The small genome of the fungus *Escovopsis weberi*, a specialized disease agent of ant agriculture

SI Appendix

Includes:

Detailed Materials and Methods

Sample collection and DNA preparation. *Escovopsis weberi* strain CC031208-10 was isolated from an *Atta cephalotes* colony in Gamboa, Panama in 2003. Fungus was serially passaged on potato dextrose agar (PDA) plates with 50 mg/L each of penicillin and streptomycin. DNA was prepared by crushing lyophilized fungal tissue in liquid nitrogen using a CTAB-based protocol (Teknova, C2190) at 40ºC.

Phenotype array analysis. To investigate carbon source utilization*, E. weberi* CC031208-10 cultures were grown on 95 different carbon sources in 96-well plates using methods described by Druzhinina et al. (1).

Genome sequencing, assembly and size estimation. Sequencing was performed using the 454 FLX Titanium pyrosequencing platform with both fragment and paired end approaches. A total of 2.5 whole-genome shotgun fragment runs and a single 8kpb insert paired-end library run were generated. The raw dataset, including both single and paired end reads (average read length of 377 bp) is deposited at DDBJ/EMBL/GenBank under PRJNA253870, and the whole genome assembly is deposited under accession LGSR00000000. The version described here is version LGSR01000000.

The genome was assembled using the *De Novo* GS Assembler v 2.6 from the Newbler software package developed by Roche. Completeness of the genome assembly was assessed using three independent methods. First, we calculated basic statistics including total length and fragmentation of the assembled sequences. Second, we identified super conserved core eukaryotic genes (CEGs) in our genome assembly using CEGMA 2.4 (2); the number of identified CEGs is an indication for the completeness of an assembly. Third, we generated a frequency distribution of

unique 31-mers in the raw sequencing reads with Jellyfish 1.1.11 (3). K-mers with more than 12 copies in the genome, which are located to the right of the inflection point (Dataset S2), were included in the computation of genome size. All K-mers left of the inflection point were considered erroneous and discounted. Therefore, those k-mers considered erroneous were subtracted from the total volume of k-mers. The remaining k-mers were then divided by the maximum k-mer coverage (44X) to generate the genome size estimate.

Transcriptome assembly via RNA-seq. To facilitate genome annotation, and to compare RNA profiles of *E. weberi* strain CC031208-10 in isolation to RNA profiles when interacting with its host fungus, *Leucoagaricus gongylophorus*, RNA-seq analysis was performed. Following protocols used for *Trichoderma* interactions with its fungal hosts (4), fungal samples were grown on cellophane disks placed on Potato Dextrose Agar (PDA) plates with 50 mg/L each of penicillin and streptomycin. For interaction plates, *L. gongylophorus* was grown on the edge of cellophane-covered PDA plates for 5-7 days and then *Escovopsis* was added to the center of the plate. Plates were monitored daily for attraction to the host fungus (5), and tissue was scraped from the plate when (i) *Escovopsis* initially started to demonstrate attraction towards the cultivated fungus (its host), and when (ii) *Escovopsis* had completely overgrown its host. Mycelium and spores from *Escovopsis* only plates were scraped from plates when the final interaction plates were scraped.

Tissue samples for RNA extraction were crushed in liquid nitrogen, homogenized in Trizol, and incubated at room temperature for five minutes. Total RNA was extracted using chloroform, and the RNA was precipitated from the aqueous layer using isopropanol. The RNA pellet was additionally washed with 75% ethanol in DEPC-treated water and dissolved in DEPC-treated water. RNA quality and quantity were

assessed using a Nanodrop spectrophotometer and Bioanalyzer 2100. Subsequent cDNA library construction followed the standard Illumina TruSeq protocol starting with 1 ug of total RNA. A 100-cycle paired-end sequence run was performed using the Illumina HiSeq platform (Illumina, San Diego, CA).

Raw RNA-seq reads were filtered for low quality with Q20 (1 in 100 chance of incorrect base call) as a cutoff value. Quality of the reads was checked after the trimming process with FastQC (6), and the high quality reads were mapped against the genome assembly using Tophat 2.0.9 (insert size used = 200 bp) (7). The aligned reads were assembled into transcripts, and abundance values for each transcript were computed using Cufflinks 2.1.0, which was used to determine significant differential gene expression between treatments (8). Results were plotted using the R package CummeRbund 2.10 (9). RNA-seq reads are deposited in the Sequence Read Archive (SRA) under accession, SRP049545.

Gene discovery and annotation. Initial *ab initio* gene discovery was performed using MAKER 2.28 (10), which combines several prediction algorithms in order to generate an annotation. First, RepeatMasker 3.3.0 (11) was utilized to identify repeat families and low complexity DNA sequences in the genome using the Repbase database (version of 04-18-2012) (12) and a custom built database of repeats found in this genome. To provide transcriptional evidence for gene predictions, Illumina RNA-seq reads were mapped to the genome assembly using the GSNAP 2013-08-19 software (13) and assembled into contiguous transcript contigs with the Program to Assemble Spliced Alignments (PASA 2.0) tool (14). These transcripts were provided to MAKER 2 and were aligned to the genome with Exonerate 2.2.0 (15) to provide support for exons. Additional exon support was provided by alignment of all available *Trichoderma* ESTs from NCBI dbEST, proteins from *Trichoderma virens*, *T. atroviride*

and *T. reesei*, a collection of fungal proteomes, and the non-redundant (NR) database from NCBI (aligned with TBLASTN 2.2.25+ (16)). These alignments were polished into spliced alignments with Exonerate 2.2.0. Next, protein coding genes were predicted with *ab initio* gene predictors Augustus 2.7 (17), SNAP 0.15.4 (18) and GeneMark 2.5 (19) using exon hits from the protein and RNAseq transcript evidence. The generated gene prediction models from MAKER 2 were loaded into EVidenceModeler (EVM) (20) together with the protein and PASA transcript coding regions to calculate the set of consensus gene models for the genome. Protein sequences from *Trichoderma virens, T. atroviride* and *T. reesei* were aligned against the *E. weberi* genome using TBLASTN 2.2.25+ to detect possibly missed gene models from MAKER 2 and EVM. The genome annotation can be visualized with GBrowse (21) hosted at http://gb2.fungalgenomes.org/.

All predicted proteins were functionally annotated using InterProScan 5-44.0 (22), which outputs Gene Ontology (GO) terms (23) and protein domains from several secondary databases. Examination of domain enrichment focused on those from the Pfam (24) identifiers since these domains are widely used and well described. Protein-coding genes were further annotated by mapping them against the KEGG pathway database (25) using the KEGG Automated Annotation Server (KAAS) (26). The annotation server constructed putative metabolic pathways that are harbored within the *E. weberi* genome. Transfer RNA genes were predicted with tRNAscan-SE 1.3.1 (27).

We scanned the genome assembly and unmapped reads for repetitive sequences that share similarity with elements in Repbase (12) using RepeatMasker 3.3.0 (11). Additionally a *de novo* repeat identification was conducted using RepeatModeler 1.0.7 (28) on the genome assembly and unmapped reads. RepeatModeler utilizes

RECON 1.07 (29) and RepeatScout 1.0.5 (30) for deriving consensus models of repeat families. Microsatellites were identified with the Tandem Repeat Finder (TRF 4.04) (31).

Orthology analysis. To identify orthologous and paralogous genes, we used Inparanoid 4.1 (32), Multiparanoid (33) and OrthoMCL 2.0.9 (34) on our predicted *E. weberi* proteome and three proteomes of close relatives (*T. reesei, T. virens and T. atroviride*). The predicted protein family (PFAM) domains and their abundances in *E. weberi* were compared with PFAM domains within the *Trichoderma* species.

Repeat Induced Point mutation (RIP) analyses. Repeat-induced point mutation (RIP) is a fungal-specific defense mechanism that can heavily influence sequence variation within repetitive regions (35). During the sexual phase, duplicated sequences in the genome that are over 400 bp in length and share over 80% sequence similarity can be subjected to RIP. This irreversible process preferentially alters C:G to T:A nucleotides and acts mainly on transposable elements; however, protein-coding genes can also be a target (36). Ratios of TA/AT > 0.89 and (CA+TG/AC+GT) < 1.03 are considered evidence for RIP (37). Evidence for repeat induced point mutation (RIP) was assessed by computing these RIP indices for the five most prevalent repeat families within the *E. weberi* genome as well as the unmapped reads utilizing RIPCAL 1.0 (38). We also searched for orthologs in *E. weberi* of genes known to be involved in the RIP process in *N. crassa* (39). Prior to RIP detection analyses, unmapped reads sharing $\geq$ 99% sequence similarity were clustered by means of UCLUST 6.0 (40) to reduce read redundancy.

Mesosynteny. Since no chromosome information for *E. weberi* or *Trichoderma* spp. was available, we investigated micro mesosynteny between scaffolds. Syntenic

regions were anchored with orthologous genes, and these regions were detected with the orthology output of Inparanoid and a custom PERL script containing an algorithm based on previously published work (41). Syntenic positions were plotted using Circos 0.63 software (42).

