrhizal fungi travel? *New*
628 *Phytol.* **222**, 645–647 (2019).
629 21. Correia, M., Heleno, R., da Silva, L. P., Costa, J. M. & Rodríguez-Echeverría, S. First
630 evidence for the joint dispersal of mycorrhizal fungi and plant diaspores by birds.
631 *New Phytol.* **222**, 1054–1060 (2019).
632 22. Templeton, A. R. The reality and importance of founder speciation in evolution.
633 *BioEssays* **30**, 470–479 (2008).
634 23. Pärtel, M. *et al.* Historical biome distribution and recent human disturbance shape
635 the diversity of arbuscular mycorrhizal fungi. *New Phytol.* **216**, 227–238 (2017).

24. Perez-Lamarque, B., Selosse, M., Öpik, M., Morlon, H. & Martos, F. Cheating in arbuscular mycorrhizal mutualism: a network and phylogenetic analysis of mycoheterotrophy. *New Phytol.* **226**, 1822–1835 (2020).
25. Sanders, I. R. Preference, specificity and cheating in the arbuscular mycorrhizal symbiosis. *Trends in Plant Science* **8**, 143–145 (2003).
26. Sauquet, H. & Magallón, S. Key questions and challenges in angiosperm macroevolution. *New Phytol.* **219**, 1170–1187 (2018).
27. Zanne, A. E. *et al.* Three keys to the radiation of angiosperms into freezing environments. *Nature* **506**, 89–92 (2014).
28. Cleal, C. J. & Cascales-Miñana, B. Composition and dynamics of the great Phanerozoic Evolutionary Floras. *Lethaia* **47**, 469–484 (2014).
29. Bredenkamp, G. J., Spada, F. & Kazmierczak, E. On the origin of northern and southern hemisphere grasslands. *Plant Ecol.* **163**, 209–229 (2002).
30. Brundrett, M. C. & Tedersoo, L. Evolutionary history of mycorrhizal symbioses and global host plant diversity. *New Phytol.* **220**, 1108–1115 (2018).
31. Werner, G. D. A. *et al.* Symbiont switching and alternative resource acquisition strategies drive mutualism breakdown. *Proc. Natl. Acad. Sci. U. S. A.* **115**, 5229–5234 (2018).
32. Werner, G. D. A., Cornwell, W. K., Sprent, J. I., Kattge, J. & Kiers, E. T. A single evolutionary innovation drives the deep evolution of symbiotic N2-fixation in angiosperms. *Nat. Commun.* **5**, 4087 (2014).
33. Fujisawa, T. & Barraclough, T. G. Delimiting species using single-locus data and the generalized mixed yule coalescent approach: A revised method and evaluation on simulated data sets. *Syst. Biol.* **62**, 707–724 (2013).
34. Pons, J. *et al.* Sequence-based species delimitation for the DNA taxonomy of undescribed insects. *Syst. Biol.* **55**, 595–609 (2006).
35. Powell, J. R., Monaghan, M. T., Öpik, M. & Rillig, M. C. Evolutionary criteria outperform operational approaches in producing ecologically relevant fungal species inventories. *Mol. Ecol.* **20**, 655–666 (2011).

36. Sepp, S. *et al.* Non-random association patterns in a plant–mycorrhizal fungal network reveal host–symbiont specificity. *Mol. Ecol.* **28**, 365–378 (2019).
37. Maliet, O., Hartig, F. & Morlon, H. A model with many small shifts for estimating species-specific diversification rates. *Nat. Ecol. Evol.* **3**, 1086–1092 (2019).
38. Stadler, T. Mammalian phylogeny reveals recent diversification rate shifts. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 6187–6192 (2011).
39. May, M. R., Höhna, S. & Moore, B. R. A Bayesian approach for detecting the impact of mass-extinction events on molecular phylogenies when rates of lineage diversification may vary. *Methods Ecol. Evol.* **7**, 947–959 (2016).
40. Varga, T. *et al.* Megaphylogeny resolves global patterns of mushroom evolution. *Nat. Ecol. Evol.* **3**, 668–678 (2019).
41. Morlon, H., Parsons, T. L. & Plotkin, J. B. Reconciling molecular phylogenies with the fossil record. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 16327–16332 (2011).
42. Moen, D. & Morlon, H. Why does diversification slow down? *Trends Ecol. Evol.* **29**, 190–197 (2014).
43. Rabosky, D. L. Ecological limits and diversification rate: Alternative paradigms to explain the variation in species richness among clades and regions. *Ecol. Lett.* **12**, 735–743 (2009).
44. Maherali, H., Oberle, B., Stevens, P. F., Cornwell, W. K. & McGlinn, D. J. Mutualism persistence and abandonment during the evolution of the mycorrhizal symbiosis. *Am. Nat.* **188**, E113–E125 (2016).
45. Meseguer, A. S. & Condamine, F. L. Ancient tropical extinctions at high latitudes contributed to the latitudinal diversity gradient. *Evolution (N. Y).* (2020). doi:10.1111/evo.13967
46. Ziegler, A. M. *et al.* Tracing the tropics across land and sea: Permian to present. *Lethaia* **36**, 227–254 (2003).
47. Read, D. J. Mycorrhizas in ecosystems. *Experientia* **47**, 376–391 (1991).
48. Condamine, F. L., Rolland, J. & Morlon, H. Macroevolutionary perspectives to environmental change. *Ecol. Lett.* **16**, 72–85 (2013).

