Phylogenomic and functional domain analysis of polyketide synthases in *Fusarium*

Daren W. Brown  
USDA-ARS, daren.brown@ars.usda.gov

Robert A. E. Butchko  
USDA-ARS

Scott E. Baker  
*Pacific Northwest National Laboratory*

Robert A. Proctor  
USDA-ARS

Follow this and additional works at: http://digitalcommons.unl.edu/usdaarsfacpub
Phylogenomic and functional domain analysis of polyketide synthases in Fusarium

Daren W. BROWNa,*, Robert A. E. BUTCHKOa, Scott E. BAKERb, Robert H. PROCTORa

aBacterial Foodborne Pathogens and Mycology Research, USDA-ARS-NCAUR, 1815 N. University St., Peoria, Illinois 61604, USA
bChemical and Biological Process Development Group, Pacific Northwest National Laboratory, Richland, Washington 99352, USA

ABSTRACT

Fusarium species are ubiquitous in nature, cause a range of plant diseases, and produce a variety of chemicals often referred to as secondary metabolites. Although some fungal secondary metabolites affect plant growth or protect plants from other fungi and bacteria, their presence in grain-based food and feed is more often associated with a variety of diseases in plants and in animals. Many of these structurally diverse metabolites are derived from a family of related enzymes called polyketide synthases (PKSs). A search of genomic sequence of Fusarium verticillioides, Fusarium graminearum, Fusarium oxysporum, and Fusarium solani identified a total of 58 PKS genes. To gain insight into how this gene family evolved and to guide future studies, we conducted phylogenomic and functional domain analyses. The resulting genealogy suggested that Fusarium PKSs represent 34 different groups responsible for synthesis of different core metabolites. The analyses indicate that variation in the Fusarium PKS gene family is due to gene duplication and loss events as well as enzyme gain-of-function due to the acquisition of new domains or of loss-of-function due to nucleotide mutations. Transcriptional analysis indicates that the 16 F. verticillioides PKS genes are expressed under a range of conditions, further evidence that they are functional genes that confer the ability to produce secondary metabolites.

Published by Elsevier Ltd on behalf of The British Mycological Society.

Introduction

A hallmark of fungi is their capacity to synthesize a wide variety of structurally diverse natural products (Fig 1) (Cole et al. 2003). The frequently associated biological activity of these metabolites has positively impacted the pharmaceutical industry (e.g. the cholesterol-lowering drug lovastatin and the antibiotic cephalosporin) and negatively impacted the agricultural industry (e.g. the mycotoxins fumonisins, zearalenone, and trichothecene). Natural products that are not required for life or for propagation are often called secondary metabolites (SMs) (Bennett & Bentley 1989; Keller et al. 2005). Consumption of mycotoxin-contaminated grain or grain-based food and feed has been associated with a variety of human and animal diseases. Economic losses caused by mycotoxin contamination of grain can be in the 100s of millions of dollars each year due to negative impacts on human and animal health and reduced value of grain (Glenn 2007; Morgavi & Riley 2007; Wu 2007). Ingesting contaminated food may have an immediate (toxic) or delayed impact. For example, fumonisin can cause leukoencephalomalacia in horses and has been epidemiologically associated with neural tube defects in newborns and esophageal cancer in adults (Sydenham et al. 1990.; Seefelder et al. 2003; Marasas et al. 2004; Voss et al. 2007).

Many fungal SMs are polyketides, including the mycotoxins fumonisins and zearalenone. Polyketides are typically

* Corresponding author. Tel.: +1 309 681 6230; fax: +1 309 681 6672.
E-mail address: daren.brown@ars.usda.gov
1878-6146/$ — see front matter Published by Elsevier Ltd on behalf of The British Mycological Society.

This document is a U.S. government work and is not subject to copyright in the United States.
synthesized by polyketide synthases (PKSs) from small carboxylic acid derivatives (acetyl-CoA and malonyl-CoA) in a manner similar to the synthesis of fatty acids by a fatty acid synthase (FAS) (Hopwood & Khosla 1992). Fungal PKSs contain five to eight functional domains that are fused into one large protein and are used iteratively (Fig 2). PKSs can be divided into two major groups based on their domain content: non-reducing PKSs (NR-PKSs) synthesize polyketides in which carbonyl groups are not reduced, and reducing PKSs (R-PKS) synthesize polyketides in which the carbonyl groups are partially or fully reduced. Polyketide synthesis is initiated by the transfer of a starter unit (typically acetyl-CoA) onto the phosphopantetheine group bound to the acyl-carrier protein (ACP) domain of the PKS by the acyltransferase (AT) domain of R-PKSs and by the starter unit AT (SAT) domain of NR-PKSs. Next, the ketoacyl synthase (KS) domain catalyzes transfer of the acetyl unit from the ACP to the KS and decarboxylative condensation with an elongation unit (typically malonyl-CoA). Prior to the condensation, an elongation unit is loaded onto the ACP domain by the AT domain of R-PKSs or the malonyltransferase domain (MAT) of NR-PKSs. R-PKSs generally remove the carbonyl group (i.e. β-keto) of the resulting four-carbon molecule via a succession of reactions: the keto reductase (KR) domain catalyzes reduction of the β-keto to a hydroxy; the dehydratase (DH) domain catalyzes reduction of the hydroxyl to an enoyl (carbon–carbon double bond); and the enoyl reductase (ER) domain catalyzes reduction of the enoyl to an alkane (carbon–carbon single bond). Because they generally lack KR, DH, and ER domains, NR-PKSs usually do not reduce the β-keto group. The cycle is initiated a second time with the condensation of the four-carbon intermediate with another malonyl-CoA. The total number of cycles (generally between two and nine), product cyclization, and release of the polyketide from the PKS are mediated by the KS, the product template (PT), and terminal thioesterase (TE) or reductase (R) domains (Crawford & Townsend 2010; Du & Lou 2010). The presence of carbonyl groups along the carbon backbone of NR-PKSs generally lead to metabolites with one or more aromatic rings, whereas R-PKSs typically give rise to metabolites with a reduced carbon chain. Additional, structural diversity of polyketides can arise from three other sources: (1) incomplete reduction of one or more β-keto groups, (2) additional modifying domains [e.g. methyltransferase (ME) domain or non-ribosomal peptide synthase module (NRPS)] in the PKS, and (3) other enzymes that catalyze structural modifications (e.g. oxygenations, reductions, acylations, or isomerizations) following synthesis of the polyketide.

