Genomic insights into the evolution of secondary metabolism of *Escovopsis* and its allies, specialized fungal symbionts of fungus-farming ants

Aileen Berasategui[1,2,3,4], Hassan Salem[1,3], Abraham G. Moller[5], Yuliana Christopher[6], Quimi Vidaurre-Montoya[7], Caitlin Conn[1,8], Timothy D. Read[5], Andre Rodrigues[7], Nadine Ziemert[2,9] and Nicole Gerardo[1]

[1] Department of Biology, Emory University, Atlanta, 30322, USA.

[2]Cluster of Excellence-Controlling microbes to fight infections, University of Tübingen, 72076, Germany

[3]Mutualisms Research Group, Max Planck Institute for Biology, 72076, Tübingen, Germany.

[4]Amsterdam Institute for Life and Environment, Vrije Universiteit Amsterdam, 1081 HV Amsterdam, The Netherlands.

[5]Division of Infectious Diseases, Emory University School of Medicine, Atlanta, 30322, USA.

[6]Instituto de Investigaciones Científicas y Servicios de Alta Tecnología, Ciudad del Saber, Clayton, 0843-01103, Panamá.

[7]Department of General and Applied Biology, São Paulo State University (UNESP), Rio Claro, 13506-900, Brazil.

[8]Department of Biology, Berry College, Mount Berry, 30149, USA

[9]Translational Genome Mining for Natural Products, Interfaculty Institute of Microbiology and Infection Medicine Tübingen (IMIT), Interfaculty Institute for Biomedical Informatics (IBMI), University of Tübingen, 72076, Germany.

Abstract

The metabolic intimacy of symbiosis often demands the work of specialists. Natural products and defensive secondary metabolites can drive specificity by ensuring infection and propagation across host generations. But in contrast to bacteria, little is known about the diversity and distribution of natural product biosynthetic pathways among fungi and how they evolve to facilitate symbiosis and adaptation to their host environment. In this study, we define the secondary metabolism of *Escovopsis* and closely related genera, members of which are specialized, diverse ascomycete fungi best known as mycoparasites of the fungal cultivars grown by fungus-growing ants. We ask how the gain and loss of various biosynthetic pathways corresponds to divergent lifestyles. Long-

read sequencing allowed us to define the chromosomal features of representative *Escovopsis* strains, revealing highly reduced genomes (21.4-38.3 Mb) composed of 7-8 chromosomes. *Escovopsis* genomes are highly co-linear, with genes localizing not only in the same chromosome, but also in the same order. Macrosynteny is high within *Escovopsis* clades, and decreases with increasing phylogenetic distance, while maintaining a high degree of mesosynteny. To explore the evolutionary history of biosynthetic pathways in this group of symbionts relative to their encoding lineages, we performed an ancestral state reconstruction analysis, which revealed that, while many secondary metabolites are shared with non-ant associated sordariomycetes, 56 pathways are unique to the symbiotic genera. Reflecting adaptation to diverging ant agricultural systems, we observe that the stepwise acquisition of these pathways mirrors the ecological radiations of attine ants and the dynamic recruitment and replacement of their fungal cultivars. As different clades encode characteristic combinations of biosynthetic gene clusters, these delineating profiles provide important insights into the possible mechanisms underlying specificity between these symbionts and their hosts. Collectively, our findings shed light on the evolutionary dynamic nature of secondary metabolism in *Escovopsis* and its allies, reflecting adaptation of the symbionts to an ancient agricultural system.

1. Introduction

All parasites are specialists. At broad scales, each parasite can exploit some host species and not others. At finer scales, many parasite strains may be specialized on particular host genotypes within a species (1). While parasite host range is constrained by different evolutionary processes, including tradeoffs and arms race coevolutionary dynamics (2, 3), the molecular mechanisms underlying pathogen specialization and the evolutionary ecology of specificity have yet to be clearly linked. Similarly, little is known about the genomic architecture underlying the evolution

of parasite specialization, the genomic consequences of host shifts and the genetic basis of shifts along the parasitism to mutualism continuum that underlies most symbioses.

Fungal symbionts are genetically tractable models for the study of host fidelity due to their diverse lifestyles and the occurrence of very closely related species that differ from each other primarily in their host range (4). Secondary metabolites, small molecules that are not necessary for the growth of an organism but aid in survival, play essential roles during fungal infection (5) and are known to affect the niche breath of fungal pathogens (4, 6, 7). Typically, specialists harbor a contracted array of specialized metabolites relative to generalists (4), reflecting the metabolic constraints that they experience in attempting to exploit different hosts. Given their role in mediating species interactions, secondary metabolites are central to arms-races dynamics (8, 9). Thus, their origin and distribution can reflect adaptation to specific host environments (7).

Escovopsis (Hypocreales: Hypocreaceae) is a specialized (10–12), diverse group of fungi found in the gardens of fungus-farming ants (Hymenoptera: Attini) (13). Currently, there are nine described species (14), some of which have been well-studied for their ability to parasitize the ants' fungal cultivars (10–12). *Escovopsis* sp. can be virulent parasites of fungus-growing ant agriculture, causing garden biomass loss and colony decline (13, 15, 16). While it is presumed that most species in the group are similarly virulent, infection by certain species appear to be not as lethal, suggesting that the ecological role and evolutionary implications of these symbionts are not fully understood (17–19). In recognition of their morphological and ecological diversity, recent work has proposed splitting the *Escovopsis* genus into multiple genera (i.e., *Escovopsis, Luteomyces and Sympodiorosea*) (14). Here, we refer to all members of the group with the common name escovopsis for simplicity.

Fungus-farming ants are a monophyletic group of obligate agriculturalists (20). Attines feed their cultivated fungi ("cultivars") with vegetative material, and in turn the cultivar represents the ants' primary food source. Different attine lineages practice different modes of agriculture, exhibiting a high degree of specificity towards their cultivars (21, 22), and these different agricultural systems are generally associated with different *Escovopsis, Sympodiorosea* and *Luteomyces* spp. (23). The ancestral system, *lower agriculture*, is practiced by a group of ants that cultivate fungi in the Agaricales family. While most of the ant species in this system grow their cultivars in the form of mycelium, some ants in the lower agriculture system subsist on Agaricales that grow in yeast form, giving rise to the name of *yeast agriculture*. While *Sympodiorosea* and *Luteomyces* infections of mycelial-growing lower attine ant gardens are common, infection of yeast gardens has never been found (20). The third agricultural system is known as *coral agriculture,* in which a group of ants within the *Apterostigma* genus exploits fungus in the *Pterulaceae* family. Infection of coral gardens are also common, and include infection by *Escovopsis*, *Luteomyces* and other related taxa (10, 14). While lower attines, practicing lower, yeast and coral agriculture, are characterized by providing their cultivars with dead vegetative material, higher attines (practicing generalized higher agriculture and leaf-cutter agriculture) provide their fungal mutualists with freshly cut vegetative material (20). The two agricultural systems of higher attines are characterized by the obligate lifestyle of the cultivar, which cannot survive without association with the ants. *Generalized higher agriculture* is practiced by ants cultivating a derived clade of agaricaceous fungi, whereas in the most derived agricultural system, that of *leaf-cutter agriculture*, a single fungal species *Leucoagaricus gongylophorus* is responsible for ant survival. Higher agriculture gardens are commonly infected with *Escovopsis*, most of which are *Escovopsis* spp. closely related to the best studied species, *E. weberi* (24).

Escovopsis symbionts show a high degree of host fidelity, being able to infect some cultivars but not others. This degree of partner specificity suggests a long history of coevolution, as demonstrated by the phylogenetic congruence between attines, their cultivars and *Escovopsis*, particularly at the broad interspecific scale (10). To manage infections, ants actively weed infected portions of garden, and many attine species associate with actinomycete *Pseudonocardia* that synthesize antifungal compounds that inhibit escovopsis growth (25) .

Despite consistent patterns of co-diversification across the tripartite interaction between the ants, their cultivars and escovopsis (10), and the outsized role of natural products in mediating fungal specialization, the secondary metabolism of escovopsis remains relatively unexplored relative to the evolutionary ecology of attine ants and their cultivars. Only a few *Escovopsis*-derived compounds have been identified (26, 27), though recent genome annotation indicates the potential to produce many more (28). Here, we performed long-read genome sequencing, assembly, and annotation to describe the chromosomal architecture, conservation, and organization of escovopsis, which will facilitate future annotation of the biosynthetic machinery. After defining the secondary metabolism across the group and spanning representative host ranges, we contextualize the distribution of biosynthetic gene clusters relative to patterns of specialization and fidelity. Through comparative genomics, extensive manual curation of biosynthetic gene clusters, and ancestral state reconstruction, we outline a symbiont whose secondary metabolism broadly reflects the dynamic patterns of cultivar recruitment and replacement by attine ants.

2. Material and Methods

a. Sample collection, isolation, DNA extraction and genome sequencing.

Strains of escovopsis infecting all attine agricultural systems were obtained from the Emory collection (Table S1). To obtain DNA, fungi were grown on PDA plates at room temperature. Genomic DNA was extracted by crushing fungal tissue with liquid nitrogen and subsequently isolating the DNA using a phenol-chloroform protocol (29). Sequencing was performed on a HiSeq 2500 Sequencing system from Illumina, utilizing the paired-end 150 bp technology. Both library preparation and DNA sequencing were carried out at Novogene. Additionally, DNA from strains NGL095 *(E. weberi),* NGL070 *(E. multiformis)* and NGL057 (*Hypocreales: Hypocreaceae*, cf. *Escovopsis*) was also sequenced with PacBio technology by OmegaBioservices.

b. Genome assembly and annotation

Strains sequenced with PacBio technology were assembled with Canu v.1.8 (30) and polished with their corresponding Illumina reads using Pilon v.1.23 (31). Those strains sequenced with Illumina alone were quality checked with FastQC (32), trimmed with trimmomatic (33), and subsequently assembled with Spades v.3.13.0 (34). Genome assembly quality was evaluated using BUSCO v.3 (35). GC content was calculated with the script GC_content.pl by Damien Richard (https://github.com/DamienFr/GC_content_in_sliding_window/ [last accessed July 2023]), using default parameters. The genomic dataset was completed with the addition of six previously sequenced *Escovopsis* genomes (26), as well as 14 closely related species from the Hypocreales family obtained from JGI Mycocosm (Table S1). The highly contiguous hybrid assemblies NGL070, NGL095 and NGL057 were screened for stretches of telomeric repeats (TTAGGG)n at

144 the end of contigs, and contigs harboring these repeats at both ends were considered complete
145 chromosomes.

146 To compare genomic architecture conservation between *Escovopsis* strains, a synteny analysis was
147 performed on the proteome sets of the most unfragmented assemblies in our dataset employing
148 GENESPACE v0.9.3 (36) as implemented in R. This dataset comprised the three hybrid
149 assemblies NGL095, NGL070 and NGL057, as well as EACOL, EAECHC, EAECHR, ETCORN
150 and EATTINE.

