bioRxiv preprint doi: https://doi.org/10.1101/375105; this version posted July 24, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

# Discovery of psychoactive plant and mushroom alkaloids in ancient fungal cicada pathogens

Greg R. Boyce[1], Emile Gluck-Thaler[2], Jason C. Slot[2], Jason E. Stajich[3], William J. Davis[4], Tim Y. James[4], John R. Cooley[5,6], Daniel G. Panaccione[1], Jørgen Eilenberg[7], Henrik H. De Fine Licht[7], Angie M. Macias[1], Matthew C. Berger[1], Kristen L. Wickert[1], Cameron M. Stauder[1], Ellie J. Spahr[1], Matthew D. Maust[1], Amy M. Metheny[1], Chris Simon[5], Gene Kritsky[8], Kathie T. Hodge[9], Richard A. Humber[9,10], Terry Gullion[11], Dylan P. G. Short[12], Teiya Kijimoto[1], Dan Mozgai[13], Nidia Arguedas[14], Matt T. Kasson[1,*]

[1]Division of Plant and Soil Sciences, West Virginia University, Morgantown, West Virginia, 26506, USA.

[2]Department of Plant Pathology, The Ohio State University, Columbus, Ohio 43210, USA.

[3]Department of Microbiology and Plant Pathology and Institute for Integrative Genome Biology, University of California, Riverside, California 92521, USA.

[4]Department of Ecology and Evolution, University of Michigan, Ann Arbor, Michigan 48109, USA.

[5]Department of Ecology and Evolutionary Biology, University of Connecticut, Storrs, Connecticut 06269, USA.

[6]College of Integrative Sciences, Wesleyan University, Middletown Connecticut USA.

[7]Department of Plant and Environmental Science, University of Copenhagen, Denmark.

[8]Department of Biology, Mount St. Joseph University, Cincinnati, Ohio 45233, USA.

[9] Plant Pathology & Plant-Microbe Biology, School of Integrative Plant Science, Cornell University, Ithaca, New York 14853, USA.

[10]USDA-ARS-NAA-BioIPM, Ithaca, New York 14853, USA.

bioRxiv preprint doi: https://doi.org/10.1101/375105; this version posted July 24, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



[11]Department of Chemistry, West Virginia University, Morgantown, West Virginia, 26506, USA.

[12]Amycel Spawnmate, Royal Oaks, California, 95067, USA.

[13]Cicadamania.com, Sea Bright, New Jersey, 07760 USA.

[14]Cleveland Metroparks, Cleveland, Ohio, USA.

***Corresponding author. Email:** mtkasson@mail.wvu.edu

**Abstract:**

Entomopathogenic fungi routinely kill their hosts before releasing infectious conidia, but select species keep their hosts alive while sporulating to enhance spore dispersal. Recent expression and metabolomics studies involving "host-killing" entomopathogens have helped unravel infection processes and host responses, yet the mechanisms underlying "active host transmission" in insects with Entomophthoralean fungal infections are completely unexplored. Here we report the discovery, through global and targeted metabolomics supported by metagenomics and proteomics, of the plant amphetamine, cathinone, in *Massospora cicadina*-infected periodical cicadas, and the mushroom tryptamine, psilocybin, in *M. platypediae*- and *M. levispora*-infected annual cicadas. The neurogenic activities of these alkaloids provide a hypothetical framework for a chemically induced extended phenotype of *Massospora* that alters cicada behavior by increasing endurance and suppressing feeding prior to death.

## Introduction

The Entomophthorales (Zoopagomycota) comprise some of the most important insect- and arthropod-destroying fungi, including *Massospora*, a genus of more than a dozen obligate, sexually transmissible pathogens of at least 21 reported cicada (Hemiptera) species worldwide (*1-5*). *Massospora* infections are conspicuous: the destruction of the intersegmental membranes incites a progressive sloughing off of abdominal sclerites, revealing an expanding mass ("plug") of infective conidia emerging from the posterior end of the cicada (*3,5*; Fig. 1A-2A, Movie S1). Over time, spores are passively disseminated during mating attempts or flights, behaviors that do not

bioRxiv preprint doi: https://doi.org/10.1101/375105; this version posted July 24, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



appear diminished as a result of infection (*2,6*). Several studies have reported hypersexual behaviors in *Massospora*-infected cicadas, where male cicadas, in addition to typical mating behavior, also attract copulation attempts from conspecific males (*5,7;* table S1). This "extended phenotype" of *Massospora* hijacks cicadas, turning them into efficient vectors for conidial transmission (*5*).

The most studied extended phenotype among entomopathogenic fungi is "summit disease" behavior, where fungi manipulate infected insects to ascend and affix to elevated substrates prior to death to maximize postmortem spore dissemination (*8-9*). Recent functional studies of the Hypocrealean summit-disease-causing zombie ant fungus, *Ophiocordyceps unilateralis*, have uncovered possible associations between host behavior modifications and fungus-derived neurogenic metabolites and enterotoxins (*10-11*). Genomic, transcriptomic, and metabolomic approaches have further helped elucidate important aspects of the evolution of summit-disease-causing Entomophthoralean fungi including expansion of genes related to host specificity (*12-13*) and changes in gene expression that coincide with disease progression (*14-15*). Nevertheless, many of the basic mechanisms that drive summit disease and other behavioral manipulations among Entomophthoralean fungi, including "active host transmission" in *Massospora* (*8-9*), remain unknown.

The inability to axenically culture a majority of the Entomophthorales necessitates the use of sensitive approaches using well-preserved, infected host materials to help elucidate the mechanisms underlying host behavior modification. Mass spectrometry, with its low detection limits and absolute quantification capability, is ideally suited for analyzing fungal biomolecules from complex biological substrates (*16*). Here, we used a combination of global and targeted liquid chromatography–mass spectrometry (LC-MS)-based metabolomics to compare three morphologically (*2-3*) and phylogenetically characterized *Massospora* species (*17*; fig. S1-S2, table S2) to identify candidate small molecules potentially contributing to a hypersexual behavioral phenotype of *Massospora* (*5*). Additionally, we sought to identify candidate molecules that may contribute more broadly to active host transmission described from several Entomophthoralean fungi including *Massospora* (*8-9,18-19*).

## Results

bioRxiv preprint doi: https://doi.org/10.1101/375105; this version posted July 24, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



We first performed LC-MS-based global metabolomics on conidial plugs excised from *M. cicadina*-infected periodical cicadas (*Magicicada* spp., Fig. 1A) and *M. platypediae*-infected wing-banger cicadas (*Platypedia putnami*; Fig. 2A), as well as posterior abdominal segments from healthy cicada controls of both species (Fig. 1B). Each *Massospora* species and the pooled controls produced unique metabolome signatures, with the two *Massospora* species displaying more similar metabolite profiles (fig. S3), with 1,176 (positive and negative ion mode) differentially detected small molecules between the two *Massospora* spp. including several notable monoamine alkaloids (*17*).

In *M. cicadina* but not *M. platypediae* conidial plugs or healthy cicada controls, we detected cathinone, an amphetamine previously reported from *Catha edulis*, a plant native to Africa and the Arabian Peninsula (*20*), along with three of its precursors (trans-cinnamic acid, benzoic acid, and 1-phenylpropane-1,2-dione) (Fig. 1B-C). A second global comparison between *M. cicadina* resting spores and conidia (fig. S4) revealed that resting spore plugs also contained cathinone, although in significantly less quantities than conidial plug counterparts (*17*). A fourth precursor, benzaldehyde, a common fungal metabolite (*21*), was upregulated in both *M. cicadina* and *M. platypediae* compared to the healthy cicada control (Fig. 1B-C). The presence of these specific intermediates coupled with the absence of Coenzyme A (CoA)-dependent intermediates (*17*) suggests *M. cicadina* may produce cathinone through the non-β-oxidative CoA-independent biosynthetic pathway (*20*; Fig. 1B-C). Fragmentation patterns for cathinone along with its intermediates were consistent with previously reported results (PubChem Compound Identifier: 62258), with cathinone fragmentation also experimentally observed for this study using a commercially available analytical standard (table S3).

bioRxiv preprint doi: https://doi.org/10.1101/375105; this version posted July 24, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



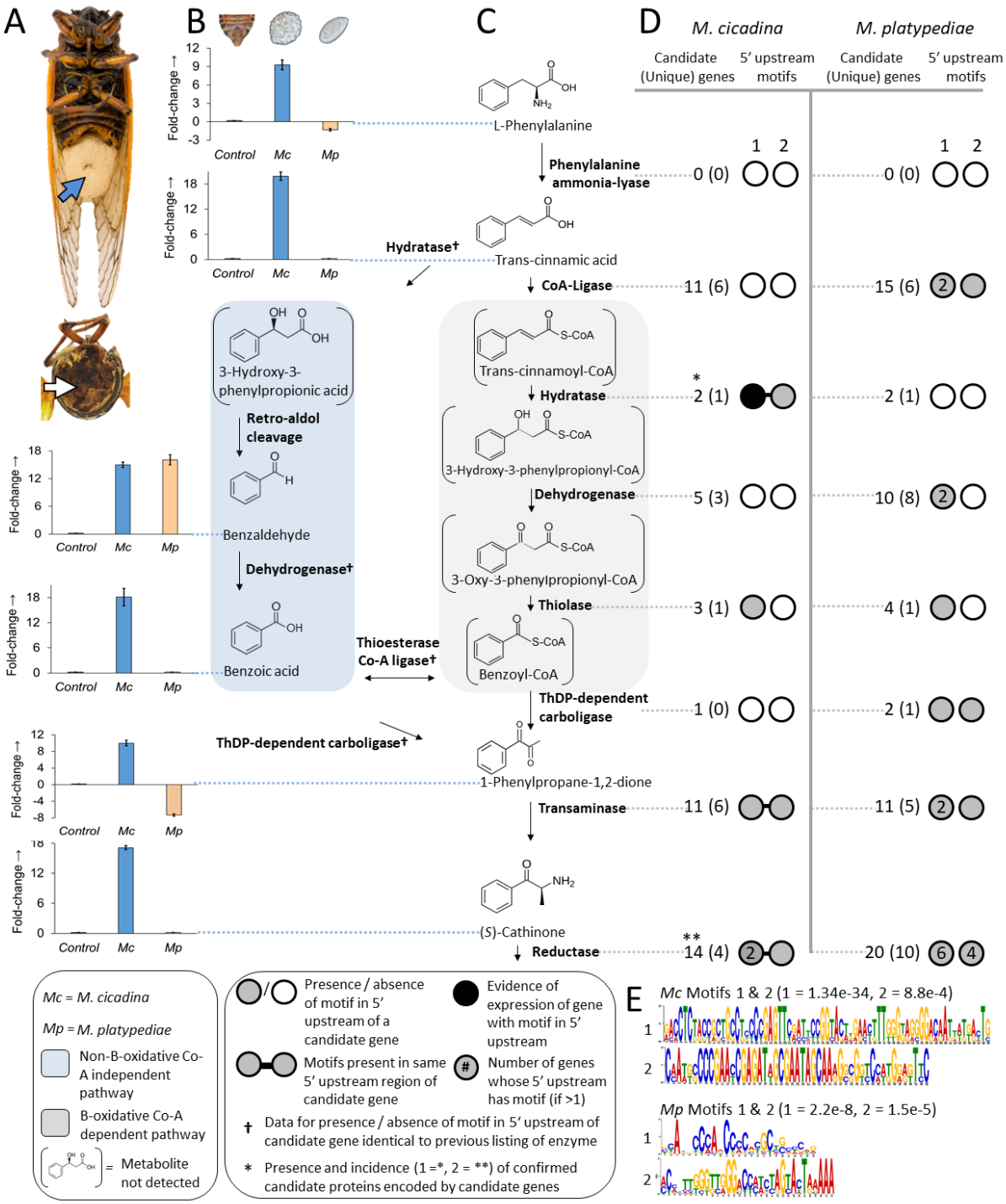


**Fig. 1. Production of the amphetamine cathinone and its proposed biosynthetic pathway in *Massospora cicadina*.**

(**A**) *M. cicadina*-infected periodical cicada (*Magicicada septendecim*) with a conspicuous conidial "plug" (blue arrow) and an internal resting spore infection (posterior cross-section, white arrow). (**B**) Global metabolomics fold-change comparisons of cathinone and its pathway intermediates among

