

RESEARCH ARTICLE

Open Access



# The genome of the truffle-parasite *Tolypocladium ophioglossoides* and the evolution of antifungal peptaibiotics

C. Alisha Quandt<sup>1,3\*</sup>, Kathryn E. Bushley<sup>2</sup> and Joseph W. Spatafora<sup>1</sup>

## Abstract

**Background:** Two major mycoparasitic lineages, the family Hypocreaceae and the genus *Tolypocladium*, exist within the fungal order, Hypocreales. Peptaibiotics are a group of secondary metabolites almost exclusively described from *Trichoderma* species of Hypocreaceae. Peptaibiotics are produced by nonribosomal peptide synthetases (NRPSs) and have antibiotic and antifungal activities. *Tolypocladium* species are mainly truffle parasites, but a few species are insect pathogens.

**Results:** The draft genome sequence of the truffle parasite *Tolypocladium ophioglossoides* was generated and numerous secondary metabolite clusters were discovered, many of which have no known putative product. However, three large peptaibiotic gene clusters were identified using phylogenetic analyses. Peptaibiotic genes are absent from the predominantly plant and insect pathogenic lineages of Hypocreales, and are therefore exclusive to the largely mycoparasitic lineages. Using NRPS adenylation domain phylogenies and reconciliation of the domain tree with the organismal phylogeny, it is demonstrated that the distribution of these domains is likely not the product of horizontal gene transfer between mycoparasitic lineages, but represents independent losses in insect pathogenic lineages. Peptaibiotic genes are less conserved between species of *Tolypocladium* and are the product of complex patterns of lineage sorting and module duplication. In contrast, these genes are more conserved within the genus *Trichoderma* and consistent with diversification through speciation.

**Conclusions:** Peptaibiotic NRPS genes are restricted to mycoparasitic lineages of Hypocreales, based on current sampling. Phylogenomics and comparative genomics can provide insights into the evolution of secondary metabolite genes, their distribution across a broader range of taxa, and their possible function related to host specificity.

**Keywords:** Secondary metabolism, Hypocreales, Mycoparasites, Lineage sorting

## Background

Hypocreales is home to a wide array of ecologically diverse fungi. Some are devastating plant pathogens (e.g., *Fusarium* spp.), while others form numerous lineages of both insect pathogens (e.g., *Cordyceps*) and mycoparasites (e.g., *Trichoderma*) [1]. At the divergence of the four most derived families (Clavicipitaceae, Cordycipitaceae, Hypocreaceae, and Ophiocordycipitaceae) of Hypocreales,

there was a major shift away from plant-based nutrition to either insect pathogenesis or fungal parasitism, i.e., mycoparasites [2]. Two major lineages of mycoparasites are found within the order, although other mycoparasites exist (e.g., several species of *Polycephalomyces*; [3, 4]). The first and larger of these two lineages is the family Hypocreaceae, most notable for mycoparasitic *Trichoderma* spp. used in biological control of plant pathogenic fungi, and *Trichoderma reesei* E.G. Simmons [5], the industrial workhorse for cellulase production [6, 7]. The second major lineage of mycoparasites, the genus *Tolypocladium*, is nested within the insect pathogenic family, Ophiocordycipitaceae [1, 8]. Most species of *Tolypocladium*

\* Correspondence: alishaq@umich.edu

<sup>1</sup>Department of Botany and Plant Pathology, Oregon State University, Corvallis, OR, USA

<sup>3</sup>Present address: Department of Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, MI, USA

Full list of author information is available at the end of the article

parasitize the truffles of *Elaphomyces* [Eurotiales, Ascomycota], ectomycorrhizal fungi closely related to *Aspergillus* and *Penicillium* [9, 10]. *Tolypocladium ophioglossoides* (Ehrh. ex J.F. Gmel.) Quandt, Kepler & Spatafora is a commonly collected truffle parasite with a broad geographic distribution throughout many parts of the Northern Hemisphere [11, 12]. There are, however, a few *Tolypocladium* species that attack insects and rotifers, and based on current multigene phylogenies some of these are inferred to be reversals to insect pathogenesis [1, 8, 13]. One of these is a beetle pathogen, *T. inflatum*, which was the first source of the immunosuppressant drug, cyclosporin A [14]. Evidence from multigene studies has also shown a close phylogenetic relationship between *T. ophioglossoides* and *T. inflatum* [1].

Secondary metabolism is defined as the synthesis of often bioactive, small molecules that are not essential to the growth of an organism. Genes related to production of secondary metabolites are often clustered together in close proximity within a genome and coregulated [15]. A wide variety of secondary metabolites including the ergot alkaloids, fumonisins, and destruxins, is produced by species of Hypocreales [16–18]. Many of these metabolites are produced by nonribosomal peptide synthetases (NRPSs), which are often large, multi-modular proteins that produce short peptides frequently incorporating non-standard amino acids. NRPS modules are composed of three primary functional domains including adenylation (A), thiolation (T), and condensation (C) domains [19]. Due to their high level of amino acid and nucleotide conservation, the A-domains are frequently used to reconstruct the evolutionary histories of these genes [20, 21]. Polyketide synthases (PKSs) are another class of secondary metabolite producing enzymes that are common in fungi and are also modular in nature. They are related to fatty acid synthases [22], and assemble small bioactive molecules based on acetyl-CoA or malonyl-CoA subunits [23]. The other major classes of secondary metabolite-producing enzymes are terpene synthases and dimethylallyltryptophan (DMAT) synthases, both of which have been reported from hypocrealean taxa. Fungal secondary metabolites clusters often include genes required for regulation of expression of the gene cluster and decoration, epimerization, and transport of the mature secondary metabolite [24, 25].

Peptaibols, or peptaibiotics, are antibiotic secondary metabolites products produced by very large NRPS enzymes (up to 21,000 amino acids in length). Their name is a derivative of their structure as they are Peptides containing the uncommon non-proteinogenic amino acid,  $\alpha$ -amino isobutyric acid (AIB), and a C-terminal amino ethanol [26]. The presence of AIB residues promotes helix formation, and several of these helices form multimeric units that in turn form voltage gated ion

channels capable of inserting into cell membranes where they disrupt membrane potential causing leakiness [27, 28]. Peptaibols are produced by *Trichoderma* spp. and other members of Hypocreaceae (The Peptaibol Database [29]), leading to the proposition that they may play a role in mycoparasitism. There is at least one empirical study to support this in *Trichoderma* [30]. Further studies have found that peptaibols function, along with cell wall degrading enzymes, to synergistically inhibit new cell wall synthesis in fungal prey of *Tr. harzianum* [31–33]. Wiest et al. [34] identified and characterized the first peptaibol synthetase NRPS modular structure (a product of the *tex1* gene) from *Tr. virens* along with the 18 residue peptaibol product. Since that time, several other peptaibol NRPS genes have been identified in *Tr. virens* and other species [21, 35, 36].

Efraeptins are another class of peptaibiotics described almost exclusively from *Tolypocladium* spp. that have antifungal and insecticidal properties [37–39]. They differ from orthodox peptaibols by the presence of a mitochondrial ATPase inhibiting C-terminal “blocking group” N-peptido-1-isobutyl-2[1-pyrrole-(1-2- $\alpha$ )-pyrimidinium,2,3,4,6,7,8-hexahydro]-ethylamine [40]. Eight of these, named efraeptins A, and C-I, have been isolated from *T. inflatum* [38, 41].

Numerous genomes from species of the mycoparasitic genus *Trichoderma* (Hypocreaceae) have been sequenced [42, 43], and more recently the genomes of several insect pathogens in Hypocreales have been completed (e.g., *Cordyceps militaris*, *Beauveria bassiana*, *Metarhizium* spp., and *Ophiocordyceps sinensis*) [44–47], including *T. inflatum*, the beetle pathogenic congener of *T. ophioglossoides* [48]. The genomes of all of these species are rich in secondary metabolite genes and clusters, ranging from 23 to 51 secondary metabolite gene clusters per genome. Comparisons of gene content and expression are beginning to shed light on mechanisms underlying host specificity and the evolution of primary and secondary metabolism. In this study, the draft genome of the truffle parasite *T. ophioglossoides* was generated to compare the gene content and secondary metabolite content of this truffle parasite to those of closely related insect pathogens and more distantly related mycoparasites. The secondary metabolite potential of *T. ophioglossoides* is characterized with a focus on understanding the evolution of gene clusters encoding for peptaibiotics.

## Results and discussion

### Genome assembly and structure

The draft genome assembly of the *T. ophioglossoides* CBS 100239 is approximately 31.2 megabases (Mb) and is assembled on 172 scaffolds. The assembly, with an n50 of 668,222 base pairs (bp), captures the majority of

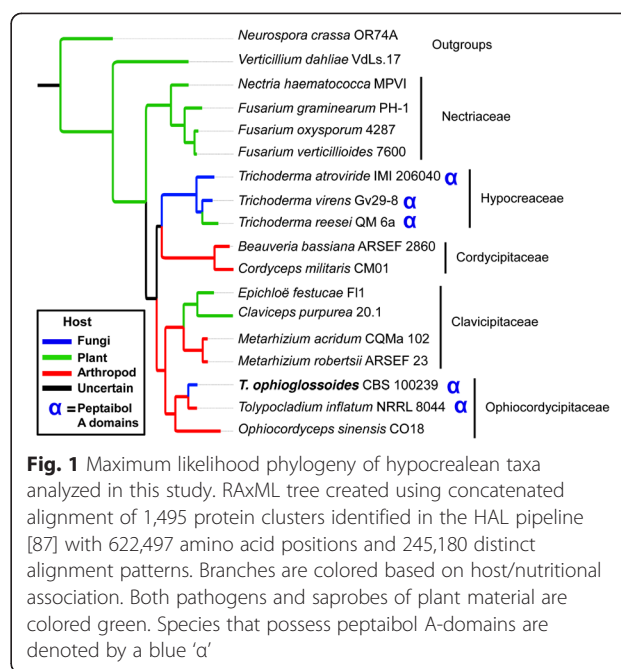
gene space and synteny [Table 1], in spite of the exclusive use of short read technology. In addition, the Core Eukaryotic Genes Mapping Approach (CEGMA) [49] identified 239 complete and 242 complete and partial core eukaryotic genes, estimating the assembly completeness at 96.4 % or 97.6 %. We predict 10,134 protein-coding genes resulting in 10,307 protein models, of which 9,476 have support from RNA, making the size and number of protein models similar to that of the beetle pathogen, *T. inflatum*, which is 30.3 Mb and has 9,998 protein models. They both also share high GC contents, 57.3 % and 58 %, for *T. ophioglossoides* and *T. inflatum*, respectively. These species are very closely related, which is reflected in the genome scale phylogeny (Fig. 1) and the large regions of shared synteny (Fig. 2). There are, however, two large scale inversions and some small rearrangements with disagreement, suggesting these species have diverged despite the relatively short branch lengths separating them in the species phylogeny (Fig. 1). The large number of small non-syntenic points in the mummerplot alignment could be due to the reliance on short read technology and the presence of short contigs in both assemblies.

