"Birth, death and horizontal transfer of the fumonisin biosynthetic gene cluster during the evolutionary diversification of fusarium"

Proctor, R.H.; Van Hove, François; Susca, A.; Stea, G.; Busman, M.; van der Lee, T.; Waalwijk, C.; Moretti, A.; Ward, T.J.

Abstract
Summary: Fumonisins are a family of carcinogenic secondary metabolites produced by members of the Fusarium fujikuroi species complex (FFSC) and rare strains of Fusarium oxysporum. In Fusarium, fumonisin biosynthetic genes (FUM) are clustered, and the cluster is uniform in gene organization. Here, sequence analyses indicated that the cluster exists in five different genomic contexts, defining five cluster types. In FUM gene genealogies, evolutionary relationships between fusaria with different cluster types were largely incongruent with species relationships inferred from primary-metabolism (PM) gene genealogies, and FUM cluster types are not trans-specific. In addition, synonymous site divergence analyses indicated that three FUM cluster types predate diversification of FFSC. The data are not consistent with balancing selection or interspecific hybridization, but they are consistent with two competing hypotheses: (i) multiple horizontal transfers of the cluster from unknown donors to...

Référence bibliographique

DOI : 10.1111/mmi.12362
Birth, death and horizontal transfer of the fumonisin biosynthetic gene cluster during the evolutionary diversification of Fusarium

Robert H. Proctor, François Van Hove, Antonia Susca, Gaetano Stea, Mark Busman, Theo van der Lee, Cees Waalwijk, Antonio Moretti† and Todd J. Ward†

1 United States Department of Agriculture, Agricultural Research Service, National Center for Agricultural Utilization Research, Peoria, IL, USA.
2 Université catholique de Louvain, Earth and Life Institute, Applied Microbiology, Mycology, Mycothèque de l’Université catholique de Louvain (BCCMTM /MUCL), Louvain-la-Neuve, Belgium.
3 National Research Council, Institute of Sciences of Food Production, Bari, Italy.

Summary

Fumonisins are a family of carcinogenic secondary metabolites produced by members of the Fusarium fujikuroi species complex (FFSC) and rare strains of Fusarium oxysporum. In Fusarium, fumonisin biosynthetic genes (FUM) are clustered, and the cluster is uniform in gene organization. Here, sequence analyses indicated that the cluster exists in five different genomic contexts, defining five cluster types. In FUM gene genealogies, evolutionary relationships between fusaria with different cluster types were largely incongruent with species relationships inferred from primary-metabolism (PM) gene genealogies, and FUM cluster types are not trans-specific. In addition, synonymous site divergence analyses indicated that three FUM cluster types predate diversification of FFSC. The data are not consistent with balancing selection or interspecific hybridization, but they are consistent with two competing hypotheses: (i) multiple horizontal transfers of the cluster from unknown donors to FFSC recipients and (ii) cluster duplication and loss (birth and death). Furthermore, low levels of FUM gene divergence in F. bulbicola, an FFSC species, and F. oxysporum provide evidence for horizontal transfer of the cluster from the former, or a closely related species, to the latter. Thus, uniform gene organization within the FUM cluster belies a complex evolutionary history that has not always paralleled the evolution of Fusarium.

Introduction

The filamentous fungus Fusarium produces toxic secondary metabolites, also known as mycotoxins, that can accumulate in food and feed crops and thereby pose health risks to humans and livestock. Fumonisins are a family of mycotoxins that can accumulate in maize and some other crops (Munkvold and Desjardins, 1997; Moretti et al., 2010; Proctor et al., 2010), cause multiple diseases in animals, and are epidemiologically associated with esophageal cancer as well as neural tube defects in some human populations (Gelderblom et al., 1988; Marasas et al., 2004). Fumonisins are linear, polyketide-derived molecules that can be classified as B or C fumonisins (FBs or FCs respectively) based on presence (FB) or absence (FC) of a terminal methyl group derived from an amino acid (Fig. 1) (Musser and Plattner, 1997; Sewram et al., 2005). Most known fumonisin-producing fusaria are members of the Fusarium (Gibberella) fujikuroi species complex (FFSC) and produce predominantly FBs (Nelson et al., 1992; Rheeder et al., 2002; Proctor et al., 2004). Although, isolates of Fusarium oxysporum, which is closely related to but not part of the FFSC, do not generally produce fumonisins, FC production has been confirmed in two isolates of this species (Seo et al., 1996; Sewram et al., 2005).

In filamentous fungi, secondary metabolite biosynthetic genes are typically located in gene clusters (Keller et al., 2005). A fumonisin biosynthetic gene (FUM) cluster has been described in the FFSC species Fusarium proliferatum (Waalwijk et al., 2004), Fusarium verticillioides (Proctor et al., 2003), one of the rare fumonisin-producing strains (FRC O-1890) of F. oxysporum (Proctor et al., 2004).
A. Structure of B and C fumonisins (FBs and FCs). R\textsubscript{1} is a methyl group (CH\textsubscript{3}) in FBs and a hydrogen atom in FCs. Within FBs and FCs, R\textsubscript{2}, R\textsubscript{3} and R\textsubscript{4} are either a hydroxyl group (OH) or a hydrogen atom (H) depending on the fumonisin. For example, in FC; and FB; R\textsubscript{2} = H, R\textsubscript{3} = OH and R\textsubscript{4} = OH; in hydroxy-FC; R\textsubscript{2} = OH, R\textsubscript{3} = OH and R\textsubscript{4} = OH.

B. Organization of the 16-gene FUM cluster described in *F. oxysporum*, *F. proliferatum* and *F. verticillioides*. A single expressed sequence tag from *F. verticillioides* and genomic DNA sequence data from *F. verticillioides* and *F. oxysporum* FRC O-1890 provide evidence for an additional small gene (FUM20) located between FUM2 and FUM13. FUM20 is unusual in that its transcript overlaps with that of FUM2 and it is predicted to encode a small (44-amino-acid) protein (Brown *et al.*, 2005).

Fig. 1. A. Structure of B and C fumonisins (FBs and FCs). R\textsubscript{1} is a methyl group (CH\textsubscript{3}) in FBs and a hydrogen atom in FCs. Within FBs and FCs, R\textsubscript{2}, R\textsubscript{3} and R\textsubscript{4} are either a hydroxyl group (OH) or a hydrogen atom (H) depending on the fumonisin. For example, in FC; and FB; R\textsubscript{2} = H, R\textsubscript{3} = OH and R\textsubscript{4} = OH; in hydroxy-FC; R\textsubscript{2} = OH, R\textsubscript{3} = OH and R\textsubscript{4} = OH.

2008), and in the more distantly related fungus *Aspergillus niger* (Khaldi and Wolfe, 2011). The *Fusarium* cluster includes 16 genes that encode biosynthetic enzymes as well as regulatory and transport proteins (Fig. 1). Functions of most of the genes in fumonisin biosynthesis in *F. verticillioides* have been determined by gene inactivation and heterologous expression analyses (Butchko *et al.*, 2003; 2006; Ding *et al.*, 2004; Yi *et al.*, 2005; Zaleta-Rivera *et al.*, 2006). Within the cluster, the FUM1 gene encodes a polyketide synthase that catalyses synthesis of a linear polyketide that forms the backbone structure of fumonisins. In addition, the FUM8 gene encodes an α-oxoamine synthase that mediates whether fusaria produce FBs or FCs by catalysing the condensation of the linear polyketide with alanine to produce FBs or with glycine to produce FCs (Branham and Plattner, 1993; Sewram *et al.*, 2005; Proctor *et al.*, 2008). The number, order and orientation of genes within FUM clusters in *F. verticillioides*, *F. oxysporum* and *F. proliferatum* are the same (Proctor *et al.*, 2003; 2008; Waalwijk *et al.*, 2004). However, the sequences flanking the clusters differ, indicating that the cluster is in a different genomic location in these three species.

