Phylogenomic analysis of polyketide synthase-encoding genes in *Trichoderma*

Scott E. Baker,1 Giancarlo Perrone,2 Nathan M. Richardson,3 Antonia Gallo2 and Christian P. Kubicek4

Members of the economically important ascomycete genus *Trichoderma* are ubiquitously distributed around the world. The mycoparasitic lifestyle and plant defence-inducing interactions of *Trichoderma* spp. make them ideal biocontrol agents. Of the *Trichoderma* enzymes that produce secondary metabolites, some of which likely play important roles in biocontrol processes, polyketide synthase (PKSs) have garnered less attention than non-ribosomal peptide synthetases such as those that produce peptaibols. We have taken a phylogenomic approach to study the PKS repertoire encoded in the genomes of *Trichoderma reesei*, *Trichoderma atroviride* and *Trichoderma virens*. Our analysis lays a foundation for future research related to PKSs within the genus *Trichoderma* and in other filamentous fungi.

**INTRODUCTION**

Given its important roles in bioenergy-related enzyme production and plant and fungal health, the genus *Trichoderma* has a significant economic footprint (Schuster & Schmoll, 2010). This genus is well known for its opportunistic lifestyle that includes saprotrophy, mycoparasitism, endophytism, and interactions with plants and animals. Because of their ability to antagonize other fungi and stimulate plant defences against phytopathogens, *Trichoderma* strains have become prominent biocontrol agents (Schuster & Schmoll, 2010). Further highlighting their mycoparasitic lifestyle, sequencing and analysis of the genomes from two mycoparasitic *Trichoderma* species (*Trichoderma virens*, *Trichoderma atroviride*) have shown that these genomes encode a rich catalogue of fungal cell wall-degrading chitinases and a relative paucity of plant cell wall-degrading enzymes (Kubicek et al., 2011; Martinez et al., 2008).

As biocontrol agents and pathogens of fungal food crops, secondary metabolite production in *Trichoderma* spp. is of particular importance. *Trichoderma* and other filamentous fungi are well known for the production of a diverse array of secondary metabolites. Non-ribosomal peptides and polyketides comprise a major portion of these products, which are synthesized by large proteins composed of various domains for the individual enzymic steps (Cox, 2007). Both broad and genus-specific phylogenomic and functional analysis of polyketide synthase (PKS)- and non-ribosomal peptide synthetase (NRPS)-encoding genes have been performed, which has accelerated the pace at which genes and pathways are being linked to specific secondary metabolites (Bushley & Turgeon, 2010; Bushley et al., 2008; Chiang et al., 2010; Cramer et al., 2006; Gaffoor et al., 2005; Kroken et al., 2003; Kubicek et al., 2011; Lee et al., 2005).

*Trichoderma* spp. are probably best known for their production of peptaibols, which are non-ribosomal peptides with antimicrobial and plant defence-stimulating activity (Viterbo et al., 2007). Relatively less studied in *Trichoderma* are PKSs. The recent analysis of the three available *Trichoderma* genome sequences has indicated that the genomes of these organisms encode only a small catalogue of PKSs: *T. atroviride* and *T. virens* each encode 18, and *Trichoderma reesei* only 11 predicted PKSs (Kubicek et al., 2011; Martinez et al., 2008). We have manually annotated and performed phylogenomic analysis of the catalogue of *Trichoderma* PKSs.
METHODS

Phylogenetic analysis. The amino acid sequences of the keto-synthase (KS) domains were aligned using CLUSTAL W and subsequently manually adjusted. The alignment was then analysed by the COBALT multiple sequence alignment tool that finds, using RPS-BLAST, the amino acid sequences of the keto-synthase (KS) domains were aligned using CLUSTAL W and subsequently manually adjusted. The alignment was then analysed by the COBALT multiple sequence alignment tool that finds, using RPS-BLAST, BLASTp and PHI-BLAST, a collection of pairwise constraints derived from a conserved domain database, protein motif database, and sequence similarity. Pairwise constraints are then incorporated into a progressive multiple alignment (Papadopoulos & Agrawala, 2007). The resulting alignment exported in FASTA format was phylogenetically analysed.