**An overview of secondary metabolites in *T. ophioglossoides***

The *T. ophioglossoides* and *T. inflatum* genomes harbor 45 and 55 core secondary metabolite genes – NRPS, PKS, terpene synthase, and DMATs – spread across 38 and 38 secondary metabolite gene clusters, respectively. Similar to the contrast observed between closely related *Metarhizium* species ([46]), *T. ophioglossoides* and *T. inflatum* differ in the types of core secondary metabolite genes they possess with 21 shared between the two species, 34 unique to *T. inflatum*, and 24 unique to *T. ophioglossoides* (Additional file 1). Notably, *T. ophioglossoides* does not contain the NRPS gene, or any of the other genes in the *simA* cluster responsible for the production of cyclosporin A in *T. inflatum* [48]. However, *T. ophioglossoides* does share a Hypocreales-conserved core set of genes that flank the *simA* region with *T. inflatum*. Cyclosporin

**Table 1** Genome statistics for *Tolypocladium ophioglossoides* compared to *T. inflatum* (Bushley et al. 2013)

	<i>T. ophioglossoides</i>	<i>T. inflatum</i>
Size (Mb)	31.2	30.3
# Scaffolds	172	101
N50	668,222	1,509,745
Longest Scaff.	2,309,933	3,562,345
%GC	57.3	58
Protein-coding genes	10,134	9,998
# SM clusters	38	38

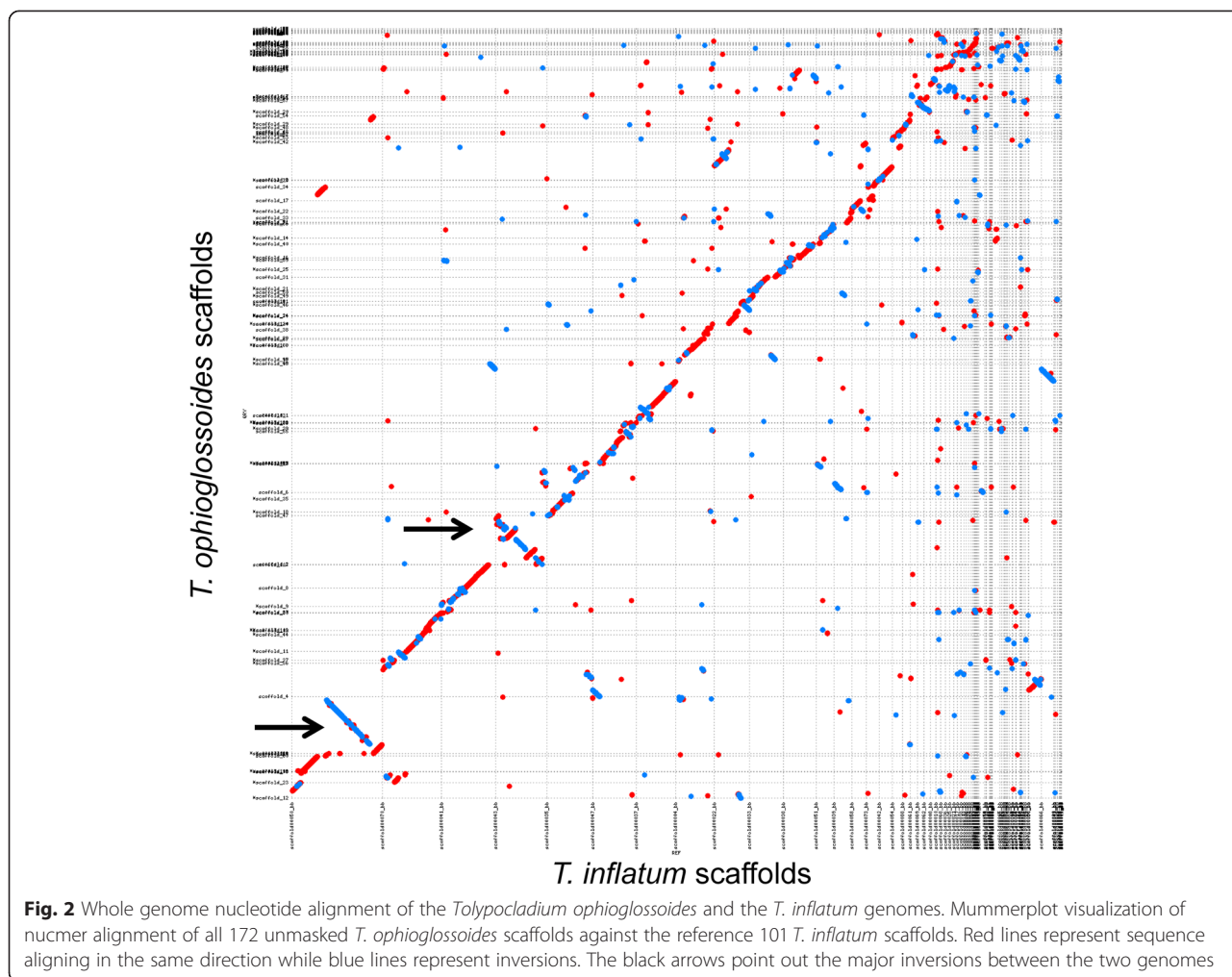


**Fig. 1** Maximum likelihood phylogeny of hypocrealean taxa analyzed in this study. RAXML tree created using concatenated alignment of 1,495 protein clusters identified in the HAL pipeline [87] with 622,497 amino acid positions and 245,180 distinct alignment patterns. Branches are colored based on host/nutritional association. Both pathogens and saprobes of plant material are colored green. Species that possess peptaibol A-domains are denoted by a blue ‘α’

production has been reported from several species of *Tolypocladium* [50, 51], however, none of the truffle parasites, including *T. ophioglossoides*, have been demonstrated to produce the compound. Whether possession of the *simA* gene is a derived character state, or whether there has been a single or multiple losses within the genus, remains unknown and requires further sampling.

Secondary metabolites core genes predicted from the *T. ophioglossoides* genome include 15 PKSs, two PKS-like genes, 15 NRPSs, six NRPS-like genes, three hybrid NRPS-PKS genes, and four terpenes (Additional file 1). No DMAT synthases were identified in the *T. ophioglossoides* genome. Based on A-domain homology, two putative siderophore synthetases (one intracellular [TOPH\_02853] and one extracellular [TOPH\_02629]) were among the predicted NRPSs (Additional files 2 and 1). This is in contrast to *T. inflatum*, which possesses three putative siderophore synthetases (two extracellular and one intracellular) [48]. The entire Pseurotin-A precursor synthetase hybrid NRPS-PKS cluster (TOPH\_07102) was identified in the *T. ophioglossoides* genome. Pseurotin-A, an antifungal compound described from several *Aspergillus* spp. [Eurotiales, Ascomycota], was also recently identified in the genome of *M. robertsii* [52]. The disjunct distribution of this secondary metabolite cluster raises several questions about the evolutionary mechanisms (e.g., horizontal gene transfer vs. complex patterns of gene loss) that may have led to this distribution.

*T. ophioglossoides* possesses the destruxins synthetase NRPS gene (TOPH\_08872) (Additional file 1). Destruxins are known for their insecticidal properties in *Metarhizium* spp. [53], and the entire destruxins synthetase



**Fig. 2** Whole genome nucleotide alignment of the *Tolypocladium ophioglossoides* and the *T. inflatum* genomes. Mummerplot visualization of nucmer alignment of all 172 unmasked *T. ophioglossoides* scaffolds against the reference 101 *T. inflatum* scaffolds. Red lines represent sequence aligning in the same direction while blue lines represent inversions. The black arrows point out the major inversions between the two genomes

cluster was characterized by Wang et al. (2012) in *M. robertsii*. Homologs of the other essential genes in the destruxins cluster are present in the *T. ophioglossoides* TOPH\_08872 cluster, except for *dtxS4* (Additional file 3), an aspartic decarboxylase responsible for producing  $\beta$ -alanine, one of the amino acids incorporated into destruxins. There are inversions in this cluster between *M. robertsii* and *T. ophioglossoides* as well, including the *dtxS2* aldo-keto reductase homolog (TOPH\_08871) and an ABC transporter (TOPH\_08869), which was not found to be essential in destruxins production in *M. robertsii* [54]. For these reasons, it is not likely that *T. ophioglossoides* produces destruxins, but possibly produces another group of related compounds. The sequenced strain of *Tr. virens* shares a homolog of the destruxins NRPS gene (Tv62540) (Additional file 3), but destruxins have not been reported to be produced by that species either [18].

To date, only two secondary metabolites have been reported to be produced by *T. ophioglossoides*: ophiocordin (also reported as balanol) and ophioretin [55–57].

Ophioretin is structurally similar to equisetin, an antibiotic with inhibitory activity of HIV-1 integrase, and both are produced by NRPS-PKS hybrid genes. Based on phylogenetic analysis of A-domains, cluster synteny with the equisetin cluster [58], and sequence homology, this study identifies the putative ophioretin synthetase cluster around the hybrid NRPS-PKS, TOPH\_07403 (Additional file 1). Further studies involving transformations and chemical verification and characterization of this cluster will be necessary to confirm this genotype-chemotype linkage. Ophiocordin is a polyketide and no putative gene or gene cluster related to its production was identified here. Except for the two peptaibiotic clusters discussed below, the remaining 33 secondary metabolite gene clusters are not yet associated with a specific gene product.