bioRxiv preprint doi: https://doi.org/10.1101/375105; this version posted July 24, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



healthy cicada controls (posterior sections), *M. cicadina*, and *M. platypediae* conidial plugs. (**C**) Previously hypothesized biosynthetic routes for cathinone production from the *Catha edulis* (20). (**D**) Candidate gene sequences from each *Massospora* assembly, a number of which were unique to each species, were retrieved using profile HMMs of enzymes predicted for each pathway step. Candidate gene co-regulation was predicted by shared nucleotide motifs within 1500bp upstream of predicted translational start sites. Only motifs associated with transaminase genes are shown. Three candidate genes from the *M. cicadina* assembly shared both Mc motif 1 and 2 in their upstream regions. Three candidate genes from *M. cicadina* showed evidence for expression based on a global proteomics analysis, including one hydratase whose upstream region contains Mc motif 1. (**E**) Logos and e-values of identified motifs from the upstream regions of candidate cathinone genes in each *Massospora* assembly. The median starting positions of Mc motifs 1 and 2 and Mp motifs 1 and 2 are 345bp, 405bp, 224bp and 796bp upstream of predicted translational start sites, respectively.

Psilocybin, a well-known psychoactive tryptamine produced by about 200 Agaricalean mushroom species, Among was among the most abundant metabolites detected in M. platypediae conidial plugs but not M. cicadina or the cicada control (*22-23*). The intermediates norbaeocystin and baeocystin from the described psilocybin biosynthetic pathway in wavy cap mushrooms (*Psilocybe cyanescens*; *22-23*) were also only detected in *M. platypediae* plugs (Fig. 2B-C). A third intermediate from an earlier step in the pathway, 4-hydroxytryptamine, was detected in both *M. platypediae* and *M. cicadina*, although more abundant in *M. platypediae* (Fig. 2B-C). Fragmentation patterns for psilocybin and its intermediates were consistent with previously reported standards (PubChem Compound Identifier: 10624), with psilocybin fragmentation experimentally observed for this study using a commercially available analytical standard (table S3).

bioRxiv preprint doi: https://doi.org/10.1101/375105; this version posted July 24, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



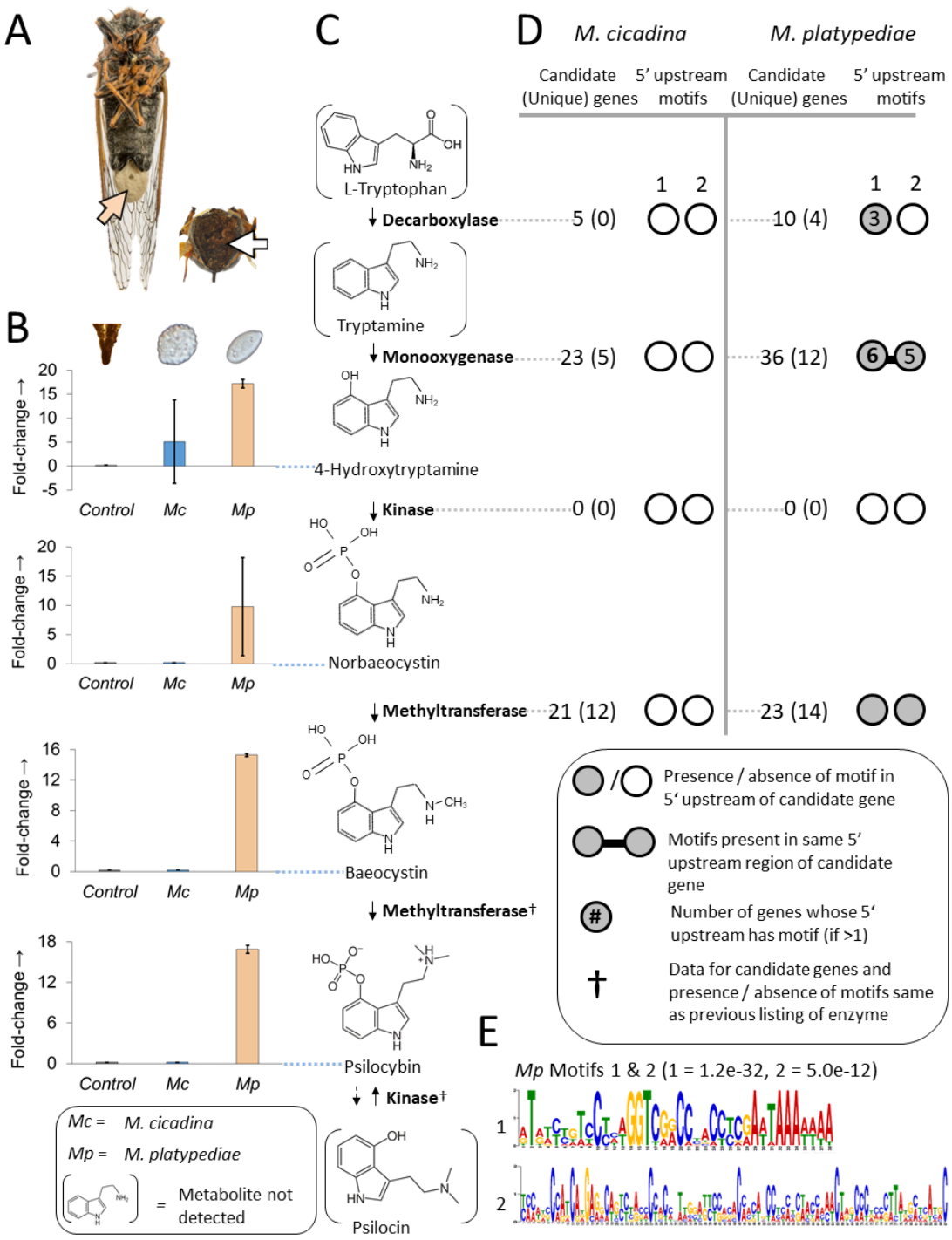


**Fig. 2. Production of the psilocybin and its proposed biosynthetic pathway in *Massospora platypediae*.**

(**A**) *M. platypediae*-infected annual cicada (*Platypedia putnami*) with a conspicuous conidial infection (beige arrow) and an internal resting spore infection (posterior cross-section, white arrow). (**B**) Global metabolomics fold-change comparisons of psilocybin and its pathway intermediates among healthy cicada controls (posterior segments), *M. cicadina*, and *M.*

bioRxiv preprint doi: https://doi.org/10.1101/375105; this version posted July 24, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



*platypediae* conidial plugs. (**C**) Proposed biosynthetic route for psilocybin production as previously described from *Psilocybe* mushrooms (22-23). (**D**) Profile HMMs and nucleotide motifs were searched as in Fig. 1D. Several candidates were unique to each assembly. Only motifs present upstream of candidate methyltransferase genes are shown. (**E**) Logos for motifs upstream of psilocybin genes are as in Fig. 1D. The median starting positions of Mp motifs 1 and 2 are 581bp and 1080bp upstream of predicted translational start sites, respectively.

We then performed targeted LC-MS/MS absolute quantification of cathinone, psilocybin, and its biologically active bioconversion product psilocin (*17*), on individual conidial and resting spore plug extracts from all three cicada-fungal pairs including *M. levispora*-infected Say's cicadas (*Okanagana rimosa*), which were excluded from the global metabolomics. The posterior abdomen of the global metabolomics controls as well as independently collected healthy *Magicicada* and *Platypedia* cicadas were also assessed for these compounds (*17*). All samples were quantified against standard curves generated from analytical standards. Cathinone was quantified in two collections of *M. cicadina*: one from Brood V (2016) in WV and a second from archived collection spanning four broods (VIII, IX, XXII, and XXIII) collected from 2001-2003 (5). Cathinone was present in 10 of 24 Brood V *M. cicadina* plugs and was quantifiable in concentrations from 44.3 − 303.0 ng / plug four of these plugs (table S4). From the archived collection cathinone, was present and quantifiable in concentrations from 68.9 to 80.1 ng / plug from 2 of 13 plugs (table S4). A Brood V *M. cicadina* plug extraction that had 4-months earlier yielded quantifiable amounts of cathinone was found to contain 11.6 ng/mL, a notable reduction from the previous time point.

Psilocybin was detected in 11 of 11 assayed *M. platypediae* plugs and 7 of 8 assayed *M. levispora* plugs while psilocin was detected in 9 of 11 plugs and 0 of 8 plugs, respectively (table S4). Of these, psilocybin was quantifiable only in *M. levispora*, with concentrations from 7.3 to 19.4 ng / plug across 6 plugs (table S4). Psilocin was detected only in *M. platypediae* in concentrations from 0.1 to 12.6 µg / plug across 9 plugs (table S4).