The genealogy of 196 PKSs on the basis of the alignment of the KS domain was inferred by neighbour-joining and maximum-parsimony (MP) analysis using MEGA version 4 (Tamura et al., 2007). Branches of zero length were collapsed and all multiple, equally parsimonious trees were saved. The majority rule consensus bootstrap tree inferred from 1000 replicates is taken to represent the evolutionary history of the taxa analysed. The MP trees were obtained using the Close-Neighbor-Interchange algorithm with search level 3, in which the initial trees were obtained with the random addition of sequences (10 replicates). The trees are drawn to scale, with branch lengths calculated using the average pathway method, and are in the units of the number of changes over the whole sequence. All positions containing gaps and missing data were eliminated from the dataset (Complete Deletion option). In addition, a more robust approach was used to infer phylogenomic relationship among the PKSs studied by performing a randomized bootstrap maximum-likelihood analysis using RAxML software (Stamatakis et al., 2008), setting the bootstrap analysis to 1000 runs and the bootstrap random seed value to 12345. There were 736 characters and the Dayhoff mutation data matrix was used for the analysis of the alignment. Protein IDs are consistent with the comparative Trichoderma genome analyses presented in Kubicek et al. (2011); revisions to protein IDs between Joint Genome Institute (JGI) genome versions are noted in Supplementary Table S1.

K/K/ ratio. Tajima’s D statistic (Tajima, 1989) was determined with DNASP 5.0 (Librado & Rozas, 2009), using a sliding-window approach. The pairwise K/K ratio [the ratio of the number of non-synonymous substitutions per non-synonymous site (K/K) to the number of synonymous substitutions per synonymous site (K/)] was determined on the respective phylogenetic tree with the aid of the K/K calculation tool (http://services.csbio.ubio.net/tools/kaks), using pre-aligned sequence datasets and MP analysis. Codon-based Fisher’s test and the codon-based Z-test, implemented in MEGA 4.0 (Tamura et al., 2007), were used to directly test the hypotheses of evolutionary models.

Codon usage was estimated using the codon adaptation index (CAI), which is a measure of the relative adaptation of the gene codon usage towards the codon usage of highly expressed genes for that organism (Sharp & Li, 1987). To estimate this index, the EMBoss (http://emboss.bioinformatics.nl/cgi-bin/emboss) program ‘ca’ was used (Rice et al., 2000).

RESULTS

Phylogenetic analysis and manual annotation of clusters

T. reesei was previously shown to contain 11 PKSs (Martinez et al., 2008). This number is exceeded by T. atroviride and T. virens, both of which have 18 PKS-encoding genes (Kubicek et al., 2011). The organizational framework for fungal PKSs based on a phylogeny of fungal PKS KS domains is well established (Kroken et al., 2003). We have therefore analysed the position of the Trichoderma PKS proteins in the phylogenetic tree (Fig. 1). The data revealed that Trichoderma contains PKS proteins that are members of the non-reducing clades I, II and III, and of the reducing clades I, III and IV. We have divided the PKS-encoding genes into orthologous groups (Table 1) and singlets (Table 2). The T. reesei genome has only two singlet PKS-encoding genes, while the T. virens and T. atroviride genomes have seven and nine singlet PKS-encoding genes, respectively (Fig. 2, Table 2).