The FFSC consists of more than 50 phylogenetically distinct species, many of which are pathogens and/or endophytes of economically important crops (O’Donnell *et al.*, 1998; Kvas *et al.*, 2009). For example, *F. proliferatum* and *F. verticillioides* are maize ear rot pathogens. The latter species is also considered a maize endophyte and is among the fungi most frequently recovered from healthy maize kernels (Munkvold and Desjardins, 1997). The FFSC has been delineated from other fusaria by DNA-based phylogenetic analyses that have also resolved the complex into three major lineages, designated as the African, American and Asian clades (O’Donnell *et al.*, 1998; 2000b; Summerell *et al.*, 2010). Although fumonisin production has been reported in one or more species in each FFSC clade (Nelson *et al.*, 1992; Rheeder *et al.*, 2002), production is discontinuous in that it is not correlated with phylogenetic relationships of species (O’Donnell *et al.*, 2000b; Rheeder *et al.*, 2002; Proctor *et al.*, 2004). Furthermore, Southern blot, PCR and sequence analyses indicate that FUM genes are absent from at least some FFSC species and *F. oxysporum* isolates that do not produce fumonisins (Mirete *et al.*, 2004; Proctor *et al.*, 2004; Glenn *et al.*, 2008; Stepien *et al.*, 2010; Van Hove *et al.*, 2011). Thus, the cluster can be present in one species, absent in a second closely related species, and present in a third distantly related species (Mirete *et al.*, 2004; Proctor *et al.*, 2004; Waalwijk *et al.*, 2004; Glenn *et al.*, 2008; Stepien *et al.*, 2010; Van Hove *et al.*, 2011).

Previous phylogenetic analyses of FUM1 and FUM8 in seven fumonisin-producing fusaria (Proctor *et al.*, 2004) indicated differences between FUM gene genealogies and those of primary-metabolism (PM) genes, i.e. species phylogenies, reported in another study (O’Donnell *et al.*, 1998). This finding indicates that the evolutionary history of FUM genes has not mirrored that of the fusaria in which the genes occur. The objectives of the current study were to (i) further evaluate the evolutionary relationships between FUM clusters in *Fusarium*, (ii) assess variation in the genomic context of the FUM cluster among species, and (iii) characterize the evolutionary processes that have shaped FUM cluster diversity in *Fusarium*. The results demonstrate that the phylogenetic discord of FUM and PM genes coincides with differences in FUM cluster genomic context; that FUM cluster diversity in the FFSC predates the diversification of this species complex; and that phylogenetic discord observed between FUM and PM genes is consistent with horizontal transfer of the cluster as well as cluster duplication, sorting and loss (birth and death).

Published 2013. This article is a U.S. Government work and is in the public domain in the USA., *Molecular Microbiology*, 90, 290–306
Results

Phylogenetic analyses

To evaluate potential discord between FUM cluster evolution and the phylogenetic relationships of fumonisin-producing Fusarium species, we conducted phylogenetic analyses of nine FUM genes and 12 PM genes from representative isolates of 10 fumonisin-producing species (Table 1). Each selected FUM gene (FUM1, FUM3, FUM6, FUM7, FUM8, FUM10, FUM13, FUM14 and FUM19) is required for wild-type fumonisin production in F. verticillioides, and together they span almost the entire length of the FUM cluster (Alexander et al., 2009). The PM genes included eight genes (CAL1, CPR1, HIS3, RED1, RPB1, RPB2, TEF1 and TUB2) that have been used previously to infer phylogenetic relationships within Fusarium (O’Donnell et al., 1998; 2004; Proctor et al., 2009); two global regulatory genes, LAE1 and FLB2 (or FlbB), that also affect fungal secondary metabolism (Lafon et al., 2006; Wiemann et al., 2010); and a dehydrogenase gene, ZBD1, that flanks the FUM cluster in F. verticillioides (Proctor et al., 2003).

Species trees inferred from maximum likelihood (ML) analysis of the combined nucleotide sequences of the 12 PM genes strongly supported F. oxysporum FRC O-1890 as distinct from a monophyletic FFSC, and resolved members of the FFSC into three well supported clades (Fig. 2) corresponding to the previously described African, American and Asian clades (O’Donnell et al., 1998; 2000b). Phylogenetic reconstructions derived from neighbour-joining (NJ), maximum parsimony (MP) and Bayesian analyses of the combined PM genes were identical to the ML tree (Fig. 2). Bootstrap analyses under ML, MP and NJ produced 100% support for every internode except the one defining a sister-group relationship between the Asian and African clades, which was supported by at least 83% of bootstrap replicates. Bayesian posterior probability (BPP) values of 1 were obtained for every internode in the PM phylogeny. Well-supported (> 70% of bootstrap replicates) internodes in individual PM gene trees (Supplemental Fig. S1) were congruent with the combined PM topology except for the placement of Fusarium phyllophilum in ML trees derived from CAL1 and RPB2 (bootstrap scores = 74% and 71% respectively).

Phylogenetic relationships inferred from combined analyses of the nine FUM genes were similar to the species phylogeny in support for monophyletic African and Asian clades (Fig. 2), but monophyly of the FFSC and the American clade were strongly contradicted in the combined and individual analyses of FUM genes (Fig. 2; Supplemental Fig. S1). Specifically, F. oxysporum was nested within the FFSC as a sister-species to Fusarium bulbicola (American clade species) in the combined (ML, MP and NJ bootstrap = 100%; BPP = 1) and individual analyses of FUM gene sequences. These two species were resolved as a sister-group to a monophyletic African clade in the combined FUM gene analyses (ML, MP and NJ bootstrap > 83%; BPP = 0.99) and in six of the nine individual FUM gene trees. The American-clade species Fusarium anthophilum was most closely related to a monophyletic Asian clade in the combined FUM gene phylogeny (ML, MP and NJ bootstrap = 100%; BPP = 1) and in eight of the nine individual FUM gene trees.

Shimodaira–Hasagawa (SH) tests were used to evaluate alternative phylogenetic hypotheses and to assess the statistical significance of the topological differences between FUM and PM trees. Constraint analyses that forced the PM topology onto the FUM data set provided a significantly worse explanation of the data than the unconstrained FUM tree (P < 0.01). Trees recovered from analyses that enforced limited constraints requiring only American clade monophyly or FFSC monophyly also provided significantly worse explanations of the FUM data than the unconstrained FUM tree (P < 0.01). Reciprocal analyses conducted on the PM data demonstrated that the optimal PM tree (species tree) offered a significantly better explanation of the PM data than trees that were consistent with (i) the FUM tree topology, (ii) a sister-species relationship between F. bulbicola and F. oxysporum, or (iii) sister relationships between F. anthophilum and the Asian clade and between F. bulbicola and the African clade (P < 0.01).

Given the incongruence between FUM cluster evolution and species relationships, we analysed intraspecific variation using a subset of four FUM genes (FUM1, FUM8, FUM14 and FUM19) from two to eight isolates of each fumonisin-producing FFSC species included in this study. Individual and combined analyses with this larger set of strains produced a topology (Fig. 3, Supplemental Fig. S2) that was completely consistent with the well-supported relationships among FUM sequences in Fig. 2. In addition, each of the FFSC species were resolved as monophyletic groups in the analyses of four FUM genes (Fig. 3) indicating that FUM cluster polymorphism was not trans-specific and that the FUM gene sequences from strains included in the nine gene analyses were representative of the FUM sequence diversity within their respective species.

Genomic context of FUM cluster

We also examined the genomic context of the FUM cluster by determining the sequence of ~3–20 kb of DNA flanking each side of the cluster. The results indicate that the cluster can occur in five genomic contexts (Fig. 4, Table 2), designated here as GC1, GC2, GC3a, GC3b and GC4. GC1 is the same context previously described for the FUM cluster in F. verticillioides (Proctor et al., 2003) and the FUM cluster remnant in Fusarium musae (Glenn et al., 2006).
Table 1. Strains of *Fusarium* examined in this study.