694 49. Condamine, F. L., Rolland, J. & Morlon, H. Assessing the causes of diversification
695 slowdowns: temperature-dependent and diversity-dependent models receive
696 equivalent support. *Ecol. Lett.* **22**, 1900–1912 (2019).
697 50. Rohde, K. Latitudinal gradients in species diversity: The search for the primary
698 cause. *Oikos* **65**, 514 (1992).
699 51. Clarke, A. & Gaston, K. J. Climate, energy and diversity. *Proceedings of the Royal*
700 *Society B: Biological Sciences* **273**, 2257–2266 (2006).
701 52. Toussaint, A. *et al.* Asymmetric patterns of global diversity among plants and
702 mycorrhizal fungi. *J. Veg. Sci.* **31**, 355–366 (2020).
703 53. Bueno, C. G. *et al.* Plant mycorrhizal status, but not type, shifts with latitude and
704 elevation in Europe. *Glob. Ecol. Biogeogr.* **26**, 690–699 (2017).
705 54. Field, K. J. *et al.* Functional analysis of liverworts in dual symbiosis with
706 Glomeromycota and Mucoromycotina fungi under a simulated Palaeozoic CO 2
707 decline. *ISME J.* **10**, 1514–1526 (2016).
708 55. Humphreys, C. P. *et al.* Mutualistic mycorrhiza-like symbiosis in the most ancient
709 group of land plants. *Nat. Commun.* **1**, 103 (2010).
710 56. Givnish, T. J. *et al.* Orchid phylogenomics and multiple drivers of their
711 extraordinary diversification. *Proc. R. Soc. B Biol. Sci.* **282**, 20151553 (2015).
712 57. Ramírez-Barahona, S., Sauquet, H. & Magallón, S. The delayed and geographically
713 heterogenous diversification of flowering plant families. *Nat. Ecol. Evol.* In press
714 (accepted 3 Jun 2020) (2020). doi:10.1038/s41559-020-1241-3
715 58. Rabosky, D. L. Challenges in the estimation of extinction from molecular
716 phylogenies: A response to Beaulieu and O'Meara. *Evolution (N. Y).* **70**, 218–228
717 (2016).
718 59. Morton, J. B. Species and clones of arbuscular mycorrhizal fungi (Glomales,
719 Zygomycetes): their role in macro- and microevolutionary processes. *Mycotaxon*
720 *(USA)* **37**, 493–515 (1990).
721 60. Davison, J. *et al.* Microbial island biogeography: isolation shapes the life history
722 characteristics but not diversity of root-symbiotic fungal communities. *ISME J.* **12**,

723 2211–2224 (2018).
724 61. Nathan, R. *et al.* Mechanisms of long-distance seed dispersal. *Trends in Ecology and*
725 *Evolution* **23**, 638–647 (2008).
726 62. Kivlin, S. N. Global mycorrhizal fungal range sizes vary within and among
727 mycorrhizal guilds but are not correlated with dispersal traits. *J. Biogeogr.* jbi.13866
728 (2020). doi:10.1111/jbi.13866
729 63. Chaudhary, V. B., Nolimal, S., Sosa-Hernández, M. A., Egan, C. & Kastens, J. Trait-
730 based aerial dispersal of arbuscular mycorrhizal fungi. *New Phytol.* nph.16667
731 (2020). doi:10.1111/nph.16667
732 64. Gómez-Rodríguez, C., Baselga, A. & Wiens, J. J. Is diversification rate related to
733 climatic niche width? *Glob. Ecol. Biogeogr.* **24**, 383–395 (2015).
734 65. Jablonski, D. Biotic interactions and macroevolution: Extensions and mismatches
735 across scales and levels. *Evolution* **62**, 715–739 (2008).
736 66. Chagnon, P. L., Bradley, R. L., Maherali, H. & Klironomos, J. N. A trait-based
737 framework to understand life history of mycorrhizal fungi. *Trends in Plant Science*
738 **18**, 484–491 (2013).
739 67. Munkvold, L., Kjøller, R., Vestberg, M., Rosendahl, S. & Jakobsen, I. High
740 functional diversity within species of arbuscular mycorrhizal fungi. *New Phytol.*
741 **164**, 357–364 (2004).
742 68. Savary, R. *et al.* A population genomics approach shows widespread geographical
743 distribution of cryptic genomic forms of the symbiotic fungus Rhizophagus
744 irregularis. *ISME J.* **12**, 17–30 (2018).
745 69. Tajima, F. Evolutionary relationship of DNA sequences in finite populations.
746 *Genetics* **105**, 437–460 (1983).
747 70. Upham, N. S., Esselstyn, J. A. & Jetz, W. Inferring the mammal tree: Species-level
748 sets of phylogenies for questions in ecology, evolution, and conservation. *PLoS Biol.*
749 **17**, e3000494 (2019).
750 71. Yildirir, G., Malar C, M., Kokkoris, V. & Corradi, N. Parasexual and Sexual
751 Reproduction in Arbuscular Mycorrhizal Fungi: Room for Both. *Trends Microbiol.*

