


Complex Evolutionary Origins of Specialized Metabolite Gene Cluster Diversity among the Plant Pathogenic Fungi of the *Fusarium graminearum* Species Complex

Sabina Moser Tralamazza^{1,2,*}, Liliana Oliveira Rocha³, Ursula Oggenfuss², Benedito Corrêa¹, and Daniel Croll ^{2,*}

¹Department of Microbiology, Institute of Biomedical Sciences, University of São Paulo, Brazil

²Laboratory of Evolutionary Genetics, Institute of Biology, University of Neuchâtel, Switzerland

³Food Engineering Faculty, Department of Food Science, University of Campinas, Av. Monteiro Lobato, Brazil

*Corresponding authors: E-mails: sabina@usp.br; daniel.croll@unine.ch.

Accepted: October 10, 2019

Data deposition: This project assembled genomes have been deposited at NCBI genome database under the accession number VSSU00000000, VSSV00000000, VSSW00000000 and VSSX00000000. Scripts and phylogenetic tree files were deposited on Figshare (https://figshare.com/projects/Data_Tralamazza_et_al_2019_/67595).

Abstract

Fungal genomes encode highly organized gene clusters that underlie the production of specialized (or secondary) metabolites. Gene clusters encode key functions to exploit plant hosts or environmental niches. Promiscuous exchange among species and frequent reconfigurations make gene clusters some of the most dynamic elements of fungal genomes. Despite evidence for high diversity in gene cluster content among closely related strains, the microevolutionary processes driving gene cluster gain, loss, and neofunctionalization are largely unknown. We analyzed the *Fusarium graminearum* species complex (FGSC) composed of plant pathogens producing potent mycotoxins and causing Fusarium head blight on cereals. We de novo assembled genomes of previously uncharacterized FGSC members (two strains of *F. austroamericanum*, *F. cortaderiae*, and *F. meridionale*). Our analyses of 8 species of the FGSC in addition to 15 other *Fusarium* species identified a pangenome of 54 gene clusters within FGSC. We found that multiple independent losses were a key factor generating extant cluster diversity within the FGSC and the *Fusarium* genus. We identified a modular gene cluster conserved among distantly related fungi, which was likely reconfigured to encode different functions. We also found strong evidence that a rare cluster in FGSC was gained through an ancient horizontal transfer between bacteria and fungi. Chromosomal rearrangements underlying cluster loss were often complex and were likely facilitated by an enrichment in specific transposable elements. Our findings identify important transitory stages in the birth and death process of specialized metabolism gene clusters among very closely related species.

Key words: head blight, wheat, fungus, pathogen, secondary metabolism.

Introduction

Fungal genomes encode highly organized structures that underlie the capacity to produce specialized (also called secondary) metabolites. The structures are composed of a tightly clustered group of nonhomologous genes that in conjunction confer the enzymatic pathway to produce a specific metabolite (Osborn 2010). Specialized metabolites (SMs) are not essential for the organism's survival but confer crucial benefits for niche adaptation and host exploitation. SMs can promote defense (e.g., penicillin), virulence (e.g., trichothecenes), or

resistance functions (e.g., melanin) (Brakhage 1998; Nosanchuk and Casadevall 2006). Gene clusters are typically composed of two or more key genes in close physical proximity. The backbone gene encodes for the enzyme defining the class of the produced metabolite and the enzyme is most often a polyketide synthase (PKS), nonribosomal peptides synthetase (NRPS), terpenes cyclase, or a dimethylallyl tryptophan synthetase. Additional genes in clusters encode functions to modify the main metabolite structure (e.g., methyltransferases, acetyltransferases, and oxidoreductases), transcription

© The Author(s) 2019. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

factors involved in the cluster regulation and resistance genes that serve to detoxify the metabolite for the producer (Keller et al. 2005). The modular nature of gene clusters favored promiscuous exchange among species and frequent reconfiguration of cluster functionalities (Rokas et al. 2018).

The broad availability of fungal genome sequences led to the discovery of a very large number of SM gene clusters (Brakhage 2013). Yet, how gene clusters are formed or reconfigured to change function over evolutionary time remains poorly understood. The divergent distribution across species (Wisecaver et al. 2014), frequent rearrangements (Rokas et al. 2018), and high polymorphism within single species (Lind et al. 2017; Wollenberg et al. 2019) complicate the analyses of gene cluster evolution. Most studies analyzed deep evolutionary timescales and focused on the origins and loss of major gene clusters (Wisecaver et al. 2014). Gene clusters often emerged through rearrangement or duplications of native genes (Wong and Wolfe 2005; Slot and Rokas 2010; Wisecaver et al. 2014). The DAL gene cluster involved in the allantoin metabolism is a clear example of this mechanism. The cluster was formed from the duplication of two genes and relocation of four native genes in the yeast *Saccharomyces cerevisiae* (Wong and Wolfe 2005). Gene clusters can also arise in species from horizontal gene transfer events (Khaldi et al. 2008; Khaldi and Wolfe 2011; Campbell et al. 2012; Slot and Rokas 2011). For example, the complete and functional gene cluster underlying the production of the aflatoxin precursor sterigmatocystin was horizontally transferred from *Aspergillus* to the unrelated *Podospora anserine* fungus (Slot and Rokas 2011). Five gene clusters underlying the hallucinogenic psilocybin production were horizontally transmitted among the distantly related fungi *Psilocybe cyanescens*, *Gymnopilus dilepis*, and *Panaeolus cyanescens* (Reynolds et al. 2018). The horizontal transfer was likely favored by the overlapping ecological niche of the involved species.

Despite evidence for high diversity in gene cluster content among closely related strains (Wiemann et al. 2013), the microevolutionary processes driving gene cluster gain, loss, and neofunctionalization are largely unknown. Closely related species or species complexes encoding diverse gene clusters are ideal models to reconstruct transitory steps in the evolution of gene clusters. The *Fusarium graminearum* species complex (FGSC) is composed of a series of plant pathogens capable to produce potent mycotoxins and cause the Fusarium head blight disease in cereals. The species complex was originally described as a single species. Based on genealogical concordance phylogenetic species recognition, members of *F. graminearum* were expanded into a species complex (O'Donnel et al. 2004). Currently, the complex includes at least 16 distinct species that vary in aggressiveness, growth rate, and geographical distribution but lack morphological differentiation (Ward et al. 2008; Puri and Zhong 2010; Aoki et al. 2012; Zhang et al. 2012). The genome of

F. graminearum sensu stricto, the dominant species of the complex, was extensively characterized for the presence of SM gene clusters (Aoki et al. 2012; Wiemman et al. 2013; Hoogendoorn et al. 2018; Brown and Proctor 2016). Based on genomics and transcriptomics analyses, Sieber et al. (2014) characterized a large number of clusters with a potential to contribute to virulence and identified likely horizontal gene transfer events.

However, the species complex harbors several other economically relevant species with largely unknown SM production potential (van der Lee et al. 2015). Diversity in metabolic capabilities within the FGSC extends to production of the potent mycotoxin trichothecene. The biosynthesis of some trichothecene variant forms (15-acetyldeoxyvalenol, 3-acetyldeoxynivalenol and nivalenol) are species-specific and associated with pathogenicity (Desjardins 2006). Comparative genomics analyses of three species of the complex (*F. graminearum* s.s., *F. asiaticum*, *F. meridionale*) identified species-specific genes associated with the biosynthesis of metabolites (e.g., PKS40 in *F. asiaticum*) (Walkowiak et al. 2016). Most species were not analyzed at the genome level for SM production potential or lack an assembled genome altogether.

In this study, we aimed to characterize exhaustively the metabolic potential of the FGSC based on comparative genomics analyses and reconstruct the evolutionary processes governing the birth and death process of gene clusters among the recently emerged species. For this, we sequenced and assembled genomes for *F. meridionale*, *F. cortaderiae*, and two strains of *F. austroamericanum*—four genomes of the most frequent members of the FGSC found in Brazilian wheat grains, after the well-characterized *F. graminearum* s.s. In total, we analyzed 11 genomes from 8 distinct species within the FGSC. We identified 54 SM gene clusters in the pangenome of the FGSC including two gene clusters not yet known from the complex. The variability in SM gene clusters was generated by multiple independent losses, horizontal gene transfer, and chromosomal rearrangements that produced novel gene cluster configurations.

Materials and Methods

Strains, DNA Preparation, and Sequencing

The fungal strains (*F. meridionale*—Fmer152; *F. cortaderiae*—Fcor153; *F. austroamericanum*—Faus151 and Faus154) were isolated from healthy and freshly harvested wheat grains from three different regions of Brazil, São Paulo State (Fmer152 and Faus151), Parana State (Fcor153), and Rio Grande do Sul State (Faus154) (Tralamazza et al. 2016). The DNA extraction was performed using a DNAeasy kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions. DNA quality was analyzed using a NanoDrop2000 (ThermoFisher Scientific, USA) and Qubit (ThermoFisher

Scientific) was used for the DNA quantification (minimal DNA concentration of 50 ng/μl). Nextera Mate Pair Sample Preparation kit (Illumina Inc.) was used for DNA Illumina library preparation. Samples were sequenced using 75 bp reads from paired-end libraries on a NextSeq500 v2 (Illumina Inc.) by the Idogene Inc. (Sao Paulo, Brazil). The software FastQC v. 0.11.7 (Andrews 2010) was used for quality control of the raw sequence reads. To perform phylogenomic analyses, whole-genome sequences of *Fusarium* species and *Trichoderma reesei* (as an outgroup) were retrieved from public databases (see [supplementary table S1, Supplementary Material](#) online for accession numbers).

Genome Assembly

De novo genome assembly was performed for the four newly sequenced genomes of the FGSC (*F. meridionale*—Fmer152; *F. cortaderiae*—Fcor153; *F. austroamericanum*—Faus151 and Faus154) and for the publicly available 150 bp paired-end raw sequence data for *F. boothii*, *F. gerlachii*, and *F. louisianense* ([supplementary table S1, Supplementary Material](#) online). We used the software Spades v.3.12.0 (Bankevich et al. 2012) to assemble Illumina short read data to scaffolds using the “careful” option to reduce mismatches. We selected the k-mer series “21, 33, 45, 67” for *F. meridionale*, *F. cortaderiae*, and *F. austroamericanum* sequences; and “21, 33, 55, 77, 99, 127” for *F. boothii*, *F. gerlachii*, and *F. louisianense*. The maximum k-mer values were adjusted according to available read length. For all other genomes included in the study (including *F. asiaticum* and *F. graminearum* s.s.), assembled scaffolds were retrieved from NCBI or Ensembl database ([supplementary table S1, Supplementary Material](#) online). The quality of draft genome assemblies was assessed using QUAST v.4.6.3 (Gurevich et al. 2013). BUSCO v.3.0.1 (Waterhouse et al. 2018) was used to assess the completeness of core fungal orthologs based on the data set fungi_odb9 which comprises 290 core orthologs of 85 species.