Phylogenetic analyses. We conducted several phylogenetic analyses. First, phylogenetic relationships between *E. weberi* and other Pezizomycotina with sequenced genomes (Fig. 1b) were estimated via Bayesian analysis (two runs of 1,000,000 generations each, partitioned by gene in Mr. Bayes (43)) of three loci (amino acid alignments of RPB1, RPB2 and EF-1 alpha). Sequences were aligned using Muscle 3.8.31 (44). A Bayesian phylogeny was estimated in Mr. Bayes 3.2.5 (43) (1,000,000 generations comparing fixed rate models of evolution, all other parameters set at defaults; confirmed stationarity and discarded the first 25% of trees). Then, divergence times were estimated in BEAST v1.8.2 (45). A model of evolution was selected for each gene partition using Prottest3 (46) (LG+I+G substitution best fit for all gene partitions based on AIC criteria). Separate substitution models were used for each gene, while clock and tree models were linked across partitions. The fossil *Paleopyrenomycites devonicus* was used to calibrate the TMRCA of all taxa excluding *Saccharomyces cervesiae*, which served as an outgroup. This 410 My old fossil is the oldest known representative of the Pezizomycotina group (47-50). A normal prior truncated at 400 My was used, with mean=400 and sd=150. This constrained the age of the most recent common ancestor of the Pezizomycotina group to be at least as old as the *P. devonicus* fossil. A lognormal relaxed clock with an uninformative uniform prior distribution (min=1e-11, max=1e100) was used to allow for rate variation among lineages, and a birth/death prior accounting for incomplete sampling (51) was implemented as a tree

prior. The phylogeney inferred with MrBayes was used as a starting tree to seed the MCMC chain. The analysis was run 4 times for 50 million generations each, logging trees and parameters every 5,000 generations. Adequate convergence and mixing, estimated by ensuring that ESS values were greater than 200, was assessed using Tracer v1.5 (52). Runs were combined using LogCombiner v1.8.2 and a maximum clade credibility tree was generated using TreeAnnotator v1.8.2 after removing the initial 1,000 trees from each run as burn-in. Second, phylogenetic placement of *E. weberi* within the Hypocreales was estimated via Bayesian analyses of five loci (SSU rDNA, LSU rDNA, RPB1, RPB2 and EF-l alpha) following methods described in Spatafora et al (53). Third, phylogenetic placement of *E. weberi* CC031208-10 relative to other *Escovopsis* spp. and fungi in the newly named genus *Escovopsiodes* was estimated via Bayesian analysis of alignable portions of three nuclear loci (LSU rDNA, EF1-alpha, ITS) in Mr. Bayes (1,000,000 generations, data partitioned by gene, all other defaults; confirmed stationarity and discarded the first 25% of trees). Models of evolution were selected in MrModeltest 2.3 (54) (GTR+I+G for each gene, based on AIC).

Supplementary Figures

Figure S1. Growth of *Escovopsis weberi* in the presence and absence of its host, the ants' cultivated fungus. Bioassays on Potato Dextrose Agar (PDA) at room temperature. For the bioassay where the ants' cultivated fungus is present (top row), the edge of the plate was inoculated with the cultivated fungus, then one week later the centers of both plates were inoculated with *E. weberi.*

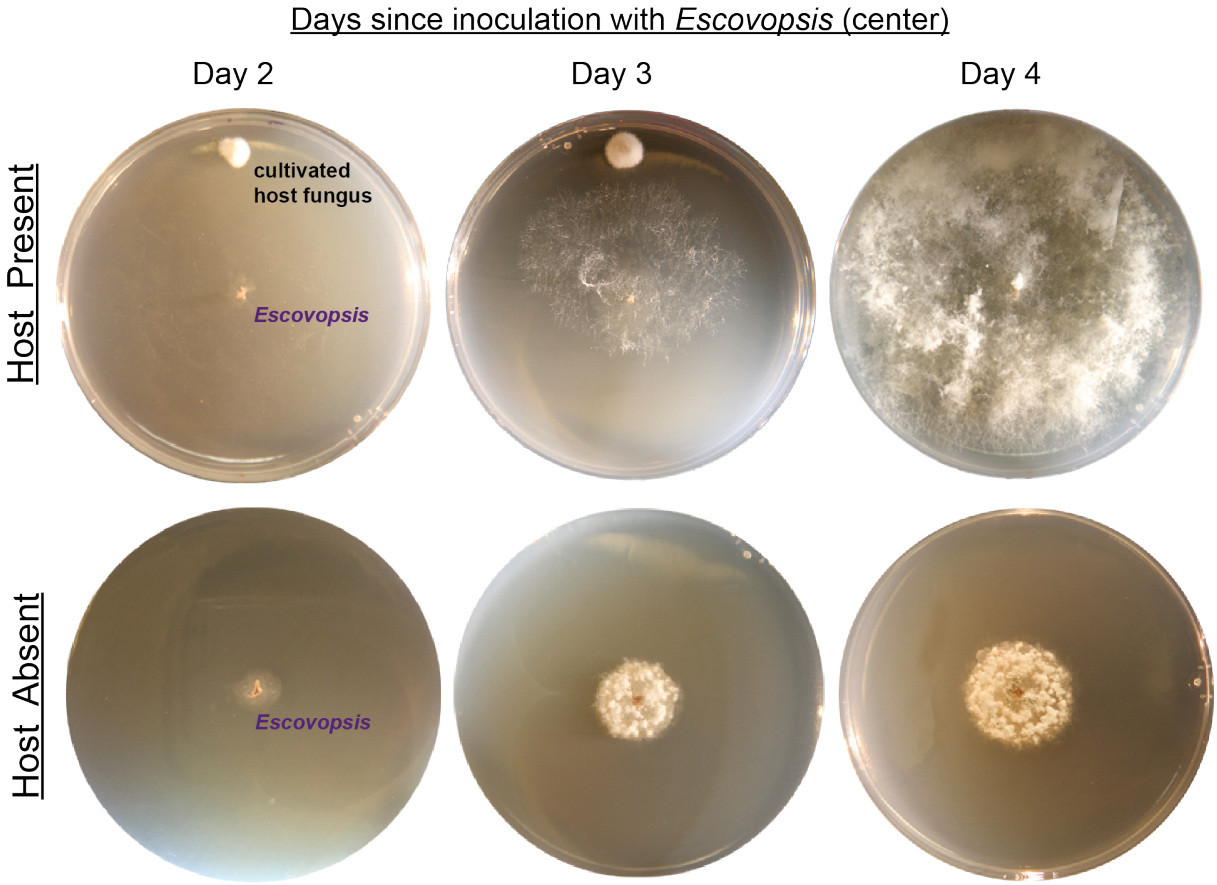
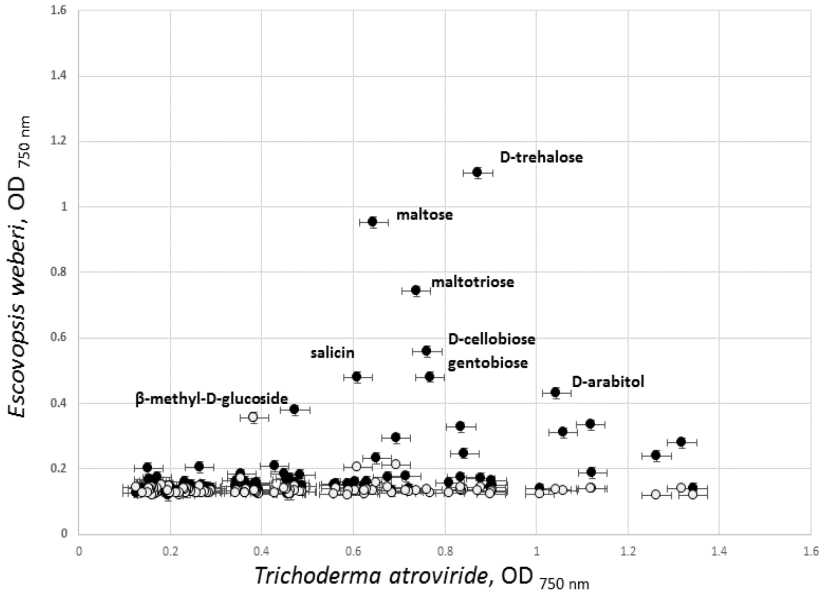


Figure S2. Growth of *Escovopsis weberi* on alternative carbon sources. a) Growth of *E. weberi* on alternative carbohydrate sources at 96 hours (white circles) and 240 hours (black circles) compared to *Trichoderma atroviride* at 96 hours. Note that at the same time point (96 hours, white circles), *E. weberi* shows less growth than *T. atroviride* on most carbon sources; one notable exception is beta-methyl-D-glucoside. b) *E. weberi* growth at two temperatures over time on representative carbon sources (water as control). The best carbon sources for *Escovopsis* growth were all disaccharides: D-trehalose and maltose at 25°C (dashed lines) and D-cellobiose at 30°C (solid lines). Data are means of four biological replicates, which differed by less than 12%. Notably, trehalose is the dominant carbohydrate constituent of *Escovopsis*' host fungus(55). Additional data in Dataset S6.

a



b

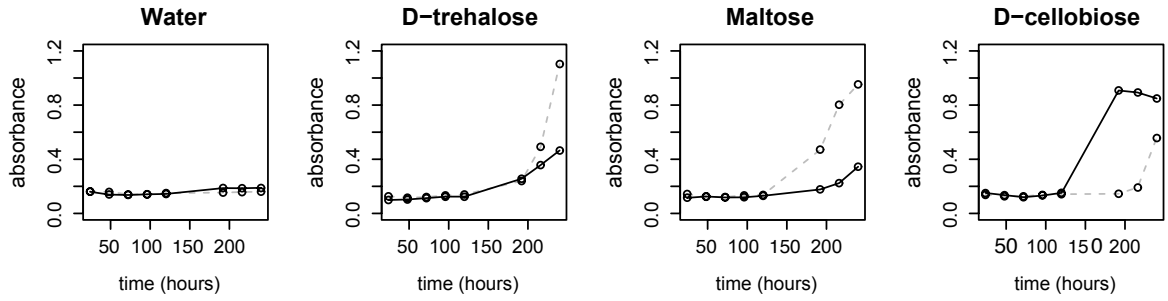


Figure S3. Phylogenetic placement of genome-sequenced strain of *Escovopsis weberi* relative to other *Escovopsis* species. Bayesian concensus tree generated using DNA sequence alignment of portions of ef1-alpha, ITS and LSU. All posterior probabilities are greater than 0.9 (not shown). Scale bar represents substitutions per site. Box indicates association with fungus-growing ant colonies of the Tribe Attini, some of which are leaf-cutting ants (*Acromyrmex* spp. and *Atta* spp.). Host ant species and country of origin (PA, Panama; BR, Brazil) indicated in parentheses. Detailed description of the morphology and phylogenetic placement of included *Escovopsis* and *Escovopsiodes* species, including NCBI accession numbers of sequences used, can be found in (56-58).