752 **28**, 517–519 (2020).
753 72. Morlon, H., Kemps, B. D., Plotkin, J. B. & Brisson, D. Explosive radiation of a
754 bacterial species group. *Evolution (N. Y).* **66**, 2577–2586 (2012).
755 73. Harmon, L. J., Schulte, J. A., Larson, A. & Losos, J. B. Tempo and mode of
756 evolutionary radiation in iguanian lizards. *Science (80-.).* **301**, 961–964 (2003).
757 74. Simpson, G. G. *The major features of evolution*. (Columbia University Press, 1953).
758 75. Katoh, K. & Standley, D. M. MAFFT Multiple sequence alignment software version
759 7: Improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
760 76. Lekberg, Y. *et al.* More bang for the buck? Can arbuscular mycorrhizal fungal
761 communities be characterized adequately alongside other fungi using general
762 fungal primers? *New Phytol.* **220**, 971–976 (2018).
763 77. Bouckaert, R. *et al.* BEAST 2: a software platform for Bayesian evolutionary
764 analysis. *PLoS Comput. Biol.* **10**, e1003537 (2014).
765 78. Louca, S. *et al.* Bacterial diversification through geological time. *Nat. Ecol. Evol.* **2**,
766 1458–1467 (2018).
767 79. Lewitus, E., Bittner, L., Malviya, S., Bowler, C. & Morlon, H. Clade-specific
768 diversification dynamics of marine diatoms since the Jurassic. *Nat. Ecol. Evol.* **2**,
769 1715–1723 (2018).
770 80. Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A. & Jermiin, L. S.
771 ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat. Methods*
772 **14**, 587–589 (2017).
773 81. Russel, P. M., Brewer, B. J., Klaere, S. & Bouckaert, R. R. Model selection and
774 parameter inference in phylogenetics using nested sampling. *Syst. Biol.* **68**, 219–233
775 (2019).
776 82. Rambaut, A., Drummond, A. J., Xie, D., Baele, G. & Suchard, M. A. Posterior
777 summarization in Bayesian phylogenetics using Tracer 1.7. *Syst. Biol.* **67**, 901–904
778 (2018).
779 83. R Core Team. R: A language and environment for statistical computing. (2019).
780 84. Simon, L., Lalonde, M. & Bruns, T. D. Specific amplification of 18S fungal

781 ribosomal genes from vesicular-arbuscular endomycorrhizal fungi colonizing roots.
782 *Appl. Environ. Microbiol.* **58**, 291–5 (1992).
783 85. Lee, J., Lee, S. & Young, J. P. W. Improved PCR primers for the detection and
784 identification of arbuscular mycorrhizal fungi. *FEMS Microbiol. Ecol.* **65**, 339–349
785 (2008).
786 86. Capella-Gutierrez, S., Silla-Martinez, J. M. & Gabaldon, T. trimAl: a tool for
787 automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*
788 **25**, 1972–1973 (2009).
789 87. Muñoz-Pajares, A. J. SIDIER: Substitution and indel distances to infer evolutionary
790 relationships. *Methods Ecol. Evol.* **4**, 1195–1200 (2013).
791 88. Price, M. N., Dehal, P. S. & Arkin, A. P. Fasttree: Computing large minimum
792 evolution trees with profiles instead of a distance matrix. *Mol. Biol. Evol.* **26**, 1641–
793 1650 (2009).
794 89. Morlon, H. *et al.* RPANDA: An R package for macroevolutionary analyses on
795 phylogenetic trees. *Methods Ecol. Evol.* **7**, 589–597 (2016).
796 90. Morlon, H. *et al.* The biogeography of putative microbial antibiotic production.
797 *PLoS One* **10**, (2015).
798 91. Ferretti, L., Raineri, E. & Ramos-Onsins, S. Neutrality tests for sequences with
799 missing data. *Genetics* **191**, 1397–1401 (2012).
800 92. Ezard, T., Fujisawa, T. & Barraclough, T. G. SPLITS: SPecies' LImits by Threshold
801 Statistics. R package (2009).
802 93. Quince, C., Curtis, T. P. & Sloan, W. T. The rational exploration of microbial
803 diversity. *ISME J.* **2**, 997–1006 (2008).
804 94. Gelman, A. & Rubin, D. B. Inference from iterative simulation using multiple
805 sequences. *Stat. Sci.* **7**, 457–472 (1992).
806 95. Höhna, S., May, M. R. & Moore, B. R. TESS: An R package for efficiently simulating
807 phylogenetic trees and performing Bayesian inference of lineage diversification
808 rates. *Bioinformatics* **32**, 789–791 (2016).
809 96. Rabosky, D. L. & Lovette, I. J. Density-dependent diversification in North

