



OPEN ACCESS

EDITED BY

Mohamed I. Husseiny Elsayed,
Beckman Research Institute, City of
Hope, United States

REVIEWED BY

Victor Meza Carmen,
Michoacana University of San Nicolás de
Hidalgo, Mexico
Yeseul Choi,
Duke University, United States

*CORRESPONDENCE

Zack Saud

✉ saudz@cardiff.ac.uk

Tariq M. Butt

✉ t.butt@swansea.ac.uk

RECEIVED 30 January 2026

REVISED 15 April 2026

ACCEPTED 20 April 2026

PUBLISHED 22 May 2026

CITATION

Saud Z, Luo Y, Wood MJ, Boostrom I,
MacLachlan BJ, Stanton RJ, Wang C and
Butt TM (2026) Discovery of a fungal Tc
toxin complex and functional
mycoserpin through a novel two-by-two
comparative genomics approach.
Front. Cell. Infect. Microbiol. 16:1800461.
doi: 10.3389/fcimb.2026.1800461

COPYRIGHT

© 2026 Saud, Luo, Wood, Boostrom,
MacLachlan, Stanton, Wang and Butt. This
is an open-access article distributed under
the terms of the [Creative Commons
Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use,
distribution or reproduction in other
forums is permitted, provided the
original author(s) and the copyright
owner(s) are credited and that the
original publication in this journal is
cited, in accordance with accepted
academic practice. No use, distribution
or reproduction is permitted which does
not comply with these terms.

Discovery of a fungal Tc toxin complex and functional mycoserpin through a novel two-by-two comparative genomics approach

Zack Saud^{1,2*}, Yujuan Luo^{3,4}, Martyn J. Wood⁵, Ian Boostrom²,
Bruce J. MacLachlan², Richard J. Stanton², Chengshu Wang^{3,4}
and Tariq M. Butt^{1*}

¹Department of Biosciences, College of Science, Swansea University, Swansea, Wales, United Kingdom, ²Infection and Immunity, School of Medicine, Cardiff University, Cardiff, United Kingdom, ³Key Laboratory of Insect Developmental and Evolutionary Biology, The Chinese Academy of Sciences (CAS) Center for Excellence in Molecular Plant Sciences, Shanghai Institute of Plant Physiology and Ecology, Chinese Academy of Sciences, Shanghai, China, ⁴School of Life Science and Technology, ShanghaiTech University, Shanghai, China, ⁵Institute of Molecular Biology and Biotechnology, Foundation for Research and Technology Hellas, Heraklion, Greece

Introduction: Fungi have been a rich source of pharmaceuticals such as antibiotics, immunosuppressants, and cholesterol-lowering drugs; however, their therapeutic potential remains largely untapped due to difficulties in culturing and elucidating the genetic basis of beneficial traits. Fungi contain ‘cryptic’ genes that are expressed under certain, and often obscure, growth conditions and can produce complex compounds that are difficult to synthesize economically. Developments in genome sequencing and DNA-synthesis technologies offer new opportunities to produce such compounds using biotechnological techniques, however, accurately identifying useful and novel genes, a prerequisite for such approaches, remains challenging.

Methods: We present a novel ‘two-by-two’ comparative genomics pipeline for comprehensive gene analysis of selected fungal groups, enabling more confident identification of unique genes across the analyzed species. The approach compares gene sets from two strains of the same species with those from two strains of different species or families within a fungal order. Self-clustering orthologs that are unique to strains from the same species provide higher confidence in identifying species-specific proteins and help reduce noise from low-quality genome assemblies and gene prediction errors.

Results: We validated our method in a well-studied fungal group, correctly identifying known species-specific genes and gene clusters, and discovering a novel functional myco-serpin and a fungal Tc toxin complex in *Metarhizium*.

Discussion: Elucidating the genes underlying beneficial traits in fungi presents significant challenges, largely due to the unique and relatively complex aspects of their lifestyles. The two-by-two approach shows broad potential for fungal genome mining and bioprospecting, demonstrated by our discovery of the first fungal Tc toxin complex and a functional myco-serpin in *Metarhizium* (Hypocreales), and is applicable to other fungal orders such as Eurotiales and Xylariales. Furthermore, the two-by-two approach can be adapted to other organisms with genome architectures similar to fungi.

KEYWORDS

bioinformatics, biosynthetic gene cluster, comparative genomics, fungi, hypocreales, infection, serpin, Tc toxin

Introduction

Fungi have played crucial roles in human industrial activities for millennia (Liu et al., 2018). The discovery of Penicillin marked the beginning of extensive exploration into fungal metabolites (Fleming, 1929; Alberti et al., 2017) that led to the development of further classes of fungal-derived pharmaceuticals such as; cyclosporine (Rüegger et al., 1976), an immunosuppressive agent, lovastatin (Alberts et al., 1980), a cholesterol-lowering drug, and various other antibiotics with differing modes of action, such as fusidic acid and cephalosporin (Newton and Abraham, 1955; Godtfredsen et al., 1962). Despite these achievements, the development of fungal therapeutics has lagged in comparison to other organisms for several reasons. Fungi exhibit intricate life cycles, encompassing the production of spores, mycelia, and fruiting bodies. Notably, the biosynthesis of bioactive compounds in fungi often occurs at specific developmental stages, with some metabolites being structurally complex and large. This inherent complexity can render their extraction and purification more cost-effective than chemical synthesis, as exemplified by the commercial production of cyclosporine (Anjum and Iram, 2015). Cultivating fungi and extracting metabolites often necessitates specialized techniques, with fewer than 5% of species having been successfully cultured to date (Kaerberlein et al., 2002). Additionally, large-scale extraction of useful compounds demands substantial investment in tailored bioreactors and monitoring systems for each species, making plant-based natural product discovery more financially attractive to pharmaceutical companies.

Fungi generally harbor more “cryptic” genes compared to other organisms that only activate in response to specific, often obscure, environmental cues or stressors, rendering them undetectable to transcriptomic, proteomic, and gene-knockout screening (Brakhage, 2013). This presents a significant opportunity for fungal natural product discovery through advanced comparative genomic

approaches, as the presence of cryptic genes indicates that the biosynthetic potential of many cultivable fungi is likely much greater than currently recognized (Scherlach and Hertweck, 2021). With the advent of NGS and improved DNA synthesis technologies, novel approaches have been developed such as the “HEX” synthetic biology platform, wherein gene clusters are cloned from hard-to-culture fungi, refactored, and optimized to express in *Saccharomyces cerevisiae*, a traditional ‘cell factory’ for pharmaceutical production due to its simple growth requirements and rapid proliferation (Harvey et al., 2018). The compact genomes of Ascomycota fungi (30–40 Mbp) (Mohanta and Bae, 2015) make full genome sequencing and assembly both affordable and feasible, as highlighted by the recent surge of fungal genome submissions to databases like GenBank, and we have completed telomere-to-telomere assemblies, including mitochondrial genomes, for Ascomycota fungi for under \$1000 (Saud et al., 2021). While some fungal genome sequences are high-quality, telomere-to-telomere assemblies or contain some complete chromosomes, many are assembled solely using short-read sequencing technologies and are thus more fragmented. The impact of this fragmentation on gene and biosynthetic gene cluster prediction accuracy remains unclear.

Despite the opportunities that have arose from technological advancements, there remains challenges as the bioinformatics analyses employed to identify functionally relevant fungal genes lack standardization and entail inherent complexity, often utilizing homology-based searches of biosynthetic gene cluster databases, likely biasing against discovery of completely novel clusters (Scherlach and Hertweck, 2021). Fungi boast higher gene densities relative to plants and animals (Galagan et al., 2005), with variation of gene sets being observed, even between closely related strains. This exacerbates the likelihood of false positive and negative discovery rates at the bioinformatics stage, especially given the intrinsic variance that can occur from differing genome assembly pipelines (sequencing technologies, read pre-processing strategies, assembler algorithm usage, polishing methods) and gene prediction approaches (e.g. *ab-initio*, homology-based, or RNA-seq guided) (Ter-Hovhannisyan et al., 2008). Given the costs and low throughput of designing, cloning, and expressing genes in heterologous hosts (Harvey et al., 2018), bioinformatics tools that confidently prioritize candidate genes are essential. There is a need for simple methods that can accurately identify strain-specific, species-wide, or genus-wide genes across diverse fungal genomes, achieving high

Abbreviations: ARSEF, ARS Collection of Entomopathogenic Fungal Cultures; Bb, *Beauveria bassiana*; BGC, Biosynthetic Gene Cluster; BUSCO, Benchmarking Universal Single-Copy Orthologs; Cc, *Cordyceps catenianmullata*; Cf, *Cordyceps fumosorosea*; Cj, *Cordyceps javanica*; Cm, *Cordyceps militaris*; Ef, *Epichloe festucae*; Mb, *Metarhizium brunneum*; Tr, *Trichoderma reesei*; Mbp, Million base pairs; NCBI, National center for biotechnology information; ORF, Open reading frame; r.m.s.d, Root mean square deviation; WT, Wild type.

confidence despite inherent genomic error rates. To address this need, we present the “two-by-two” approach, which compares gene sets from two strains of the same species with those from two strains of different species or families within a fungal order. Self-clustering orthologs between strains of the same species provide higher confidence in identifying species-specific proteins and help reduce noise from low-quality genome assemblies and gene prediction errors in comparison to comparing species using only a single representative strain. While pan-genome analysis examines all strains of a species, the two-by-two approach compares only two strains to identify species-specific genes, making it more suitable for organisms with less abundant genomic data.

We conducted a proof-of-concept study on Hypocreales fungi, an order with several well-studied species and multiple genomes of varying quality, including some telomere-length assemblies, available in GenBank. Notable members of Hypocreales include *Trichoderma* spp., widely used in agriculture as biocontrol agents against plant pathogens (Woo et al., 2023); *Cordyceps* spp., entomopathogens valued in traditional medicine, with compounds under study for immunomodulatory and anticancer effects (Tuli et al., 2014); *Beauveria bassiana*, a common pest management agent (Zimmermann, 2007); and *Metarhizium* spp., also used in pest control, with conidia that have been demonstrated to be larvicidal against disease-vectoring mosquitoes (Butt et al., 2013) and shown to produce compounds with immunomodulatory (Lv et al., 2020) and anticancer (Dornetshuber-Fleiss et al., 2013) potential. Many genes involved in producing bioactive compounds in these fungi are well-characterized, allowing effective benchmarking of our approach. Additionally, the diverse ecological niches occupied by these closely related fungi enable assessment of the hits by plausibility. The test group includes five closely related species from Cordycipitaceae, two distantly related species from Clavicipitaceae, and one from Hypocreaceae, allowing evaluation of the approach in distinguishing unique genes across varying levels of relatedness.

Using our two-by-two approach, we identified two novel proteins in *Metarhizium brunneum*: a functional serpin-like protein with homology to *Drosophila melanogaster* serpin 42Dd, an immune modulating protein, and a fungal Tc toxin complex with structural homology to bacterial Tc toxin complexes. Our findings provide a valuable reference for genome mining within the Hypocreales order and demonstrate that this analysis method can be applied to other fungal orders or organisms with high gene densities and significant interspecies gene variation, where potentially useful but under-characterized gene products are present.

