Inventory of the Secondary Metabolite Biosynthetic Potential of Members within the Terminal Clade of the *Fusarium solani* Species Complex

Ambika Pokhrel [1,2] **and Jeffrey J. Coleman** [1,*]

1 Department of Entomology and Plant Pathology, Auburn University, Auburn, AL 36849, USA; azp0091@auburn.edu
2 The Donald Danforth Plant Science Center, St. Louis, MO 63132, USA
* Correspondence: jjcoleman@auburn.edu; Tel.: +1-334-844-1971

Abstract: The *Fusarium solani* species complex (FSSC) constitutes at least 77 phylogenetically distinct species including several agriculturally important and clinically relevant opportunistic pathogens. As with other Fusaria, they have been well documented to produce many secondary metabolites—compounds that are not required for the fungus to grow or develop but may be beneficial to the organism. An analysis of ten genomes from fungi within the terminal clade (clade 3) of the FSSC revealed each genome encoded 35 (*F. cucurbitcola*) to 48 (*F. tenucristatum*) secondary metabolite biosynthetic gene clusters (BGCs). A total of seventy-four different BGCs were identified from the ten FSSC genomes including seven polyketide synthases (PKS), thirteen nonribosomal peptide synthetases (NRPS), two terpene synthase BGCs, and a single dimethylallytryptophan synthase (DMATS) BGC conserved in all the genomes. Some of the clusters that were shared included those responsible for producing naphthoquinones such as fusarubins, a red pigmented compound, squalestatin, and the siderophores malonichrome, ferricrocin, and triacetylfusarinine. Eight novel NRPS and five novel PKS BGCs were identified, while BGCs predicted to produce radicicol, gibberellin, and fusaoctaxin were identified, which have not previously described in members of the FSSC. The diversity of the secondary metabolite repertoire of the FSSC may contribute to the expansive host range of these fungi and their ability to colonize broad habitats.

Keywords: nonribosomal peptide synthetase (NRPS); polyketide synthase (PKS); terpene synthase/cyclase; siderophore; radicicol; gibberellin

1. Introduction

Many fungi are capable of synthesizing complex compounds with bioactive properties that are not essential for the organism. These compounds, termed secondary metabolites, can have diverse chemical properties and are usually produced by a cluster of genes within the fungal genome [1]. These biosynthetic gene clusters (BGCs) frequently contain a key gene encoding the enzyme responsible for the synthesis of the "core" or "backbone" unit of the compound and may include accessory genes that are responsible for further modification(s) to the compound. The two most common classes of key genes found in BGCs encode either polyketide synthases (PKSs) or nonribosomal peptide synthetases (NRPSs). PKS enzymes generate a compound from linking ketides into a chain and are further divided into non-reducing and reducing PKS proteins, which are responsible for generating aromatic and fatty acid-like molecules, respectively. NRPSs catalyze the production of peptide-based compounds from individual amino acids. Other key enzymes in BGCs responsible for secondary metabolite biosynthesis include terpene synthase/cyclases, which produce compounds with the formula $(C_5H_8)_n$, dimethylallytryptophan synthases (DMATS) producing derivatives of tryptophan, as well as BGCs encoding a phosphoenolpyruvate phosphomutase responsible for producing phosphonate compounds. The

most prolific fungal producers of secondary metabolites are members of the Ascomycota, which produce well-known secondary metabolites such as the medically relevant compounds penicillin and lovastatin and the mycotoxins aflatoxin and deoxynivalenol (also known as vomitoxin).

Members of the ascomycete genus *Fusarium* produce a myriad of secondary metabolites including several mycotoxins. As a genus, *Fusarium* is composed of hundreds of species separated into at least 23 phylogenetically distinct species complexes [2,3]. While the genus is estimated to have originated ~91 million years ago during the Cretaceous period [4], the group referred to as the *F. solani* species complex (FSSC) diverged from the rest of the Fusaria ~67 million years ago [4]. Currently, the FSSC is divided into at least 77 phylogenetically distinct species and are found worldwide [3,5]. This species complex is composed of three clades (clades 1–3) where the terminal clade 3 is the largest with at least 61 distinct species and is composed of many agriculturally important species as well as those associated with clinical infections [3,5–8].

Many secondary metabolites have been identified from members of the FSSC. Primarily these have been napthoquinones including fusarubin and a multitude of derivatives and similarly structured compounds (javanicin, solaniol, matricin, bikaverin, trichodermaol, etc.) [9,10]. Other confirmed secondary metabolites include citreoisocoumarin, cyclosporin, gibepyrone, lucilactaene, N-carbenzoxy-L-phenylalaninol, sansalvadmide, YCM1008A, as well as unknown metabolites including a red pigment [10,11]. Of these compounds, only the production of sansalvadmide (*PKS30*) and the red pigment (*PKS35* also referred to as *PKSN*) has been experimentally linked to specific BGCs in isolates from the FSSC [12,13].

Secondary metabolites play an important role in fungal pathogenicity on plants and animals. Additionally, many secondary metabolites have inhibitory activity against other microbes, and therefore may influence the soil microbiome through their production. As members of the FSSC are soil-borne plant pathogens, evaluation of their secondary metabolite biosynthetic potential is essential to better understand these pathogens. Analysis of ten FSSC genomes revealed BGCs that were shared between all the FSSC genomes while some were unique to a single genome. Despite the abundance of these BGCs in the FSSC, the biosynthetic product of a vast majority of these BGCs remains unknown.

2. Materials and Methods

2.1. Genomic Data of FSSC Isolates

The relevant information of the FSSC genomes included in this study are listed in Table 1. The genome sequence and annotation of seven FSSC isolates were obtained from Mycocosm at the Joint Genome Institute (JGI) and three genomes were obtained from the National Center for Biotechnology Information (NCBI). A revised version of the *F. vanettenii* 77-13-4 genome assembly and annotation was used in this study resulting in a discrepancy in the protein IDs of the PKS and NRPS proteins provided in Tables S2–S5 and protein IDs reported in previous studies [14]. The predicted number of BGCs in the newer version of the *F. vanettenii* 77-13-4 genome was different (total 39) compared to the previous version (total 36).

In addition to the genome data listed in Table 1, the PKS and NRPS reference protein sequences from other *Fusarium* spp. and fungal species were obtained from GeneBank and included in the phylogenetic analyses for comparison (Tables S2 and S3).

Table 1. List and details of FSSC genomes used in this study.

Species	Strain	Genome Size (Mbp)	Number of Proteins	Source/Reference
F. vanettenii	77-13-4	54.59	16,929	[15]
F. vanettenii	T23	61.89	18,783	[16]
F. tenuicristatum	NRRL 22470	61.51	17,800	[16]
F. mori	NRRL 22230	43.64	14,523	[16]
F. cucurbiticola	NRRL 22165	42.45	12,147	[16]
F. bataticola	NRRL 22400	50.41	16,777	[16]
F. solani	FS5	52.93	17,656	[17]
F. ambrosium	NRRL 20438	49.04	17,262	NCBI
F. euwallaceae	UCR1854	50.55	17,630	NCBI
F. floridanum	NRRL 62606	47.42	16,762	NCBI

2.2. Species Phylogeny of FSSC Isolates

The gene sequences of translation elongation factor (*TEF1*) and the subunits of RNA-dependent polymerase (*RPB1* and *RPB2*) for all the FSSC isolates were identified using BLAST+ [18]. A multiple gene sequence alignment was conducted in MEGA11 [19], and the gaps and non-conserved regions were trimmed using Gblocks [20]. Model selection and a maximum likelihood phylogeny with 1000 bootstrap replicates were constructed using MEGA11 [19].

2.3. Prediction of Secondary Metabolite Gene Clusters and Phylogenetic Analysis of PKS and NRPS Proteins

The secondary metabolite gene clusters in the FSSC genomes were predicted using the fungal version of antiSMASH 6.0 using the options KnownClusterBlast, ClusterBlast, and SubClusterBlast [21]. The number of biosynthetic gene clusters (BGCs) for each genome was calculated and the PKS and NRPS protein sequences from these predicted clusters were extracted. The PKS and NRPS proteins from both the previous and revised versions of the *F. vanettenii* 77-13-4 genomes were included in the analysis to compare the differences in annotation. The PKS and NRPS proteins were classified into various groups separately using a phylogeny-based approach. Already known PKS and NRPS reference protein sequences from various fungal genomes (Tables S2 and S3) were also included in the analysis to aid in the classification [14,22–24]. For the phylogenetic analysis, PKS and NRPS protein sequences from all FSSC isolates and reference protein sequences were aligned using MAFFT [25]. All the gaps and non-conserved regions were removed from the alignment using BMGE [26]. The best protein model selection was conducted using the program ProtTest 3.4 [27], and maximum likelihood phylogeny trees with 1000 bootstrap replicates were generated using RAxML 8.0 [28]. Visualization of the phylogenetic trees was completed in iTOL v6 [29]. The PKS and NRPS clades were classified based on the reference proteins.