**Peptaibiotics of *Tolypocladium***

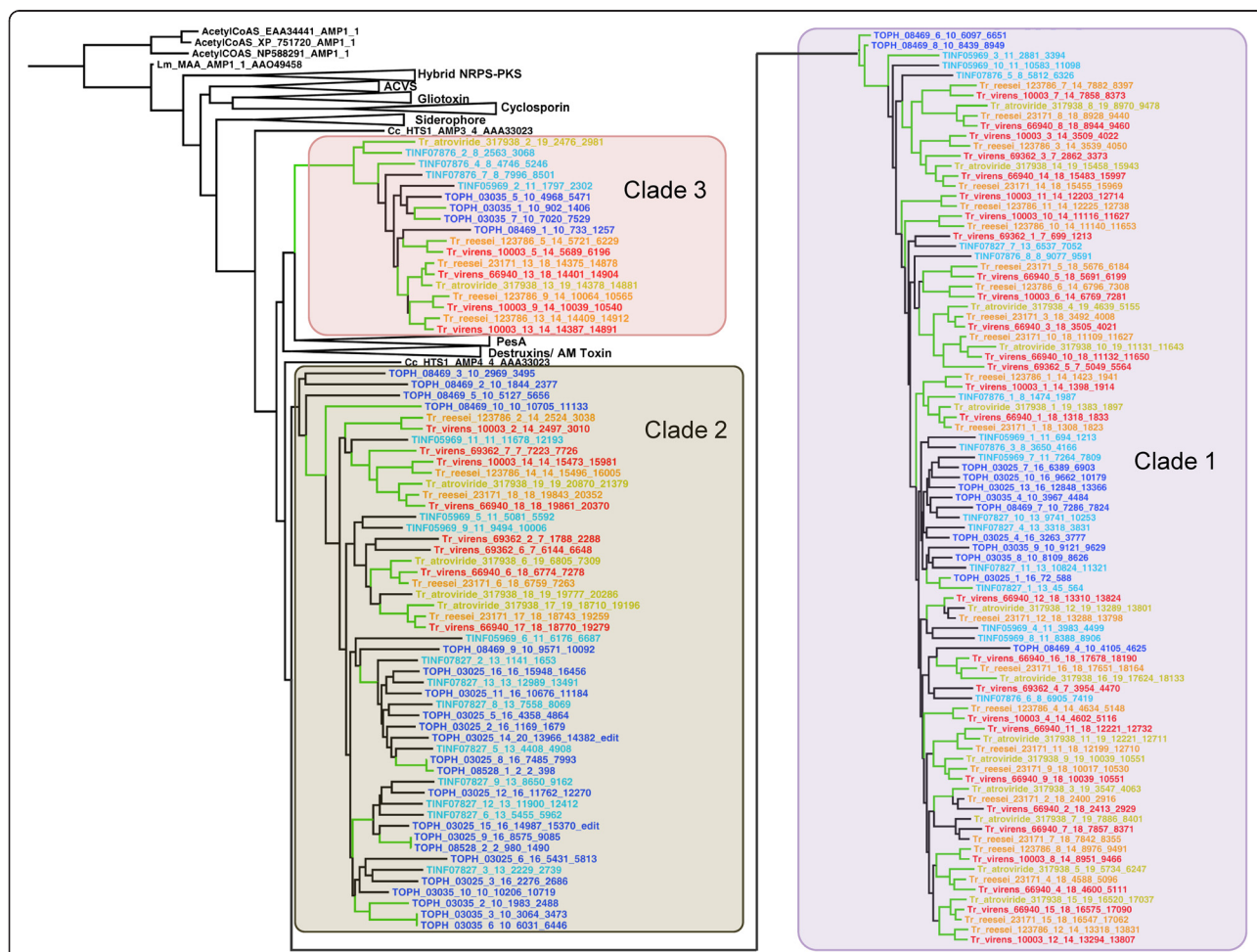
Among the wide assortment of secondary metabolite genes and gene clusters, the draft genome of *T. ophioglossoides* possesses four peptaibiotic NRPS genes located



within three gene clusters. Phylogenetic analyses of fungal A-domains from a variety of NRPS genes known to produce specific products and from whole genome mining of the hypocrealean species in Fig. 1, revealed that *Trichoderma* peptaibol A-domains group into three clades, one of which (Clade 3) is well supported in all analyses (Fig. 3, Additional file 2), the other two clades have strong support with the exception of a few A-domains from one of the *T. ophioglossoides* peptaibiotic genes (TOPH\_08469) (see below). All A-domains from the four peptaibiotic genes in *T. ophioglossoides* fall within these three clades (Fig. 3), and representation in the three peptaibol clades is exclusive to A-domains from *Trichoderma* and *Tolypocladium*. *T. inflatum* also possesses three peptaibiotic NRPS genes. Strikingly, this limits the presence of peptaibiotic A-domains and genes to the two sampled mycoparasitic lineages in Hypocreales.

There are differences between species and gene membership within A-domain clades, however. For instance,

Clade 1 is enriched in A-domains from the *Trichoderma* peptaibols (62 *Trichoderma* A-domains vs. 27 *Tolypocladium* A-domains), while Clade 2 contains more A-domains from *Tolypocladium* spp. (17 *Trichoderma* A-domains vs. 33 *Tolypocladium* A-domains). Importantly Clade 1 contains A-domains that encode for incorporation of AIB, as well as other A-domains that encode for incorporation of isovaline, leucine, isoleucine, alanine, glycine, valine, and serine ([21, 35]). Clade 2 is known to include A-domains that encode for valine, glutamine, asparagine, leucine, and isoleucine ([21, 35]). Clade 3, which is known only to incorporate a single amino acid, proline ([21, 35]), has a relatively equal distribution of both *Trichoderma* (10) and *Tolypocladium* (8) A-domains. Prolines are proposed to play an important structural role in peptaibols by creating a kink in the peptaibol chain, and Clade 3 occupies a long branch within the tree, suggesting it is highly diverged from the other A-domains (Additional file 2).



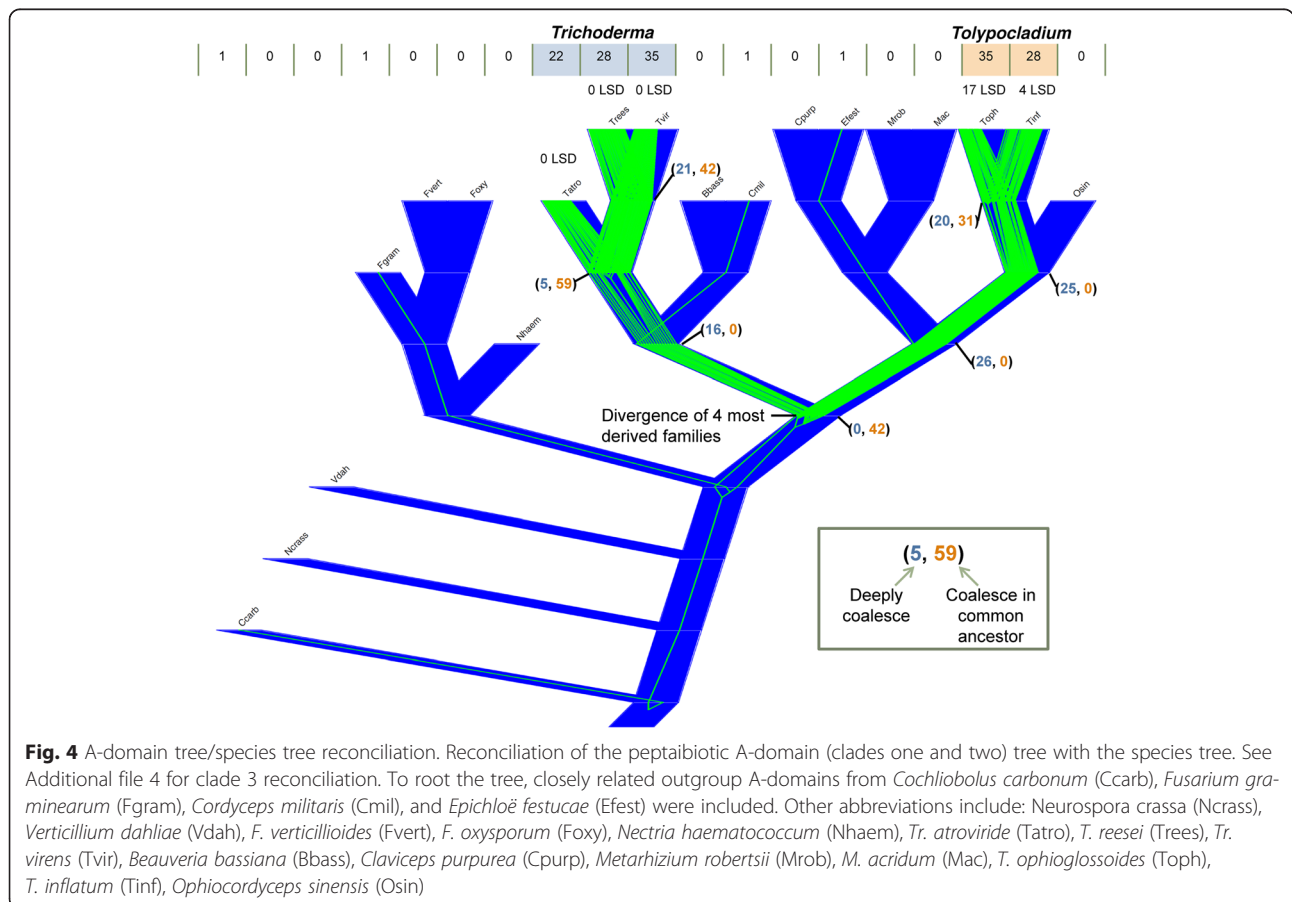
**Fig. 3** Phylogeny of the three peptaibol A-domain clades. Peptaibol focused, A-domain phylogeny created using RAxML, showing the 3 major clades. Green branches represent those supported by  $\geq 70\%$  bootstrap support. *Tolypocladium* spp. A-domains are colored in blue and teal, and *Trichoderma* spp. A-domains are colored in yellow, red, or orange

Peptaibols were reported from *T. geodes* [59] based on chemical isolation, but all previous reports of the compounds were identified from fungi in Hypocreaceae [29] (or Boletaceae [Basidiomycota] which are likely produced by hypocreaceous mycoparasites of Boletaceae fruiting bodies [60, 61]). A gene cluster responsible for the production of these peptaibols has not been identified, and to date no genomic sequence data have been produced for *T. geodes* from which to predict which genes or clusters may be responsible for its production. Efrapeptins, which are peptaibiotics originally described from *T. inflatum*, have been reported from several species of *Tolypocladium* [38], and it remains unknown whether the products produced by these clusters in *T. ophioglossoides* are of the efrapeptin class, or more traditional class of peptaibols. Regardless, the phylogenetic diversity of peptaibiotic NRPSs, as revealed by phylogenomic analyses of A-domains, supports a greater chemical diversity of peptaibiotics than currently known from chemical analyses.

To interpret the evolutionary history of the peptaibol A-domains within Hypocreales, the peptaibol A-domain clade tree was reconciled with the species tree (Fig. 4). Because Clade 3 is inferred to have an independent

origin from Clades 1 and 2 within the A-domain phylogeny, it was reconciled with the species tree separately (Additional file 4). We have included closely related A-domains from a complete A-domain phylogeny of Hypocreales that are not a part of the three peptaibol A-domain clades in order to root the peptaibol A-domain clades. The deep coalescence of peptaibol A-domains from all three clades at the common ancestor of *Trichoderma* and *Tolypocladium* (Fig. 4) suggests that the presence of peptaibiotics in *Tolypocladium* is an ancient attribute of *Tolypocladium* genomes and not a product of a more recent horizontal transfer. Analyzing Clades 1 and 2, only 31 of the A-domains in *T. ophioglossoides* and *T. inflatum* coalesce at their most recent common ancestor, whereas 20 coalesce more deeply at the divergence of the four most derived families in Hypocreales (Fig. 4).