Having confidently detected and quantified cathinone and psilocybin/psilocin from independently collected *Massospora* populations (table S4), we next searched for candidate genetic mechanisms underlying their biosynthesis in assembled *M. cicadina* and *M. platypediae* metagenomes. After failing to retrieve *Massospora* sequences orthologous to genes from the characterized *Psilocybe* psilocybin or predicted *Catha* cathinone pathways using BLAST, we instead searched instead for all sequences containing protein domains found among the enzymes

bioRxiv preprint doi: https://doi.org/10.1101/375105; this version posted July 24, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



of each biosynthetic pathway (*17,20,22;* table S5). In each metagenome, we identified multiple genes representative of all classes of enzymes in each pathway with two exceptions (table S6, table S7). Although we identified at least one homolog for 7 of the 8 genes predicted to play a role in cathinone biosynthesis in the *M. cicadina* (cathinone +) and *M. platypediae* (cathinone -) metagenomes, we failed to detect genes encoding phenylalanine ammonia-lyase (*PAL*) (Fig. 1D). Similarly, in the *M. platypediae* (psilocybin +) and *M. cicadina* (psilocybin -) metagenomes, we identified at least one homolog for 3 of the 4 genes participating in psilocybin biosynthesis, the exception being 4-hydroxytryptamine kinase (*PsiK*) (Fig. 2D). The failure to detect *PAL* and *PsiK* homologs could be attributed to the fragmented and incomplete nature of the metagenomes, by the evolution of an alternate enzyme mechanism (see methods), or the execution of these steps either directly or indirectly by the host; further investigations with higher quality genome assemblies of fungi and hosts are needed to test these hypotheses.

To narrow down our list of candidate genes, we employed a two-pronged approach to gather additional lines of evidence that would suggest particular candidate genes participate in a common secondary metabolite pathway (*24-25;* fig. S5): 1) evidence of coordinated regulation from shared nucleotide sequence motifs in the upstream non-coding region of each candidate sequence and 2) evidence of gene clustering among homologs of the candidate sequences across a phylogenetically diverse database of 375 fungal genomes, since clustering could not be assessed in the fragmented *Massospora* metagenomes (table S8). The two candidate regulatory motifs recovered upstream of candidate cathinone biosynthesis genes in the *M. cicadina* assembly (cathinone +) were found upstream of the same predicted hydratase, transaminase and reductase genes (Fig. 1E, table S9). Two of the seven candidate regulatory motifs in the *M. platypediae* assembly (cathinone -) were only found upstream of transaminase genes (Fig. 1E, table S9). Two of the five motifs found upstream of candidate psilocybin biosynthesis genes in the *M. platypediae* assembly (psilocybin +) were present upstream of methyltransferase genes, while the single motif found in the *M. cicadina* assembly (psilocybin -) was not (Fig. 2E, table S9).

Searches of the fungal database identified 1548 gene clusters containing homologs of at least three candidate cathinone genes, but none contained a transaminase, which is predicted to catalyze the formation of cathinone itself. We also found 309 clusters containing homologs of at least three candidate psilocybin genes, 123 of which contain a methyltransferase, which catalyzes

bioRxiv preprint doi: https://doi.org/10.1101/375105; this version posted July 24, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



the formation of psilocybin (table S10). One notable locus in *Conidiobolus coronatus* (Entomophthorales) contained four aromatic-L-amino-acid decarboxylases and one methyltransferase (both functions are required for psilocybin biosynthesis), in addition to one helix-loop-helix transcription factor, one inositol phosphatase, and various amino acid, sugar and inorganic ion transporters (table S11). However, we were unable to confirm presence of psilocybin from eight diverse *C. coronatus* isolates (table S12), in culture or inoculated into a susceptible insect host, wax moth larvae (*17*).

Since genes encoding the synthesis of specialized secondary metabolites are occasionally inherited through non-vertical means, we assessed the potential for horizontal gene transfer of cathinone and psilocybin biosynthesis genes by constructing maximum likelihood phylogenetic trees on 22 high quality candidate genes from *M. cicadina* and *M. platypediae* (see methods; table S13-S14; fig. S6). High quality candidate genes were defined as sharing putative regulatory motifs in their 5' upstream non-coding regions with either a transaminase or methyltransferase gene (which encode enzymes that produce cathinone or psilocybin, respectively), or homologous to genes in the cluster of interest in *C. coronatus*. Gene trees of these candidates had topologies consistent with vertical inheritance among the Entomophthorales and highly similar homologs in Entomophthoralean fungi. The phylogenies of candidate psilocybin pathway sequences gleaned from the preliminary metagenomic data are consistent with an origin of psilocybin biosynthesis independent from Basidiomycetes, but this cannot be confirmed until the exact genes underlying this pathway are found.

We also analyzed the *Massospora cicadina* proteome in an attempt to validate and extend the metagenomic findings. The proteome of *M. platypediae* could not be studied in detail because of insufficient high quality sample material. Search results from the proteomic analysis of a single Brood V conidial plug from *M. cicadina* indicated the expression of three proteins with predicted functions linked to 2 of the 8 steps in the cathinone biosynthesis pathway - an aldo-keto reductase, an enoyl-(acyl carrier protein) reductase, and a 3-hydroxyacyl-CoA dehydrogenase (syn. hydratase), which had one of the detected motifs (*Mc* motif 1) in its 5' upstream gene region (Fig. 1, table S15). Four additional proteins with less confident IDs (only one unique peptide identified) involved in the cathinone pathway were recovered including a second aldo-keto reductase, two separate CoA thiolases, and one CoA dehydrogenase, which, together with the more confident

bioRxiv preprint doi: https://doi.org/10.1101/375105; this version posted July 24, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



identifications, provide additional candidate gene sequences that may participate in cathinone production via the CoA-dependent biosynthesis pathway (*17*).

The combined genomics and proteomics data provide complementary lines of evidence that support the metabolomics findings of amphetamine and psilocybin production in *Massospora* via previously characterized pathways. Additionally, their discovery raises interesting questions about the broader evolutionary history of behavior-modifying traits across the Entomophthorales as well as their functional role within infected cicadas. In order to model evolution of such traits and establish whether active host transmission behavior (*8*) seen in *Massospora* and other close relatives, including the genus *Strongwellsea* that also discharges conidia from living hosts (*8,19*), reflects a single origin or multiple independent origin events, we conducted a phylogenetic analysis of a genome-scale data set of > 200 conserved orthologous proteins from the *M. cicadina* and *M. platypediae* metagenomes along with 7 other representative taxa (table S16). Mapping of behavioral traits onto the inferred phylogeny from the concatenated genome-wide alignment revealed active host transmission behavior in both clades suggesting three possible parsimonious reconstructions of this behavioral trait in Entomophthoralean fungi (*17*). Those fungi that induced active host transmission behavior in their living hosts shared several key traits including 1) localized external sporulation patterns, and 2) increased host specificity compared to their summit disease counterparts (Fig. 3).

bioRxiv preprint doi: https://doi.org/10.1101/375105; this version posted July 24, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



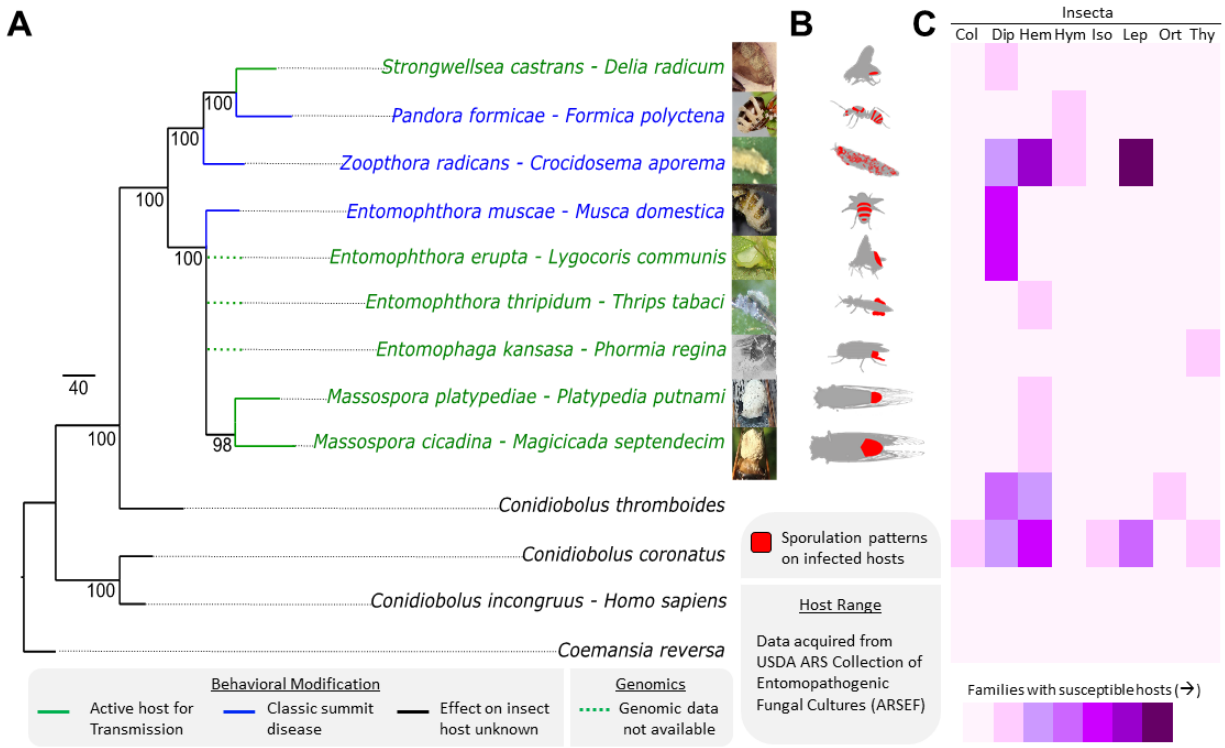


**Fig. 3. Phylogenomic relationships among behavior-modifying Entomophthoralean fungi, the behavioral and morphological host modifications they impose, and their host specificity.**

(**A**) RAxML phylogenetic tree of Entomopthoralean fungi (Zoopagomycota) based on the concatenated alignment of >200 conserved orthologous protein sequences with overlaid behavioral modification in infected hosts. (**B**) External sporulation patterns, (**C**) Heat map showing both the host range within and across select insect orders: Col: Coleoptera, Dip: Diptera, Hem: Hemiptera, Hym: Hymenoptera, Iso: Isoptera, Lep: Lepidoptera, Ort: Orthoptera, and Thy: Thysanoptera. The species *Z. radicans* and *C. thromboides* may represent broader species complexes, of which each phylogenetic species may have a narrower host spectrum. Photos used with permission from co-authors as well as Florian Freimoser, Joanna Malagocka, Judith Pell, and Ruth Ahlburg.

## Discussion

Without exposure studies involving healthy cicadas, the mechanism(s) by which monoamine alkaloids alone modify cicada behaviors cannot be fully resolved. However, previous arthropod exposure studies have provided extensive insight into the effects of amphetamine and psilocybin on various behaviors (*26-28*). For example, low doses (1-12 µg) of amphetamine similar to the cathinone levels detected in *M. cicadina* increased intraspecific aggression in ants (*26*) and significantly depleted biogenic amines, altering feeding behavior in adult blow flies (*Phormia regina*) by decreasing responsiveness to weak gustatory stimuli (*28*). Psilocybin

bioRxiv preprint doi: https://doi.org/10.1101/375105; this version posted July 24, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



ingestion had less observable impact on behavior compared to amphetamine for the limited number of arthropods observed (*27,29*). While these secondary metabolites from *Massospora* may help improve endurance to levels needed to engage with other conspecifics given their debilitating infections, activity-level effects of amphetamine and psilocybin cannot easily explain the display of female-typical courtship behaviors by male cicadas (5), except possibly in *P. putnami* (*7,17*). The atypical mating behaviors of other cicada species may be better explained by *Massospora*-mediated hormonal changes incidental to or interacting with alkaloid production (17, fig. S7). The differentially expressed "hypersexualization" between conidial- and resting spore-infected cicadas seems to support the latter hypothesis since changes in male behaviors cannot be incidental effects of castration or infection, or the observed behaviors would be seen in both spore types. The basis of hypersexualization remains unknown; the compounds we identified may be involved in this phenomenon, or may serve other roles including parasitoid deterrence (*30*).