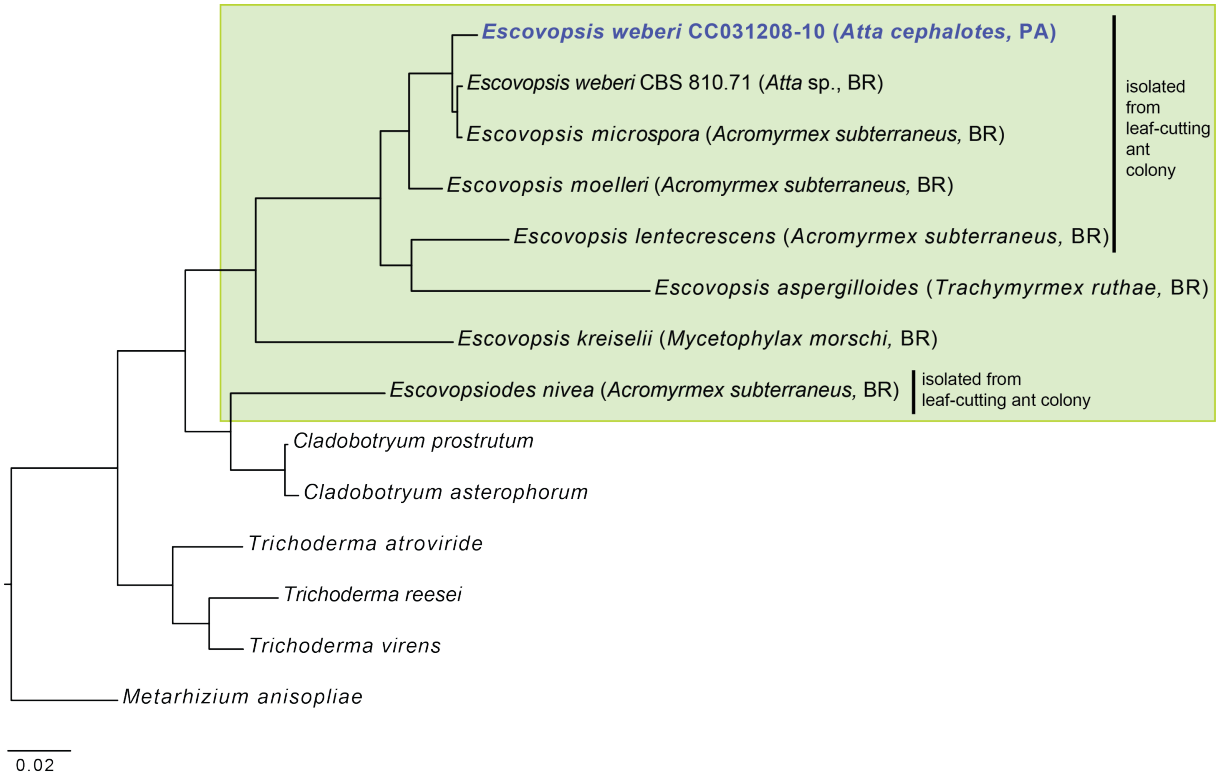


Figure S4. Repeat content in *Escovopsis weberi*. Unmapped reads sharing ≥ 99% sequence similarity were clustered by means of UCLUST for reducing sequencing read redundancy.

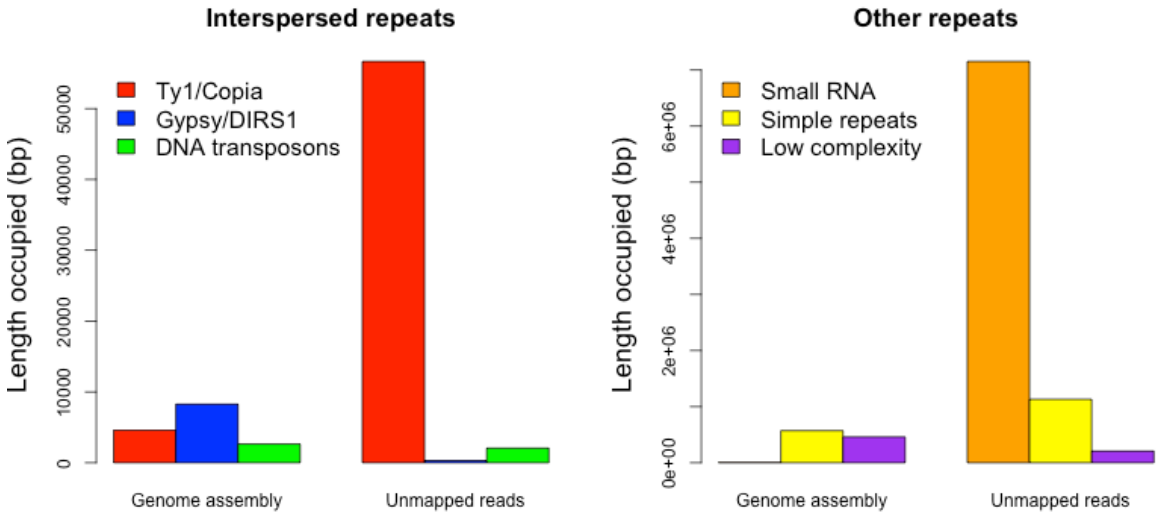


Figure S5. Genome Browser view of putative *MAT1-2* and *Mat1-1*-like loci. a) For *Mat1-2*, we identified a gene (ESCO_005440) with an HMG box homolog like that found in *Mat1-2-1*, but it falls within a locus with little similarity to *Mat1-2*. Specifically, the gene is not adjacent to a DNA lyase gene, a shared feature of this region in Sordariomycetes. Additionally, aside from the HMG box, the protein sequence has little similarity to that of other *Mat1-2-1*s, and indeed, based on BlastP, matches an unknown HMG-box protein from close relatives but no *Mat1-2* proteins; sequence similarity across *Mat1-2* make these genes easily identified in other ascomycetes using Blast (SI Appendix, Table S4). b) For *Mat1-1*, when examining the genome assembly scaffolds and gene expression data, a transcriptionally active *Mat1-1-1* gene was observed. Potential orthologs to *Mat1-1-2* or *Mat1-1-3*, however, do not appear to be transcriptionally active. These genes, while variable, are easily detected in sexual ascomycetes (SI Appendix, Table S4). Deletion of these two genes leads to strongly decreased fertility in *Neurospora crassa* (59) and complete arrest of fruiting-body development in *Podospora anserina* (60).

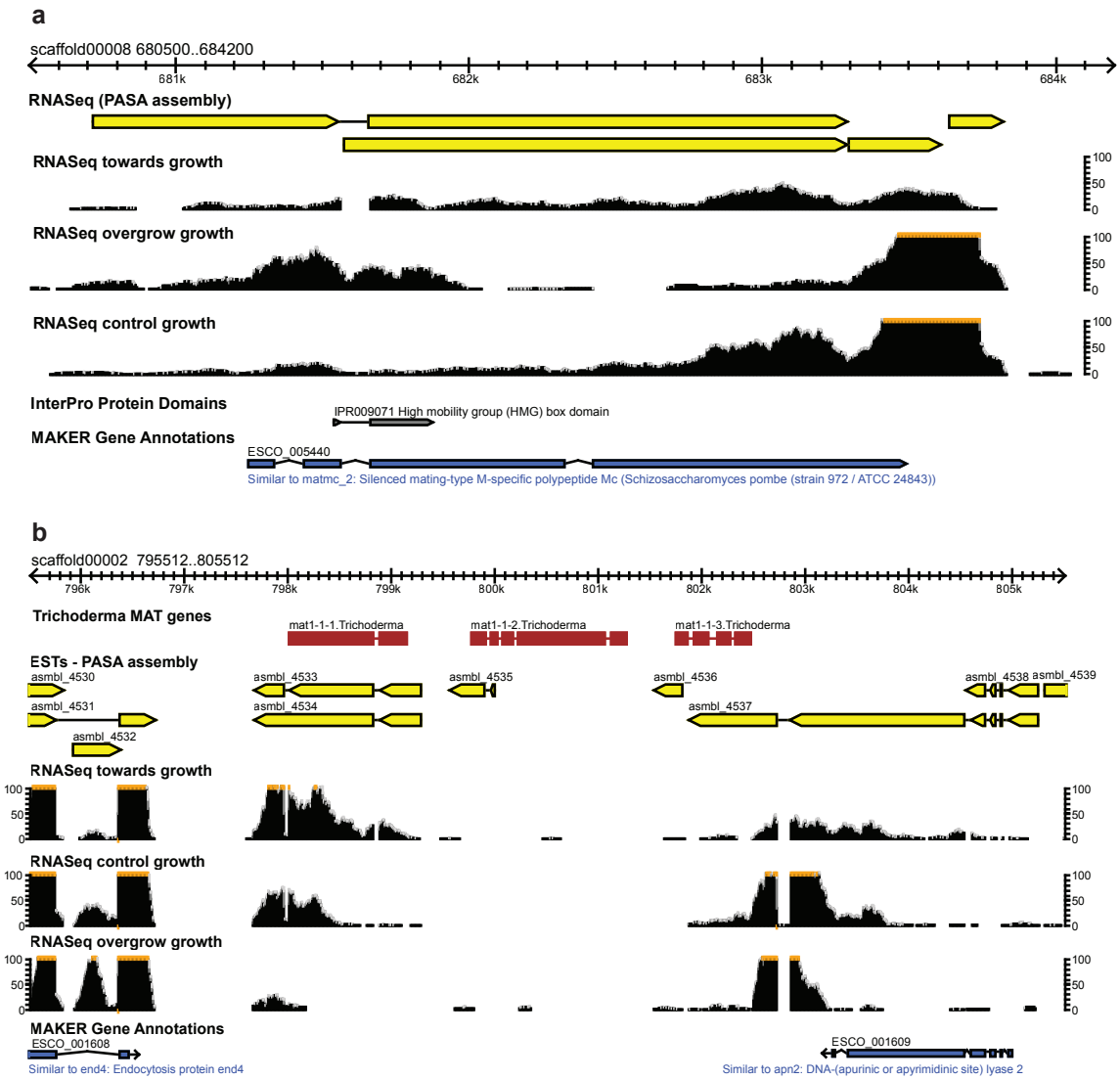


Figure S6. Phylogenetic placement of genome-sequenced *Escovopsis weberi* relative to other Hypocreales fungi. Bayesian analysis of *rpb1*, *rpb2* and *ef1-alpha* nuclear sequences. Triangle nodes indicate posterior probability > 0.98. Genome sizes and number of protein coding genes indicated for three *Trichoderma* species and for *E. weberi*.