There exist at least two possible explanations to these findings in *Tolypocladium* spp. First, the divergence of the A-domains could be significant enough so as to distort the evolutionary history as represented in the phylogenetic tree, especially since many of the branches of the *Tolypocladium* A-domain tree do not have strong bootstrap support. This explanation is not supported by the



**Fig. 4** A-domain tree/species tree reconciliation. Reconciliation of the peptaibiotic A-domain (clades one and two) tree with the species tree. See Additional file 4 for clade 3 reconciliation. To root the tree, closely related outgroup A-domains from *Cochliobolus carbonum* (Ccarb), *Fusarium graminearum* (Fgram), *Cordyceps militaris* (Cmil), and *Epichloë festucae* (Efest) were included. Other abbreviations include: *Neurospora crassa* (Ncrass), *Verticillium dahliae* (Vdah), *F. verticillioides* (Fvert), *F. oxysporum* (Foxy), *Nectria haematococcum* (Nhaem), *Tr. atroviride* (Tatro), *T. reesei* (Trees), *Tr. virens* (Tvir), *Beauveria bassiana* (Bbass), *Claviceps purpurea* (Cpupp), *Metarhizium robertsii* (Mrob), *M. acridum* (Mac), *T. ophioglossoides* (Toph), *T. inflatum* (Tinf), *Ophiocordyceps sinensis* (Osin)

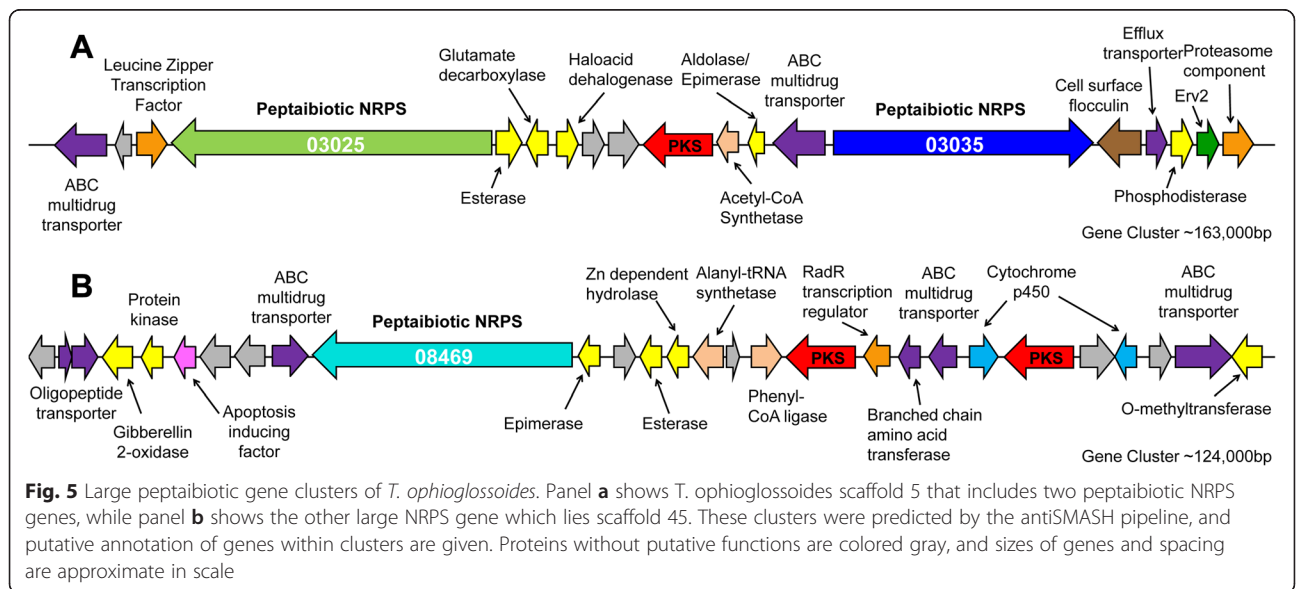
data, because based on the species phylogeny amino acid divergence levels between *Tolypocladium* species is less than that of *Trichoderma* species (Fig. 1). Second, incomplete lineage sorting could lead to this pattern of coalescence, suggesting that the ancestor to the genus possessed a multitude of these A-domains within one or several peptaibiotic genes that have undergone a complex history of ancient gains (duplications) and losses.

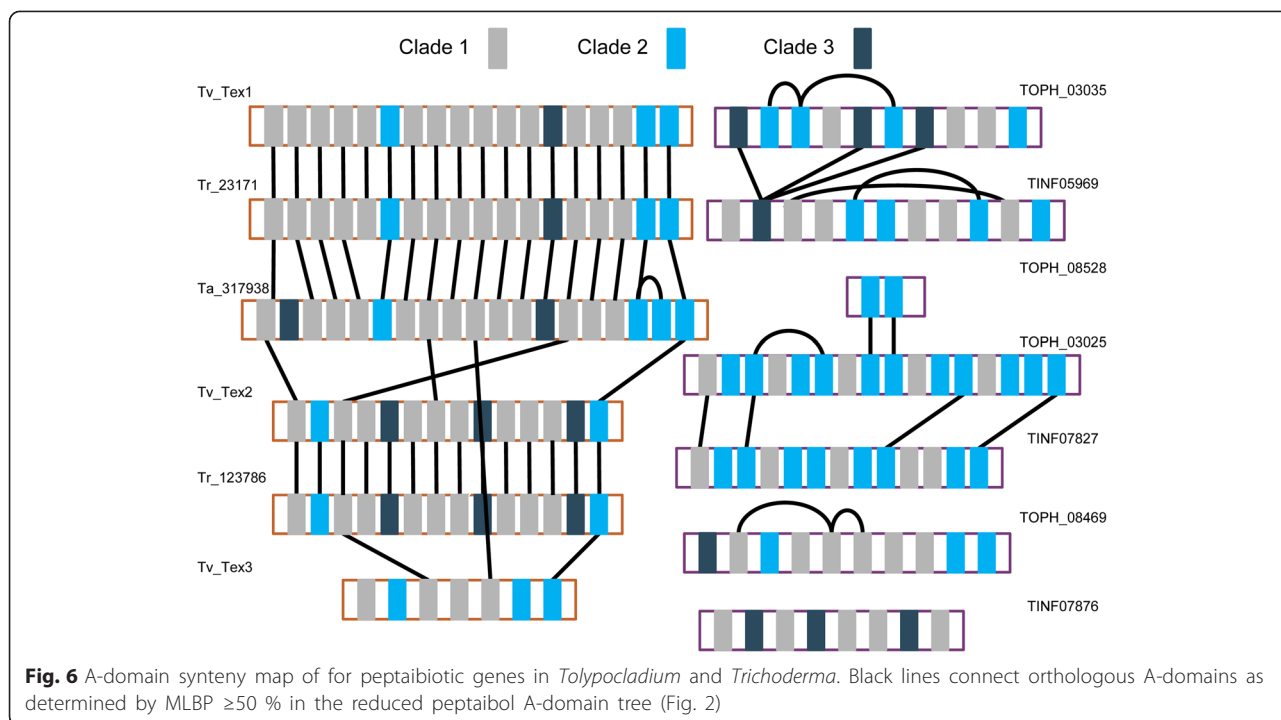
In contrast, five of the A-domains inferred to be present in the common ancestor of *Trichoderma* spp. deeply coalesce, and 59 are shared in the common ancestor. This pattern indicates a higher degree of domain tree – species tree congruence, which could be explained by two different mechanisms including 1) vertical descent with expansion and maintenance or 2) horizontal gene transfer from *Tolypocladium* to common ancestor of *Trichoderma*. To test for signatures of horizontal gene transfer, the A-domain tree was reconciled, with a modified species tree in which *Trichoderma* was sister to *Tolypocladium*, and out-group A-domains were included to root the domain tree. This produced a smaller deep coalescent cost (212 v. 275), but this is due to the fewer number of extinctions (in Cordycipitaceae, Clavicipitaceae, and *O. sinensis*) required (118 v. 181), because the number of duplications remained the same (91 v. 91). In this simulated reconciliation, 14 A-domains were inferred to deeply coalesce, and of these, seven were inherited in each lineage. Taken together, this suggests that the diversity of *Trichoderma* A-domains cannot solely be characterized as the product of horizontal transfer from *Tolypocladium* A-domains. Instead it suggests that the common ancestor of *Trichoderma* possessed a small number of peptaibiotic NRPSs and A-domains

that largely diversified in a manner consistent with speciation of the genus.

**T. ophioglossoides peptaibiotic gene clusters**

The four peptaibiotic genes (TOPH\_03025, TOPH\_03035, TOPH\_08469 and TOPH\_08528) are located in three gene clusters on three different scaffolds (Fig. 5, Additional file 1). Genes TOPH\_03035 (10 modules) and TOPH\_03025 (16 modules) are located within the same gene cluster and separated by only nine genes, including one PKS. In total, the cluster contains 18 genes, many of which are typically found within secondary metabolite clusters including two multidrug transporters (TOPH\_03022 and TOPH\_3034), a decarboxylase (TOPH\_03027), an esterase (TOPH\_03026), an epimerase (TOPH\_03033), and a leucine zipper transcription factor (TOPH\_03024) among others (Fig. 5a). At more than 17,000 amino acids in length, TOPH\_03025 encodes for the largest peptaibiotic NRPS produced by *T. ophioglossoides*, and is the largest gene (52.5 kb) in the genome. Gene TOPH\_08528 (2 modules), located on a separate scaffold, is only 2,587 amino acids in length and shares a high degree of amino acid identity with modules 8 and 9 of TOPH\_03025 (Fig. 6). The genes surrounding TOPH\_08528 include those typically found in an intact secondary metabolite cluster, including an ABC transporter, transcription factors, and an esterase, and there are no indications that TOPH\_08528 is a nonfunctional pseudogene. However, no short peptaibiotics have been described so far, but the two A-domains of this gene both fall within peptaibiotic Clade 2 (Fig. 3), and it is clearly orthologous to part of TOPH\_03025. So while it may not be producing a peptaibiotic, it is analyzed here only within the evolutionary context of these genes.