151 All assemblies were subjected to gene prediction and annotation using the Funannotate v.1.8.3
152 (37) pipeline. Repeats were identified with RepeatModeler and soft masked using RepeatMasker.
153 Protein evidence from a UniprotKB/Swiss-Prot-curated database (38), and the proteomes from
154 *Trichoderma sp., Cladobotryum sp., Hypomyces rosellus* and *H. perniciosus* were aligned to the
155 genomes using TBlastN and Exonerate (39). Three gene prediction tools were used: AUGUSTUS
156 v3.3.3(40), snap (41) and GlimmerHMM v3.0.4 (42). tRNAs were predicted with tRNAscan-SE
157 (43). Consensus gene models were found with EvidenceModeler (44). Functional annotation was
158 conducted using BlastP to search the Uniprot/SwissProt protein database. Protein families (Pfam)
159 and Gene Ontology (GO) terms were assigned with InterProScan5 (45). Additional predictions
160 were inferred by alignments to the eggnog orthology database (46), using emapper v3 (47). The
161 secretome was predicted using Phobius v.1.01 (48), which identifies proteins carrying a signal
162 peptide. Carbohydrate active enzymes were identified using HMMER v3.3 (49) and family
163 specific HMM profiles of the dbCAN2 server (50). Proteases and protease inhibitors were
164 predicted using the MEROPS database v (51), and biosynthetic gene clusters were annotated using
165 fungiSMASH v6 (52) with relaxed parameters.

c. Phylogenetic reconstruction

The phylogenetic relationship of *Escovopsis* spp. and their close relatives was reconstructed using the BUSCO_phylogenomics pipeline (53). In short, single copy orthologues for each genome were identified by running BUSCO v5 (35) with the Ascomycota_odb10 lineage database. This analysis identified 660 single-copy orthologs shared by all 34 strains in the dataset. Gene sequences were aligned with MUSCLE (54), and the alignment was trimmed with TrimAl (55). Output alignments were concatenated into a supermatrix. A maximum likelihood tree was built with IQ-TREE (56) , allowing ModelFinder (57) to predict the best evolutionary model for partitioning the alignment. The resulting tree was rooted using *Trichoderma spp.* and visualized with iTol v6 (58).

In order to place escovopsis strains in a broader phylogenetic context, we performed a multi-locus analysis employing different molecular markers (i.e., ITS, TEF and LSU). Sequences were aligned in MAFFT v.7 (59) and a phylogenetic tree was reconstructed using maximum likelihood (ML) in RAxML (60). The GTR nucleotide substitution model was selected by independent runs in jModelTest2 (Darriba et al. 2012) using the Akaike Information Criterion (AIC) with 95% confidence intervals. 1000 independent trees and 1000 bootstrap replicates were performed and the final tree was edited in FigTree v.1.4 and Adobe Illustrator CC v.17.1.

Gene Cluster Family (GCF) identification

Biosynthetic gene clusters (BGCs) of all fungal strains were identified again using FungiSMASH 6.1 (52) with relaxed parameters, utilizing as input the GenBank files obtained after genome annotation. With the aid of cblaster v.1.3.12 (61), BGCs split onto different contigs, especially those located on contig edges, were manually assembled based on homology with other BGCs in the dataset. Likewise, fused BGCs were manually split into separate BGCs. The final BGC set was

analyzed using BiG-SCAPE v1.0.1 (62) to identify homologous BGCs across all strains and to cluster related BGCs into gene cluster families (GCFs). BGCs from the MiBIG database 2.0 (63) were included in the analysis with the –mibig flag to identify already described BGCs. The scikit-learn package was downgraded to v.0.19.1, and the following parameters were enabled: –mix, --hybrids-off, and –include_singletons. The program was run in 'glocal' alignment mode with edge-length cutoffs from 0.1 to 0.9, with step increments of 0.1. After inspection, networks at thresholds 0.5-0.6 were found to be similar and further analyses were based on a cutoff of 0.5. The resulting sequence similarity matrixes were visualized using Cytoscape v.3.9.0 (64). A presence/absence matrix was built to evaluate BGC distribution, with 1 representing presence and 0 absence of a GCF in a fungal strain and was visualized as a heatmap employing R. To assess whether BGC profiles can delineate groups of escovopsis, a Jaccard distance matrix was computed using the presence/absence table. The distance matrix was then used to construct nonmetric multidimensional scaling (NMDS) ordination plots to detect grouping patterns and subjected to a PERMANOVA analysis (permutational multivariate analysis of variance) to identify significant factors underlying observed groupings. To assess the adequacy of our sampling, and to provide an estimate of GCF richness for the given sequencing effort, rarefaction curves were built at the genus level, and at both levels of attine agricultural systems (*i.e.,* lower and higher agriculture, as well as lower, coral, general higher and leaf-cutter agriculture).

d. Co-cladogenesis analyses

To assess whether BGC profiles delineate escovopsis, the GCF presence/absence was subjected to a hierarchical clustering analysis using a correlation-centered similarity metric with the complete linkage clustering method. A tanglegram was built in R to evaluate the congruency between the symbiont phylogeny and strain BGC profiles using the package "dendextend" v 1.17.1.

e. Ancestral State Reconstruction

To assess the evolutionary history of the GCFs, the ancestral node for each GCF was inferred in the species tree using the trace character history function implemented in Mesquite (65). In some cases, BiG-SCAPE split BGCs into multiple GCFs that were highly homologous, suggesting they may be involved in the biosynthesis of related compounds. Data exploration with different BiG-SCAPE similarity cutoffs did not resolve these relationships, prompting the manual grouping of GCFs into pathways (66, 67). GCFs were considered to belong to the same pathway if: (i) the BGCs shared similar architecture, (ii) the majority of the genes in the cluster had the same function, albeit not necessarily in the same order, and (iii) the majority of genes in the BGC had a BLAST similarity of more than 50% over 80% coverage rate (67). A pathway presence/absence table was used as a character matrix, and likelihood calculations were performed using the Mk1 model. Likelihood scores >50% were used to infer the points of pathway acquisition in the species tree.

3. Results and Discussion

To characterize the genomic features and secondary metabolism potential of this diverse group of specialized symbionts, we sequenced the genomes of 14 strains across the symbiont phylogeny, spanning all ant-agriculture ecologies (Table S1) (20), with the exception of yeast agriculture where escovopsis have never been found. Three strains belonging to different clades were sequenced with PacBio and Illumina technologies, whereas the rest were sequenced with Illumina alone (Table S2). We expanded our dataset with the addition of 24 additional escovopsis strains that were publicly available) (26, 28), and a number of other closely related fungal species from the Hypocreaceae family (Table S1).

The quality of the genomic assemblies generated in this study was high, with an average BUSCO (Benchmarking Universal Single-Copy Orthologs) score of 94.7% for the Ascomycota lineage dataset (Table S2). GC content ranged from 47.2% to 56.4%, with an average of 52.3% (Table S2), consistent with recent reports (24, 28) and other Pezizomycotina fungi (68) .

Phylogenetics of *Escovopsis* and relatives

To infer a genome-scale phylogeny of *Escovopsis* and relatives, we employed a concatenation approach using single-copy genes. The inferred proteomes of all 52 species in our dataset were subjected to an orthology analysis, resulting in 2314 single-copy orthologous genes that were subsequently utilized to infer a phylogeny. Our data reveals that the attine-associated symbionts form a monophyletic group, sister to a clade composed of *Cladobotryum* sp. and *Hypomyces rosellus*, both mycoparasites (Figure 1A). The evolutionary history of escovopsis suggested by this phylogeny generally reflects that of the fungus-growing ants (20). As such, strains infecting gardens of lower attines appear basal to the rest, whereas most recently diverging lineages are

246 associated with higher attine agriculture and leaf-cutter ants (Figure 1A). The shift experienced by
247 some lower attines to cultivating Pterulaceae fungi is also mirrored by the phylogeny, with an
248 intermediate clade exploiting coral agriculture, represented by strains NGL070, ICBG726,
249 ICBG1054, ICBG1065 and ICBG1075. Highlighting the diversity of escovopsis associated with
250 coral agriculture, a clade including four strains associated with coral fungi (ICBG712, ICBG721,
251 NGL057 and NGL216), appear within the basal members of this monophyletic group (Figure 1A).
252 The presence of these two distinct coral agriculture-associated clades, therefore, break congruence
253 of the ant and escovopsis phylogenies. Recent studies have proposed to split the genus *Escovopsis*
254 intro three different genera (*Escovopsis*, *Sympodiorosea*, and *Luteomyces*) (14). To assess whether
255 these two coral agriculture-associated clades may in fact represent two distinct taxonomical
256 genera, we inferred the phylogenetic position of the escovopsis in this study among those from
257 previous studies (14). Our results (Figure S1) suggest that strains within these two clades indeed
258 belong to different genera. Together with strains exploiting higher agriculture (*E. weberi, E.*
259 *moelleri* and *E. aspergilloides*), the intermediate clade exploiting coral agriculture are true
260 *Escovopsis* (*E. multiformis*). However, its sister clade contains strains closely related to the newly
261 described *Sympodiorosea*. Interestingly, the basal-most clade comprises strains most closely
262 related to *Luteomyces* and to strains belonging to an as of yet undescribed genus (Figure S1).
263 Overall, these results highlight the need for further work to fully resolve the taxonomical diversity
264 withing these group of parasites.

265 To estimate the evolutionary distance between strains, we performed a POCP (Percentage of
266 Conserved Proteins) analysis. As expected, with increased phylogenetic distance POCP values
267 decrease. For instance, escovopsis infecting leaf-cutter agriculture share on average 96% of their
268 proteins among each other, whereas only around 88% are shared with *Luteomyces*, *Sympodiorosea*

and the newly undescribed genus (Table S3, Figure 1B). Despite appearing in the same clade in our phylogeny, *Luteomyces* strains and those from the undescribed genus, share as many proteins between each other (88%) as they do with strains parasitizing any other agricultural system. This suggests that there is as much phylogenetic divergence between these two groups, as there is between them and any other escovopsis clade, supporting the notion that what has been traditionally considered *Escovopsis* is in fact at least three, and possibly four, different genera. Furthermore, POCP values lower than 91% segregate our dataset into the recently proposed genera, whereas values above 91% and 95% delineate distinct species and strains within a species respectively (Table S3). Mirroring our phylogenetic placement of *M. zeteki*-associated *Escovopsis*, in POCP analysis, NGL179 shares more proteins (95.1%) with strains infecting leaf-cutter agriculture, than with those exploiting general higher agriculture (92%). POCP analyses have been useful to resolve bacterial groups at genus level, which correlate with POCP values < 50%. While some studies have implemented the method in some fungi at the family level (POCP values < 70%) (70), this strategy cannot be widely employed yet for delineating fungal groups, as genome sampling in fungi remains scarce. However, our POCP analysis reveals a significant degree of genetic diversity between *Escovopsis* (sensu lato) clades, and suggests a protein similarity threshold of 87-91% to delineate different genera in this group of parasites. Further efforts are required to elucidate whether the POCP differences can delineate distinct genera in a diversity of fungi.