American wood warblers. *Proc. R. Soc. B Biol. Sci.* **275**, 2363–2371 (2008).

97. Royer, D. L., Berner, R. A., Montanez, I. P., Tabor, N. J. & Beerling, D. J. CO2 as a primary driver of Phanerozoic climate. *GSA Today* **14**, 4–10 (2004).

98. Clavel, J. & Morlon, H. Accelerated body size evolution during cold climatic periods in the Cenozoic. *Proc. Natl. Acad. Sci. U. S. A.* **114**, 4183–4188 (2017).

99. Foster, G. L., Royer, D. L. & Lunt, D. J. Future climate forcing potentially without precedent in the last 420 million years. *Nat. Commun.* **8**, 14845 (2017).

100. Alroy, J. Geographical, environmental and intrinsic biotic controls on Phanerozoic marine diversification. *Palaeontology* **53**, 1211–1235 (2010).

101. Rees, A. M. *et al.* Taxonomic occurrences of Embryophyta recorded in Fossilworks, the Evolution of Terrestrial Ecosystems database, the Paleobiology Database, and the Paleogeographic Atlas Project database. *Fossilworks* (2020). Available at: http://fossilworks.org.

102. Kembel, S. W. *et al.* Picante: R tools for integrating phylogenies and ecology. *Bioinformatics* **26**, 1463–1464 (2010).

103. Dormann, C. F., Gruber, B. & Fründ, J. Introducing the bipartite package: analysing ecological networks. *R News* **8**, 8–11 (2008).

104. Buja, A. & Eyuboglu, N. Remarks on parallel analysis. *Multivariate Behav. Res.* **27**, 509–540 (1992).

105. Fruciano, C. GeometricMorphometricsMix: Miscellaneous functions useful for geometric morphometrics, R package version 1.0-0. (2020).

106. Hadfield, J. D. MCMC methods for multi-response generalized linear mixed models: The MCMCglmm R package. *J. Stat. Softw.* **33**, 1–22 (2010).

107. Janzen, T. & Etienne, R. S. Inferring the role of habitat dynamics in driving diversification: evidence for a species pump in Lake Tanganyika cichlids. *bioRxiv* **11**, 1–18 (2016).

Acknowledgment:

The authors acknowledge C. Strullu-Derrien, M. Elias, D. de Vienne, J.-Y. Dubuisson, C. Quince, S.-K. Sepp, and M. Chase for helpful discussions. They also thank L. Aristide, S. Lambert, J. Clavel, I. Quintero, I. Overcast, and G. Sommeria for comments on an early version of the manuscript and David Marsh for English editing. BPL acknowledges B. Robira, F. Foutel--Rodier, F. Duchenne, E. Faure, E. Kerdoncuff, R. Petrolli, and G. Collobert for useful discussions and C. Fruciano and E. Lewitus for providing codes.

This work was supported by a doctoral fellowship from the École Normale Supérieure de Paris attributed to BPL and the École Doctorale FIRE – Programme Bettencourt. MÖ was supported by the European Regional Development Fund (Centre of Excellence EcolChange) and the Estonian Research Council (IUT20-28). Funding of the research of FM was from the Agence Nationale de la Recherche (ANR-19-CE02-0002). HM acknowledges support from the European Research Council (grant CoG-PANDA).

Author contributions:

All the authors designed the study. MÖ gathered the data and BPL performed the analyses. OM and ACAS provided some codes. BPL and HM wrote the first version of the manuscript and all authors contributed substantially to the revisions.

Competing Interests statement:

The authors declare that there is no conflict of interest.