The remaining *T. ophioglossoides* peptaibiotic NRPS gene, TOPH\_08469 (10 modules), is located within a cluster containing two PKS genes (TOPH\_08457 and TOPH\_08462), several ABC transporters (TOPH\_08453, TOPH\_08459, and TOPH\_08470), an esterase (TOPH\_08466), an epimerase (TOPH\_08468), a hydrolase (TOPH\_08465), two cytochrome p450s (TOPH\_08458 and TOPH\_08455), and a RadR transcription regulator (TOPH\_08461) (Fig. 5b). It remains to be seen if the products of the PKSs are incorporated into the peptide created by TOPH\_08469. Some of the A-domains within TOPH\_08469 are divergent (Fig. 3), especially those that group (without support) as the earliest diverging lineages of Clade 2, and the presence of these A-domains in the tree, causes the support for this clade to weaken substantially.

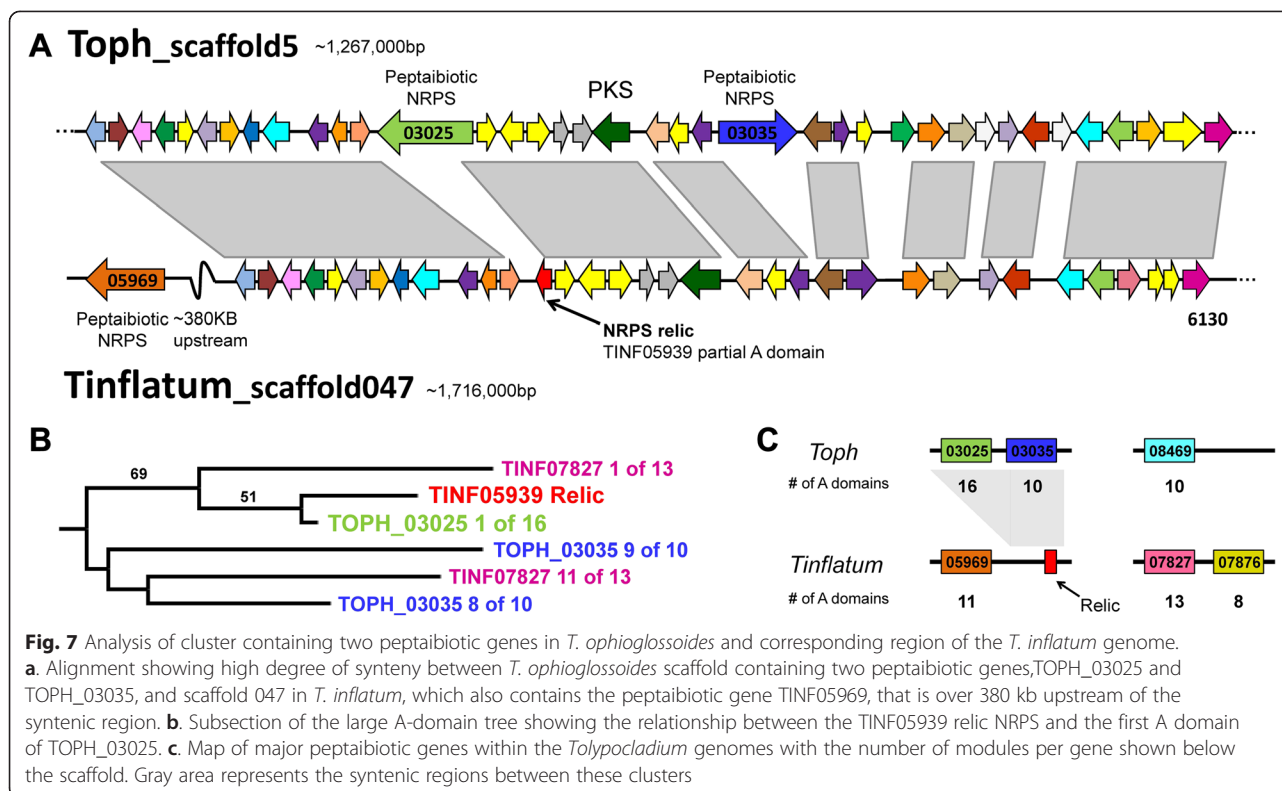
**Peptaibiotic cluster synteny between mycoparasitic and insect pathogenic *Tolypocladium* spp.**

Despite also possessing three large peptaibiotic genes in two clusters, *T. inflatum* peptaibiotic genes are highly divergent from *T. ophioglossoides* and the gene clusters are not located in the same regions of the genome (Fig. 7c). Alignment of *T. ophioglossoides* scaffold containing two peptaibiotic NRPS genes (TOPH\_03025 and TOPH\_03035) with the *T. inflatum* scaffold containing the single peptaibiotic NRPS gene (TINF05969), revealed high synteny in the *T. ophioglossoides* two-peptaibiotic gene cluster region, except for the absence of the two peptaibiotic genes, themselves (Fig. 7a). No other genes are present on the *T. inflatum* scaffold in the positions where the

peptaibiotic NRPS genes are located in the *T. ophioglossoides* genome. This alignment did reveal, however, the presence of a truncated N terminal portion of an NRPS (TINF05939) that aligns with the N terminal region of TOPH\_03025. When included in the entire hypocrealean A-domain phylogeny from Additional file 1, this *T. inflatum* NRPS “relic”, TINF05939, is most closely related to the first A-domain of TOPH\_03025 (Fig. 7b). While retrotransposon relics are reported from secondary metabolite gene clusters [62, 63] and secondary metabolite gene modules may be truncated [25], this is the first report of a relic NRPS domain that remains within an otherwise intact secondary metabolite cluster including the PKS located within the cluster. Due to the truncated nature of this protein model (only 68 amino acids in length) and lack of other functional domains, it is unlikely that this relic produces a secondary metabolite peptide. There is no evidence of transposable elements in or near this cluster in either of the *Tolypocladium* genomes. As mentioned above, this *T. inflatum* scaffold which is syntenous with the gene cluster in *T. ophioglossoides* containing two peptaibiotic NRPS genes, also contains the NRPS gene TINF05969, but this gene is located approximately 380 Kb downstream of the relic cluster (Fig. 7a).

The region of the scaffold containing the *T. ophioglossoides* third large peptaibiotic gene, TOPH\_08469, does not align well with any portion of the *T. inflatum* genome. Similarly, the final peptaibol cluster in *T. inflatum*, containing two peptaibiotic NRPS genes (TINF07827 and TINF07876), does not align well to any portion of the





*T. ophioglossoides* genome. The lack of synteny between these clusters in *Tolypocladium* spp. highlights the significant amount of genomic rearrangements between these closely related taxa. Campbell et al. [64] observed patterns of differential gene loss in *Botrytis* spp. within an ancient, horizontally-transferred, secondary metabolite gene cluster, leading to a patchy distribution of the genes within the clusters. This is not the pattern seen in the peptaibiotic clusters in *Tolypocladium* spp., in which the protein models are not reciprocal best BLAST hits (except for the protein models in the “relic” cluster in *T. inflatum*). Thus, despite the fact that their products may have similar functions, these peptaibiotic NRPS genes are divergent and located within nonhomologous gene clusters.

**Mixed homology of peptaibol A-domains in Hypocreales**

Using the moderate to strongly supported nodes ( $\geq 50$  maximum likelihood bootstrap percentage [MLBP]) in the A-domain phylogeny as a guide for module homology, the peptaibol NRPS genes are more conserved among the *Trichoderma* species examined as compared to species of *Tolypocladium* (Fig. 6), a finding reflected in the domain tree – species tree reconciliation analyses (Fig. 4). Using whole genome data of the sampled species of *Trichoderma*, A-domains from the *Tr. virens* three peptaibol NRPS genes (*tex1*, *tex2*, and *tex3*) [65] were identified for the phylogenetic analyses; *Tr. virens*

peptaibols contain 18, 14, and 7 modules respectively. *Tr. reesei* has two peptaibol NRPSs (*Tr\_23171* and *Tr\_123786*) which possess 18 and 14 modules. In the annotation of *Tr. atroviride* IMI 206040, the A-domain HMM identified one 19 module peptaibol NRPS (*Ta\_317938*), and several single A-domain protein models that group within the three peptaibiotic clades and are all located on scaffold 29. Further examination of this *Tr. atroviride* gene region using the JGI genome browser (Grigoriev et al. 2014) revealed that all of the *ab initio* gene predictions of that region predict a single protein model that is the approximate length of the 14 module peptaibol genes in *Tr. virens* and *Tr. reesei*. Degenkolb et al. [36] identified a homolog of the 14 modular peptaibol gene from a different strain of *Tr. atroviride*, and thus this is likely a mis-annotation of *Tr. atroviride* scaffold 29. Alignment of this scaffold in *Tr. atroviride* (scaffold 29) to those of *Tr. virens* and *Tr. reesei* (Additional file 5) revealed high nucleotide homology. However, the flanking regions did not align well, and a BLAST search of the *Tr. atroviride* genome using the nucleotide sequences from *Tr. virens* and *Tr. reesei* revealed that the 14 modular peptaibol gene in *Tr. atroviride* is located within a different portion of the genome than in the other two *Trichoderma* spp. (Additional file 5).

Comparing the *Trichoderma* A-domains, each species possesses one ortholog of the large 18 or 19 module

NRPS gene (Fig. 6), and at the nucleotide level these regions of their genomes also align and are syntenous (Additional file 5). The A-domains are syntenous in their arrangement within the large peptaibol NRPS genes across the species, except for: (a) the insertion of a Clade 3 domain at the third module, and (b) a duplication of either the seventeenth or eighteenth A-domains, which are most closely related. Within the two 14 module NRPS genes in *Tr. virens* and *Tr. reesei*, there is complete synteny of the A-domains. In *Tr. virens*, it has been demonstrated that this 14 module NRPS, Tex2, is responsible for two different sizes of peptaibols (11 and 14 residues in length) [35]. Due to differences in annotation (see above), the 14 module peptaibol gene in *Tr. atroviride* is not compared in this analysis. The short 7 module peptaibol synthetase from *Trichoderma* spp. is found only in *Tr. virens*. Between these three groups of peptaibol NRPSs, the terminating residues are all orthologous, as well as the initiating residues in the larger classes of peptaibol NRPSs.

In *Tolypocladium*, there is a very different pattern of homology and synteny between the peptaibiotic NRPSs of the two species. Only a few of the A-domain relationships within *Tolypocladium* are statistically supported ( $\geq 50$  MLBP) (Figs. 3 and 6). The first, third, and last A-domains of the largest NRPS in both species (TOPH\_03025 and TINF07827) are orthologous, but not the other domains within those two genes. There are several instances of intragenic module duplications which are known to occur within NRPS genes and have been proposed to play a role in the evolution of novel metabolites [23]. Within TOPH\_03035, for example, there is strong support for a shared ancestry between modules 2, 3 and 6 (Fig. 6), indicating that these modules are the product of lineage specific duplications (Fig. 4). This indicates a more complicated evolutionary history of these genes in *Tolypocladium*.