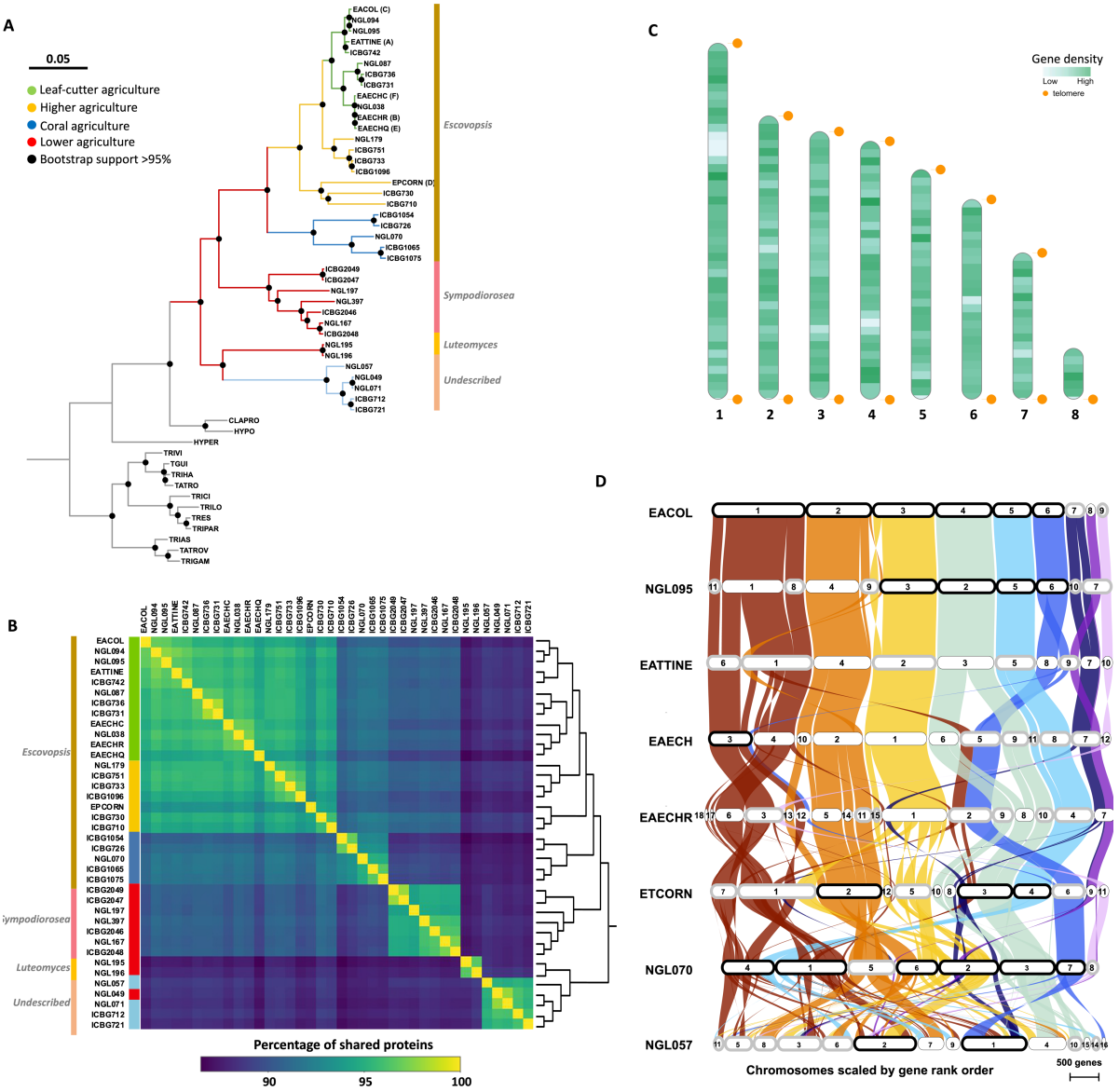


Figure 1. Genomic features of escovopsis. A) Phylogenomic tree of *Escovopsis* and allies constructed with a supermatrix approach on 2314 single-copy orthologous genes. Black dots represent bootstrap support higher than 90%. Branch colors describe different attine agricultural systems: green, leaf-cutter agriculture; yellow, general higher agriculture; blue, coral agriculture (divided in A, blue; and B, light blue; reflecting most derived and more basal coral agriculture); and red, lower agriculture. Side colored bars represent new taxonomical affiliations based on (14) . (B) Heatmap depicting the percentage of conserved proteins across escovopsis strains. Lighter colors represent high levels of shared proteins, whereas dark colors depict fewer shared proteins. The dendrogram on the right represents a hierarchical clustering analysis. (C) Ideogram representing the chromosomal level assembly of an *Escovopsis* sp. strain isolated from an *Apterostigma dentigerum* nest (NGL070). Green colored bands represent regions with high gene density; lighter colors depict low gene density regions. Orange dots represent areas harboring telomeric repeats. (D) Synteny plot depicting the collinearity between different escovopsis genomes across attine agriculture. Highly syntenic regions are connected by colored bands.

Contigs in black boxes represent complete chromosomes, whereas those in grey harbor telomeric repeats just at one chromosomal end.

Escovopsis genomes are organized into highly syntenic chromosomes.

To elucidate the genomic organization of escovopsis, we screened the genomes of the four most contiguous assemblies for telomeric repeats. In NGL070, stretches of (TTAGGG)n were detected at both ends of six contigs, representing complete chromosomes (Figure 1C). The two remaining contigs harbored telomeric repeats only at one end, constituting either two fragments of the same chromosome, or two distinct uncomplete chromosomes. A similar pattern was observed for the highly contiguous EACOL, NGL095 and NGL057 strains, harboring 6, 4 and 2 complete chromosomes, and 2, 5 and 7 fragmented ones with telomeric repeats at one end respectively (Figure 1D). These observations suggest that escovopsis has 7-8 chromosomes, in agreement with other members of the Hypocreales family, such as *Trichoderma reesei, Neurospora crassa* (71), and *Metarhizium brunneum* (72), which organize their genomes in seven chromosomes.

To assess the conservation of escovopsis genomic architecture, we performed a synteny analysis of the eight most continuous genomes available. Our ortholog-based analysis reveals that strains share a high degree of collinearity, with 87.83% of the genes appearing in the same chromosome and in the same order (Figure 1D). This is particularly apparent among strains of the same clade, as evidenced by those parasitizing leaf-cutter agriculture (EACOL, NGL095, EATTINE, EAECHC, and EAECHR). As expected, collinearity has a positive correlation with phylogenetic relatedness, with distant strains exhibiting increasingly different genomic organization. Chromosomes 1, 2, 3, 4 and 5 (nomenclature relative to strain EACOL), are extremely well conserved, extending beyond *Escovopsis* spp. infecting leaf cutter agriculture and including those

involved in general higher agriculture. Chromosome 6, although well conserved in strains exploiting general higher agriculture and coral agriculture-associated NGL070, has experienced recent rearrangements, as evidenced by its fusion with a fragment of chromosome 1 occurring in the clade represented by EAECHC and EAECHR. Previous reports revealed a high degree of micro mesosynteny between genomes of *Escovopsis* and *Trichoderma* (24), suggesting that both genomes are organized in genome segments with similar gene content but rearranged in order and orientation.

Escovopsis harbor a reduced genome

Fungi vary extensively in genome size, spanning three orders of magnitude and ranging from the small genomes of some Microsporidia (2Mb) to the large ones in Picciniales fungi (2Gb). Some of the smallest genomes are found in obligate parasites (73). Escovopsis genome sizes range between 21.4 and 38.3 Mbp (40.7Mb), with an average of 28.7 Mb, corroborating previous studies (24, 28) that estimated their genome sizes around 24.7-27.2 Mb. These genomes are reduced in size relative to those of closely related Sordariomycetes (Figure 2A, 2B, S2A, S2B). Interestingly, *Escovopsis* represents three of the five smallest genomes from all Sordariomycetes strains publicly available in Mycocosms (https://mycocosm.jgi.doe.gov) (Figure 2A). The other two belong to *Ophiocordyceps camponoti-rufipedis* and *O. australis* strain 1348a, both highly specific parasites of ants (74). Within escovopsis, lower attine *Luteomyces* spp. strains harbor significantly smaller genomes than those infecting higher attine nests (Figure 2B, Table S4A). No differences in genome size were detected in escovopsis exploiting other agricultural systems (Figure 2B, Table S4B), with the exception of strains infecting higher agriculture, which are highly variable.

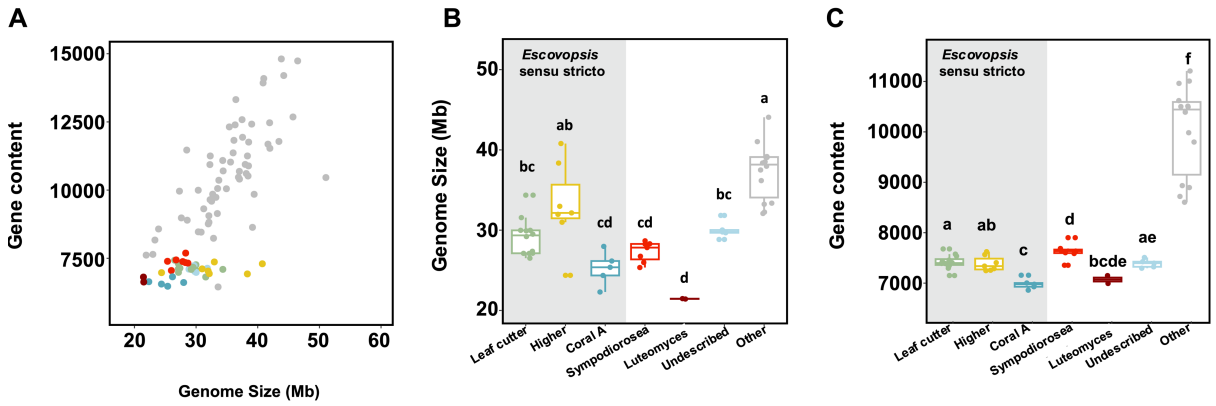


Figure 2. Escovopsis harbor reduced genomes with fewer genes than their non-ant associated relatives. (A) Relationship between genome size and gene content for sequenced fungal genomes. Genome size (B) and gene content (C) of escovopsis strains across different attine agricultural systems. Box colors denote attine clades systems: green, leaf-cutter agriculture (*Escovopsis* spp.); yellow, general higher agriculture (*Escovopsis* spp.); blue, coral agriculture A (*Escovopsis* spp.); red, lower agriculture (*Sympodiorosea* spp.); and light blue, coral agriculture B (undescribed genus), grey other *Sordariomycetes*.

Gene number in escovopsis ranged between 6477 and 7693 (Table S2), representing 9 out of the 10 species in Mycocosm with the fewest genes within the Sordariomycetes. Unlike other fungi in the family, where gene content positively correlates with genome size ($r^2 = 0.32$, $p < 0.0001$), gene number in escovopsis is stable and does not associate with genome size ($r^2 = 0.06$, $p = 0.07$) (Figure 2A). While escovopsis harbor fewer genes than their relatives (Kruskal-Wallis rank sum test, $X^2 = 30.11$, d.f. = 1, $p < 0.001$, Figure S3A), there is no difference in gene content between escovopsis exploiting the nests of lower and higher attines (Table S4C, Figure S3B). However, those within the coral A clade have a slightly different gene content than those associated with lower and leaf-cutter agriculture (Figure 2C, Table S4D). These results are congruent with recent surveys (28) revealing that total coding sequence (CDS) length and intron number in escovopsis genomes are low relative to free-living relatives, consistent with reduced gene content.

In addition to gene number, we investigated two drivers of fungal genome size: repeat content and Repeat-Induced Point mutation. First, while transposable elements are often associated with fungal pathogens (75, 76), their number in escovopsis is significantly lower than in non-ant associated relatives (Kruskal-Wallis, $X^2 = 14.19$, d.f. = 1, $p < 0.001$), which in part, explains escovopsis' small genomes. Second, fungi have evolved a genome defense mechanism to mitigate the potentially detrimental consequences of transposable elements and other repeated genomic regions (77). By altering nucleotide ratios, Repeat-Induced Point mutation (RIP) leads to gene inactivation and genome reduction. Deactivation of RIP, therefore, can lead to genome expansion due to retrotransposon proliferation (78). Previous reports based on the analysis of just one strain, suggested that *E. weberi* may have lost genes involved in RIP. BLAST analyses with the sequences of the two canonical genes known to mediate the RIP pathway revealed that all *Escovopsis* genomes in our dataset harbored orthologs for one gene essential to the RIP process (RID, RIP deficient) but lacked orthologs to the other RIP canonical gene (DIM2, defective in methylation) (Table S5). Genome-wide RIP analyses using The RIPper's sliding window approach revealed that all escovopsis strains show hallmarks of RIP (Table S6), although they vary greatly in the proportion of their genomes that are affected by it. While some strains harbored little evidence of RIP (ICBG1096, 1.01%), others are highly affected by it, with the most extreme case being ICBG1075, where 23.26% of its genome present hallmarks of RIP. This variation across escovopsis genomes of similar size indicates that RIP is not solely responsible for genome reduction in this group, but it may play some role in some species.

Escovopsis' small genomes and the genomic traces of RIP, together with the generalized loss of DIM2, support previous studies (24) that proposed RIP as a genomic defensive mechanism that limited transposon proliferation in *Escovopsis* in the past. A consequence of RIP is the relative

absence of duplicated genes (77). Therefore, the loss of this defense mechanism, may represent an opportunity for this parasite to evolve new metabolic functions through gene duplication and subfuncionalization.