Fig 1 – Diverse chemical structures of fungal polyketide-derived metabolites. The type of core enzyme involved in metabolite synthesis and the fungi associated with their synthesis are indicated. Fv = F. verticillioides; Fg = F. graminearum; Fo = F. oxysporum; Nh = F. solani; Asp = Aspergillus; Pen = Penicillium. NR-PKS–NRPS = NR-PKS with an NRPS module. The metabolites are made by the Fusarium species indicated while the two metabolites in the inset are examples of fungal derived pharmaceutical.
Studies examining Fusarium focus primarily on their plant pathogenic nature and their ability to contaminate grains and grain products with mycotoxins (Fig 1). Fusarium is a diverse group of fungi, and, in addition to causing significant crop disease, may live unnoticed in plants and soils as endophytes and saprophytes. The name Fusarium refers to the asexual phase (anamorph) of these fungi. Their sexual phase (teleomorph) is Gibberella. Nectria haematococca (synonym Haemato-nectria haematococca) is the teleomorphic name for the anamorph Fusarium solani, so named because like fusaria with a Gibberella teleomorph it produces multi-celled, spindle-shaped asexual spores (macroconidia). Nevertheless, fusaria with Gibberella and Nectria teleomorphs represent two phylogenetically distinct lineages within the fungal family Nectriaceae (order Hypocreales, phylum Ascomycota) (O’Donnell et al. 2008).

A recent comparative genomic study of Fusarium graminearum, Fusarium oxysporum, and Fusarium verticillioides identified 46 potential secondary metabolite biosynthesis gene clusters, of which 87 % include a PKS gene (Ma et al. 2010). The number of polyketide-derived metabolites (>15) potentially synthesized by any one of these species is significantly greater than the number of metabolites currently attributed to them [reviewed in (Desjardins & Proctor 2007)]. For example, to date, the function of five of the 16 F. verticillioides PKSs are known. FUM1 is involved in the synthesis of a fumonisin pigment (Proctor et al. 2007), BIK1 is involved in bikaverin synthesis (Brown et al. 2008) and FAB1 is involved in fusaric acid synthesis (Brown et al. submitted for publication). Phylogenetic analysis resolved the 58 Fusarium PKSs into four major clades and revealed a high degree of diversity with few closely related orthologues (Ma et al. 2010), results consistent with previous studies examining PKSs from multiple, diverse genera of the Ascomycota (Kroken et al. 2003; Baker et al. 2006).

In this study, we used previously reported (Ma et al. 2010) phylogenetic analysis of 58 Fusarium PKSs to: (1) examine the content of functional domains in the PKSs and (2) identify species-specific orthologues and predict chemical products of PKSs based on their relatedness to PKSs for which the corresponding SM has been identified. The identification of species-specific PKSs and orthologous groups of PKSs in two, three or all four Fusarium will help direct further studies aimed at characteristics of PKSs that are unique to one or a few or common to many Fusarium species (Table 1).

### Materials and methods

#### Search for PKSs

Fusarium verticillioides, Fusarium graminearum, and Fusarium oxysporum PKS genes were identified by Basic Local Alignment...
Search Tool (BLAST) analysis against all predicted proteins (Altschul et al. 1990) at the Broad Institute (http://www.broad institute.org/annotation/genome/fusarium_verticillioides/ MultiHome.html), and Fusarium solani PKS genes were identified at the Joint Genome Institute (JGI) (http://genome.jgi-psf.org/) websites as described (Ma et al. 2010).

Phylogenetic analysis

The phylogram representing the phylogenetic relationships between Fusarium PKSs (Fig 3) was constructed based on an alignment of amino acid (AA) sequences of the KS and AT/MAT domains from all 58 PKSs as described (Ma et al. 2010). AA sequence alignments were done with the ClustalW+ program in the Genetics Computer Group (GCC) analysis software package Version 11.1.3-UNIX, Accelys Inc., San Diego, CA, USA (Devereux et al. 1984) using the Blosum multiple-sequence alignment scoring matrix. Phylogenetic analyses were done using the GCGs PaupSearch and PAUP* version 4.0b10-Unix [Phylogenetic Analysis Using Parsimony (PAUP)]. Statistical support for branches within the phylogenetic trees was generated by bootstrap analysis with 1000 pseudoreplicates. Two-AA sequence comparisons were carried out using the DNAMAN software (Lynnon BioSoft, Vaudreuil, Quebec, Canada).