Results

Genome attributes of species used in analyses

Genomes were chosen from the Hypocreales families *Cordycipitaceae*, *Clavicipitaceae*, and *Hypocreaceae*, utilizing criteria such as genome quality as assessed by BUSCO score and assembly N50, giving preference to assemblies that were less contiguous.

A total of 16 genomes comprising two strains from each species within three families in the order Hypocreales were subjected to orthologous protein and biosynthetic gene cluster analyses (Table 1). 6 of the genomes in the group had all 7 chromosomes sequenced telomere length, and these included; both strains of *Trichoderma reesei*, *Cordyceps javanica* isolate Apopka 97, *Cordyceps militaris* isolate ATCC 34164 (Kramer and Nodwell, 2017), *Metarhizium brunneum* isolate ARSEF 4556 (Saud et al., 2021), and *Epichloe festucae* isolate FL1. We performed gene predictions for species with genbank entries that lacked any gene information and the included; *Cordyceps fumosorosea* grCorFumo1 (10158 genes), *Cordyceps cateniannulata* FRD 24 (10850 genes), *C. cateniannulata* MBC 234 (11452 genes), *E. festucae* RoseCity (6979), *T. reesei* CBS999.97 (9196 genes), *T. reesei* QM6a (9298 genes).

Two-by-two biosynthetic gene cluster analysis of hypocreales

The two-by-two approach used two strains from each species to perform *in-silico* biosynthetic gene cluster and orthogonal protein analyses against other species in the group (Figure 1). We performed an *in-silico* analysis using the fungal version of antiSMASH to quantify predicted biosynthetic gene clusters in each fungal species and to identify shared and species-specific clusters that had homology to known BGCs in the MIBiG database, providing a resource to support future research (Figure 2). *M. brunneum* exhibited the highest number of predicted biosynthetic gene clusters (BGCs), followed by *B. bassiana*, *Cordyceps cateniannulata*, and *C. javanica* (Figure 2a). Approximately one-third to one-quarter of the predicted gene clusters for each species had genes matched to known BGCs. Out of the 64 BGCs that had hits to known BGC clusters, 43 (67.2%) were unique to a single species within the Hypocreales fungi analyzed. For BGCs with known MIBiG database hits, we present a heat map showing the presence or absence across species and a color-coded key of the proportion of genes with significant BLAST matches in the predicted gene cluster (Figure 2b). The most prevalent gene cluster detected across all species of Hypocreales resembled that of the BGC for choline biosynthesis, a metabolite essential for fungal growth (Hai et al., 2019). Clusters resembling those of the siderophore metachelin C and antimicrobial poly-amino acid ϵ -poly-L-lysine were also detected in most fungal species analyzed (Figure 2b), and these compounds are known to be commonly produced by fungi (Giuliano Garisto Donzelli et al., 2015; Wang et al., 2021; Zhang et al., 2021; Pokhrel and Coleman, 2023).

Interestingly, a lower scoring hit (only some genes from the known cluster present) was observed for a gene cluster associated with the synthesis of the squalene-synthesis inhibiting compound squalenyl S1 within the Cordycipitaceae and Hypocreaceae families, but genes from this cluster were absent in the genomes of Clavicipitaceae species (Koulman et al., 2007; Bonsch et al., 2016). BGCs containing genes for the pathway synthesizing the steroid clavaric acid were also absent from species in the family Clavicipitaceae, and were also absent from *C. javanica* (Jayasuriya et al., 1998). Hits for genes resembling the genes in the cluster for fumosorinone synthesis, a compound found to inhibit protein tyrosine phosphatase 1B, a major negative regulator of the insulin

TABLE 1 Assembly and annotation statistics for Hypocreales fungi used in this study.

Species	Family	Isolate	Assembly accession	Release year	Total length	Scaffolds (contigs)	N50	Full chromosomes (plasmids)	Predicted proteins	Protein busco (N = 4494)
<i>Cordyceps javanica</i>	Cordycipitaceae	Apopka 97	GCA_051103115.1	2024	35.14	–	–	7 (1)	10469	96.90%
<i>Cordyceps javanica</i>		IJ2G	GCA_006981975.1	2019	34.97	173 (658)	1822545	0	11142	94.40%
<i>Cordyceps fumosorosea</i>		ARSEF 2679	GCA_001636725.1	2016	33.49	430 (685)	872179	0	10061	95.60%
<i>Cordyceps fumosorosea</i>		grCorFumo1	GCA_963580265.1	2023	34.32	–	–	9 (1)	*10158	94.00%
<i>Cordyceps militaris</i>		ATCC 34164	GCA_008080495.1	2017	33.62	–	–	7 (0)	9287	92.40%
<i>Cordyceps militaris</i>		CM01	GCA_000225605.1	2011	32.27	32 (597)	4551492	0	9651	95.90%
<i>Cordyceps cateniannulata</i>		FRD 24	GCA_028828415.1	2023	33.63	0 (16)	4260819	0	*10850	96.90%
<i>Cordyceps cateniannulata</i>		MBC 234	GCA_030411495.1	2023	33.99	0 (1,338)	104394	0	*11452	95.90%
<i>Beauvaria bassiana</i>		ERL836	GCA_010099065.1	2020	35.40	0 (14)	3988868	0	10399	91.30%
<i>Beauvaria bassiana</i>		JEF-350	GCA_021365345.1	2022	35.70	0 (8)	5649914	0	7829	95.50%
<i>Metarhizium brunneum</i>	Clavicipitaceae	ARSEF 4556	GCA_013426205.1	2020	37.80	–	–	7 (1)	11420	99.10%
<i>Metarhizium brunneum</i>		ARSEF 3297	GCF_000814965.1	2015	37.10	92 (180)	1825569	0	10689	97.10%
<i>Epichloe festucae</i>		FII	GCA_003814445.1	2018	35.00	–	–	7 (1)	7034	87.40%
<i>Epichloe festucae</i>		RoseCity	GCA_016859245.1	2021	31.00	0 (19)	2735662	0	*6979	84.50%
<i>Trichoderma reesei</i>	Hypocreaceae	CBS999.97	GCA_016806755.1	2021	34.30	–	–	7 (0)	*9196	98.70%
<i>Trichoderma reesei</i>		QM6a	GCA_002006585.1	2017	34.90	–	–	7 (0)	*9298	98.00%
		Bold Isolate = RefSeq							* = Gene predictions performed in this study	

Bold isolate names indicates that the assembly is the reference genome sequence for that species. An asterisk in front of the predicted proteins number indicates that the gene prediction was performed in this study.

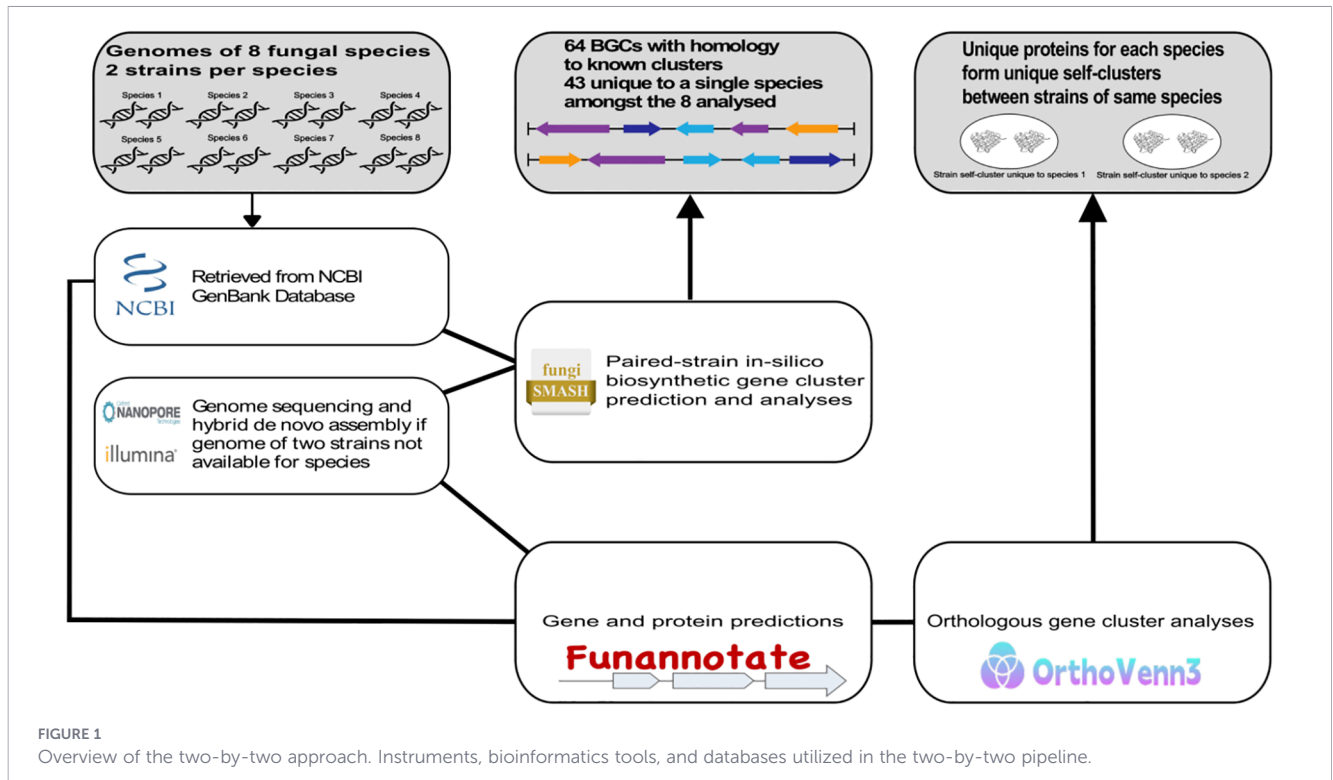


FIGURE 1 Overview of the two-by-two approach. Instruments, bioinformatics tools, and databases utilized in the two-by-two pipeline.

signaling pathway, were detected in both species of *Cordyceps fumosorosea* from which the compound was first identified. Interestingly, we also detected a high homology hit in both *C. cateniannulata* genomes analyzed, to our knowledge, the first time this gene cluster has been reported in the species. A BGC resembling that of beauvericin, a cyclic hexadepsipeptide mycotoxin that has insecticidal, antimicrobial, antiviral, and cytotoxic activities (Wang and Xu, 2012), was detected in all genomes of species

within Cordycipitaceae (with a high percent score in *Beauveria* species and lower percentage hits in *Cordyceps* species), but no genes for the cluster were detected in *C.javanica*. Genes resembling those found in the iron-binding siderophore epichloenin A (Koulman et al., 2012) were detectable in both genomes of *C.javanica* and *C.cateniannulata*, but only present in one genome of *E. festucae*. Some of the genes for the known BGC synthesizing the neurotogenic polyketide Shimalactone A/B gene cluster (Sofiyev