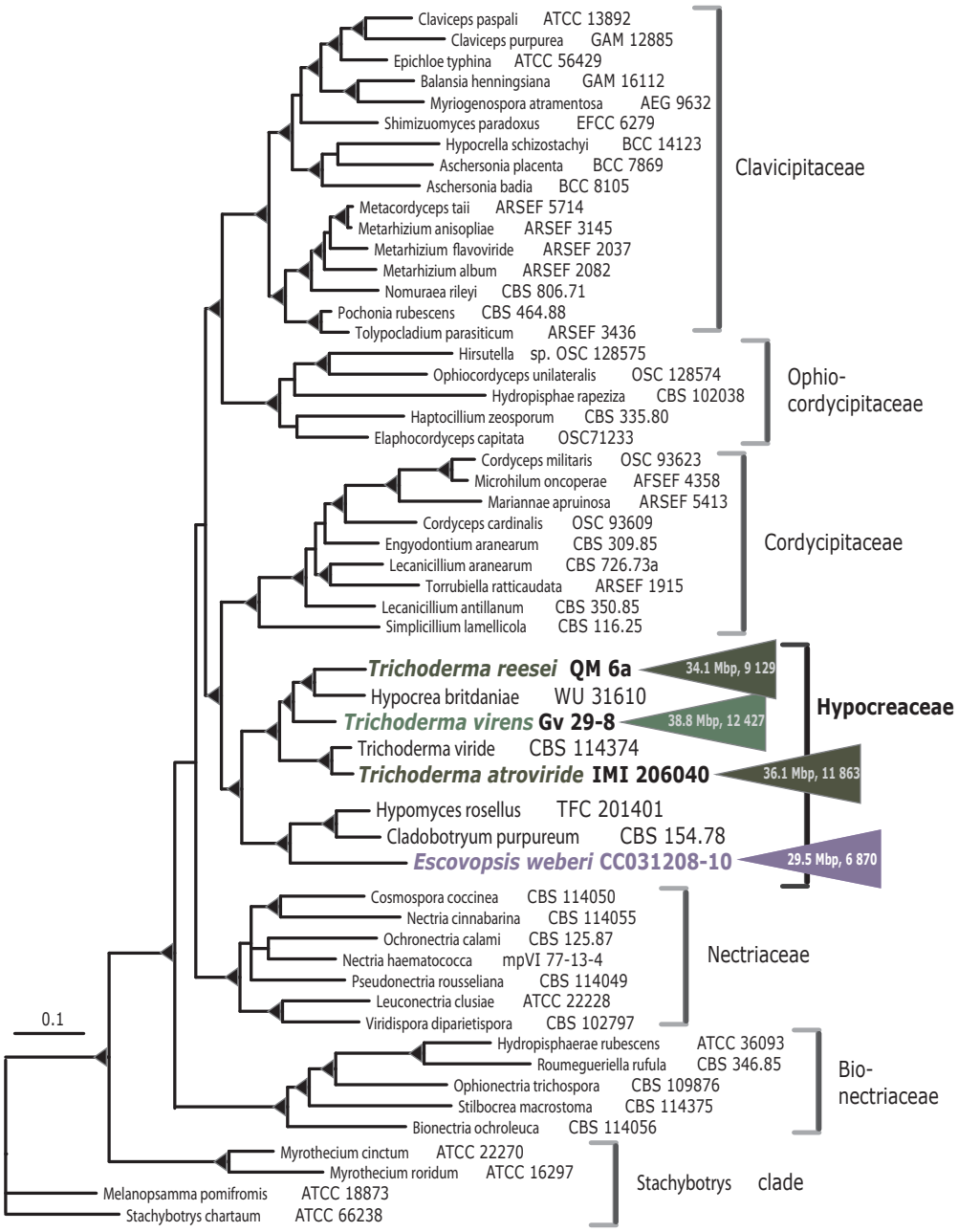


Figure S7. Age Estimates for the Divergence of *Escovopsis weberi* from other Pezizomycotina. Bayesian phylogeny estimated using *rpb1*, *rpb2* and *ef1-alpha* amino acid sequences. All posterior probabilities are greater than 0.98. Age estimates are given in millions of years before present, and node bars indicate the 95% highest posterior density interval surrounding the mean node age.

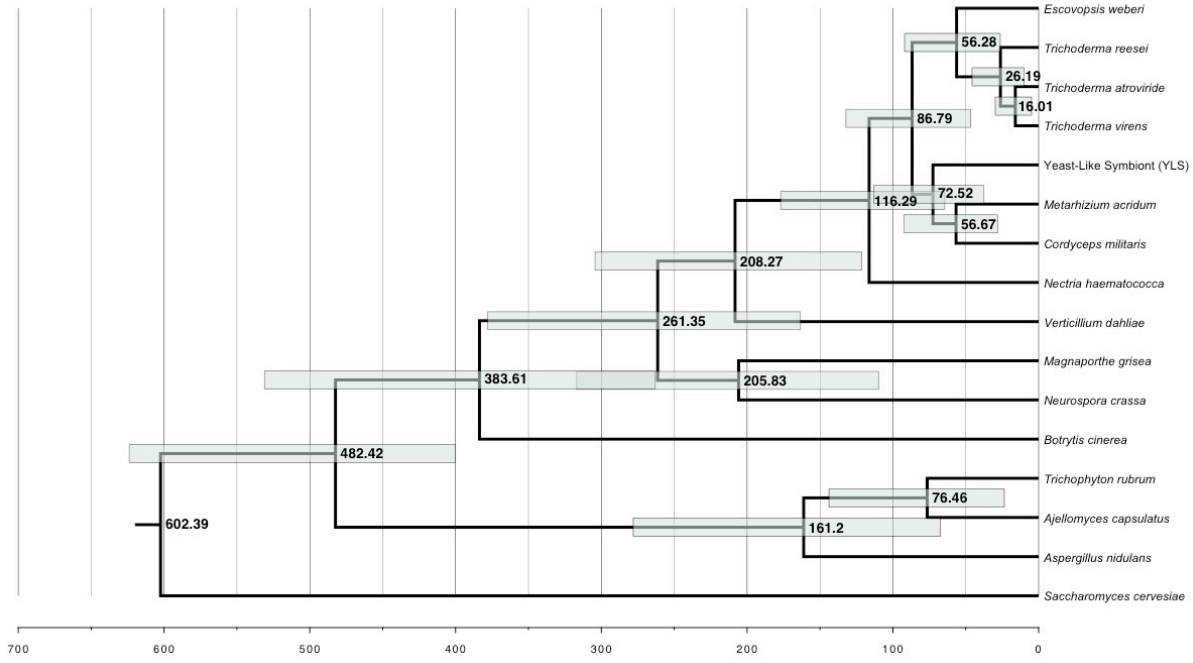


Figure S8. *Escovopsis weberi's* gene content overlap with the genomes of two other specialized fungal symbionts with small genomes. Comparison of the gene content of the *E. weberi* genome (29.5Mb, 6870 genes) with that of the human fungal pathogen *Trichophyton rubrum* (22.5Mb, 8707 genes), and the aphid-associated fungal endosymbiont YLS (25.4Mb, 6960 genes) based on orthology analysis indicates an overlapping core set that includes ~50% of *E. weberi's* genes.

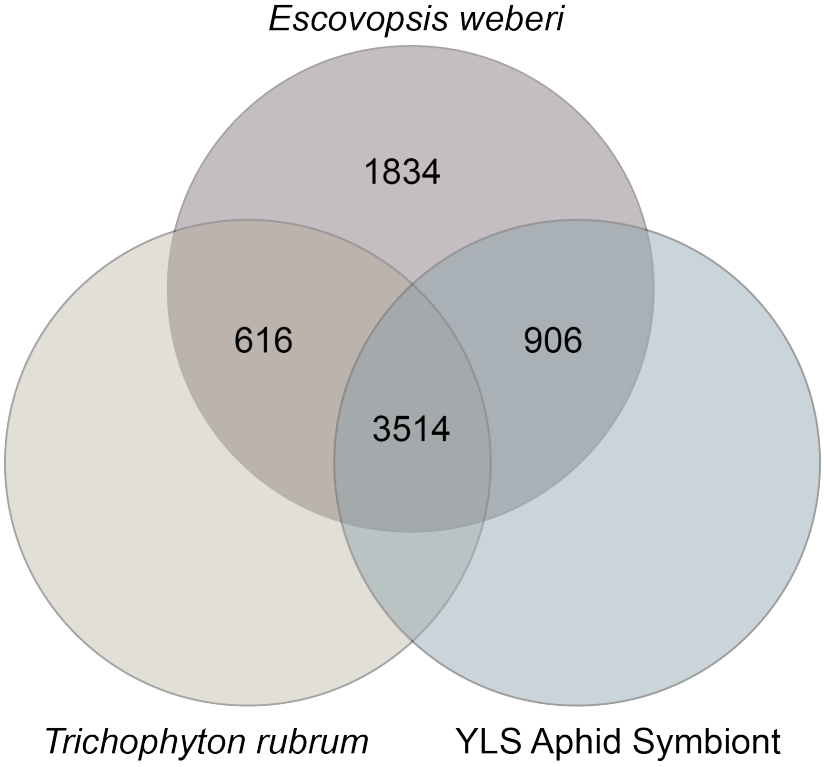
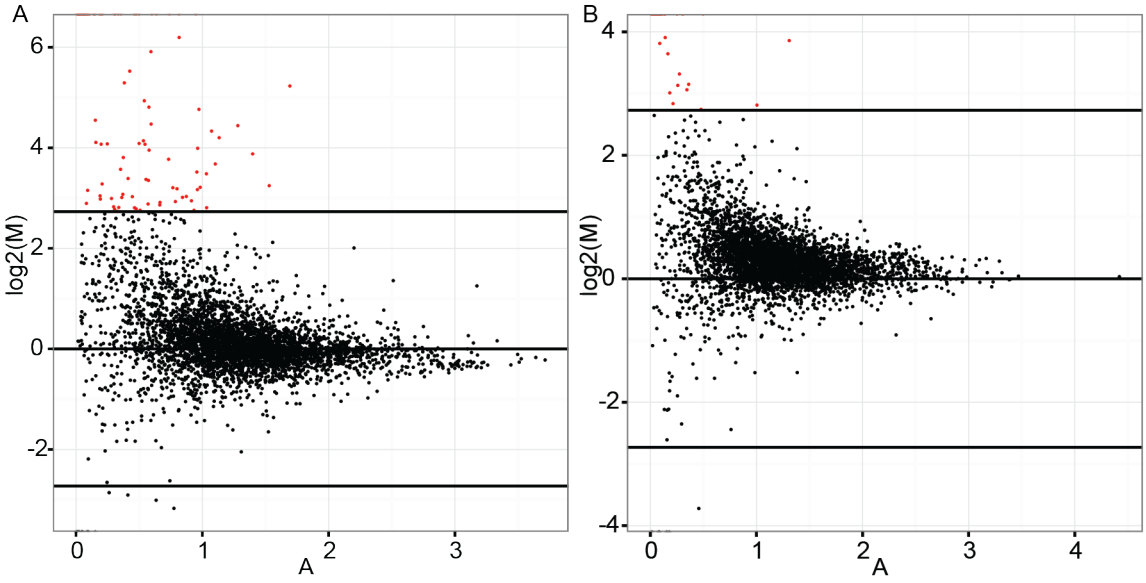