Expression analysis

Gene expression in wild-type Fusarium verticillioides strain FRC M-3125 was analyzed by Expressed Sequence Tags (ESTs) and by microarray under different developmental, nutritional, and pathogenic conditions. ESTs (transcripts) corresponding to the F. verticillioides PKS genes were retrieved from the F. verticillioides Gene Index database at the Dana-Farber Cancer Institute (http://compbio.dfci.harvard.edu/cgi/ cgi-bin/tgi/gimain.pl?gudb=f_verticill). The Gene Index is composed of 13 350 sequences derived from over 86 908 ESTs collected from ten F. verticillioides cDNA libraries (Brown et al. 2005). Libraries FvF, FvG, FvM, and FvO were generated from cultures grown onGYAM (0.24 M glucose, 0.05 % yeast extract, 8 mM L-asparagine, 5 mM malic acid, 1.7 mM NaCl, 4.4 mM K2HPO4, 2 mM MgSO4, and 8.8 mM CaCl2, pH 3.0) for 24 h (FvF), 48/72 h (FvM), 96 h (FvG) from M-3125 and 90 h (FvO) with the F. verticillioides fumonisins non-producing strain 57-7-7 (Desjardins et al. 1996). Library FvIV was generated from M-3125 grown on corn meal medium for 4 and 6 d, library FvM from M-3125 grown on excised maize seedling root and shoots and library FvO from germinating spores of M-3125 after 10 h in water extracts of maize seedlings (Brown et al. 2005).

Microarrays were designed and produced by Roche NimbleGen (Madison, WI, USA). Each sequence is represented on the array by between six and 12 unique 60-mer probes. Total RNA was extracted using the RNeasy Mini kit (Qiagen Inc., Valencia, CA, USA). Microarray hybridization, data acquisition, and initial analysis were conducted by Roche NimbleGen, Iceland. Data were normalized by robust multi-array average (Bolstad et al. 2003; Irizarry et al. 2003) and were compared using the Acuity 4.0 microarray analysis software package (Molecular Devices Corp, Sunnyvale, USA). A brief description of the ten different microarray experiments referred to in Fig 8 are as follows. In general, samples for RNA extraction were collected after growth on different substrates for different intervals of time, flash frozen by immersion in liquid nitrogen and stored at –80 °C prior to RNA extraction. The first two experiments were conducted on liquid media with shaking. In the GYAM experiment, samples were collected after 24, 48, 72, and 96 h of growth. (Gene Expression Omnibus (GEO) accession no. GSE16900 (Brown 2011)). In the potato dextrose broth (PDB), Difco, Detroit, MI, USA) experiment, samples were collected after 24 and 48 h of growth. The next two experiments were conducted on autoclaved maize kernels. In the En/Germ experiment, samples were collected after growth on either excised kernel endosperm or germ tissue (Shim et al. 2003). In the CC 48 experiment, samples were collected after growth on cracked corn for 48 h. The next three experiments were conducted on live maize ears. In the Ear 3–7 experiment, infected kernels were collected 3 and 7 d after inoculation. In the Ear Symptomatic experiment, infected kernels displaying significant disease symptoms were collected 21 d after inoculation. In the Live Seeds experiment, live maize seeds were inoculated and collected after 4 and 8 d. The next three experiments were conducted after growth in maize seedlings roots. In the SRIWV experiment, infected maize seedling roots were collected after 2, and 4 weeks of growth; in the SRIWV experiment, infected maize seedling roots were collected after 5, 8, 14, and 21 d of growth; and in SRIWWS experiment, infected maize seedling roots were also collected after 5, 8, 14, and 21 d of growth except that in this case, the seedlings were not watered between day 8 and day 13. PKS genes were assigned as ‘expressed’ if the expression level changed >2-fold over the course of one experiment or if expression changed >4-fold in one experiment relative to another.

Results and discussion

PKS genealogy

Phylogenomic analysis of PKSs can delineate distinct orthologous groups of PKSs shared by one or more species. Such delineation can provide insight into characteristics of PKSs that affect metabolite production and into biosynthetic potential of species. Previously, we surveyed the four Fusarium genome sequence databases and identified 58 PKS genes (Table 1) (Ma et al. 2010). Maximum parsimony analysis of alignments of the AA sequences of the combined KS and AT/MAT domains resolved Fusarium PKSs into two major clades that correspond to NR-PKSs and R-PKSs based on functional genealogy reported here is consistent with previously reported KS genealogies from multiple fungal genera (Kroken et al. 2003; Baker et al. 2006). The resolution of NR-PKSs into a single clade in the KS-AT/MAT tree supports the hypothesis that Fusarium NR-PKSs are monophyletic; i.e. they share a common ancestor (Fig 3). However, based on previous analyses with multiple fungal genera (Kroken et al. 2003; Baker et al. 2006), the putative ancestral NR-PKS existed before divergence of Fusarium from other fungi. In the Fusarium genealogy, the R-PKSs were resolved into three clades with significant bootstrap support
Fig 3 — Genealogy of the Fusarium PKSs inferred by maximum parsimony analysis of deduced AA sequences of the KS and AT/MAT domains. Major clades are demarked by horizontal lines to the right of the tree. Numbers below branches indicate percentage bootstrap support for each branch point based on 1000 pseudoreplicates. The FAS from Gallus gallus served as an outgroup. Designations beginning with FGSG, FOXG, and FVEG correspond to the F. graminearum, F. oxysporum, and F. verticillioides, respectively, gene/protein models from the Broad Institute’s Fusarium Comparative Database. Designations beginning with JGI correspond to F. solani gene/protein models from JGI’s genome sequence database. The dashes in the Mutant Domain column under ME and ER heading refer to the presence of a non-functional ME or ER domain, respectively, in the predicted PKSs. The carboxy-terminal domains (TE, R, and ACP) are indicated in the Terminal column for the NR-PKS clade. The presence of an additional NRPS or CAT (carnitine acyltransferase) domain in R-PKSs is indicated by a plus sign. When known, the secondary metabolites corresponding to a PKS are listed. Metabolite names that are italicized and in bold text are predicted products.
Orthologous PKSs