2.4. Similarity Network and Phylogenetic Analysis of the Terpene Synthase/Cyclase, DMATS, and Phosphoenolpyruvate Phosphomutase

The similarity network analysis and the diversity of the BGCs responsible for the production of terpenes, dimethylallytryptophan derivatives, and phosphonate compounds were explored using the program BiG-SCAPE [30]. BiG-SCAPE is a genome mining tool that facilitates the fast and interactive analysis of BGCs from multiple genomes, develops similarity networks, and classifies the clusters into various gene cluster families (GCFs). All the reference BGCs from the MIBiG database [31] were included in the analysis to identify the similarities with known natural products.

In addition to developing the similarity networks of BGCs and classifying various GCFs, BiG-SCAPE elucidates the phylogenetic relationship within these families using a multi-locus phylogeny approach. The phylogenetic analysis of the terpenes, dimethylallytryptophan derivatives, and phosphonate-producing BGCs was accomplished using the

program BiG-SCAPE [30], which also generated the genetic organization figures for each of these BGCs. The phylogenetic tree files from the BiG-SCAPE analyses were visualized using MEGA [19] and arranged manually.

2.5. Analysis of the Genetic Organization of the Conserved and Unique Biosynthetic Gene Clusters with Known Products

The GeneBank files of the conserved BGCs from PKS and NRPS groups, as well as the unique BGCs: radicicol, gibberellin, and fusaoctaxin A, were extracted from the antiSMASH output. If the BGC was similar to any previously characterized cluster as determined by antiSMASH results, the corresponding GeneBank file was obtained from the MIBiG repository of known BGCs [31]. These homologous BGCs were compared and visualized using the program clinker [32], generating the gene cluster comparison figures.

3. Results

3.1. Phylogenetic Analysis of FSSC Isolates Used in This Study

To better understand the evolutionary relationship between the FSSC isolates used in this study, a phylogenetic analysis was generated using three phylogenetically informative loci including the coding genes of *TEF1*, *RPB1*, and *RPB2*. A maximum-likelihood phylogenetic tree using the concatenated three gene sequences allowed for the visualization of the relationship between the ten FSSC genomes (Figure 1). The organization of the FSSC isolates in the phylogenetic tree was in agreement with previous studies [3], placing the three FSSC isolates (*F. floridanum*, *F. euwallaceae*, and *F. ambrosium*) in the ambrosia *Fusarium* clade.
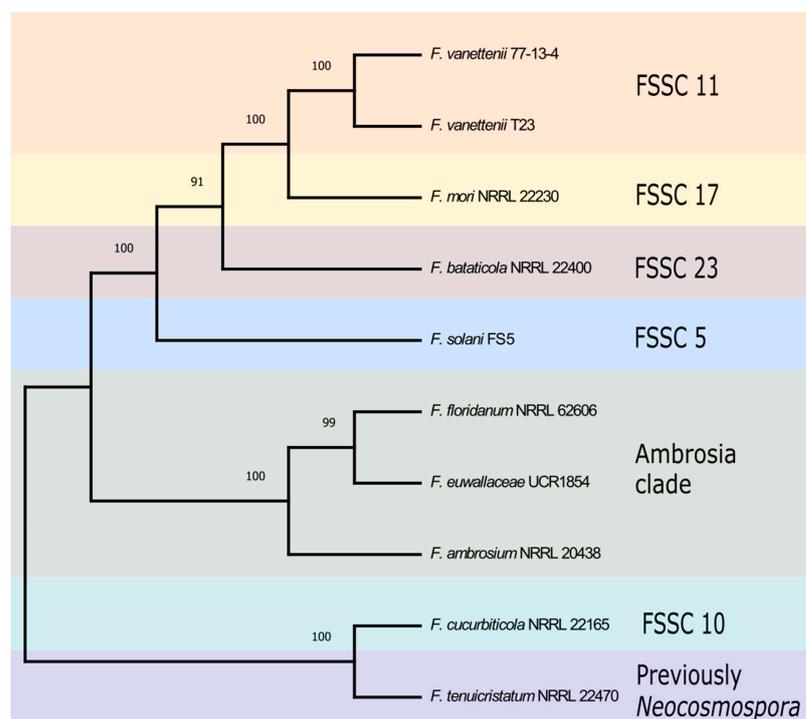
Figure 1. Phylogenetic analysis of ten FSSC isolates analyzed in this study. The maximum-likelihood phylogenetic tree of ten FSSC isolates from the terminal clade was constructed using the coding sequences of *TEF1*, *RPB1*, and *RPB2*. The robustness of the tree was assessed using 1000 bootstrap replicates.

3.2. Identification of Secondary Metabolite Biosynthetic Clusters

The secondary metabolite biosynthetic gene clusters (BGCs) for all ten FSSC genomes were identified using the fungal version of antiSMASH 6.0 [21]. The antiSMASH analysis revealed the FSSC genomes encoded an average of 41 BGCs, with the least in *F. cucurbitcola*

with 35 and the most in *F. tenuicristatum* with 48 (Table 2). The predicted number of BGCs in the two genomes of *F. vanettenii*, 77-13-4 and T23, were 39 and 42, respectively, indicating that diversity in the secondary metabolite potential exists within the same species (Table 2). The revised version of the *F. vanettenii* 77-13-4 genome identified 39 BGCs, 3 more than in the prior version. The FSSC isolates were predicted to produce PKS, NRPS, terpene synthase/cyclase, DMATS, phosphonate biosynthetic, and hybrid BGCs. The BGCs containing a PKS or NRPS gene as the main biosynthetic core gene were the most abundant gene clusters followed by terpene synthase/cyclase- containing clusters. Among these BGCs, only *F. bataticola* NRRL 22400 was predicted to encode a PKS–terpene hybrid cluster. Two ambrosia *Fusarium* clade genomes, *F. floridanum* NRRL 62606 and *F. euwallaceae* UCR 1854, were capable of producing an NRPS–terpene hybrid cluster and the *F. bataticola* and *F. solani* FS5 genomes were predicted to carry an NRPS–DMATS hybrid cluster.

Table 2. Number and type of predicted secondary metabolite producing BGCs in ten clade 3 FSSC genomes.

Species	Strain	Total BGC [a]	Total [b]					Hybrid BGCs					
			PKS	NRPS	Terpene	DMATS	Phosphonate	PKS–PKS Hybrid	NRPS–NRPS Hybrid	NRPS–PKS Hybrid	PKS–Terpene Hybrid	NRPS–Terpene Hybrid	NRPS–DMATS Hybrid
Fusarium vanettenii	77-13-4	39	14	18	6	2	1	1		1			
Fusarium vanettenii	T23	42	16	17	8	1	1	1					
Fusarium tenuicristatum	NRRL 22470	48	18	18	10	2	2	1		1			
Fusarium mori	NRRL 22230	41	15	17	9	1	1	1		1			
Fusarium cucurbiticola	NRRL 22165	35	10	18	4	2	1						
Fusarium bataticola	NRRL 22400	45	17	22	7	2	1	1	1		1		1
Fusarium solani	FS5	43	15	22	6	2	1			2			1
Fusarium ambrosium	NRRL 20438	40	14	19	5	2	1			1			
Fusarium euwallaceae	UCR1854	40	13	21	5	3	1		1	1		1	
Fusarium floridanum	NRRL 62606	38	14	16	6	2	1					1	

[a] Total BGC number is the number of BGCs in each genome and can be calculated as = (the sum of core biosynthetic genes from each BGC class—the number of hybrid BGCs). [b] The total is the number of the indicated core protein and does not necessarily reflect the number of BGCs with that type of core biosynthetic protein. Some NRPS–PKS hybrids consisted of a single core biosynthetic protein and were classified only under the PKS category. These BGCs were not included in the NRPS–PKS hybrid BGC list.