Predicted pigment genes

The genomes of all three Trichoderma genomes encode PKS genes that group in the non-reducing fungal PKS clade I, which includes the genes associated with Fusarium graminearum aurofusarin (Frandsen et al., 2011; Kim et al., 2005; Malz et al., 2005), Fusarium fukuiroii bikaverin (Linnemannstöns et al., 2002; Wiemann et al., 2009) and Aspergillus spp. DHN melanin (Baker, 2008; Chiang et al., 2011; Jørgensen et al., 2011; Langfelder et al., 1998; Tsai et al., 1998, 2001; Watanabe et al., 1999, 2000). In addition, all three genomes encode putative multi-copper oxidases that are clustered with the PKS genes and distantly related to Aspergillus fumigatus brown2/Arp2 and F. graminearum gip1 pigment pathway-associated genes (Fig. 3). The third gene in the putative pigment clusters shows amino acid sequence similarity to the recently characterized protein product of F. graminearum aurZ, which has been shown to convert the compound YWA1 to nor-rubofusarin (Frandsen et al., 2011). None of the Trichoderma genomes has genes that are related to known hydroxynaphthalene (HN) reductases, which catalyse the conversion of 1,3,6,8-tetrahydroxynaphthalene to scytalone and the reduction of 1,3,8-trihydroxynaphthalene to vermelone. It is therefore intriguing that both the T. atroviride and the T. virens genomes contain
genes, not clustered with the putative pigment PKS genes, which are predicted to encode scytalone dehydratases.

### Evolution of the PKS genes

In order to examine whether the *Trichoderma* PKS-encoding genes would indeed be products of horizontal gene transfer (HGT) or be subject to other evolutionary mechanisms, we first used Tajima’s D test (Tajima, 1989). To this end, we grouped the *Trichoderma* PKS-encoding genes according to their position in the phylogenetic tree into five datasets: (a) non-reducing PKS clades I and II; (b) non-reducing PKS clade III; (c) reducing PKS clade I and II (lovastatin-type); (d) reducing PKS clade III (fumonisin-type); and (e) a small clade that occupied a basal position to the reducing PKS clades. Tajima’s D was calculated with a

<table>
<thead>
<tr>
<th>Table 1. <em>Trichoderma</em> PKS orthologous groups</th>
</tr>
</thead>
<tbody>
<tr>
<td>Orthologous group or singlet designation</td>
</tr>
<tr>
<td>PKS group 1</td>
</tr>
<tr>
<td>PKS group 2</td>
</tr>
<tr>
<td>PKS group 3</td>
</tr>
<tr>
<td>PKS group 4</td>
</tr>
<tr>
<td>PKS group 5</td>
</tr>
<tr>
<td>PKS group 6</td>
</tr>
<tr>
<td>PKS group 7</td>
</tr>
<tr>
<td>PKS group 8</td>
</tr>
<tr>
<td>PKS group 9</td>
</tr>
<tr>
<td>PKS group 10</td>
</tr>
<tr>
<td>PKS group 11</td>
</tr>
</tbody>
</table>
sliding-window approach over the length of the KS domain-encoding portion of the gene. The data are shown in Fig. 4: Tajima’s D was between 0 and \( -0.15 \) throughout in the genes in clades (a) to (d), and only the basal branch (e) showed three areas of slightly positive values. \( P \) values were all \( >0.05 \), however, and we therefore conclude that the evolution of the PKS-encoding genes is neutral. We also used the \( K_s/K_a \) ratio as an alternative means to test whether the PKS-encoding genes are under any evolutionary selection. To this end, we calculated the \( K_s/K_a \) ratio for all pairwise combinations of genes. With all five groups, we only obtained ratios that were significantly less than 1, which indicates the operation of negative or purifying selection, the actual ratios varying between 0.02 and 0.6. When the \( K_s/K_a \) ratios were plotted on the nodes of the phylogenetic tree of the five PKS clades, the lowest values were observed in most (albeit not all) of the terminal clades, whereas the highest values were associated with internal nodes that represented ancient splitting events (Supplementary Figs S2–S5).

We further used the codon-based Z-test and the codon-based Fishers’s exact test to compare the relative abundance of synonymous and non-synonymous substitutions that have occurred in the gene sequences. The first test did not reject the null hypothesis of purifying selection, whereas that of neutral and positive selection was rejected; in Fisher’s exact test, a \( p \) of 1.0 also supported purifying selection, rather than positive selection. We therefore conclude that the PKS-encoding genes of *Trichoderma* evolve under purifying selection.