The phylogenetic analysis of KS-AT/MAT domains also indicated the presence of closely related orthologous PKSs in two or more of the species examined (Figs 3 and 4). For example, the set of four closely related PKSs (with 100 % bootstrap support and identical functional domains) FGSG_09182, FVEG_03695, FOXG_05816, and JGI_101778 includes two required for synthesis of the violet perithecial pigment in Fusarium verticillioides and Fusarium graminearum (Gaffoor et al. 2005; Proctor et al. 2007) and fusarubin in F. verticillioides (Busman and Proctor, unpublished observations). Likewise, the set of three orthologues (FGSG_07798, FVEG_11086, and JGL_70660) include two required for fusarin production in F. verticillioides and F. graminearum (Gaffoor et al. 2005; Brown et al. submitted for publication). Although the closely related PKS orthologues are expected to synthesize the same polyketide, the structure and, by extension, ecological function of the SM end-product may differ in each fungus. For example, the putative gene cluster associated with the PKS gene required for the violet perithecial pigment consists of the same six genes in F. graminearum, Fusarium oxysporum, and F. verticillioides (Gaffoor et al. 2005; Proctor et al. 2007; Brown et al. submitted for publication). However, the putative cluster in Fusarium solani includes the PKS gene (JGI_101778) and only two of the other five genes present in the Fusarium species (Proctor et al. 2007). Thus, any putative product of the putative F. solani cluster is likely to be different than that of the clusters in the Fusarium species. In addition, a red pigment rather than a violet one accumulates in the perithecial walls of F. solani, and a different PKS (JGI_33672) is required for its synthesis (Graziani et al. 2004).

We identified three PKS genes with closely related orthologues in three of four species examined. The function of only one of these PKSs has been determined; the F. graminearum gene FGSG_07798 and its F. verticillioides orthologue (FVEG_11086) are required for production of fusarin mycotoxins (Song et al. 2004; Gaffoor et al. 2005; Brown et al. submitted for publication). Again, the genes flanking the fusarin PKS in F. graminearum and F. verticillioides (Brown et al. submitted for publication) do not exactly match the sets of genes flanking the F. solani orthologue JGI_70660, suggesting that once the polyketide is synthesized it undergoes different chemical modifications in F. solani versus F. verticillioides and F. graminearum. Because of the high degree of similarity shared by FOXG_15886 and FOXG_14850 (97 % identity) and the 100 % bootstrap support shared between this set and FVEG_13715 and JGI_106233, these four PKSs are considered an orthologous set of three PKSs (i.e. a pair from F. oxysporum and one each from F. verticillioides and F. solani).

We also identified eight PKS orthologues that occur only in two of the species examined. Six of these orthologous PKSs occur in F. verticillioides and F. oxysporum, which are the most closely related species included in this analysis (Ma et al. 2010). Overall, 12 of the PKSs in these two fungi are closely related orthologues, implying that most of the polyketides produced by these species are structurally very similar. In contrast, almost half of the PKSs from F. graminearum and F. solani are unique to each fungus (Fig 4D). The last two orthologous pairs are shared by F. oxysporum and F. graminearum and F. oxysporum and F. solani (Fig 4C) and presumably reflects a shared function unique to these fungi. In contrast, the lack of a shared pair of orthologous PKSs by F. verticillioides and F. graminearum, both pathogens of maize, suggests that a common PKS product may not be critical for pathogenesis on maize.

The phylogenetic analysis has resolved Fusarium PKSs into 14 clades of two (8), three (3), and four (3) likely orthologous PKSs (adjacent PKSs with branch points of 100 % bootstrap support) (Figs 3 and 4A–C) and 20 PKSs that lack closely related orthologues among the species examined (Fig 4D).
Thus, the analysis suggests that there are 34 distinct groups of PKSs among the four species. Percent identity between putative orthologues ranges from 31% to 94%. In some cases, orthologous PKSs required for synthesis of the same metabolites have been characterized in multiple Fusarium species (e.g., fusarin PKS; 64–71%). Based on the percent identity of these characterized putative orthologues, it is likely that the 34 groups of PKSs are responsible for synthesis of at least 34 structurally distinct groups of polyketides, only seven of which would be produced by both Fusarium and Nectria.