Of the putative BGCs, natural products for only a few were able to be predicted using the KnownClusterBlast function in antiSMASH (Table S1). The natural products squalestatin, a red pigment, and fusarubins/oxyjavanicin as well as several potential siderophores including fusarinine, ferricrocin, malonichrome, and metachelin, were predicted to be produced by all ten FSSC isolates analyzed in the study. Sansalvamide (NRPS30) and clavaric acid were predicted to be produced by all the FSSC isolates with the exception of those included in the ambrosia *Fusarium* clade (Table S1). An NRPS–PKS hybrid BGC in *F. solani* FS5 and *F. tenuicristatum* was predicted to be responsible for the synthesis of cyclosporin, a secondary metabolite that has previously been described from members of the FSSC [10,33].

### 3.3. Comparison of the Polyketide Synthase (PKS)-Containing Clusters

PKS reference proteins from *Fusarium* and other fungal species (Table S2; [14,22–24]) and the PKS proteins (including those in hybrid PKS BGCs) from our FSSC study group (Table S4) were included in the analysis for a total of 290 PKS amino acid sequences. The phylogenetic analysis of the PKS proteins revealed 27 different groups of PKS orthologues in the FSSC genomes used in this study (Figures 2,S1 and Table S4), and were represented by 23 previously known PKS proteins in other *Fusarium* spp. and 4 that were previously

undescribed (PKSA-D; Figure 2B). BGCs containing *PKS3*, *PKS7*, *PKS22*, *PKS32*, *PKS33*, *PKS35*, and the previously undescribed *PKSB* were conserved in all the FSSC genomes included in this study. Among the 27 PKS BGC groups, only the products of 10 were able to be predicted (Figure 2) based on antiSMASH results and previous studies [11,14]. The secondary metabolites fusarubins (PKS3), gibepyrone (PKS8), and the red pigment (PKS35) are the currently known conserved secondary metabolites from the FSSC [11]. While the core biosynthetic gene *PKS3* is moderately conserved across the ten FSSC isolates in this study, an analysis of the genetic organization of the entire *PKS3* BGC responsible for fusarubin/oxyjavanicin synthesis revealed that the accessory genes of the ambrosia FSSC genomes were different compared to the other FSSC isolates (Figure S2). Previously, the gene encoding a PKS responsible for the production of a red pigment (termed *PKSN/PKS35*) was characterized [13], and the genetic organization of the BGC encoding *PKS35* was conserved across all the FSSC isolates (Figure S3).
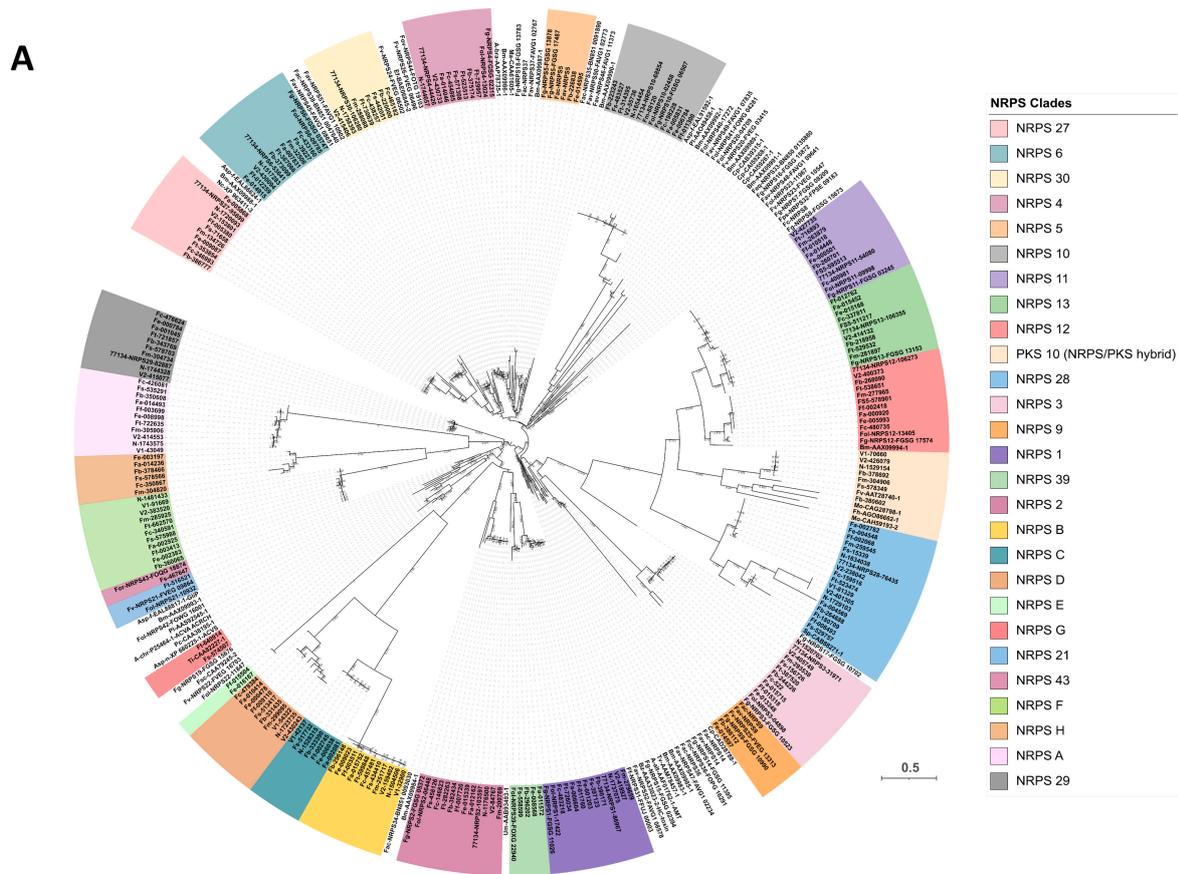
3.4. Comparison of the Nonribosomal Peptide Synthetase (NRPS)-Containing Clusters

A phylogenetic analysis of 332 NRPS proteins was conducted representing NRPS proteins identified from the antiSMASH analysis of the FSSC genomes (including the NRPSs in hybrid BGCs) and using NRPS reference proteins from other *Fusarium* and fungal species (Table S3) [23,24]. Twenty-six different clades of NRPS proteins were distinguished (Figure 3 and Table S5), of which eighteen were previously identified in other *Fusarium* spp. and eight (NRPSA-H) were previously undescribed NRPS proteins found in the FSSC genomes.