Culturable fungi have yielded many antimicrobial drugs, medications, and industrial enzymes over the last century. The discovery of amphetamine production in fungi as well as preliminary evidence for the independent evolution of psilocybin production in the Zoopagomycota is consistent with the breadth of diverse secondary metabolite synthesis pathways that have evolved in fungi. However, the discovery of cathinone in *Massospora*-infected cicadas raises questions about the biological origin of amphetamine production in *Catha* plants and the possibility of cryptic fungal partners involved in its biosynthesis. Over the last few decades, previously cryptic fungal endophytes have been identified as the source of ergot alkaloid production in fescue (*Festuca* spp.) and morning glories (*Ipomoea* spp.) (*31-32*). Culture-based studies of *Catha edulis* (*33-34*) and ephedrine-producing *Ephedra nebrodensis* (*35*), have uncovered a diverse assortment of fungi, none of which have been screened for cathinone production, though most plant tissues remain unexplored. Interestingly, at least two cicada genera have been reported feeding on *Ephedra* plants, possibly indicating a detoxification mechanism present in some cicadas (*36-37*). The results of this study provide a strong impetus for enhanced screening of Ephedraceous and Celastraceous plants for alkaloid-producing endophytes as well as Entomophthoralean fungi for possible facultative plant associations (*17*). Despite the very recent discovery of Hypocrealean fungal symbionts from the fat bodies surrounding the bacteriomes and/or the well-developed surface sheath cells of more than a dozen cicada species native to Japan (*38*), our results do not support the presence of cryptic monoamine alkaloid-producing fungal

bioRxiv preprint doi: https://doi.org/10.1101/375105; this version posted July 24, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



symbionts in the posterior region of healthy or outwardly asymptomatic wing-banger and periodical cicadas (*17*).

The obligate lifestyle and ephemeral nature of many Entomophthoralean fungi have long constrained our understanding of these early diverging entomopathogens despite up to a century of observational studies into their behavior-modifying capabilities. We anticipate these discoveries will foster a renewed interest in early diverging fungi and their pharmacologically important secondary metabolites, which may serve as the next frontier for novel drug discovery and pave the way for a systems biology approach for obligate fungal pathogens spanning the known diversity of Kingdom Fungi.

## Materials and Methods

Sampling of healthy and fungus-infected cicadas

Sampling of freshly infected and asymptomatic wing-banger cicadas and Brood VII periodical cicadas was carried out on private lands in CA (2018), NM (2017), and NY (2018) with permission from individual landowners. Infected Say's cicada and Brood V periodical cicadas were collected from university-owned properties in MI (1998) and WV (2016), respectively. Collection of infected and healthy cicadas from all locations did not require permits. In CA, cicadas were collected under Exemption 5 of California Code of Regulations Section 650, Title 14 (https://nrm.dfg.ca.gov/FileHandler.ashx?DocumentID=157389&inline).

From May 28 to June 6, 2016, newly emerged Brood V 17-year periodical cicadas were visually inspected for *Massospora* fungal infections on the campus of West Virginia University in Morgantown, West Virginia, USA. Individuals with visible fungal plugs were transferred individually to sterile 20 mL vials, transported to the laboratory, and stored in a -20 C freezer until processing. Samples were obtained from 95 infected cicadas, of which 47 were from *M. cassini*, 41 from *Magicicada septendecim*, and seven from *M. septendecula*. A similar number of males and females were sampled across the three Brood V species, with 53 from females and 50 from males.

On June 1, 2016, additional geographically separated Brood V cicada populations were visually inspected for fungal infections. Seven infected specimens of *M. septendecim* and one of *M. cassini* were recovered from Hocking Co., OH. Concurrently, a collaborator provided an additional 34 infected Brood V specimens of *M. septendecim* collected from across five OH

bioRxiv preprint doi: https://doi.org/10.1101/375105; this version posted July 24, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



counties including Ashland, Cuyahoga, Hocking, Licking, and Lorain counties. Eight female, *Massospora*-infected, Brood VI specimens of *Magicicada septendecim* were collected in June, 2017 from North Carolina. All samples were collected and stored as previously described.

Twenty-three *M. cicadina*-infected specimens of *Magicicada* collected from 1978-2002 were provided from Dr. Chris Simon at the University of Connecticut, where they had been stored at room temperature, dried and pinned, since their collection. Dr. Gene Kritsky from Mount St. Joseph University provided fungal plugs from seven *M. cicadina* infected specimens of *Magicicada* collected from 1989-2008, three of which were thirteen year cicadas and four were 17-year cicadas.

In addition to periodic cicadas, two species of annual cicadas also were sampled and included in this study. On June 4, 2017, newly emerged banger-wing cicadas (*Platypedia putnami*) were inspected for *Massospora* infections 12 km northeast of Gallina, NM along the Continental Divide Trail. Sixteen infected individuals and one outwardly asymptomatic individual (to serve as a negative control) were identified and sampled as described above.

Nine *Massospora levispora*-infected specimens of *Okanagana rimosa* were collected by John Cooley from the University of Michigan Biological Station, Cheboygan and Emmet Counties, MI on June 20, 1998 (5,*39*). Infected individuals as well as outwardly asymptomatic individuals were placed immediately in individual 20-ml screw-cap vials containing 95% ethanol. Specimens were completely submerged and stored at 4 C until October 2017, when specimens were sent to WVU for processing.

Two independent populations of healthy cicadas, wing-banger cicadas from Trinity County, CA, USA and Brood VII periodical cicadas (*Magicicada septendecim*) from Onondaga Hill, NY were sampled in early June, 2018 to serve as negative controls for targeted metabolomics. Five males and five female cicadas were collected for each cicada species, transferred individually to sterile 25-50 mL vials, and stored in a -80 C freezer until processing.

Cicadas were processed aseptically, with half of the visible fungal plug removed with a sterile scalpel and transferred to a 1.5 ml microcentrifuge tube for DNA extraction. Controls were processed similarly, where half of the most posterior segments as well as underlying host tissues were sampled. A majority of the remaining plug from infected individuals were used for global metabolomics analyses. Dry plugs were transferred aseptically to sterile 1.5 mL microcentrifuge

bioRxiv preprint doi: https://doi.org/10.1101/375105; this version posted July 24, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



tubes and stored at 20 C. A small sampling of spores was retained for morphological studies, although not all specimens were examined.

Morphological studies of *Massospora* spp.

Given the scarcity of *Massospora* sequences from NCBI GenBank, morphological comparisons were necessary to determine if data on putatively assigned *Massospora* species agreed with that of previously reported measurements for conidial and resting spore stages (*2-3*). A portion of select fungal spore masses was harvested with a sterile scalpel and mounted in lactophenol for examination with light field microscopy. Slip covers were fastened with nail polish to allow slides to be archived and reexamined when necessary. Mean conidial/resting spore length x width and cicada host provided the main criteria to differentiate *Massospora* species. Slides were examined and photographed using a Nikon Eclipse E600 compound microscope (Nikon Instruments, Melville, New York) equipped with a Nikon Digital Sight DS-Ri1 high-resolution microscope camera. A sampling of 25 spores from each slide mount were measured using Nikon NIS-Elements BR3.2 imaging software. A total of 51 *Massospora*-infected cicadas were examined including 26 *M. cicadina* (15 conidial and 11 resting spore isolates), nine *M. levispora* (eight conidial and one resting spore isolates), and 15 *M. platypediae* (14 conidial and one aygospore isolates).

DNA extraction, amplification, sequencing, and assembly

Fungal genomic DNA was extracted from harvested fungal plugs using a modified Wizard kit (Promega, Madison, WI, USA) as previously described (*40*). DNA was suspended in 75 µl Tris-EDTA (TE) buffer and stored at -20 C. Portions of the nuclear 18S (SSU) and 28S (LSU) ribosomal DNA regions were amplified with the primer combinations NS24/NSSU1088R (*41*) and LR0R/LR5 (*42*) using BioLine PCR kits (Bioline USA Inc.,Taunton, MA, USA). EFL, an EF-like gene found in some groups of early diverging fungi including the Entomphthorales, was amplified using primers EF1-983F amd EF1aZ-1R (*43*). Three newly designed primers pairs (table S17) targeting three psilocybin biosynthesis enzymes, PsiD (l-tryptophan decarboxylase), PsiK (4-hydroxytryptamine kinase), and PsiM (norbaeocystin methyltransferase), described from hallucinogenic mushrooms, *Psilocybe cubensis* and *Psilocybe cyanescens* (*22-23*), were used to attempt to amplify *Psilocybe* psilocybin biosynthesis genes in *Massospora platypediae*.

bioRxiv preprint doi: https://doi.org/10.1101/375105; this version posted July 24, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



PCR conditions were as follows. LSU: initial denaturation at 94°C for 2 minutes, 35 cycles of denaturation at 94°C for 30 seconds, annealing at 51.1°C for 45 seconds, and elongation at 72°C for 90 seconds, and final elongation at 72°C for 5 minutes; SSU: initial denaturation at 94°C for 5 minutes, 36 cycles of denaturation at 94°C for 30 seconds, annealing at 49°C for 30 seconds, and elongation at 72°C for 60 seconds, and final elongation at 72°C for 7 minutes. EFL: initial denaturation at 94°C for 5 minutes, 30 cycles of denaturation at 95°C for 60 seconds, annealing at 54°C for 60 seconds, and elongation at 72°C for 90 seconds, and final elongation at 72°C for 7 minutes. PsiD: initial denaturation at 95°C for 4 minutes, 30 cycles of denaturation at 95°C for 45 seconds, annealing at 61°C for 45 seconds, and elongation at 72°C for 45 seconds, and final elongation at 72°C for 7 minutes. PsiK: initial denaturation at 95°C for 5 minutes, 35 cycles of denaturation at 95°C for 45 seconds, annealing at 52.5°C for 50 seconds, and elongation at 72°C for 50 seconds, and final elongation at 72°C for 5 minutes. PsiM: initial denaturation at 95°C for 5 minutes, 35 cycles of denaturation at 95°C for 45 seconds, annealing at 54°C for 50 seconds, and elongation at 72°C for 40 seconds, and final elongation at 72°C for 5 minutes.