Figure S9. MA plot based on RNAseq expression data from *Escovopsis* grown in the absence of its host fungus and while growing towards its host (A) and in the absence of its host fungus and when overgrowing and consuming its host (B). To facilitate genome annotation and to begin to assess *Escovopsis*' response to host fungi, we conducted RNAseq transcriptome analysis using RNA isolated from *E. weberi* CC031208-10 under three conditions: growing on Potato Dextrose Agar (PDA) in the absence of the host (control), growing on media towards signals produced by host fungi, and overgrowing the host (see methods). Substantially more genes exhibited significantly increased expression when *E. weberi* was attracted to and growing towards the fungal cultivar compared to when it was consuming its host (Dataset S7). This is consistent with the hypothesis that *E. weberi* is producing necessary enzymes prior to feeding on the cultivar. Genes exhibiting significantly increased expression and high FPKM (Fragments Per Kilobase Of Exon Per Million Fragments Mapped) values when *E. weberi* was growing towards its host included one encoding for an alpha-glucosidase (ESCO_001987), which break down starch and disaccharides to glucose, major facilitator superfamily (MFS) transporters (ESCO_001467, ESCO_005830, ESCO_002122), and an endo-beta-1,4-glucanase (ESCO_000002). When feeding on its host, *E. weberi* exhibited significantly increased transcriptional levels in association with an alpha-glucosidase (ESCO_001987), and hydrolyzing enzymes (ESCO_003566), amongst other genes. In the MA plots below, log2(M) greater than zero indicates increased expression of the gene in the presence of its host fungus relative to in the absence of its host fungus. Results of follow-up qPCR with multiple biological replicates are in SI Appendix, Table S14.



Supplementary Tables

Table S1. General features of *Escovopsis weberi* compared to representative fungal genomes within the Pezizomycotina

Organism	Taxonomy (phylum, order)	Lifestyle	Size (Mb)	# genes	% GC	REF
Escovopsis weberi	**Ascomycota, Pezizomycotina Hypocreales**	**Mycoparasite of fungus-growing ant cultivars**	**29.5**	**6,870**	**55.74**	**this paper**
Trichoderma reesei	Ascomycota, Pezizomycotina Hypocreales	Mycoparasite, saprophyte	34.1	9,143	52	(61)
Trichoderma virens	Ascomycota, Pezizomycotina Hypocreales	Mycoparasite, saprophyte	38.8	12, 518	49.25	(62)
Trichoderma atroviride	Ascomycota, Pezizomycotina Hypocreales	Mycoparasite, saprophyte	36.1	11,865	49.75	(62)
YLS symbiont	Ascomycota, Pezizomycotina Hypocreales	Obligate aphid endosymbiont (specialist)	25.4	6,960	54	(63)
Cordyceps militaris	Ascomycota, Pezizomycotina Hypocreales	Entomopathogen (generalist)	32.2	9,684	51.4	(64)
Metarhizium acridum	Ascomycota, Pezizomycotina Hypocreales	Entomopathogen (grasshopper specialist)	38.1	9,849	49.9	(65)
Metarhizium anisopliae	Ascomycota, Pezizomycotina Hypocreales	Entomopathogen (generalist)	39.1	10,582	51.5	(65)
Nectria haematococca	Ascomycota, Pezizomycotina Hypocreales	Saprophyte, plant pathogen, animal pathogen	54.4	15,707	not reported	(66)
Verticillium dahliae	Ascomycota, Pezizomycotina Hypocreales	Plant pathogen	33.8	10,535	55.85	(67)
Fusarium graminearum	Ascomycota, Pezizomycotina Hypocreales	Plant pathogen	36.1	11,640	48.3	(68)
Magnaporthe grisea	Ascomycota, Pezizomycotina Magnaporhales	Plant pathogen (rice specialist)	39.4	12,841	52.0	(69)
Chaetomium thermophilum	Ascomycota, Pezizomycotina Sordariales	Thermophilic saprophyte	28.3	7,227	52.6	(70)
Neurospora crassa	Ascomycota, Pezizomycotina Sordariales	Saprophyte	38.7	10,620	49.6	(36)
Botrytis cinerea	Ascomycota, Pezizomycotina, Helotiales	Plant pathogen (generalist)	39.5 - 42.3	16,360-16,448	43.2	(71)
Sclerotinia sclerotiorum	Ascomycota, Pezizomycotina, Helotiales	Plant pathogen (generalist)	38.3	11,860	41.8	(71)
Trichophyton rubrum	Ascomycota, Pezizomycotina, Onygenales	Human pathogen	22.5	8,707	48.3	(72)
Microsporum canis	Ascomycota, Pezizomycotina, Onygenales	Zoophile (generalist)	23.1	8,915	47.5	(72)
Ajellomyces capsulatus	Ascomycota, Pezizomycotina, Onygenales	Human pathogen	33.0	9,390	42.8	(73)
Aspergillus nidulans	Ascomycota, Pezizomycotina Eurotiales	Saprophyte	30.1	10,701	50.3	(74)

Table S2. Biological assembly statistics for the *Escovopsis weberi* genome and its *Trichoderma* relatives. bp, base pairs; AA, amino acids.

	E. weberi	*T. reesei*	*T. virens*	*T. atroviride*
Number of Genes	**6870**	9143	12518	11865
Average exon per gene (bp)	**2.74**	3.06	2.98	2.93
Average gene length (bp)	**1623**	1793	1710	1747
Average exon length (bp)	**514**	508	506	528
Average intron length (bp)	**112**	120	105	104
Average protein length (AA)	**469**	492	479	471

Table S3. Fifteen most abundant PFAM domains in *Escovopsis weberi* and their associated abundances in *Trichoderma reesei* and *T. virens.* Of note, the genome encodes 42 proteins with ankyrin domains. Interestingly, *E. weberi* ankyrins belong to the PF12796 group, whereas *Trichoderma* ankyrins are members of PF00023 (62), which is depleted in *E. weberi*. Proteins containing the ankyrin domain can have a wide range of functions and have not been studied systematically in Pezizomycotina. Their presence in transcription factors of secondary metabolism, in a sensor for depletion of inorganic phosphate, and in the well known heterochromatin protein 1, however, have been reported (75-77). Also of note, the *E. weberi* genome encodes 35 proteins with the SEL1 domain (solenoid proteins, which are distinguished from general globular proteins by their modular architectures). SEL1 domain proteins can have diverse functions: the eukaryotic Sel1 and Hrd3 proteins are involved in ER-associated protein degradation, whereas bacterial LpnE, EnhC, HcpA, ExoR, and AlgK proteins mediate interactions between bacterial and eukaryotic host cells (78).

PFAM ID	Annotation	*Escovopsis*	*Trichoderma*	
		E. weberi	*T. reesei*	*T. virens*
PF00400	WD40 repeat	330	121	152
PF00069	Protein kinase domain	95	112	134
PF00153	Mitochondrial carrier	86	35	38
PF07690	MFS	83	180	255
PF00076	RNA-binding	74	62	62
PF00271	Helicase C	65	77	88
PF04082	Zn2Cys6 transcription factor, central	50	113	206
PF00005	ABC transporter	47	47	64
PF00270	DEAD box	46	53	53
PF00004	AAA+ ATPase	44	52	54
PF00083	Sugar transporter	43	142	227
PF12796	Ankyrin 2 repeat	42	0	0
PF00106	ADH short	40	142	227
PF00067	Cytochrome P450	40	71	118
PF08238	SEL-1 repeat	35	0	0

Table S4. Mating type genes (MAT) and Repeat Induced Point mutation (RIP) presence for several fungi.
MAT and RIP genes were identified using BLASTP and TBLASTN 2.3+ against a proteome or genome assembly, respectively. When a protein gene is present we recorded the protein length in amino acids (aa). Absence is indicated by a -.