The lack of module synteny and orthology between *Tolypocladium* peptaibiotic gene modules is comparable to the lack of genomic synteny observed between their clusters. Part of this is due to the deep coalescence of the *Tolypocladium* A-domains. This evidence indicates that *Tolypocladium* peptaibiotic genes are not highly similar but are the products of more ancient divergences. This is notable, because in contrast to *Trichoderma* spp., all of which exhibit some degree of mycoparasitism (Druzhinina et al. 2011), *T. ophioglossoides* and *T. inflatum* have different ecologies, which are characterized by mycoparasitism and insect pathogenicity, respectively. Thus, if peptaibols are important in successful mycoparasitism (as the case has been made in *Trichoderma* spp. [30, 31], then there may be less selective pressure to maintain a specific mycoparasitic function of these extremely large (>10,000

amino acid) NRPS genes in more ecologically diverse lineages. Future genome-scale studies sampling other hypocrealean lineages containing both mycoparasites and insect pathogens, including the genus *Polycephalomycetes*, could further test this hypothesis.

## Conclusions

The genome of *T. ophioglossoides* is rich in secondary metabolite gene clusters, and 31 out of 38 of these clusters have no putative product. Given this potential and its life history as a mycoparasite, this species should be targeted for future studies to discover novel natural compounds with potential antibiosis, including antifungal, activity. The *simA* NRPS gene cluster, responsible for the production of the immunosuppressant cyclosporin, is not present in the *T. ophioglossoides* genome, but three large peptaibiotic genes are present within two clusters. These are the first data to suggest the potential for peptaibiotic production from a mycoparasitic species of *Tolypocladium*. This study confirms the presence of three phylogenetic clades of peptaibiotic NRPS A-domains from *Tolypocladium* and *Trichoderma* spp., and that peptaibiotics in general are limited to the mycoparasitic lineages of Hypocreales, based on current sampling. Reconciliation of the A-domain tree with the organismal phylogeny reveals that the peptaibiotic NRPSs of *Trichoderma* and *Tolypocladium* are likely the product of different mechanisms of diversification. *Trichoderma* is characterized by A-domain diversification that is largely consistent with speciation whereas *Tolypocladium* is characterized by A-domain diversity that shows patterns of deep coalescence. Deep coalescence is inconsistent with peptaibiotic NRPS diversity being the product of HGT to *Tolypocladium*, rather it is the product of complex patterns of lineage sorting and gains and losses of A-domains from hypocrealean ancestors. While the diversity of peptaibiotic NRPSs in *Trichoderma* could possibly be explained by HGT in the common ancestor of the three species, none of the *Tolypocladium* peptaibiotic NRPSs analyzed here are candidates for HGT. Further research is required to identify the structures of specific metabolites of the *Tolypocladium* gene clusters and to determine if these peptaibiotics are produced during mycoparasitism by *T. ophioglossoides* or if these genes are present in other mycoparasitic lineages of Hypocreales.

## Methods

### Genome sequencing

*T. ophioglossoides* strain CBS 100239 was grown for 7 days in a shaking incubator in potato dextrose broth (PDB) inoculated with plugs of tissue growing on potato dextrose agar for collection of tissue for DNA extraction. Tissue was harvested via filtration, frozen at  $-80$  °C in 1.5 mL tubes, and then lyophilized for 24 h. Lyophilized

tissue was ground using a mortar and pestle, and DNA was extracted using a Qiagen DNeasy Plant Mini kit following the standard protocol starting at the step with the addition of lysis buffer AP1 and eluted in 50  $\mu$ L water. Tissue for RNA extraction was grown in Yeast Malt (YM) broth, minimal media (MM) containing autoclaved insect cuticle with proteins removed using the protocol in Andersen (1980) [48], and MM containing lyophilized *Elaphomyces muricatus* peridium for 24 h and harvested into liquid nitrogen and stored at  $-80^{\circ}$  until extraction. RNA was extracted using the Qiagen RNeasy Plant kit following the manufacturer's protocol. The small insert DNA library was prepared using New England Biolabs NEBNext reagents, and size selection (350 bp) was performed using gel extraction. Nextera Mate Pair Sample Preparation of a large insert (6800 bp) library and sequencing was conducted by the Core Labs at the Center for Genome Research and Biocomputing (CGRB) at Oregon State University. The Illumina TruSeq RNA Sample Preparation Kit v2 was used for RNA library construction, using the manufacturer's suggested protocols including Agencourt AMPure magnetic beads for cleaning steps. All libraries were sequenced on the Illumina HiSeq2000 at the Core Labs of the CGRB with paired-end 101 cycles for DNA libraries and single-end 51 cycles for RNA libraries.

#### Assembly, annotation, and bioinformatic analyses

Using scripts in the fastx toolkit [66], raw reads were trimmed (to 50 bp in length) and filtered based on quality score (all bases  $\geq$  q20). Initial de novo assembly of the short insert reads was conducted in Velvet v. 1.19 [67] with over 156 million reads where the assembly had a median coverage depth of 74.45. The final trim length (50 bp) used in the assembly was chosen after trimming to different lengths (40–80 bp) followed by quality filtering and then the assembly with highest n50 and fewest number of contigs was selected as the “best” assembly. From that assembly, 50 million overlapping 150 bp paired reads were simulated with a 250 bp insert size using the program wgsim v. 0.3.1-r13 in “haploid” mode [68]. Final assembly using the simulated overlapping short insert library reads and the mate pair reads from the 6 kb library was conducted in AllPaths-LG with default settings [69]. The Core Eukaryotic Mapping Genes Approach (CEGMA) was used to estimate the completeness of the *T. ophioglossoides* genome [49]. Scripts in the Mummer3 package were used to create a mummerplot between *T. ophioglossoides* and *T. inflatum*; specifically nucmer v. 3.07 was run with default settings [70].

Gene model predictions were created using the Maker annotation pipeline [71] incorporating RNA data assembled in Trinity [72] using the Jellyfish v. 2.0 method of kmer counting [73]. Other information given to Maker included

a custom hidden Markov model (HMM) for *T. ophioglossoides* built by Genemark-ES v 2.0 [74], a SNAP HMM [75] trained on *Fusarium graminearum*, which was also set as the species model for AUGUSTUS [76], and protein and/or EST data from the following hypocrealean taxa: *F. graminearum*, *N. haematococca*, *Tr. reesei*, *Tr. virens*, *M. robertsii*, *T. inflatum*, *C. militaris*, *B. bassiana*. Annotation of transposable elements was performed in RepeatMasker v 3.2.8 with organism set to “fungi” [77], and custom repeat content was estimated using RepeatScout v 1.0.3 and scripts associated with that package [78]. Non-overlapping *ab initio* protein models were aligned using BLAST [79] against a custom database of all the protein models of all the hypocrealean taxa used in this study. Any of these protein models with a significant hit ( $\leq 1e^{-5}$ ) were included in the final protein set and used for downstream analyses.

Using a set of NRPS A-domains from a wide array of published fungal genomes [20] an HMM was created for this study using the program Hmmer 3.0 [80]. This HMM was then used to mine the 18 hypocrealean genomes used for this study for the identification of A-domains. Putative A-domains identified were filtered for short sequences (less than 100 bp), and where applicable cross referenced with published reports of NRPS from those species (e.g., *Tr. virens* Tex1) [34]. Additional annotation of secondary metabolite clusters was completed using the antiSMASH [81] and SMURF [82] pipelines. A-domain trees were reconciled with species trees in Mesquite v. 2.75 with the contained tree treated as unrooted [83].

Whole scaffold alignments were performed in the program Mauve [84] with default progressivemaue alignment settings.

#### Phylogenetic analyses

Predicted A-domain amino acid sequences were aligned using MUSCLE v 3.8.31 [85] under default settings. Gaps were removed manually, and all alignments were analyzed using RAxML v 7.2.6 [86] using the Gamma model of rate heterogeneity and the WAG substitution matrix with 100 bootstrap replicates.

Whole genome phylogenomic analyses were executed in the HAL pipeline [87]. Orthologous clusters of proteins were identified in MCL [88] across inflation parameters 1.2, 3 and 5. Briefly, orthologous clusters were filtered for retention of clusters with one sequence per genome and removal of any redundant clusters. The resulting unique, single-copy orthologous clusters of proteins were aligned in MUSCLE [85] with default settings; poorly aligned regions were identified using Gblocks ([89]; gap removal setting = c, for conservative) and excluded from subsequent analyses. The aligned clusters were concatenated into a superalignment and maximum likelihood analysis was performed using RAxML v 7.2.6 with the Gamma

model of rate heterogeneity and the WAG substitution matrix with 100 bootstrap replicates.

### Availability of supporting data

This Whole Genome Shotgun project has been deposited at DDBJ/EMBL/GenBank under the accession LFRF00000000. The version described in this paper is version LFRF01000000.