Gene Prediction and Annotation

Genes were predicted using Augustus v.2.5.5 (Stanke and Morgenstern 2005). We used the pretrained gene prediction database for the *F. graminearum* s.s. genome as provided by the Augustus distribution for all annotations and used default parameters otherwise. Predicted proteomes were annotated using InterProScan v.5.19 (Jones et al. 2014) identifying conserved protein domains and gene ontology. Secreted proteins were defined according to the absence of transmembrane domains and the presence of a signal peptide based on Phobius v.1.01 (Kall et al. 2007), SignalP v.4.1 (Petersen et al. 2011), and TMHMM v.2.0 (Krog et al. 2001) concordant results. We identified the predicted secretome with a machine learning approach implemented in EffectorP v2.0 (Sperschneider et al. 2018). We used the package Codon

Adaptation Index of the Jemboss v. 1.5 software to analyze codon usage variation (Carver and Bleasby 2003).

Genome Alignment and Phylogenomic Analyses

For the phylogenomic analyses, we used OrthoMCL (Li et al. 2003) to identify single-copy orthologs conserved among all strains. High accuracy alignment of orthologous sequences was performed using MAFFT v.7.3 (Katoh et al. 2017) with parameters `–maxiterate 1000 –localpair`. To construct a maximum-likelihood phylogenetic tree for each alignment, we used RAXML v.8.2.12 (Stamatakis 2014) with parameters `–m PROTGAMMAAUTO` and bootstrap of 100 replicates). The whole-genome phylogeny tree was constructed using Astral III v.5.1.1 (Zhang et al. 2017) which uses the multi-species coalescent model and estimates a species tree given a set of unrooted gene trees. We used Figtree v.1.4.0 for the visualization of phylogenetic trees (Rambaut 2012).

SM Gene Cluster Prediction

To retrieve SM gene clusters from genome assemblies, we performed analyses using antiSMASH v.3.0 (Blin et al. 2017) and matched predicted gene clusters with functional predictions based on InterProScan v. 5.29-68 (Jones et al. 2014). For the *F. graminearum* reference genome (FgramR), we retrieved SM gene clusters identified in a previous study, which used evidence from multiple prediction tools and incorporated expression data (Sieber et al. 2014). We selected only clusters with a defined class/function, identified backbone gene and annotated cluster size. We made an exception for cluster SM45, which was predicted by antiSMASH but not characterized by Sieber et al. (2014) likely due to discrepancies in gene annotation.

Pangenome SM Gene Cluster Map and Synteny Analysis

We constructed a pangenome of SM gene clusters in the FGSC by mapping the backbone genes of each distinct cluster against all other genomes. BLAST+ v.2.8 (Camacho et al. 2009) local alignment search (Blastp with default parameters) was performed and matches with the highest bitscores were retrieved. For each unique cluster in FGSC, we selected the backbone gene of a specific genome as a reference for the presence/absence analyses within the complex. We used FgramR backbone sequences for the majority of the clusters (clusters SM1–SM45), for SM46 we used FasiR2, for SM47–SM52 FasiR, for SM53 we used Fcor153 and for SM54 we used Faus154 ([supplementary table S3, Supplementary Material](#) online). We considered a gene cluster as present if the Blastp identity of the backbone gene was above 90% (threshold for FGSC members). For strains outside of the FGSC (i.e. all other *Fusarium* species), we used a cutoff of 70%. Heatmaps were drawn using the R package ggplot2 (Wickham 2016) and syntenic regions of the gene clusters

were drawn using the R package *genoplR* (Guy et al. 2010). For SMGC with taxonomical distribution mismatching the species phylogeny, we performed additional phylogenetic analyses. For this, we queried each encoded protein of a cluster in the NCBI protein database (see [supplementary table S2, Supplementary Material](#) online for accession numbers). We reconstructed the most likely evolutionary history of a gene cluster using the maximum-likelihood method based on the JTT matrix-based amino acid substitution model (Jones et al. 1992). We performed 1,000 bootstrap replicates and performed all analyses using the software *MEGA v.7.0.26* (Kumar et al. 2016).

Repetitive Elements Annotation

We performed *de novo* repetitive element identification of the complete genome of *F. graminearum* (FgramR) using RepeatModeler 1.0.11 (Smit and Hubley 2008). We identified conserved domains of the coding region of the transposable elements using BlastX and the nonredundant NCBI protein database. One predicted transposable element family was excluded due to the high sequence similarity to a major facilitator superfamily gene and low copy number ($n=2$), which strongly suggests that a duplicated gene was misidentified as a transposable element. We then annotated the repetitive elements with RepeatMasker v.4.0.7 (Smit et al. 2015). One predicted transposable element family (element 4-family1242) showed extreme length polymorphism between the individual insertions and no clearly identifiable conservation among all copies. The consensus sequence of family1242 also contained several large poly-A islands, tandem repeats and palindromes. Using BlastN, we mapped the sequences of all predicted insertions against the consensus sequence and identified five distinct regions with low sequence similarity between them. We created new consensus sequences for each of these five regions based on the genomes of *F. graminearum* and *F. austroamericanum* (Faus154) (Zhang et al. 2000; Morgulis et al. 2008). We filtered all retrieved sequences for identity $>80\%$ and $>80\%$ alignment length. We added flanking sequences of 3,000 bp and visually inspected all retrieved hits with Dotter v.3.1 (Sonnhammer and Durbin 1995). Then, we performed a multiple sequence alignment using Clustalw (Higgins and Sharp 1988; Altschul 1997) to create new consensus sequences. Finally, we replaced the erroneous element 4-family1242 with the five identified subregions. We used the modified repeat element library jointly with the Dfam and Repbase database to annotate all genomes using RepeatMasker (Smit and Hubley 2008). Transposable element locations in the genome were visualized with the R package *genoPlotR v0.8.9* (Guy et al. 2010). We performed transposable element density analyses of the genomes in 10 kb windows using *bedtools v.2.27* (Quinlan and Hall 2010).

Results

Genomic Sampling of the FGSC

We analyzed genomes of 11 strains of 8 different species of the FGSC in order to resolve species relationships and detect divergence in their specialized metabolism. We performed the first *de novo* assembly and genome annotation for two strains of *F. austroamericanum* (Faus151 and Faus154), a strain of *F. cortaderiae* (Fcor153), and a strain of *F. meridionale* (Fmer152). We included 15 other species of the *Fusarium* genus including the *Fusarium fujikuroi* species complex (FFSC) and the *Fusarium sambucinum* species complex (FSAMSC) to distinguish between gene gains and losses. We first assessed the genome assembly quality within FGSC ([supplementary table S1, Supplementary Material](#) online). N50 values of the newly sequenced genomes ranged from 220 to 442 kb. The N50 of previously sequenced genomes of the FGSC ranged from 149 to 9,395 kb including the fully finished assembly of the reference genome *F. graminearum* PH-1 (FgramR). By analyzing the completeness of all assemblies, we found the percentage of recovered BUSCO orthologues to be above 99.3% for all FGSC members. The genome sizes within the FGSC ranged from 35.02 to 38.0 Mb. All genomes shared a similar GC content (47.84–48.39%) and number of predicted genes (11,484–11,985) excluding the reference genome. The *F. graminearum* reference genome showed a higher number of predicted genes (14,145) most likely due to the completeness of the assembly and different gene annotation procedures. The percentage of repetitive elements in the genome varied from 0.47 to 4.85% among members of the *Fusarium* genus with a range of 0.97–1.99% within the FGSC. Genomes of strains falling outside of the FGSC showed N50 values and a BUSCO recovery of 31–9,395 kb and 93–100%, respectively.

Phylogenomic Reconstruction

We analyzed the phylogenetic relationships of eight distinct species within the FGSC and 15 additional members of *Fusarium*. We included *Trichoderma reesei* as an outgroup species. Using OrthoMCL, we identified 4,191 single-copy orthologs conserved in all strains and used these to generate a maximum-likelihood phylogenomic tree (fig. 1). The three species complexes included in our analyses (FFSC, FSAMSC, and FGSC) were clearly differentiated with high bootstrap support (100%). All FGSC members clustered as a monophyletic group and *F. culmorum* was the closest species outside of the complex. The cluster of *F. graminearum*, *F. boothii*, *F. gerlachii*, and *F. louisianense*, as well *F. cortaderiae*, *F. austroamericanum*, and *F. meridionale* each formed well-supported clades. The FGSC species clustered together consistent with previous multi-locus phylogenetic studies based on 11 combined genes (Aoki et al. 2012) apart from *F. asiaticum* clade that was found separated from the clade



FIG. 1.—Phylogenomic tree of the *Fusarium graminearum* species complex (FGSC) and other *Fusarium* strains, inferred from a coalescence-based analysis of 4192 single-copy orthologues and bootstrap of 100 replicates. *Trichoderma reesei* was used as an outgroup. Tree nodes without values have a bootstrap of 100%. Substrate/host information was retrieved from Aoki et al. (2012) and Bottalico (1998). **Fusarium oxysporum* lineages are usually host specific. FFSC, *Fusarium fujikuroi* species complex; FSAMSC, *Fusarium sambucinum* species complex.

of *F. graminearum*, *F. boothi*, *F. gerlachii*, and *F. louisianense*. The tree clearly resolves the FSAMSC as a monophyletic group, which includes *F. culmorum*, *F. pseudograminearum*, *F. langsethiae*, *F. poae*, and *F. sambucinum*, together with all members of the FGSC. The members of the FFSC (*F. fujikuroi*, *F. verticillioides*, *F. bulbicola*, *F. proliferatum*, and *F. mangiferae*) also formed a monophyletic group.