Species	Order	Propagation	MAT1-1-1	MAT1-1-2	MAT1-1-3	MAT1-2	RIP
E. weberi	Hypocreales	Asexual	Not found in proteome but found on scaffold, see Fig S5 for details	-	-	-	Lacks several RIP genes
T. reesei	Hypocreales	Sexual	379 aa	434 aa	205 aa	241 aa	Contains all RIP genes
T. virens	Hypocreales	Sexual	-	-	-	371 aa	Contains all RIP genes
T. atroviride	Hypocreales	Sexual	-	-	-	235 aa	Contains all RIP genes
YLS	Hypocreales	Asexual?	168 aa (partial)	-	-	-	Lacks several RIP genes
C. militaris	Hypocreales	Sexual	456 aa	337 aa	-	239 aa	Contains all RIP genes
M. acridum	Hypocreales	Sexual	-	-	-	246 aa	Contains all RIP genes
M. anisopliae	Hypocreales	Sexual	376 aa	321 aa	193 aa	54 aa (partial)	Contains all RIP genes
N. haematococca	Hypocreales	Sexual	225 aa (partial)	436 aa	Na	150 aa (partial)	Contains all RIP genes
V. dahliae	Hypocreales	Sexual	421 aa	-	-	156 aa	Contains all RIP genes
F. graminearum	Hypocreales	Sexual	345 aa	446 aa	181 aa	253 aa	Contains all RIP genes, RIP tested in the lab
M. grisae	Magnaporhales	Sexual	327 aa	-	-	442 aa	Lacks all RIP genes
C. thermophilum	Sordariales	Sexual	-	-	-	420 aa	Contains all RIP genes
N. crassa	Sordariales	Sexual	293 aa	373 aa	-	382 aa	Contains all RIP genes
B. cinerea	Helotiales	Sexual	353 aa	-	-	376 aa	Contains all RIP genes
S. sclerotiorum	Helotiales	Sexual	361 aa	-	-	394 aa	Contains all RIP genes
T. rubrum	Onygenales	Sexual	384 aa	-	-	369 aa	Contains all RIP genes
M. canis	Onygenales	Sexual	380 aa	-	-	659 aa	Contains all RIP genes
A. capsulatus	Onygenales	Sexual	404 aa	-	-	421 aa	Contains all RIP genes
A. nidulans	Eurotiales	Sexual	361 aa	-	-	318 aa	Contains all RIP genes

Table S5. Presence and absence of RIP-associated genes in *Escovopsis weberi.* *Trichoderma reesei* identifiers are from JGI. Corresponding gene names in *Neurospora crassa*, the fungus in which the RIP mechanism was discovered, are provided for reference. The absence of several RIP associated genes (and lack of mating genes) suggests RIP is not possible in extant *E. weberi.*

Present in *Escovopsis*			
***T. reesei* ID**	***E. weberi* ID**	**Annotation**	***N. crassa* gene name**
49832	ESCO_000076	QDE2, Argonaute-like protein, essential for quelling	*qde-2*
111216	ESCO_001976	DIM5, Histone 3	*dim-5*
69494	ESCO_005318	DCL1, Dicer-like protein, involved in quelling	*dcl-1*
79823	ESCO_001305	DCL2, Dicer-like protein, involved in quelling	*dcl-2*
Not Found in *Escovopsis*			
57424	N/A	QIP, Putative exonuclease protein, involved in quelling	*qip*
67742	N/A	QDE1, RdRP, essential for quelling	*qde-1*
102458	N/A	QDE3, RecQ helicase, essential for quelling	*qde-3*
103470	N/A	SAD1, RdRP essential for MSUD	*sad-1*

Table S6. Identified paralogous genes in the *Escovopsis weberi* genome assembly. The confidence score, which shows how closely related it is to its seed ortholog (ESCO_001354), is listed for each paralog.

Gene ID	Inparanoid confidence score
ESCO_001354	1.0
ESCO_001356	0.18
ESCO_001357	0.07
ESCO_001359	0.11
ESCO_001360	0.05

Table S7. Evidence of RIP in most abundant repeat families. Top, five most abundant repeat families in unmapped reads. **Bottom,** five most abundant repeat families in assembly. Repeat families starting with the 'rnd' prefix are *de novo* predictions from RepeatModeler. Ratios of TA/AT > 0.89 and (CA+TG/AC+GT) < 1.03 are considered evidence for RIP.

Five Most Abundant Repeat Families, Unmapped Reads			
Name	**TA/AT**	**(CA+TG)/(AC+GT)**	**RIP evidence**
rnd-3_family-5	1.5	0.29	Yes
LSU rRNA	1	1	Yes
SSU rRNA	0.94	1.12	No
rnd-1_family-107	0.35	1.46	No
rnd-1_family-57	1.02	1.06	No
Five Most Abundant Repeat Families, Assembly			
rnd-2_family-41	1.75	0.07	Yes
Copia-14_BG-I-int	0.16	0.68	No
Gypsy-15_LBS-I-int	0.11	0.71	No
SUBTEL_sa	0	1.74	No
Copia-3_SCH-I-int	0.09	0.57	No

Table S8. Non-*Trichoderma* Pezizomycotina fungal genera to which *Escovopsis weberi* proteins show BLAST matches.

Genus	Number of genes	Taxonomy
Metarhizium	32	Hypocreales; Clavicipitaceae
Fusarium	16	Hypocreales; Nectriaceae
Colletotrichum	15	Glomerellales; Glomerellaceae
Beauveria	13	Hypocreales; Cordycipitaceae
Ophiocordyceps	10	Hypocreales; Cordycipitaceae
Cordyceps	4	Hypocreales; Ophiocordycipitaceae

Table S9. Gene categories of those found to be shared between the *E. weberi* genome and two small genomes based on orthology analysis (bidirectional best blast hits) relative to those found in *E. weberi* but not in the two other small genomes. Those genes unique to *Escovopsis* in the three-way, small genome comparison between *E. weberi,* YLS and *Trichophyton rubrum* are then further broken down into those found in *Trichoderma virens* and those not found in *T. virens.*
*****Enrichment indicates the relative higher number of genes in a given category that are shared between the three small genomes compared to those genes found in *Escovopsis* but not in either of the other two small genomes. For example, 3.7-fold enrichment of Zn2Cys6-associated genes is calculated by taking the number of these genes shared between *E. weberi*, YLS and *T. rubrum* (25) and dividing it by the total number of shared genes (3514). This 'shared' ratio is then divided by the 'unique' ratio, the number of these genes found in *Escovopsis* but not either of the two other genomes (48) divided by the total number of *Escovopsis*-unique genes (1834).

		Genes based on functional category						
	Total No.	**Zn2Cys6 genes**	**C2H2 genes**	**F-box genes**	**Chitinase genes**	**GH genes**	**Other genes**	**Unknown genes**
Genes shared between *Escovopsis*-YLS-*Trichophyton*	3514	25	10	2	3	21[a]	2397	1056
Gene in *Escovopsis* that are not found in either YLS or *Trichophyton*	1834	48	20	6	9	49	820	882
Found in *Trichoderma*	1064	37	17	0	7	47[b]	497	459
Not Found in *Trichoderma*	770	11	3	6	2	2	323	423
Enrichment *		3.7-fold	3.8-fold	5.7-fold	5.7-fold	4.5-fold		

[a] Glycoside Hydrolase (GH) families abundant in those shared between the three small genomes include GH13 and GH16.

[b] Glycoside Hydrolase (GH) families abundant in those shared between *Trichoderma virens* and *E. weberi* but not YLS or *Trichophyton rubrum* include GH3, GH5, GH12, GH18 and GH31.

Table S10. Gene families encoding polysaccharide depolymerizing enzymes (a.k.a., Carbohydrate Active Enzymes, CAZmyes). Data for *E. weberi* (Esco), *Trichoderma reesei* (Tr), *T. virens* (Tv), and *T. atroviride* (Ta) are based on genome assemblies. Numbers indicate number of each enzyme family. Gray highlights families significantly depleted in *Escovopsis* relative to *Trichoderma*. Proteomic, transcriptomic and draft genome sequencing have identified some of these enzymes in the ants' culitvated fungus, *Leucoagaricus gongylophorus*. For *L. gonylophorus* (the cultivar), cells are dark orange when it is known that at least one member of this family is present in the cultivar and is highly expressed and light orange when a member has been identified but there is no published evidence of it being highly expressed; white indicates no members have yet to be identified (79, 80). (A) Glycoside Hydrolases (GH), (B) Carbohydrate Binding Molecules (CBM), (C) Carbohydrate Esterases (CE), (D) Polysaccharide Lyases (PL) and (E) Axillary Activity (AA).

A) Glycoside Hydrolases (GH)

GH		1	2	3	5	6	7	9	10	11	12	13	15	16	17	18	20	24	25	27	28	29	30	31	35	36	37	38	43	44	45	47
	Esco	2	3	10	5	0	1	0	0	0	2	2	1	15	3	16	2	0	0	1	1	0	0	5	1	1	1	1	1	0	0	8
	Tr	2	3	21	8	1	2	0	1	3	2	4	2	16	1	15	3	1	1	4	4	0	4	4	1	1	2	1	3	0	1	8
	Tv	2	5	35	11	1	2	0	2	4	4	5	2	17	0	28	3	1	1	6	6	0	5	5	1	1	2	2	4	0	2	8
	Ta	4	10	15	12	1	2	0	1	4	3	4	3	17	4	28	3	0	1	8	6	0	5	7	1	1	2	1	7	0	1	8
	cultivar																															

GH		49	54	55	61	62	63	65	67	71	72	74	75	76	78	79	81	85	88	89	92	95	105	109	115	125	127	128	132
	Esco	1	0	4	0	0	2	1	1	1	4	1	1	6	0	1	1	0	1	1	5	3	0	7	1	2	1	3	2
	Tr	0	2	6	3	1	2	2	1	4	5	7	3	8	1	4	1	0	0	1	7	4	0	10	1	2	1	4	2
	Tv	1	2	9	3	3	2	2	2	5	6	8	5	9	3	4	2	0	3	1	7	4	0	19	1	3	1	5	2
	Ta	0	2	7	3	2	2	2	2	4	5	6	6	9	3	4	2	0	2	1	8	4	0	15	1	2	2	5	2
	cultivar																												

(continued on next page)

Table S10 Continued…

B) Carbohydrate Binding Molecules (CBM)

CBM		1	5	13	18	20	24	27	21	32	43	48	50	57	66
	Esco	5	0	3	4	2	1	0	1	0	1	1	7	0	1
	Tr	14	0	7	6	2	2	0	1	0	2	1	12	0	3
	Tv	24	0	12	22	2	4	0	1	0	3	2	29	0	8
	Ta	22	0	9	21	4	2	0	1	0	2	2	13	0	4
	cultivar														

C) Carbohydrate Esterases (CE)

CE		1	3	4	5	7	8	9	10	12	14	15
	Esco	14	1	3	2	2	0	1	19	1	1	0
	Tr	19	2	4	5	3	0	1	24	2	2	0
	Tv	36	4	4	7	1	2	1	59	2	2	0
	Ta	31	7	6	6	1	1	1	41	1	2	0
	cultivar											

D) Polysaccaride Lyases (PL)

PL		1	3	4	5	7	8
	Esco	2	0	0	2	1	1
	Tr	0	0	0	1	2	1
	Tv	0	0	0	2	3	1
	Ta	2	0	0	1	3	1
	cultivar						