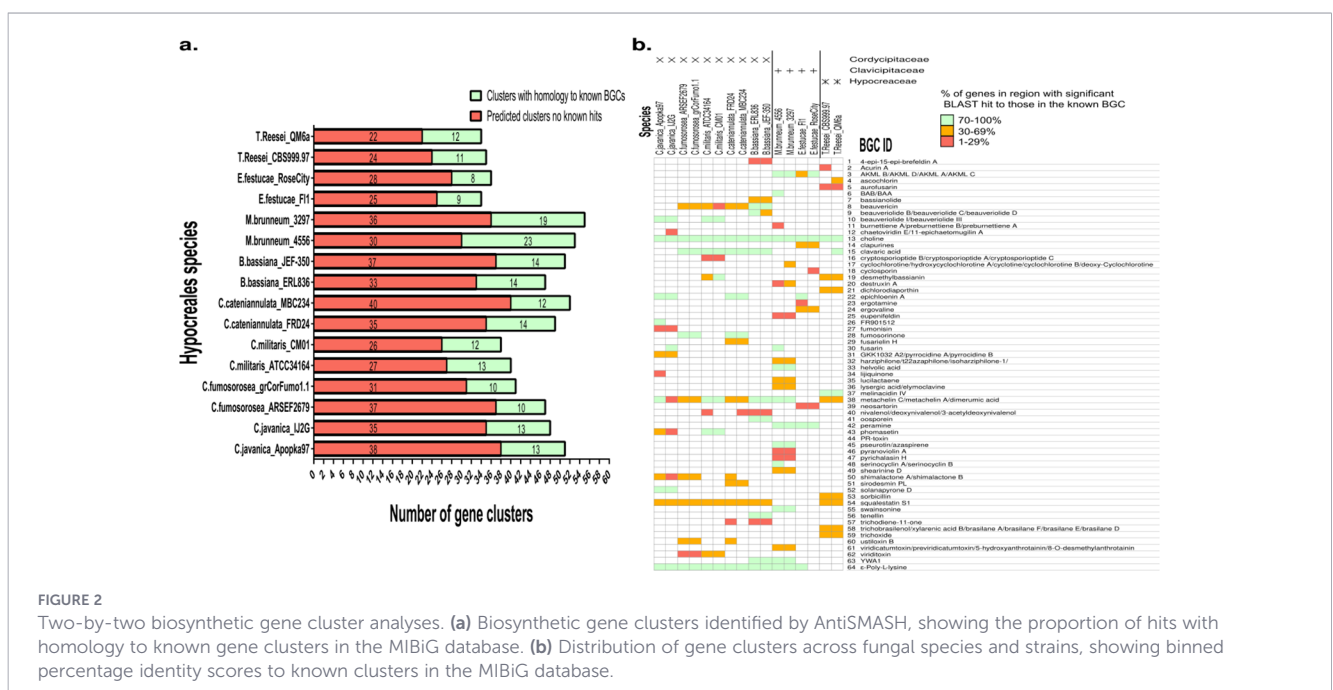


FIGURE 2 Two-by-two biosynthetic gene cluster analyses. (a) Biosynthetic gene clusters identified by AntiSMASH, showing the proportion of hits with homology to known gene clusters in the MIBiG database. (b) Distribution of gene clusters across fungal species and strains, showing binned percentage identity scores to known clusters in the MIBiG database.

et al., 2008) were detected in the Cordyceps species excepting for *C. militaris* wherein no identity matches were detected. Genes matching those that synthesize the beauveriolide compounds were present in Cordyceps and Beauveria, as previously reported (Yin et al., 2020). Gene clusters that were uniquely identified in two strains of a single species and were known to be unique from that species (or at least absent from the other species analyzed) include; clusters resembling those of swainsonine, destruxin A and helvolic acid in *M. brunneum*, clusters resembling those of oosporein and tenellin in *B. bassiana*, clusters resembling those of melinacidin IV in *T. reesei*, and a cluster resembling that of ergovaline in *E. festucae*. Some predicted clusters had high percentage identity hits to known clusters, but were detected in only one of the two strains analyzed. These include clusters for; (2Z,4E,6E,10E)-9-hydroxydodeca-2,4,6,10-tetraenoic acid/(2E,4E,6E,10E)-9-hydroxydodeca-2,4,6,10-tetraenoic acid (BAA/BAB), serinocyclin A/serinocyclin B and fusarin only detected in *M. brunneum* 4556, epichloenin A only present in *E. festucae* Fl1, and clavatic acid only present in *T. reesei* QM6a. It is not clear whether these differences are due to a genuine gene variation between the two strains, or whether they are present in the genome of the second strain but not predicted due to sequencing and/or assembly issues. In this regard, it is worth noting that in the case of *M. brunneum*, the genome of 4556 is telomere length and was assembled using hybrid sequencing technology, whilst the 3297 genome was assembled as part of pioneering efforts to utilize massively parallel sequencing technologies on fungi for the first time (see dates, BUSCO scores and contig information in Table 1). Together, these results underscore the complexity of BGC prediction and demonstrate the value of the two-by-two approach in validating putative clusters through confirmation in at least two strains per species. *In-silico* BGC predictions cannot determine the compounds produced; functional validation requires knockout or heterologous expression, both non-trivial methods that

benefit from better-validated clusters identified by the two-by-two approach.

Two-by-two analyses of orthogonal proteins

To identify unique proteins in each species, orthologous protein analysis was performed with OrthoVenn3 using the OrthoFinder algorithm and using the two-by-two approach to identify strain self-clusters that were absent from other species (Figure 3).

A total of 3891 clusters were seen to contain orthologous proteins from the 16 strains comprising 8 species that were analyzed, and this represents the core genes for the group shared across all the species. The highest number of species self-clusters was observed between the two strains of *M. brunneum* analyzed (1269 clusters), followed by the two strains of *T. reesei* (847 clusters). *B. bassiana* formed 559 clusters, *C. militaris* 383 clusters, *C. cateniannulata* 330 clusters, *E. festucae* 273 clusters, *C. javanica* 236 clusters and *C. fumosorosea* 198 clusters (Supplementary Table 1). Singlet proteins that did not form a cluster with any other proteins ranged in number for each species from 1411 for *C. javanica* strain IJ2G to 56 for *M. brunneum* ARSEF 3297 with an average of around 250 proteins across all 16 strains analyzed (Supplementary Table 2). Clusters that were found to form uniquely within a single strain (strain self-orthologs) included; 21 clusters for *B. bassiana* JEF350, 17 clusters for *C. javanica* Apopka97, 11 clusters for *C. cateniannulata* MBC234, 9 clusters for *M. brunneum* ARSEF4556, 7 clusters for *C. cateniannulata* FRD24, 6 clusters for *C. javanica* IJ2G, 5 clusters for both *C. fumosorosea* grCorFumo1 and *B. bassiana* ERL836, 4 clusters for *E. festucae* Fl1, 3 clusters for *E. festucae* RoseCity, 2 clusters for *C. fumosorosea* ARSEF2679, 1 cluster for both *C. militaris* CM01 and *T. reesei* QM6a and no unique strain self-clusters were observed for *C. militaris*

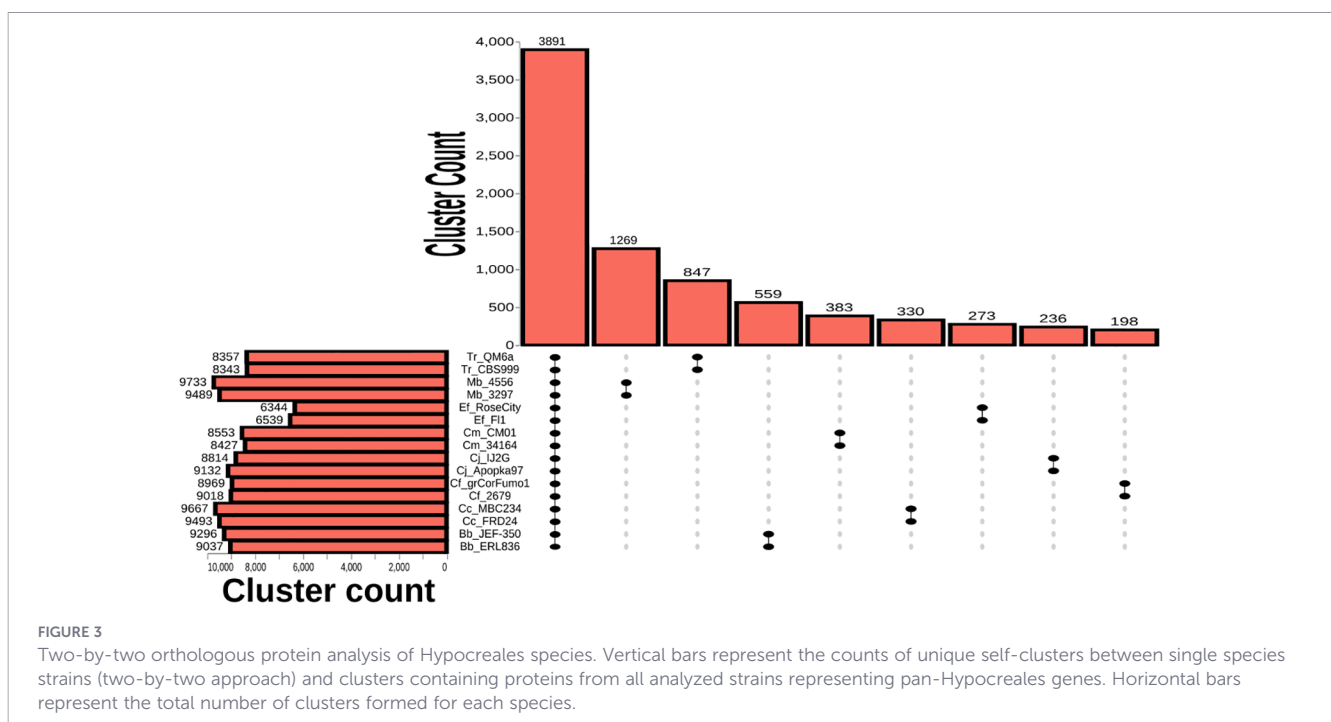


TABLE 2 Species-specific self-clusters identified via the two-by-two approach with previously characterized unique functions.