Ascomycota with genome sizes between 25 and 70 Mb, and in particular Sordariomycetes, often exhibit positive correlations between genome size and gene content (73, 79, 80). *Escovopsis* evades this trend (Figure 2A), suggesting that a different evolutionary processes may be affecting this genus. Symbiosis often leads to the streamlining of microbial genomes through genome reduction and gene loss, as epitomized by the tiny genomes of many bacterial endosymbionts of insects (81). Genome streamlining in bacteria can be explained by the loss of redundant genes with drift (82), or by selection against non-essential genes (83). Similar dynamics can occur in fungal mutualists and parasites (84, 85). In particular, fungal parasites associated with insects have been shown to be particularly prone to gene loss (80). Within the Sordariomycetes, the smallest genomes belong almost exclusively to endosymbionts, endoparasites, or fungal parasites vectored by insects (80).

BGC Diversity and distribution

Secondary metabolites in fungi can define ecological niches (86), delimit host ranges (4, 87, 88), and provide selective advantages under specific ecological conditions (89). The metabolic pathways responsible for the synthesis of microbial toxins and other secondary metabolites are typically encoded by biosynthetic gene clusters (BGCs). BGCs encode for backbone enzymes responsible for the synthesis of the core structure of a metabolite, as well as tailoring enzymes that modify this assembly, along with transcription factors and transporters (90). To assess the biosynthetic potential of the escovopsis, we performed a computational genome mining analysis

using the program fungiSMASH (52). The most common backbone enzymes in fungi include polyketide synthases (PKSs), nonribosomal peptide synthetases (NRPSs), terpene synthases, and dimethylallyltransferases (DMATs) (91). All escovopsis genomes analyzed harbored a diversity of BGCs belonging to the major biosynthetic classes (Table S7). Escovopsis' chemical potential content ranged from 16 BGCs in NGL197, to 33 in NGL057. On average, each genome featured 23 BGCs, and an average metabolic diversity of 28.7% NRPs, 25.6% PKS, 21.3% terpenoids, 16.3% hybrids, 2.4% betalactones and 3.6% others. There was no correlation between the number of BGCs and the number of contigs or scaffolds per genome ($R^2 = 0.03$, $p = 0.13$), suggesting that our dataset was robust and that the different sequencing technologies employed did not bias our BGC survey. In addition, no correlation was found between the number of BGCs in each escovopsis strain and genome size ($R^2=0.004$, $p = 0.7$).

While fungi within the Hypocreales are prolific secondary metabolite producers, with an average of 43 BGCs per genome, escovopsis have significantly fewer BGCs than their non-fungus-farming ant associated relatives (Kruskal-Wallis $X^2 = 28.17$, d.f. = 1, $p < 0.001$, Figure S4A), corroborating recent findings using fewer escovopsis genomes (28). We found no statistical differences in BGC abundance between escovopsis strains infecting higher or lower attine nests (Figure S4B), nor between the majority of strains attacking different agricultural systems (Kruskal-Wallis p=0.67, Figure 3A), with the exception of small differences in BGC number in strains infecting general higher agriculture and leaf-cutter agriculture. Upon graphical inspection, we observed a clear bimodal distribution in BGC abundance in strains infecting lower agriculture (Figure 3A), that unequivocally divided the dataset in distinct phylogenetic taxa. We therefore explored whether there is a correlation between BGC content and phylogeny by assessing differences in BGC number across escovopsis groups (Figure S4C). All clades harbored significantly different number

435 of BGCs, with the exception of *Leutomyces* spp. and the undescribed genera, both composed of
436 strains infecting lower agriculture (Kruskal-Wallis, $X^2 = 47.91$, d.f. = 6, $p < 0.001$). Strains within
437 *Luteomyces* and the undescribed new genus (i.e., NGL195, NGL196, NGL049, NGL057,
438 NGL216, ICBG712 and ICBG721) harbor more BGCs than more derived strains. Within the
439 *Escovopsis* spp. there is an increase in BGC abundance from the most basal strains (coral A) to the
440 more derived (leaf cutter agriculture) (Figure S4C).

441 The reduction in BGC abundance in escovopsis relative to other non-ant associated Hypocreales
442 is consistent with its shift in lifestyle to an obligate symbiont of ant nests. Transitions from free-
443 living states to obligate symbioses can often be accompanied by gene loss due to relaxed selection
444 on genes that are no longer necessary in a stable, predictable environment (24, 92). Additionally,
445 some specialist parasites are known to harbor a narrower suit of BGCs relative to generalist ones.
446 For instance, *Metarhizium* strains that acquired the *dtx* biosynthetic gene cluster, responsible for
447 the synthesis of a diversity of toxins, have broader host ranges (infecting hundreds of insect
448 species) compared with non-toxigenic strains (lacking the BGC) that present much narrower host
449 ranges, infecting only locusts and grasshoppers (4). Correlating with a higher content in
450 biosynthetic gene clusters, *Escovopsis* spp. strains infecting higher agriculture (e.g. *E. weberi*) are
451 thought to be more virulent than escovopsis infecting lower agriculture (19).

452

453

454

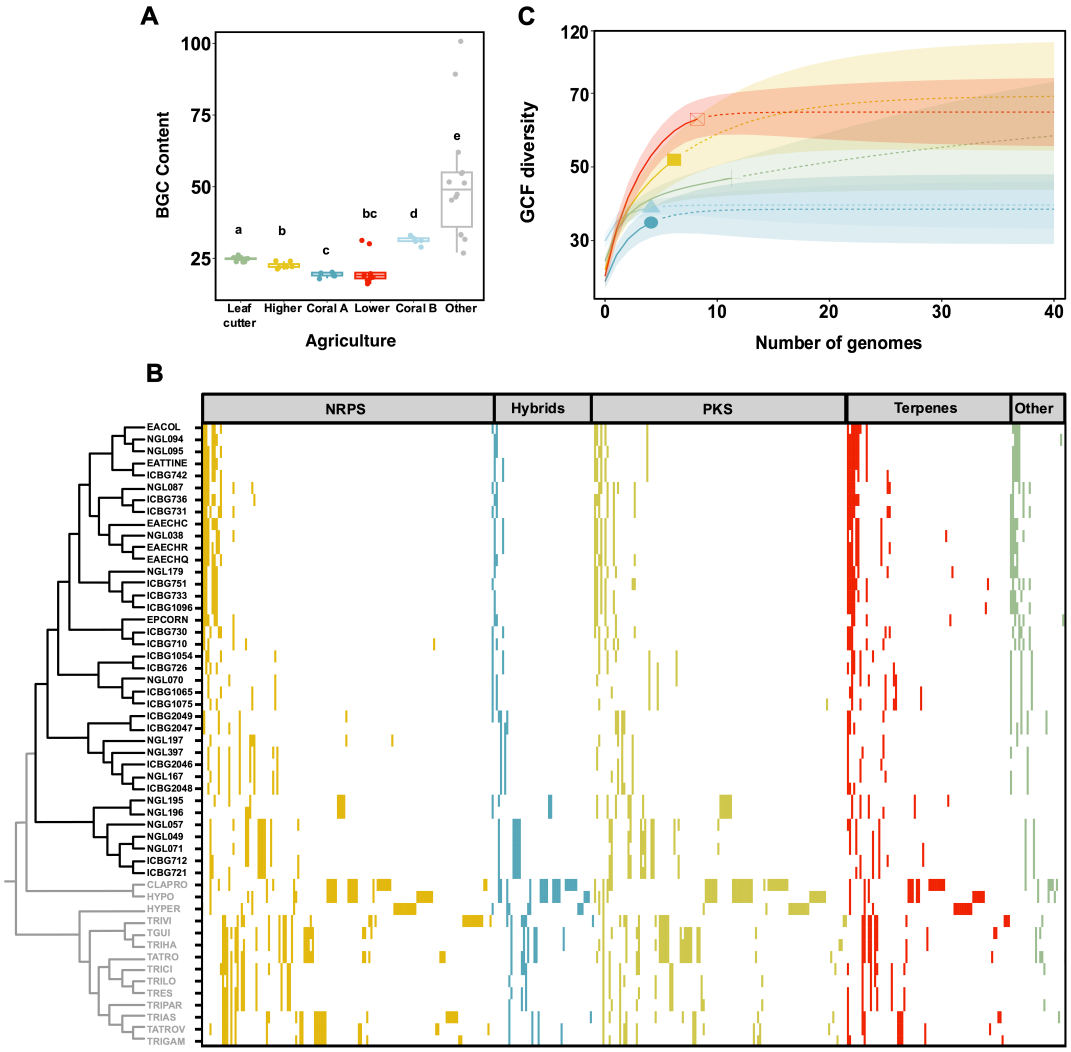


Figure 3. (A) Total number of biosynthetic gene clusters (BGCs) identified across escovopsis strains infecting different attine agricultural systems and non-ant associated relatives. Box colors denote attine clades systems: grey, free-living; green, leaf-cutter agriculture (*Escovopsis* spp.); yellow, general higher agriculture (*Escovopsis* spp.); blue, coral agriculture A (*Escovopsis* spp.); red, lower agriculture (*Sympodiorosea* spp.); and light blue, coral agriculture B (undescribed genus). Graph background is colored according to the general agriculture strains exploit: yellow, higher agriculture; and red lower agriculture. (B) Gene Cluster Family (GCF) distribution across escovopsis and closely related fungi. Each column in the heatmap represents a GCF. The presence of a GCF in a strain is highlighted by colored blocks according to BGC class: yellow, NRPs blue, PKS-NRPS hybrids; light green, PKS; red, Terpenes; and green, others (including RiPPs, indoles, siderophores and others). The absence of a GCF is represented by white spaces. (C) Rarefaction curves assessing GCF richness in escovopsis strains across different attine agricultural systems for the given sequencing effort. Continuous lines represent observed diversity, dashed lines inferred diversity. Shaded areas denote confidence intervals. Colors denote agricultural systems: green, leaf-cutter agriculture, yellow, general higher agriculture; blue, coral agriculture A red, lower agriculture; and light blue, coral agriculture B.

To compare BGC composition across strains, we grouped BGCs into gene cluster families (GCFs) based on sequence homology and cluster architecture employing the BiG-SCAPE algorithm. The resulting sequence similarity network built with a similarity score cutoff of 0.5, clustered 1595 BGCs into 411 gene cluster families (GCFs). We visualized the GCF distribution across escovopsis through the construction of a presence/absence table (Figure 3B). One hundred and twenty-eight GCFs were present in the sampled escovopsis, and 102 of them were unique to escovopsis relative to non-Attine associated fungi. Only 26 GCFs were shared between escovopsis and other Hypocreales species (Table S8, Figure 3B). A rank-abundance curve demonstrates that 27 GCFs occur only once in escovopsis, and an additional 27 are present in just two strains (Figure S5). Surprisingly, no GCF as defined by BigSCAPE was ubiquitous in escovopsis, and therefore characteristic of the group of symbionts as a whole. Rarefaction curves provide an assessment of GCFs richness for the given sequencing effort and reveal that although our sampling was largely adequate, additional chemical diversity is yet to be discovered, especially within the most basal clade of escovopsis infecting lower attine gardens (Figure 3C). Further sequencing efforts in strains from this group will surely reveal additional GCFs.

To distinguish novel BGCs from already described ones, we supplemented our dataset with characterized gene clusters from the MIBiG database as a reference, which at the date of publication contained 1923 BGCs, out of which 207 were of fungal origin. Only five GCFs in our escovopsis dataset are homologous to BGCs in the database. Three families comprising highly similar BGCs cluster together with the MIBiG cluster BGC0001585, responsible for the synthesis of melinacidin IV, suggesting they represent slightly different variants of the same biosynthetic pathway. The other two GCFs are homologous to BGC0001583 and BGC0001777, which encode

for emodin and shearinines respectively. The distribution of all three GCFs is discrete. While most strains of escovopsis harbor the BGC responsible for the production of melinacidin IV, those encoding for shearinine and emodin are restricted to more derived clades (i.e., *Escovopsis* spp. for shearinine, and *Escovopsis* spp, with the exception of those exploiting coral agriculture for emodin).