(continued on next page)

Table S10 Continued…

E) Axillary Activity (AA)

AA		1	2	3	4	5	6	7	9
	Esco	2	7	5	2	1	1	15	1
	Tr	2	9	17	4	1	1	22	5
	Tv	4	10	21	3	1	1	40	4
	Ta	6	9	15	5	1	2	37	3
	cultivar								

Table S11. Seventeen putative secondary metabolite synthase clusters in *Escovopsis weberi* identified with antiSMASH2.0

Cluster	Type	Scaffold	Start	End	From	To
1	Terpene	1	ESCO_001120	ESCO_001125	4508687	4529730
2	T1pks	2	ESCO_001453	ESCO_001484	115408	224343
3	NRPS	2	ESCO_001520	ESCO_001536	367254	422171
4	Putative	2	ESCO_002065	ESCO_002097	2536720	2652079
5	Putative	2	ESCO_002308	ESCO_002333	3442148	3509314
6	T1pks	2	ESCO_002340	ESCO_002365	3543148	3656705
7	Putative	3	ESCO_002738	ESCO_002785	1402537	1548094
8	NRPS-T1pks	3	ESCO_003295	ESCO_003317	3505629	3603956
9	Terpene	4	ESCO_003535	ESCO_003540	640849	662308
10	Terpene	7	ESCO_005021	ESCO_005026	220627	241592
11	Putative	7	ESCO_005074	ESCO_005132	427818	628464
12	Putative	8	ESCO_005306	ESCO_005316	159054	190423
13	Terpene	9	ESCO_005701	ESCO_005738	569840	740562
14	T1pks-terpene	9	ESCO_005786	ESCO_005814	901165	993409
15	T1-pks	10	ESCO_005823	ESCO_005829	1310	49917
16	Terpene	10	ESCO_006046	ESCO_006052	834345	856793
17	Putative	11	ESCO_006081	ESCO_006108	2528	123057

Table S12. *Escovopsis weberi* genes containing a 'Common in several Fungal Extracellular Membrane' (CFEM) domain

Protein ID	Gene function
ESCO_002825	hypothetical protein
ESCO_004691	hypothetical protein
ESCO_001469	hypothetical protein (part of T1pks cluster)
ESCO_001980	hypothetical protein
ESCO_004195	hypothetical protein
ESCO_005464	hypothetical protein
ESCO_000678	hypothetical protein

Table S13. Fifteen most abundant PFAM domains in 1066 genes unique to *Escovopsis weberi* when comparing to three *Trichoderma* species.

PFAM ID	Annotation	Occurence
PF05730	CFEM	10
PF00069	Protein kinase domain	8
PF07728	AAA proteins	8
PF00076	RNA recognition motif	8
PF08238	Sel1 repeat	7
PF07690	MFS transporter	7
PF12796	Ankyrin2 repeat	7
PF00067	Cytochrome P450	7
PF00664	Transmembrane domain of ABC transporters	5
PF00172	Zinc finger domain	4
PF01822	WSC domain	4
PF13465	Zinc finger double domain	4
PF00083	Sugar transporter	4
PF00106	Short-chain dehydrogenase	4
PF00005	ATP-binding domain of ABC transporters	4

Table S14. Relative quantitative PCR (qPCR) of eight genes that exhibited significantly increased expression and high FPKM values in presence of the host relative to in the absence of the host in the RNAseq transcriptome analysis. Cellophane-PDA bioassays and RNA extractions were performed as for RNAseq using *E. weberi* CC031208-10 and four other strains of *Escovopsis* isolated from different *Atta spp.* gardens [CC030328-05 (*A. sexdens*), NMG030611-01 (*A. columbica*), SES030113-01 (*A. mexicana*), ST041017-01 (*A. columbica*)]. cDNA were prepared and primer efficiencies were tested following previously published protocols (81). Expression of the eight genes of interest was standardized relative to the endogenous control gene *ef1-alpha* using the delta-delta CT method. **Values listed are the average relative expression (RQ) values across three technical replicates**. **An RQ greater than one indicates increased expression in the presence of the host (either growing towards it or overgrowing it) relative to in the absence of the host.** Gene annotations are as follows: ESCO_001467 - Major Facilitator Superfamily (MSF) transporter; ESCO_001468 - FAD linked oxidase domain protein; ESCO_001469 - cell wall Thr-rich mannoprotein; ESCO_001509 - PutA NAD-dependent aldehyde dehydrogenase; ESCO_001987 - alpha-glucosidase; ESCO_002122 - MSF H+/oligopeptide transporter; ESCO_003842 - short chain dehydrogenase/reductase; ESCO_001413 - G protein–coupled receptor (GPCR), contains regulator of G protein signaling (RGS) domain. Interestingly, not all genes amplified in all samples, suggesting that presence or sequence of some of these loci may vary. Strains varied substantially in their expression patterns for the same gene. Overall, seven of the eight genes exhibited increased expression in the presence compared to the absence of the host in more than one of the strains tested.

	***Escovopsis* Strain**				
	CC031208-10	CC030328-05	NMG030611-01	SES030113-01	ST041017-01
Condition and Gene					
Growing Towards Host:					
ESCO_001467	1.42	did not amplify	did not amplify	did not amplify	5.17
ESCO_001468	1.08	did not amplify	did not amplify	did not amplify	1.10
ESCO_001469	3.41	7.31	did not amplify	did not amplify	8.32
ESCO_001509	0.98	1.83	1.04	did not amplify	1.00
ESCO_001987	1.04	3.83	0.25	2.17	2.17
ESCO_002122	1.27	5.17	did not amplify	0.94	2.49
ESCO_003842	1.10	1.50	1.40	1.24	0.30
When Overgrowing Host:					
ESCO_001413	2.20	did not amplify	did not amplify	did not amplify	8.56
ESCO_001987	0.50	1.51	0.51	2.17	2.17

Description of Supplementary Datasets Available as Excel Files

- **Dataset S1.** List of all tRNAs and their positions; tRNA gene intron locations are also listed. List of microsatellites and their locations, organized by scaffold.
- **Dataset S2.** *E. weberi* genome size estimation based on the 31-mer distribution in raw 454 reads.
- **Dataset S3.** Functional annotation of *E. weberi* genes based on: 1) best blast hits to *Trichoderma virens* genome (less than e-30 and 70% or more coverage), 2) best blast hits to *T. virens*, *T. atroviride* or *T. reesei* genome, 3) manual curation using
- **Dataset S4**. Interproscan analysis results. The output lists different protein signatures, including PFAM domains and GO terms.
- **Dataset S5**. Best hits for the 128 *E. weberi* proteins not found in *Trichoderma* but with hits to other Pezizomycotina fungi.
- **Dataset S6.** *Escovopsis* growth on alternative carbon sources. Includes comparative data for *T. atroviride.*
- **Dataset S7**. RNA-seq analysis results. Differently expressed genes when comparing growth towards host fungus to growth in absence of host, and differently expressed genes when comparing overgrowth of host fungus to growth in absence of host. Only those genes exhibiting significantly increased expression (positive fold change greater than two) in the presence than absence of the host are listed.
- **Dataset S8.** Genes found in *T. reesei* and *T. virens* that are not present in *E. weberi*.
- **Dataset S9**. SignalP analysis of secreted proteins. Hits without a transmembrane (TM) domain were accepted.

References

1. Druzhinina IS, Schmoll M, Seiboth B, Kubicek CP (2006) Global carbon utilization profiles of wild-type, mutant, and transformant strains of *Hypocrea jecorina. Appl Environ Microbiol* 72(3):2126–2133.

2. Parra G, Bradnam K, Korf I (2007) CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. *Bioinformatics* 23(9):1061–1067.

3. Marçais G, Kingsford C (2011) A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics* 27(6):764–770.

4. Seidl V, et al. (2009) Transcriptomic response of the mycoparasitic fungus *Trichoderma atroviride* to the presence of a fungal prey. *BMC Genomics* 10(1):567.

5. Gerardo NM, Jacobs SR, Currie CR, Mueller UG (2006) Ancient host-pathogen associations maintained by specificity of chemotaxis and antibiosis. *Plos Biol* 4(8):e235.

6. Andrews S Fastqc. Available at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

7. Kim D, et al. (2013) TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol* 14(4):R36.

8. Trapnell C, et al. (2010) Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol* 28(5):511–515.

9. Goff L, Trapnell C, Kelley DR cummeRbund: Analysis, exploration, manipulation, and visualization of Cufflinks high-throughput sequencing data. Available at: http://compbio.mit.edu/cummeRbund/.

10. Holt C, Yandell M (2011) MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. *BMC Bioinformatics* 12(1):491.

11. Smit AF, Hubley R, Green P RepeatMasker Open-4.0.1. Available at: www.repeatmasker.org.

12. Jurka J, et al. (2005) Repbase Update, a database of eukaryotic repetitive elements. *Cytogenet Genome Res* 110(1-4):462–467.

13. Wu TD, Nacu S (2010) Fast and SNP-tolerant detection of complex variants and splicing in short reads. *Bioinformatics* 26(7):873–881.

14. Haas BJ, et al. (2003) Improving the *Arabidopsis* genome annotation using maximal transcript alignment assemblies. *Nucleic Acids Res* 31(19):5654–5666.

15. Slater GSC, Birney E (2005) Automated generation of heuristics for biological

sequence comparison. *BMC Bioinformatics* 6:31.

16. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 215(3):403–410.

17. Stanke M, Waack S (2003) Gene prediction with a hidden Markov model and a new intron submodel. *Bioinformatics* 19 Suppl 2(Suppl 2):ii215–25.

18. Korf I (2004) Gene finding in novel genomes. *BMC Bioinformatics* 5:59.

19. Lomsadze A, Ter-Hovhannisyan V, Chernoff YO, Borodovsky M (2005) Gene identification in novel eukaryotic genomes by self-training algorithm. *Nucleic Acids Res* 33(20):6494–6506.

20. Haas BJ, et al. (2008) Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments. *Genome Biol* 9(1):R7.

21. Stein LD, et al. (2002) The generic genome browser: a building block for a model organism system database. *Genome Res* 12(10):1599–1610.