SM Gene Clusters Diversity in the FGSC

We analyzed all genome assemblies for evidence of SM gene clusters based on physical clustering and homology-based inference of encoded functions. Out of 54 SM gene cluster within the FGSC, seven were absent from the *F. graminearum* reference (fig. 2). The class of NRPS was

the most frequent SM gene cluster category ($n = 19$), followed by PKS ($n = 13$) and TPS ($n = 11$). We also found several cases of hybrid clusters, containing more than one class of backbone gene (fig. 2). We found substantial variation in the presence or in the absence of SM gene clusters within the FGSC and among *Fusarium* species in general. We classified gene clusters into three distinct categories based on the phylogenetic conservation of the backbone gene in FGSC (fig. 2). Out of the 54 clusters, 43 SM gene clusters were common to all FGSC members (category 1; fig. 2). The SM gene clusters shared within the species complex were usually also found in the heterothallic species *F. culmorum* (86.4% of all clusters) and in *F. pseudograminearum* (79.7% of all clusters), the most closely related species outside of the FGSC (fig. 1). The gene cluster responsible for the production of the

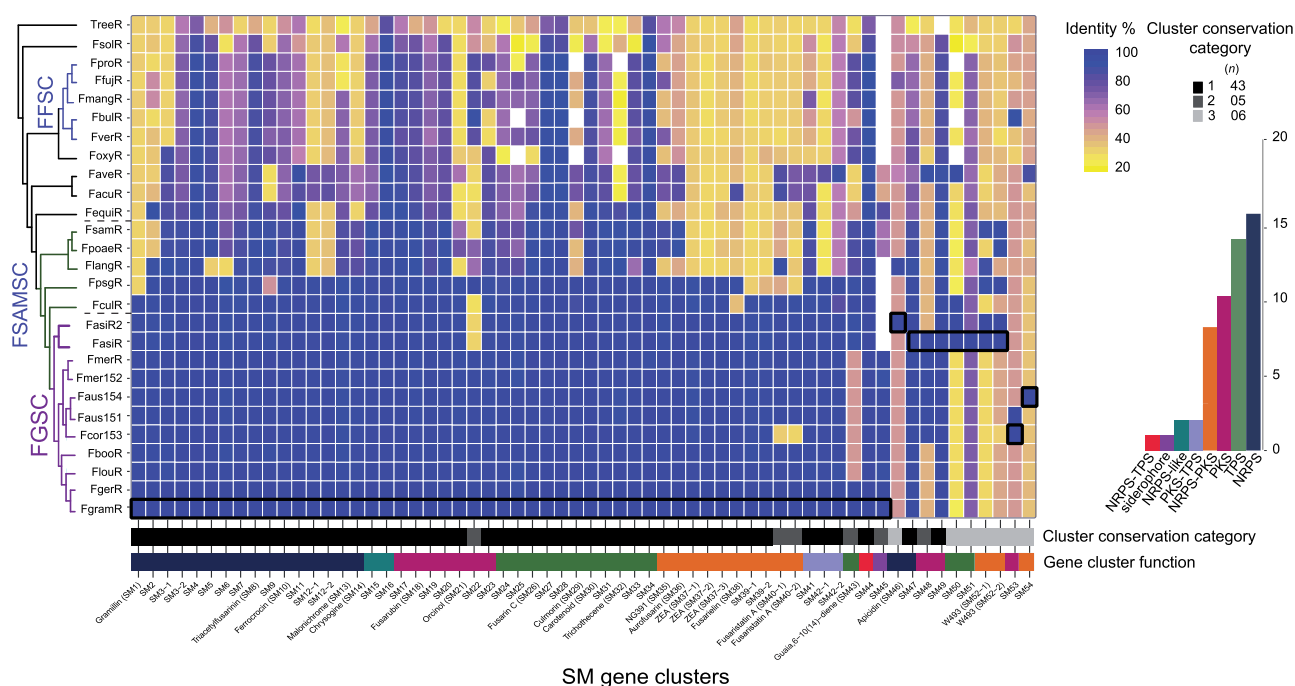


Fig. 2.—Secondary metabolite gene cluster pangenome of the *Fusarium graminearum* species complex (FGSC) based on evidence for backbone genes. Squares with black lines in the heatmap correspond to genomes used for comparative genomic analyses: FgramR (SM1–SM45), FasiR2 (SM46), FasiR (SM47–SM52), Fcor153 (SM53), and Faus154 (SM54). The bar chart is identifying frequencies clusters types. Colored bars below the heatmap correspond to the cluster type. Black/gray bars correspond to the category conservation cluster table. PKS, polyketide synthase; NRPS, nonribosomal peptide synthetase; TPS, terpene synthase; FFSC, *Fusarium fujikuroi* species complex; FSAMSC, *Fusarium sambucinum* species complex.

metabolite gramillin was shared among all FGSC species and *F. culmorum* (fig. 2). We found five SM gene clusters (SM22, SM43, SM45, and SM48) that were not shared by all FGSC members but present in more than 20% of the strains (category 2; fig. 2). Six SM gene clusters (SM46, SM50, SM51, SM52, SM53, and SM54) were rare within the FGSC or even unique to one analyzed genome (category 3; fig. 2). We also found 13 highly conserved SM gene clusters among members of the *Fusarium* genus with 24 of the 26 analyzed genomes encoding the backbone gene (>70% amino acid identity; supplementary table S3, Supplementary Material online). An example of such a conserved cluster is SM8 underlying the production of the siderophore triacetylfusarine, which facilitates iron acquisition both in fungi and bacteria (Charlang et al. 1981).

Multiple Gene Cluster Rearrangements and Losses within the FGSC

We analyzed the mechanisms underlying gene cluster presence–absence polymorphism within the FGSC (categories 2 and 3; fig. 2). These clusters were encoding the machinery for the production of both known and uncharacterized metabolites. We considered a gene cluster to be lost if at least the backbone gene was missing or suffered pseudogenization. Both, SM45, underlying siderophore production, and SM33,

a PKS cluster, were shared among all FGSC members except *F. asiaticum* (FasiR). The cluster of fusaristatin A (SM40), a metabolite with antibiotic activities and expression associated with infection in wheat (Sieber et al. 2014) was another example of cluster loss in a single species, *F. cortaderiae* (Fcor153). We found that the cluster encoding for the production of the metabolite guaia, 6-10(14)-diene (SM43) is conserved in different species within FGSC but the cluster suffered independent losses in *Fusarium*. The TPS class gene cluster identified in *F. fujikuroi* (Burkhardt et al. 2016) was shared among different species complexes (FFSC and FSAMSC; fig. 3). In the FFSC, the species *F. fujikuroi*, *F. proliferatum*, *F. bulbicola*, and *F. mangiferae* share the cluster. In the FSAMSC, the parent complex that includes also FGSC, the guaia, 6-10(14)-diene cluster was found to be rearranged compared with the cluster variant found in the FFSC. Gene cluster synteny analyses among strains within the FGSC showed that several members (*F. cortaderiae*, *F. austroamericanum*, *F. meridionale*, and *F. louisianense*) lost two segments of the cluster. The gene cluster variant with partial deletions retained only the gene encoding for the biosynthesis of pyoverdine and the genes flanking the cluster (fig. 3). To retrace the evolutionary origins of the guaia, 6-10(14)-diene cluster, we performed a phylogenetic analysis of each gene within the cluster. The backbone gene encoding for the terpene synthase and the pyoverdine biosynthesis

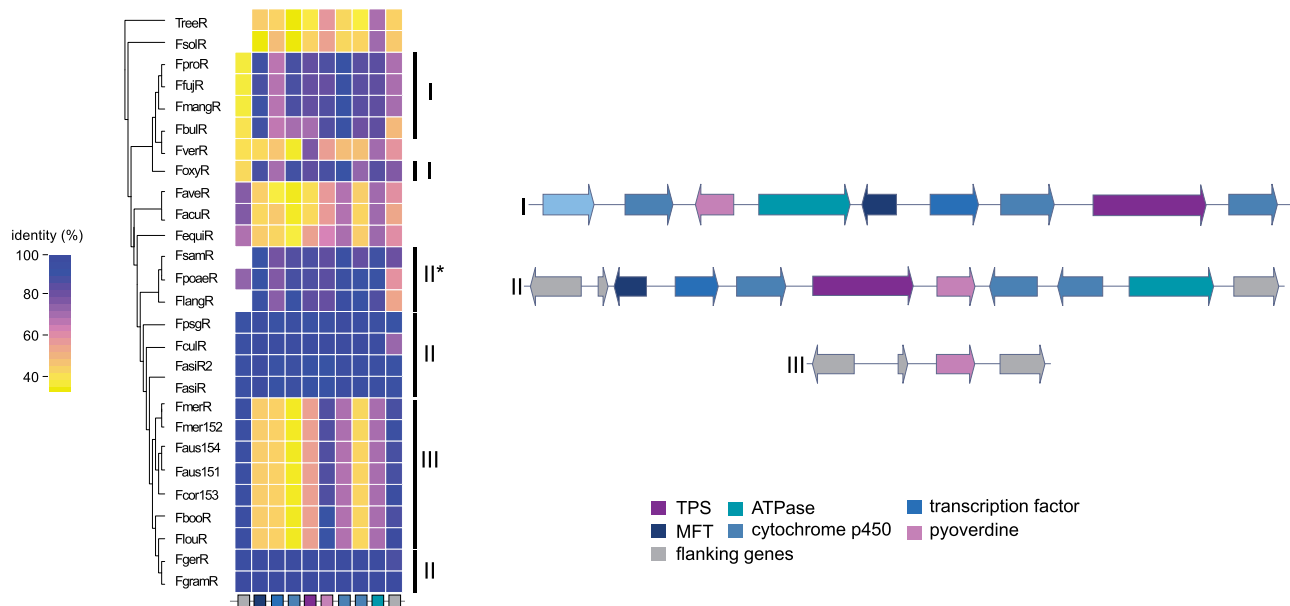


Fig. 3.—Synteny plot of the SM46 (guaia-6, 10-diene) gene cluster and heatmap of protein identity based on the *Fusarium graminearum* FgramR reference genome. Rectangles below the heatmap correspond to the genes shown in the synteny plot. Arrows of identical color correspond to homologous genes and identify the predicted protein function. TPS, terpene synthase; MFT, major facilitator superfamily transporter.

genes show congruent phylogenetic relationships. However, the gene phylogenies showed discrepancies compared with the species tree (supplementary fig. S1, Supplementary Material online). Both gene trees showed that orthologs found within the FGSC grouped with species outside of the complex. *Fusarium graminearum* and *F. gerlachii* formed a subclade with the sister species *F. culmorum* as did *F. asiaticum* with the FSAMSC species *F. pseudograminearum*.

We found the cluster underlying the apicidin metabolite production (SM46) present within the FGSC (fig. 4). The cluster was first discovered in *F. incarnatum* (former *F. semitectum*; Jin et al. 2010) and was found to underlie the production of metabolites with antiparasitic properties (Darkin-Ratway et al. 1996). Our analysis showed that the cluster suffered multiple independent losses across the *Fusarium* genus including a near complete loss within the FGSC, except in the strain of *F. asiaticum* (FasiR2), which shares a complete and syntenic cluster with the distantly related species *F. incarnatum* and *F. sporotrichioides*. *Fusarium langsethiae* is known to produce apicidin A (Lysøe et al 2016) yet it showed a distinct rearrangement or possibly suffered a partial cluster inversion (fig. 4). Surprisingly, the *F. asiaticum* strain FasiR maintained only a pseudogenized NRPS backbone gene and the flanking genes on one end of the cluster. *Fusarium fujikuroi* is missing *aps10* encoding a ketoreductase and is known to produce a similar metabolite called apicidin-F (Niehaus et al. 2014). We performed a phylogenetic analysis of the genes *aps1* encoding an NRPS, *aps5* encoding a transcription factor, *aps10* and *aps11* encoding a fatty acid synthase to investigate a scenario of horizontal gene transfer.