### Additional files

#### Additional file 1: Table of secondary metabolite genes and clusters in *T. ophioglossoides*.

Genes belonging to the same clusters are highlighted in the same color. Cluster prediction based on antiSMASH. Siderophores are marked as either intracellular (i) or extracellular (e). Orthologs of *T. inflatum* (*Ti*) were annotated as reciprocal best BLAST hits and further based on amino acid alignment and A-domain phylogeny. (DOCX 22 kb)

#### Additional file 2: RAxML phylogeny of A-domains mined from the hypocrealean genomes samples (Fig. 1).

A-domains from *T. ophioglossoides* are highlighted in blue. Taxonomic abbreviations: *Alternaria alternata* (Aalt), *Aspergillus fumigatus* (Afu), *A. nidulans* (Anid), *Beauveria bassiana* (Bbass), *Claviceps purpurea* (Cpurp), *Cochliobolus carbonum* (Ccarb), *Cordyceps militaris* (CCM), *Epichloë festucae* (Efest), *Fusarium equiseti* (Fe), *F. graminearum* (FGSG), *F. heterosporum* (Fh), *F. oxysporum* (FOXG), *F. verticillioides* (FVEG), *Gibberella fujikuroi* (Gf), *Leptosphaeria maculans* (Lmac), *Magnaporthe grisea* (MGG), *Metarhizium acridum* (MAC), *M. robertsii* (MAA), *Nectria haematococca* (Nhaem), *Neurospora crassa* (Ncrass), *Ophiocordyceps sinensis* (Osin), *Penicillium chrysogenum* (PC), *Rhizopus oryzae* (RO), *Tolypocladium inflatum* (TINF), *T. ophioglossoides* (TOPH), *Trichoderma atroviride* (Tr\_atro), *Tr. reesei* (Tr\_reesei), *Tr. virens* (Tr\_virens), *Ustilago Maydis* (UM), *Verticillium dahliae* (VDAG). (PDF 57 kb)

#### Additional file 3: Destruxins NRPS and cluster in *T. ophioglossoides* and other taxa.

A. Excerpts from larger A-domain phylogeny (Additional file 2) showing the phylogenetic relationships of destruxins A-domains (which group into two areas within the A-domain phylogeny; the PerA clade, and then *T. inflatum simA* clade) in *M. robertsii*, *T. ophioglossoides*, and *Tr. virens*. B. Nucleotide alignment of *M. robertsii* destruxins cluster contig with homologous region in *T. ophioglossoides* genome. Abbreviations as in Additional file 2. (TIFF 409 kb)

#### Additional file 4: A-domain Clade 3 tree/species tree reconciliation.

Reconciliation of the peptaibiotic A-domain tree with the species tree. Abbreviations as in Fig. 4. (TIFF 1364 kb)

#### Additional file 5: Alignment of *Trichoderma* peptaibol scaffolds.

Top panel depicts alignments of the regions of *Trichoderma* genomes that contain the large (18–19 module) peptaibol genes (which are boxed in blue). Bottom panel shows alignments of scaffold containing the *Trichoderma* 14 module peptaibol genes. one demonstrates that the region around the 14 module peptaibols in *Tr. reesei* and *Tr. virens* are conserved and syntenic with *Tr. atroviride* contig 25, which does not contain an NRPS. Alignment two shows the same two scaffolds in *Tr. reesei* and *Tr. virens*, but aligned with *Tr. atroviride* contig 29, which does not share synteny with the other two species except for the peptaibol NRPS. (TIFF 10788 kb)

### Competing interests

The authors declare that they have no competing interests.

### Authors' contributions

CAQ participated in the conception and design of the study, carried out the molecular sequencing, conducted the sequence analysis and bioinformatics, and wrote the manuscript. KEB participated in the design of the study and helped to draft the manuscript. JWS participated in the conception of the study, contributed to its coordination and design, participated in results interpretation, and helped to draft the manuscript. All authors read and approved the final manuscript.

### Acknowledgements

The authors thank Mark Desanko, Cedar Hesse, and Chris Sullivan for providing technical and computational assistance for this project. This project was supported with funds from the NSF grant DEB-0732993 to JWS, and CAQ was supported by an NSF Graduate Research Fellowship.

### Author details

<sup>1</sup>Department of Botany and Plant Pathology, Oregon State University, Corvallis, OR, USA. <sup>2</sup>Department of Plant Biology, University of Minnesota, St. Paul, MN, USA. <sup>3</sup>Present address: Department of Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, MI, USA.

Received: 15 January 2015 Accepted: 14 July 2015

Published online: 28 July 2015

### References

- Sung G-H, Hywel-Jones NL, Sung J-M, Luangsang-Ard JJ, Shrestha B, Spatafora JW. Phylogenetic classification of *Cordyceps* and the clavicipitaceous fungi. *Stud Mycol.* 2007;57:5–59.
- Spatafora JW, Sung G-H, Sung J-M, Hywel-Jones NL, White JF. Phylogenetic evidence for an animal pathogen origin of ergot and the grass endophytes. *Mol Ecol.* 2007;16:1701–11.
- Kepler RM, Sung G-H, Harada Y, Tanaka K, Tanaka E, Hosoya T, et al. Host jumping onto close relatives and across kingdoms by *Tyrannicordyceps* (Clavicipitaceae) gen. nov. and *Ustilagoidea* (Clavicipitaceae). *Am J Bot.* 2012;99:552–61.
- Kepler R, Ban S, Nakagiri A, Bischoff J, Hywel-Jones N, Owensby CA, et al. The phylogenetic placement of hypocrealean insect pathogens in the genus *Polycephalomyces*: an application of One Fungus One Name. *Fungal Biol.* 2013;117:611–22.
- Hjeljord L, Tronsmo A: *Trichoderma* and *Gliocladium* in biological control: an overview. In *Trichoderma Gliocladium. Enzymes Biological Control and Commercial Applications*. Edited by Harman GE, Kubicek CP; 2002:115–133
- Durand H, Clanet M, Tiraby G. Genetic improvement of *Trichoderma reesei* for large scale cellulase production. *Enzyme Microb Technol.* 1988;10:341–6.
- Sukumaran RK, Singhanian RR, Pandey A. Microbial cellulases: Production, applications and challenges. *J Scientific and Industrial Reseach.* 2005;64:832–44.
- Quandt CA, Kepler RM, Gams W, Araujo JPM, Ban S, Evans HC, et al. Phylogenetic-based nomenclatural proposals for Ophiocordycipitaceae (Hypocreales) with new combinations in *Tolypocladium*. *IMA Fungus.* 2014;5:121–34.
- Landvik S, Shailer NFJ, Eriksson OE. SSU rDNA sequence support for a close relationship between the Elaphomycetales and the Eurotiales and Onygenales. *Mycoscience.* 1996;37:237–41.
- LoBuglio KF, Berbee ML, Taylor JW. Phylogenetic origins of the asexual mycorrhizal symbiont *Cenococcum geophilum* Fr. and other mycorrhizal fungi among the ascomycetes. *Mol Phylogenet Evol.* 1996;6:287–94.
- Mains EB. Species of *Cordyceps* parasitic on *Elaphomyces*. *Bull Torrey Bot Club.* 1957;84:243–51.
- Kobayasi Y, Shimizu D. Monographic studies on *Cordyceps* 1. Group parasitic on *Elaphomyces*. *Bull Natl Sci Museum, Tokyo.* 1960;5:69–85.
- Sung G-H, Poinar GO, Spatafora JW. The oldest fossil evidence of animal parasitism by fungi supports a Cretaceous diversification of fungal-arthropod symbioses. *Mol Phylogenet Evol.* 2008;49:495–502.
- Borel JF. History of the discovery of cyclosporin and of its early pharmacological development. *Wien Klin Wochenschr.* 2002;114:433.
- Keller NP, Turner G, Bennett JW. Fungal secondary metabolism - from biochemistry to genomics. *Nat Rev Microbiol.* 2005;3:937–47.
- Isaka M, Kittakoop P, Thebtaranonth Y: Secondary Metabolites of Clavicipitacean Fungi. In *Clavicipitacean Fungi Evolutionary Biology, Chemistry, Biocontrol And Cultural Impacts*. Edited by White JF, Bacon CW, Hywel-Jones NL, Spatafora JW. New York, NY: Marcel Dekker; 2003:355–98.
- Desjardins AE, Proctor RH. Molecular biology of *Fusarium* mycotoxins. *Int J Food Microbiol.* 2007;119:47–50.
- Molnár I, Gibson DM, Krasnoff SB. Secondary metabolites from entomopathogenic Hypocrealean fungi. *Nat Prod Rep.* 2010;27:1241–75.
- Marahiel MA, Stachelhaus T, Mootz HD. Modular Peptide Synthetases Involved in Nonribosomal Peptide Synthesis. *Chem Rev.* 1997;97:2651–74.