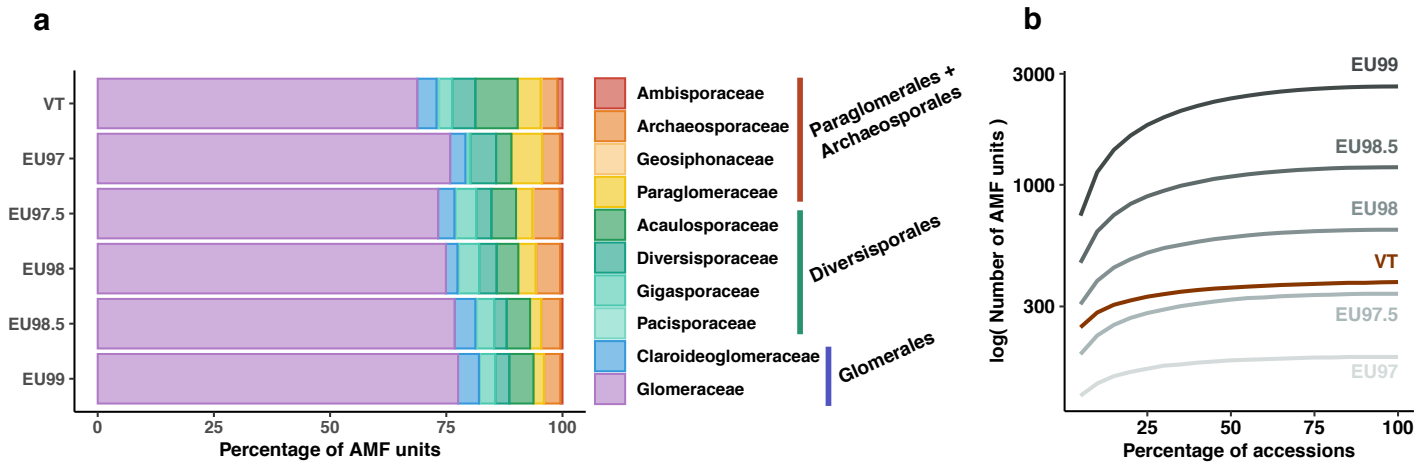
Figures

Figure 1: Molecular-based species delineations of arbuscular mycorrhizal fungi (AMF) give consistent results and indicate a nearly complete sampling.

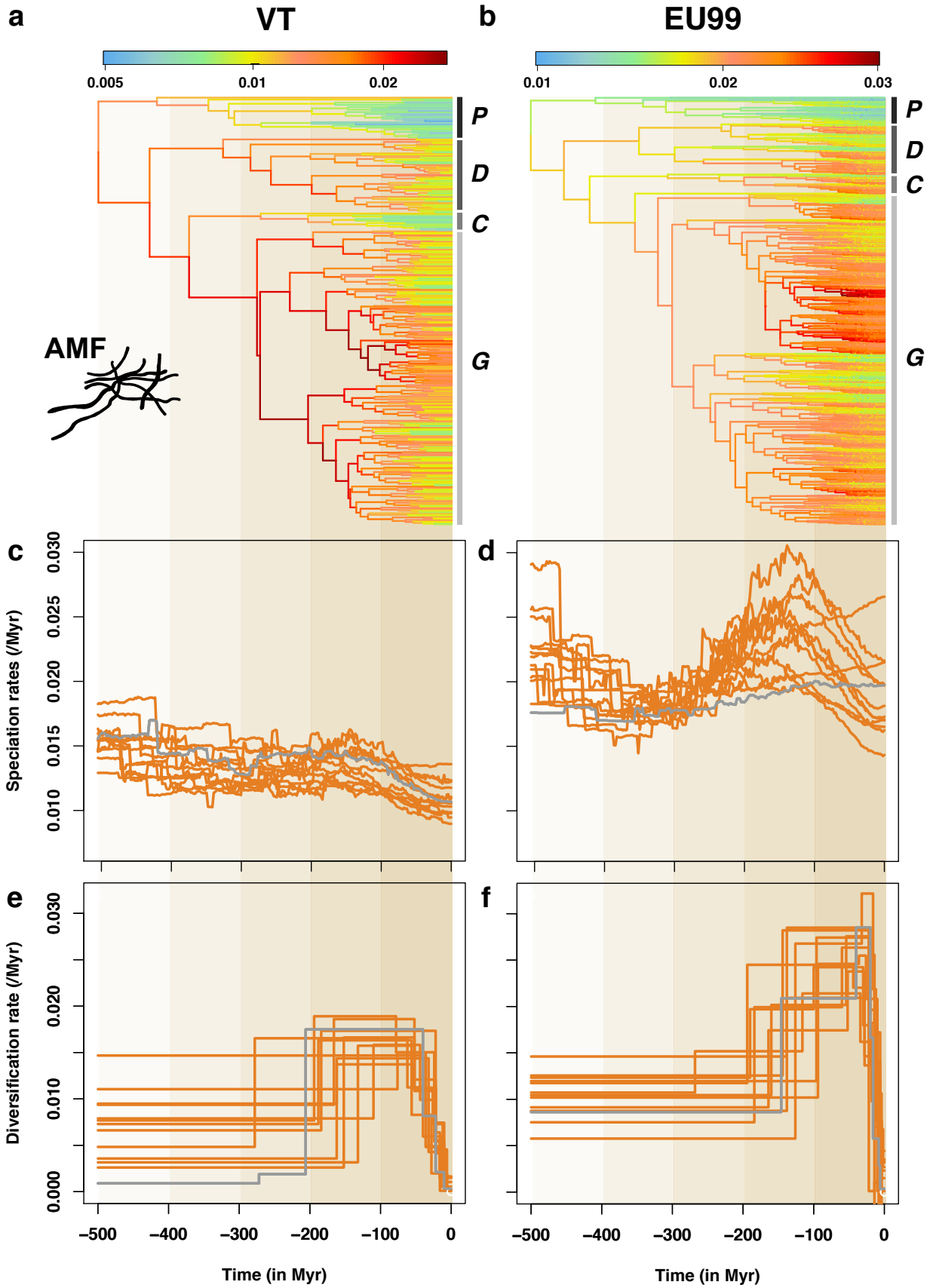
We compared the *virtual taxa* (VT) delineation from [12] with newly-developed automatic delineations into *evolutionary units* (EUs) based on an average threshold of similarity and a criterion of monophyly.