**CAI**

To estimate deviations in codon usage, the CAI was calculated for all five datasets. CAI is a measure of the relative adaptiveness of the codon usage of a gene towards the codon usage of highly expressed genes for that organism: the higher the index value, the greater the codon usage bias (Sharp & Li, 1987). CAI was previously determined for 100 orthologous and syntenic genes in *T. reesei*, *T. virens* and *T. atroviride*, and the mean value was around 0.72 (Kubicek et al., 2011). We also checked CAI for the strongly expressed fungal chitinase gene CHI18-5 (Seidl et al., 2005) and arrived at values between 0.65 and 0.8. The corresponding values for the five PKS clades of this study were between 0.65 and 0.7 (Supplementary Table S2). Therefore, there is a moderate codon usage bias in

---

**Table 2. PKS singlets**

<table>
<thead>
<tr>
<th>Singlet</th>
<th>Protein ID</th>
<th>Clade</th>
</tr>
</thead>
<tbody>
<tr>
<td>PKS singlet 1</td>
<td>Ta 79482</td>
<td>Non-reducing fungal PKS clade III</td>
</tr>
<tr>
<td>PKS singlet 2</td>
<td>Ta 32458</td>
<td>Reducing PKS clade IV fumonisins</td>
</tr>
<tr>
<td>PKS singlet 3</td>
<td>Ta 52962</td>
<td>Reducing clade I: lovastatin/citrinin diketide</td>
</tr>
<tr>
<td>PKS singlet 4</td>
<td>Ta 85006</td>
<td>Reducing clade I: lovastatin/citrinin diketide</td>
</tr>
<tr>
<td>PKS singlet 5</td>
<td>Ta 134224</td>
<td>Reducing clade I: lovastatin/citrinin diketide</td>
</tr>
<tr>
<td>PKS singlet 6</td>
<td>Ta 51503</td>
<td>Reducing clade I: lovastatin/citrinin diketide</td>
</tr>
<tr>
<td>PKS singlet 7</td>
<td>Ta 51726</td>
<td>Non-reducing fungal PKS clade II–III</td>
</tr>
<tr>
<td>PKS singlet 8</td>
<td>Ta 80219</td>
<td>Non-reducing fungal PKS clade I–II</td>
</tr>
<tr>
<td>PKS singlet 9</td>
<td>Ta 45540</td>
<td>Fungal MSAS PKS</td>
</tr>
<tr>
<td>PKS singlet 10</td>
<td>Tr 73621</td>
<td>Non-reducing fungal PKS clade III</td>
</tr>
<tr>
<td>PKS singlet 11</td>
<td>Tr 73618</td>
<td>Reducing clade I: lovastatin/citrinin diketide</td>
</tr>
<tr>
<td>PKS singlet 12</td>
<td>Tv 53518</td>
<td>Reducing clade I: lovastatin/citrinin diketide</td>
</tr>
<tr>
<td>PKS singlet 13</td>
<td>Tv 53633</td>
<td>Reducing PKS clade IV: fumonisins</td>
</tr>
<tr>
<td>PKS singlet 14</td>
<td>Tv 70876</td>
<td>Reducing PKS clade IV: fumonisins</td>
</tr>
<tr>
<td>PKS singlet 15</td>
<td>Tv 62549</td>
<td>Reducing clade I: lovastatin/citrinin diketide</td>
</tr>
<tr>
<td>PKS singlet 16</td>
<td>Tv 64883</td>
<td>Reducing clade I: lovastatin/citrinin diketide</td>
</tr>
<tr>
<td>PKS singlet 17</td>
<td>Tv 20101</td>
<td>Non-reducing fungal PKS clade III</td>
</tr>
<tr>
<td>PKS singlet 18</td>
<td>Tv 39352</td>
<td>Non-reducing fungal PKS clade II–III</td>
</tr>
</tbody>
</table>

---

**Fig. 2.** Venn diagram illustrating distribution of orthologous *T. reesei*, *T. virens* and *T. atroviride* PKS-encoding genes.
the PKS-encoding genes, probably indicative of moderate levels of expression. However, the values are high enough to conclude that none of these genes was recently acquired as the result of a horizontal transfer event.