Fermentation experiments using *E. weberi* have led to the detection and potential functional role of all three metabolites and some derivatives (26). *E. weberi*-produced shearinine derivatives can deter ants and are lethal at high concentrations, preventing insect workers from weeding their garden, thus allowing the parasite to persist in the nest (26). The production of the epipolythiodiketopiperazine melinacidin IV inhibits the growth of the ant defensive mutualist *Pseudonocardia*, whereas the synthesis of emodin has detrimental effects on the cultivar (26) and other co-occurring actinobacteria, such as *Streptomyces*. While the production of these metabolites has been detected in *Escovopsis* strains parasitizing leaf cutter ant gardens, our results demonstrate that the distribution of these BGCs is broader than previously thought and extends to strains exploiting other agricultural systems. Whereas shared GCFs with other fungal genera suggests that they may play a general role in fungal physiology, the presence of GCFs characteristic of specific clades correlating with different attine agricultural systems likely reflects the distinct selective pressures exerted on the symbionts by these different ecosystems. These results are consistent with an on-going arms-race where these symbionts must constantly evolve new adaptations to overcome not only cultivar defenses, but also, very likely, ant defenses, those exerted by protective symbionts such as *Pseudonocardia* and those exerted by other microbes that inhabit these complex microbial communities. For example, the defensive symbiont of beewolves, *Streptomyces* spp., produces different antibiotic cocktails (both in composition and concentration)

in association with each insect species, but also in distinct geographical regions (93), presumably as an adaptation to defend their hosts against different local pathogen communities. Furthermore, the variation metabolic profiles of escovopsis could be a a reflection of them having different impacts on the symbiosis; while some (e.g., *E. weberi*) have been shown to be highly virulent parasites of the ants' cultivars, experimental tests of the impacts of most species are not known. More experimental work is required to assess the specific roles that individual metabolites may play in the ecology of this diverse group of symbionts..

Recent surveys reveal that less than 3% of the biosynthetic space represented by fungal genomes has been linked to metabolites (91, 94). As such, most of the GCFs identified by BGC similarity networks could not be correlated with known compounds, suggesting *Escovopsis* is a promising source for the discovery of new bioactive compounds.

Gene cluster families (GCFs) delineate groups of *Escovopsis*

To assess differences in biosynthetic profiles between *Escovopsis* strains exploiting different attine agricultural systems we performed a non-metric multidimensional scaling analysis. Our results demonstrates that *Escovopsis* strains harbor very different GCF profiles than related non-ant associated fungi, and that these profiles differ between *Escovopsis* infecting higher and lower agriculture (Figure 4A, ANOSIM, $R = 0.59$, $p < 0.001$, 999 permutations). Likewise, GCF profiles are sufficient to cluster strains into separate groups based on phylogenetic lineage (Figure 4B, ANOSIM, $R = 0.87$, $p < 0.001$, 999 permutations). A PERMANOVA analysis reveals that most of the variation (95%) is explained by the interaction between symbiont genus and ant species (Figure 4B, adonis2, 999 permutations, $R^2=0.952$, $p=0.001$).

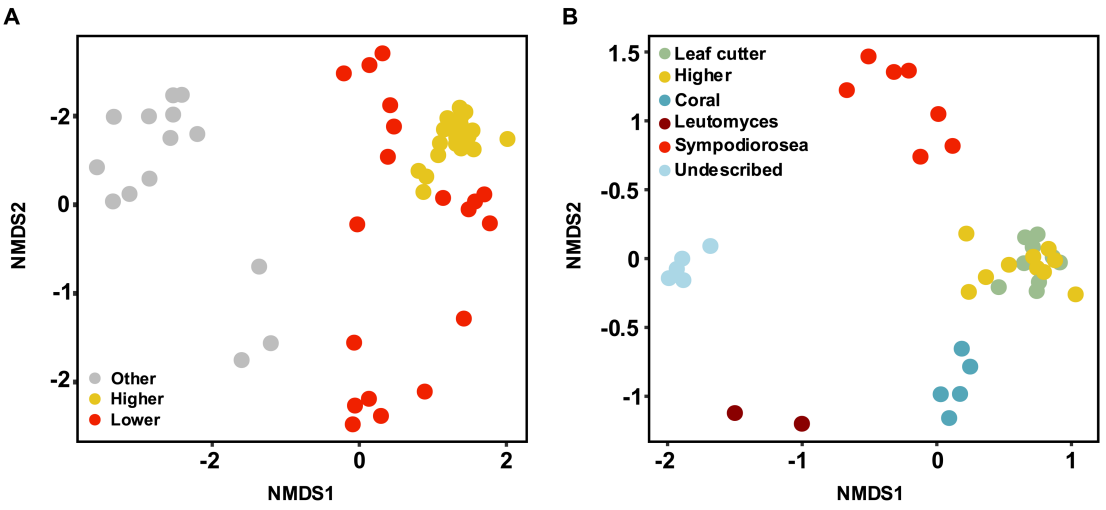


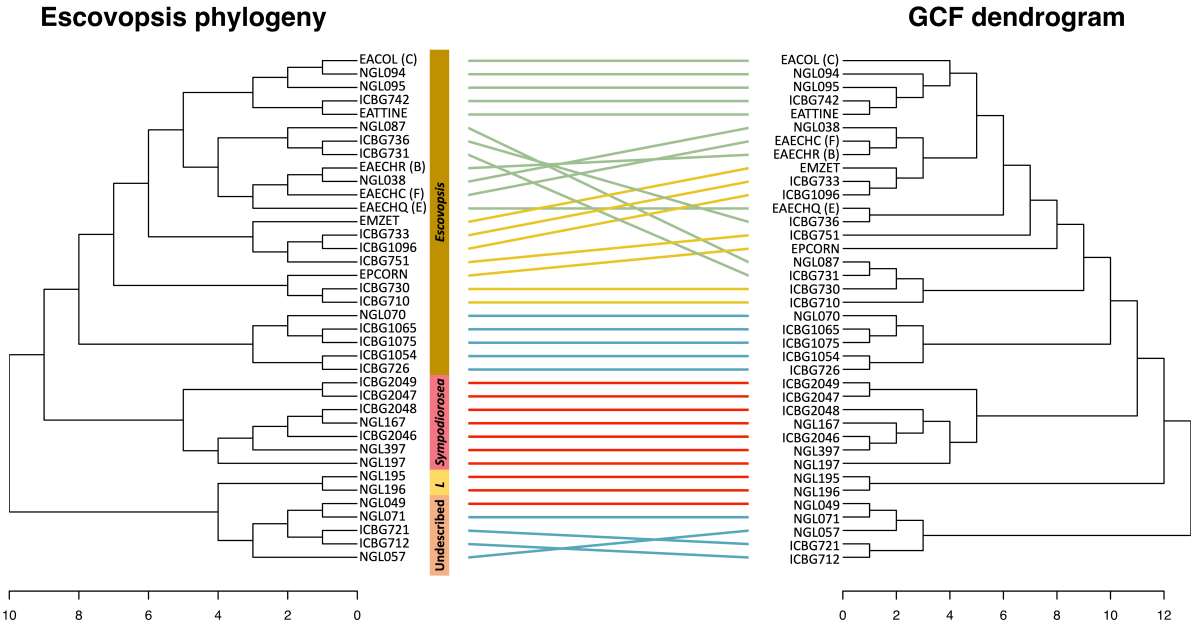
Figure 4. Gene cluster families (GCFs) delineate groups of escovopsis strains. (A) Non-metric multidimensional scaling (NMDS) plot showing differences in GCF composition among escovopsis infecting higher agriculture (yellow) and lower agriculture (red) and non-ant associated fungal relatives (gray). (B) NMDS plot depicting GCF composition in escovopsis strains across ant clades: *Escovopsis* spp. (green, leaf-cutter agriculture, yellow, general higher agriculture; blue, coral agriculture A), dark red, *Luteomyces*; red, *Sympodiorosea*; and light blue, the undescribed genus.

Based on the presence/absence matrix of GCFs across strains, we constructed a hierarchical clustering analysis. *Escovopsis*' phylogeny, based on all orthologs, and the GCF dendrogram are highly congruent (Figure 5A), with the exception that the most basal clade of strains parasitizing coral agriculture and lower agriculture appear as paraphyletic in the GCF dendrogram. An entanglement analysis, which gives a visual approximation of the level of agreement between two dendrograms (95), yielded a score of 0.02, suggesting a high degree of congruence between escovopsis*'* phylogeny and the GCF dendrogram.

These results suggest that the parasite's biosynthetic potential is a phylogenetic trait and can be employed to delineate groups of escovopsis, particularly at broad taxonomical levels. Christopher et al. (69) demonstrated that phylogenetic analyses based on chemical profiles of *Escovopsis*, resulted in similar tree topologies to gene-based phylogenies, confirming that chemical profiles can be considered phylogenetic traits. Additionally, the congruency between the species phylogeny

559 and the BGC profile dendrogram, suggest that BGCs are evolving in parallel with escovopsis
560 species, and that pathway gains and subsequent vertical inheritance, as well as losses are the main
561 forces driving BGC diversification, given that common horizontal transfer of BGCs between
562 strains would result in incongruent topologies.

563



564

565 **Figure 5. GCF profiles in *Escovopsis* are a phylogenetic trait**. Tanglegrams revealing
566 congruence between (A) *Escovopsis* phylogeny and strain biosynthetic potential. Lines connect
567 strains with their GCF profile and colors denote attine clades systems: grey, free-living; green,
568 leaf-cutter agriculture, yellow, general higher agriculture; blue, coral agriculture A red, lower
569 agriculture; and light blue, coral agriculture B. Branches have been rotated for maximum
570 congruency. The maximum likelihood tree was built with 681 single copy orthologous genes. The
571 chemical dissimilarity dendrogram was generated using hierarchical cluster analysis on the
572 presence and absence of GCFs using Jaccard distance and UPGMA (unweighted pair group
573 method with arithmetic mean) as the agglomeration method.

574

575

576

577

a. Pathway evolution: Ancestral state reconstruction

To explore the evolutionary history of escovopsis' biosynthetic pathways relative to their encoding strains, we performed an ancestral state reconstruction analysis. We clustered GCFs into pathways (Ps) based on the assumption that they produce related compounds (See methods section, Table S9). The 411 GCFs detected in our escovopsis dataset were clustered into many different pathways. Their distribution was overlaid onto a simplified parasite phylogeny, generated by collapsing certain branches on the species tree, resulting in 8 lineages A-H, that correspond with the newly described genera (A, undescribed genera; B, *Luteomyces*; C, *Sympodiorosea*; and D-H *Escovopsis* spp.) Figure 6). 67 pathways were present in escovopsis, out of which 56 were unique to this group of symbionts and 11 were shared with other hypocreales. The analysis revealed that 15 pathways were present in the common ancestor of escovopsis , and 11 of those were shared with the closely related genus *Cladobotryum.* The transition from a non-ant associated lifestyle to a fungal garden parasite correlates with the loss of one pathway (P67), involved in the biosynthesis of an uncharacterized PKS, that is present in all close relatives but absent in every *Escovopsis* strain. Five pathways (P7, P8, P10, P14 and P15) evolved early in the evolutionary history of these fungal symbionts and are present in most strains. However, none of them are ubiquitous, as there have been some clade-specific loses.