Of the 26 NRPSs in the FSSC genomes, 13 were conserved in all 10 genomes (*NRPS1*, *NRPS2*, *NRPS3*, *NRPS6*, *NRPS10*, *NRPS11*, *NRPS12*, *NRPS13*, *NRPS27*, *NRPS28*, *NRPSA*, *NRPSB*, and *NRPSF*; Figure 3). Conversely, *NRPS21* and *NRPS43* were found in only a single FSSC genome, *F. tenuicristatum* NRRL 22470 and *F. solani* FS5, respectively. Many of the conserved NRPS BGCs were predicted to be responsible for the biosynthesis of the siderophores malonichrome (*NRPS1*), ferricrocin (*NRPS2*), triacelyfusarinine (*NRPS6*), and metachelin (*NRPS27*) (Figures S4–S7). The only NRPS in the FSSC that has been experimentally confirmed is NRPS30, responsible for sansalvamide biosynthesis [12], and it was present in all FSSC genomes included in this study except for those in the ambrosia *Fusarium* clade (Figure 3). *NRPS43* is predicted to contribute to the biosynthesis of fumarylalanine, and a single BGC containing this NRPS was identified in the genome of *F. solani* FS5 (Figure 3). Two NRPS-encoding genes, *NRPS5* and *NRPS9*, were found to reside in a single BGC in the genomes of *F. bataticola* and *F. euwallaceae* and homologues of these NRPS genes have been characterized in *F. graminearum* and are known to produce the octapeptide fusaoctaxin A [34].
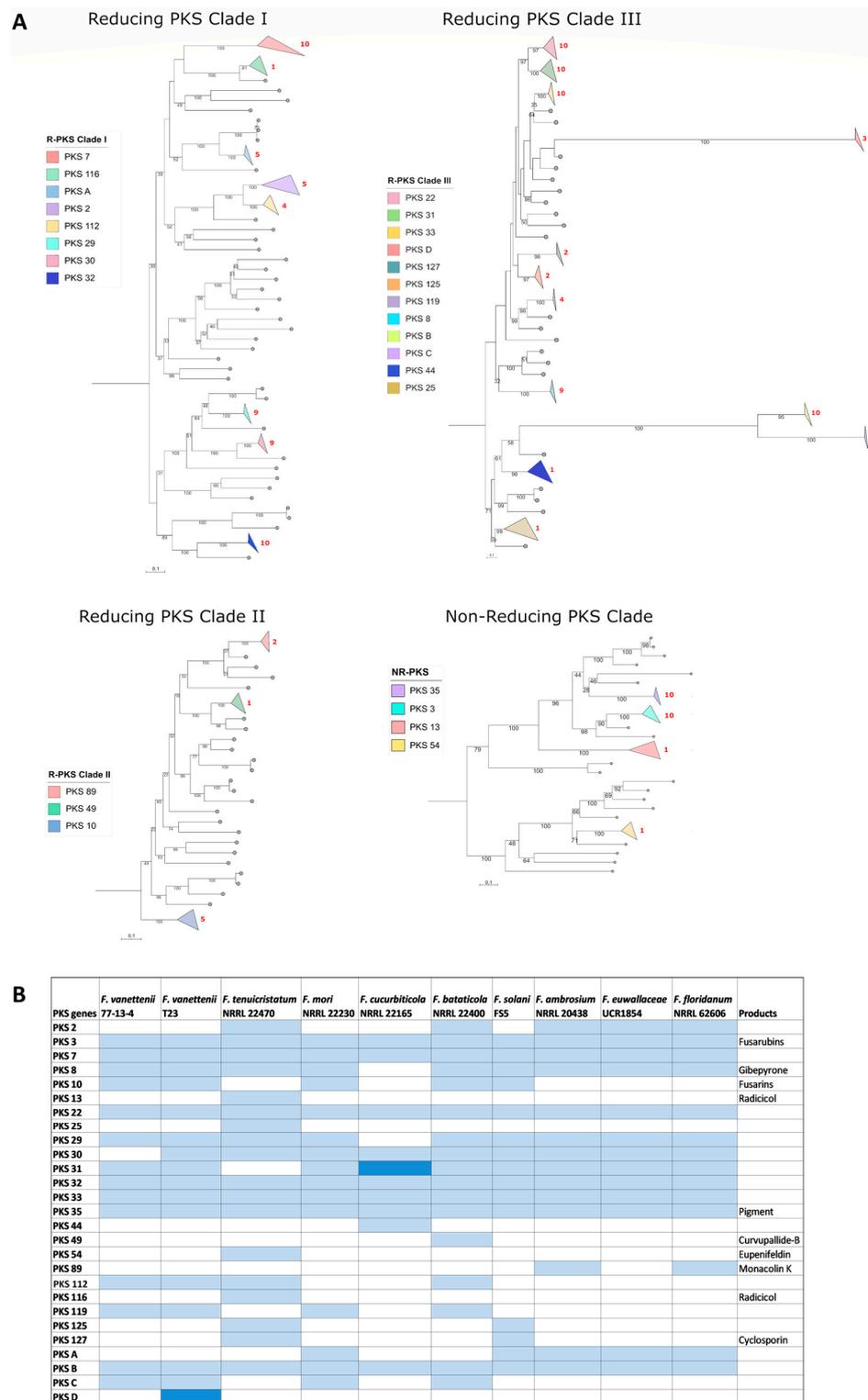
Figure 2. Overview of the distribution of PKS-encoding genes across FSSC genomes. (**A**) Maximum likelihood phylogenetic tree of 290 PKS proteins from FSSC genomes and reference PKS proteins from *Fusarium* and various other fungal species. The robustness of the tree was assessed using 1000 bootstrap replicates. The clades with PKS proteins from 10 FSSC genomes are highlighted in various colors in the phylogenetic tree. (**B**) The detailed distribution of PKS genes in FSSC genomes. Known products of the PKS-containing BGC are also provided. Light blue color indicates the presence of one core biosynthetic gene and the dark blue color indicates the presence of two or more core biosynthetic genes. Previously undescribed PKS genes are indicated with letters, i.e., PKS A-D.

Figure 3. Overview of the distribution of NRPS-encoding genes across FSSC genomes. (**A**) Maximum likelihood phylogenetic tree of 332 NRPS proteins from FSSC genomes and reference NRPS proteins from *Fusarium* and various other fungal species. The robustness of the tree was assessed using 1000 bootstrap replicates. The clades with NRPS proteins from 10 FSSC genomes are highlighted in various colors in the phylogenetic tree. (**B**) The detailed distribution of NRPS genes in FSSC genomes. Known products of the NRPS-containing BGC are also provided. Light blue color indicates the presence of one core biosynthetic gene and the dark blue color indicates the presence of two or more core biosynthetic genes. Previously undescribed NRPS genes are indicated with letters, i.e., NRPS A-H.

### 3.5. Comparison of the Terpene Synthase/Cyclase, DMATS, and Phosphonate-Producing BGCs

A total of 98 BGCs responsible for the production of terpenes, dimethylallytryptophan derivatives, and phosphonate compounds were predicted from the 10 FSSC genomes using antiSMASH and were further analyzed using BiG-SCAPE [30]. Of all these BGCs within these biosynthetic groups, only two terpene- synthase/cyclase-containing clusters and a single DMATS were conserved between all ten FSSC genomes (Figure 4). The multi-locus phylogenetic analysis of these GCFs (Figures S8–S10) revealed that their phylogenetic grouping is mostly in accordance with their species phylogeny (Figure 1).

A

Terpene GCF	*F. vanettenii* 77-13-4	*F. vanettenii* T23	*F. tenuicristatum* NRRL 22470	*F. mori* NRRL 22230	*F. cucurbiticola* NRRL 22165	*F. bataticola* NRRL 22400	*F. solani* FS5	*F. ambrosium* NRRL 20438	*F. euwallaceae* UCR1854	*F. floridanum* NRRL 62606	Products
T-GCF1	●	●	●	●	●	●	●	●	●	●	
T-GCF2	●	●	●	●		●	●	●	●	●	
T-GCF3	●	●	●	●		●	●	●	●	●	
T-GCF4	●	●	●	●	●	●	●	●	●	●	Squalestatin S1
T-GCF5	●	●	●	●	●	●		●	●	●	Clavaric acid
T-GCF6	●	●		●		●		●			Aspterric acid
T-GCF7	●	●	●				●				
T-GCF8		●●									
T-GCF9									●		
T-GCF10				●●							
T-GCF11				●							Gibberellin
T-GCF12		●									
T-GCF13			●								
T-GCF14			●								
T-GCF15			●								
T-GCF16								●			
T-GCF17										●	
T-GCF18					●						

B

Indole GCF	*F. vanettenii* 77-13-4	*F. vanettenii* T23	*F. tenuicristatum* NRRL 22470	*F. mori* NRRL 22230	*F. cucurbiticola* NRRL 22165	*F. bataticola* NRRL 22400	*F. solani* FS5	*F. ambrosium* NRRL 20438	*F. euwallaceae* UCR1854	*F. floridanum* NRRL 62606
I-GCF1	●	●	●	●	●	●	●	●	●	●
I-GCF2	●	●	●	●	●	●	●	●	●	●
I-GCF3	●		●		●					
I-GCF4						●				
I-GCF5									●	

C

Phosphonate GCF	*F. vanettenii* 77-13-4	*F. vanettenii* T23	*F. tenuicristatum* NRRL 22470	*F. mori* NRRL 22230	*F. cucurbiticola* NRRL 22165	*F. bataticola* NRRL 22400	*F. solani* FS5	*F. ambrosium* NRRL 20438	*F. euwallaceae* UCR1854	*F. floridanum* NRRL 62606
P-GCF1	●	●	●	●	●	●	●	●		
P-GCF2			●							
P-GCF3					●					

Figure 4. Distribution of terpene synthases/cyclases, DMATS, and BGCs responsible for the synthesis of phosphonate compounds across the FSSC genomes. Similarity network analysis of the terpene synthases/cyclases, DMATS, and phosphonate-generating BGCs was conducted to identify their distribution and conservation across the FSSC genomes. The product name is indicated alongside the gene cluster family (GCF) if predicted or known. GCF grouping of (**A**) terpene synthases/cyclases (T-GCF), (**B**) DMATS (I-GCF), and (**C**) phosphonate-producing BGCs (P-GCF) are shown across the FSSC genomes. Light blue color indicates the presence of one core biosynthetic gene and the dark blue color indicates the presence of two or more core biosynthetic genes.