<table>
<thead>
<tr>
<th>Species</th>
<th>Straina</th>
<th>Other designations</th>
<th>Fumonisina</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Fusarium (Gibberella) fujikuroi</em> species complex</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>F. anthophilum</em></td>
<td>NRRL 25214&lt;br&gt;NRRL 13393&lt;br&gt;NRRL 13602&lt;br&gt;NRRL 25217&lt;br&gt;Fant-1</td>
<td>BBA 62266&lt;br&gt;FRC M-1211&lt;br&gt;CBS 737.97, FRC M-1355&lt;br&gt;BBA 64252</td>
<td>FC&lt;br&gt;None&lt;br&gt;FC&lt;br&gt;None</td>
<td>O’Donnell et al. (1998)&lt;br&gt;O’Donnell et al. (1998)&lt;br&gt;O’Donnell et al. (1998)&lt;br&gt;O’Donnell et al. (1998)</td>
</tr>
<tr>
<td><em>F. bulbicola</em></td>
<td>NRRL 13618&lt;br&gt;NRRRL 13600&lt;br&gt;NRRRL 22947&lt;br&gt;NRRRL 25176&lt;br&gt;F. bulbicola</td>
<td>BBA 63628, CBS 220.76&lt;br&gt;BBA 63620, FRC M-1354&lt;br&gt;BBA 63620, CBS 245.59&lt;br&gt;None</td>
<td>NONE&lt;br&gt;nd&lt;br&gt;FC</td>
<td>O’Donnell et al. (1998)&lt;br&gt;O’Donnell et al. (1998)&lt;br&gt;O’Donnell et al. (1998)</td>
</tr>
<tr>
<td><em>F. fujikuroi</em></td>
<td>HKM 41</td>
<td>FGSC 8382</td>
<td>FB</td>
<td>Desjardins et al. (2000)</td>
</tr>
<tr>
<td></td>
<td>41-79</td>
<td>FB</td>
<td>Desjardins et al. (2007)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>41-84</td>
<td>FB</td>
<td>Desjardins et al. (2007)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Fp0</td>
<td>nd</td>
<td>Proctor et al. (2004)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>GGM 118</td>
<td>nd</td>
<td>Desjardins et al. (2007)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>GL 24</td>
<td>None</td>
<td>Proctor et al. (2004)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>GL 28</td>
<td>None</td>
<td>Proctor et al. (2004)</td>
<td></td>
</tr>
<tr>
<td><em>F. globosum</em></td>
<td>NRRL 26131&lt;br&gt;NRRRL 26133</td>
<td>CBS 428.97, FRC M-8014, MRC 6647&lt;br&gt;CBS 430.97, MRC 6657</td>
<td>FB&lt;br&gt;FB</td>
<td>O’Donnell et al. (1998); Moses et al. (2010)&lt;br&gt;O’Donnell et al. (1998); Moses et al. (2010)</td>
</tr>
<tr>
<td><em>F. nygamai</em></td>
<td>FRC M-7492</td>
<td>MUCL 47026, FGSC 8934</td>
<td>FB</td>
<td>Klaassen and Nelson (1996)</td>
</tr>
<tr>
<td></td>
<td>FRC M-7491</td>
<td>FGSC 8933</td>
<td>FB</td>
<td>Klaassen and Nelson (1996)</td>
</tr>
<tr>
<td></td>
<td>NRRRL 13448&lt;br&gt;NRRRL 25549&lt;br&gt;NRRRL 25596&lt;br&gt;NRRRL 25178&lt;br&gt;NRRRL 25341</td>
<td>BBA 69862, CBS 749.97, FRC M-1375&lt;br&gt;BBA 65175, CBS 413.97&lt;br&gt;BBA 7893, CBS 246.61&lt;br&gt;BBA 63625&lt;br&gt;MUCL 46369</td>
<td>nd&lt;br&gt;nd&lt;br&gt;nd&lt;br&gt;nd&lt;br&gt;nd</td>
<td>O’Donnell et al. (1998)&lt;br&gt;O’Donnell et al. (1998)&lt;br&gt;O’Donnell et al. (1998)&lt;br&gt;O’Donnell et al. (1998)&lt;br&gt;O’Donnell et al. (1998)</td>
</tr>
<tr>
<td><em>F. phyllophilum</em></td>
<td>NRRL 13617</td>
<td>MUCL 43905, CBS 216.76</td>
<td>FB</td>
<td>O’Donnell et al. (1998)</td>
</tr>
<tr>
<td></td>
<td>NRRL 25053&lt;br&gt;NRRRL 25178&lt;br&gt;NRRRL 25341</td>
<td>BBA 7893, CBS 246.61&lt;br&gt;BBA 63625&lt;br&gt;MUCL 46369</td>
<td>nd&lt;br&gt;nd&lt;br&gt;nd</td>
<td>O’Donnell et al. (1998)&lt;br&gt;O’Donnell et al. (1998)&lt;br&gt;O’Donnell et al. (1998)</td>
</tr>
<tr>
<td><em>F. verticilliioides</em></td>
<td>FRC M-3125&lt;br&gt;41-109&lt;br&gt;AMRF-1&lt;br&gt;ISU-3&lt;br&gt;ISU-94040&lt;br&gt;FRC M-3120&lt;br&gt;FRC M-5500</td>
<td>NRRL 20956, FGSC 7600&lt;br&gt;AMRF-1&lt;br&gt;FB&lt;br&gt;FB&lt;br&gt;FB&lt;br&gt;FB&lt;br&gt;None</td>
<td>FB&lt;br&gt;FB&lt;br&gt;FB&lt;br&gt;FB&lt;br&gt;FB&lt;br&gt;FB&lt;br&gt;None</td>
<td>Leslie et al. (1992)&lt;br&gt;Leslie et al. (1992)&lt;br&gt;Desjardins et al. (2007)&lt;br&gt;Desjardins et al. (2007)&lt;br&gt;Current study&lt;br&gt;Desjardins et al. (2007)&lt;br&gt;Searle et al. (2008)</td>
</tr>
</tbody>
</table>

*Fusarium oxysporum*

<table>
<thead>
<tr>
<th>Species</th>
<th>Straina</th>
<th>Other designations</th>
<th>Fumonisina</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>F. oxysporum</em></td>
<td>FRC O-1890&lt;br&gt;CAR</td>
<td>NRRL 39464</td>
<td>FC</td>
<td>Proctor et al. (2008); Seo et al. (1996)</td>
</tr>
</tbody>
</table>

a. BBA indicates the Biologische Bundesanstalt fur Land- und Forstwirtschaft collection, Berlin, Germany; CBS indicates the Centraal Bureau voor Schimmelcollections collection, Utrecht, the Netherlands; FGSC indicates the Fungal Genetics Stock Center culture collection at the University of Missouri, Kansas City (Missouri), USA; FRC indicates the *Fusarium* Research Center collection at Pennsylvania State University, University Park, USA; ITEM indicates the collection at the Institute of Sciences of Food Production, National Research Council, Bari, Italy; MRC indicates the Medical Research Council collection, Tygerberg, South Africa; MUCL indicates the collection at the Mycothèque de l’Université catholique de Louvain, Louvain-la-Neuve, Belgium; NRRL indicates collection at the US Department of Agriculture, Agriculture Research Service, National Center for Agricultural Utilization Research (USDA ARS NCAUR), Peoria (Illinois), USA. Strain HKM 41 was obtained from A.E. Desjardins at USDA ARS NCAUR. *F. oxysporum* strain FRC O-1890 (= CAR, NRRL 39464) was obtained from Y.-W. Lee, Seoul National University. Strains GL 24 and GL 28 were provided by Thomas Gordon, University of California, Davis, California. Strains ISU-3 and ISU-94040 were provided by Gary Munkvold, Iowa State University, Ames, Iowa.
b. FB and FC indicate that B or C fumonisins, respectively, were the predominant fumonisins produced by the strain indicated. Production by most species listed has been reported previously (Nelson et al., 1992; Fotso et al., 2002; Rheeder et al., 2002; Proctor et al., 2004; Glenn et al., 2007; Moretti et al., 2010). Data on production by strains of *F. anthophilum* and *F. bulbicola* were determined during the course of the current study. *F. anthophilum* strains NRRL 13602 and NRRL 25214 produced FC1, hydroxy-FC1 and FC2 at 3–11, 12–45 and 7–45 μg per g cracked maize kernel culture respectively. *F. bulbicola* strain NRRL 22947 produced FC1, FC2 and FC3 at 117, 2 and 19 μg per g culture respectively.
c. Strains used for construction of GenomeWalker library for analysis of *FUM* cluster flanking regions.

Published 2013. This article is a U.S. Government work and is in the public domain in the USA., *Molecular Microbiology*, 90, 290–306
GC2 was observed in all African-clade species examined (Fig. 4, Table 2). The organization of genes in GC2 was confirmed by examination of the two recently released genome sequences of *F. fujikuroi* (Jeong et al., 2013; Wiemann et al., 2013).