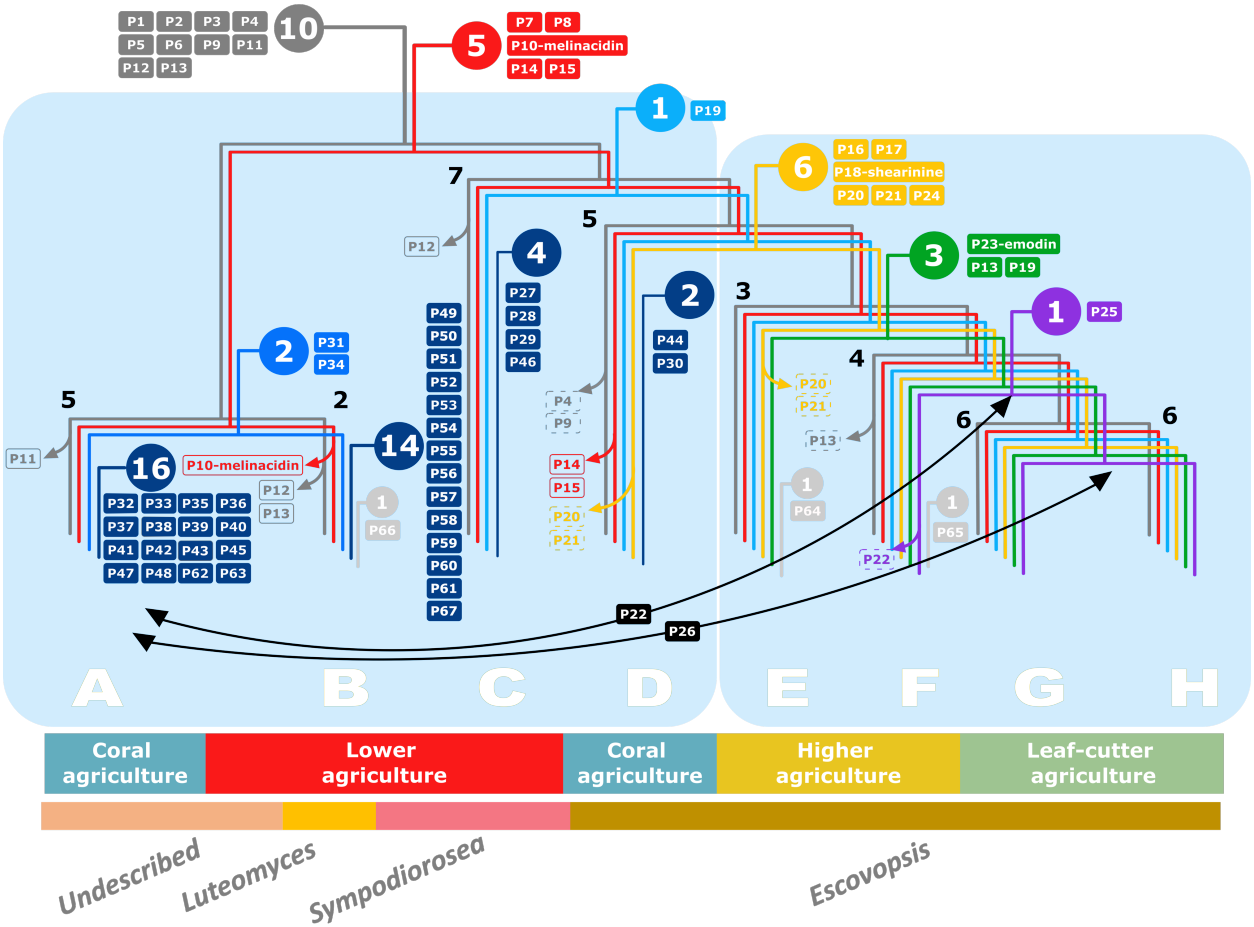


Figure 6. Phylogeny depicting the inferred ancestry of secondary metabolite biosynthetic pathways present in escovopsis. A simplified species phylogenomic tree depicts 8 major *Escovopsis* lineages A-H, that correspond to the newly proposed taxonomical divisions. The number of the strains in each group is indicated in black adjacent to branch nodes. Circles indicate the number of pathways originated at various points in the species tree, whereas filled boxes state pathways next to the point of acquisition. Transparent boxes represent pathway losses in all strains within a clade (continuous outline) or more than 50% of the strains in the clade (dashed outline). Grey, pathways shared with the sister clade; red, shared the common ancestor of the genus; indigo, present in basal strains; light blue, present in derived clade; yellow; shared between strains infecting derived coral agriculture and general higher agriculture; green, shared by all higher agriculture; purple, shared by derived general higher agriculture and leaf-cutter agriculture; dark blue, clade-specific pathways; light gray, strain-specific pathways. Black arrows describe HGT events.

The remaining 50 pathways were acquired at various points during the evolution of the genus either through horizontal gene transfer (HGT) or *de novo*. An average of three pathways are

acquired by *Escovopsis* with every transition to a new ant agricultural system. However, the transition from strains within the three most ancestral escovopsis clades (A-C, *Sympodiorosea*, *Leutomyces* and an undescribed genus) to the most derived super-clade including clades D-H (*Escovopsis* spp.) correlate with the acquisition of 5 pathways, including P18, responsible for the biosynthesis of shearinine D and indicates that these pathways are unique for *Escovopsis* spp. Four pathways, including P24 which synthesizes emodin, evolved early in the divergence of *Escovopsis* to infect higher attine agriculture. Interestingly, no pathway is unique to the most derived clade of the parasite (i.e., strains parasitizing leaf-cutter ant nests, clades G and H).

Phylogenetic analysis of key biosynthetic genes from each pathway confirms, based on congruence with the species tree, vertical inheritance for most of the pathways following acquisition. However, they also suggest that some pathways were exchanged between strains. P22, encoding for a terpenoid, has been transferred between the ancestor of parasite strains exploiting higher attines (ancestor of clades F-H) and the basal clade parasitizing coral agriculture (clade A). Similarly, P26, encoding a PKS, has been shared between the ancestor of strains infecting leaf-cutter agriculture and the most derived clade parasitizing coral agriculture. In both cases, the direction of the exchange remains unclear. However, once transferred, these pathways have subsequently been vertically inherited by all members of the clades.

The evolution of *Escovopsis'* biosynthetic potential has not only evolved through pathway acquisition, but also through BGC loses. Six pathways have been lost in strains infecting lower attines: three that were already present in the sister clade represented by *Cladobotryum* and *Hypomyces rosellus* (P10, 11, 12, and P13) and three that evolved in the common ancestor of all escovopsis strains (P10, P14 and P15). P12 appears to have been lost twice, once in clade B (*Luteomyces*), as well as in clade C (*Sympodiorosea*). P4 and P9 have also been lost in 4 and 2

636 strains respectively. Within *Escovopsis* parasitizing higher attines, no pathway has been lost
637 completely. Only three pathways have been lost in some strains: the ancient P13 in clade F, and
638 the more recently evolved P18, encoding for shearinine, in one clade E strain (EPCORN). While
639 the loss of this BGC in EPCORN, and its inability to synthesize the resulting compound was
640 already described through both bioinformatic and chemical assays (26), our results suggest it is
641 not a widespread event, given that all the remaining strains still conserve the BGC. A number of
642 pathways (P4, P9, P14, P15, P20, P21) have been lost in *Escovopsis* spp. strains that experienced
643 a host-shift, from exploiting a *Leucocoprineae* to a *Pterulaceae* host. It is plausible that these
644 pathway losses represent an adaptation and specialization to exploit a new host. In general, more
645 pathways have been lost in *Escovopsis* strains infecting lower attine gardens than those attacking
646 the cultivars of higher attines, and those pathways were most often ancient, suggesting that newly
647 acquired BGCs either (i) have not had enough evolutionary time to be selected against, or (ii) may
648 be adaptive and thus maintained in *Escovopsis*. These results oppose patterns described in other
649 fungi, where generalist parasites harbor more BGCs than specialist ones (4). Escovopsis strains
650 attacking lower attines are thought to be less specialized than those infecting higher attine gardens
651 (19). However, our results suggest that they may be more specialized than previously thought.
652 Furthermore, the colonies of lower attines are smaller than those of higher attines, consisting of a
653 handful and millions of workers respectively. Given the insecticidal properties of some BGCs, it
654 is plausible that parasite strains attacking bigger colonies, require a more diverse cocktail of
655 bioactive compounds relative to those infecting smaller colonies. In fact, studies have
656 demonstrated that the proportion of ant nests harboring fungal contaminants (fungi other than the
657 cultivar) is highest in lower attines (13). However, the proportion of those contaminants made up
658 by *Escovopsis* and its allies is highest for higher attines (13). This could be the result of *Escovopsis*

659 attacking higher attine nests being able to better fend off both fungal competitors than those
660 infection lower attine nests, and inhibiting ant-weeding behavior, given their higher content in
661 BGCs. Additionally, *Escovopsis* strains infecting small colonies may encounter less diverse
662 microbial communities in compared to those encountered in bigger gardens, and may not require
663 as many antibiotic compounds to outcompete other microbes.

664 Our results suggest that *Escovopsis* acquired the capability to synthesize the antimicrobial
665 compound melinacidin IV early in the evolution of this group of symbionts and was subsequently
666 lost in lineage B (P10), i.e. *Luteomyces* parasitizing lower attines. The evolution of the pathway,
667 is however, uncertain. Although we did not detect the presence of the core biosynthetic enzymes
668 in the sister clade to escovopsis, consisting of *Cladobotryum* and *Hypomyces* strains, other
669 Hypocreales are known to synthesize this metabolite, such as *Acrostalagmus* sp, a rare fungal
670 genus that has been found associated with soil (96), mushrooms (97) and plant material (98),
671 suggesting this BGC may have been acquired horizontally. However, while the closely related
672 genus *Trichoderma*, has never been described to synthesize this antibiotic, strains within this genus
673 harbor a number of homologous genes to the melinacidin IV BGC, including the backbone enzyme
674 (99). Therefore, it is alternatively plausible that the pathway responsible for the production of
675 melinacidin IV evolved early within the hypocreal family and was lost in the *Cladobotryum-H.*
676 *rosellus* clade, accumulating enough changes (or requiring fewer genes than previously thought)
677 that we have classified them as different GCFs in our survey.

678 The inferred ancestry for the pathway responsible for shearinine (P18) biosynthesis, suggests it
679 was characteristic of *Escovopsis* spp. While absent from other Hypocreales, a BGC encoding for
680 shearinine D, has been described for the distantly related fungus *Penicillium janthinellum* (100),
681 suggesting that it may evolved through HGT in these symbionts. Emodin, encoded by pathway

682 P24, was one of the last BGCs to evolve within *Escovopsis*, appearing in the ancestor of strains
683 parasitizing general higher agriculture and leaf-cutter ants.

684 The evolutionary transition between lower to higher agriculture in attine ants correlates not only
685 with an increment colony size (from hundreds to millions of workers) (84), but also with an
686 incipient division of labor between worker ants that culminates with the cast system in leaf-cutter
687 ants (101). The transition of escovopsis from attacking lower to higher attine agriculture coincided
688 with the evolution of a new suit of biosynthetic gene clusters, possibly explaining the increase in
689 complexity required by escovopsis to survive in this environment, and the split of these strains into
690 different genera.

691

692 **4. Conclusion**

693 Parasites evade and counter host defenses through a remarkable array of secondary metabolites
694 and natural products. Here, we highlight the highly streamlined genomic features of escovopsis,
695 defined by 7 chromosomes, harboring few repetitive sequences. Despite a high degree of metabolic
696 conservation, we observe limited variation in the parasite's potential to produce secondary
697 metabolites. As the phylogenetic distribution of the encoding biosynthetic gene clusters coincides
698 with attine transitions in agricultural systems and cultivar types, we highlight the likely role of
699 these metabolites in mediating adaptation by a highly specialized pathogen. Future efforts will
700 shed light on the mode-of-action and mechanistic basis of *Escovopsis* secondary metabolites, their
701 role in underlying virulence and pathogenicity in an ancient agricultural system.

702

5. Acknowledgements

We acknowledge funding from the German Research Foundation (Deutsche Forschungsgemeinschaft) under the individual grant program (BE6922/1-1) and Germany's Excellence Strategy (EXC 2124 – 390838134) for AB. This work was also funded by the National Science Foundation (award 1711545) for CC, (award 1754595) for HS, NMG and TR, and (award1927411) for NMG and AR, QVM and AR would like to thank the São Paulo Research Foundation (Fundação de Amparo à Pesquisa do Estado de São Paulo [FAPESP], grant number 2019/03746-0. We are very thankful to Dr. Martina Adamek for advice on BGC clustering and Dr. J Lovell and Dr. Navarro for help in troubleshooting software functioning.

6. Data accessibility

Accession numbers will be provided upon acceptance. BigScape results and other supplementary material can be found in FigShare under the title's project.