The analysis of 67 terpene-synthase-/cyclase-containing BGCs revealed 18 different terpene synthase/cyclase gene cluster families (T-GCFs), including 10 singletons (Figure 4A). Of these 18 T-GCFs, a product of 5 of these terpene synthase/cyclase-containing clusters was predicted. Squalestatin was predicted to be produced by all ten FSSC isolates; its amino acid similarity and the synteny of the cluster indicates that it was conserved during vertical inheritance throughout the terminal clade of the species complex (Figures 4, 5A and S8). Aspterric acid was predicted to be produced by five FSSC isolates, while a BGC from *F.*

bataticola was a hybrid PKS–terpene cluster (Figures 4, 5B and S8). Seven FSSC genomes were predicted by antiSMASH to have the terpene synthase BGC responsible for the production of clavaric acid (T-GCF-6; Figures 4 and 5C).
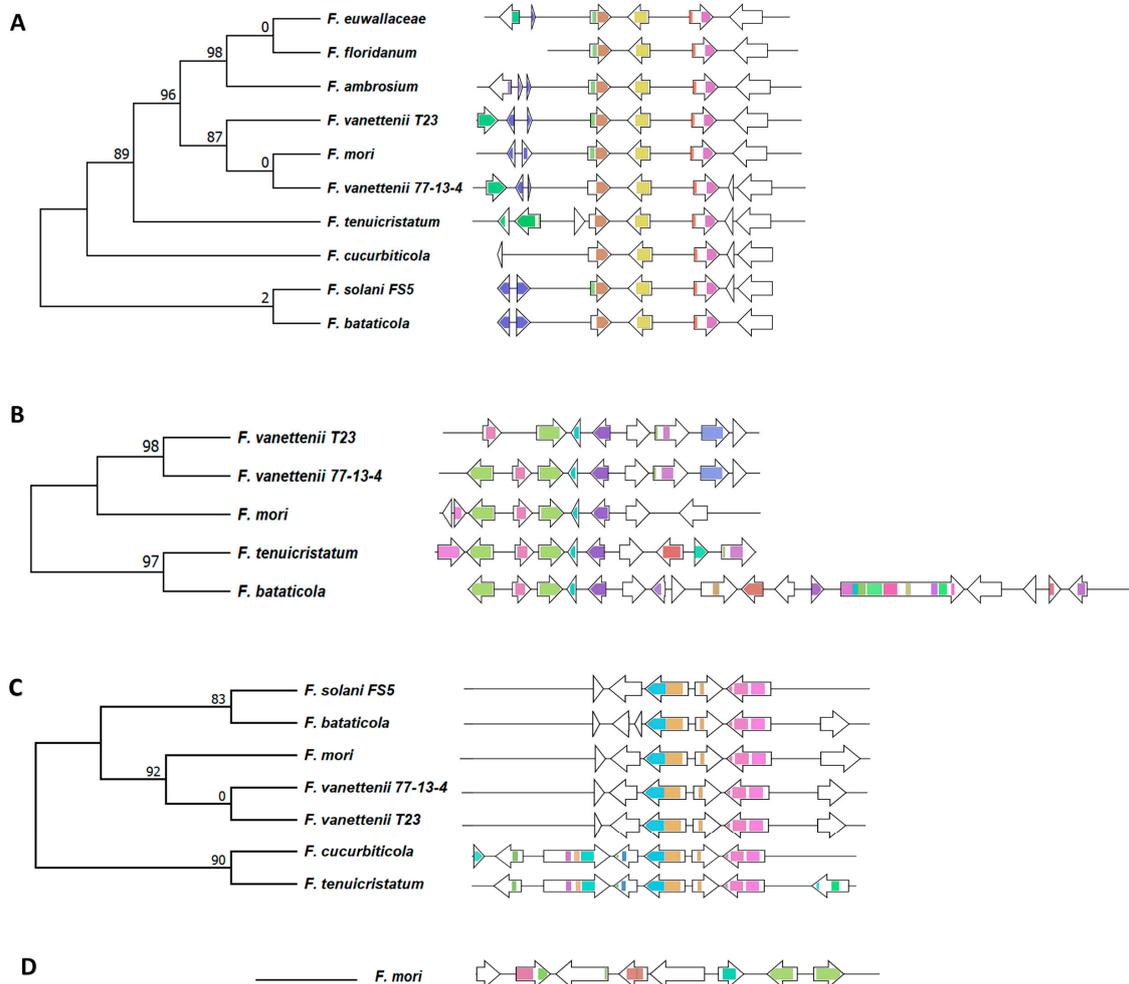

Figure 5. Analysis of the terpene GCFs with predicted or known products. (**A**) Squalestatin, (**B**) aspterric acid, (**C**) clavaric acid, and (**D**) gibberellin. The multi-locus phylogenetic analysis of the GCFs generated by BiG-SCAPE is shown along with the gene cluster organization within the FSSC genomes. Each arrow represents a different gene, and the color represents various functional domains.

Nineteen total BGCs containing a DMATS were identified across the ten FSSC genomes, and only one DMATS-encoding BGC was conserved between all FSSC genomes in the analysis (Figures 4B and S9). A single conserved BGC encoding a putative phosphoenolpyruvate phosphomutase was present in all FSSC genomes with the exception of *F. cucurbiticola* NRRL 22165, which had a unique phosphonate-producing BGC Figures 4C and S10).

3.6. Unique and Specific BGCs of Interest

Phylogenetic analysis indicated that the genome of *F. mori* carries a BGC that is similar to the BGC responsible for gibberellin biosynthesis in *F. fujikouri* (Figure 5D). Further analysis between these two gibberellin BGCs uncovered four homologous genes and the organization of the cluster was conserved (Figure 6A). The homologue encoding a geranylgeranyl diphosphate synthase may be responsible for the production of the precursor to gibberellin, geranylgeranyl diphosphate (GGDP), while the homologue to the CPS/KS-encoding enzyme is necessary for the cyclization of the final product.
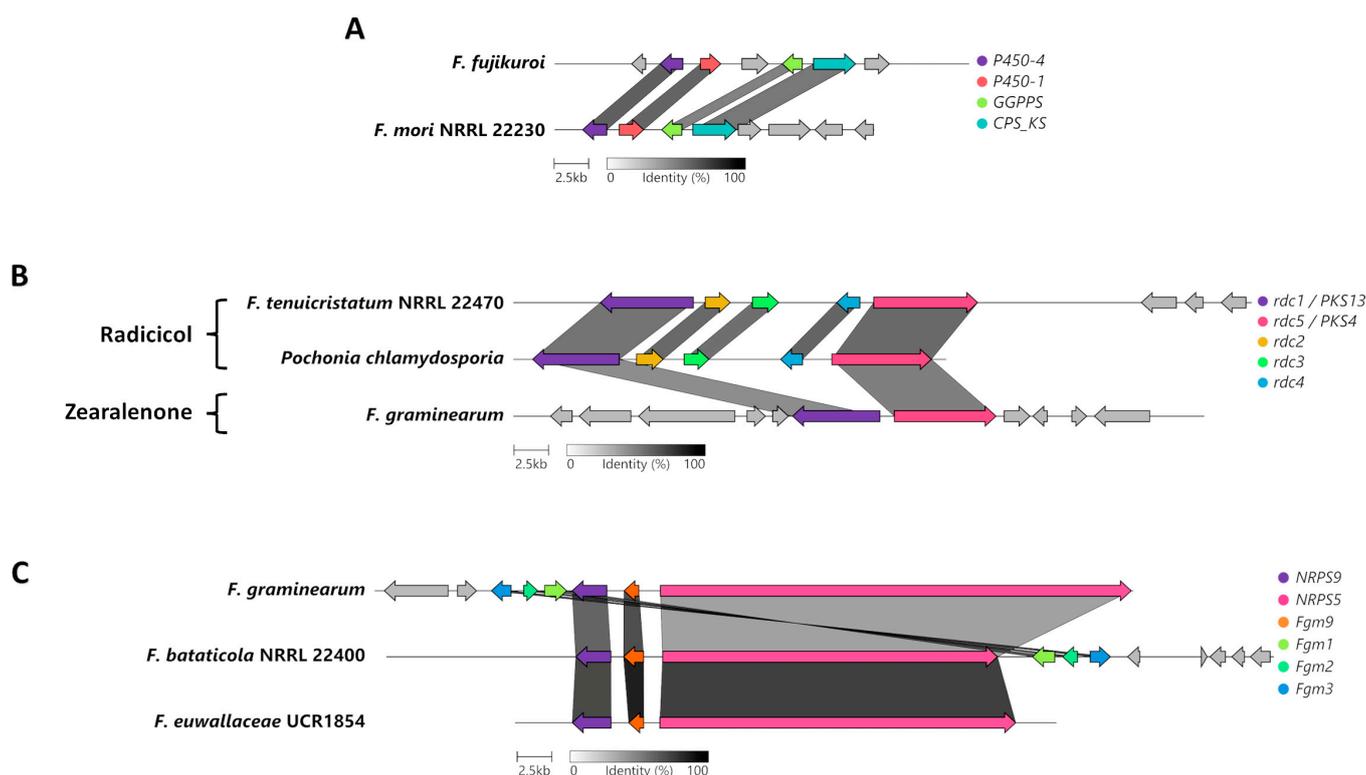
Figure 6. Genetic organization and conservation of the BGCs of interest. Conservation and biosynthesis of (**A**) gibberellin, (**B**) radicicol, and (**C**) fusaoctaxin A. Each color represents a different gene and are named based on the gene names of the originally described or characterized cluster.