GC3a and GC3b were observed in American-clade species *F. anthophilum* and *F. bulbicola* respectively. GC3a and GC3b are similar in that the three genes (*CPM1*, *MFS2* and *DOX1*) flanking the *FUM19* side of the cluster are homologous (Fig. 4, Table 2). In contrast, GC3a and GC3b differ in the *FUM21*-flanking region. In GC3a (*F. anthophilum*), *FUM21* is flanked by genes *CPM2* and *TSP1* (Fig. 4, Table 2). We obtained only ~1500 bp of sequence data upstream of *FUM21* in GC3b (*F. bulbicola*), and within this region there was no evidence for genes. Within the ~300 bp immediately upstream of *FUM21*, the DNA sequences of GC3a and GC3b shared multiple 20- to 87-bp-long regions with up to 80% identity. More than 300 bp upstream of *FUM21*, there was no significant sequence identity. We identified a *F. bulbicola* homologue of *CPM2* with 85% identity to the *F. anthophilum CPM2*. However, analysis of the *CPM2* flanking region in *F. bulbicola* revealed that *CPM2* and *FUM21* are not located within 17 kb of one another in this species. BLASTX analysis indicated there are five genes located within the 15.5 kb of sequence immediately upstream of the *F. bulbicola CPM2* homologue. In *F. anthophilum*, in contrast, the intergenic region between *CPM2* and *FUM21* is 4 kb, and there is no evidence for other genes between them. Thus, although the DNA sequences of *F. anthophilum* and *F. bulbicola* exhibit significant similarity within the first 300 bp upstream of *FUM21*, the sequences are completely different further upstream. This indicates that the *FUM21*-flanking regions in these species are not the same and, therefore, constitute different genomic contexts. One attribute common to the *FUM21* flanking region in these two species is AT-rich sequences. *F. anthophilum* has a 1 kb AT-rich segment beginning 600 bp upstream of *FUM21*, and *F. bulbicola* has two shorter segments (~350 bp) within the 1500 nucleotides upstream of the *FUM21* start codon.

In GC4, the genomic context of the *FUM* cluster in *F. oxysporum* strain FRC O-1890 (Proctor et al., 2008), there is no evidence for a full-length gene within the ~2800 bp region upstream of *FUM21*. However, this region includes three apparent gene fragments that are 140–377 bp in length and are located adjacent to one another. The deduced amino acid sequence of one of these fragments (377 bp) exhibits 65% identity to putative transposase genes from other fungi, including other strains of *F. oxysporum*. The region upstream of *FUM21* in GC4 also includes two AT-rich regions of ~340 and ~400 bp. GC4 is like GC3a (*F. anthophilum*) and GC3b (*F. bulbicola*) in that the *FUM19* side of the cluster is flanked by a homologue of *CPM1*. The 770 bp of sequence available upstream of *CPM1* in GC4 (*F. oxysporum*) does not include sequences...
with similarity to any previously described genes and lacks significant identity to the sequence upstream of $CPM_1$ in GC3a and GC3b.

Examination of sequence data at the National Center for Biotechnology Information (NCBI) and the Broad Institute indicate that the genes flanking the two ends of the FUM cluster in GC1 or GC2 are adjacent to or near one another in other Fusarium genomes. For example, the genome sequences of $F. fujikuroi$ (Jeong et al., 2013; Wiemann et al., 2013) and $F. oxysporum$ strain FGSC 4287 (Ma et al., 2010) has homologues of the $ZNF1$, $ZBD1$, ORF20 and ORF21 arranged in the same order as in GC1. However, unlike GC1, there is no FUM cluster located between the $F. oxysporum$ FGSC 4287 homologues of $ZBD1$ and ORF20. Instead, these two genes are located only 1550 bp apart. Likewise, the $F. oxysporum$ FGSC 4287 and $F. verticillioides$ (Ma et al., 2010) genome sequences have homologues of $MFS1$, $ZCB1$ and $GAT1$ arranged in the same order as in GC2, but without a FUM cluster between $ZCB1$ and $GAT1$. A similar arrangement of FUM cluster-flanking genes from GC3a was not evident in other Fusarium genome sequences. In fact, there is no evidence for closely related homologues of the GC3a/ GC3b FUM19 flanking genes $CPM1$, $MFS2$ and $DOX1$ in any of the Fusarium genome sequences examined. However, there is a $\sim$ 220 bp remnant of $CPM1$ located $\sim$ 10 bp downstream of FUM19 in GC1 (Fig. 4).

Together, these results provide evidence that the FUM cluster can exist in five different genomic contexts in the fusaria examined. The differences in genomic contexts are concordant with genealogies inferred from PM genes (Fig. 2), but discordant with genealogies inferred from FUM genes (Fig. 4), indicating an association between differences in genomic context of the cluster and the phylogenetic discord.

Analysis of synonymous site divergence

The timing of FUM cluster divergence relative to divergence of species and clades in which the cluster occurs was assessed by comparing the number of synonymous differences per synonymous site ($d_s$) in FUM versus PM genes. Non-synonymous substitutions were excluded from the analyses in order to minimize the impact of variation in substitution rates resulting from gene-specific differences in selective constraint. Within the FFSC, $d_s$
values for FUM genes were significantly greater ($P < 0.001$) than for PM genes in five of the six pairwise comparisons involving species or clades that differ in genomic context of the FUM cluster (Table 3). For example, in a comparison of the American clade species F. anthophilum (GC3a) and F. bulbicola (GC3b), divergence of FUM genes ($d_s = 0.365$) was more than 10-fold greater than PM genes ($d_s = 0.034$). These differences suggest that divergence of some FUM cluster types (defined by genomic context) predate divergence of the species or clades being compared. The comparison of the Asian clade (GC2) and F. anthophilum (GC3a) was an exception, in that divergence of FUM ($d_s = 0.189$) and PM genes ($d_s = 0.197$) was not significantly different, which is consistent with the expected pattern for orthologous FUM clusters. FUM gene divergence values were equal to or lower than values for PM gene in seven of the nine pairwise comparisons of species in which the FUM cluster is in the same genomic context (i.e. within the Asian or African clades) (Supplemental Table S1). These comparisons and the similar levels of divergence of FUM and PM genes in comparisons of Asian clade species and F. anthophilum indicate the evidence for an ancient divergence of some FUM cluster types is likely not due to elevated rates of synonymous substitution in FUM cluster genes. Thus, differences in $d_s$ values of FUM and PM genes likely reflect different lengths of time that the genes have been diverging rather than different rates of synonymous substitution.

Interestingly, divergence of FUM genes ($d_s = 0.028$) was significantly less ($P < 0.001$) than PM genes ($d_s = 0.179$) when comparing F. oxysporum FRC O-1890 (GC4) and F. bulbicola (GC3b). Combined with the phylogenetic results demonstrating that FUM cluster sequences from F. oxysporum are nested within the FFSC, the very recent divergence of the FUM clusters in GC3b and GC4 provides
strong evidence for horizontal transfer of the FUM cluster between F. oxysporum and F. bulbicola or a closely related member of the FFSC. This transfer event also appears to have included the CPM1 gene, which is adjacent to FUM19 in American clade species and F. oxysporum FRC O-1890 (Fig. 4). Genetic divergence of CPM1 sequences in F. bulbicola and F. oxysporum FRC O-1890 (dS = 0.037) was not significantly different from that of FUM gene sequences (dS = 0.028), and was significantly less (P < 0.001) than the divergence of PM gene sequences used to infer species relationships (dS = 0.174). Similarly, the CPM1, MFS2 and DOX1 genes, which are adjacent to FUM19 in both F. bulbicola and F. anthophilum, displayed levels of synonymous divergence (dS = 0.365) consistent with those observed for FUM gene sequences in these species (dS = 0.305), and significantly greater (P < 0.001) than was

### Table 2. Descriptions of genes in FUM cluster flanking regions.

<table>
<thead>
<tr>
<th>Genomic context</th>
<th>Region flanking</th>
<th>Tentative designation</th>
<th>Fusarium homologues</th>
<th>Predicted functions</th>
</tr>
</thead>
<tbody>
<tr>
<td>GC1</td>
<td>FUM21</td>
<td>ZNF1</td>
<td>FVEG_00313</td>
<td>Zinc (RING) finger protein – transcription factor</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ZBD1</td>
<td>FVEG_00314</td>
<td>Zinc-binding dehydrogenase</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ORF20</td>
<td>FVEG_00330</td>
<td>No known protein domains</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ORF21</td>
<td>FVEG_00331</td>
<td>No known protein domains</td>
</tr>
<tr>
<td>GC2</td>
<td>FUM21</td>
<td>MFS1</td>
<td>FVEG_10012</td>
<td>Major facilitator superfamily transport protein</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ZCB1</td>
<td>FVEG_10011</td>
<td>Zinc(2)-Cysteine(6) transcription factor</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ANK1</td>
<td>None</td>
<td>Ankyrin repeat protein</td>
</tr>
<tr>
<td></td>
<td></td>
<td>GAT1</td>
<td>None</td>
<td>Glutamine amidotransferase-like domain</td>
</tr>
<tr>
<td>GC3a</td>
<td>FUM21</td>
<td>CPM2</td>
<td>FVEG_10523</td>
<td>Cytochrome P450 monooxygenase</td>
</tr>
<tr>
<td></td>
<td></td>
<td>TSP1 (P)</td>
<td>None</td>
<td>Transposase</td>
</tr>
<tr>
<td>GC3a &amp; GC3b</td>
<td>FUM19</td>
<td>CPM1</td>
<td>None</td>
<td>Cytochrome P450 monooxygenase</td>
</tr>
<tr>
<td></td>
<td></td>
<td>MFS2</td>
<td>FGSG_03360</td>
<td>Major facilitator superfamily transport protein</td>
</tr>
<tr>
<td></td>
<td></td>
<td>DOX1</td>
<td>None</td>
<td>Dioxygenase</td>
</tr>
<tr>
<td>GC4</td>
<td>FUM21</td>
<td>None</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td></td>
<td>FUM19</td>
<td>None</td>
<td>None</td>
<td>Cytochrome P450 monooxygenase</td>
</tr>
</tbody>
</table>