7. Author contributions

AB, HS, NMG and NZ conceived of the study. NMG and YC collected samples. AB, HS and CC performed DNA extractions. AB, HS and AGM assembled genomes. AB, QV and AR performed analyses. AB wrote the manuscript. All authors provided valuable comments to the manuscript.

1. References

1. Barrett LG, Heil M. 2012. Unifying concepts and mechanisms in the specificity of plant–enemy interactions. Trends Plant Sci 17:282–292.

2. Woolhouse MEJ, Taylor LH, Haydon DT. 2001. Population biology of multihost pathogens. Science 292:1109–1112.

3. Jaenike J, Perlman SJ. 2002. Ecology and evolution of host-parasite associations: mycophagous *Drosophila* and their parasitic nematodes. Am Nat 160:S23–S39.

4. Wang B, Kang Q, Lu Y, Bai L, Wang C. 2012. Unveiling the biosynthetic puzzle of destruxins in Metarhizium species. Proceedings of the National Academy of Sciences 109:1287–1292.

5. Möbius N, Hertweck C. 2009. Fungal phytotoxins as mediators of virulence. Curr Opin Plant Biol 12:390–398.

6. Wang Y, Wu J, Yan J, Guo M, Xu L, Hou L, Zou Q. 2022. Comparative genome analysis of plant ascomycete fungal pathogens with different lifestyles reveals distinctive virulence strategies. BMC Genomics 23:34.

7. Drott MT, Rush TA, Satterlee TR, Giannone RJ, Abraham PE, Greco C, Venkatesh N, Skerker JM, Glass NL, Labbé JL, Milgroom MG, Keller NP. 2021. Microevolution in the pansecondary metabolome of *Aspergillus flavus* and its potential macroevolutionary implications for filamentous fungi. Proceedings of the National Academy of Sciences 118.

8. Speed MP, Fenton A, Jones MG, Ruxton GD, Brockhurst MA. 2015. Coevolution can explain defensive secondary metabolite diversity in plants. New Phytol 208:1251–1263.

9. Raffa N, Keller NP. 2019. A call to arms: Mustering secondary metabolites for success and survival of an opportunistic pathogen. PLoS Pathog 15:e1007606.

10. Gerardo NM, Mueller UG, Currie CR. 2006. Complex host-pathogen coevolution in the *Apterostigma* fungus-growing ant-microbe symbiosis. BMC Evol Biol 6:88.

11. Gerardo NM, Jacobs SR, Currie CR, Mueller UG. 2006. Ancient host–pathogen associations maintained by specificity of chemotaxis and antibiosis. PLoS Biol 4:e235.

12. Gerardo NM, Mueller UG, Price SL, Currie CR. 2004. Exploiting a mutualism: parasite specialization on cultivars within the fungus–growing ant symbiosis. Proceedings of the Royal Society of London Series B: Biological Sciences 271:1791–1798.

13. Currie CR, Mueller UG, Malloch D. 1999. The agricultural pathology of ant fungus gardens. Proceedings of the National Academy of Sciences 96:7998–8002.

14. Montoya QV, Martiarena MJS, Bizarria R Jr, Gerardo NM, Rodrigues A. 2021. Fungi inhabiting attine ant colonies: reassessment of the genus *Escovopsis* and description of *Luteomyces* and *Sympodiorosea* gens. nov. IMA Fungus 12:23.

15. Currie CR. 2001. Prevalence and impact of a virulent parasite on a tripartite mutualism. Oecologia 128:99–106.

16. Reynolds HT, Currie CR. Pathogenicity of *Escovopsis weberi*: the parasite of the attine ant-microbe symbiosis directly consumes the ant-cultivated fungus 5.

17. Mello D, Silva MC, Leite Martins G, Costa Moreira C, Gechel Kloss T, Elliot SL. 2021. Low virulence of the fungi Escovopsis and Esocovopsioides to leaf-cutting ant-fungus symbiosis. Frontiers in Microbiology 12.

18. Marfetan JA, Romero AI, Folgarait PJ. 2015. Pathogenic interaction between *Escovopsis weberi* and *Leucoagaricus* sp.: mechanisms involved and virulence levels. Fungal Ecol 17:52–61.

19. Bizarria R Jr, Nagamoto NS, Rodrigues A. 2020. Lack of fungal cultivar fidelity and low virulence of Escovopsis trichodermoides. Fungal Ecology 45.

20. Schultz TR, Brady SG. 2008. Major evolutionary transitions in ant agriculture. Proceedings of the National Academy of Sciences 105:5435–5440.

21. Branstetter MG, Ješovnik A, Sosa-Calvo J, Lloyd MW, Faircloth BC, Brady SG, Schultz TR. 2017. Dry habitats were crucibles of domestication in the evolution of agriculture in ants. Proceedings of the Royal Society B: Biological Sciences 284:20170095.

22. Schultz TR, Sosa-Calvo J, Brady SG, Lopes CT, Mueller UG, Bacci M, Vasconcelos HL. 2015. The Most Relictual Fungus-Farming Ant Species Cultivates the Most Recently Evolved and Highly Domesticated Fungal Symbiont Species. Am Nat 185:693–703.

23. Huey Yek S, Boomsma JJ, Poulsen M. 2012. Towards a better understanding of the evolution of specialized parasites of fungus-growing ant crops. Psyche.

24. de Man TJB, Stajich JE, Kubicek CP, Teiling C, Chenthamara K, Atanasova L, Druzhinina IS, Levenkova N, Birnbaum SSL, Barribeau SM, Bozick BA, Suen G, Currie CR, Gerardo NM. 2016. Small genome of the fungus *Escovopsis weberi* , a specialized disease agent of ant agriculture. Proc Natl Acad Sci USA 113:3567–3572.

25. Currie CR, Scott JA, Summerbell RC, Malloch D. 1999. Fungus-growing ants use antibiotic-producing bacteria to control garden parasites. Nature 398:701–704.

26. Heine D, Holmes NA, Worsley SF, Santos ACA, Innocent TM, Scherlach K, Patrick EH, Yu DW, Murrell JC, Vieria PC, Boomsma JJ, Hertweck C, Hutchings MI, Wilkinson B. 2018. Chemical warfare between leafcutter ant symbionts and a co-evolved pathogen. Nat Commun 9:2208.

27. Dhodary B, Schilg M, Wirth R, Spiteller D. 2018. Secondary metabolites from *Escovopsis weberi* and their role in attacking the garden fungus of leaf-cutting ants. Chemistry – A European Journal 24:4445–4452.

28. Gotting K. 2022. Genomic diversification of the specialized parasite of the fungus-growing ant symbiosis. Proceedings of the National Academy of Sciences.

29. Blumer LS, Beck C. 2005. DNA prep for MiSeq Analysis. https://www.beanbeetle.org/microbiome/dna-prep-for-miseq-analysis/. Retrieved 3 July 2023.

30. Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. 2017. Canu: scalable and accurate long-read assembly via adaptive *k* -mer weighting and repeat separation. Genome Res 27:722–736.

31. Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q, Wortman J, Young SK, Earl AM. 2014. Pilon: An Integrated Tool for Comprehensive Microbial Variant Detection and Genome Assembly Improvement. PLoS One 9:e112963.

32. Andrews, S. (2010). FastQC: A Quality Control Tool for High Throughput Sequence Data. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

33. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120.

34. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. J Comput Biol 19:455–477.

35. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 31:3210–3212.

813 36. Lovell JT, Sreedasyam A, Schranz ME, Wilson MA, Carlson JW, Harkess A, Emms D,
814 Goodstein D, Schmutz J. 2022. GENESPACE: syntenic pan-genome annotations for
815 eukaryotes. bioRxiv.

816 37. Palmer J (2020) Funannotate: Fungal genome annotation.
817 https://zenodo.org/record/4054262#.ybe-3lmo8zu.

818 38. The UniProt Consortium. 2019. UniProt: a worldwide hub of protein knowledge. Nucleic
819 Acids Res 47:D506–D515.

820 39. Slater G, Birney E. 2005. Automated generation of heuristics for biological sequence
821 comparison. BMC Bioinformatics 6:31.

822 40. Hoff KJ, Stanke M. 2019. Predicting Genes in Single Genomes with AUGUSTUS. Curr
823 Protoc Bioinformatics 65:e57.

824 41. Korf I. 2004. Gene finding in novel genomes. BMC Bioinformatics 5:59.

825 42. Majoros WH, Pertea M, Salzberg SL. 2004. TigrScan and GlimmerHMM: two open source
826 ab initio eukaryotic gene-finders. Bioinformatics 20:2878–2879.

827 43. Lowe TM, Eddy SR. tRNAscan-SE: a program for improved detection of transfer RNA genes
828 in genomic sequence 10.

829 44. Haas BJ, Salzberg SL, Zhu W, Pertea M, Allen JE, Orvis J, White O, Buell CR, Wortman JR.
830 2008. Automated eukaryotic gene structure annotation using EVidenceModeler and the
831 Program to Assemble Spliced Alignments. Genome Biol 9:R7.

832 45. Jones P, Binns D, Chang H-Y, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J,
833 Mitchell A, Nuka G, Pesseat S, Quinn AF, Sangrador-Vegas A, Scheremetjew M, Yong S-
834 Y, Lopez R, Hunter S. 2014. InterProScan 5: genome-scale protein function classification.
835 Bioinformatics 30:1236–1240.

836 46. Huerta-Cepas J, Szklarczyk D, Heller D, Hernández-Plaza A, Forslund SK, Cook H, Mende
837 DR, Letunic I, Rattei T, Jensen LJ, von Mering C, Bork P. 2019. eggNOG 5.0: a hierarchical,
838 functionally and phylogenetically annotated orthology resource based on 5090 organisms and
839 2502 viruses. Nucleic Acids Res 47:D309–D314.

840 47. Huerta-Cepas J, Forslund K, Coelho LP, Szklarczyk D, Jensen LJ, von Mering C, Bork P.
841 2017. Fast Genome-Wide Functional Annotation through Orthology Assignment by
842 eggNOG-Mapper. Mol Biol Evol 34:2115–2122.

48. Kall L, Krogh A, Sonnhammer ELL. 2007. Advantages of combined transmembrane topology and signal peptide prediction--the Phobius web server. Nucleic Acids Res 35:W429–W432.

49. Eddy SR. 2011. Accelerated Profile HMM Searches. PLoS Comput Biol 7:e1002195.

50. Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B. 2014. The carbohydrate-active enzymes database (CAZy) in 2013. Nucleic Acids Res 42:D490–D495.

51. Rawlings ND, Barrett AJ, Finn R. 2016. Twenty years of the *MEROPS* database of proteolytic enzymes, their substrates and inhibitors. Nucleic Acids Res 44:D343–D350.

52. Blin K, Shaw S, Kloosterman AM, Charlop-Powers Z, van Wezel GP, Medema MH, Weber T. 2021. antiSMASH 6.0: improving cluster detection and comparison capabilities. Nucleic Acids Res 49:W29–W35.

53. McGowan J, Fitzpatrick DA. 2020. Recent advances in oomycete genomics. Adv Genet 105:175–228.

54. Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32:1792–1797.

55. Capella-Gutierrez S, Silla-Martinez JM, Gabaldon T. 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics 25:1972–1973.

56. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, Lanfear R. 2020. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. Mol Biol Evol 37:1530–1534.

57. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. Nat Methods 14:587–589.

58. Letunic I, Bork P. 2021. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. Nucleic Acids Res 49:W293–W296.

59. Katoh K, Misawa K, Kuma K-I, Miyata T. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res 30:3059–3066.

60. Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30:1312–1313.

61. Gilchrist CLM, Booth TJ, van Wersch B, van Grieken L, Medema MH, Chooi Y-H. 2021. cblaster: a remote search tool for rapid identification and visualization of homologous gene clusters. Bioinformatics Advances 1:vbab016.

874 62. Navarro-Muñoz JC, Selem-Mojica N, Mullowney MW, Kautsar SA, Tryon JH, Parkinson EI,
875 De Los Santos ELC, Yeong M, Cruz-Morales P, Abubucker S, Roeters A, Lokhorst W,
876 Fernandez-Guerra A, Cappelini LTD, Goering AW, Thomson RJ, Metcalf WW, Kelleher NL,
877 Barona-Gomez F, Medema MH. 2020. A computational framework to explore large-scale
878 biosynthetic diversity. Nat Chem Biol 16:60–68.