(a) The proportion of AMF units (VT or EUs) in each AMF family reveals constant proportions across delineations, although Glomeraceae tend to be relatively less numerous compared to the other AMF family in the VT delineation. The main AMF orders are indicated on the right of the charts: Paraglomerales + Archaeosporales, Diversisporales, and Glomerales (Glomeraceae + Claroideoglomeraceae).

(b) Rarefaction curves indicating the number of AMF units as a function of the percentage of sampled AMF accession revealed that the AMF sampling in MaarjAM is close to saturation for all delineations (VT or EUs). Rarefactions were performed 100 times every 5 percent and the median of the 100 replicates is represented here.



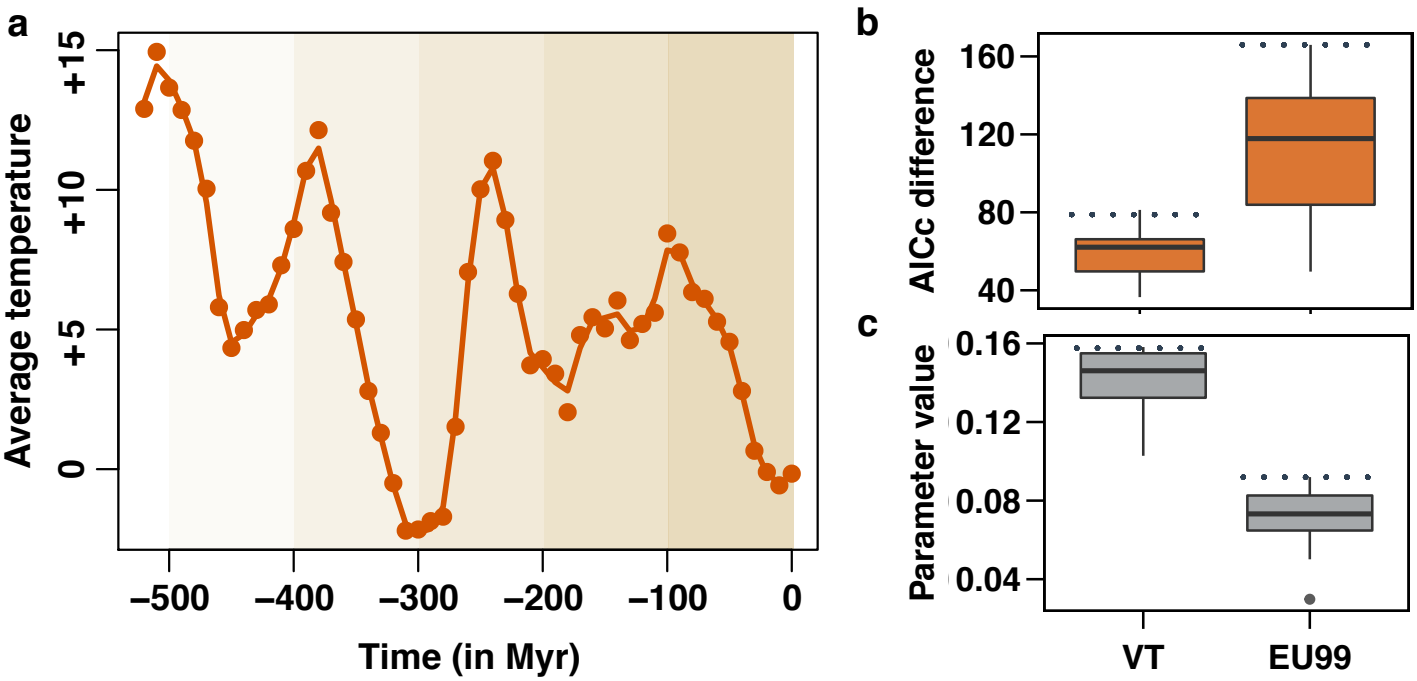
880 **Figure 2: The diversification dynamic of arbuscular mycorrhizal fungi (AMF) varies**
881 **significantly through time and between lineages.**
882
883 **(a-b):** AMF consensus phylogenetic trees corresponding to the VT (a) and EU99 (b) species
884 delineations. Branches are colored according to the lineage-specific speciation rates
885 estimated by ClaDS using the BDES estimated sampling fraction: lineages with low and
886 high speciation rates are represented in blue and red, respectively.
887 The main AMF clades are indicated with the following letters: *P* = Paraglomerales +
888 Archaeosporales, *D* = Diversisporales, *C* = Claroideoglomeraceae, and *G* = Glomeraceae.
889
890 **(c-d):** Average speciation rates (MAPS) through time estimated by ClaDS, for the VT (c)
891 and EU99 (d) delineations and using the BDES estimated sampling fraction. Orange and
892 grey lines represent the 12 independent replicate trees and the consensus tree, respectively.
893 Unlike most replicate trees, the EU99 consensus tree tends to present constant rates toward
894 the present, which reinforces the idea that consensus trees can be a misleading
895 representation [107].
896
897 **(e-f):** Net diversification rates (speciation rates minus extinction rates) through time
898 estimated by TreePar, for the VT (c) and EU99 (d) delineations and using the BDES
899 estimated sampling fraction. Orange and grey lines represent the 12 independent replicate
900 trees and the consensus tree, respectively.