20. Bushley KE, Turgeon BG. Phylogenomics reveals subfamilies of fungal nonribosomal peptide synthetases and their evolutionary relationships. *BMC Evol Biol.* 2010;10:26.
21. Wei X, Yang F, Straney DC. Multiple non-ribosomal peptide synthetase genes determine peptaibol synthesis in *Trichoderma virens*. *Can J Microbiol.* 2005;51:423–9.
22. Jenke-Kodama H, Sandmann A, Müller R, Dittmann E. Evolutionary implications of bacterial polyketide synthases. *Mol Biol Evol.* 2005;22:2027–39.
23. Fischbach MA, Walsh CT, Clardy J. The evolution of gene collectives: How natural selection drives chemical innovation. *Proc Natl Acad Sci.* 2008;105:4601–8.
24. Zhang YQ, Wilkinson H, Keller NP, Tsitsigiannis D, An Z. Secondary metabolite gene clusters. In: An Z, editor. *Handbook of industrial microbiology*. New York: Dekker; 2004. p. 355–86.
25. Hoffmeister D, Keller NP. Natural products of filamentous fungi: enzymes, genes, and their regulation. *Nat Prod Rep.* 2007;24:393–416.
26. Chugh JK, Wallace BA: Peptaibols: models for ion channels Sequence alignments into subfamilies (SFs). 2001:565–570.
27. Fox RO, Richards FM. A voltage-gated ion channel model inferred from the crystal structure of alamethicin at 1.5-Å resolution. *Nature.* 1982;300:325–30.
28. Chugh JK, Bruckner H, Wallace BA. Model for a Helical Bundle Channel Based on the High-Resolution Crystal Structure of Trichotoxin \_ A50E. *Biochemistry.* 2002;41:12934–41.
29. Whitmore L, Wallace BA. The Peptaibol Database: a database for sequences and structures of naturally occurring peptaibols. *Nucleic Acids Res.* 2004;32(Database issue):D593–4.
30. Röhrich CR, Iversen A, Jaklitsch WM, Voglmayr H, Berg A, Dörfelt H, et al. Hypopulvins, novel peptaibiotics from the polyporicolous fungus *Hypocrea pulvinata*, are produced during infection of its natural hosts. *Fungal Biol.* 2012;116:1219–31.
31. Schimbock M, Lorito M, Wang Y, Hayes CK, Arisan-atac I, Scala F, et al. Parallel Formation and Synergism of Hydrolytic Enzymes and Peptaibol Antibiotics, Molecular Mechanisms Involved in the Antagonistic Action of *Trichoderma harzianum* against Phytopathogenic Fungi. *Appl Environ Microbiol.* 1994;60:4364–70.
32. Lorito M, Peterbauer C, Hayes CK, Harman GE. Synergistic interaction between fungal cell wall degrading enzymes and different antifungal compounds enhances inhibition of spore germination. *Microbiology.* 1994;140(Pt 3):623–9.
33. Lorito M, Woo SL, D'Ambrosio M, Harman GE, Hayes CK, Kubicek CP, et al. Synergistic interaction between cell wall degrading enzymes and membrane affecting compounds. *Mol Plant Microbe Interact.* 1996;9:206–13.
34. Wiest A, Grzegorski D, Xu B-W, Goulard C, Rebuffat S, Ebbold DJ, et al. Identification of peptaibols from *Trichoderma virens* and cloning of a peptaibol synthetase. *J Biol Chem.* 2002;277:20862–8.
35. Mukherjee PK, Wiest A, Ruiz N, Keightley A, Moran-Diez ME, McCluskey K, et al. Two classes of new peptaibols are synthesized by a single non-ribosomal peptide synthetase of *Trichoderma virens*. *J Biol Chem.* 2011;286:4544–54.
36. Degenkolb T, Agheheh RK, Dieckmann R, Neuhof T, Baker SE, Druzhinina IS, et al. The production of multiple small peptaibol families by single 14-module peptide synthetases in *Trichoderma/Hypocrea*. *Chem Biodivers.* 2012;9:499–535.
37. Krishna K, Sukumar M, Balaran P, Unit MB. Structural chemistry and membrane modifying activity of the fungal polypeptides zervamicins, antiamoebins and efrapeptins. *Pure Appl Chem.* 1990;62:1417–20.
38. Krasnoff SB, Gupta S. Efrapeptin production by *Tolypocladium* fungi (Deuteromycotina: Hyphomycetes): Intra- and interspecific variation. *J Chem Ecol.* 1992;18:1727–41.
39. Bandani AR, Khambay BPS, Faull JL, Newton R, Deadman M, Butt TM. Production of efrapeptins by *Tolypocladium* species and evaluation of their insecticidal and antimicrobial properties. *Mycol Res.* 2000;104:537–44.
40. Gupta S, Krasnoff SB, Roberts DW, Renwick JAA. Structure of Efrapeptins from the fungus *Tolypocladium niveum*: peptide inhibitors of mitochondrial ATPase. *J Org Chem.* 1992;57:2306–13.
41. Nagaraj G, Uma MV, Shivayogi MS. Antimalarial Activities of Peptide Antibiotics Isolated from Fungi. *Antimicrob Agents Chemother.* 2001;45:145–9.
42. Martinez D, Berka RM, Henrissat B, Saloheimo M, Arvas M, Baker SE, et al. Genome sequencing and analysis of the biomass-degrading fungus *Trichoderma reesei* (syn. *Hypocrea jecorina*). *Nat Biotechnol.* 2008;26:553–60.
43. Kubicek CP, Herrera-Estrella A, Seidl-Seiboth V, Martinez DA, Druzhinina IS, Thon M, et al. Comparative genome sequence analysis underscores mycoparasitism as the ancestral life style of *Trichoderma*. *Genome Biol.* 2011;12:R40.
44. Zheng P, Xia Y, Xiao G, Xiong C, Hu X, Zhang S, et al. Genome sequence of the insect pathogenic fungus *Cordyceps militaris*, a valued traditional Chinese medicine. *Genome Biol.* 2011;12:R116.
45. Xiao G, Ying S-H, Zheng P, Wang Z-L, Zhang S, Xie X-Q, et al. Genomic perspectives on the evolution of fungal entomopathogenicity in *Beauveria bassiana*. *Sci Rep.* 2012;2:483.
46. Gao Q, Jin K, Ying SH, Zhang Y, Xiao G, Shang Y, et al. Genome sequencing and comparative transcriptomics of the model entomopathogenic fungi *Metarhizium anisopliae* and *M. acridum*. *PLoS Genet.* 2011;7:e1001264.
47. Hu X, Zhang Y, Xiao G, Zheng P, Xia Y, Zhang X, et al. Genome survey uncovers the secrets of sex and lifestyle in caterpillar fungus. *Chinese Sci Bull.* 2013;58:2846–54.
48. Bushley KE, Raja R, Jaiswal P, Cumbie JS, Nonogaki M, Boyd AE, et al. The genome of *Tolypocladium inflatum*: evolution, organization, and expression of the cyclosporin biosynthetic gene cluster. *PLoS Genet.* 2013;9, e1003496.
49. Parra G, Bradnam K, Ning Z, Keane T, Korf I. Assessing the gene space in draft genomes. *Nucleic Acids Res.* 2009;37:289–97.
50. Sedmera P, Havlicek V, Jegorov A, Segre AL. Cyclosporin D Hydroperoxide, a new metabolite of *Tolypocladium terricola*. *Tetrahedron Lett.* 1995;36:6953–6.
51. Traber R, Dreyfuss MM. Occurrence of cyclosporins and cyclosporin-like peptolides in fungi. *J Ind Microbiol.* 1996;17:397–401.
52. Wiemann P, Guo C-J, Palmer JM, Sekonyela R, Wang CCC, Keller NP. Prototype of an intertwined secondary-metabolite supercluster. *Proc Natl Acad Sci U S A.* 2013;110:17065–70.
53. Kershaw M, Moorhouse E, Bateman R, Reynolds S, Charnley A. The role of destruxins in the pathogenicity of *Metarhizium anisopliae* for three species of insect. *J Invertebr Pathol.* 1999;74:213–23.
54. Wang B, Kang Q, Lu Y, Bai L, Wang C. Unveiling the biosynthetic puzzle of destruxins in *Metarhizium* species. *Proc Natl Acad Sci U S A.* 2012;109:1287–92.
55. Kneifel H, König WA, Loeffler W, Müller R. Ophiocordin, an antifungal antibiotic of *Cordyceps ophioglossoides*. *Arch Microbiol.* 1977;113:121–30.
56. Boros C, Hamilton SM, Katz B, Kulanthaivel P. Comparison of Balanol from *Verticillium balanoides* and ophiocordin from *Cordyceps ophioglossoides*. *J Antibiot (Tokyo).* 1994;47:1010–6.
57. Putri SP, Kinoshita H, Ihara F, Igarashi Y, Nihira T. Ophiosetin, a new tetramic acid derivative from the mycopathogenic fungus *Elaphocordyceps ophioglossoides*. *J Antibiot (Tokyo).* 2010;63:195–8.
58. Sims JW, Fillmore JP, Warner DD, Schmidt EW. Equisetin biosynthesis in *Fusarium heterosporum*. *Chem Commun (Camb).* 2005;186–8.
59. Tsantrizos YS, Pischos S, Sauriol F, Widden P. Peptaibol metabolites of *Tolypocladium geodes*. *Can J Chem.* 1996;172:165–72.
60. Lee S-J, Yeo W-H, Yun B-S, Yoo I-D. Isolation and sequence analysis of new peptaibol, boletusin, from *Boletus* spp. *J Pept Sci.* 1999;5:374–8.
61. Lee S-J, Yun B-S, Cho D-H, Yoo I-D. Tylopeptins A and B, new antibiotic peptides from *Tylophilus neofelleus*. *J Antibiot (Tokyo).* 1999;52:998–1006.
62. Young C, McMillan L, Telfer E, Scott B. Molecular cloning and genetic analysis of an indole-diterpene gene cluster from *Penicillium paxilli*. *Mol Microbiol.* 2001;39:754–64.
63. Fleetwood DJ, Scott B, Lane GA, Tanaka A, Johnson RD. A complex ergovaline gene cluster in *Epichloë* endophytes of grasses. *Appl Environ Microbiol.* 2007;73:2571–9.
64. Campbell MA, Staats M, van Kan JAL, Rokas A, Slot JC. Repeated loss of an anciently horizontally transferred gene cluster in *Botrytis*. *Mycologia.* 2013;105:1126–34.
65. Mukherjee PK, Horwitz BA, Kenerley CM. Secondary metabolism in *Trichoderma*—a genomic perspective. *Microbiology.* 2012;158(Pt 1):35–45.
66. FASTQ/A short-reads pre-processing tools [[http://hannonlab.cshl.edu/fastx\\_toolkit/](http://hannonlab.cshl.edu/fastx_toolkit/)]
67. Zerbino DR, Birney E. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res.* 2008;18:821–9.
68. Wgsim - read simulator for next generation sequencing [<http://github.com/lh3/wgsim>]
69. Gnerre S, Maccallum I, Przybylski D, Ribeiro FJ, Burton JN, Walker BJ, et al. High-quality draft assemblies of mammalian genomes from massively parallel sequence data. *Proc Natl Acad Sci U S A.* 2011;108:1513–8.

70. Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, Antonescu C, et al. Versatile and open software for comparing large genomes. *Genome Biol.* 2004;5:R12.
71. Cantarel BL, Korf I, Robb SMC, Parra G, Ross E, Moore B, et al. MAKER: an easy-to-use annotation pipeline designed for emerging model organism genomes. *Genome Res.* 2008;18:188–96.
72. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol.* 2011;29:644–52.
73. Marçais G, Kingsford C. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics.* 2011;27:764–70.
74. Ter-Hovhannisyanyan V, Lomsadze A, Chernoff YO, Borodovsky M. Gene prediction in novel fungal genomes using an *ab initio* algorithm with unsupervised training. *Genome Res.* 2008;18:1979–90.
75. Korf I. Gene finding in novel genomes. *BMC Bioinformatics.* 2004;5:59.
76. Stanke M, Morgenstern B. AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints. *Nucleic Acids Res.* 2005;33(Web Server issue):W465–7.
77. RepeatMasker [<http://www.repeatmasker.org>]
78. Price AL, Jones NC, Pevzner PA. De novo identification of repeat families in large genomes. *Bioinformatics.* 2005;21 Suppl 1:i351–8.
79. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic Local Alignment Search Tool. *J Mol Biol.* 1990;215:403–10.
80. Eddy SR. A new generation of homology search tools based on probabilistic inference. *Genome informatics.* 2009;23:205–11.
81. Blin K, Medema MH, Kazempour D, Fischbach MA, Breitling R, Takano E, et al. antiSMASH 2.0—a versatile platform for genome mining of secondary metabolite producers. *Nucleic Acids Res.* 2013;41(Web Server issue):W204–12.
82. Khaldi N, Seifuddin FT, Turner G, Haft D, Nierman WC, Wolfe KH, et al. SMURF: Genomic mapping of fungal secondary metabolite clusters. *Fungal Genet Biol.* 2010;47:736–41.
83. Mesquite: a modular system for evolutionary analysis [<http://mesquiteproject.org>]
84. Darling AE, Mau B, Perna NT. ProgressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. *PLoS One.* 2010;5, e11147.
85. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 2004;32:1792–7.
86. Stamatakis A. RAXML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics.* 2006;22:2688–90.
87. Robbertse B, Yoder RJ, Boyd A, Reeves J, Spatafora JW. Hal: an automated pipeline for phylogenetic analyses of genomic data. *PLoS Curr.* 2011;3, RRR1213.
88. Enright AJ, Van DS, Ouzounis CA. An efficient algorithm for large-scale detection of protein families. *Nucleic Acid Res.* 2002;30:1575–84.
89. Talavera G, Castresana J. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Syst Biol.* 2007;56:564–77.

**Submit your next manuscript to BioMed Central and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at  
[www.biomedcentral.com/submit](http://www.biomedcentral.com/submit)