The antiSMASH analysis revealed that the genome of *F. tenuicristatum* has a PKS BGC that is similar to the BGC responsible for radicicol biosynthesis in *Pochonia chlamydosporia* (Figure 6B). Further analysis between these two radicicol BGCs uncovered five orthologous genes, and the organization of this cluster was also conserved. Our study indicates that *PKS116/rdc5* and *PKS13/rdc1* in *F. tenuicristatum*n confer the ability to likely synthesize radicicol as opposed to zearalenone, a secondary metabolite from *F. graminearum* with a similar chemical structure.

A phylogenetic analysis of various NRPS proteins indicated the presence of *NRPS5* and *NRPS9* in *F. bataticola* and *F. euwallaceae* (Figure 3). These two NRPS proteins are known to be involved in the production of the octapeptide fusaoctaxin A in *F. graminearum* [34]. Further analysis between these three homologous BGCs for fusaoctaxin A uncovered conservation in the organization of this cluster, indicating that fusaoctaxin might potentially be produced by *F. bataticola* and *F. euwallaceae*.

4. Discussion

Overall, the FSSC isolates included in this analysis revealed that there was a diverse array of potential secondary metabolites that could be synthesized. Collectively, a total of 74 different BGCs (19 PKSs, 2 PKS–PKS hybrids, 21 NRPSs, 1 NRPS–NRPS hybrid, 17 terpene synthases/cyclases, 4 DMATS, 3 phosphonate-producing BGCs, 4 NRPS–PKS hybrids, 1 PKS–terpene hybrid, 1 NRPS–terpene hybrid, and 1 NRPS–DMATS hybrid) were identified in the 10 clade 3 FSSC genomes. The secondary metabolite BGC repertoire for each isolate was unique where no two genomes had the same secondary metabolite biosynthetic potential, even between the two *F. vanettenii* genomes. There were seven PKSs, thirteen NRPSs, two terpene synthases/cyclases, and one DMATS BGC common between all ten FSSC genomes included in this analysis. *PKS3*, *PKS7*, and *PKS8* are frequently found in most *Fusarium* genomes [11,14], and all 10 FSSC isolates in this study carried these BGCs. Conversely, *PKS32*, *PKS33*, and *PKS35* are primarily found in genomes of the FSSC and all

10 FSSC genomes had these BGCs. While the biosynthetic products of PKS32 and PKS33 are not known, PKS35 is responsible for the production of a red pigment [13]. The presence of these PKS BGCs in an unidentified *Fusarium* sp. could aid in their identification as a member of the FSSC; however, it should be noted that *PKS35* has recently been identified in closely related *Fusarium* species complexes such as *F. staphlyeae*, *F. decemcellulare*, and *F. dimerum* [14].

The production of a vast array of secondary metabolites likely contributes to the diverse environmental niche that members of the FSSC can be isolated from. For instance, many naphthoquinones have been isolated from members of the FSSC [3]. PKS3 is involved in the production of naphthoquinones such as fusarubin and oxyjavanicin, and could provide a competitive advantage to the FSSC isolates producing them as napthoquinones have antibacterial activity, especially against Gram-positive bacteria as well as inhibitory activity against protozoa and fungi [9,35].

Members of the FSSC are well-established plant pathogens and collectively have a broad host range, while some within the terminal clade of the FSSC are also associated with clinical infections resulting in fusariosis [3,5–8]. Some secondary metabolites have been demonstrated to be important virulence factors in fungi other than the FSSC [36]. The NRPS BGCs responsible for the synthesis of siderophores have been documented to be important for virulence for the closely related phytopathogen *F. graminearum* [37]. NRPS6 is conserved in ascomycetes and is responsible for the synthesis of the extracellular siderophore triaceylfusarinine, which is critical for virulence in several plant pathogenic fungi [38]. NRPS1 is responsible for the synthesis of another extracellular siderophore malonichrome, although it appears not to be as important for virulence as NRPS6, while NRPS2 is responsible for the production of the intracellular siderophore ferricrocin. In addition to these three siderophores, another might be produced by NRPS27, which is closely related to NRPS6; this may synthesize metachelins, which have been characterized in *Metarhizium* spp. [39].

In addition to siderophore production, two FSSC genomes included in this study also carried a BGC encoding *NRPS5* and *NRPS9*, which are responsible for fusaotaxin A production in *F. graminearum*. Fusaotaxin A is a virulence factor involved in the cell-to-cell invasion of wheat [34]; however, mutants lacking the cluster display virulence similar to wild-type when inoculated in a maize stalk rot assay, indicating that this secondary metabolite is host-specific [34]. The *NRPS5* homologues in the FSSC isolates are shorter, and therefore, it is hypothesized that they are responsible for producing a smaller product, although the similarity between the *NRPS9* homologues suggests that the initial unit is likely γ–amino butyl acid (GABA), as seen with fusaotaxin A.

Many secondary metabolites have phytotoxic activity and can aid in pathogenicity; in particular, the previously mentioned naphthoquinones are reported to arrest root growth [9]. Another example is radicicol (also known as monorden), which was first identified as a phytotoxin in *F. virguliforme*, a member of clade 2 of the FSSC and is the causative agent of soybean sudden death syndrome (SDS) [40]. Radicicol is capable of producing SDS-like symptoms such as marginal curling and interveinal necrosis on soybean leaves. The secondary metabolite is an inhibitor of heat shock protein 90 [41], and its production in the fungus *Colletotrichum graminicola* is hypothesized to suppress competing microorganisms and the plant defense response [42]. The two core radicicol PKS biosynthetic genes from *F. tenuicristatum* are similar to the *PKS4*- and *PKS13*-containing BGC in *F. graminearum*, which are responsible for the synthesis of the mycotoxin zearalenone [43,44]. A recent genus-wide phylogenetic analysis of PKS proteins from *Fusarium* resolved the radicicol-producing *PKS4* into a separate clade, *PKS116*, and identified structural differences in radicicol and zearalenone that were caused by reducing PKSs and post-PKS modification by other enzymes [14]. While there is similarity between the zearalenone BGC of *F. graminearum* and the radicicol BGC in *F. tenuicristatum*, radicicol biosynthesis could provide a competitive advantage in rhizosphere colonization and/or be directly involved in suppressing the plant host immune response.

The presence of some of the genes from the gibberellin BGC have been previously described in genomes of other *Fusarium* spp. that are more evolutionarily related than members of the FSSC (i.e., *F. mangiferae*, *F. circinatum*, and *F. oxysporum*; [45]). The *F. mori* genome has four of the seven genes in the gibberellin BGC including a homologue of the key *CPS/KS* gene (Figure 6A). While other *Fusarium* genomes have been identified to have intact gibberellin BGCs, these *Fusarium* isolates did not produce gibberellins under standard laboratory conditions [45]. Therefore, it is unlikely that *F. mori* is capable of producing gibberellins and the product of this similar BGC, if any, is unknown.