Species	Cluster ID	Number of proteins in cluster	Swiss-prot hit (Entry - species)	Swiss-prot hit protein name	Members of cluster	Function	Reference
<i>B. bassiana</i>	cluster10663	3	OPS1 - <i>Beauveria bassiana</i>	Orsellinic acid synthase	Bb_ERL836 KAF1732856.1; Bb_JEF-350 KAH8713103.1; Bb_JEF-350 KAH8713104.1	Gene cluster that mediates the biosynthesis of the bibenzoquinone oosporein, a metabolite required for fungal virulence that acts by evading host immunity to facilitate fungal multiplication in insects	(Feng et al., 2015)
	cluster11284	2	OPS7 - <i>Beauveria bassiana</i>	Oxidoreductase OpS7	Bb_ERL836 KAF1732862.1; Bb_JEF-350 KAH8713097.1		
	cluster11283	2	J5J924 - <i>Beauveria bassiana</i>	MFS transporter OpS2	Bb_ERL836 KAF1732858.1; Bb_JEF-350 KAH8713102.1		
	cluster11282	2	OPS3 - <i>Beauveria bassiana</i>	Oosporein cluster regulator OpS3	Bb_ERL836 KAF1732857.1; Bb_JEF-350 KAH8713101.1		
<i>C. militaris</i>	cluster12734	2	No hit		Cm_34164 ATY65128.1; Cm_CM01 EGX93064.1	Product encoded by gene CNS1 of the Cordycepin biosynthetic gene cluster	(Xia et al., 2017)
	cluster12735	2	O05502 - <i>Bacillus subtilis</i>	Uncharacterized protein Ydhj	Cm_34164 ATY65129.1; Cm_CM01 EGX93065.1	Product encoded by gene CNS2 of the Cordycepin biosynthetic gene cluster	
	cluster12736	2	No hit		Cm_34164 ATY65134.1; Cm_CM01 EGX93066.1	Product encoded by gene CNS3 of the Cordycepin biosynthetic gene cluster	
<i>M. brunneum</i>	cluster13200	2	P29138 - <i>Metarhizium anisopliae</i>	Cuticle-degrading protease PR1	Mb_3297 KID60942.1; Mb_4556 QLI68484.1	Cuticle-degrading protease, Chymoelastase, PR1. Capable of breaching the insect cuticle	(St Leger et al., 1992)
	cluster13207	2	L8G6I7 - <i>Pseudogymnoascus destructans</i>	Subtilisin-like protease 1	Mb_3297 KID61215.1; Mb_4556 QLI74370.1	Major secreted subtilisin-like serine endopeptidase. Degrades the nonhelical regions of collagen that function in the cross-linking of the helical components	(Bagga et al., 2004; Pannkuk et al., 2015)
	cluster13386	2	L8GD75 - <i>Pseudogymnoascus destructans</i>	Subtilisin-like protease 3	Mb_3297 KID66974.1; Mb_4556 QLI70477.1	Major secreted subtilisin-like serine endopeptidase. Degrades the nonhelical regions of collagen that function in the cross-linking of the helical components	(Bagga et al., 2004; Pannkuk et al., 2015)
	cluster13323	2	A0A0B4GDU5 - <i>Metarhizium brunneum</i>	Scytalone dehydratase-like protein Arp1	Mb_3297 KID65390.1; Mb_4556 QLI69143.1	Scytalone dehydratase-like protein; part of the Pks2 gene cluster that mediates the formation of infectious structures (appressoria), enabling these fungi to kill insects faster	(Zeng et al., 2018)
	cluster13990	2	D4AX50 - <i>Arthroderma benhamiae</i>	Subtilisin-like protease 8	Mb_3297 KID76020.1; Mb_4556 QLI67847.1	Secreted subtilisin-like serine protease with keratinolytic activity that contributes to pathogenicity	(Bagga et al., 2004)
	cluster10467	4	E9F8M0 - <i>Metarhizium robertsii</i>	Transmembrane transporter swnT	Mb_3297 KID64719.1; Mb_3297 KID73970.1; Mb_4556 QLI68254.1; Mb_4556 QLI74702.1	Gene cluster that mediates the biosynthesis of swainsonine, a cytotoxic fungal alkaloid and a potential cancer therapy drug.	(Cook et al., 2017)
	cluster13587	2	D4AU27 - <i>Arthroderma benhamiae</i>	Swainsonine transporter swnT	Mb_3297 KID72725.1; Mb_4556 QLI71645.1		
cluster10517	4	D4AU26 - <i>Arthroderma benhamiae</i>	Dioxygenase swnH1	Mb_3297 KID73870.1; Mb_3297 KID73969.1; Mb_4556 QLI67169.1; Mb_4556 QLI67423.1			

(Continued)

TABLE 2 Continued

Species	Cluster ID	Number of proteins in cluster	Swiss-prot hit (Entry - species)	Swiss-prot hit protein name	Members of cluster	Function	Reference
<i>T. reesei</i>	cluster13740	2	E9F8M4 - <i>Metarhizium robertsii</i>	Dioxygenase swmH2	Mb_3297 KID73974.1; Mb_4556 QLI68135.1	Short chain dehydrogenase; part of the SOR gene cluster that mediates the biosynthesis of sorbicillinoids, a diverse group of yellow secondary metabolites that restrict growth of competing pathogenic fungi but not of bacteria.	(Derntl et al., 2017)
	cluster14860	2	G0R6S7 - <i>Trichoderma reesei</i>	Short chain dehydrogenase sor7	Tr_CBS999 TreiseiCBS999.97_006656-T1; Tr_QM6a TreiseiQM6a_006735-T1		
<i>E. festucae</i>	cluster12871	2	A0A024FA41 - <i>Lecanicillium</i> sp	Fatty acid hydroxylase vImA	Ef_FL QPG9328.1; Ef_RoseCity EfestucaefRoseCity_000440-T1	Fatty acid hydroxylase; part of the gene cluster that mediates the biosynthesis of verlamelin, a lipopeptide that exhibits antifungal activity against plant pathogenic fungi. This gene product has previously been observed in <i>E. festucae</i> and has been detected in <i>E. festucae</i> infected <i>Festuca rubra</i>	(Wang et al., 2019)
	cluster13039	2	A0A024F910 - <i>Lecanicillium</i> sp	Nonribosomal peptide synthetase vImS	Ef_FL QPH06327.1; Ef_RoseCity EfestucaefRoseCity_000031-T1		

ATCC34164, *M. brunneum* ARSEF3296 or *T. Reesei* CBS999 (Supplementary Table 3).

Of the self-clusters that formed uniquely between two strains, 18 clusters were found to contain genes that produce 10 previously characterized natural products that have been demonstrated to be unique to the species (Table 2). These included: proteins responsible for synthesizing oosporein in *B. bassiana*, a metabolite required for fungal virulence that acts by evading host immunity to facilitate fungal multiplication in insects (Feng et al., 2015); proteins that produce the compound cordycepin in *C. militaris*, an adenosine derivative shown to have antibiotic, anti-inflammatory, and anti-cancer activities (Xia et al., 2017); the *Metarhizium* proteins Cuticle-degrading protease PR1 (St Leger et al., 1992) and various subtilisin-like proteins known to be involved in cuticle and collagen degradation (Bagga et al., 2004; Pannkuk et al., 2015), a known protein involved in the formation of appressoria (Zeng et al., 2018), and various proteins involved in the production of swainsonine (Cook et al., 2017), a cytotoxic fungal alkaloid and a potential cancer therapy drug. In the non-entomopathogenic fungal species, a protein known to take part in synthesis of sorbicillinoids was detected in *T. reesei*- an anti-fungal metabolite known to be produced by this species (Derntl et al., 2017). Two proteins known to be involved in the synthesis of the antifungal verlamelin were detected in *E. festucae* (Wang et al., 2019). Genes for oosporein, sorbicillin, and swainsonine were also detected in the BGC analyses (Figure 2c), however, it was encouraging to see that multiple proteins in this pathway were also detected as unique species self-clusters in the orthogonal protein analysis pipeline, validating the approach.

BGC algorithms such as antiSMASH have previously been shown to be unable to detect the cordycepin *Cns* cluster (Hong et al., 2024), and we were also unable to detect it using the latest version of this tool but were able to detect 3 of the 4 genes in the cluster using the two-by-two orthologous protein analysis (Table 2). Given that the two-by-two approach was able to detect this gene cluster that could not be detected by BGC identification algorithms, we mapped all the species self-cluster proteins for fungi with telomere length genome assemblies back to their respective genomes, and then compiled a list of potential gene clusters by mapping genes unique to each species that were within 5kb of each other in the genome (Supplementary Table 4). We present this list as a resource that likely contains other novel BGCs for these species that could not be detected by antiSMASH. Overall, the two-by-two approach when used for individual proteome analysis effectively identified known distinct proteins for each species within the group, unaffected by the degree of relatedness or genetic distance among the species analyzed and, compared to the BGC analysis, was seemingly less affected by the quality of the genomes utilized.

Identification of a fungal Tc toxin complex and functional myco-serpin

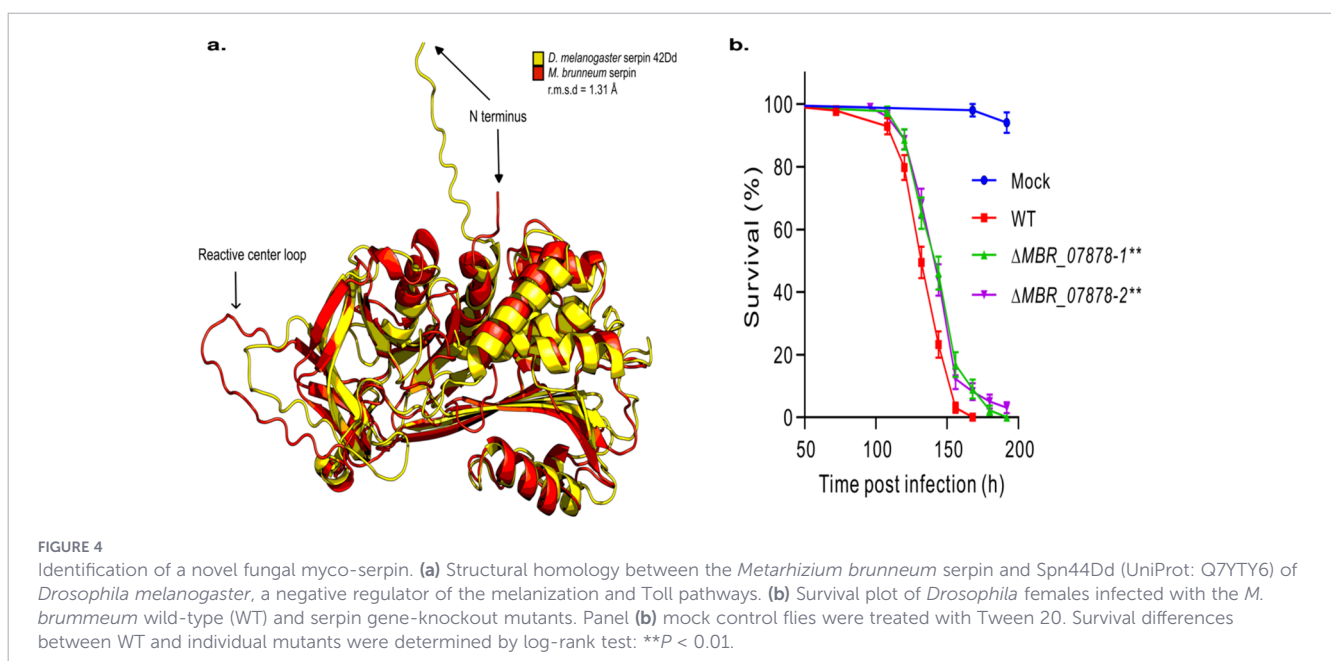
Next, we assessed self-cluster hits that were found to contain proteins with high homology to previously characterized proteins that are involved in pathogenesis, antibiotic production, or distinctive catabolic processes, and we have listed these in Supplementary Table 5.