a
VT
b
EU99
0.005
0.01
0.02
0.01
0.02
0.03
P
D
C
G
AMF
c
d
e
f
Speciation rates (/Myr)
Diversification rate (/Myr)
0.010
0.015
0.020
0.025
0.030
0.000
−500
−400
−300
−200
−100
0
Time (in Myr)

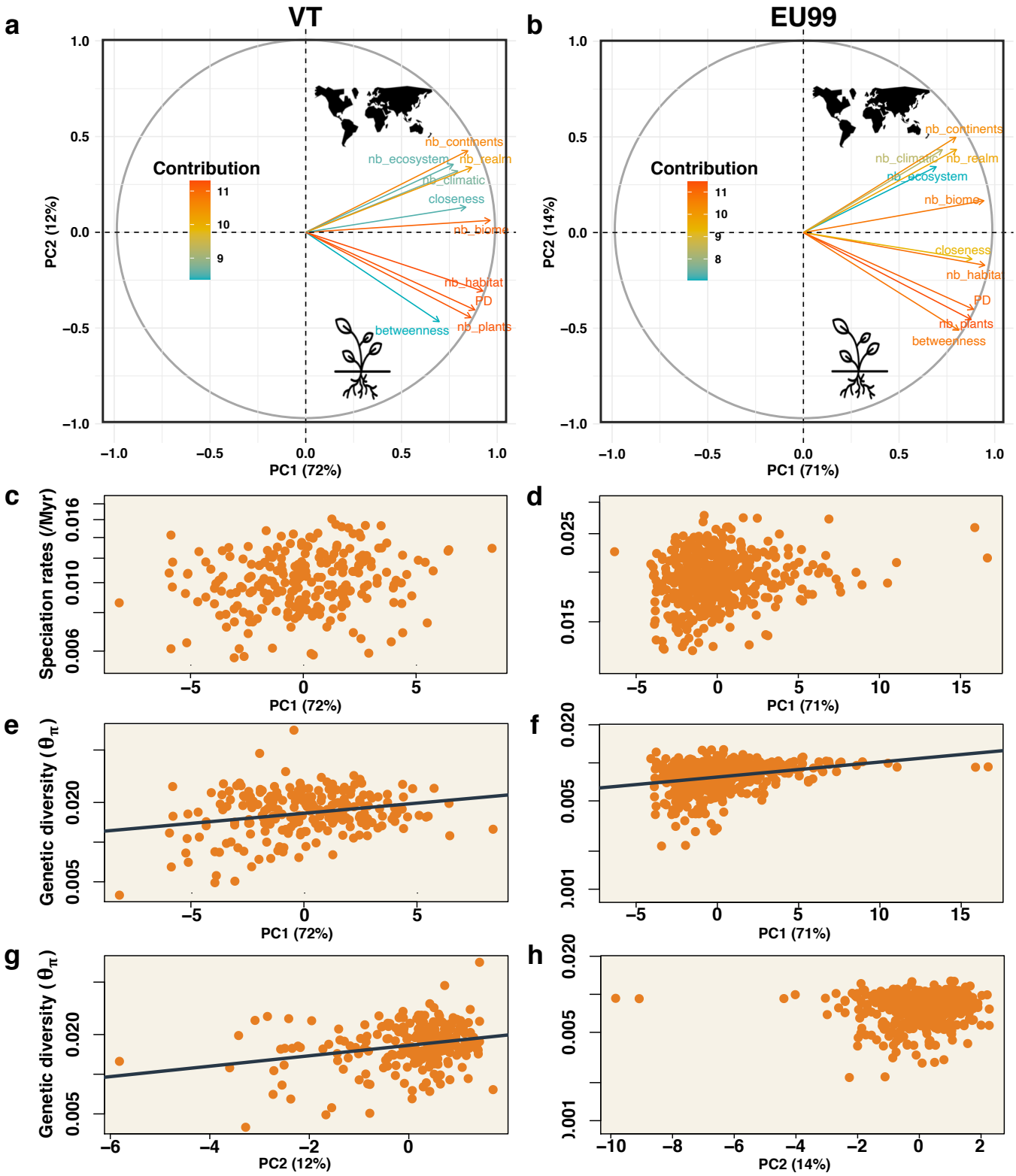
901

902 **Figure 3: Temperature-dependent diversification models reveal that global temperature**
903 **positively associates with the speciation rates of arbuscular mycorrhizal fungi (AMF) in**
904 **the last 500 million years.**
905
906 **(a):** Average global temperature in the last 500 million years (Myr) compared to the mean
907 temperature between 1960 and 1990. The smoothed orange line represents cubic splines
908 with 33 degrees of freedom used to fit temperature-dependent models of AMF
909 diversification with RPANDA. This default smoothing was estimated using the R function
910 *smooth.spline*.
911
912 **(b):** AICc difference between the best-supported time-dependent model and the
913 temperature-dependent model in RPANDA, for the VT (left) and EU99 (right) delineations,
914 using the BDES estimated sampling fraction. An AICc difference greater than 2 indicates
915 that there is significant support for the temperature-dependent model.
916
917 **(c):** Parameter estimations of the temperature-dependent models (speciation rate ~
918 exp(parameter * temperature)). A positive parameter value indicates a positive effect of
919 temperature on speciation rates.
920
921 For both delineations, the boxplots represent the results obtained for the consensus tree
922 and the 12 independent replicate trees. Boxplots indicate the median surrounded by the
923 first and third quartiles, and whiskers extend to the extreme values but no further than 1.5
924 of the inter-quartile range. The horizontal dotted lines highlighted the values estimated for
925 the consensus trees. Compared to the replicate trees, the consensus trees tend to present
926 extreme values (stronger support for temperature-dependent model), which reinforces the
927 idea that consensus trees can be a misleading representation [107].

928


a
Average temperature
0
+5
+10
+15
−500
−400
−300
−200
−100
0