- **a.** Genomic contexts of the FUM cluster as indicated in the text.
- **b.** Tentative designations for genes are based on predicted functions, which are in turn based on similarity/identity to gene/proteins reported to have functions in the NCBI/GenBank database. Similarity was determined by BLASTX and/or BLASTP analysis against the NCBI database. (P) indicates that sequence for only a part of a coding region was obtained.
- **c.** Homologues in the Fusarium Comparative Database for which the per cent identity of the predicted amino acid sequence is > 55% based on BLAST analysis. FGSG, FOXG and FVEG numbers indicate sequences from F. graminearum, F. oxysporum and F. verticillioides respectively.

### Table 3. Synonymous site divergence estimates (dS) from concatenated sequences of 12 PM genes (above diagonal) and nine FUM genes (below diagonal) for species/clades with the FUM cluster in different genomic contexts.

<table>
<thead>
<tr>
<th></th>
<th>F. anthophilum</th>
<th>F. bulbicola</th>
<th>Asian clade</th>
<th>African clade</th>
<th>F. oxysporum O-1890</th>
</tr>
</thead>
<tbody>
<tr>
<td>F. anthophilum (GC3a)</td>
<td>0.365 ± 0.017</td>
<td>0.034 ± 0.004</td>
<td>0.197 ± 0.010</td>
<td>0.179 ± 0.009</td>
<td>0.179 ± 0.009</td>
</tr>
<tr>
<td>F. bulbicola (GC3b)</td>
<td>0.189 ± 0.011</td>
<td>0.534 ± 0.023</td>
<td>0.192 ± 0.010</td>
<td>0.173 ± 0.009</td>
<td>0.181 ± 0.009</td>
</tr>
<tr>
<td>Asian clade (GC2)</td>
<td>0.429 ± 0.019</td>
<td>0.242 ± 0.012</td>
<td>0.589 ± 0.026</td>
<td>0.179 ± 0.009</td>
<td>0.173 ± 0.009</td>
</tr>
<tr>
<td>African clade (GC1)</td>
<td>0.341 ± 0.016</td>
<td>0.028 ± 0.004</td>
<td>0.509 ± 0.022</td>
<td>0.173 ± 0.009</td>
<td>0.221 ± 0.012</td>
</tr>
</tbody>
</table>

- **a.** Designations in parentheses indicate genomic contexts of the FUM cluster as described in the text and shown in Fig. 4. 
  dS values were estimated by two methods, Codeml-Maximum likelihood and modified Nie-Gojobori. The two methods yielded highly similar results, and therefore, only the results of Codeml-Maximum likelihood are presented.
Comparisons of two residues in the deduced amino acid sequence of Fum8 that exhibit a consistent difference in FB- versus FC-producing Fusarium species.

<table>
<thead>
<tr>
<th>Species</th>
<th>Genomic context</th>
<th>Residue 552</th>
<th>Residue 580</th>
</tr>
</thead>
<tbody>
<tr>
<td>F. nygamai</td>
<td>GC1</td>
<td>LFPA</td>
<td>GPCS</td>
</tr>
<tr>
<td>F. phyllophillum</td>
<td>GC1</td>
<td>LEAK</td>
<td>GPSS</td>
</tr>
<tr>
<td>F. ramigenum</td>
<td>GC1</td>
<td>LEAP</td>
<td>GPSS</td>
</tr>
<tr>
<td>F. verticillioides</td>
<td>GC1</td>
<td>LEAP</td>
<td>GPSS</td>
</tr>
<tr>
<td>F. fujikuroi</td>
<td>GC2</td>
<td>LEAG</td>
<td>GPSS</td>
</tr>
<tr>
<td>F. globosum</td>
<td>GC2</td>
<td>LEGIN</td>
<td>GPSS</td>
</tr>
<tr>
<td>F. proliferatum</td>
<td>GC2</td>
<td>LEGIN</td>
<td>GPSS</td>
</tr>
<tr>
<td>F. anthophilum</td>
<td>GC3a</td>
<td>LFPA</td>
<td>GPCS</td>
</tr>
<tr>
<td>F. bulbicola</td>
<td>GC3b</td>
<td>LEAP</td>
<td>GPSS</td>
</tr>
<tr>
<td>F. oxysporum</td>
<td>GC4</td>
<td>LEAF</td>
<td>GPSS</td>
</tr>
<tr>
<td>A. niger</td>
<td>FB</td>
<td>LEAF</td>
<td>GPSS</td>
</tr>
</tbody>
</table>

a. Except for F. anthophilum and F. bulbicola, fumonisin production was determined previously (Rheeder et al., 2002; Proctor et al., 2004; Sewram et al., 2005). Despite two attempts using our standard crack maize kernel assay and liquid chromatography-mass spectroscopy analysis, we were not able to confirm the previous report (Nelson et al., 1992) of low-level FB production in F. anthophilum strains FRC M-854 and FRC M-1238; the strains did not produce detectable levels of fumonisin in the assay. In addition, we found that the previous report (Nelson et al., 1992) of high FB production in a third F. anthophilum strain (FRC M-1139) was most likely the result of species misidentification; FRC M-1139 produced high levels of FBs in our standard assay, but analysis of its TEF1 sequence at the Fusarium-ID database (Geiser et al., 2004) revealed that the strain was most likely F. proliferatum (data not shown).

b. The letters are standard abbreviations for amino acids (e.g. A = alanine, V = valine).

The residues correspond to amino acids 552 and 580 (highlighted in black) in the F. verticillioides Fum8 homologue (NCBI GenBank accession AAG27130.3).

observed for PM genes (d = 0.034). As such, it appears that divergence of the CPM1, MFS2 and DOX1 homologues in F. anthophilum and F. bulbicola predates divergence of the FFSC and that these three genes have evolved in concert with the FUM cluster in the two species.

Fumonisin production and variation in FUM8

With the exception of F. anthophilum and F. bulbicola, fumonisin production has been confirmed in multiple reports in the literature for the FFSC species shown in Table 1. To clarify the fumonisin production phenotypes of F. anthophilum and F. bulbicola, we subjected three to four strains of each species to a fumonisin production assay by growing them on cracked maize kernel medium for 10 days and analysing culture extracts by liquid chromatography-mass spectroscopy. F. anthophilum strains NRRL 13602 and NRRL 25214 produced fumonisins FC1, hydroxy-FC1 and FC2, but no or only trace amounts of FBs; whereas F. anthophilum strains NRRL 13293 and NRRL 25217 produced no detectable fumonisins (Table 1). F. bulbicola strain NRRL 22947 produced FC1, FC2 and FC3, but no FBs; and F. bulbicola strains NRRL 13618 or NRRL 25176 produced no detectable fumonisins (Table 1).

Of the fumonisin-producing species examined in this study, the majority produce predominantly FBs, and three (F. anthophilum, F. bulbicola and F. oxysporum) produce predominantly FCs (Table 1). Because the FUM8-encoded enzyme, Fum8, determines FB versus FC production (Proctor et al., 2008), we compared the predicted amino acid sequence of Fum8, corresponding to the region of FUM8 used in the phylogenetic analysis, in order to identify differences in the sequences that might contribute to production of FBs versus FCs. The Fum8 homologue in the distantly related FB-producing fungus A. niger was also included in this analysis. Within the Fum8 region examined, only two residues differed consistently between FB and FC-producing fusaria. The first residue corresponded to amino acid 552 in the F. verticillioides Fum8 homologue and was an alanine in all FB producers and a proline in all FC producers (Table 4). However, this difference was not conserved in A. niger. In fact, the A. niger Fum8 sequence exhibits no identity to the Fusarium Fum8 sequence over a 17-amino-acid region that includes residue 552. The second residue corresponded to amino acid 580 in the F. verticillioides Fum8 homologue and was a valine in the FB-producing fusaria as well as in A. niger. In contrast, the corresponding residue was a valine in the Fum8 homologues of FC-producing fusaria (Table 4).