Amplified products were visualized with SYBR gold (Invitrogen, Grand Island, NY, USA), which then were loaded onto a 1-1.5%, wt/vol agarose gel made with 0.5-1% Tris-borate-EDTA buffer (Amresco, Solon, OH, USA). Psilocybin biosynthesis (Psi) products were cleaned with Zymo Research (Irvine, CA, USA) DNA Clean and Concentrator-5 kit. All products were purified with ExoSAP-IT (Affymetrix, Inc., Santa Clara, CA, USA) and submitted to Eurofins MWG Operon (Huntsville, AL, USA) for Sanger sequencing. Chromatograms were assembled and inspected using Codon Code aligner and phylogentic analyses conducted using MEGA 7 (*44*).

The *M. platypediae* metagenome assembly was generated from 390 Million 2x150 bp raw reads. Reads were quality trimmed with sickle using default parameters (*45*). Metagenome assembly was performed with megahit with "--presets meta-sensitive" option and utilizing GPU cores for accelerated matching speed to produce a genome assembly of ~3 Gb (2.1 million contigs; N50= 2784 bp). The *M. cicadina* assembly was generated with SPAdes using default parameters. Both assemblies were further cleaned for vector and primer contamination with NCBI vecscreen protocol (https://www.ncbi.nlm.nih.gov/tools/vecscreen/) and implemented in package (https://github.com/hyphaltip/autovectorscreen). Several rounds of BLASTN based screening were applied to remove contaminated contigs or regions. Further screening of contigs were performed during genome submission process to NCBI to remove highly similar matches to plant

bioRxiv preprint doi: https://doi.org/10.1101/375105; this version posted July 24, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



or insect sequences. This Whole Genome Shotgun project has been deposited at DDBJ/ENA/GenBank under the accessions QKRY00000000 (*M. platypediae*) and QMCF00000000 (*M. cicadina*). The versions described in this paper are versions QKRY01000000, QMCF01000000.

Phylogenetic analysis of ribosomal RNA genes

Phylogenetic trees were constructed using LSU, SSU, and EFL sequences generated from 22, 26, and 20 *Massospora* sequences, respectively, representing three *Massospora* species, *M. cicadina*, *M. levispora* and *M. platypediae* (data not shown). The concatenated set (LSU+SSU) consisted of four sequences, two from *M. cicadina* and one each from *M. levispora* and *M. platypediae*. Nine additional published sequences spanning the known diversity of the Entompohthorales were included to help resolve relationships among *Massospora* and its close allies (*41*). One isolate of *Conidiobolus pumilus* (strain ARSEF 453) served as an outgroup. Sequences were aligned using CLUSTAL-W (*46*). Alignments were visually inspected, after which optimized nucleotide substitution model (Tamura 3-parameter + G) was selected using ModelTest (*47*). Partial deletion with site coverage cutoff of 95% was used. Maximum likelihood (ML) trees were estimated using MEGA 7. Phylogenetic support was assessed by bootstrap analysis with 1000 replicates using MEGA 7.

Phylogenomic analyses

Multi-gene phylogenetic analysis of protein coding genes was also performed to assess confidence in the phylogenetic relationships. Since the metagenome assemblies were highly fragmented due to high repetitive sequence content, a complete genome annotation was not undertaken. Instead a targeted similarity search identified contigs encoding homologs of a collection of generally single copy protein coding genes developed for the 1000 Fungal genomes project. These 434 "JGI_1086" markers are available as Hidden Markov Models (https://github.com/1KFG/Phylogenomics_HMMs). Consensus sequences generated from the JGI_1086 markers using hmmemit (HMMER v3) (*48*) were used as queries in a translated search of the *Massospora* metagenome assemblies with TFASTY (*49*) using a permissive expectation cutoff value of '0.001'. These candidate contigs were then annotated with funannotate pipeline

bioRxiv preprint doi: https://doi.org/10.1101/375105; this version posted July 24, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



(https://github.com/nextgenusfs/funannotate) with informant protein sequences of a collection Zoopagomycota fungi, *de novo* gene prediction tools Augustus (*50*) and GeneMark ES+ET (*51*).

The predicted proteins for the targeted *Massospora* contigs were analyzed in the PHYling pipeline with proteins from the annotated genomes of the Kickxellomycotina fungus *Coemansia reversa* NRRL 1564, the Entomophthoromycotina fungi, *Conidiobolus coronatus* NRRL 28638, *Conidiobolus incongruus* B7586, *Conidiobolus thromboides* FSU 785) and the predicted ORFs from assembled transcriptomes of Entomophthoromycotina *Entomophthora muscae*, *Pandora formicae*, *Strongwellsea castrans*, *Zoophthora radicans* ATCC 208865. Briefly PHYling (https://github.com/stajichlab/PHYling_unified; DOI: 10.5281/zenodo.1257002) uses a preselected marker set of HMMs (https://github.com/1KFG/Phylogenomics_HMMs; DOI: 10.5281/zenodo.1251477) to identify best homologous sequences, aligns, trims, and builds a concatenated phylogenetic tree or individual gene trees. Searches of the annotated genomes of most species identified between 397 (*Con. coronatus*) and 427 (*S. castrans*) of the 434 marker genes in each species. Searches of the *M. cicadina* and *M. platypediae* targeted annotation set identified 202 and 237 of the 434 marker sets respectively. This lower number indicates a less complete and highly fragmented metagenome assembly but these ~200 provide sufficient phylogenetic information to reconstruct species relationships. The 10 taxon tree was constructed with IQ-TREE (v1.6.3; *52-53*) using the concatenated alignment and 433 partitions and 1000 ultrafast bootstrap support computation (-bb 1000 -st AA -t BIONJ -bspec GENESITE). All marker gene sequences, alignment and partition files are available at DOI: 10.5281/zenodo.1306939.

Fungal metabolite extraction

Fungal plugs were weighed and transferred into sterile 1.5 mL microcentifuge tubes. Metabolite extracts from each fungal sample were normalized to a final concentration of 10 mg of fungal tissue to 1 mL and polar metabolites were extracted in a buffer consisting of 80% HPLC grade methanol and 20% water. Samples then were sonicated for 10 minutes. The sonicated samples were centrifuged at 14,000 x g for 15 minutes at 4℃ to pellet proteins and cellular debris. The supernatant was removed and dried down using lyophilization prior to analysis. Samples were reconstituted in 100 µL of 50% methanol and 50% water for mass spectrometry analysis.

bioRxiv preprint doi: https://doi.org/10.1101/375105; this version posted July 24, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



Mycelia from cultures of *Conidiobolus coronatus* (*17*) were harvested by scraping 7-day old cultures grown on ½ strength SDAY. Whole *Con. coronatus*--infected greater wax moth (*Galleria mellonella*) were harvested at 48 hours post inoculation. Both tissues were macerated in a buffer as previously described.

Analytical procedure and LC-MS conditions for metabolomics

The Agilent LC-MS system (Agilent Technologies, Santa Clara, CA, USA) included an Agilent 1290 Infinity quaternary ultra-high pressure liquid chromatography (UHPLC) pump. This LC system was coupled to an Agilent 6530 quadrupole time of flight (Q-ToF) mass spectrometer with electrospray ionization. The chromatographic separations were performed using a Merck polymeric bead based ZIC-pHILIC column (100mmx2.1mm, 5µm).

Mobile phase A consisted of 10 mM ammonium acetate and mobile phase B consisted of 100% acetonitrile. Metabolites were separated on a gradient that went from 95% mobile phase to 5% mobile phase B over 20 minutes with 7-minute re-equilibration at the end of the chromatographic run.

The mass spectrometer was operated in both positive and negative ion mode for the analysis. Technical replicates were acquired for each sample. Data dependent MS/MS was performed on triplicate pooled samples for fragmentation and library searching. For *M. cicadina*, pooled samples consisted of 5 independently extracted Brood V periodical cicadas collected on the campus of WVU in 2016. For *M. platypediae* pooled sampled consisted of 5 independently extracted *Platypedia putnami* annual cicadas collected in NM in 2017. Controls consisted of excised posterior segments from healthy brood V periodical cicadas (n=3) and a single healthy *P. putnami* from NM.

Data processing and metabolite identification for global metabolomics

The LC-MS/MS data acquired using Agilent Mass Hunter Workstation (*.d files) were processed in Agilent Profinder software version 2.3.1 for batch recursive analysis. The datasets were subjected to spectral peak extraction with a minimum peak height of 600 counts and the charge state for each metabolite was restricted to two. Further, retention time and mass alignment corrections were performed using Agilent Profinder software version 2.3.1 on the runs to remove non-reproducible signals. The resulting features then were exported as *.cef files to Mass Profiler

bioRxiv preprint doi: https://doi.org/10.1101/375105; this version posted July 24, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



Professional (MPP) software version 2.4.3 (Agilent Technologies, Santa Clara, CA, USA) for statistical analysis.

Statistical data analysis for metabolomics

The extracted features were imported into MPP software for statistical analysis. Principle Component Analysis (PCA) was performed to check the quality of the samples after which the data containing filtered features were processed by one way ANOVA to ascertain differences between control and fungal groups. Only the analytes with p values < 0.05 and fold change (FC) > 20 were treated as statistically significant. Additionally, multiple test corrections using Bonferroni were applied to reduce false positives and false negatives in the data.

Targeted metabolomics

The absolute quantification of cathinone, psilocin and psilocybin from the *M. cicadina* and *M. platypediae* fungal plug metabolite extracts described above, *M. levispora* extracts that were stored long term were not able to be analyzed for global metabolite changes were able to be analyzed in this targeted experiment. This analysis was performed using a 5500 QTRAP mass spectrometer (SCIEX Inc., Concord, Ontario, Canada). Chromatographic separations were performed using an HPLC system consisting of two Shimadzu LC 20 AD pumps that included a degasser and a Shimadzu SIL 20 AC auto sampler (Shimadzu Corp. Kyoto, Kyoto Prefecture, Japan). Analytes were separated on a 3.0x150 mm Imtakt Scherzo SW-C18 column (Imtakt USA, Portland, Oregon). The HPLC column oven was set to 40℃ for the analysis, with an 8 µL injection volume for each sample. Mobile phase A consisted of HPLC grade H2O with 0.1% formic acid, while mobile B consisted of acetonitrile with 0.1% formic acid. The LC gradient started 5% for 0.5 minutes and then increased to 95% B over 5 minutes followed by a re-equilibration to starting conditions for a period of 2.1 minutes. The flow rate was set to 0.3 mL/minute.