Several DMATSs and phosphonate BGCs were identified in the FSSC; however, after an analysis of these clusters, none of the final products could be predicted. DMATSs catalyze the prenylation of L-tryptophan to generate dimethylallytrypthohan (DMAT), which is then predicted to be further modified by the accessory proteins also encoded within the cluster [46]. One of the best-known examples of secondary metabolites synthesized by a DMATS encoded in BGCs are the ergot alkaloids of *Claviceps* spp. Phosphonate-producing BGCs are capable of aiding an organism to sequester phosphorus when it is scarce or can function as an antimicrobial metabolite by producing toxic phosphonate compounds [47]. Further experimental characterization of these BGCs is necessary to identify the natural products synthesized and evaluate their biological relevance.

Fungal secondary metabolites have a diverse array of bioactivity as evidenced in the number of compounds that have been developed for clinical use. As a majority of the secondary metabolite BGCs from members of the FSSC produce unknown products, further research into this area could provide alternative therapeutics. Additionally, better knowledge of these compounds could provide alternative options for disease management. Overall, the vast armamentarium utilized by members of the FSSC not only plays a critical role in pathogenesis but likely provides a competitive advantage to these fungi for further expanding their environmental niches.

their core biosynthetic protein IDs; Table S7: Distribution of the DMATS gene cluster families (I-GCF) identified in ten clade 3 FSSC genomes and their core biosynthetic protein IDs; Table S8: Distribution of the phosphonate-producing gene cluster families (P-GCF) identified in ten clade 3 FSSC genomes and their core biosynthetic protein IDs.

References

1. Keller, N.P. Fungal secondary metabolism: Regulation, function and drug discovery. *Nat. Rev. Microbiol.* **2019**, *17*, 167–180. [CrossRef] [PubMed]
2. Geiser, D.M.; Aoki, T.; Bacon, C.W.; Baker, S.E.; Bhattacharyya, M.K.; Brandt, M.E.; Brown, D.W.; Burgess, L.W.; Chulze, S.; Coleman, J.J.; et al. One fungus, one name: Defining the genus *Fusarium* in a scientifically robust way that preserves longstanding use. *Phytopathology* **2013**, *103*, 400–408. [CrossRef]
3. Geiser, D.M.; Al-Hatmi, A.; Aoki, T.; Arie, T.; Balmas, V.; Barnes, I.; Bergstrom, G.C.; Bhattacharyya, M.K.K.; Blomquist, C.L.; Bowden, R.; et al. Phylogenomic analysis of a 55.1 kb 19-gene dataset resolves a monophyletic *Fusarium* that includes the *Fusarium solani* Species Complex. *Phytopathology* **2020**, *111*, 1064–1079. [CrossRef] [PubMed]
4. O'Donnell, K.; Rooney, A.P.; Proctor, R.H.; Brown, D.W.; McCormick, S.P.; Ward, T.J.; Frandsen, R.J.N.; Lysoe, E.; Rehner, S.A.; Aoki, T.; et al. Phylogenetic analyses of *RPB1* and *RPB2* support a middle Cretaceous origin for a clade comprising all agriculturally and medically important fusaria. *Fungal Genet. Biol.* **2013**, *52*, 20–31. [CrossRef]
5. Coleman, J.J. The *Fusarium solani* species complex: Ubiquitous pathogens of agricultural importance. *Mol. Plant Pathol.* **2016**, *17*, 146–158. [CrossRef] [PubMed]
6. Muhammed, M.; Anagnostou, T.; Desalermos, A.; Kourkoumpetis, T.K.; Carneiro, H.A.; Glavis-Bloom, J.; Coleman, J.J.; Mylonakis, E. *Fusarium* infection: Report of 26 cases and review of 97 cases from the literature. *Medicine* **2013**, *92*, 305–316. [CrossRef]
7. Zhang, N.; O'Donnell, K.; Sutton, D.A.; Nalim, F.A.; Summerbell, R.C.; Padhye, A.A.; Geiser, D.M. Members of the *Fusarium solani* species complex that cause infections in both humans and plants are common in the environment. *J. Clin. Microbiol.* **2006**, *44*, 2186–2190. [CrossRef]
8. O'Donnell, K.; Sutton, D.A.; Fothergill, A.; McCarthy, D.; Rinaldi, M.G.; Brandt, M.E.; Zhang, N.; Geiser, D.M. Molecular phylogenetic diversity, multilocus haplotype nomenclature, and in vitro antifungal resistance within the *Fusarium solani* species complex. *J. Clin. Microbiol.* **2008**, *46*, 2477–2490. [CrossRef]
9. Medentsev, A.G.; Akimenko, V.K. Naphthoquinone metabolites of the fungi. *Phytochemistry* **1998**, *47*, 935–959. [CrossRef]
10. Short, D.P.G.; O'Donnell, K.; Thrane, U.; Nielsen, K.F.; Zhang, N.; Juba, J.H.; Geiser, D.M. Phylogenetic relationships among members of the *Fusarium solani* species complex in human infections and the descriptions of *F. keratoplasticum* sp nov and *F. petroliphilum* stat. nov. *Fungal Genet. Biol.* **2013**, *53*, 59–70. [CrossRef]
11. Nielsen, M.R.; Sondergaard, T.E.; Giese, H.; Sørensen, J.L. Advances in linking polyketides and non-ribosomal peptides to their biosynthetic gene clusters in *Fusarium*. *Curr. Genet.* **2019**, *65*, 1263–1280. [CrossRef]
12. Romans-Fuertes, P.; Sondergaard, T.E.; Sandmann, M.I.H.; Wollenberg, R.D.; Nielsen, K.F.; Hansen, F.T.; Giese, H.; Brodersen, D.E.; Sørensen, J.L. Identification of the non-ribosomal peptide synthetase responsible for biosynthesis of the potential anti-cancer drug sansalvamide in *Fusarium solani*. *Curr. Genet.* **2016**, *62*, 799–807. [CrossRef] [PubMed]
13. Graziani, S.; Vasnier, C.; Daboussi, M.J. Novel polyketide synthase from *Nectria haematococca*. *Appl. Environ. Microbiol.* **2004**, *70*, 2984–2988. [CrossRef] [PubMed]
14. Brown, D.W.; Kim, H.-S.; McGovern, A.E.; Probyn, C.E.; Proctor, R.H. Genus-wide analysis of *Fusarium* polyketide synthases reveals broad chemical potential. *Fungal Genet. Biol.* **2022**, *160*, 103696. [CrossRef] [PubMed]

15. Coleman, J.J.; Rounsley, S.D.; Rodriguez-Carres, M.; Kuo, A.; Wasmann, C.C.; Grimwood, J.; Schmutz, J.; Taga, M.; White, G.J.; Zhou, S.; et al. The genome of *Nectria haematococca*: Contribution of supernumerary chromosomes to gene expansion. *PLOS Genet.* **2009**, *5*, e1000618. [CrossRef]

16. Coleman, J.J. *Exploring the Supernumerary Chromosomes of the Fusarium Solani and Fusarium oxysporum Species Complexes: Reservoirs for Function Genetic Diversity*; USDDOE Joint Genome Institute (JGI): Walnut Creek, CA, USA, 2013.