Both the individual gene trees and a concatenated tree (with *aps1*, *aps5*, and *aps11*) showed that the genes follow the species tree phylogeny except for *F. avenaceum* (fig. 4). The phylogeny of *aps10* included a homologous gene of *F. acuminatum*, which together with *F. avenaceum*, is part of the *Fusarium tricinctum* species complex. The phylogeny of *aps10* diverged from the species tree, with *F. asiaticum* and *F. sporotrichioides* clustering together. The apicidin amino acid sequences of *F. asiaticum* showed overall closer identity to *F. sporotrichioides* than to *F. langsethiae* or other species (supplementary table S4, Supplementary Material online). We found codon usage differences between the full genome and the genes composing the apicidin cluster in *F. asiaticum*, *F. sporotrichioides*, and *F. langsethiae*, however, no difference was found between the three species (supplementary table S5, Supplementary Material online). An analysis of gene cluster synteny showed that the *F. avenaceum* gene cluster is missing the genes *aps12*, *aps6*, and *aps3* and underwent a drastic gene order rearrangement compared with the other species. The phylogeny of g666 showed the presence of divergent paralogues in *F. avenaceum*. The rearrangement and divergency may be the consequence of a partial gene cluster duplication and may have led to a neofunctionalization of the gene cluster in *F. avenaceum*. The discontinuous taxonomic distribution and codon usage could be suggestive of a horizontal gene transfer event from *F. sporotrichioides* to *F. asiaticum*. However, multiple independent losses across the *Fusarium* genus combined with a possible advantage to maintain the cluster in the *F. asiaticum* strain FasiR2 could explain the observed patterns as well (fig. 4).

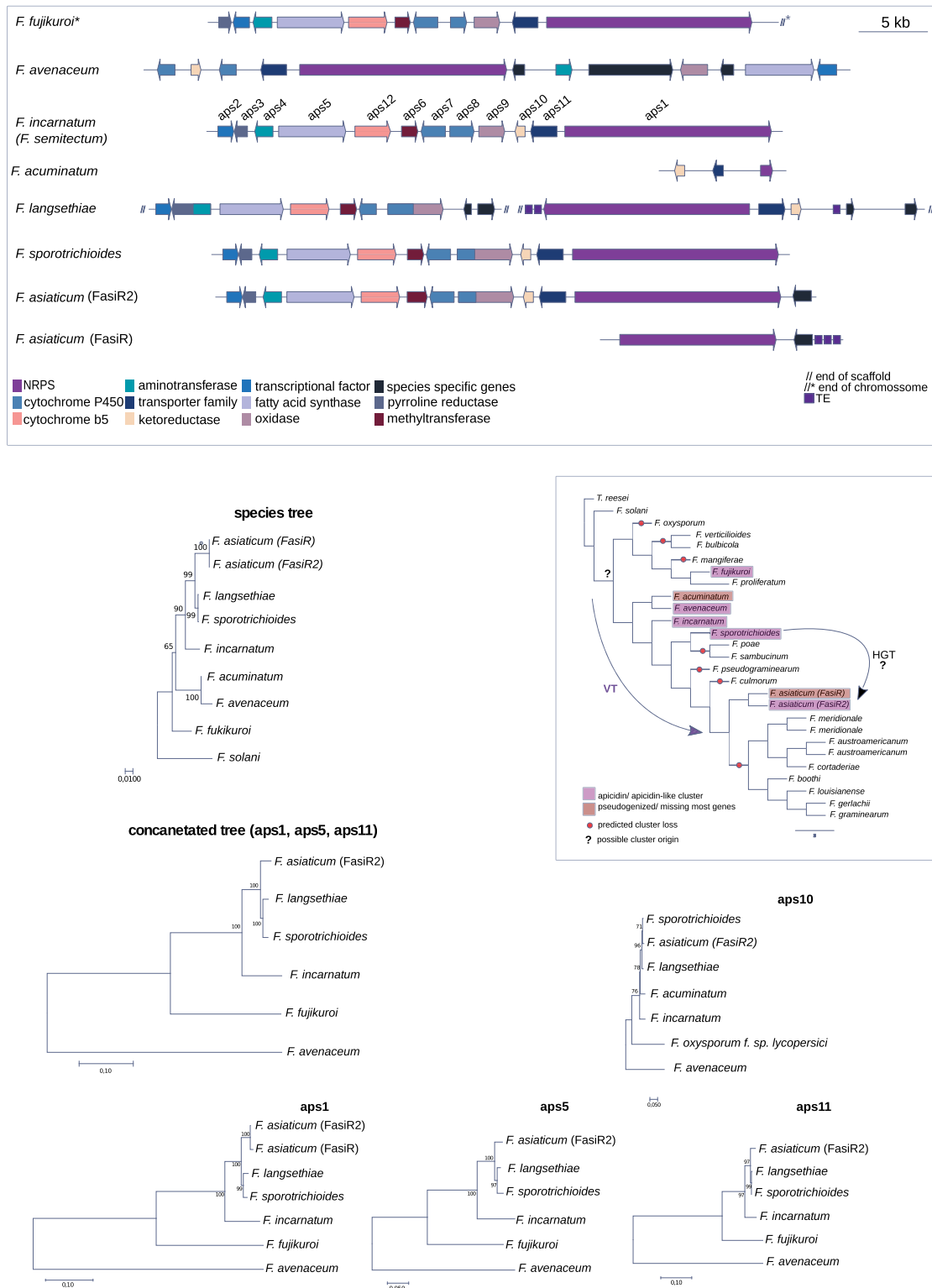


FIG. 4.—Syntenic plot of the SM46 apicidin metabolite gene cluster. Arrows of identical color correspond to homologous genes and identify the predicted protein function. **Fusarium fujikuroi* is an apicidin-F producer. Phylogenetic trees were constructed using maximum-likelihood and the JTT matrix-based amino acid model with 1,000 bootstrap replicates. The species tree was based on the concatenated analysis of the EF-1a, RBP1, and RBP2 genes. *Fusarium solani* was used as the outgroup. Gray boxes indicate the presence, independent loss, and possible origin of the apicidin cluster. HGT, horizontal gene transfer; VT, vertical transmission.

Signatures Consistent with Multiple Horizontal Gene Transfer Events

We found phylogenetic patterns consistent with a recent horizontal transfer of six genes among fungi and a single ancient bacterial transfer event in the formation of the SM54 gene cluster. The rare cluster (category 3), with a predicted size of 11 genes, was found in the FGSC strain *F. austroamericanum* (Faus154). Across *Fusarium* species, six genes of the cluster are shared with *F. avenaceum* (fig. 5). Of the six genes, the backbone gene encoding the PKS, a cytochrome P450 and a methyltransferase gene share homology with the genes *fdsS*, *fdsH*, and *fdsD*, respectively, constituting the Fusaridione A cluster in *F. heterosporum*. A homology search of the genes shared between *F. austroamericanum* and *F. avenaceum* showed *F. avenaceum* to be the only hit with a high percentage of identity (>80%) to the analyzed genes (supplementary table S6, Supplementary Material online). The phylogenetic analyses of the six genes, consistently grouped *F. austroamericanum* with *F. avenaceum*. This clustering was conserved if the tree included also orthologs found in *F. heterosporum*, which is a species more closely related to *F. avenaceum* than *F. austroamericanum* (fig. 5). The phylogenetic distribution of the gene cluster and high homology suggest that at least a segment of the cluster was horizontally transferred from the *F. avenaceum* lineage to *F. austroamericanum* to create the SM54 gene cluster.

Interestingly, a second gene of the SM54 cluster (Faus154_g659), encoding a NAD(P)/FAD-binding protein was gained most likely through horizontal transfer from bacteria. A homology search identified a homolog in the Actinobacteria *Streptomyces antibioticus* with 44.3% identity and 57.4% similarity followed by several other *Streptomyces* spp. strains as the next best hits (supplementary table S6, Supplementary Material online). The homologs in *F. austroamericanum* and *S. antibioticus* share the same NAD(P)/FAD-binding domains (supplementary fig. S2, Supplementary Material online). Among fungi, hits to the *F. austroamericanum* homolog were of lower percentage identity, the best hit was found in the Eurotiomycetes *Aspergillus wentii* with 40.6% identity (supplementary table S6, Supplementary Material online). A constrained search within the Sordariomycetes (including *F. austroamericanum*) revealed a hit in *Metarhizium robertsi* with 43.2% identity and 57.1% similarity (supplementary table S7, Supplementary Material online). The search for *S. antibioticus* homologs among eukaryotes identified a high identity (>67%) and similarity (>78%) hits in *Aspergillus* species and weaker hits in other members of the Eurotiomycetes and Sordariomycetes (supplementary table S8, Supplementary Material online). This is indicative of a horizontal transfer event between an ancestor of *Streptomyces* and most likely Pezizomycotina. Even though Faus154_g659 has no clear homologs, the lack of close orthologues in other fungi of the same class

(Sordariomycetes), the phylogenetic incongruences, and the amino acid similarity and functional homology from bacteria is consistent with an ancient bacterial origin of this gene via a horizontal transfer event.

Gene Cluster Reconfiguration across Diverse Fungi

The cluster SM53 is shared among two FGSC strains, *F. cortaderiae* (strain Fcor153) and *F. austroamericanum* (strain Faus151). In the second *F. austroamericanum* strain (Faus154), the cluster is missing most genes and suffered pseudogenization (fig. 6). We conducted a broad homology search across fungi and found SM53 to be present in *F. bulbicola*, which is not a member of the FGSC. In *F. bulbicola*, the core gene set clusters with at least six additional genes that are typically associated with a fumonisin gene cluster including a cytochrome P450 homolog identified as the fumonisin gene *cpm1*. Even though *F. bulbicola* has the capacity to produce fumonisin C, the specific strain analyzed here was shown to be a non-producer (Proctor et al. 2013). To investigate possible gaps in the genome assembly near the gene cluster, we searched the *F. bulbicola* genome for additional fumonisin genes. We analyzed homology at the nucleotide and amino acid level between *F. bulbicola* and the *F. oxysporum* strain RFC O-1890. RFC O-1890 is a fumonisin C producer (Proctor et al. 2008) and the most closely related available strain to *F. bulbicola* (supplementary table S9, Supplementary Material online). We identified fumonisin cluster elements on 4 different *F. bulbicola* scaffolds with the exception of *FUM11* and *FUM17*.

We found additional evidence for the SM53 core cluster in distantly related fungi including *Metarhizium*, *Aspergillus*, and *Zygomycetes*. The cluster variant identified in the entomopathogenic fungus *M. anisopliae* was identified as a Mapks12 cluster (Sbaraini et al. 2016). Although, the full cluster size in *M. anisopliae* is still unknown, transcriptomic data showed expression of the gene encoding the PKS and adjacent genes in culture media (Sbaraini et al. 2016). In the wheat pathogen *Z. tritici*, the core gene set is forming a larger functional cluster and transcriptomic data shows coordinated upregulation, and high expression upon infection of wheat (Palma-Guerrero et al. 2016). Phylogenetic analyses of the backbone gene encoding a PKS showed broad congruence with the species tree consisted with long-term maintenance despite widespread losses in other species (supplementary fig. S3, Supplementary Material online). The highly conserved core cluster segment may constitute a functional cluster because it encodes a typical complement of cluster functions including a PKS, a cytochrome P450, a dehydrogenase, a methyltransferase, a transcription factor, and a major facilitator superfamily transporter.