Time (in Myr)
b
AICc difference
160
120
80
40
c
Parameter value
0.16
0.12
0.08
0.04
VT
EU99

929 **Figure 4: Abiotic and biotic drivers of the species diversification and differentiation of**
930 **arbuscular mycorrhizal fungi (AMF)**
931 **(a-b):** Projection of 10 abiotic and biotic variables on the two principal coordinates
932 according to the VT (a) or EU99 (b) delineations. Principal coordinate analysis (PCA) was
933 performed for the AMF units represented by at least 10 sequences. Colors represent the
934 contribution of the variable to the principal coordinates. The percentage for each principal
935 coordinate (PC) indicates its amount of explained variance.
936 Tested variables were: the numbers of continents on which the AMF unit occurs
937 (nb_continent), of realms (nb_realm), of ecosystems (nb_ecosystems), of habitats
938 (nb_habitats), of biomes (nb_biomes), and climatic zones (nb_climatic) [12], as well as
939 information about the associated plant species of each unit, such as the number of plant
940 partners (nb_plants), the phylogenetic diversity of these plants (PD), and the betweenness
941 and closeness measurement of each fungal unit in the plant-fungus interaction network
942 (see Methods).
943
944 **(c-d):** Speciation rates as a function of the PC1 coordinates for each VT (c) or EU99 (d) unit.
945 Only the AMF consensus tree is represented here (other replicate trees are presented in
946 Supplementary Fig. 32).
947
948 **(e-h):** Genetic diversity (Tajima's $\theta\pi$ estimator) as a function of the PC1 (e-f) or PC2 (g-h)
949 coordinates for each VT (e-g) or EU99 (f-h) unit. Only the AMF consensus tree is
950 represented here (other replicate trees are presented in Supplementary Fig. 32). The grey
951 lines indicate the statistically significant linear regression between the two variables
952 inferred using MCMCglmm.
953


a
VT
b
EU99
Contribution
11
10
9
8
nb_continents
nb_ecosystem
nb_realm
nb_climatic
closeness
nb_biome
nb_habitat
PD
nb_plants
betweenness
PC1 (72%)
PC2 (12%)
PC1 (71%)
PC2 (14%)
c
d
Speciation rates (/Myr)
e
f
Genetic diversity (θ_π)
g
h
PC2 (12%)
PC2 (14%)

954