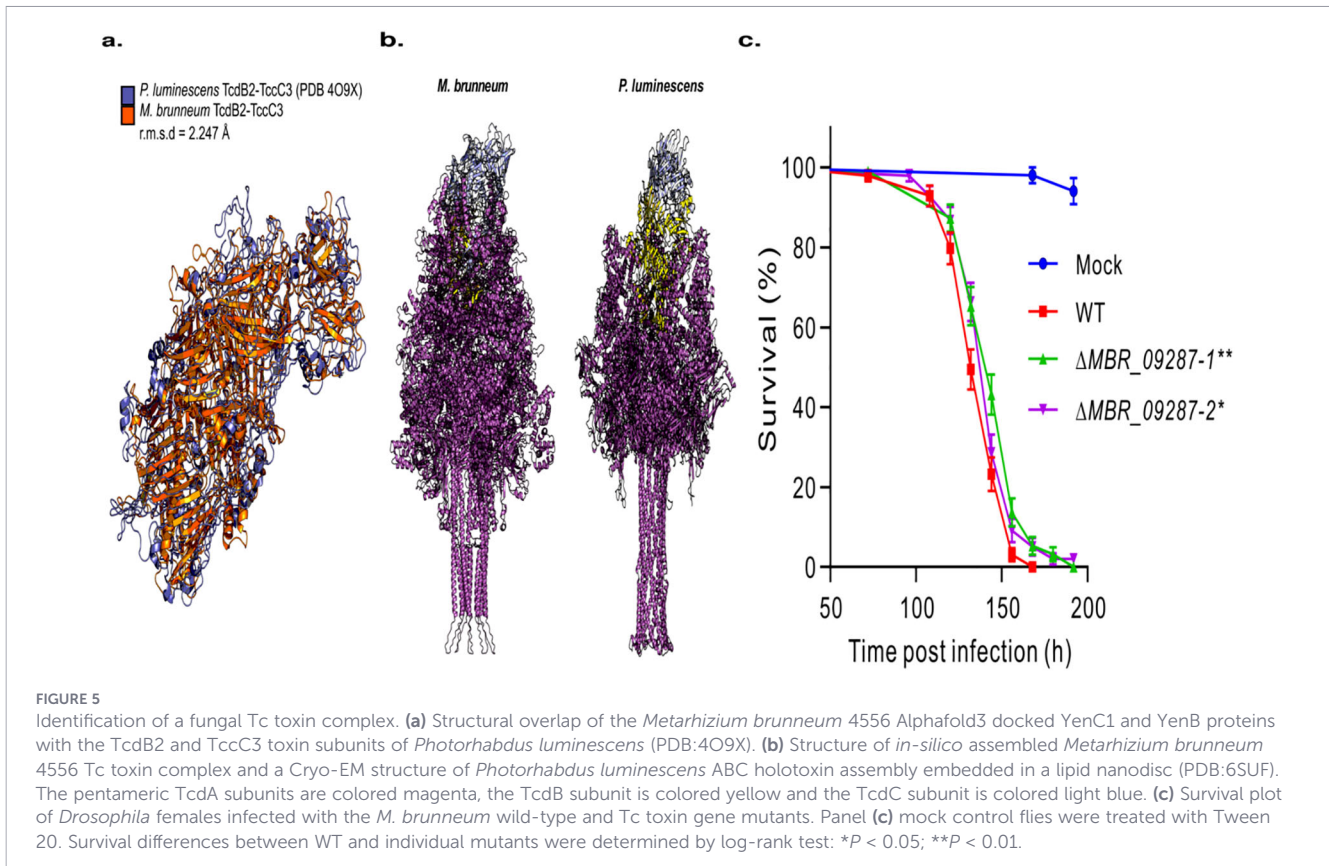
Interestingly, we observed a previously undescribed protein in *M. brunneum* with homology to serpin 42Dd of the fruit fly, *Drosophila melanogaster* (cluster 13612, ORF G6M90_00g004700 of *M. brunneum* 4556). The protein was predicted to have a classical serpin structural domain (IPR023796) spanning from amino acids 13–365 of the 365 amino acid sequence. A structural alignment revealed close structural homology between the two proteins and the predicted protein structure was found to have the classical conserved serpin folds and a reactive center loop (Figure 4a). This protein was not present in any of the other entomopathogenic fungi, but a blast search revealed it to be present in multiple species of *Trichophyton*, a mammalian pathogen, as well as being present in multiple species of *Aspergillus* and *Fusarium*, which contain species that are opportunistic mammalian pathogens. To determine the potential function of this gene in fungal virulence, a gene knockout line of this serpin, designated MBR_07878, was made. Topical infection of *D. melanogaster* revealed that the virulence of null mutants was significantly impaired (log-rank test, $P < 0.01$) compared to the wild-type strain (Figure 4b), thus implicating the protein in the infection process. By 132 hours post-infection (hpi), the wild-type strain exhibited significantly higher virulence than the mutant strains, with survival rates of approximately 50% and >60%, respectively. Furthermore, while the wild-type strain achieved 100% mortality by 168 hpi, approximately 10% of the host population remained viable following infection with the mutant strains.

Another hit in the *M. brunneum* species self-cluster showed a hit to subunit C1 of the Tc toxin complex of *Yersinia entomophaga* (cluster 10481, ORF G6M90_00g045720 of *M. brunneum* 4556). Tc toxins complexes are virulence factors of many infectious and entomopathogenic bacteria (Roderer and Raunser, 2019). The toxins are composed of three subunits that assemble to perforate a host membrane and translocate a toxic enzyme into the cell (Roderer and Raunser, 2019). To date, no such toxin complex has been described in a fungal species. The

Y. entomophaga Tc toxin subunit C1 ortholog found in *Metarhizium* mapped to chromosome 2 (*M. brunneum* strain 45556- regions 5382531–5386727). This 1258 amino acid protein was found to contain an Insecticide toxin TcdB middle/C-terminal region domain (IPR022044). Interestingly, a downstream neighboring (*M. brunneum* strain 45556- regions 5387051–5388929) gene was found to have homology to *Y. entomophaga* Tc toxin subunit B (ORF G6M90_00g045730 of *M. brunneum* 4556). Using AlphaFold3, we computationally docked the *M. brunneum* toxin subunits B and C (Figure 5a). The predicted multimer showed good structural homology to a crystal structure of TcdB2-TcdC3 subunits of the bacteria *Photorhabdus luminescens* [PDB:4O9X (Meusch et al., 2014)], with an RMSD of 2.247 Å (a measure of structural divergence). This metric represents the spatial average of the distances between equivalent atom pairs in the superimposed structures, where a lower RMSD value indicates higher structural conservation, and a value of less than 2.0 Å showing excellent similarity.

We then conducted a protein blast of the *M. brunneum* proteome using the sequence for *Yersinia Entomophaga* Tc toxin subunits A1 and A2. A hit was located for subunit A2 that mapped to chromosome 4 of *M. brunneum* 4556 (regions 1096054–1104757). This protein was found to be 2882 amino acids long and contained a Tc toxin complex TcA C-terminal TcB-binding domain (IPR040840) at positions 2440–2726, a Neuraminidase-like domain (IPR041079) at positions 1495–1655, and an ABC toxin N-terminal region domain (IPR046839) at positions 1344–1465 (ORF G6M90_00g069900 of *M. brunneum* 4556). A blast search of this protein confirmed its presence in multiple other *Metarhizium* species including; *M. robertsii*, *M. anisopliae*, *M. humberti*, *M. guizhouense*, and *M. album*. Previous studies in bacteria have demonstrated the A subunit alone is readily toxic but its effect is enhanced with the other two subunits docked (Waterfield et al., 2005). We used AlphaFold3 to predict structures of the various subunits of the *M. brunneum* Tc toxin complex and input these





smaller subunits into CombFold to generate the 16,253 amino acid toxin complex and compared it to the a Cryo-EM structure of *Photorhabdus luminescens* ABC holotoxin assembly embedded in a lipid nanodisc (Figure 5b). Both toxin complexes have the distinct common architecture of Tc toxins. An A2 subunit Tc toxin gene knockout *M. brunneum* strain (MBR_09287) was made, which showed an increased survival rate for the null mutants when compared to wild-type when after topical challenge of *D. melanogaster* (Figure 5c), thus implicating the protein in the infection process. By 132 hours post-infection (hpi), the wild-type strain exhibited significantly higher virulence than the mutant strains, with survival rates of approximately 50% and >60%, respectively. Furthermore, while the wild-type strain achieved 100% mortality by 168 hpi, >5% of the host population remained viable following infection with the mutant strains. The identification of these previously unreported functional genes highlights the utility of the two-by-two approach in discovering novel functional genes.

Discussion

We put forward a simple, novel, validated approach to determine the unique genes more accurately within a pre-selected group of fungi and is capable of distinguishing genes that are unique to a strain. We validated the method by selecting a group of fungi that have been subjected to decades of wet-lab experimentation, and confirmed that the approach could successfully detect genes and

gene clusters that have previously been demonstrated to be unique to certain species within the group. Using the two-by-two approach combined with *in silico* BGC prediction, we generated a heatmap depicting the biosynthetic potential of species in the families Cordycipitaceae, Clavicipitaceae, and Hypocreaceae within the order Hypocreales, providing a framework for future functional and comparative BGC analyses in these and related fungi. Two-by-two analyses of BGC clusters revealed that strains within a species harbored similar numbers of total BGCs, with a general trend for predicted BGCs to be shared between strains, irrespective of genome quality. For instance, the two *C. javanic* strains analyzed differed markedly in genome quality (173 scaffolds and 658 contigs for the IJ2G strain versus seven complete chromosomes for the Apopka 97 strain), yet the total numbers of predicted BGCs were similar (48 and 51, respectively). Of the 15 BGCs with matches in the MIBiG database, 11 were shared by both strains, while the remaining four were strain-specific—two unique to Apopka 97 and two unique to IJ2G. Moreover, the *B. bassiana* strains differed in protein BUSCO scores (91.3% versus 95.5%), a standard metric of genome completeness, yet all predicted BGCs with matches to known clusters in the MIBiG database were recovered in both strains. The ability of the fungal version of antiSMASH to predict BGCs appeared largely unaffected by genome quality. In *T. reesei*, where genome quality metrics were comparable between strains (seven complete chromosomes and BUSCO scores of 98.0 and 98.7), certain BGC predictions were strain-specific—for instance, a high-quality match to clavric acid was detected only in strain QM6a. These findings underscore the complexity of BGC evolution in *Hypocreales* fungi and highlight the value of two-by-two analyses

for robust identification of putative BGCs prior to resource-intensive functional studies, as they increase confidence that a BGC is truly present in multiple strains of a species. A further limitation is that differences in algorithm parameters—for example, running the tool with relaxed versus strict settings—can produce varying results, further complicating comparisons across studies. Integrating our current framework with multi-omic approaches (transcriptomics and metabolomics) would provide a more robust characterization of the infection process and facilitate the precise identification of alternative splice variants. However, a significant limitation is that key virulence genes may remain transcriptionally silent under *in vitro* conditions. Conversely, *in vivo* analyses are inherently complicated by the high background of host-derived transcripts and metabolites, which can interfere with the detection of fungal-specific signals.