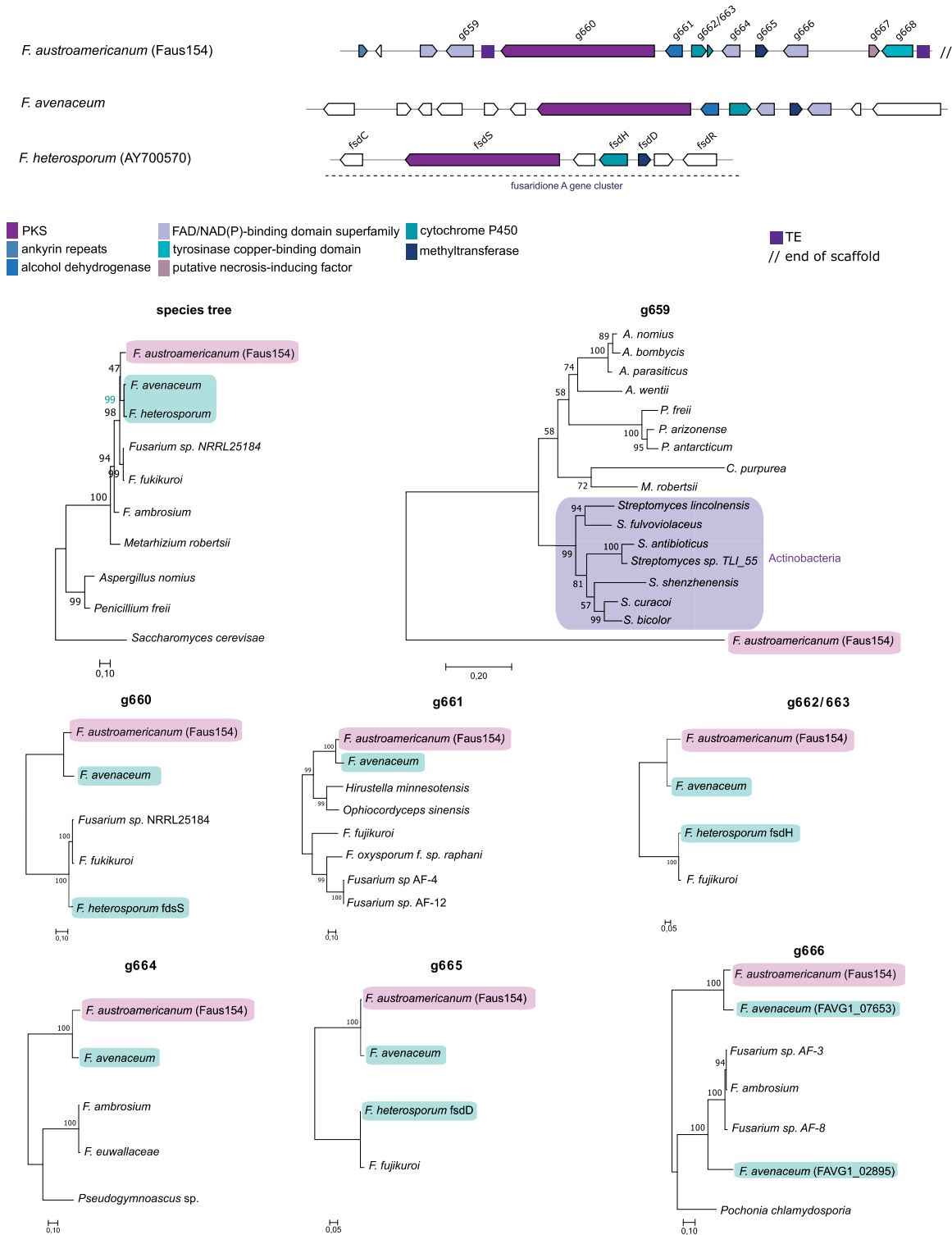


FIG. 5.—Synteny plot of the SM54 gene cluster. Arrows of identical color correspond to homologous genes and identify the predicted protein function. White arrows identify genes without a homolog in corresponding strain. Phylogenetic trees were built using maximum likelihood and the JTT matrix-based model with 1,000 bootstrap replicates. The species tree was based on the concatenated genes EF-1 α , RPB1, and RPB2. *Saccharomyces cerevisiae* was used as the outgroup.

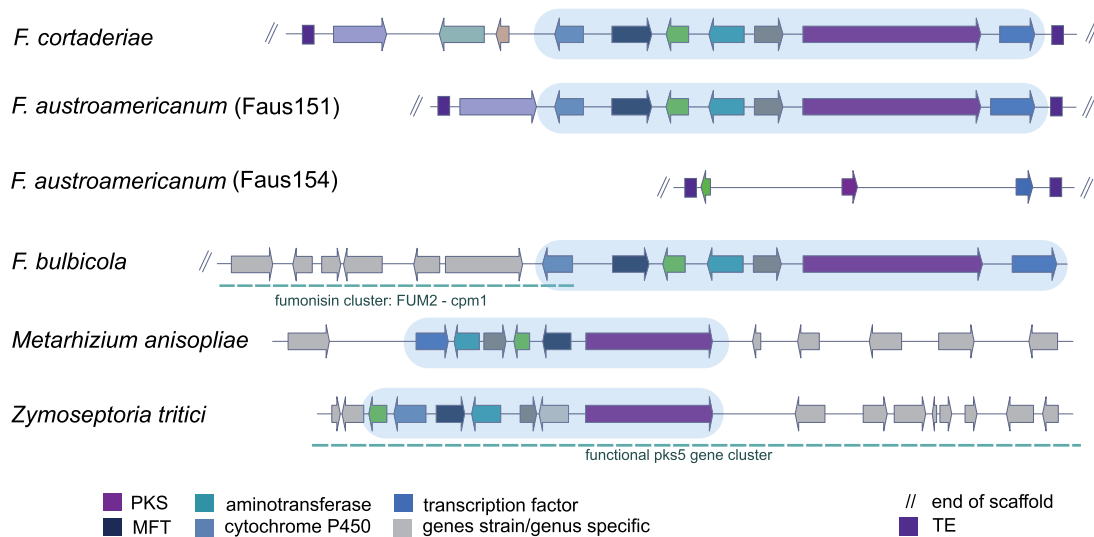


Fig. 6.—Synteny plot of the SM53 gene cluster. Arrows of identical color correspond to homologous genes and identify the predicted protein function. Light gray arrows correspond to genes lacking homology among analyzed strains. Light blue identifies the conserved core set of genes. Blue dotted lines in *Fusarium bulbicola* correspond to the fumonisin cluster adjacent to the core set and in *Zymoseptoria tritici* to the PKS5 gene cluster upregulated during infection in wheat. PKS, polyketide synthase; MFT, major facilitator superfamily transporter; TE, transposable element.

Transposable Elements Associated with Gene Cluster Rearrangements

We found evidence for the gene cluster SM48 in four different species of the FGSC (*F. cortaderiae*, *F. austroamericanum*, *F. meridionale*, and *F. asiaticum*). In *F. graminearum* s.s., the PKS backbone gene is absent. However, we found evidence for five additional genes of SM48 in four different chromosomal locations and two different chromosomes (fig. 7). A gene encoding a homeobox-like domain protein, a transporter gene, and the flanking genes clustered together on chromosome 2, but in two different loci at ~60 and 50 kb from each other, respectively. The gene encoding the glycosyl hydrolase, which is next to the backbone gene encoding the PKS in the canonical SM48 gene cluster configuration, was found as an individual gene in the subtelomeric region of chromosome 4. *F. avenaceum* is the only analyzed species outside the FGSC that shared the PKS gene (fig. 7). Interestingly, the SM48 gene cluster contained a series of transposable elements integrated either next to the gene encoding the PKS and/or the gene encoding the glycosyl hydrolase. Furthermore, a phylogenetic analysis showed a patchy taxonomic distribution of homologs across the *Fusarium* genus (supplementary table S10, Supplementary Material online). The gene cluster SM48 was most likely vertically inherited by the FGSC as indicated by the patchy presence of homologs across *Fusarium* and evidence for at least segments of the cluster in *F. avenaceum*. Disrupted cluster variants are present in the clade formed by *F. graminearum* s.s., *F. boothii*, *F. louisianense*, and *F. gerlachii*. The high density of transposable elements might have facilitated the rearrangement of the gene cluster.

Transposable Element Families in the Genomic Environment of Gene Clusters

Several gene clusters of categories 2 and 3 (SM46, SM48, SM48, and SM54; fig. 2), which showed various levels of reconfigurations were flanked by transposable elements. To understand broadly how transposable elements may have contributed to gene cluster evolution, we analyzed the identity of transposable elements across the genomes and in close association with gene clusters. We found overall no difference in transposable element density in proximity to gene clusters compared with the rest of the genome with the exception of the *F. asiaticum* strain FasiR (supplementary fig. S4, Supplementary Material online). FasiR showed about twice the transposable element density in proximity to clusters (9.9%) compared with genome-wide average (4.1%). Next, we analyzed the frequency of individual transposable element families within 10 kb of gene clusters and compared this with the frequency in all 10 kb windows across the genomes of the FGSC (fig. 8A). We found a series of transposable element families that were more frequent in proximity to gene clusters (fig. 8B). The most abundant elements in the genomes of the FGSC are the unclassified elements 3-family-62 (mean frequency of 0.147 per 10 kb window) followed by 2-family-17 (mean frequency of 0.124). In proximity to SM gene clusters, the frequency of the 2-family-17 was higher than 3-family-62 in 54% of the strains, with an overall mean of 0.174 and 0.160, respectively. The element 4-family-882, which is enriched in the clade comprising *F. graminearum* s.s., *F. gerlachii*, *F. boothii*, and *F. louisianense*, as well as the strain *F. cortaderiae*, is seven times more frequent near SM gene clusters compared with

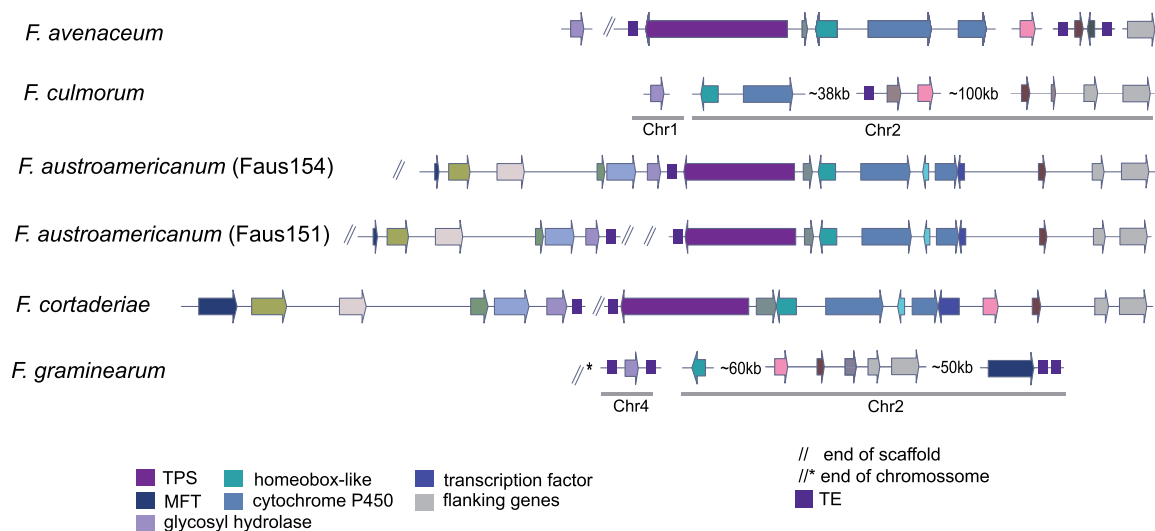


Fig. 7.—Synteny plot of the SM48 gene cluster. Arrows of identical color correspond to homologous genes and identify the predicted protein function. Values adjacent to disrupted clusters define physical distances and gray bars below genes define to chromosomal locations. TPS, terpene synthase; MFT, major facilitator superfamily transporter; TE, transposable element.