**Gene duplication**

Two *Fusarium oxysporum* PKSs, FOXG_14850, and FOXG_15886, appear to have arisen by a gene duplication event. The PKS coding sequences, as well as flanking sequences (2.5 kb upstream and 3.2 kb downstream), share 98% identity for a total duplicated genomic segment of 14.5 kb. The entire segment may represent a three-gene cluster as the upstream and downstream sequences are each predicted to encode a different methyltransferase (FOXG_14851 and FOXG_15882 downstream and FOXG_14849 and FOXG_15887 upstream from each PKS respectively), enzymes that are often involved in secondary metabolism. A 767-bp insertion just upstream of the predicted ATG of FOXG_14850 suggests that these two PKSs are expressed differently, although we have no evidence supporting this possibility. The lack of apparent PKS gene duplication events in the other three species could be due to repeat-induced point mutation (RIP). RIP occurs during sexual development and has been described in *Fusarium graminearum* and *Fusarium verticillioides* (Miller et al. 2003; Ansari et al. 2008). The putative non-functional or remnant ME domains had no or very limited similarity to pfam08242 (average E-value = 1 × 10⁻¹⁵) to the conserved ME domain, pfam0842, described in the National Center for Biotechnology Information (NCBI) Conserved Domain Database (CDD) (Marchler-Bauer et al. 2005). The putative non-functional or remnant ME domains of the four chemical products of the Fusarium NR-PKSs match the predicted functionality of their ME domain. The chemical structures of fusaric acid and zearalenone do not include any candidate ME-derived methyl groups, and the ME domains in the fusaric acid and zearalenone R-PKSs are likely non-functional based on the presence of poorly conserved ME motifs. In contrast, the chemical structures of fumonisins and fusarins include methyl groups (Figs 1), and the ME domains in the R-PKSs required for their production are predicted to be functional based on the presence of well-conserved ME motifs. A comparison of the putative non-functional ME domain of the fusaric acid PKS (FVEG_12523) with the functional domain of the fumonisin PKS (FVEG_00316) indicates that the two regions lack any sequence identity, which suggests that the functions of these regions are not the same.

**Non-functional R-PKS domains**

Examination of each domain from each PKS, by BLAST analysis and AA alignments, indicate that all domains present in the 11 NR-PKSs are functional. In contrast, 16 ME domains and 11 ER domains in the 47 R-PKSs are likely to be non-functional. Alignment of all ME domains, spanning approximately 120 AAs, identified missing AAs and AA differences within one, two, three or all four of the AA motifs critical for ME function (Miller et al. 2003; Ansari et al. 2008) in the 16 PKSs. Figs 5A and B are representative alignments of the putative functional and non-functional ME domains from the *Fusarium verticillioides* NR-PKSs. The functional ME domains contain four well-conserved motifs (Fig 5A) while comparable regions in the putative non-functional ME domains are very poorly conserved (Fig 5B) (Miller et al. 2003; Ansari et al. 2008). The dramatic differences in AA sequence emphasize the conclusion drawn from the phylogenetic analysis that the loss-of-function occurred on multiple occasions over time. BLASTp analysis of each ME domain provided additional evidence supporting whether or not the domains are functional. The functional ME domains shared significant deduced AA identity (average E-value = 4 × 10⁻⁶). It is important to note that the structure of the four chemical products of the *Fusarium graminearum* NR-PKSs match the predicted functionality of their ME domain. The chemical structures of fumonisins and fusarins share methyl groups (Figs 1), and the ME domains in the R-PKSs required for their production are predicted to be functional based on the presence of well-conserved ME motifs. A comparison of the putative non-functional ME domain of the fusaric acid PKS (FVEG_12523) with the functional domain of the fumonisin PKS (FVEG_00316) indicate that the two regions lack any sequence identity, which suggests that the functions of these regions are not the same.

---

**Fig 5** – Alignment of ME domains from *F. verticillioides* R-PKSs. (A) Putative functional ME domains. The four motifs previously identified in small molecule ME sequences are enclosed in boxes. (B) Putative non-functional ME domains. Conserved AAs are in bold text. (C) FVEG CON list the conserved AA residues where a capital letter indicates complete conservation and a small case letter indicates significant conservation. FGSG_04588 indicates the AAs corresponding to the ME domain present at the carboxy-terminal end of the *F. graminearum* R-PKS FVEG_04587. Conserved AAs are in bold text.
Alignment of all putative ER domains, which span approximately 230 AAs, also identified numerous gaps and differences in AA sequences among PKSs. Fig 6A depicts the AAs critical to ER function associated with binding reduced nicotine adenine dinucleotide phosphate (NADPH) (Song et al. 2004). This domain is well-conserved in 36 putative functional ER domains (Fig 6B) and poorly conserved in 11 putative non-functional ER domains (Fig 6C). BLASTp analysis of the putative functional ER domains indicated that they are significantly similar (average E-value = 1 × 10⁻⁴⁵) to the conserved ER domain, cd05195, described in the NCBI CDD. In contrast, the putative non-functional ER domains were markedly less similar (average E-value = 4 × 10⁻⁸) to cd05195. The loss of ER function is also reflected in the percent identity shared between non-functional and functional domains. The putative non-functional ER domain of the fusarin PKS (FVEG_11086) shares 24.6 % identity with the functional ER domain of the fumonisin PKS (FVEG_00316), whereas the fumonisin PKS ER domain shares 45.2 % identity with the functional ER domain of the fusaric acid PKS (FVEG_12523). The functionality of an ER domain cannot be predicted from a visual inspection of the polyketide product (Fig 1). Although the structure of lovastatin would suggest an ER is involved, the NR-PKS ER domain was found to be non-functional (Kennedy et al. 1999). The ER activity involved in lovastatin synthesis is coded for by a different gene located adjacent to the PKS (Kennedy et al. 1999). In contrast, the structure of fusarins suggests that their synthesis does not require ER activity, and this is reflected in the lack of a functional ER domain in the fusarin PKSs according to the current analysis. These results are consistent with the findings of Song et al. 2004.