Targeted metabolites were quantified using multiple reaction monitoring. The transitions that were monitored for cathinone were m/z 150.1→106.3. The transitions that were monitored for psilocybin were m/z 284.5→205.2 and 240.1. Psilocin was monitored at m/z 205.2→116.3. Mass spectrometer source conditions were set to the following: curtain gas = 15, ion spray voltage= 4800 V, temperature= 600 °C, gas 1= 45, gas 2= 50. Calibration samples were prepared in 50 % methanol/$H_2O$. Calibration concentrations ranged from 0.05 to 10 ng/mL of both psilocin and

bioRxiv preprint doi: https://doi.org/10.1101/375105; this version posted July 24, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



psilocybin and 1.25-25 ng/mL for cathinone. In order to assess variation in the assay duplicate standard curves were analyzed at the beginning and at the end of the analytical run. Absolute quantification data was processed using Analyst software ver. 1.6.3 (SCIEX Inc., Concord, Ontario, Canada).

Targeted quantification of CoA dependent pathway intermediates

Targeted detection of CoA dependent pathway intermediates including Trans-cinnamoyl-CoA, 3-Hydroxy-3-phenylpropionyl-CoA, 3-Oxy-3-phenylpropionyl-CoA, and Benzoyl-CoA was performed using a 5500 QTRAP mass spectrometer (SCIEX Inc., Concord, Ontario, Canada). Chromatographic separations were performed using an HPLC system consisting of two Shimadzu LC 20 AD pumps that included a degasser and a Shimadzu SIL 20 AC auto sampler (Shimadzu Corp. Kyoto, Kyoto Prefecture, Japan). Analytes were separated on a 3.0x150 mm Imtakt Imtrada column (Imtakt USA, Portland, Oregon). Mobile phase A consisted of 10 mM ammonium acetate at pH 9.0 and mobile phase B consisted of 100% acetonitrile. CoA intermediates were separated on a gradient that went from 95% mobile phase B to 5% mobile phase B over 15 minutes with 3-minute re-equilibration at the end of the chromatographic run.
Absolute quantification data was processed using Analyst software ver. 1.6.3 (SCIEX Inc., Concord, Ontario, Canada).

Targeted quantification of cathinone, psilocybin and psilocin in *Conidiobolus coronatus* and 2018-collected healthy *Platypedia putnami* and *Magicicada septendecim* (Brood VII)

Targeted cathinone, psilocybin and psilocin experiments were performed using an Accela 1250 UHPLC coupled to a Q Exactive Orbitrap Mass Spectrometer (Thermo Scientific, San Jose, CA) operating in positive ion mode. An Agilent Zorbax C18 column (2.1 mm diameter by 10cm length) was loaded with 2 µL of extracts. Separations were performed using a linear gradient ramping from 95% solvent A (water, 0.1% formic acid) to 90% solvent B (acetonitrile, 0.1% formic acid) over 6 minutes, flowing at 300 µL/min.

The mass spectrometer was operated in targeted-SIM/data-dependent MS2 acquisition mode. Precursor scans were acquired at 70,000 resolution with a 4 m/z window centered on 285.1005 m/z and 205.1341 m/z (5e5 AGC target, 100 ms maximum injection time). Tandem MS spectra were acquired using HCD scans triggered by 285.1005 m/z and 205.1341 m/z (NCE 35, 2e5 AGC

bioRxiv preprint doi: https://doi.org/10.1101/375105; this version posted July 24, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



target, 100 ms maximum injection time, 4.0 m/z isolation window). Charge states 2 and higher were excluded for fragmentation and dynamic exclusion was set to 5.0 s.

Proteomics of *M. cicadina* and *M. platypediae* plugs

A bicinchoninic acid assay (BCA) assay was used to normalize protein prior to LC-MS analysis. Several plugs of *M. cicadina* and *M. platypediae,* both conidial and resting spore isolates, were assessed, after which only a single conidial isolate for *Massospora cicadina* was chosen for proteomics based on BCA results. Peptide sequencing experiments were performed using an EASY-nLC 1000 coupled to a Q Exactive Orbitrap Mass Spectrometer (Thermo Scientific, San Jose, CA) operating in positive ion mode. An EasySpray C18 column (2 µm particle size, 75 µm diameter by 15cm length) was loaded with 500 ng of protein digest in 22 µL of solvent A (water, 0.1% formic acid) at a pressure of 800 bar. Separations were performed using a linear gradient ramping from 5% solvent B (75% acetonitrile, 25% water, 0.1% formic acid) to 30% solvent B over 120 minutes, flowing at 300 nL/min.

The mass spectrometer was operated in data-dependent acquisition mode. Precursor scans were acquired at 70,000 resolution over 300-1750 m/z mass range (3e6 AGC target, 20 ms maximum injection time). Tandem MS spectra were acquired using HCD of the top 10 most abundant precursor ions at 17,500 resolution (NCE 28, 1e5 AGC target, 60 ms maximum injection time, 2.0 m/z isolation window). Charge states 1, 6-8 and higher were excluded for fragmentation and dynamic exclusion was set to 20.0 s.

Mass spectra were searched for peptide identifications using Proteome Discoverer 2.1 (Thermo Scientific) using the Sequest HT and MSAmanda algorithms, peptide spectral matches were validated using Percolator (target FDR 1%). Initial searches were performed against the complete UniProt database (downloaded 19 March, 2018). Proteins identified from this database were used to generate a new database which included the psilocybin and cathinone pathway sequences from the metagenomics analysis. Peptide matches were restricted to 10 ppm MS1 tolerance, 20 mmu MS2 tolerance, and 2 missed tryptic cleavages. Fixed modifications were limited to cysteine carbamidomethylation, and dynamic modifications were methionine oxidation and protein N-terminal acetylation. Peptide and protein grouping and results validation was performed using Scaffold 4.8.4 (Proteome Software, Portland, OR) along with the X! Tandem algorithm against the previously described database. Proteins were filtered using a 99% FDR threshold.

bioRxiv preprint doi: https://doi.org/10.1101/375105; this version posted July 24, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



Search for candidate psilocybin and cathinone biosynthetic genes

We used multiple approaches to identify candidate psilocybin and cathinone biosynthetic genes (Figure S5). We first performed BLASTx searches of *Massospora* metagenome assemblies using four characterized psilocybin biosynthesis proteins from *Psilocybe cubensis* (*22*) (norbaeocystin methyltransferase (accession P0DPA9.1); L-tryptophan decarboxylase (accession P0DPA6.1); 4-hydroxytryptamine kinase (accession P0DPA8.1); cytochrome P450 monooxygenase (accession P0DPA7.1)) and all 42 candidate cathinone biosynthesis proteins identified in a transcriptomic study of *Catha edulis* (retrieved from S1 dataset and Table 2 of *20*). We next subtracted aligned sequences between the two assemblies using NUCmer and PROmer from MUMMER v.4.0.0beta2 in order to identify sequence unique to each assembly (*54*). Finally, we used profile HMM of the protein family (Pfam) domains found in characterized psilocybin and cathinone biosynthetic genes to search the 6-frame amino acid translations of the *Massospora* metagenomic assemblies with HMMER3 v.3.1b2 (*55*; Table S5). We searched with all SAM and SAM-like methyltransferase domains because many are similar to the S-adenosyl-l-methionine (SAM)-dependent methyltransferase domain in the psilocybin pathway's norbaeocystin methyltransferase. Unless stated otherwise, all BLAST searches were conducted using the BLAST suite v.2.2.25+ with a max. evalue threshold=1e-4 (*56*) against a local database containing 375 publically available fungal genomes (Table S8).

All subsequent steps were performed on the set of sequences retrieved by profile HMM search. We filtered out bacterial, protist and insect sequences by aligning hits to NCBI's non-redundant protein database (last accessed March 14$^{th}$, 2018) using BLASTp. We annotated contigs containing filtered hits using WebAUGUSTUS with Con*idiobolus coronatus* species parameters using the option "predict any number of (possibly partial) genes" and reporting genes on both strands (*57*). We manually verified the coordinates of CDS regions by using the hits as queries in a BLASTx search (min similarity=45%) to the local proteome database. The predicted amino acid sequences of all candidate genes, as well as the nucleotide sequences and coordinate files of annotated contigs are available at DOI: 10.5281/zenodo.1306939.

To investigate putative ungapped regulatory motifs associated with candidate gene sequences, we used a custom Perl script to extract up to 1500bp upstream of each gene's start codon and submitted each set of extracted sequences to commandline MEME v.4.12.0 (max evalue=5e-2;

bioRxiv preprint doi: https://doi.org/10.1101/375105; this version posted July 24, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



min width=25; max width=100; max motifs=10) (*58*). Motifs were manually inspected and those consisting solely of homopolymer repeats were discarded.

Because the *Massospora* assemblies were themselves too fragmented to assess if candidate genes were found in clusters, we searched for evidence of clusters containing candidate gene homologs retrieved from the local fungal genome database using BLASTp and a custom Perl script (available at https://github.com/egluckthaler/cluster_retrieve). Homologs were considered clustered if separated by 6 or fewer intervening genes. Only clusters with at least 3 genes were retained. Genes at clustered loci of interest were annotated with eggNOG-mapper v.0.99.1 (*59*) based on orthology data from the fungal-specific fuNOG database (*60*).

We designated genes as "high-quality" candidates of interest if either a) their upstream regions contained a motif also present in the upstream region of a predicted transaminase or methyltransferase gene or b) if they were homologous to the methyltransferase and decarboxylase sequences from the one notable gene cluster retrieved from *Conidiobolus coronatus* (Results). We used these criteria to help define candidates of interest because methyltransferase and transaminase enzymes catalyze the formation of psilocybin and cathinone, respectively.

To investigate the evolution of the 22 high quality candidate gene sequences (Table S13), we first divided sequences into groups based on the profile HMM that initially retrieved them. Sequences from each group were used as queries to retrieve hits from the local fungal genome database (Table S8) which was modified at this point to also include all candidate genes identified from both *Massospora* assemblies, as well as the metatranscriptomes of *Conidiobolus incongruus* B7586, *Entomophthora muscae* (ARSEF 13376), *Zoophthora radicans* ATCC 208865, *Strongwellsea* sp. (HHDFL040913-1), and *Pandora formicae (14*) using BLASTp.

The *Strongwellsea* sp. transcriptome was based on total RNA extracted from a single infected *Delia radiucm* caught in an organically grown cabbage field in Denmark: Sørisgaard (55.823706, 12.171149). The fly had the characteristic *Strongwellsea*-caused open hole on the side of the abdomen and was snap-frozen in liquid nitrogen as quickly as possible. Illumina sequencing yielded ca. 20 million 150 bp paired-end reads that were assembled using Trinity 2.2.0 (*61*). Transcripts were filtered so that only the highest expressed isoform for each transcript and transcripts assigned to kingdom Fungi using MEGAN4 (*62*) were retained.

bioRxiv preprint doi: https://doi.org/10.1101/375105; this version posted July 24, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



Each set of queries and associated hits (with two exceptions, see below) were then aligned with MAFFT v.7.149b using the --auto strategy (*63)*, trimmed with TrimAL v.1.4 using the –automated1 strategy (*64*), and had sequences composed of >70% gaps removed. Sets of sequences retrieved using candidates with the 'p450' or 'aldo_ket_red' Pfam HMMs were too large to construct initial alignments. Instead, all retrieved hits and queries were first sorted in homolog groups by submitting the results of an all vs. all BLASTp to OrthoMCL v.2.0 (inflation = 1.5) (*65*). All sequences in the homolog groups containing the query candidate genes were then aligned and trimmed as above. Each trimmed alignment was then used to construct a phylogenetic tree with FASTTREE v.2.1.10 using default settings that was then midpoint rooted (*66*). For each high quality query sequence present in the tree, a strongly supported clade containing the query sequence and closely related Entomophthoralean sequences was extracted. Sequences in the extracted clade were re-aligned and trimmed (as above) and used to build final maximum likelihood trees using RAxML v8.2.11 with 100 rapid bootstraps (*67*) (fig. S6). Final RAxML trees were rooted based on the outgroup in the clade extracted from the midpoint rooted FASTTREE.