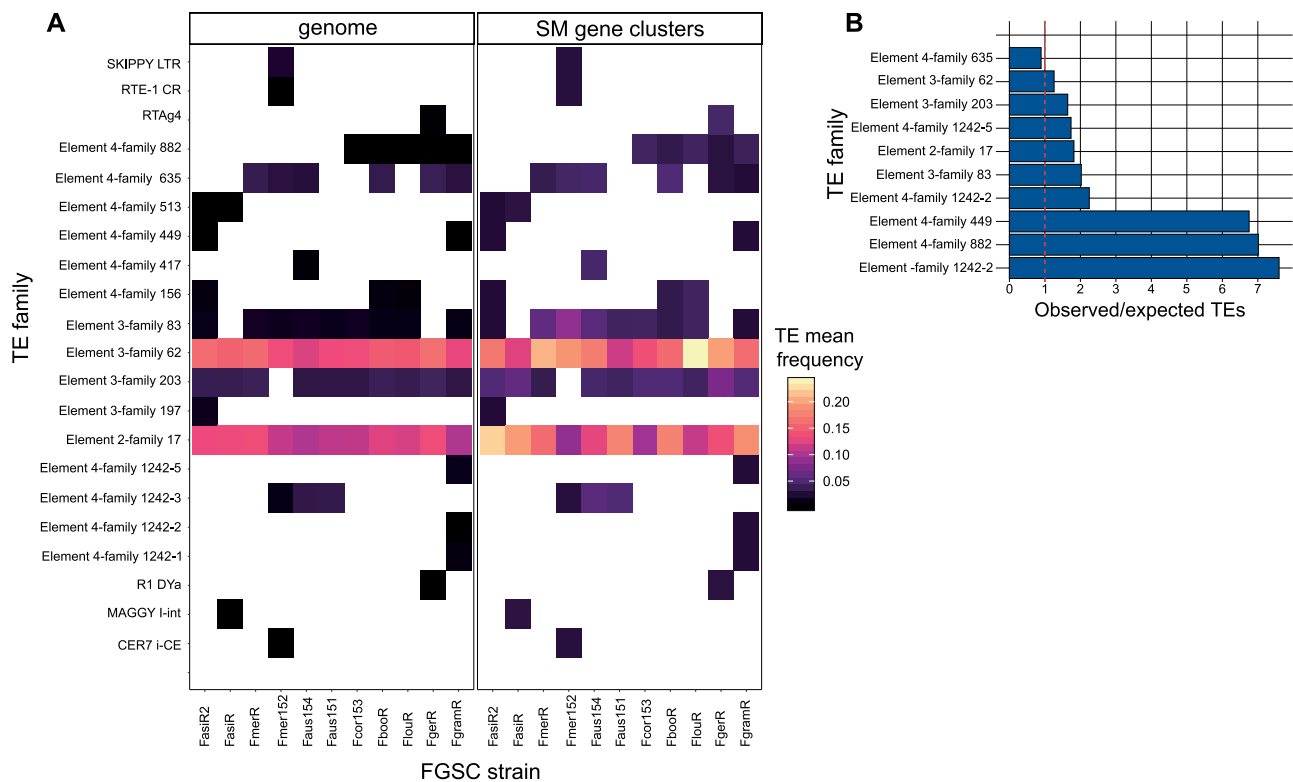


Fig. 8.—(A) Heatmap with the most frequent transposable element families flanking FGSC gene clusters and the overall genome in 10 kb windows. (B) Bar chart showing the ratio of the observed (SM gene cluster) over the expected transposable elements (genome) in the *F. graminearum* reference genome (FgramR). Red dotted line marks the ratio of one representing no difference.

the whole genome (FgramR; fig. 8B). The analyses of transposable elements in the vicinity of gene clusters do not establish a mechanistic link for gene cluster

rearrangements, but the over-representation of specific transposable elements raises intriguing questions about the unique genomic environment of gene clusters.

Discussion

We assembled and analyzed a comprehensive set of genomes representative of the FGSC diversity. Our phylogenomic analyses corroborated previous multilocus studies and refined our understanding of the evolutionary relationships within the complex (O'Donnel et al. 2004; Aoki et al. 2012). The recent speciation among members of the FGSC led to differentiation in host range, genome size, gene and transposable element content. Our analyses of SM gene clusters within the FGSC revealed more complexity than previously reported (Walkowiak et al. 2016). Individual gene clusters underwent independent gene losses, sequence rearrangements associated with transposable elements and multiple horizontal transfer events, leading to the presence/absence polymorphism and chemical diversity within the FGSC.

A Diverse SM Gene Cluster Pangenome of the FGSC

We performed pangenome analyses of 8 species of FGSC (11 isolates) to exhaustively characterize the presence of known and unknown SM gene clusters. The emergence of the FGSC was accompanied by the loss and rearrangement of several SM gene clusters. The most recent common ancestor with other members of the *Fusarium* clade likely carried more SM gene clusters. The recently lost clusters may underlie the adaptation to wheat as a primary host. Among the fully conserved gene clusters within the FGSC, we found clusters underlying the production of siderophores including triacetylfusarin and ferricrocin that facilitate iron acquisition (Charlang et al. 1981). We also found conserved clusters underlying the production of virulence factors, for example, gramillin on maize (Bahadoor et al. 2018). The conservation likely reflects the essential functions of these metabolites in the life cycle of the fungi. The SM gene clusters not fixed within the FGSC spanned a surprisingly broad number of types including TPS, NRPS, NRPS-TPS, and NRPS-PKS. Segregating gene clusters may reflect adaptation to niches specific to a subset of the FGSC. Such adaptation may explain the conservation of the apicidin cluster in the *F. asiaticum* strain FasiR2 isolated from maize and the lack of the cluster in the strain FasiR isolated from barley (O'Donnel et al. 2000).

How the environmental heterogeneity selects for diversity in SM gene clusters among closely related species is poorly understood, yet studies have found strong associations of SM gene clusters with different lifestyles and geographical distribution (Reynolds et al. 2017; Wollenberg et al. 2019). The fusaristatin A gene cluster, thought to be missing in *F. pseudograminearum* (but present in FGSC), was recently found to be functional in a Western Australian population of *F. pseudograminearum* (Wollenberg et al. 2019). In FGSC, trichothecenes are key adaptations to exploit the host. Different forms of trichothecenes (i.e. deoxynivalenol, 3-acetyldeoxynivalenol, 15-acetyldeoxynivalenol, and nivalenol chemotypes) are segregating in pathogen populations

due to balancing selection (Ward et al. 2002). The trichothecene polymorphism is likely adaptive with the role in pathogenesis depending both on the crop host (Desjardins et al. 1992; Proctor et al. 2002; Cuzick et al. 2008) and the specific trichothecene produced (Carter et al. 2002, Ponts et al. 2009; Spolti et al. 2012). For example, nivalenol production is associated with pathogenicity on maize and deoxynivalenol is essential to *Fusarium* head blight in wheat spikelets but seems to play no role for pathogenicity on maize (Maier et al. 2006). Both toxins play no role in pathogenicity on barley. A variable pangenome of metabolic capacity maintained among members of the FGSC may, hence, also serve as a reservoir for adaptive introgression among species.

Mechanisms Generating Chemical Diversity in *Fusarium*

Our study revealed a complex set of mechanisms underlying SM gene cluster diversity in FGSC. We found that multiple independent losses are a key factor generating extant cluster diversity within the FGSC and *Fusarium*. The SM43 (guaia,6-10(14)-diene) and the apicidin clusters were lost multiple times within *Fusarium* and in different lineages of the FGSC. Independent losses are frequently associated with the evolutionary trajectory of SM gene clusters (Patron et al. 2007; Khaldi et al. 2008). The evolution of the galactose cluster in yeasts was characterized by multiple independent losses and at least 11 times among the subphyla of *Saccharomycotina* and *Taphrinomycotina* (Riley et al. 2016). Similarly, Campbell et al. (2012) showed that the bikaverin gene cluster was repeatedly lost in the genus *Botrytis* after receiving the cluster horizontally from a putative *Fusarium* donor. A gene cluster loss is typically favored by either a decreased benefit to produce the metabolite or an increase in production costs (Rokas et al. 2018). Along these lines, the *black queen hypothesis* conveys the idea that the loss of a costly gene (cluster) can provide a selective advantage by conserving an organism's limited resources (Morris et al. 2012). Such loss-of-function mutations (e.g. abolishing metabolite production) are viable in an environment where other organisms ensure the same function (Morris et al. 2012; Mas et al. 2016). The *black queen hypothesis* may at least partially explain the metabolite diversity and high level of cluster loss in the FGSC if different lineages and species frequently coexist in the same environment or host.

Horizontal gene transfer is an important source of gene cluster gain in fungi (Khaldi et al. 2008; Khaldi and Wolfe 2011; Slot and Rokas 2011; Campbell et al. 2012) and likely contributed to the FGSC gene cluster diversity. Here, we report an unusual case of multiple, independent horizontal transfer events involving an ancient transfer from bacteria and a more recent fungal donor. The horizontal transfer contributed to the formation of the SM54 gene cluster found in the strain *F. austroamericanum* (Faus154). Horizontal transfer events have been proposed as an important form of

pathogenicity emergence. A gene cluster of *F. pseudograminearum* was most likely formed by three horizontally acquired genes from other pathogenic fungi. An additional gene of the cluster encoding an amidohydrolase was received from a plant-associated bacterial donor and associated with pathogenicity on wheat and barley (Gardiner et al. 2012). Similarly, the *Metarhizium* genus of entomopathogens acquired at least 18 genes by independent horizontal transfer events that contribute to insect cuticle degradation (Zhang et al. 2019).