22. Zdobnov EM, Apweiler R (2001) InterProScan--an integration platform for the signature-recognition methods in InterPro. *Bioinformatics* 17(9):847–848.

23. Ashburner M, et al. (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* 25(1):25–29.

24. Finn RD, et al. (2013) Pfam: the protein families database. *Nucleic Acids Res* 42(D1):D222–D230.

25. Kanehisa M, Goto S, Sato Y, Furumichi M, Tanabe M (2012) KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Res* 40(Database issue):D109–14.

26. Moriya Y, Itoh M, Okuda S, Yoshizawa AC, Kanehisa M (2007) KAAS: an automatic genome annotation and pathway reconstruction server. *Nucleic Acids Res* 35(Web Server issue):W182–5.

27. Lowe TM, Eddy SR (1997) tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res* 25(5):955–964.

28. Smit AF, Hubley R RepeatModeler. Available at: www.repeatmasker.org.

29. Bao Z, Eddy SR (2002) Automated de novo identification of repeat sequence families in sequenced genomes. *Genome Res* 12(8):1269–1276.

30. Price AL, Jones NC, Pevzner PA (2005) De novo identification of repeat families in large genomes. *Bioinformatics* 21 Suppl 1(suppl 1):i351–8.

31. Benson G (1999) Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res* 27(2):573–580.

32. Remm M, Storm CE, Sonnhammer EL (2001) Automatic clustering of orthologs and in-paralogs from pairwise species comparisons. *J Mol Biol* 314(5):1041–1052.

33. Alexeyenko A, Tamas I, Liu G, Sonnhammer ELL (2006) Automatic clustering of orthologs and inparalogs shared by multiple proteomes. *Bioinformatics* 22(14):e9–15.

34. Li L, Stoeckert CJJ, Roos DS (2003) OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res* 13(9):2178–2189.

35. Galagan JE, Selker EU (2004) RIP: the evolutionary cost of genome defense. *Trends Genet* 20(9):417–423.

36. Galagan JE, et al. (2003) The genome sequence of the filamentous fungus *Neurospora crassa*. *Nature* 422(6934):859–868.

37. Margolin BS, et al. (1998) A methylated *Neurospora* 5S rRNA pseudogene contains a transposable element inactivated by repeat-induced point mutation. *Genetics* 149(4):1787–1797.

38. Hane JK, Oliver RP (2008) RIPCAL: a tool for alignment-based analysis of repeat-induced point mutations in fungal genomic sequences. *BMC Bioinformatics* 9:478.

39. Borkovich KA, et al. (2004) Lessons from the genome sequence of *Neurospora crassa*: tracing the path from genomic blueprint to multicellular organism. *Microbiol Mol Biol Rev* 68(1):1–108.

40. Edgar RC (2010) Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 26(19):2460–2461.

41. Hoberman R, Sankoff D, Durand D (2005) The statistical analysis of spatially clustered genes under the maximum gap criterion. *J Comput Biol* 12(8):1083–1102.

42. Krzywinski M, et al. (2009) Circos: an information aesthetic for comparative genomics. *Genome Res* 19(9):1639–1645.

43. Ronquist F, Huelsenbeck JP (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19(12):1572–1574.

44. Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32(5):1792–1797.

45. Drummond AJ, Rambaut A (2007) BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol* 7(1):214–8.

46. Darriba D, Taboada GL, Doallo R, Posada D (2011) ProtTest 3: fast selection of best-fit models of protein evolution. *Bioinformatics* 27(8):1164–1165.

47. Taylor TN, Hass H, Kerp H, Krings M, Hanlin RT (2005) Perithecial

ascomycetes from the 400 million year old Rhynie chert: an example of ancestral polymorphism. *Mycologia* 97(1):269–285.

48. Taylor TN, Hass H, Kerp H (1999) The oldest fossil ascomycetes. *Nature*.

49. Prieto M, Wedin M (2013) Dating the diversification of the major lineages of Ascomycota (Fungi). *PLoS ONE* 8(6):e65576.

50. Beimforde C, et al. (2014) Estimating the Phanerozoic history of the Ascomycota lineages: Combining fossil and molecular data. *Mol Phylogenet Evol* 78(C):386–398.

51. Stadler T (2009) On incomplete sampling under birth–death models and connections to the sampling-based coalescent. *J Theo Biol* 261(1):58–66.

52. Rambaut A, Suchard MA, Xie D, Drummond AJ (2014) Tracer. Available at: http://beast.bio.ed.ac.uk/Tracer.

53. Spatafora JW, et al. (2006) A five-gene phylogeny of Pezizomycotina. *Mycologia* 98(6):1018–1028.

54. Nylander JAA (2004) MrModeltest v2. Available at: https://github.com/nylander/MrModeltest2.

55. Martin MM, Carman RM, Macconnell JG (1969) Nutrients derived from the fungus cultured by the fungus-growing ant *Atta colombica tonsipes*. *Ann Entomol Soc Am* 62(1):11–13.

56. Augustin JO, et al. (2013) Yet more “Weeds” in the garden: fungal novelties from nests of leaf-cutting ants. *PLoS ONE* 8(12):e82265.

57. Meirelles LA, Montoya QV, Solomon SE, Rodrigues A (2015) New light on the systematics of fungi associated with attine ant gardens and the description of *Escovopsis kreiselii* sp. nov. *PLoS ONE* 10(1):e0112067.

58. Masiulionis VE, Cabello MN, Seifert KA, Rodrigues A, Pagnocca FC (2015) *Escovopsis trichodermoides* sp. nov., isolated from a nest of the lower attine ant *Mycocepurus goeldii*. *Antonie van Leeuwenhoek*. doi:10.1007/s10482-014-0367-1.

59. Ferreira AV, An Z, Metzenberg RL, Glass NL (1998) Characterization of mat A-2, mat A-3 and deltamatA mating-type mutants of *Neurospora crassa*. *Genetics* 148(3):1069–1079.

60. Berteaux-Lecellier V, Silar P, Debuchy R (2010) Mating systems and sexual morphogenesis in ascomycetes. *Cellular and Molecular Biology of Filamentous Fungi*, eds K B, D E (Washington, D.C.), pp 501–535.

61. Martinez D, et al. (2008) Genome sequencing and analysis of the biomass-degrading fungus *Trichoderma reesei* (syn. *Hypocrea jecorina*). *Nat Biotechnol* 26(5):553–560.

62. Kubicek CP, et al. (2011) Comparative genome sequence analysis underscores mycoparasitism as the ancestral life style of *Trichoderma*. *Genome Biol* 12(4):R40.

63. Vogel KJ, Moran NA (2013) Functional and evolutionary analysis of the genome of an obligate fungal symbiont. *Genome Biol Evol* 5(5):891–904.

64. Zheng P, et al. (2011) Genome sequence of the insect pathogenic fungus *Cordyceps militaris*, a valued traditional chinese medicine. *Genome Biol* 12(11):R116.

65. Gao Q, et al. (2011) Genome sequencing and comparative transcriptomics of the model entomopathogenic fungi *Metarhizium anisopliae* and *M. acridum*. *PLoS Genet* 7(1):e1001264.

66. Coleman JJ, et al. (2009) The genome of *Nectria haematococca*: contribution of supernumerary chromosomes to gene expansion. *PLoS Genet* 5(8):e1000618.

67. Klosterman SJ, et al. (2011) Comparative genomics yields insights into niche adaptation of plant vascular wilt pathogens. *PLoS Pathog* 7(7):e1002137.

68. Cuomo CA, et al. (2007) The *Fusarium graminearum* genome reveals a link between localized polymorphism and pathogen specialization. *Science* 317(5843):1400–1402.

69. Dean RA, et al. (2005) The genome sequence of the rice blast fungus *Magnaporthe grisea*. *Nature* 434(7036):980–986.

70. Amlacher S, et al. (2011) Insight into structure and assembly of the nuclear pore complex by utilizing the genome of a eukaryotic thermophile. *Cell* 146(2):277–289.

71. Amselem J, et al. (2011) Genomic analysis of the necrotrophic fungal pathogens *Sclerotinia sclerotiorum* and *Botrytis cinerea*. *PLoS Genet* 7(8):e1002230.

72. Martinez DA, et al. (2012) Comparative genome analysis of *Trichophyton rubrum* and related dermatophytes reveals candidate genes involved in infection. *mBio* 3(5):e00259–12–e00259–12.

73. Sharpton TJ, et al. (2009) Comparative genomic analyses of the human fungal pathogens Coccidioides and their relatives. *Genome Res* 19(10):1722–1731.

74. Galagan JE, et al. (2005) Sequencing of *Aspergillus nidulans* and comparative analysis with *A. fumigatus* and *A. oryzae*. *Nature* 438(7071):1105–1115.

75. Gras DE, et al. (2009) Transcriptional changes in the nuc-2A mutant strain of *Neurospora crassa* cultivated under conditions of phosphate shortage. *Microbiol Res* 164(6):658–664.

76. Pedley KF, Walton JD (2001) Regulation of cyclic peptide biosynthesis in a

plant pathogenic fungus by a novel transcription factor. *Proc Natl Acad Sci USA* 98(24):14174–14179.

77. Lorentz A, Heim L, Schmidt H (1992) The switching gene swi6 affects recombination and gene expression in the mating-type region of *Schizosaccharomyces pombe*. *Mol Gen Genet* 233(3):436–442.

78. Mittl PRE, Schneider-Brachert W (2007) Sel1-like repeat proteins in signal transduction. *Cell Signal* 19(1):20–31.

79. Grell MN, et al. (2013) The fungal symbiont of *Acromyrmex* leaf-cutting ants expresses the full spectrum of genes to degrade cellulose and other plant cell wall polysaccharides. *BMC Genomics* 14:928.

80. Aylward FO, et al. (2013) *Leucoagaricus gongylophorus* produces diverse enzymes for the degradation of recalcitrant plant polymers in leaf-cutter ant fungus gardens. *Appl Environ Microbiol* 79(12):3770–3778.

81. Altincicek B, Kovacs JL, Gerardo NM (2012) Horizontally transferred fungal carotenoid genes in the two-spotted spider mite *Tetranychus urticae*. *Biol Lett* 8(2):253–257.