**Acknowledgments:** Special thanks to the Drug Enforcement Administration (DEA) for guidance on discovery and handling of Schedule 1 controlled substances. Genomes and transcriptome sequencing of *C. thromboides*, *C. coronatus*, and *Z. radicans* were supported by the Department of Energy's Joint Genomes Institute under the JGI Community sequencing program project 1978 'Genomics of the early diverging lineages of fungi and their transition to terrestrial, plant-based ecologies' supported by the Office of Science off the US Department of Energy under contract DE-AC02-05CH11231. **Funding:** G.R.B. supported by Protea Biosciences and NSF grant DBI-1349308, J.C.S. by NSF DEB 1638999, J.E.S. by NSF DEB 1441715, W.J.D. and T.Y.J. by NSF DEB 1441677, D.G.P by NIH R15GM114774, H.H.D.F.L. by the Villum Foundation (grant number 10122), T.G. by NSF grant CHE-1608149, J. R. C. and C. S. by NSF DEB 1655891, and MTK by funds from West Virginia Agricultural and Forestry Experiment Station and NSF DBI-1349308. **Author contributions:** G.R.B., J.S., W.J.D., T.Y. J., D.G.P., and M.T.K. conceived of the study. J.R.C., C.S., A.M.M., M.C.B., G.K., K.T.H., R.A.H., T.G., D.M., and N.A. provided archived material and/or fresh specimens. G.R.B., A.M.M., M.C.B., K.L.W., C.M.S., E.J.S., A.M.M., T.K. and M.T.K. performed laboratory work with the help of D.P.G.S. and D.G.P.

bioRxiv preprint doi: https://doi.org/10.1101/375105; this version posted July 24, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



G.R.B., E.G.-T., J.C.S., J.S., W.J.D., T.Y.J., J.E., H.D.V.L., M.D.M., and M.T.K. analyzed data. G.R.B., E.G.-T., J.C.S., J.S., W.J.D., J.R.C., J.E., C.S., and M.T.K. wrote the manuscript with input from all coauthors; **Competing interests:** Authors declare no competing interests. **Data and materials availability:** Whole Genome Shotgun projects have been deposited at DDBJ/ENA/GenBank under the accessions QKRY00000000 (*M. platypediae*) and QMCF00000000 (*M. cicadina*). The versions described in this paper are QKRY01000000 and QMCF01000000. All phylogenomic marker and candidate cathinone and psilocybin gene sequences are available at DOI: 10.5281/zenodo.1306939.

**References and Notes**

1. Spatafora, J.W., Chang, Y., Benny, G.L., Lazarus, K., Smith, M.E., Berbee, M.L., Bonito, G., Corradi, N., Grigoriev, I., Gryganskyi, A. and James, T.Y., 2016. A phylum-level phylogenetic classification of zygomycete fungi based on genome-scale data. *Mycologia*, 108(5), pp.1028-1046.
2. Soper, R.S., 1963. *Massospora levispora*, a new species of fungus pathogenic to the cicada, *Okanagana rimosa*. *Canadian Journal of Botany*, 41(6), pp.875-878.
3. Soper, S.R., 1974. The genus *Massospora*, entomopathogenic for cicadas. Part I. Taxonomy of the genus. *Mycotaxon*, 1, pp.13-40.
4. Soper, S. R., 1981. New cicada pathogens: *Massospora cicadettae* from Australia and *Massospora pahariae* from Afghanistan. *Mycotaxon*, 13, pp.50-58.
5. Cooley, J.R., Marshall, D.C. and Hill, K.B., 2018. A specialized fungal parasite (*Massospora cicadina*) hijacks the sexual signals of periodical cicadas (Hemiptera: Cicadidae: Magicicada). *Scientific Reports*, 8(1), p.1432.
6. White, J.A., Ganter, P., McFarland, R., Stanton, N. and Lloyd, M., 1983. Spontaneous, field tested and tethered flight in healthy and infected *Magicicada septendecim* L. *Oecologia*, 57(3), pp.281-286.
7. Murphy, J. M., and Redden, G. A., 2003. Fungal infection and gender confusion in the wing-banger cicada *Platypedia putnami*. *Norse Scientist* (*NKU*), 65–68.
8. Roy, H.E., Steinkraus, D.C., Eilenberg, J., Hajek, A.E. and Pell, J.K., 2006. Bizarre interactions and endgames: entomopathogenic fungi and their arthropod hosts. *Annu. Rev. Entomol.*, 51, pp.331-357.

bioRxiv preprint doi: https://doi.org/10.1101/375105; this version posted July 24, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



9. Hughes, D.P., Araújo, J.P.M., Loreto, R.G., Quevillon, L., De Bekker, C. and Evans, H.C., 2016. From so simple a beginning: the evolution of behavioral manipulation by fungi. In Advances in Genetics (Vol. 94, pp. 437-469). Academic Press.
10. de Bekker, C., Quevillon, L.E., Smith, P.B., Fleming, K.R., Ghosh, D., Patterson, A.D. and Hughes, D.P., 2014. Species-specific ant brain manipulation by a specialized fungal parasite. *BMC Evolutionary Biology*, 14(1), p.166.
11. De Bekker, C., Ohm, R.A., Evans, H.C., Brachmann, A. and Hughes, D.P., 2017. Ant-infecting *Ophiocordyceps* genomes reveal a high diversity of potential behavioral manipulation genes and a possible major role for enterotoxins. *Scientific Reports*, 7(1), p.12508.
12. Grell, M.N., Jensen, A.B., Olsen, P.B., Eilenberg, J. and Lange, L., 2011. Secretome of fungus-infected aphids documents high pathogen activity and weak host response. *Fungal Genetics and Biology*, 48(4), pp.343-352.
13. Arnesen, J.A., Malagocka, J., Gryganskyi, A., Grigoriev, I.V., Voigt, K., Stajich, J.E. and Licht, H.H.H.D.F., 2018. Early diverging insect pathogenic fungi of the order Entomophthorales possess diverse and unique subtilisin-like serine proteases. *bioRxiv*, p.247858.
14. Małagocka, J., Grell, M.N., Lange, L., Eilenberg, J. and Jensen, A.B., 2015. Transcriptome of an entomophthoralean fungus (*Pandora formicae*) shows molecular machinery adjusted for successful host exploitation and transmission. *Journal of Invertebrate Pathology*, 128, pp.47-56.
15. Elya, C., Lok, T.C., Spencer, Q.E., McCausland, H., Martinez, C.C. and Eisen, M.B., 2017. Robust manipulation of the behavior of *Drosophila melanogaster* by a fungal pathogen in the laboratory. *bioRxiv* p.232140.
16. Şenyuva, H.Z., Gilbert, J. and Öztürkoğlu, Ş., 2008. Rapid analysis of fungal cultures and dried figs for secondary metabolites by LC/TOF-MS. *Analytica Chimica Acta*, 617(1-2), pp.97-106.
17. Supplemental results are available as supplementary materials.
18. Hutchison, J.A., 1962. Studies on a new *Entomophthora* attacking calyptrate flies. *Mycologia*, 54(3), pp.258-271.

bioRxiv preprint doi: https://doi.org/10.1101/375105; this version posted July 24, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



19. Batko, A. and Weiser, J., 1965. On the taxonomic position of the fungus discovered by Strong, Wells, and Apple: *Strongwellsea castrans* gen. et sp. nov.(Phycomycetes; Entomophthoraceae). *Journal of Invertebrate Pathology*, 7(4), pp.455-463.
20. Groves, R.A., Hagel, J.M., Zhang, Y., Kilpatrick, K., Levy, A., Marsolais, F., Lewinsohn, E., Sensen, C.W. and Facchini, P.J., 2015. Transcriptome profiling of khat (*Catha edulis*) and *Ephedra sinica* reveals gene candidates potentially involved in amphetamine-type alkaloid biosynthesis. *PLoS One*, 10(3), p.e0119701.
21. Lomascolo, A., Stentelaire, C., Asther, M. and Lesage-Meessen, L., 1999. Basidiomycetes as new biotechnological tools to generate natural aromatic flavours for the food industry. *Trends in Biotechnology*, 17(7), pp.282-289.
22. Fricke, J., Blei, F. and Hoffmeister, D., 2017. Enzymatic synthesis of psilocybin. *Angewandte Chemie International Edition*, 56(40), pp.12352-12355.
23. Reynolds, H.T., Vijayakumar, V., Gluck-Thaler, E., Korotkin, H.B., Matheny, P.B. and Slot, J.C., 2018. Horizontal gene cluster transfer increased hallucinogenic mushroom diversity. *Evolution Letters*, 2(2), pp.88-101.
24. Wolf, T., Shelest, V., Nath, N. and Shelest, E., 2015. CASSIS and SMIPS: promoter-based prediction of secondary metabolite gene clusters in eukaryotic genomes. *Bioinformatics*, 32(8), pp.1138-1143.
25. Slot, J.C., 2017. Fungal gene cluster diversity and evolution. In Advances in genetics (Vol. 100, pp. 141-178). Academic Press.
26. Frischknecht, H.R. and Waser, P.G., 1978. Actions of hallucinogens on ants (*Formica pratensis*)—II. Effects of amphetamine, LSD and delta-9-tetrahydrocannabinol. *General Pharmacology: The Vascular System*, 9(5), pp.375-380.
27. Witt, P.N., 1971. Drugs alter web-building of spiders: A review and evaluation. *Behavioral Science*, 16(1), pp.98-113.
28. Brookhart, G.L., Edgecomb, R.S. and Murdock, L.L., 1987. Amphetamine and reserpine deplete brain biogenic amines and alter blow fly feeding behavior. *Journal of Neurochemistry*, 48(4), pp.1307-1315.
29. Masiulionis, V.E., Weber, R.W. and Pagnocca, F.C., 2013. Foraging of *Psilocybe* basidiocarps by the leaf-cutting ant *Acromyrmex lobicornis* in Santa Fé, Argentina. *SpringerPlus*, 2(1), p.254.

bioRxiv preprint doi: https://doi.org/10.1101/375105; this version posted July 24, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