We performed two-by-two orthogonal protein analyses using protein sets that were either deposited in GenBank, when available, or predicted *in silico* in this study. We showed that among orthologous proteins that uniquely self-clustered between two strains, we could identify proteins known to be species-specific (listed in Table 2), thereby validating the technique's ability to detect genuinely unique genes in species. Building on this validation, we compiled a list of orthologous proteins that uniquely self-cluster between two strains of the same species that have high sequence homology to previously characterized proteins known to have roles in pathogenesis, antibiotic production, or distinctive catabolic processes (Supplementary Table 5). Supplementary Table 1 contains a full list of self-clusters between two strains of each species analysed in this study. Additionally, we compiled a list of singlet proteins that did not form any orthologous clusters (Supplementary Table 2), as well as a list of proteins that formed clusters exclusively with proteins from the same strain (Supplementary Table 3). We also mapped genes unique to each species that were located within 5 kb of one another in the genome (Supplementary Table 4); these may represent novel gene clusters not detected by antiSMASH, as exemplified by the cordycepin BGC (Hong et al., 2024). These resources can be used for bioprospecting novel genes and gene clusters within these species, and may also serve to subtract common fungal genes or clusters when analyzing species outside our study group. Given the gene richness of fungal genomes, declining sequencing costs, and improving genome assembly algorithms, we anticipate that comparative genomics, including our two-by-two approach, will continue to enhance fungal gene discovery. This method complements RNA-Seq by simplifying differential gene expression analyses and identifying species-specific genes irrespective of their expression, thereby facilitating the discovery of novel cryptic genes.

Two *M. brunneum* self-clusters represent novel genes that have never been described before in fungi. We have identified the first functional myco-serpin that has significant homology to a known *D. melanogaster* immunomodulatory serpin. A major function of serpins in *Drosophila* is to confine immune reactions and prevent excessive activation (Levashina et al., 1999; Shan et al., 2023). Serpin 42Dd is a negative regulator of the Toll and melanization cascade pathways, blocking the serine proteases ModSP and Grass, with mutations in its gene leading to spontaneous melanization and constitutive activation of the Toll pathway (Shan et al., 2023). A recent study has shown that *Metarhizium robertsii* produces an immune modulating protein during infection that can also block the Toll pathway (Lu et al., 2024). The

Metarhizium serpin may play a role in immune evasion, as previous studies have shown that a serpin-like compound in the Microsporidia, *Nosema bombycis*, can suppress insect host hemolymph melanization (Bao et al., 2019). Parasitic insects have also been shown to use serpin-like proteins to dampen the melanization response against eggs they inject into other insects (Asgari et al., 2003), and poxviruses produce secreted serpin-like proteins to dampen inflammatory responses against mammals (Nash et al., 1997). Furthermore, a serpin-like protein from a symbiotic human gut bacteria has been shown to protect against exogenous proteolysis by gut proteins (Ivanov et al., 2006). Moreover, the serpin shares homology with other previously undescribed fungal serpins in *Fusarium*, *Aspergillus*, *Microsporium* and *Trichophyton*, with the two former species being opportunistic human pathogens and the latter two being parasitic mammalian pathogens. It would be of great interest to assess the impact on infection of knocking out the gene from a mammalian parasite, as mammals are also known to utilize serpins to regulate the immune system. From the gene knockout mutant, we cannot exclude the possibility that the serpin acts on a fungal serine protease, as proteins can exhibit pleiotropic effects; however, no detrimental impact on fungal growth was observed in the knockout strain. Prior to this study, the only other fungi to have been reported to have a serpin-like protein is the anaerobic fungus *Piromyces* sp., which contains a multi-domain protein that forms part of the cellulosome and contains a smaller serpin-like domain (Steenbakkers et al., 2008).

Finally, utilizing the two-by-two approach to analyze orthogonal proteins, we show that *M. brunneum* contains multiple subunits of a Tc toxin complex that exhibit significant structural homology to the well-characterized bacterial Tc toxin complexes (Gatsogiannis et al., 2013; Leidreiter et al., 2019; Roderer et al., 2020; Song et al., 2021; Belyy et al., 2022; Xu et al., 2022; Sitsel et al., 2024). Two of these subunits were found to be neighboring one another on the same chromosome, whilst the third subunit was found on a separate chromosome. The composition of this protein complex and its similarity to bacterial Tc toxins may indicate acquisition by fungi through horizontal gene transfer, a well-documented process (Bruto et al., 2014; Habig et al., 2024). The fact that the two subunits are located on chromosome 2, whilst the third is located on chromosome 4 may be attributable to a peculiar form of chromosomal evolution termed mesosynteny that has only been reported in Ascomycete fungi (Hane et al., 2011). The serpin and Tc complex are upregulated in *Metarhizium brunneum* during infection of *Myzus persicae*, as revealed by a recent transcriptomic study of differentially expressed genes [in Supplementary Figure 1 of reference (Reingold et al., 2024)]. Further work should assess the functional activity of the fungal toxin complex in purified form to assess its efficacy in comparison to bacterial Tc toxin complexes. Since the fungal proteins were predicted from spliced genes, investigating the potential enhancements in the fungal protein upon expression would be intriguing. Given the recent recognition of horizontal gene transfer from bacteria to fungi (Bruto et al., 2014), it would be interesting to investigate whether splicing of these acquired genes can generate enhanced protein variants, potentially representing a novel mechanism of protein evolution.

In this study, we have limited our search to truly unique genes between species by looking at strain self-clusters, however, the approach would also allow one to determine interesting genes shared by two very closely related species but potentially not the

rest of the species within a test group (e.g. by looking at clusters formed between two paired strains from 2 separate species). The species selected for analyses will likely affect the outputs, however, genes/gene clusters that are truly unique to only one species of fungi will likely always form strain self-clusters, regardless of what other fungi are added to the group. This was underscored by our finding that the serpin protein identified in *Metarhizium* is absent from other fungi within this group, yet present in more distantly related lineages. We envision that the approach can be used to elucidate unique genes in organisms from other kingdoms that have genome architecture and species diversity comparable to fungi.

Methods

Gene predictions and annotations

Biosynthetic gene cluster predictions were performed using the fungal version of the antiSMASH web server (v7.1) (Blin et al., 2023), setting the detection strictness to 'relaxed' and activating all extra features. The presence of signal peptides was predicted using SignalP 5.0 (Almagro Armenteros et al., 2019). Protein domain analyses was performed using InterProScan (v5.65-97.0). FunAnnotate (v1.8.15) was used to predict and annotate genes for the Hypocreales fungal genomes used in this study that lacked gene predictions in Genbank. BUSCO analyses were performed with BUSCO version 5.5.0 (Simão et al., 2015), using the hypocreales_odb10 lineage gene set. To determine the core genes shared across the Hypocreales species analyzed, comparison of orthologous gene clusters between the protein sets for each of the Hypocreales fungi were performed with a standalone version of OrthoVenn3 using the OrthoFinder algorithm (Sun et al., 2023). The AlphaFold3 server was used to predict protein structure and dock multimers (Abramson et al., 2024). Overlaps of protein structure were performed using the align tool in PyMOL Version 3.0 (The PyMOL Molecular Graphics System, Version 3.0). *In-silico* structural prediction of the *M. brunneum* 4556 was Tc toxin complex was performed using CombFold (Shor and Schneidman-Duhovny, 2024), using multiple AlphaFold3 docked dimer combinations as input.

Gene deletions and fly survival assays

Deletion of the serpin and Tc genes were conducted in *M. brunneum* by homologous replacement (Lu et al., 2024). In brief, the 5'- and 3'-flanking regions of each gene were amplified by PCR using different primer pairs (Supplementary Table 6), and the purified products were cloned into the binary vector pDHT-bar (conferring resistance to ammonium glufosinate) (Lu et al., 2024). Constructs were individually transformed into the *Agrobacterium tumefaciens* AGL1 strain, which was used to infect the wild-type spores of *M. brunneum*. The drug-resistance colonies were selected and verified by PCR analysis. At least, two independent mutant isolates were selected for survival assays against the females of *D. melanogaster* (3 days post eclosion). Two-week-old conidia of the wild-type and mutants were harvested from the potato dextrose agar, and suspended in 0.01% Tween-20. The spore suspensions were adjusted to 5×10^5 conidia/ml, and used for topical infection of

female flies (Lu et al., 2024). Flies treated with Tween 20 carrier were included as a mock control. There were more than 70 flies used for each treatment, and insect survivals were recorded every 12 hours. The difference between wild-type and mutant strains was examined by Kaplan-Meier analysis utilizing the log-rank test using GraphPad Prism version 10.1 (GraphPad software Inc., San Diego, CA).

Data and materials availability

Output files have been deposited in the following GitHub repository - <https://github.com/zacksaud/Cordyceps-javanica-Assembly-Project>. The following genomes and/or information on the genome assemblies were retrieved from NCBI's GenBank database; *Beauveria bassiana* ERL836 (accession number: GCA_010099065.1), *Beauveria bassiana* JEF-350 (accession number: GCA_021365345.1), *Cordyceps javanica* IJ2G (accession number: GCA_006981975.1), *Cordyceps javanica* Apopka 97 (accession number: GCA_051103115.1) *Cordyceps fumosorosea* ARSEF 2679 (accession number: GCA_001636725.1), *Cordyceps fumosorosea* grCorFumo1 (accession number: GCA_963580265.1), *Cordyceps militaris* ATCC 34164 (accession number: GCA_008080495.1), *Cordyceps militaris* CM01 (accession number: GCA_000225605.1), *Cordyceps cateniannulata* FRD 24 (accession number: GCA_028828415.1), *Cordyceps cateniannulata* MBC 234 (accession number: GCA_030411495.1), *Metarhizium brunneum* ARSEF 4556 (accession number: GCA_013426205.1), *Metarhizium brunneum* ARSEF 3297 (accession number: GCF_000814965.1), *Epichloe festucae* F11 (accession number: GCA_003814445.1), *Epichloe festucae* RoseCity (accession number: GCA_016859245.1), *Trichoderma reesei* CBS999.97 (accession number: GCA_016806755.1), *Trichoderma reesei* QM6a (accession number: GCA_002006585.1).

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/Supplementary Material.

Author contributions

ZS: Visualization, Data curation, Methodology, Formal analysis, Investigation, Writing – review & editing, Conceptualization, Software, Writing – original draft. YL: Visualization, Formal analysis, Writing – review & editing, Writing – original draft, Investigation. MW: Writing – review & editing, Investigation, Methodology, Writing – original draft. IB: Software, Writing – review & editing, Writing – original draft. BM: Software, Writing – review & editing, Writing – original draft. RS: Resources, Writing – original draft, Writing – review & editing. CW: Supervision, Formal analysis, Visualization, Methodology, Writing – review & editing, Conceptualization, Writing – original draft, Resources, Investigation. TB: Writing – original draft, Resources, Investigation, Visualization,

Funding acquisition, Formal analysis, Project administration, Data curation, Validation, Conceptualization, Writing – review & editing, Methodology, Supervision.