The distribution of non-functional ME and ER domains among Fusarium R-PKSs is significantly different. Non-functional ME domains are present in all three major clades of R-PKSs (Clades I–III), while non-functional ER domains occur only in R-PKS Clade II. In the case of the non-functional ME domains, their varied location within the tree and the pattern of AA insertions and deletions (indels) suggest that the initial DNA mutation leading to loss of ME function occurred multiple times during the evolution of Fusarium R-PKSs. JGI_98127, FGSG_08795, FVEG_01914, and FOXG_03051 are closely related orthologous R-PKSs that occur in the four species examined here and have putative non-functional ME domains. In the case of the PKSs with non-functional ER domains, all are located in a single clade (Clade II) suggesting that the loss-of-function may have been due to a single event, tracing back to a single PKS progenitor.

**Variability in NR-PKS domain organization**

The organization of the first five domains (SAT, KS, MAT, PT, and ACP) is the same in all 11 Fusarium NR-PKSs, but the type and number of domains near the carboxy-terminal region varies (Fig 3). For example, five of the NR-PKSs include an additional ACP domain; four of these are the PKS orthologues involved in synthesis of fusarubin. An additional ACP domain does not appear to affect product synthesis as site-directed mutagenesis of a dual-ACP-containing PKS in the fungus Aspergillus revealed that each ACP domain was functional and able to operate independently of the other (Fujii et al. 2001). Here, the tree generated by phylogenetic analysis of Fusarium ACP domains from both dual and single-ACP PKSs was similar in topology with the tree generated with KS-AT/MAT from NR-PKSs (data not shown). In addition, the eight ACP domains from the four fusarubin PKSs orthologues form two clades; the more carboxy-terminal ACPs form one clade (with 99 % bootstrap support) and the more amino-terminal ACPs form the second clade (with 69 % bootstrap support). The positions of NR-PKSs relative to each other in the ACP-based tree were the same as in the KS-AT-based tree (Fig 3). Results from a previous phylogenomic analysis of PKSs from diverse fungal genera (Kroken et al. 2003; Baker et al. 2006) suggest that the dual-ACP NR-PKSs that currently exist in Fusarium and Aspergillus species did not arise independently in the two genera but instead are descended from a common ancestor.

The second difference near the carboxy terminus is the last domain. The C-terminal domain in all functionally characterized NR-PKSs is involved in determining chain length, final product cyclization and product release (Bailey et al. 2007; Du & Lou 2010). We found that the terminal domains for the 11 Fusarium NR-PKSs to be either a TE (6), an R (4); or an ME (1). TE and R domains are the most commonly described PKS terminal domains and their role in product release have been extensively studied (Du & Lou 2010). TE domains belong to the α/β-hydrolase superfamily and may catalyze the release of a linear product with a terminal carboxylic acid or a cyclized product (Du & Lou 2010). In contrast, R domains may catalyze release of a linear product with a terminal aldehyde or alcohol functionality via reduction of the acyl thioester, or the release of a product with one or more aromatic rings via a non-redox condensation reaction (Du & Lou 2010). Thus, the presence of either a TE or R terminal domain can provide information about the chemical structure of polyketide synthesized by an NR-PKS. BLASTp analysis of the terminal domain of FGSG_04588 indicated that it shares significant identity (E = 4.49 × 10⁻¹³) with the conserved ME domain (pfam08242) described in the NCBI CDD database. In addition to similar BLASTp scores, all four ME motifs are well-conserved (Fig 5C). Although PksCT, involved in citrinin synthesis in Monascus purpureus has a similar type ME domain, it also
includes a terminal TE domain (Shimizu et al. 2005). To our knowledge, no PKS with a C-terminal ME domain has been described in the literature and how or whether this terminal ME domain is involved in product release is not known.

**Acquisition of C-terminal domains**

Acquisition of an NRPS module and a putative carnitine/choline AT (CAT) domain at the carboxy-terminal end of R-PKSs appear to have occurred only once each, because NRPS domains occur only in R-PKS Clade II and CAT domains occur only in R-PKS Clade I (Fig 3). The NRPS module is a collection of four domains that span approximately 1400 AAs. The organization and putative functions of the four domains are completely conserved within all NRPS modules of the Fusarium R-PKSs. The four domains are responsible for condensation of the polyketide chain and an amino acid and then release of the resulting hybrid polyketide–amine molecule. Three PKSs (FGSG_10464, FVEG_10497, and FOXG_01189) within Clade II (the PKS–NRPS clade) lack an NRPS module. The positions of these three PKSs within Clade II suggest that the absence of the NRPS module can be explained by loss on at least two occasions: once in the branch leading to FGSG_10464 and once in the branch leading to FVEG_10497 and FOXG_01189. Loss of the NRPS module from FGSG_10464 appears to have occurred after divergence from two pairs of PKS–NRPSs in Fusarium verticillioides and Fusarium oxysporum (81 % bootstrap support) (Fig 3). In contrast, at what point the NRPS module was lost in the branch leading to FVEG_10497 and FOXG_01189 is less clear because of the poor bootstrap support for branches leading to the well supported FVEG_10497/FOXG_01189 branch. A previous analysis also indicates that NRPS modules have been lost repeatedly from fungal R-PKSs (Kroken et al. 2003).

CAT enzymes are generally involved in the transport of acetyl-CoA across mitochondria and peroxisomal membranes (van der Leij et al. 2000; Roze et al. 2011). They function by covalently attaching carnitine (L-3-hydroxy-4-aminobutyrobetaine) to the carboxylic acid portion of acetyl-CoA. A carnitine acylcarnitine translocase then moves the carnitine complex across a membrane after which another CAT reformats the two starting molecules (van Roermund et al. 1999; Roze et al. 2011). Aspergillus nidulans facC encodes a cytoplasmic CAT required for acetate utilization and acuL encodes a mitochondrial and peroxisomal CAT required for fatty acid utilization (Hynes et al. 2011). BLAST analysis of the four Fusarium genomic sequences with facC and acuL as query sequences identified two CAT encoding genes in each genome which we designated CFT1 and CFT2 for carnitine (fatty acid) acyltransferase. The deduced AA sequences of these genes are >90 % identical to each other and 68 % identical to FacC and AcuL respectively (Hynes et al. 2011). The deduced AA sequences of CFT1 and CFT2 from the same species share ~25 % identity.