17. Mesny, F.; Miyauchi, S.; Thiergart, T.; Pickel, B.; Atanasova, L.; Karlsson, M.; Hüttel, B.; Barry, K.W.; Haridas, S.; Chen, C.; et al. Genetic determinants of endophytism in the Arabidopsis root mycobiome. *Nat. Commun.* **2021**, *12*, 7227. [CrossRef]

18. Camacho, C.; Coulouris, G.; Avagyan, V.; Ma, N.; Papadopoulos, J.; Bealer, K.; Madden, T.L. BLAST+: Architecture and applications. *BMC Bioinform.* **2009**, *10*, 421. [CrossRef]

19. Tamura, K.; Stecher, G.; Kumar, S. MEGA11: Molecular evolutionary genetics analysis version 11. *Mol. Biol. Evol.* **2021**, *38*, 3022–3027. [CrossRef]

20. Castresana, J. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol. Biol. Evol.* **2000**, *17*, 540–552. [CrossRef]

21. Blin, K.; Shaw, S.; Kloosterman, A.M.; Charlop-Powers, Z.; van Wezel, G.P.; Medema, M.H.; Weber, T. AntiSMASH 6.0: Improving Cluster Detection and Comparison Capabilities. *Nucleic Acids Res.* **2021**, *49*, W29–W35. [CrossRef]

22. Brown, D.W.; Proctor, R.H. Insights into natural products biosynthesis from analysis of 490 polyketide synthases from Fusarium. *Fungal Genet. Biol.* **2016**, *89*, 37–51. [CrossRef] [PubMed]

23. Hansen, F.T.; Gardiner, D.M.; Lysøe, E.; Fuertes, P.R.; Tudzynski, B.; Wiemann, P.; Sondergaard, T.E.; Giese, H.; Brodersen, D.E.; Sørensen, J.L. An Update to Polyketide Synthase and Non-Ribosomal Synthetase Genes and Nomenclature in *Fusarium*. *Fungal Genet. Biol.* **2015**, *75*, 20–29. [CrossRef] [PubMed]

24. Bushley, K.E.; Turgeon, B.G. Phylogenomics reveals subfamilies of fungal nonribosomal peptide synthetases and their evolutionary relationships. *BMC Evol. Biol.* **2010**, *10*, 26. [CrossRef]

25. Katoh, K.; Standley, D.M. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol. Biol. Evol.* **2013**, *30*, 772–780. [CrossRef]

26. Criscuolo, A.; Gribaldo, S. BMGE (Block Mapping and Gathering with Entropy): A new software for selection of phylogenetic informative regions from multiple sequence alignments. *BMC Evol. Biol.* **2010**, *10*, 210. [CrossRef]

27. Darriba, D.; Taboada, G.L.; Doallo, R.; Posada, D. ProtTest 3: Fast selection of best-fit models of protein evolution. *Bioinformatics* **2011**, *27*, 1164–1165. [CrossRef] [PubMed]

28. Stamatakis, A. RAxML Version 8: A Tool for Phylogenetic Analysis and Post-Analysis of Large Phylogenies. *Bioinformatics* **2014**, *30*, 1312–1313. [CrossRef] [PubMed]

29. Letunic, I.; Bork, P. Interactive Tree of Life (iTOL) V5: An online tool for phylogenetic tree display and annotation. *Nucleic Acids Res.* **2021**, *49*, W293–W296. [CrossRef]

30. Navarro-Muñoz, J.C.; Selem-Mojica, N.; Mullowney, M.W.; Kautsar, S.A.; Tryon, J.H.; Parkinson, E.I.; De Los Santos, E.L.C.; Yeong, M.; Cruz-Morales, P.; Abubucker, S.; et al. A computational framework to explore large-scale biosynthetic diversity. *Nat. Chem. Biol.* **2020**, *16*, 60–68. [CrossRef]

31. Terlouw, B.R.; Blin, K.; Navarro-Muñoz, J.C.; Avalon, N.E.; Chevrette, M.G.; Egbert, S.; Lee, S.; Meijer, D.; Recchia, M.J.J.; Reitz, Z.L.; et al. MIBiG 3.0: A community-driven effort to annotate experimentally validated biosynthetic gene clusters. *Nucleic Acids Res.* **2023**, *51*, D603–D610. [CrossRef]

32. Gilchrist, C.M.; Chooi, Y.-H. clinker & clustermap.js: Automatic generation of gene cluster comparison figures. *Bioinformatics* **2021**, *37*, 2473–2475. [CrossRef] [PubMed]

33. Sugiura, Y.; Barr, J.R.; Barr, D.B.; Brock, J.W.; Elie, C.M.; Ueno, Y.; Patterson, D.G.; Potter, M.E.; Reiss, E. Physiological characteristics and mycotoxins of human clinical isolates of *Fusarium* species. *Mycol. Res.* **1999**, *103*, 1462–1468. [CrossRef]

34. Jia, L.-J.; Tang, H.-Y.; Wang, W.-Q.; Yuan, T.-L.; Wei, W.-Q.; Pang, B.; Gong, X.-M.; Wang, S.-F.; Li, Y.-J.; Zhang, D.; et al. A linear nonribosomal octapeptide from *Fusarium graminearum* facilitates cell-to-cell invasion in wheat. *Nat. Commun.* **2019**, *10*, 922. [CrossRef] [PubMed]

35. Coleman, J.J.; Ghosh, S.; Okoli, I.; Mylonakis, E. Antifungal activity of microbial secondary metabolites. *PLoS ONE* **2011**, *6*, e25321. [CrossRef] [PubMed]

36. Haas, H.; Eisendle, M.; Turgeon, B.G. Siderophores in fungal physiology and virulence. *Annu. Rev. Phytopathol.* **2008**, *46*, 149–187. [CrossRef]

37. Oide, S.; Berthiller, F.; Wiesenberger, G.; Adam, G.; Turgeon, B.G. Individual and combined roles of malonichrome, ferricrocin, and TAFC siderophores in *Fusarium graminearum* pathogenic and sexual development. *Front. Microbiol.* **2015**, *5*, 759. [CrossRef]

38. Lee, B.-N.; Kroken, S.; Chou, D.Y.T.; Robbertse, B.; Yoder, O.C.; Turgeon, B.G. Functional analysis of all nonribosomal peptide synthetases in *Cochliobolus heterostrophus* reveals a factor, NPS6, involved in virulence and resistance to oxidative stress. *Eukaryot. Cell* **2005**, *4*, 545–555. [CrossRef]

39. Krasnoff, S.B.; Keresztes, I.; Donzelli, B.G.G.; Gibson, D.M. Metachelins, mannosylated and N-oxidized coprogen-type siderophores from *Metarhizium robertsii*. *J. Nat. Prod.* **2014**, *77*, 1685–1692. [CrossRef]

40. Chang, H.-X.; Domier, L.L.; Radwan, O.; Yendrek, C.R.; Hudson, M.E.; Hartman, G.L. Identification of multiple phytotoxins produced by *Fusarium virguliforme* including a phytotoxic effector (FvNIS1) associated with sudden death syndrome foliar symptoms. *Mol. Plant-Microbe Interact.* **2016**, *29*, 96–108. [CrossRef]

41. Piper, P.W.; Millson, S.H. Spotlight on the microbes that produce heat shock protein 90-targeting antibiotics. *Open Biol.* **2012**, *2*, 120138. [CrossRef]

42. Wicklow, D.T.; Jordan, A.M.; Gloer, J.B. Antifungal metabolites (monorden, monocillins I, II, III) from *Colletotrichum graminicola*, a systemic vascular pathogen of maize. *Mycol. Res.* **2009**, *113*, 1433–1442. [CrossRef] [PubMed]

43. Kim, Y.-T.; Lee, Y.-R.; Jin, J.; Han, K.-H.; Kim, H.; Kim, J.-C.; Lee, T.; Yun, S.-H.; Lee, Y.-W. Two different polyketide synthase genes are required for synthesis of zearalenone in *Gibberella zeae. Mol. Microbiol.* **2005**, *58*, 1102–1113. [CrossRef] [PubMed]

44. Gaffoor, I.; Trail, F. Characterization of two polyketide synthase genes involved in zearalenone biosynthesis in *Gibberella zeae*. *Appl. Environ. Microbiol.* **2006**, *72*, 1793–1799. [CrossRef] [PubMed]

45. Wiemann, P.; Sieber, C.M.K.; von Bargen, K.W.; Studt, L.; Niehaus, E.-M.; Espino, J.J.; Huß, K.; Michielse, C.B.; Albermann, S.; Wagner, D.; et al. Deciphering the cryptic genome: Genome-wide analyses of the rice pathogen *Fusarium fujikuroi* reveal complex regulation of secondary metabolism and novel metabolites. *PLOS Pathog.* **2013**, *9*, e1003475. [CrossRef]

46. Xu, W.; Gavia, D.J.; Tang, Y. Biosynthesis of fungal indole alkaloids. *Nat. Prod. Rep.* **2014**, *31*, 1474–1487. [CrossRef] [PubMed]

47. Yu, X.; Doroghazi, J.R.; Janga, S.C.; Zhang, J.K.; Circello, B.; Griffin, B.M.; Labeda, D.P.; Metcalf, W.W. Diversity and abundance of phosphonate biosynthetic genes in nature. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 20759–20764. [CrossRef]