Our analyses revealed the SM53 gene cluster core segment that is conserved across distantly related genera. The core section underlies the formation of superclusters through the rearrangement with a separate cluster and likely led to neofunctionalization. The backbone and adjacent genes in the conserved segment were found to be expressed in *M. anisopliae* in culture medium (Sbaraini et al. 2016). In the wheat pathogen *Z. tritici*, the core segment was associated with additional genes forming a larger cluster with coordinated upregulation upon host infection (Palma-Guerrero et al. 2016). A study in *A. fumigatus* identified a similar event, where the clusters underlying pseurotin and fumagillin production were rearranged to form a supercluster (Wiemann et al. 2013). Similar to the gene cluster SM53, the segments of the supercluster were conserved in *A. fischeri* and in the more distantly related species *M. robertsii*. Taxonomically widespread conserved gene cluster segments may represent functional but transitory gene cluster variants that can give rise to superclusters. Viable, transitory stages are an efficient route to evolve new metabolic capacity across fungi (Lind et al. 2017; Rokas et al. 2018).

Transposable Elements as Possible Drivers of Gene Cluster Rearrangements

Our analyses revealed that gene cluster gains and losses in the FGSC may be influenced by the presence of specific transposable elements. We found an enrichment in transposable elements adjacent or integrated within different clusters (i.e. SM1, SM21, SM48, SM53, and SM54). Our data suggest that the cluster SM48 emerged within FGSC and may have suffered transposable element-associated chromosomal rearrangements in the *F. graminearum* s.s. clade followed by functional loss. The SM53 pseudogenization and gene loss in the *F. austroamericanum* strain Faus154 was coinciding with transposable element insertions adjacent to the cluster. Transposable elements play an important role in the evolution of fungal pathogens (Gardiner et al. 2013; Fouché et al. 2018; Sánchez-Vallet et al. 2018). Transposable elements can induce gene cluster rearrangements due to nonhomologous recombination among repeat copies (Boutanaev and Osbourn 2018), but also impact genome structure and function by causing gene inactivation, copy number variation, and expression polymorphism (Manning et al. 2013;

Hartmann et al. 2017; Krishnan et al. 2018). For example, flanking transposable elements likely caused transposition events of a specialized cluster in *A. fumigatus* (Lind et al. 2017). The enriched transposable elements near gene clusters in FGSC genomes were likely overall an important driver of gene cluster loss, rearrangement, and neofunctionalization.

Our study provides insights into the evolutionary origins of SM gene clusters in a complex of closely related species. The recency of speciation within the FGSC is reflected by the predominant number of conserved gene clusters. Nevertheless, the FGSC accumulated previously underappreciated gene cluster diversity, which originated from a broad spectrum of mechanisms including parallel gene losses, rearrangements and horizontal acquisition. Independent losses within the complex were likely due to ecological drivers and strong selection. Hence, environmental heterogeneity may play an important role in gene cluster evolution (Rokas et al. 2018). Chromosomal rearrangements underlying cluster loss were often complex and were likely facilitated by transposable elements. At the same time, chromosomal rearrangements contributed to gene cluster neofunctionalization. The extant chemical diversity of FGSC highlights the importance of transitory stages in the evolution of specialized metabolism among very closely related species.

Supplementary Material

Supplementary data are available at Genome Biology and Evolution online.

Acknowledgments

We thank Dr Robert Proctor from the National Center for Agricultural Utilization Research (United States Department of Agriculture) for kindly providing the genomic sequences of *F. bulbicola*. This research was supported by FAPESP (Fundação de Amparo a Pesquisa do Estado de São Paulo) grant process 2017/22369-7 and 2016/04364-5. D.C. receives support from the Swiss National Science Foundation (grants 31003A_173265 and IZCOZO_177052).

Authors' Contributions

S.M.T., L.O.R., B.C., and D.C. conceived the study; S.M.T., L.O.R., and B.C. provided samples and data sets; S.M.T. and U.O. analyzed the data; S.M.T. and D.C. wrote the manuscript; L.O.R., U.O., and B.C. edited the manuscript.

Literature Cited

- Rambaut A. 2012. Figtree. Available from: <http://tree.bio.ed.ac.uk/software/figtree/>; last accessed April 01, 2019.
- Altschul S. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25(17):3389–3402.

- Andrews S. 2010. FastQC: a quality control tool for high throughput sequence data. Available from: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>; last accessed October 12, 2018.
- Aoki T, Ward TJ, Kistler HC, O'Donnell K. 2012. Systematics, phylogeny and trichothecene mycotoxin potential of fusarium head blight cereal pathogens. *Mycotoxins* 62(2):91–102.
- Bahadoor A, et al. 2018. Gramillin A and B: cyclic lipopeptides identified as the nonribosomal biosynthetic products of *Fusarium graminearum*. *J Am Chem Soc.* 140(48):16783–16791.
- Bankevich A, et al. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol.* 19(5):455–477.
- Blin K, et al. 2017. antiSMASH 4.0—improvements in chemistry prediction and gene cluster boundary identification. *Nucleic Acids Res.* 45(W1):W36–W41.
- Bottalico A. 1998. Fusarium diseases of cereals: species complex and related mycotoxin profiles, in Europe. *J Plant Pathol.* 80:85–103.
- Boutanaev AM, Osbourn AE. 2018. Multigenome analysis implicates miniature inverted-repeat transposable elements (MITEs) in metabolic diversification in eudicots. *Proc Natl Acad Sci U S A.* 115(28):E6650–E6658.
- Brakhage AA. 1998. Molecular regulation of beta-lactam biosynthesis in filamentous fungi. *Microbiol. Microbiol Mol Biol Rev.* 62(3):547–585.
- Brakhage AA. 2013. Regulation of fungal secondary metabolism. *Nat Rev Microbiol.* 11(1):21–32.
- Brown DW, Proctor RH. 2016. Insights into natural products biosynthesis from analysis of 490 polyketide synthases from *Fusarium*. *Fungal Genet Biol.* 89:37–51.
- Burkhardt I, et al. 2016. Mechanistic characterization of two sesquiterpene cyclases from the plant pathogen *Fusarium fujikuroi*. *Angew Chem Int Ed.* 55(30):8748–8751.
- Camacho C, et al. 2009. BLAST+: architecture and applications. *BMC Bioinformatics.* 10(1):421.
- Campbell MA, Rokas A, Slot JC. 2012. Horizontal transfer and death of a fungal secondary metabolic gene cluster. *Genome Biol Evol.* 4(3):289–293.
- Carter JP, et al. 2002. Variation in pathogenicity associated with the genetic diversity of *Fusarium graminearum*. *Eur J Plant Pathol.* 108(6):573–583.
- Carver T, Bleasby A. 2003. A. The design of Jemboss: a graphical user interface to EMBOSS. *Bioinformatics* 19(14):1837–1843.
- Charlang G, Ng B, Horowitz NH, Horowitz RM. 1981. Cellular and extracellular siderophores of *Aspergillus nidulans* and *Penicillium chrysogenum*. *Mol Cell Biol.* 1(2):94–100.
- Cuzick A, Urban M, Hammond-Kosack K. 2008. *Fusarium graminearum* gene deletion mutants map1 and tri5 reveal similarities and differences in the pathogenicity requirements to cause disease on *Arabidopsis* and wheat floral tissue. *New Phytologist.* 177(4):990–1000.
- Darkin-Rattray SJ, et al. 1996. Apicidin: a novel antiprotozoal agent that inhibits parasite histone deacetylase. *Proc Natl Acad Sci U S A.* 93(23):13143–13147.
- Desjardins AE. 2006. *Fusarium* mycotoxins: chemistry, genetics, and biology. St. Paul, MN: American Phytopathological Society (APS Press).
- Desjardins AE, Hohn TM, McCormick SP. 1992. Effect of gene disruption of trichodiene synthase on the virulence of *Gibberella pulicaris*. *Mol Plant-Microbe Interact.* 5(2):4–222.
- Fouché S, Plissonneau C, Croll D. 2018. The birth and death of effectors in rapidly evolving filamentous pathogen genomes. *Curr Opin Microbiol.* 46:34–42.
- Gardiner DM, et al. 2012. Comparative pathogenomics reveals horizontally acquired novel virulence genes in fungi infecting cereal hosts. *PLoS Pathog.* 8(9):e1002952.
- Gardiner DM, Kazan K, Manners JM. 2013. Cross-kingdom gene transfer facilitates the evolution of virulence in fungal pathogens. *Plant Sci.* 210:151–158.
- Gurevich A, Saveliev V, Vyahhi N, Tesler G. 2013. QUAST: quality assessment tool for genome assemblies. *Bioinformatics* 29(8):1072–1075.
- Guy L, Roat Kultima J, Andersson S. 2010. genoPlotR: comparative gene and genome visualization in R. *Bioinformatics* 26(18):2334–2335.
- Hartmann FE, Sánchez-Vallet A, McDonald BA, Croll D. 2017. A fungal wheat pathogen evolved host specialization by extensive chromosomal rearrangements. *ISME J.* 11(5):1189–1204.
- Higgins DG, Sharp PM. 1988. CLUSTAL: a package for performing multiple sequence alignment on a microcomputer. *Gene* 73(1):237–244.
- Hoogendoorn K, et al. 2018. Evolution and diversity of biosynthetic gene clusters in *Fusarium*. *Front Microbiol.* 9:1158.
- Jin J-M, et al. 2010. Functional characterization and manipulation of the apicidin biosynthetic pathway in *Fusarium semitectum*. *Mol Microbiol.* 76(2):456–466.
- Jones DT, Taylor WR, Thornton JM. 1992. The rapid generation of mutation data matrices from protein sequences. *Comput Appl Biosci.* 8:275–282.
- Jones P, et al. 2014. InterProScan 5: genome-scale protein function classification. *Bioinformatics* 30(9):1236–1240.
- Kall L, Krogh A, Sonnhammer E. 2007. Advantages of combined transmembrane topology and signal peptide prediction—the Phobius web server. *Nucleic Acids Res.* 35(Web Server):W429–W432.
- Katoh K, Rozewicki J, Yamada KD. 2017. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Brief Bioinform.* doi:10.1093/bib/bbx108
- Keller NP, Turner G, Bennett JW. 2005. Fungal secondary metabolism—from biochemistry to genomics. *Nat Rev Microbiol.* 3(12):937–947.
- Khalidi N, Collemare J, Lebrun M-H, Wolfe KH. 2008. Evidence for horizontal transfer of a secondary metabolite gene cluster between fungi. *Genome Biol.* 9(1):R18.
- Khalidi N, Wolfe KH. 2011. Evolutionary origins of the fumonisin secondary metabolite gene cluster in *Fusarium verticillioides* and *Aspergillus niger*. *Int J Evol Biol.* 2011:1.
- Krishnan P, et al. 2018. Transposable element insertions shape gene regulation and melanin production in a fungal pathogen of wheat. *BMC biology.* 6(1):78.
- Krogh A, Larsson B, von Heijne G, Sonnhammer EL. 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol.* 305(3):567–580.
- Kumar S, Stecher G, Tamura K. 2016. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol Biol Evol.* 33(7):1870–1874.
- Li L, Stoeckert CJ, Roos DS. 2003. OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res.* 13(9):2178–2189.
- Lind AL, et al. 2017. Drivers of genetic diversity in secondary metabolic gene clusters within a fungal species. *PLOS Biol.* 15(11):e2003583.
- Lysøe E, et al. 2016. Draft genome sequence and chemical profiling of *Fusarium langsethiae*, an emerging producer of type A trichothecenes. *Int J Food Microb.* 221:29–36.
- Manning VA, et al. 2013. Comparative genomics of a plant-pathogenic fungus, *Pyrenophora tritici-repentis*, reveals transduplication and the impact of repeat elements on pathogenicity and population divergence. *G3: Genes, Genomes, Genetics.* 3(1):41–63.
- Maier FJ, et al. 2006. Involvement of trichothecenes in fusarioses of wheat, barley and maize evaluated by gene disruption of the trichodiene synthase (Tri5) gene in three field isolates of different chemotype and virulence. *Mol Plant Pathol.* 7(6):449–461.
- Mas A, Jamshidi S, Lagadeuc Y, Eveillard D, Vandenkoornhuysen P. 2016. Beyond the Black Queen Hypothesis. *ISME J.* 10(9):2085–2091.
- Morgulis A, et al. 2008. Database indexing for production MegaBLAST searches. *Bioinformatics* 24(16):1757–1764.
- Morris JJ, Lenski RE, Zinser ER. 2012. The Black Queen Hypothesis: evolution of dependencies through adaptive gene loss. *MBio.* 3(2):e00036–12.