Discussion

The results of this study extend previous evidence of discord between the evolutionary histories of two FUM cluster genes (Proctor et al., 2004) and species relationships within the FFSC (O’Donnell et al., 1998) by providing direct comparisons of phylogenies inferred from nine FUM genes and 12 PM genes. Discord between FUM cluster evolution and the phylogenetic relationships of fumonisin-producing fusaria was strongly supported by
bootstrap analyses and Bayesian posterior probabilities (Fig. 2), and by direct tests of alternative phylogenetic hypotheses via constraint analyses. In addition, our analyses indicated an association between phylogenetic discord and differences in FUM cluster genomic context and that FUM cluster diversity in the FFSC predates diversification of this species complex. Cases of phylogenetic discord and unexpectedly high levels of genetic divergence were previously observed in analyses of the aflatoxin (afl) and trichothecone (TRI) biosynthetic gene clusters in Aspergillus and Fusarium respectively (Ward et al., 2002; Carbone et al., 2007). This discord was attributed to trans-species evolution of afl and TRI polymorphism caused by balancing selection to maintain diversity of metabolite production phenotypes (chemotypes). In contrast to these examples, FUM cluster polymorphism was not trans-specific (Fig. 3), phylogenetic incongruence and high levels of synonymous divergence were directly related to differences in FUM cluster genomic context (Figs 2 and 4 and Table 3), and relationships inferred from FUM cluster genes were not consistent with fumonisin chemotype differences (Table 1). These findings indicate that balancing selection and trans-species evolution are not likely explanations for the phylogenetic discord and unexpectedly high levels of synonymous divergence between FUM cluster types occupying different genomic locations in Fusarium.

Interspecies hybridization is another potential cause of phylogenetic discord. However, the affects of hybridization should be evident across the genome, and not restricted to a single locus. Two hallmarks for such hybridization are discord among PM genes and/or multiple copies of individual PM genes per genome (O’Donnell et al., 2000a; Moon et al., 2004). We found no evidence for either of these phenomena among species with different FUM cluster types. Therefore, hybridization between the Fusarium species examined, or their close relatives, is not a likely explanation for the observed phylogenetic discord.

Horizontal transfer and FUM cluster evolution

Recent documentation of horizontal transfer (HT) of secondary metabolite gene clusters between fungi (Khaldi et al., 2008; Slot and Rokas, 2011; Campbell et al., 2012) indicates that HT is another potential cause of the phylogenetic discord of FUM and PM genes (Fig. 2). The observation that FUM cluster sequences from F. oxysporum strain FRC O-1890 were deeply nested within the FFSC in individual and combined FUM gene trees (Figs 2 and 3, and Supplemental Figs S1 and S2) is consistent with HT of the FUM cluster between F. bulbicola or a closely related member of the FFSC and F. oxysporum. The relative magnitude of synonymous divergence (Table 3) suggests the transfer occurred relatively recently, probably following the split between the American clade species F. bulbicola and F. anthophilum. Multiple lines of evidence suggest that the transfer occurred from the American clade to F. oxysporum rather than the reverse. First, if transfer had occurred from F. oxysporum to F. bulbicola, FUM sequences from these two species should be resolved on a branch that formed an outgroup relative to all other members of the FFSC, in the same manner that F. oxysporum is resolved as an outgroup to the FFSC in PM gene genealogies. Transfer from the FFSC American clade to F. oxysporum is also supported by the fact that fumonisn production and the FUM cluster occur in multiple species of the FFSC, but production has been confirmed in only two isolates of F. oxysporum (Seo et al., 1996; Sewram et al., 2005) and the cluster has been reported only in strain FRC O-1890 (Proctor et al., 2008). HT-mediated acquisition of the FUM cluster by F. oxysporum is also consistent with evidence that ~25% of the F. oxysporum genome is composed of DNA acquired from other organisms (Ma et al., 2010).

Phylogenetic relationships among FUM cluster sequences (Figs 2 and 3) were also suggestive of two HT events among members of the FFSC; one between the African clade and F. bulbicola, and a second between the Asian clade and F. anthophilum. To test the hypotheses that such HT events occurred, we assessed the relative levels of synonymous divergence of FUM and PM genes between these species or clades. If the HT events had occurred, FUM cluster genes would be expected to have lower levels of sequence divergence than PM genes in comparisons of putative donor and recipient species, as was the case for F. bulbicola and F. oxysporum (Table 3) and in recently reported instances of HT involving secondary metabolite biosynthetic (SMB) gene clusters in fungi (Slot and Rokas, 2011; Campbell et al., 2012). These lower levels of divergence of transferred SMB clusters occur because the split between donor and recipient species necessarily predates HT of the clusters. However, branch lengths (Fig. 2) and synonymous divergence values (Table 3) strongly contradicted hypotheses of HT between members of the FFSC. In fact, these analyses indicated that FUM cluster diversity among some FFSC species predates the diversification of the FFSC, which is the opposite of what would be expected if HT of the FUM cluster had occurred among members of this species complex.

A model based on HT can still be used to explain the phylogenetic incongruence (Fig. 2) and deep divergence of FUM cluster sequences (Table 3), but it requires multiple independent transfers of the FUM cluster into multiple lineages of the FFSC from unidentified donors outside of the FFSC (Fig. 5A). In this model, the FUM cluster was acquired by an ancestor of the FFSC via HT from an unknown donor, followed by vertical inheritance and retention of the cluster in the Asian clade and F. anthophilum as
well as its loss from \( F. \) \textit{bulbicola} and the ancestor of the African clade. This model also requires a second transfer of the \( FUM \) cluster from a second unknown donor to an ancestor of the African clade, and a third transfer from a third unknown donor to an ancestor of \( F. \) \textit{bulbicola}. Because of the distant relationship of \( FUM \) genes in \( A. \) \textit{niger} and \( Fusarium \) (Fig. 2), \( A. \) \textit{niger} could not have been one of the donors. Alternative HT scenarios are possible but are not less complicated than the one presented here. Thus, an HT model that explains the discord between \( FUM \) gene and species phylogenies within the FFSC requires a minimum of three HGT events from unknown donors to multiple lineages within the FFSC. There are an increasing number of studies documenting HT of genes and entire biosynthetic gene clusters between fungi (Khaldi \textit{et al.}, 2008; Mehrabi \textit{et al.}, 2011; Slot and Rokas, 2011; Campbell \textit{et al.}, 2012; Gardiner \textit{et al.}, 2012). However, the number of independent HT events depicted in Fig. 5A involving closely related species of fungi would be unprecedented and suggestive of a strong selective advantage associated with the acquisition of the \( FUM \) cluster by species of \( Fusarium \). The little and conflicting information available on the role of fumonisins in the ecology of \( Fusarium \) makes it difficult to speculate about the adaptive value of the cluster. Furthermore, fumonisin production appears to have been lost in multiple species or lineages of the FFSC (Proctor \textit{et al.}, 2004; Glenn \textit{et al.}, 2008; Van Hove \textit{et al.}, 2011), suggesting frequent loss of selection. Given this, we contend that a model of \( FUM \) cluster evolution within the FFSC that relies solely on horizontal transfers is improbable. Nevertheless, such a model cannot be ruled out given existing information.

**Birth, death and \( FUM \) cluster evolution**

The birth-and-death model of evolution that has been applied to several multigene families (Nei and Rooney, 2005; Rooney and Ward, 2005) offers an alternative explanation for the observed phylogenetic incongruence (Fig. 2) and deep divergence of \( FUM \) cluster sequences (Table 3). Birth-and-death evolution involves duplication and subsequent maintenance or loss of gene copies to describe expansion, diversification and contraction of multigene families (Nei and Rooney, 2005). When this model is used to explain \( FUM \) cluster evolution, the three highly divergent \( FUM \) cluster types found in the African clade (GC1), \( F. \) \textit{bulbicola} (GC3b), and the Asian clade plus \( F. \) \textit{anthophilum} (GC2/GC3a) represent paralogues derived from ancient duplication events rather than homologues of the cluster derived from speciation (orthologous clusters) or HT (xenologous clusters) events (Fig. 5B). Cluster duplication and paralogy could explain the incongruence between \( FUM \) gene trees and species trees within the FFSC because phylogenetic relationships and levels of divergence among paralogues would reflect the pattern and timing of cluster duplication rather than
speciation events. However, FUM cluster types in the Asian clade and F. anthophilum are interpreted as orthologous because the level of divergence observed between their FUM gene sequences was not significantly different than levels of divergence between PM genes. This suggests that divergence of the FUM cluster in the Asian-clade species and F. anthophilum coincided with the split between the Asian and American clades. Furthermore, the presence of the F. anthophilum and Asian-clade FUM cluster orthologues in different genomic contexts suggests that relocation of the FUM cluster within a genome can occur in the absence of duplication or HT.