R-PKS Clade I includes three PKSs with a terminal CAT domain (PKS–CAT): JGI_78513 and JGI_37058 from Fusarium solani and FGSG_05794 from Fusarium graminearum. Each CAT domain is ~600 AAs and has a high level of identity to AcuL by BLASTp analysis (183–243 bits, E = 9 × 10^{-46} to 2 × 10^{-63}) and to the conserved choline/carnitine O-acyltransferase, pfam00755, (E = 1 × 10^{-63} to 2 × 10^{-69}) described in the CDD. FOXG_10805 is likely a forth PKS–CAT in F. oxysporum, based on phylogenetic data (100 % bootstrap support) but, the nucleotide sequence coding for the CAT domain is missing from the genomic sequence data. A search of sequenced fungal genomes (at NCBI and JGI) identified 14 additional putative PKS–CATs from ten other fungi. Fungi containing this PKS group are present in five different classes from across the Ascomycota. It is interesting to note that most (83 %) of the fungi with PKS–CATs are pathogens of either plants (7), humans (1) or other fungi (2). Phylogenetic analysis of the 17 CAT domains from the PKS–CATs, CFT1 and CFT2 from all four fusaria, FacC and AcuL from A. nidulans, two CATs from Candida glabrata, and one CAT from Saccharomyces cerevisiae, with a human CAT (EAW87874) serving as an outgroup, indicate that the CAT domain from fungal PKSs form a monophyletic group (100 % bootstrap support) (Fig 7). Our analysis also suggests that the ancestral PKS–CAT acquired the CAT domain from a cytoplasmic FacC-like CAT after the latter diverged from the mitochondrial AcuL-like CAT.

To our knowledge, no polyketide synthesized by a PKS–CAT domain has been described in the literature. Although the function of the CAT domain likely involves the attachment of carnitine or carnitine-like moiety to the polyketide backbone generated by the PKS, the role of the moiety can only be speculated. One possibility is that the presence of the moiety enables the transport of the PKS product into a vesicle via a translocase. This possible mechanism to localize a secondary metabolite within an organelle is different than the mechanism described for synthesis of aflatoxin, where biosynthetic proteins are transported to vesicles designated aflatoxinsomes (Roze et al. 2011). If PKS–CATs do indeed function to attach a carnitine to the nascent polyketide, it will interesting to determine if they are also capable of removing the carnitine, a function that both CFT1 and CFT2 confer.

**Transcriptional analysis**

Expression of the 16 Fusarium verticillioides PKSs was examined by analysis of EST and microarray data. Analysis of ESTs in the F. verticillioides Gene Index revealed transcripts for ten of the 16 PKSs in seven of ten cDNA libraries (Fig 8). Previously, we reported the expression of nine PKSs in a single microarray experiment (Ma et al. 2010). The corresponding secondary metabolites for five of the 16 PKSs are known. Here, we found evidence for the expression of all 16 F. verticillioides PKSs in nine additional microarray studies (Fig 8). Although the pattern of expression of each gene is distinct, in some cases, the results have directed further studies. For example, the only PKS found to be expressed in live, symptomatic maize kernels during an ear rot infection assay was FVEG_13715. Preliminary studies found that FVEG_13715 insertional mutants are not affected in their ability to cause maize ear rot (Proctor, unpublished observations). Taken together, the expression data of the 16 PKSs is consistent with their being functional and involved in synthesis of polyketide product(s). A previous study also provided evidence for expression of 14 of 15 Fusarium graminearum PKSs (Gafour et al. 2005).

**Predicting PKS function**

Based on results from the phylogenetic analysis, coupled with previous published reports describing Fusarium metabolites, we predict the function of two Fusarium solani PKSs and
three *Fusarium oxysporum* PKSs (Fig 3). Two PKSs, *F. solani* JGI_101778 and *F. oxysporum* FOXG_05816, belong to a group of four orthologous PKSs where the function of the other two has been determined; *Fusarium verticillioides* FVEG_03695 and *Fusarium graminearum* FGSG_09182 are involved in the synthesis of the violet perithecial pigment (*Gaffoor et al. 2005; Proctor et al. 2007*). The recent discovery that FVEG_03695 is also required for fusarubin synthesis (Busman and Proctor, 2015).