- Niehaus E-M, et al. 2014. Apicidin F: characterization and genetic manipulation of a new secondary metabolite gene cluster in the rice pathogen *Fusarium fujikuroi*. *PLoS One*. 9(7):e103336.
- Nosanichuk JD, Casadevall A. 2006. Impact of melanin on microbial virulence and clinical resistance to antimicrobial compounds. *Antimicrob Agents Chemother*. 50(11):3519–3528.
- O'Donnell K, Kistler HC, Tacke BK, Casper HH. 2000. Gene genealogies reveal global phylogeographic structure and reproductive isolation among lineages of *Fusarium graminearum*, the fungus causing wheat scab. *Proc Natl Acad Sci U S A*. 97:7905–7910.
- O'Donnell K, Ward TJ, Geiser DM, Corby Kistler H, Aoki T. 2004. Genealogical concordance between the mating type locus and seven other nuclear genes supports formal recognition of nine phylogenetically distinct species within the *Fusarium graminearum* clade. *Fungal Genet Biol*. 41:600–623.
- Osborn A. 2010. Secondary metabolic gene clusters: evolutionary toolkits for chemical innovation. *Trends Genet*. 26(10):449–457.
- Palma-Guerrero J, et al. 2016. Comparative transcriptomic analyses of *Zymoseptoria tritici* strains show complex lifestyle transitions and intra-specific variability in transcription profiles. *Mol Plant Pathol*. 17(6):845–859.
- Patron NJ, et al. 2007. Origin and distribution of epipolythiodioxopiperazine (ETP) gene clusters in filamentous ascomycetes. *BMC Evol Biol*. 7(1):174.
- Petersen TN, Brunak S, von Heijne G, Nielsen H. 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat Methods* 8(10):785–786.
- Ponts N, et al. 2009. *Fusarium* response to oxidative stress by H₂O₂ is trichothecene chemotype-dependent. *FEMS Microbiol Lett*. 293(2):255–262.
- Proctor RH, et al. 2013. Birth, death and horizontal transfer of the fumonisin biosynthetic gene cluster during the evolutionary diversification of *Fusarium*. *Mol Microbiol*. 90(2):290–306.
- Proctor RH, Busman M, Seo J-A, Lee YW, Plattner RD. 2008. A fumonisin biosynthetic gene cluster in *Fusarium oxysporum* strain O-1890 and the genetic basis for B versus C fumonisin production. *Fungal Genet Biol*. 45(6):1016–1026.
- Proctor RHH, et al. 2002. Genetic analysis of the role of trichothecene and fumonisin mycotoxins in the virulence of *Fusarium*. *Eur J Plant Pathol*. 108(7):691–698.
- Puri KD, Zhong S. 2010. The 3ADON population of *Fusarium graminearum* found in North Dakota is more aggressive and produces a higher level of DON than the prevalent 15ADON population in spring wheat. *Phytopathology* 100(10):1007–1014.
- Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26(6):841–842.
- Reynolds HT, et al. 2017. Differential retention of gene functions in a secondary metabolite cluster. *Mol Biol Evol*. 34(8):2002–2015.
- Reynolds HT, et al. 2018. Horizontal gene cluster transfer increased hallucinogenic mushroom diversity. *Evol Lett*. 2(2):88–101.
- Riley R, et al. 2016. Comparative genomics of biotechnologically important yeasts. *Proc Natl Acad Sci U S A*. 113(35):9882–9887.
- Rokas A, Wisecaver JH, Lind AL. 2018. The birth, evolution and death of metabolic gene clusters in fungi. *Nat Rev Microbiol*. 16:731–744.
- Sánchez-Vallet A, et al. 2018. The genome biology of effector gene evolution in filamentous plant pathogens. *Annu Rev Phytopathol*. 56(1):21–40.
- Sbaraini N, et al. 2016. Secondary metabolite gene clusters in the entomopathogen fungus *Metarhizium anisopliae*: genome identification and patterns of expression in a cuticle infection model. *BMC Genomics* 17(S8):736.
- Sieber CMK, et al. 2014. The *Fusarium graminearum* genome reveals more secondary metabolite gene clusters and hints of horizontal gene transfer. *PLoS One*. 9(10):e110311.
- Slot JC, Rokas A. 2010. Multiple GAL pathway gene clusters evolved independently and by different mechanisms in fungi. *Proc Natl Acad Sci U S A*. 107(22):10136–10141.
- Slot JC, Rokas A. 2011. Horizontal transfer of a large and highly toxic secondary metabolic gene cluster between fungi. *Curr Biol*. 21(2):134–139.
- Smit A, Hubley R. 2008. 2010 RepeatModeler Open-1.0. Available from: <http://www.repeatmasker.org/RepeatModeler/>; last accessed April 03, 2019.
- Smit AFA, Hubley R, Green P. 2015. RepeatMasker Open-4.0. Available from: <http://www.repeatmasker.org>; last accessed April 03, 2019.
- Sonnhammer ELL, Durbin R. 1995. A dot-matrix program with dynamic threshold control suited for genomic DNA and protein sequence analysis. *Gene* 167(1–2):GC1–GC10.
- Sperschneider J, Dodds PN, Gardiner DM, Singh KB, Taylor JM. 2018. Improved prediction of fungal effector proteins from secretomes with EffectorP 2.0. *Mol Plant Pathol*. 19(9):2094–2110.
- Spolti P, Barros NC, Gomes LB, dos Santos J, Del Ponte EM. 2012. Phenotypic and pathogenic traits of two species of the *Fusarium graminearum* complex possessing either 15-ADON or NIV genotype. *Eur J Plant Pathol*. 133(3):621–629.
- Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30(9):1312–1313.
- Stanke M, Morgenstern B. 2005. AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints. *Nucleic Acids Res*. 33(Web Server):W465–W467.
- Tralamazza SM, Bemvenuti RH, Zorzete P, Garcia FS, Corrêa B. 2016. Fungal diversity and natural occurrence of deoxynivalenol and zearalenone in freshly harvested wheat grains from Brazil. *Food Chem*. 196:445–456.
- van der Lee T, Zhang H, van Diepeningen A, Waalwijk C. 2015. Biogeography of *Fusarium graminearum* species complex and chemotypes: a review. *Food Addit Contam Part A Chem Anal Control Expo Risk Assess*. 32(4):453–460.
- Walkowiak S, Rowland O, Rodrigue N, Subramaniam R. 2016. Whole genome sequencing and comparative genomics of closely related *Fusarium* head blight fungi: *Fusarium graminearum*, *F. meridionale* and *F. asiaticum*. *BMC Genomics* 17(1):1014.
- Ward TJ, et al. 2008. An adaptive evolutionary shift in *Fusarium* head blight pathogen populations is driving the rapid spread of more toxic *Fusarium graminearum* in North America. *Fungal Genet Biol*. 45(4):473–484.
- Ward TJ, Bielawski JP, Kistler HC, Sullivan E, O'Donnell K. 2002. Ancestral polymorphism and adaptive evolution in the trichothecene mycotoxin gene cluster of phytopathogenic *Fusarium*. *Proc Natl Acad Sci U S A*. 99(14):9278–9283.
- Waterhouse RM, et al. 2018. BUSCO applications from quality assessments to gene prediction and phylogenomics. *Mol Biol Evol*. 35(3):543–548.
- Wickham H. 2016. ggplot2: elegant graphics for data analysis. New York, USA: Springer.
- Wiemann P, et al. 2013. Prototype of an intertwined secondary-metabolite supercluster. *Proc Natl Acad Sci U S A*. 110(42):17065–17070.
- Wisecaver JH, Slot JC, Rokas A. 2014. The evolution of fungal metabolic pathways. *PLoS Genet*. 10(12):e1004816.
- Wollenberg RD, et al. 2019. There it is! *Fusarium pseudograminearum* did not lose the fusaristatin gene cluster after all. *Fungal Biol*. 123:10–17.
- Wong S, Wolfe KH. 2005. Birth of a metabolic gene cluster in yeast by adaptive gene relocation. *Nat Genet*. 37(7):777–782.
- Zhang C, Sayyari E, Mirarab S. 2017. ASTRAL-III: increased Scalability and Impacts of Contracting Low Support Branches. Cham: Springer. p. 53–75.

Zhang H, et al. 2012. Population analysis of the *Fusarium graminearum* species complex from wheat in China show a shift to more aggressive isolates. PLoS One. 7(2):e31722.

Zhang Q, et al. 2019. Horizontal gene transfer allowed the emergence of broad host range entomopathogens. Proc Natl Acad Sci U S A. 116(16):7982–7989.

Zhang Z, Schwartz S, Wagner L, Miller W. 2000. A Greedy algorithm for aligning DNA sequences. J Comput Biol. 7(1–2):203–214.

Associate editor: Laura Rose