Synonymous divergence between the three putative FUM cluster paralogues was significantly greater than between PM genes in comparisons of F. oxysporum and the FFSC (Table 3), indicating that all three paralogues were present in the common ancestor of the F. oxysporum and FFSC lineages. However, PCR and Southern data from multiple species (Proctor et al., 2004; Stepien et al., 2010) and analyses of the F. fujikuroi and F. verticillioides genomes (Ma et al., 2010; Jeong et al., 2013) indicate that there is only one copy of the FUM cluster in extant, fumonisin-producing species of Fusarium that have been examined. Therefore, acceptance of the birth-and-death model requires us to postulate that following duplication events of the FUM cluster in an ancestral Fusarium genome, the evolutionary history of the FFSC has been marked by reductions in FUM cluster copy number (Fig. 5B). The discontinuous distribution of fumonisin production within the FFSC combined with PCR and Southern data indicating that fumonisinnonproducing FFSC species lack an intact FUM cluster (Proctor et al., 2004; Glenn et al., 2008; Van Hove et al., 2011) provides indirect evidence of FUM cluster loss. When evidence for the presence and absence of FUM genes is mapped onto FFSC species phylogenies (O’Donnell et al., 1998; Kvas et al., 2009) it appears that the FUM cluster has been lost at least 15 times during the evolution of the FFSC (Supplemental Fig. S3). Some of these losses map to relatively short branches such that closely related or sibling species (e.g. F. musae–F. verticillioides, F. napiforme–F. ramigenum and F. phyllophorum–F. udum) now differ in the presence and absence of the FUM cluster. In addition, there is direct evidence for FUM cluster loss in F. musae, which lacks a cluster but has remnants of FUM21 and FUM19 in the same genomic context (GC1) as the intact clusters in other African clade species (Glenn et al., 2008; Van Hove et al., 2011). Although our analysis did not identify any fusaria with multiple paralogues of the FUM cluster, there are examples of two or more copies of SMB clusters in other fungal species: two paralogues of a putative polyketide biosynthetic cluster in F. oxysporum (Brown et al., 2012); three paralogues of a putative non-ribosomal peptide biosynthetic cluster in Pyrenophora tritici-repentis (Manning et al., 2013); and two paralogues of the all cluster in Aspergillus parasiticus (Chang and Yu, 2002).

Under a birth-and-death model, the discontinuous distribution of fumonisins production and the phylogenetic discord between FUM gene and species phylogenies would have resulted from differential loss of FUM cluster paralogues in different lineages through a process known as paralogue sorting (Rooney and Ward, 2008). In such a scenario, some relatively distantly related lineages (e.g. F. bulbicola and the African clade) retained more closely related paralogues, whereas some closely related lineages (e.g. F. anthophilum and F. bulbicola) retained more distantly related paralogues, and some lineages or species lost all copies of the FUM cluster. Paralogue sorting is similar to the well-characterized process of lineage sorting (Avise et al., 1983), which describes the non-phylogenetic transmittal of alleles at a single locus during speciation, and failure to consider evidence of paralogue sorting has resulted in false inferences of HT (Rooney and Ward, 2008).

Loss of cluster paralogues could also account for absence of the FUM cluster in most strains of F. oxysporum despite FUM gene divergence data indicating that multiple FUM cluster paralogues were present in the common ancestor of F. oxysporum and the FFSC. The absence of the cluster is evident from an inability to detect FUM genes in 11 F. oxysporum genome sequences at the Broad Institute’s Comparative Fusarium Database or by Southern blot or PCR of 15 strains of this species (Proctor et al., 2004; R.H. Proctor and A. Moretti, unpublished). The exception to this is strain FRC O-1890, which our data indicate acquired the FUM cluster via HT from the American clade. Thus, it appears that HT reintroduced the FUM cluster and fumonisin production into a lineage of Fusarium that had previously lost this metabolic capability. A similar mechanism was responsible for the reacquisition of the biotin biosynthetic pathway in Saccharomyces cerevisiae (Hall and Dietrich, 2007). In this context, the results of the current study suggest a mixed model of FUM cluster evolution that incorporates HT as well as cluster duplication and loss (birth and death) with paralogue sorting. A similar pattern of evolution was described for the virulence-associated internalin gene family in Listeria, where a relatively recent HT event was overlaid on a background of gene duplication and loss events consistent with birth-and-death evolution (Rooney and Ward, 2008). To our knowledge, the current study is the first to explicitly apply the birth-and-death model with paralogue sorting to the evolution of an SMB gene cluster, and in fact, the first application of the model to a biosynthetic pathway. However, the model is very similar to that used to describe the evolution of the avirulence gene (ACE1) cluster, in which ACE1 cluster dupli-
Evolution of FC production

Comparison of FB and FC production among fusaria indicates a correlation between fumonisin chemotype and the three major clades of the FFSC. All fumonisin-producing African and Asian-clade species that have been examined produce predominantly FBs, whereas all fumonisin-producing American-clade species that have been examined produce predominantly FCs (Table 1) (Proctor et al., 2004; Sewram et al., 2005). FB but not FC production has also been reported in two other fungal genera: A. niger (Frisvad et al., 2007) and three species of Tolypocladium (Mogensen et al., 2010). Based on the occurrence of FB production in Aspergillus, Fusarium and Tolypocladium but FC production only in Fusarium, we propose that FB production is ancestral and that a change(s) in FUM8 sequence in the American clade of the FFSC resulted in a switch from FB to FC production. Because American clade species F. anthophilum and F. bulbicola have non-orthologous FUM clusters (e.g. different paralogues in the birth-and-death model or xenologues in the HT model) the FUM8 mutation(s) leading to the change from FB to FC production would have had to occur twice, once in each homologue type, to account for the current distribution of FC production in the American clade and its absence in the African and Asian clades.

Comparison of deduced amino acid sequences of FUM8 homologues revealed a perfect correlation between differences at residue 580 and FB versus FC production (Table 4). Residue 580 was previously proposed (Proctor et al., 2008) to play a role in determining FB versus FC production in Fusarium because mutation of the corresponding residue in another α-oxoamine synthase changed the amino acid substrate specificity of the enzyme (Shoolingin-Jordan et al., 2003). Thus, the current data provide further correlative evidence for the role of this variation in determining whether fungi produce FBs or FCs. In an extension of the hypothesis for evolution of FC production from FB production, residue 580 would have changed independently from an alanine to a valine in the Fum8 encoded by the FUM-cluster homologues in F. anthophilum and F. bulbicola. Such a change in amino acid sequence would require a change in only one nucleotide in the corresponding codon of the FUM8 coding region: GCN (alanine) to GTN (valine). Mutation analysis of residue 580 as well as other Fum8 residues that are consistently different in FC and FB-producing fungi should aid in determining the role of the residues in FB versus FC production.

Conclusion

The results presented here demonstrate that FUM cluster evolution in Fusarium has been complex and is not adequately represented by the species phylogeny. Current data from the FUM cluster do not permit definitive rejection of the multiple HT model of FUM cluster evolution or the mixed model of birth and death with HT. Further evaluation of these models will be facilitated by the availability of an increasing number of genome sequences from Fusarium and other filamentous fungi, which will enable us to search for potential FUM cluster donors (in support of the HT model) or the presence of multiple FUM cluster paralogues or their remnants within a single genome (in support of the birth-and-death model). Regardless of which model is correct, this study adds to a number of previous studies in collectively demonstrating that SMB gene cluster diversity in filamentous fungi has been influenced by a variety of dynamic processes, including cluster duplication and loss (birth and death), balancing selection, shifts in functional constraint, translocation and HT (Ward et al., 2002; Carbone et al., 2007; Khaldi et al., 2008; Khaldi and Wolfe, 2011; Slot and Rokas, 2011). Furthermore, it appears that a combination of these processes has shaped the evolution and current distribution of some SMB clusters and contributed to metabolic diversity in fungi.