Funding

The author(s) declared that financial support was received for this work and/or its publication. ZS was supported by a Wellcome Trust grant (226615/Z/22/Z) awarded to RJS. YL was supported by grants from the National Natural Science Foundation of China (No. 32021001 and 32230087) awarded to CW.

Acknowledgments

We thank Professor Stephen C. Graham for guidance on in-silico protein structural modelling very large complexes.

Conflict of interest

The author(s) declared that this work was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

- Abramson, J., Adler, J., Dunger, J., Evans, R., Green, T., Pritzel, A., et al. (2024). Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature*. 630, 493–500. doi: 10.1038/s41586-024-07487-w
- Alberti, F., Foster, G. D., and Bailey, A. M. (2017). Natural products from filamentous fungi and production by heterologous expression. *Appl. Microbiol. Biotechnol.* 101, 493–500. doi: 10.1007/s00253-016-8034-2
- Alberts, A. W., Chen, J., Kuron, G., Hunt, V., Huff, J., Hoffman, C., et al. (1980). Mevinolin: a highly potent competitive inhibitor of hydroxymethylglutaryl-coenzyme A reductase and a cholesterol-lowering agent. *Proc. Natl. Acad. Sci. U.S.A.* 77, 3957–3961. doi: 10.1073/pnas.77.7.3957
- Almagro Armenteros, J. J., Tsirigos, K. D., Sønderby, C. K., Petersen, T. N., Winther, O., Brunak, S., et al. (2019). SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nat. Biotechnol.* 37, 420–423. doi: 10.1038/s41587-019-0036-z
- Anjum, T., and Iram, W. (2015). “Production of cyclosporine A by submerged fermentation,” in *Fungal metabolites*. Eds. J. M. Merillon and K. Ramawat (Springer, Cham). doi: 10.1007/978-3-319-19456-1_4-1
- Asgari, S., Zhang, G., Zareie, R., and Schmidt, O. (2003). A serine proteinase homolog venom protein from an endoparasitoid wasp inhibits melanization of the host hemolymph. *Insect Biochem. Mol. Biol.* 33, 1017–1024. doi: 10.1016/s0965-1748(03)00116-4
- Bagga, S., Hu, G., Screen, S. E., and St Leger, R. J. (2004). Reconstructing the diversification of subtilisins in the pathogenic fungus *Metarhizium anisopliae*. *Gene* 324, 159–169. doi: 10.1016/j.gene.2003.09.031
- Bao, J., Liu, L., An, Y., Ran, M., Ni, W., Chen, J., et al. (2019). *Nosema bombycis* suppresses host hemolymph melanization through secreted serpin 6 inhibiting the prophenoloxidase activation cascade. *J. Invertebr. Pathol.* 168, 107260. doi: 10.1016/j.jip.2019.107260
- Belyy, A., Lindemann, F., Roderer, D., Funk, J., Bardiaux, B., Protze, J., et al. (2022). Mechanism of threonine ADP-ribosylation of F-actin by a Tc toxin. *Nat. Commun.* 13, 4202. doi: 10.1038/s41467-022-31836-w
- Blin, K., Shaw, S., Augustijn, H. E., Reitz, Z. L., Biermann, F., Alanjary, M., et al. (2023). antiSMASH 7.0: new and improved predictions for detection, regulation,

Generative AI statement

The author(s) declared that generative AI was not used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fcimb.2026.1800461/full#supplementary-material>

- chemical structures and visualisation. *Nucleic Acids Res.* 51, W46–W50. doi: 10.1093/nar/gkad344
- Bonsch, B., Belt, V., Bartel, C., Duensing, N., Koziol, M., Lazarus, C. M., et al. (2016). Frequent, independent transfers of a catabolic gene from bacteria to contrasted filamentous eukaryotes. *Chem. Commun. (Camb)* 52, 6777–6780. doi: 10.1039/c6cc02130a
- Brakhage, A. A. (2013). Regulation of fungal secondary metabolism. *Nat. Rev. Microbiol.* 11, 21–32. doi: 10.1038/nrmicro2916
- Bruto, M., Prigent-Combaret, C., Luis, P., Moëgne-Loccoz, Y., and Muller, D. (2014). Frequent, independent transfers of a catabolic gene from bacteria to contrasted filamentous eukaryotes. *Proc. Biol. Sci.* 281, 20140848. doi: 10.1098/rspb.2014.0848
- Butt, T. M., Greenfield, B. P., Greig, C., Maffei, T. G., Taylor, J. W., Piasecka, J., et al. (2013). *Metarhizium anisopliae* pathogenesis of mosquito larvae: a verdict of accidental death. *PLoS One* 8, e81686. doi: 10.1371/journal.pone.0081686
- Cook, D., Donzelli, B. G. G., Creamer, R., Baucom, D. L., Gardner, D. R., Pan, J., et al. (2017). Swainsonine biosynthesis genes in diverse symbiotic and pathogenic fungi. *G3 (Bethesda)* 7, 1791–1797. doi: 10.1534/g3.117.041384
- Derntl, C., Guzmán-Chávez, F., Mello-de-Sousa, T. M., Busse, H. J., Driessen, A. J. M., Mach, R. L., et al. (2017). *In vivo* study of the sorbicillinoid gene cluster in trichoderma reesei. *Front. Microbiol.* 8, 2037. doi: 10.3389/fmicb.2017.02037
- Dornetshuber-Fleiss, R., Heffeter, P., Mohr, T., Hazemi, P., Kryeziu, K., Seger, C., et al. (2013). Destruxins: fungal-derived cyclohexadepsipeptides with multifaceted anticancer and antiangiogenic activities. *Biochem. Pharmacol.* 86, 361–377. doi: 10.1016/j.bcp.2013.05.022
- Feng, P., Shang, Y., Cen, K., and Wang, C. (2015). Fungal biosynthesis of the bibenzoquinone oosporein to evade insect immunity. *Proc. Natl. Acad. Sci. U.S.A.* 112, 11365–11370. doi: 10.1073/pnas.1503200112
- Fleming, A. (1929). On the antibacterial action of cultures of a penicillium, with special reference to their use in the isolation of *B. influenzae*. *Br. J. Exp. Pathol.* 10, 226. doi: 10.1093/climids/2.1.129
- Galagan, J. E., Henn, M. R., Ma, L. J., Cuomo, C. A., and Birren, B. (2005). Genomics of the fungal kingdom: insights into eukaryotic biology. *Genome Res.* 15, 1620–1631. doi: 10.1101/gr.3767105