*Fig 7* – Genealogy inferred by maximum parsimony analysis of deduced AA sequences of CAT domains from 17 filamentous fungal PKS–CATs and 13 CATs from *Fusarium*, *Aspergillus nidulans*, *Candida glabrata*, and *Saccharomyces cerevisiae*. The *Fusarium* PKS–CAT domains are indicated by an asterisk.
personal communication) coupled with the observation that F. solani may produce fusarubin (Ammar et al. 1979) suggest that JGI_101778 is involved in fusarubin synthesis. The third PKS, F. oxysporum FOXG_04757 is part of a pair of orthologues of which its partner, F. verticillioides FVEG_03379 is likely involved in bikaverin synthesis (Brown et al. 2008; Brown et al. submitted for publication). FVEG_03379 shares 94.4 % identity with the well characterized Fusarium fujikuroi bikaverin PKS (Wiemann et al. 2009). Coupled with the observation that F. oxysporum is reported to synthesize bikaverin (McInnes et al. 1976; Bell et al. 2003), we predict that FOXG_04757 is involved in bikaverin synthesis in F. oxysporum. Likewise, the fourth PKS, F. oxysporum FOXG_15248 is a closely related orthologue of F. verticillioides FVEG_11086, which is required for production of fusaric acid (Brown et al. submitted for publication). This, coupled with the observation that F. oxysporum is reported to produce fusaric acid (Bacon et al. 1996), we predict that FOXG_15248 is required for fusaric acid synthesis in F. oxysporum. The fifth PKS, F. solani JGI_70660 is a closely related orthologue of F. verticillioides FVEG_11086 and F. graminearum FGSG_07798, which are reported for fusarin production in these fungi (Gaffoor et al. 2005; Brown et al. submitted for publication). The conservation of sequences flanking FVEG_11086, FGSG_07798, and JGI_101778 in the respective fungi (Gaffoor et al. 2005; Brown et al. submitted for publication) further supports the hypothesis that JGI_101778 is involved in synthesis of fusarins or a structurally related metabolite(s) by F. solani. Demonstration of the predicted functions of these PKSs will require gene deletion and chemical analysis of the resulting mutants. The significant variability in polyketide structures observed, even for PKSs within the same clade, do not allow any further predictions to be made with confidence.

In a previous analysis of PKSs from a wide range of genera of the Ascomycota, the diversity and distribution of PKSs was attributed to gene duplication, divergence and differential gene loss rather than horizontal gene transfer (HGT) (Kroken et al. 2003). In contrast, the current analysis of Fusarium PKSs suggests that HGT could be responsible for significant PKS diversity in these fungi. If duplication was a major contributor to PKS diversity in Fusarium, one could expect to observe multiple examples of paralogous PKS genes in the Fusarium species examined. However, we observed evidence for only one pair of paralogous PKSs, FOXG_14850 and FOXG_15886 in F. oxysporum. In addition, of the 15 F. graminearum PKSs and 16 F. verticillioides PKSs, only five are closely related orthologues, and most of the rest are more closely related to PKSs from other fungal genera than they are to one another (Kroken et al. 2003; Baker et al. 2006). This indicates that most F. graminearum and F. verticillioides PKS genes began diverging from one another before Fusarium began diverging from other fungal genera. BLAST analyses also provide evidence that some F. oxysporum and F. solani PKSs are also more closely related to PKSs from other fungal genera than they are to other Fusarium PKSs. If divergence within Fusarium was a major contributor to PKS diversity in these fungi, we would expect that most Fusarium PKSs to be more similar to other Fusarium PKSs than they are to those in other fungi. However, this is not the case for most PKSs analyzed to date. HGT provides an explanation for the observed relationships of PKS from Fusarium and other fungi. An alternative explanation is that most Fusarium PKSs were present in the ancestral fusarium. If this were the case, the ancestral Fusarium would have had ~34 different PKSs corresponding to the 34 groups of PKSs presented in Fig 4. However, these species represent only a portion of the genetic diversity within Fusarium and, therefore, probably only a portion of the PKS diversity in these fungi. Thus, the ancestral Fusarium would have had to have >34 PKSs. Given that the Fusarium species examined have only 12–16 PKSs, it seems unlikely that the genome of their common ancestor included at least 34 PKS genes. Although it is a complicated explanation for the diversity of PKSs in Fusarium, HGT is consistent with existing phylogenetic data on PKSs and does not require that
the common ancestor was fundamentally different from extant species with respect to the number of PKS genes in its genome. The potential of HGT contributing to diversity of PKS genes in *Fusarium* is also consistent with two other observations: (1) multiple examples of HGT of SM biosynthetic gene clusters that include PKS gene between fungi (Khalid & Wolfe 2011; Slot & Rokas 2011); and (2) the genomes of *F. oxysporum* and *F. solani* include a significant percentage of DNA that is not closely related to DNA in other *Fusarium* species but instead is more closely related to DNA in more distantly related organisms (Coleman et al. 2009; Ma et al. 2010). The later observation indicates that *Fusarium* includes species that have the ability to acquire DNA from other organisms, and this in turn could have facilitated HGT of PKS genes into *Fusarium* from other fungi.

Functional characterization of fungal PKSs is challenging and requires significant resources. Despite the considerable increase in *F. graminearum* research efforts since the release of its genome sequence (Trail 2009), the function of only one *F. graminearum* PKS in synthesis of a polyketide has been demonstrated since 2005 (Sørensen et al. 2011). We have performed a phylogenetic and domain structure analysis of PKSs encoded by four *Fusarium* genomes with the purpose of laying a foundation on which to design future experiments to elucidate the biological functions of polyketides produced by these fungi.

**Acknowledgements**

We would like to thank Chris McGovern, Crystal Probyn, and Marcie Moore for technical assistance. Mention of trade names or commercial products in this article is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the US Department of Agriculture. USDA is an equal opportunity provider and employer.

**REFERENCES**


Brown, Butchko, Busman, Proctor RH. Identification of gene clusters associated with fusaric acid, fusarin, and perithecial pigment production in *Fusarium verticillioides*. Fungal Genetics and Biology, submitted for publication.


Morgavi DP, Riley RT, 2007. An historical overview of field disease outbreaks known or suspected to be caused by consumption of feeds contaminated with *Fusarium* toxins. *Animal Feed Science and Technology* 137: 201–212.