Experimental procedures

Strains and media

Fusarium strains used in this study are listed in Table 1. Species selected for the analysis either were previously reported to produce fumonisins or were identified as fumonisin producers during the course of this study. Strains were stored as 15% glycerol stocks at −80°C and grown on V-8 juice agar medium (Tuite, 1969) to prepare spores and mycelia to inoculate other media. For isolation of genomic DNA, Fusarium strains were grown in liquid GYEP medium (2% glucose, 0.1% yeast extract, 0.1% peptone) for 2–4 days at 200 r.p.m. and 28°C. For fumonisin production, strains were grown in cracked maize kernel medium at 25°C as described previously (Seo et al., 2001).

Nucleic acid manipulations

To isolate genomic DNA, cultures grown in liquid GYEP medium were harvested by vacuum filtration, lyophilized, ground to a powder and suspended in extraction buffer (200 mM Tris-Cl, pH 8, 250 mM NaCl, 25 mM EDTA pH 8 and 0.5% SDS) at ~50 mg per 250 μl buffer. DNA used for PCR or to prepare GenomeWalker libraries was purified from this suspension with the DNeasy Plant Mini Kit as described by the manufacturer (Qiagen). Some DNA used only for PCR was prepared with a modified UltraClean protocol (Mo Bio Laboratories). Briefly, the mycelia suspension described above was extracted with an equal volume of a 1:1 (v/v) mixture of
TRIS-equilibrated phenol and chloroform:isoamyl alcohol (24:1). Genomic DNA was purified from the resulting aqueous phase with the UltraClean DNA Purification kit as specified by the manufacturer except that the aqueous phase was mixed with two volumes of sodium iodide solution and 5 μl of UltraBind solution. For genome sequence analysis of *F. anthophilum* and *F. bulbicola*, genomic DNA was obtained by growing the fungi in liquid GYEP medium and extracting DNA with the ZR Soil Microbe DNA MidiPrep kit (Zymo Research Corporation) following the manufacturer’s specifications.

Oligonucleotide primers used to amplify fragments of FUM and PM genes are listed in Supplemental Table S2. Primers were synthesized by Integrated DNA Technologies or by Sigma Life Science-Genosys. Standard PCR amplifications employed Taq DNA polymerase (Qiagen) or Platinum PCR SuperMix polymerase (Invitrogen Life Technologies) and conditions recommended by the manufacturers. PCR amplification employed DNA polymerase contained in the Platinum SuperMix High Fidelity reagents (Invitrogen) using conditions recommended by the manufacturer. PCR products were analysed by agarose gel electrophoresis and were purified from the agarose with the UltraClean DNA Purification protocol (Mo Bio Laboratories) prior to their being used in nucleotide sequencing reactions.

For *F. anthophilum*, *F. bulbicola*, *F. fujikuroi*, *F. globosum*, *F. nygami*, *F. phyllobium* and *F. ramigene*, DNA sequences of regions flanking the FUM cluster were obtained using the Universal GenomeWalker Kit (Clontech Laboratories) as previously described (Proctor et al., 2008; 2009). The initial gene-specific primers for this protocol were designed based on the sequences of fragments of FUM19 and FUM21. For most species, several rounds of GenomeWalker PCR were performed. When the protocol did not yield a PCR product, we obtained sequence data for a given species by standard PCR methods using primers complimentary to genes flanking the FUM cluster in one or more closely related species. Nucleotide sequence analysis employed BigDye Terminator version 3.1 (Applied Biosystems) reagents and UltraClean-purified PCR products (Mo Bio Laboratories) as DNA templates. Sequence reactions were purified with the BigDye Xterminator Purification protocol and analysed with a 3730 DNA analyser at the USDA-ARS-NCAUR DNA Sequence Facility. Sequence data were viewed and edited with Sequencher version 5.0 (Gene Codes Corporation). Additional sequence data for the FUM cluster flanking regions in *F. anthophilum* NRRL 25214 and *F. bulbicola* NRRL 13618 were obtained from genome sequences of these strains. The genome sequence data were acquired with an Ion Torrent system (Life Technologies). Sequence reads were trimmed and assembled and the resulting data were analysed with the CLC Genomics Workbench software (CLC Bio).

**Phylogenetic analysis and genetic distance estimation**

The 12 PM genes used to infer species phylogenies and the proteins they encode were: *CAL1* calmodulin; *CPR1*, NADPH-cytochrome P450 reductase; *FLB1*, regulator of G protein signalling; *HIS3*, histone H3; *LAE1*, velvet complex methyltransferase, *RED1*, reductase; *RPB1*, RNA polymerase II largest subunit; *TPS1*, α,α-trehalose-phosphate synthase; *TUB2*, β-tubulin; and *ZBD1*, zinc-binding dehydrogenase (Steenkamp et al., 2002; O’Donnell et al., 2004; 2007; Malonek et al., 2005). Nucleotide sequences of *A. niger* FUM genes were retrieved from the Joint Genome Institute (http://genome.jgi-psf.org) via BLAST analysis with the corresponding *F. verticillioides* homologues.

Nucleotide sequences were aligned by CLUSTALW+ or MUSCLE as implemented in the program MEGA5 (Tamura et al., 2011). Phylogenetic relationships were inferred from aligned sequences using maximum likelihood (ML), neighbour-joining (NJ) and maximum parsimony (MP) as implemented in MEGA5, as well as by Bayesian estimation using MrBayes 3.1.2 (Ronquist and Huelsenbeck, 2003) as implemented on the CIPRES Science Gateway web server (Miller et al., 2010). The best fit model of molecular evolution for use in ML, NJ and Bayesian analyses was selected based on estimation of the Bayesian Information Criterion (BIC) for each of the 24-nucleotide substitution models included in MEGA5. MP analyses were conducted with CNI branch-swapping on 10 random starting trees. In all three analyses all gaps and missing data were eliminated, and support for individual nodes was estimated via non-parametric bootstrapping with 1000 pseudoreplications of the data. Bayesian estimations of phylogeny employed the best fit model of molecular evolution for each with two independent runs of four simultaneous chains of 10 000 000 generations, sampling every 1000 generations and using a burnin period of 100 000 generations. In order to assess the significance of the discord observed between the FUM and PM gene trees, we conducted Shimodaira–Hasagawa (SH) tests (Shimodaira and Hasagawa, 1999) using RAxML 7.2.8 (Stamatakis, 2006) implemented on the CIPRES Science Gateway web server.

Estimates of genetic divergence were based on the number of synonymous substitutions per synonymous site (*dS*) estimated from concatenated sequence alignments of FUM and PM genes used in the phylogenetic inference analyses described above, and from individual alignments of genes flanking the FUM cluster. Two methods were employed to estimate *dS* values: 1) Maximum likelihood estimates of *dS* were obtained using the CODEML program of PAML 4 (Yang, 2013) implemented on the Phylemon2 web server (Sanchez et al., 2011); and 2) estimates of *dS* were also obtained using the modified Nei-Gojobori method with Jukes-Cantor correction (Zhang et al., 1998) implemented in MEGA5. Standard errors were estimated using the bootstrap method with 1000 pseudoreplications of the data. All positions containing gaps were eliminated from the analyses. The significance of differences in genetic divergence estimates was assessed using Z tests.

**Fumonisins production assay**

Two-week-old cultures grown on 10 g cracked maize kernel medium were extracted with 25 ml of a 50% acetonitrile-50% water solution for 3 h with shaking at 200 r.p.m. Extracts were centrifuged at 380 g for 5 min. The supernatant was then analysed for fumonisins by reverse-phase liquid chromatography coupled to a mass spectrometer (LC-MS) as previously described (Plattner et al., 1996; Proctor et al., 2006). Fumonisins were identified and quantified by comparison of retention
times, masses and mass spectra to those of FB and FC standards (Plattner et al., 1996).

Acknowledgements

We thank Céline Bivort, Stephanie Folmar, Stéphanie Huret, Marcie Moore, Nathane Orwig and Crystal Probyn for their excellent technical support. In addition, we thank Alejandro Rooney for insightful discussions and Christopher Dunlap for assistance with genome sequence analysis. F.V.H. received financial support from the Belgian Federal Science Policy Office (contracts BCCM C2/10/007 and C3/10/003).

References


**Supporting information**

Additional supporting information may be found in the online version of this article at the publisher’s web-site.

Published 2013. This article is a U.S. Government work and is in the public domain in the USA., *Molecular Microbiology*, **90**, 290–306.