879 63. Kautsar SA, Blin K, Shaw S, Navarro-Muñoz JC, Terlouw BR, van der Hooft JJJ, van Santen
880 JA, Tracanna V, Suarez Duran HG, Pascal Andreu V, Selem-Mojica N, Alanjary M,
881 Robinson SL, Lund G, Epstein SC, Sisto AC, Charkoudian LK, Collemare J, Linington RG,
882 Weber T, Medema MH. 2019. MIBiG 2.0: a repository for biosynthetic gene clusters of
883 known function. Nucleic Acids Res gkz882.

884 64. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B,
885 Ideker T. 2003. Cytoscape: A Software Environment for Integrated Models of Biomolecular
886 Interaction Networks. Genome Res 13:2498–2504.

887 65. Maddison, W. P. and D.R. Maddison. 2021. Mesquite: a modular system for evolutionary
888 analysis. Version 3.70 http://www.mesquiteproject.org.

889 66. Ziemert N, Lechner A, Wietz M, Millan-Aguinaga N, Chavarria KL, Jensen PR. 2014.
890 Diversity and evolution of secondary metabolism in the marine actinomycete genus
891 Salinispora. Proceedings of the National Academy of Sciences 111:E1130–E1139.

892 67. Adamek M, Alanjary M, Sales-Ortells H, Goodfellow M, Bull AT, Winkler A, Wibberg D,
893 Kalinowski J, Ziemert N. 2018. Comparative genomics reveals phylogenetic distribution
894 patterns of secondary metabolites in Amycolatopsis species. BMC Genomics 19.

895 68. Shen X-X. 2020. Genome-scale phylogeny and contrasting modes of genome evolution in the
896 fungal phylum Ascomycota. Science Advances 6.

897 69. Christopher Y, Aguilar C, Gálvez D, Wcislo WT, Gerardo NM, Fernández-Marín H. 2021.
898 Interactions among *Escovopsis*, antagonistic microfungi associated with the fungus-growing
899 ant symbiosis. Journal of Fungi 7:1007.

900 70. Wibberg D, Stadler M, Lambert C, Bunk B, Spröer C, Rückert C, Kalinowski J, Cox RJ,
901 Kuhnert E. 2021. High quality genome sequences of thirteen Hypoxylaceae (Ascomycota)
902 strengthen the phylogenetic family backbone and enable the discovery of new taxa. Fungal
903 Divers 106:7–28.

904 71. Druzhinina IS, Kopchinskiy AG, Kubicek EM, Kubicek CP. 2016. A complete annotation of
905 the chromosomes of the cellulase producer Trichoderma reesei provides insights in gene
906 clusters, their expression and reveals genes required for fitness. Biotechnol Biofuels 9:75.

907 72. Saud Z, Kortsinoglou AM, Kouvelis VN, Butt TM. 2021. Telomere length de novo assembly
908 of all 7 chromosomes and mitogenome sequencing of the model entomopathogenic fungus,
909 Metarhizium brunneum, by means of a novel assembly pipeline. BMC Genomics 22:87.

910 73. Stajich JE. 2017. Fungal Genomes and Insights into the Evolution of the Kingdom.
911 Microbiology Spectrum https://doi.org/10.1128/microbiolspec.FUNK-0055-2016.

912 74. de Bekker C, Ohm RA, Evans HC, Brachmann A, Hughes DP. 2017. Ant-infecting
913 Ophiocordyceps genomes reveal a high diversity of potential behavioral manipulation genes
914 and a possible major role for enterotoxins. Sci Rep 7:12508.

915 75. Fouché S, Oggenfuss U, Chanclud E, Croll D. 2022. A devil's bargain with transposable
916 elements in plant pathogens. Trends Genet 38:222–230.

917 76. Sánchez-Vallet A, Fouché S, Fudal I, Hartmann FE, Soyer JL, Tellier A, Croll D. 2018. The
918 Genome Biology of Effector Gene Evolution in Filamentous Plant Pathogens. Annu Rev
919 Phytopathol 56:21–40.

920 77. Galagan JE, Selker EU. 2004. RIP: the evolutionary cost of genome defense. Trends Genet
921 20:417–423.

922 78. Galagan JE, Calvo SE, Borkovich KA, Selker EU, Read ND, Jaffe D, FitzHugh W, Ma L-J,
923 Smirnov S, Purcell S, Rehman B, Elkins T, Engels R, Wang S, Nielsen CB, Butler J, Endrizzi
924 M, Qui D, Ianakiev P, Bell-Pedersen D, Nelson MA, Werner-Washburne M, Selitrennikoff
925 CP, Kinsey JA, Braun EL, Zelter A, Schulte U, Kothe GO, Jedd G, Mewes W, Staben C,
926 Marcotte E, Greenberg D, Roy A, Foley K, Naylor J, Stange-Thomann N, Barrett R, Gnerre
927 S, Kamal M, Kamvysselis M, Mauceli E, Bielke C, Rudd S, Frishman D, Krystofova S,
928 Rasmussen C, Metzenberg RL, Perkins DD, Kroken S, Cogoni C, Macino G, Catcheside D,
929 Li W, Pratt RJ, Osmani SA, DeSouza CPC, Glass L, Orbach MJ, Berglund JA, Voelker R,
930 Yarden O, Plamann M, Seiler S, Dunlap J, Radford A, Aramayo R, Natvig DO, Alex LA,
931 Mannhaupt G, Ebbole DJ, Freitag M, Paulsen I, Sachs MS, Lander ES, Nusbaum C, Birren
932 B. 2003. The genome sequence of the filamentous fungus *Neurospora crassa*. Nature
933 422:859–868.

934 79. Kelkar YD, Ochman H. 2012. Causes and Consequences of Genome Expansion in Fungi.
935 Genome Biol Evol 4:13–23.

936 80. Fijarczyk A, Hessenauer P, Hamelin RC, Landry CR. 2022. Lifestyles shape genome size and
937 gene content in fungal pathogens. bioRxiv https://doi.org/10.1101/2022.08.24.505148.

938 81. McCutcheon JP, Moran NA. 2012. Extreme genome reduction in symbiotic bacteria. Nat Rev
939 Microbiol 10:13–26.

940 82. Moran NA, McCutcheon JP, Nakabachi A. 2008. Genomics and Evolution of Heritable
941 Bacterial Symbionts. Annu Rev Genet 42:165–190.

942 83. Giovannoni SJ, Cameron Thrash J, Temperton B. 2014. Implications of streamlining theory
943 for microbial ecology. ISME J 8:1553–1565.

944 84. Kooij PW, Pellicer J. 2020. Genome Size Versus Genome Assemblies: Are the Genomes
945 Truly Expanded in Polyploid Fungal Symbionts? Genome Biol Evol 12:2384–2390.

946 85. Katinka MD, Duprat S, Cornillot E, Méténier G, Thomarat F, Prensier G, Barbe V,
947 Peyretaillade E, Brottier P, Wincker P, Delbac F, El Alaoui H, Peyret P, Saurin W, Gouy M,
948 Weissenbach J, Vivarès CP. 2001. Genome sequence and gene compaction of the eukaryote
949 parasite Encephalitozoon cuniculi. Nature 414:450–453.

950 86. Schimek C. 2011. Evolution of special metabolism in fungi: Concepts, mechanisms, and
951 pathways, p. 293–329. *In* Pöggeler, S, Wöstemeyer, J (eds.), Evolution of Fungi and Fungal-
952 Like Organisms. Springer, Berlin, Heidelberg.

953 87. Hu X, Xiao G, Zheng P, Shang Y, Su Y, Zhang X, Liu X, Zhan S, St. Leger RJ, Wang C.
954 2014. Trajectory and genomic determinants of fungal-pathogen speciation and host
955 adaptation. Proceedings of the National Academy of Sciences 111:16796–16801.

956 88. Gan P, Ikeda K, Irieda H, Narusaka M, O'Connell RJ, Narusaka Y, Takano Y, Kubo Y,
957 Shirasu K. 2013. Comparative genomic and transcriptomic analyses reveal the
958 hemibiotrophic stage shift of Colletotrichum fungi. New Phytologist 197:1236–1249.

959 89. Balancing selection for aflatoxin in *Aspergillus flavus* is maintained through interference
960 competition with, and fungivory by insects https://doi.org/10.1098/rspb.2017.2408.

961 90. Keller NP. 2019. Fungal secondary metabolism: regulation, function and drug discovery. Nat
962 Rev Microbiol 17:167–180.

963 91. Robey MT, Caesar LK, Drott MT, Keller NP, Kelleher NL. 2021. An interpreted atlas of
964 biosynthetic gene clusters from 1,000 fungal genomes. Proceedings of the National Academy
965 of Sciences 118:e2020230118.

966 92. Conlon BH, Gostinčar C, Fricke J, Kreuzenbeck NB, Daniel J-M, Schlosser MSL, Peereboom
967 N, Aanen DK, de Beer ZW, Beemelmanns C, Gunde-Cimerman N, Poulsen M. 2021.
968 Genome reduction and relaxed selection is associated with the transition to symbiosis in the
969 basidiomycete genus Podaxis. iScience 24:102680.

970 93. Engl T, Kroiss J, Kai M, Nechitaylo TY, Svatoš A, Kaltenpoth M. 2018. Evolutionary
971 stability of antibiotic protection in a defensive symbiosis. Proc Natl Acad Sci U S A
972 115:E2020–E2029.

973 94. Li YF, Tsai KJS, Harvey CJB, Li JJ, Ary BE, Berlew EE, Boehman BL, Findley DM, Friant
974 AG, Gardner CA, Gould MP, Ha JH, Lilley BK, McKinstry EL, Nawal S, Parry RC,
975 Rothchild KW, Silbert SD, Tentilucci MD, Thurston AM, Wai RB, Yoon Y, Aiyar RS,
976 Medema MH, Hillenmeyer ME, Charkoudian LK. 2016. Comprehensive curation and
977 analysis of fungal biosynthetic gene clusters of published natural products. Fungal Genet Biol
978 89:18–28.

979 95. de Vienne DM. 2019. Tanglegrams Are Misleading for Visual Evaluation of Tree
980 Congruence. Mol Biol Evol 36:174–176.

981 96. Mohammadi A, Amini Y. Molecular Characterization and Identification of Acrostalagmus
982 luteoalbus from Saffron in Iran 3.

983 97. Zhang GZ, Tang CY. 2015. First Report of Acrostalagmus luteo-albus Causing Red Rust of
984 Needle Mushroom (Flammulina velutipes) in China. Plant Dis 99:158–158.

985 98. Rubini MR, Silva-Ribeiro RT, Pomella AWV, Maki CS, Araújo WL, dos Santos DR,
986 Azevedo JL. 2005. Diversity of endophytic fungal community of cacao (*Theobroma cacao*
987 L.) and biological control of *Crinipellis perniciosa,* causal agent of Witches' Broom Disease.
988 Int J Biol Sci 24–33.

989 99. Venice F, Davolos D, Spina F, Poli A, Prigione VP, Varese GC, Ghignone S. 2020. Genome
990 Sequence of Trichoderma lixii MUT3171, A Promising Strain for Mycoremediation of PAH-
991 Contaminated Sites. Microorganisms 8:1258.

992 100. Xu M, Gessner G, Groth I, Lange C, Christner A, Bruhn T, Deng Z, Li X, Heinemann SH,
993 Grabley S, Bringmann G, Sattler I, Lin W. 2007. Shearinines D–K, new indole triterpenoids
994 from an endophytic Penicillium sp. (strain HKI0459) with blocking activity on large-
995 conductance calcium-activated potassium channels. Tetrahedron 63:435–444.

996 101. Currie CR. 2001. A community of ants, fungi, and bacteria: a multilateral approach to
997 studying symbiosis. Annu Rev Microbiol 55:357–380.

998

999