**DISCUSSION**

Phylogenetic analysis of protein family coding sequences is a valuable aid for gene annotation and hypothesis generation. We have performed such an analysis on PKS-encoding genes from three genome-sequenced *Trichoderma* species and those examined by Kroken et al. (2003). We find that the majority of PKS-encoding genes (29 out of a total of 47) fall into an orthologous group. It may be that *Trichoderma* as a genus has a limited catalogue of PKSs and that diversity in the respective metabolites may be low or is expanded at the level of ‘accessory enzymes’ versus the backbone synthesizing enzyme. However, as the number of sequenced *Trichoderma* genomes increases, the diversity of the PKS-encoding gene family within this genus will expand as well.

Our analysis has also allowed us to predict the PKS-encoding gene most likely responsible for the characteristic *Trichoderma* yellow-green pigment. Indeed, the value of this study and others like it is in the generation of hypotheses to be tested. Ultimately, our prediction must be validated by genetic and biochemical characterization of the *Trichoderma* pigment biosynthesis system. It is also interesting to note that there is synteny only at the level of the putative pigment biosynthetic cluster; the surrounding genes are not conserved. Most of the PKS-encoding genes present in *T. reesei* are also found in *T. virens* and *T. atroviride*, but about half of the genes remaining in the latter two are novel for the respective species, and similar to the case of the putative pigment cluster, form non-syntenic islands (C. P. Kubicek & S. E. Baker, unpublished data) in the genome, suggesting transfer by recombination.

Most of the *Trichoderma* PKS genes occur as orthologues in all three species, implying that several nearly identical polyketides are produced by *T. reesei*, *T. virens* and *T. atroviride*. In this regard, the almost complete lack of duplicated copies was remarkable: this could be due to repeat-induced point mutation (RIP) in *Trichoderma*, which occurs during sexual development and alters
nucleotides in both copies of duplicated genes and renders them nonfunctional. Evidence for the operation of RIP in *Trichoderma* has been presented (Kubicek et al., 2011). The operation of RIP would also be consistent with the much lower number of PKSs in *T. reesei* than in *T. atroviride* and *T. virens*, because *T. reesei* is almost always observed as a teleomorph while the other two species are frequently sampled as anamorphs. RIP may therefore also explain the – in comparison with other Pezizomycota – low number of PKS genes; indeed, only *Neospora crassa* is known to have a lower number (Galagan et al., 2003), for which strong RIP is known (Selker & Garrett, 1988; Selker et al., 1987a, b).

There is also a still-ongoing debate about whether or not HGT contributes or has contributed to the evolution of PKS genes. Modular PKS genes appear to be transferred horizontally between bacteria similarly to iterative PKS genes (Ridley et al., 2008). Kroken et al. (2003), however, demonstrated that there is little evidence for HGT with the fungal PKS genes (an exception being MSAS-encoding genes, which represent an HGT event from bacteria to fungi), and stressed the major importance of gene duplications eventually followed by gene loss. Our evolutionary analysis is consistent with the latter scenario. First, examination of the codon usage pattern showed that the codon bias is similar to that of other moderately expressed genes of *Trichoderma*. Further, the overall substitution ratios of the *Trichoderma* PKS ketoacyl synthase module revealed that it is under strong purifying selection. Similar findings have also been reported for PKS from bacteria (Zucko et al., 2011) and lichenized fungi (Muggia et al., 2008).

Our analysis lays the groundwork for future polyketide studies in *Trichoderma*. As more PKSs are linked to metabolite structures, the predictive value of this type of analysis will increase.

**ACKNOWLEDGEMENTS**

S. E. B. is funded by the DoE Office of the Biomass Program. N. M. R. is funded by an undergraduate research fellowship from Washington State University Pullman.

**REFERENCES**


Edited by: B. A. Horwitz