30. Soper, R.S., Shewell, G.E. and Tyrrell, D., 1976. *Colcondamyia auditrix* nov. sp.(Diptera: Sarcophagidae), a parasite which is attracted by the mating song of its host, *Okanagana rimosa* (Homoptera: Cicadidae). The *Canadian Entomologist*, 108(1), pp.61-68.
31. Lyons, P.C., Plattner, R.D. and Bacon, C.W., 1986. Occurrence of peptide and clavine ergot alkaloids in tall fescue grass. *Science*, 232(4749), pp.487-489.
32. Kucht, S., Groß, J., Hussein, Y., Grothe, T., Keller, U., Basar, S., König, W.A., Steiner, U. and Leistner, E., 2004. Elimination of ergoline alkaloids following treatment of *Ipomoea asarifolia* (Convolvulaceae) with fungicides. *Planta*, 219(4), pp.619-625.
33. Alhubaishi, A.A.A. and Abdel-Kader, M.I.A., 1991. Phyllosphere and phylloplane fungi of qat in Sana'a, Yemen Arab Republic. *Journal of Basic Microbiology*, 31(2), pp.83-89.
34. Mahmoud, A.L., 2000. Mycotoxin-producing potential of fungi associated with qat (*Catha edulis*) leaves in Yemen. *Folia Microbiologica*, 45(5), pp.452-456.
35. Peláez, F., Collado, J., Arenal, F., Basilio, A., Cabello, A., Matas, M.D., Garcia, J.B., Del Val, A.G., González, V., Gorrochategui, J. and Hernández, P., 1998. Endophytic fungi from plants living on gypsum soils as a source of secondary metabolites with antimicrobial activity. *Mycological Research*, 102(6), pp.755-761.
36. Wang, X., He, Z. and Wei, C., 2017. A new cicada species of *Psalmocharias* Kirkaldy feeding on an *Ephedra* plant from China (Hemiptera: Cicadidae). *Zootaxa*, 4290(2), pp.367-372.
37. Sanborn, A.F. and Phillips, P.K., 2011. Elevation of a subspecies of *Tibicen* (Hemiptera: Cicadoidea: Cicadidae) to a full species. *The Southwestern Naturalist*, 56(3), pp.363-368.
38. Matsuura, Y., Moriyama, M., Łukasik, P., Vanderpool, D., Tanahashi, M., Meng, X.Y., McCutcheon, J.P. and Fukatsu, T., 2018. Recurrent symbiont recruitment from fungal parasites in cicadas. *Proceedings of the National Academy of Sciences*, p.201803245.
39. Cooley, J.R., 1999. Sexual behavior in North American cicadas of the genera *Magicicada* and *Okanagana* (Doctoral dissertation, University of Michigan).
40. Short, D.P., Double, M., Nuss, D.L., Stauder, C.M., MacDonald, W. and Kasson, M.T., 2015. Multilocus PCR assays elucidate vegetative incompatibility gene profiles of *Cryphonectria parasitica* in the United States. *Applied and Environmental Microbiology*, pp.AEM-00926.

bioRxiv preprint doi: https://doi.org/10.1101/375105; this version posted July 24, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



41. Hodge, K.T., Hajek, A.E. and Gryganskyi, A., 2017. The first entomophthoralean killing millipedes, *Arthrophaga myriapodina* n. gen. n. sp., causes climbing before host death. *Journal of Invertebrate Pathology*, 149, pp.135-140.
42. Vilgalys, R. and Hester, M., 1990. Rapid genetic identification and mapping of enzymatically amplified ribosomal DNA from several *Cryptococcus* species. *Journal of Bacteriology*, 172(8), pp.4238-4246.
43. James, T.Y., Kauff, F., Schoch, C.L., Matheny, P.B., Hofstetter, V., Cox, C.J., Celio, G., Gueidan, C., Fraker, E., Miadlikowska, J. and Lumbsch, H.T., 2006. Reconstructing the early evolution of Fungi using a six-gene phylogeny. *Nature*, 443(7113), p.818.
44. Kumar, S., Stecher, G. and Tamura, K., 2016. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Molecular Biology and Evolution*, 33(7), pp.1870-1874.
45. Joshi, N.A. and Fass, J.N., 2011. Sickle: A sliding-window, adaptive, quality-based trimming tool for FastQ files (Version 1.33)[Software].
46. Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R. and Thompson, J.D., 2007. Clustal W and Clustal X version 2.0. *Bioinformatics*, 23(21), pp.2947-2948.
47. Posada, D. and Crandall, K.A., 1998. Modeltest: testing the model of DNA substitution. *Bioinformatics* (Oxford, England), 14(9), pp.817-818.
48. Eddy, S.R., 1998. Profile hidden Markov models. *Bioinformatics* (Oxford, England), 14(9), pp.755-763.
49. Pearson, W.R., Wood, T., Zhang, Z. and Miller, W., 1997. Comparison of DNA sequences with protein sequences. *Genomics*, 46(1), pp.24-36.
50. Stanke, M. and Waack, S., 2003. Gene prediction with a hidden Markov model and a new intron submodel. *Bioinformatics*, 19: ii215–ii225.
51. Lomsadze, A., Ter-Hovhannisyan, V., Chernoff, Y.O. and Borodovsky, M., 2005. Gene identification in novel eukaryotic genomes by self-training algorithm. *Nucleic Acids Research*, 33(20), pp.6494-6506.
52. Hoang, D.T., Chernomor, O., von Haeseler, A., Minh, B.Q. and Vinh, L.S., 2017. UFBoot2: improving the ultrafast bootstrap approximation. *Molecular Biology and Evolution*, 35(2), pp.518-522.

bioRxiv preprint doi: https://doi.org/10.1101/375105; this version posted July 24, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



53. Nguyen, L.T., Schmidt, H.A., von Haeseler, A. and Minh, B.Q., 2014. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution*, 32(1), pp.268-274.
54. Marçais, G., Delcher, A.L., Phillippy, A.M., Coston, R., Salzberg, S.L. and Zimin, A., 2018. MUMmer4: A fast and versatile genome alignment system. *PLoS Computational Biology*, 14(1), p.e1005944.
55. Eddy, S.R., 2011. Accelerated profile HMM searches. *PLoS Computational Biology*, 7(10), p.e1002195.
56. Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J., 1990. Basic local alignment search tool. *Journal of Molecular Biology*, 215(3), pp.403-410.
57. Hoff, K.J. and Stanke, M., 2013. WebAUGUSTUS—a web service for training AUGUSTUS and predicting genes in eukaryotes. *Nucleic Acids Research*, 41(W1), pp.W123-W128.
58. Bailey, T.L. and Elkan, C., 1994. Fitting a mixture model by expectation maximization to discover motifs in bipolymers. 28-36.
59. Huerta-Cepas, J., Forslund, K., Coelho, L.P., Szklarczyk, D., Jensen, L.J., von Mering, C. and Bork, P., 2017. Fast genome-wide functional annotation through orthology assignment by eggNOG-mapper. *Molecular Biology and Evolution*, 34(8), pp.2115-2122.
60. Huerta-Cepas, J., Serra, F. and Bork, P., 2016. ETE 3: reconstruction, analysis, and visualization of phylogenomic data. *Molecular Biology and Evolution*, 33(6), pp.1635-1638.
61. Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q. and Chen, Z., 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology*, 29(7), p.644.
62. Huson, D.H., Mitra, S., Ruscheweyh, H.J., Weber, N. and Schuster, S.C., 2011. Integrative analysis of environmental sequences using MEGAN4. *Genome Research*, pp.gr-120618.
63. Katoh, K. and Standley, D.M., 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution*, 30(4), pp.772-780.
64. Capella-Gutiérrez, S., Silla-Martínez, J.M. and Gabaldón, T., 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*, 25(15), pp.1972-1973.

bioRxiv preprint doi: https://doi.org/10.1101/375105; this version posted July 24, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



65. Li, L., Stoeckert, C.J. and Roos, D.S., 2003. OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Research*, 13(9), pp.2178-2189.
66. Price, M.N., Dehal, P.S. and Arkin, A.P., 2010. FastTree 2–approximately maximum likelihood trees for large alignments. *PloS one*, 5(3), p.e9490.
67. Stamatakis, A., 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, 30(9), pp.1312-1313.

## Supplementary Figure Legends

**Fig. S1.** Mean (A) conidia and (B) resting spore dimensions for three *Massospora* species sampled from infected cicadas. White bars indicate overall mean values while black bars indicate the range of variation in spore dimensions. Twenty-five conidia or resting spores were measured for each specimen except for *M. levispora* and *M. platypediae* resting spores, in which 50 spores were measured. Sampling sizes were as follows: For conidia *Mc* = 15, *Ml* = 8, and *Mp* = 14; For resting spores, *Mc* = 11, *Ml* = 1, *Mp* = 1.

**Fig. S2.** Concatenated LSU+SSU maximum likelihood (ML) tree consisting of *Massospora* species and related species in the Entomophthorales. **B**ootstrap support is indicated near each node and only values greater than 70% are shown.

**Fig. S3.** Representative total base peak chromatograms (BPC) from the global metabolomics analysis of *M. cicadina*, *M. platypediae*, and a healthy cicada controls.

**Fig. S4.** Representative total base peak chromatograms (BPC) from the global metabolomics analysis of *M. cicadina* resting spores versus *M. cicadina* conidia.

**Fig. S5.** Psilocybin and cathinone biosynthesis pathway gene discovery pipeline.

**Fig. S6.** Candidate ML gene phylogenies potentially involved in cathinone and psilocybin biosynthesis pathway from *M. cicadina* and *M. platypediae*. Phylogenies include (A-C) three Group II pyridoxal-dependent decarboxylases (Pyridoxal_deC), (D) one Phosphatidylserine

bioRxiv preprint doi: https://doi.org/10.1101/375105; this version posted July 24, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



decarboxylase (PS_Dcarbxylase), (E-G) three methyltransferases (Methyltransf), (H) one acetyl-coenzyme A acetyltransferase (Thiolase_N), (I) one Enoyl-(Acyl carrier protein) reductase (adh_short_C2), (J) one Aminotransferase class I and II (Aminotran_1_2), (K) one 3-hydroxyacyl-CoA dehydrogenase, NAD binding domain (3HCDH_N), (L) one Aldo-keto reductase (Aldo_ket_red), (M-P) ten Cytochrome P450 (p450).

**Fig. S7.** Differential fold change and regulation of 2-Hydroxyestradiol and 3-O-acetylecdysone 2-phosphate expression in conidial plugs of *M. platypediae* and *M. cicadina* using global metabolomics. The former metabolite is an endogenous steroid and the latter is a product of ecdysone, a steroidal prohormone of the major insect molting hormone 20-hydroxyecdysone.

**Movie S1.** Living *Platypedia putnami* with conidial and resting spore infections.
10.6084/m9.figshare.6855062

**Movie S2.** Living Brood V *Magicicada septendecim* with conidial spore infection.

10.6084/m9.figshare.6855215