- Gatsogiannis, C., Lang, A. E., Meusch, D., Pfaumann, V., Hofnagel, O., Benz, R., et al. (2013). A syringe-like injection mechanism in *Photorhabdus luminescens* toxins. *Nature* 495, 520–523. doi: 10.1038/nature11987
- Giuliano Garisto Donzelli, B., Gibson, D. M., and Krasnoff, S. B. (2015). Intracellular siderophore but not extracellular siderophore is required for full virulence in *Metarhizium robertsii*. *Fungal Genet. Biol.* 82, 56–68. doi: 10.1016/j.fgb.2015.06.008
- Godtfredsen, W., Jahnsen, S., Lorck, H., Roholt, K., and Tybring, L. (1962). Fusidic acid: a new antibiotic. *Nature* 193, 987. doi: 10.1038/193987a0
- Habig, M., Grasse, A. V., Müller, J., Stukenbrock, E. H., Leitner, H., and Cremer, S. (2024). Frequent horizontal chromosome transfer between asexual fungal insect pathogens. *Proc. Natl. Acad. Sci. U.S.A.* 121, e2316284121. doi: 10.1073/pnas.2316284121
- Hai, Y., Huang, A. M., and Tang, Y. (2019). Structure-guided function discovery of an NRPS-like glycine betaine reductase for choline biosynthesis in fungi. *Proc. Natl. Acad. Sci. U.S.A.* 116, 10348–10353. doi: 10.1073/pnas.1903282116
- Hane, J. K., Rouxel, T., Howlett, B. J., Kema, G. H., Goodwin, S. B., and Oliver, R. P. (2011). A novel mode of chromosomal evolution peculiar to filamentous Ascomycete fungi. *Genome Biol.* 12, R45. doi: 10.1186/gb-2011-12-5-r45
- Harvey, C. J. B., Tang, M., Schlecht, U., Horecka, J., Fischer, C. R., Lin, H. C., et al. (2018). HEx: A heterologous expression platform for the discovery of fungal natural products. *Sci. Adv.* 4, eaar5459. doi: 10.1126/sciadv.aar5459
- Hong, S., Shang, J., Sun, Y., and Wang, C. (2024). “Genetics and infection biology of the entomopathogenic fungi,” in *Fungal associations*, vol. 9. Eds. Y. P. Hsueh and M. Blackwell (Springer, Cham). doi: 10.1007/978-3-031-41648-4_13
- Ivanov, D., Emonet, C., Foata, F., Affolter, M., Delley, M., Fisseha, M., et al. (2006). A serpin from the gut bacterium *Bifidobacterium longum* inhibits eukaryotic elastase-like serine proteases. *J. Biol. Chem.* 281, 17246–17252. doi: 10.1074/jbc.M601678200
- Jayasuriya, H., Silverman, K. C., Zink, D. L., Jenkins, R. G., Sanchez, M., Pelaez, F., et al. (1998). Clavarinic acid: a triterpenoid inhibitor of farnesyl-protein transferase from *Clavariadelphus truncatus*. *J. Nat. Prod.* 61, 1568–1570. doi: 10.1021/np980200c
- Kaerberlein, T., Lewis, K., and Epstein, S. S. (2002). Isolating “uncultivable” microorganisms in pure culture in a simulated natural environment. *Science* 296, 1127–1129. doi: 10.1126/science.1070633
- Koulman, A., Lane, G. A., Christensen, M. J., Fraser, K., and Tapper, B. A. (2007). Peramine and other fungal alkaloids are exuded in the guttation fluid of endophyte-infected grasses. *Phytochemistry* 68, 355–360. doi: 10.1016/j.phytochem.2006.10.012
- Koulman, A., Lee, T. V., Fraser, K., Johnson, L., Arcus, V., Lott, J. S., et al. (2012). Identification of extracellular siderophores and a related peptide from the endophytic fungus *Epichloë festucae* in culture and endophyte-infected *Lolium perenne*. *Phytochemistry* 75, 128–139. doi: 10.1016/j.phytochem.2011.11.020
- Kramer, G. J., and Nodwell, J. R. (2017). Chromosome level assembly and secondary metabolite potential of the parasitic fungus *Cordyceps militaris*. *BMC Genomics* 18, 1–10. doi: 10.1186/s12864-017-4307-0
- Leidreiter, F., Roderer, D., Meusch, D., Gatsogiannis, C., Benz, R., and Raunser, S. (2019). Common architecture of Tc toxins from human and insect pathogenic bacteria. *Sci. Adv.* 5, eaax6497. doi: 10.1126/sciadv.aax6497
- Levashina, E. A., Langley, E., Green, C., Gubb, D., Ashburner, M., Hoffmann, J. A., et al. (1999). Constitutive activation of toll-mediated antifungal defense in serpin-deficient *Drosophila*. *Science* 285, 1917–1921. doi: 10.1126/science.285.5435.1917
- Liu, L., Wang, J., Rosenberger, D., Zhao, H., Lengyel, G., and Nadel, D. (2018). Fermented beverage and food storage in 13,000 y-old stone mortars at Raqefet Cave, Israel: Investigating Natufian ritual feasting. *J. Archaeological Science: Rep.* 21, 783–793. doi: 10.1016/j.jasrep.2018.08.008
- Lu, M., Wei, D., Shang, J., Li, S., Song, S., Luo, Y., et al. (2024). Suppression of *Drosophila* antifungal immunity by a parasite effector via blocking GGBP3 and GGBP-like 3, the dual receptors for β -glucans. *Cell Rep.* 43, 113642. doi: 10.1016/j.celrep.2023.113642
- Lv, H., Zhang, S., and Hao, X. (2020). Swainsonine protects H9c2 cells against lipopolysaccharide-induced apoptosis and inflammatory injury via down-regulating miR-429. *Cell. Cycle* 19, 207–217. doi: 10.1080/15384101.2019.1706902
- Meusch, D., Gatsogiannis, C., Efremov, R. G., Lang, A. E., Hofnagel, O., Vetter, I. R., et al. (2014). Mechanism of Tc toxin action revealed in molecular detail. *Nature* 508, 61–65. doi: 10.1038/nature13015
- Mohanta, T. K., and Bae, H. (2015). The diversity of fungal genome. *Biol. Proced. Online* 17, 8. doi: 10.1186/s12575-015-0020-z
- Nash, P., Lucas, A., and McFadden, G. (1997). SERP-1, a poxvirus-encoded serpin, is expressed as a secreted glycoprotein that inhibits the inflammatory response to myxoma virus infection. *Adv. Exp. Med. Biol.* 425, 195–205. doi: 10.1007/978-1-4615-5391-5_19
- Newton, G. G., and Abraham, E. P. (1955). Cephalosporin C, a new antibiotic containing sulphur and D-alpha-aminoacidic acid. *Nature* 175, 548. doi: 10.1038/175548a0
- Pannkuk, E. L., Risch, T. S., and Savary, B. J. (2015). Isolation and identification of an extracellular subtilisin-like serine protease secreted by the bat pathogen *Pseudogymnoascus destructans*. *PLoS One* 10, e0120508. doi: 10.1371/journal.pone.0120508
- Pokhrel, A., and Coleman, J. J. (2023). Inventory of the Secondary Metabolite Biosynthetic Potential of Members within the Terminal Clade of the *Fusarium solani* Species Complex. *J. Fungi (Basel)* 9, 799. doi: 10.3390/jof9080799
- Reingold, V., Faigenboim, A., Matveev, S., Haviv, S., Belasov, E., Vilcinskis, A., et al. (2024). Transcriptional reprogramming in the entomopathogenic fungus *Metarhizium brunneum* and its aphid host *Myzus persicae* during the switch between saprophytic and parasitic lifestyles. *BMC Genomics* 25, 917. doi: 10.1186/s12864-024-10824-y
- Roderer, D., Bröcker, F., Sitsel, O., Kaplonek, P., Leidreiter, F., Seeberger, P. H., et al. (2020). Glycan-dependent cell adhesion mechanism of Tc toxins. *Nat. Commun.* 11, 2694. doi: 10.1038/s41467-020-16536-7
- Roderer, D., and Raunser, S. (2019). Tc toxin complexes: assembly, membrane permeation, and protein translocation. *Annu. Rev. Microbiol.* 73, 247–265. doi: 10.1146/annurev-micro-102215-095531
- Rüegger, A., Kuhn, M., Lichti, H., Loosli, H. R., Huguenin, R., Quiquerez, C., et al. (1976). Cyclosporin A, ein immunsuppressiv wirksamer Peptidmetabolit aus *Trichoderma polysporum* (Link ex Pers.) Rifai [Cyclosporin A, a Peptide Metabolite from *Trichoderma polysporum* (Link ex Pers.) Rifai, with a remarkable immunosuppressive activity. *Helv. Chim. Acta* 59, 1075–1092. doi: 10.1002/hlca.19760590412
- Saud, Z., Kortsinoglou, A. M., Kouvelis, V. N., and Butt, T. M. (2021). Telomere length de novo assembly of all 7 chromosomes and mitogenome sequencing of the model entomopathogenic fungus, *Metarhizium brunneum*, by means of a novel assembly pipeline. *BMC Genomics* 22, 87. doi: 10.1186/s12864-021-07390-y
- Scherlach, K., and Hertweck, C. (2021). Mining and unearthing hidden biosynthetic potential. *Nat. Commun.* 12, 3864. doi: 10.1038/s41467-021-24133-5
- Schrödinger, (2021). *The PyMOL Molecular Graphics System, Version 3.0* (New York, NY: Schrödinger, LLC). Available online at: <https://www.frontiersin.org/journals/molecular-biosciences/articles/10.3389/fmolb.2022.867241/full>.
- Shan, T., Wang, Y., Bhattarai, K., and Jiang, H. (2023). An evolutionarily conserved serine protease network mediates melanization and Toll activation in *Drosophila*. *Sci. Adv.* 9, eadk2756. doi: 10.1126/sciadv.adk2756
- Shor, B., and Schneidman-Duhovny, D. (2024). CombFold: predicting structures of large protein assemblies using a combinatorial assembly algorithm and AlphaFold2. *Nat. Methods* 21(3), 477–487. doi: 10.1038/s41592-024-02174-0
- Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., and Zdobnov, E. M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31, 3210–3212. doi: 10.1093/bioinformatics/btv351
- Sitsel, O., Wang, Z., Janning, P., Kroczeck, L., Wagner, T., and Raunser, S. (2024). Yersinia entomophaga Tc toxin is released by T10SS-dependent lysis of specialized cell subpopulations. *Nat. Microbiol.* 9, 390–404. doi: 10.1038/s41564-023-01571-z
- Sofiyev, V., Navarro, G., and Trauner, D. (2008). Biomimetic synthesis of the shimalactones. *Org. Lett.* 10, 149–152. doi: 10.1021/ol702806v
- Song, N., Chen, L., Zhou, Z., Ren, X., Liu, B., Zhou, S., et al. (2021). Genome-wide dissection reveals diverse pathogenic roles of bacterial Tc toxins. *PLoS Pathog.* 17, e1009102. doi: 10.1371/journal.ppat.1009102
- Steenbakkens, P. J., Irving, J. A., Harhangi, H. R., Swinkels, W. J., Akhmanova, A., Dijkerman, R., et al. (2008). A serpin in the cellulosome of the anaerobic fungus *Piriformospora* sp. strain E2. *Mycol. Res.* 112, 999–1006. doi: 10.1016/j.mycres.2008.01.021
- St Leger, R. J., Frank, D. C., Roberts, D. W., and Staples, R. C. (1992). Molecular cloning and regulatory analysis of the cuticle-degrading-protease structural gene from the entomopathogenic fungus *Metarhizium anisopliae*. *Eur. J. Biochem.* 204, 991–1001. doi: 10.1111/j.1432-1033.1992.tb16721.x
- Sun, J., Lu, F., Luo, Y., Bie, L., Xu, L., and Wang, Y. (2023). OrthoVenn3: an integrated platform for exploring and visualizing orthologous data across genomes. *Nucleic Acids Res.* 51, W397–W403. doi: 10.1093/nar/gkad313
- Ter-Hovhannissyan, V., Lomsadze, A., Chernoff, Y. O., and Borodovsky, M. (2008). Gene prediction in novel fungal genomes using an ab initio algorithm with unsupervised training. *Genome Res.* 18, 1979–1990. doi: 10.1101/gr.081612.108
- Tuli, H. S., Sandhu, S. S., and Sharma, A. K. (2014). Pharmacological and therapeutic potential of *Cordyceps* with special reference to *Cordycepin*. *3 Biotech.* 4, 1–12. doi: 10.1007/s13205-013-0121-9
- Wang, R., Clarke, B. B., and Belanger, F. C. (2019). Transcriptome Analysis of Choke Stroma and Asymptomatic Inflorescence Tissues Reveals Changes in Gene Expression in Both *Epichloë festucae* and Its Host Plant *Festuca rubra* subsp. *rubra*. *Microorganisms* 7, 567. doi: 10.3390/microorganisms7110567
- Wang, Q., and Xu, L. (2012). Beauvericin, a bioactive compound produced by fungi: a short review. *Molecules* 17, 2367–2377. doi: 10.3390/molecules17032367
- Wang, L., Zhang, C., Zhang, J., Rao, Z., Xu, X., Mao, Z., et al. (2021). Epsilon-poly-L-lysine: recent advances in biomanufacturing and applications. *Front. Bioeng. Biotechnol.* 9, 748976. doi: 10.3389/fbioe.2021.748976

- Waterfield, N., Hares, M., Yang, G., Dowling, A., and French-Constant, R. (2005). Potentiation and cellular phenotypes of the insecticidal toxin complexes of *Photorhabdus* bacteria. *Cell. Microbiol.* 7, 373–382. doi: 10.1111/j.1462-5822.2004.00467.x
- Woo, S. L., Hermosa, R., Lorito, M., and Monte, E. (2023). Trichoderma: a multipurpose, plant-beneficial microorganism for eco-sustainable agriculture. *Nat. Rev. Microbiol.* 21, 312–326. doi: 10.1038/s41579-022-00819-5
- Xia, Y., Luo, F., Shang, Y., Chen, P., Lu, Y., and Wang, C. (2017). Fungal cordycepin biosynthesis is coupled with the production of the safeguard molecule pentostatin. *Cell Chem. Biol.* 24, 1479–1489.e4. doi: 10.1016/j.chembiol.2017.09.001
- Xu, Y., Viswanatha, R., Sitsel, O., Roderer, D., Zhao, H., Ashwood, C., et al. (2022). CRISPR screens in *Drosophila* cells identify Vsg as a Tc toxin receptor. *Nature* 610, 349–355. doi: 10.1038/s41586-022-05250-7
- Yin, Y., Chen, B., Song, S., Li, B., Yang, X., and Wang, C. (2020). Production of diverse beauveriolide analogs in closely related fungi: a rare case of fungal chemodiversity. *mSphere* 5, e00667-20. doi: 10.1128/mSphere.00667-20
- Zeng, G., Zhang, P., Zhang, Q., Zhao, H., Li, Z., Zhang, X., et al. (2018). Duplication of a Pks gene cluster and subsequent functional diversification facilitate environmental adaptation in *Metarhizium* species. *PLoS Genet.* 14, e1007472. doi: 10.1371/journal.pgen.1007472
- Zhang, J., Zhang, P., Zeng, G., Wu, G., Qi, L., Chen, G., et al. (2021). Transcriptional differences guided discovery and genetic identification of coprogen and dimeric acid siderophores in *metarhizium robertsii*. *Front. Microbiol.* 12, 783609. doi: 10.3389/fmicb.2021.783609
- Zimmermann, G. (2007). Review on safety of the entomopathogenic fungi *Beauveria bassiana* and *Beauveria brongniartii*. *Biocontrol Sci. Technol.* 17, 553–596. doi: 10.1080/09583150701309006