European Journal of Plant Pathology **110:** 533–544, 2004.

Synteny in toxigenic *Fusarium* species: The fumonisin gene cluster and the mating type region as examples

Cees Waalwijk[1], Theo van der Lee[1], Ineke de Vries[1], Thamara Hesselink[1], Joop Arts[2] and Gert H.J. Kema[1]
[1]*Plant Research International BV, Business unit Biointeractions and Plant Health, Droevendaalsesteeg 1, P.O. Box 16, 6700 AA Wageningen, The Netherlands (Fax: +31317418094; E-mail: cees.waalwijk@wur.nl);* [2]*Current address: Department of Cell Biology and Immunology, Wageningen University, Wageningen, The Netherlands*

Key words: genome, *Gibberella*, plant pathogen, toxin

Abstract

A comparative genomic approach was used to study the mating type locus and the gene cluster involved in toxin production (fumonisin) in *Fusarium proliferatum*, a pathogen with a wide host range and a complex toxin profile. A BAC library, generated from *F. proliferatum* isolate ITEM 2287, was used to identify chromosomal regions flanking the mating type locus and the gene cluster involved in the biosynthesis of fumonisin. These regions were sequenced and compared with corresponding sequences in other ascomycetes. The results demonstrated that the level of synteny between ascomycetes can vary greatly for different genomic regions and that the level of similarity of genes within a region can also fluctuate strongly. Synteny was found in the regions flanking the mating type idiomorph among ascomycetes that supposedly diverged 280 million years ago. The fumonisin gene clusters of *F. proliferatum* and *F. verticillioides* were completely syntenic but absent in *F. graminearum*. The regions flanking the fumonisin gene clusters were highly dissimilar between *F. proliferatum* and *F. verticillioides*, whereas they formed a continuous region in *F. graminearum*. This indicates that the fumonisin gene cluster has been inserted at different genome locations in both species. Surprisingly low similarity was found between the corresponding genes within the fumonisin cluster of *F. proliferatum* and *F. verticillioides*, compared to other genomic sequences indicative for two independent acquisition events from distinct genetic sources. The results demonstrate the power of comparative genomics for gene annotation and for studies on the evolution of genes, gene-clusters and species.

Introduction

Filamentous ascomycetes and related fungi imperfecti represent a wide range of organisms, many of which are pathogenic to plants. The diversity of lifestyles is enormous, ranging from strictly vegetative to primarily sexually reproducing and from specialized biotrophs to necrotrophs and saprophytes. The differences in host range among these fungi are of special interest to plant pathologists. While some ascomycetes infect a wide variety of plant species, others are strongly restricted in their host range. The origin of such differences has been studied for several decades but is in most cases poorly understood. Yet, these various lifestyles are likely to be reflected in the genome content of the different species. With the availability of large amounts of sequence data, these questions can now be addressed from a genomic perspective. Genome comparisons can be used to address specific questions about fungal life style or specialization and can help to elucidate the evolution of species, specialized forms and races. The level of conservation of gene sequences and gene order will be instrumental in reconstructing past events and seeing which regions are under the stronger evolutionary selection pressure.

Several classes of genes are of particular interest. The first group comprises those common to all organisms, such as rRNA, the ITS region, elongation factor 1α, β-tubulin, histone H3 and actin, which have been used to construct phylogenetic trees (e.g. Glass and Donaldson, 1995; O'Donnell et al., 2000b). A second class is formed by genes that are involved in mating as they may restrict the number of possible partners in the sexual cycle and therefore are important in speciation. Thirdly, in plant pathogens, genes involved in pathogenicity, particularly those that may be involved in host specificity are important. The genus *Fusarium* forms an excellent starting point for comparative studies. Several hundred different species have been identified within this genus, encompassing important plant pathogens, saprophytes and even antagonists (Larkin and Fravel, 1999). Their phylogenetic relations have been determined by sequence analysis of various genes (O'Donnell et al., 2000a). The genus comprises species with a wide range of life styles. Many *Fusarium* species are notorious for the production of mycotoxins and many of these toxins are specific to particular *Fusarium* species (Bottalico, 1998). Some of these toxins may be involved in pathogenicity. Trichothecenes clearly play a role in virulence since disruption of the trichodiene synthase gene reduced virulence (Proctor et al., 1995; Desjardins et al., 1996). For other mycotoxins such as fumonisin, however, the situation is less clear. Field isolates that do not produce fumonisin are very rare, suggesting an ecological function for this secondary metabolite. Moreover, an association was found between virulence and the level of fumonisin production (Desjardins et al., 1995). However, no effect on virulence was found after disruption of essential genes in the fumonisin pathway. Isolates of *F. verticillioides* that were unable to produce fumonisin were still virulent on maize (Desjardins and Plattner, 2000). Although the role of toxins in pathogenesis is not always clear, the production of mycotoxins by *Fusarium* species is a major concern for food and feed safety. Fumonisins are hepato- and nephro-carcinogenic in animal models (Gelderblom et al., 1991; Howard et al., 2001) and consumption of fumonisin is associated with high incidence of oesophageal cancer in South Africa (Rheeder et al., 1992) and China (Wang et al., 2000).

In this study, we analyzed *F. proliferatum,* a member of the *Gibberella fujikuroi* complex (O'Donnell et al., 2000b), which is known for its wide host range (Elmer, 1995; Moretti et al., 1997b; Ocamb et al., 2002) and its capacity to produce a variety of toxins (Moretti et al., 1997a; Munkvold et al., 1998; Moss, 2002). A BAC library of *F. proliferatum* was produced and the chromosomal regions containing the mating type locus and the fumonisin biosynthesis gene cluster were identified and sequenced. These sequences were used to study synteny among filamentous fungi in order to improve understanding of the phylogenetic relationships between these fungi.

Materials and methods

Fungal strains

Fusarium proliferatum isolates ITEM 2287 and ITEM 2400 have opposite mating types and were selected for the isolation of the idiomorphs in this species (Munkvold et al., 1998). Based on its capacity to produce several mycotoxins, including fumonisin, beauvericin and fusaproliferin, isolate ITEM 2287 was selected to generate a BAC library of this species.

Preparation of high-molecular-weight DNA from Fusarium proliferatum *ITEM 2287*

Spores of isolate ITEM 2287 were grown overnight in Potato Dextrose Broth and mycelium was collected by filtration and subsequently washed with water and twice with protoplast buffer (1.2 M $MgSO_4$, 10 mM NaPi pH 5.8) before incubation at 30 °C for 1.5 h in a 2.5% glucanex (Novozymes, Dittingen, CH) protoplast buffer solution. Protoplasts were collected by centrifugation (10 min at 3000 rpm) after addition of one volume STC (1 M sorbitol, 50 mM $CaCl_2$, 50 mM Tris–HCl pH 8.0). DNA plugs were produced by re-suspending the protoplasts in protoplast buffer, heating to 40 °C and adding an equal volume of 1.2% InCert agarose (final concentration 1.0E8 per ml.). The plugs were incubated twice for 24 h at 50 °C with lysisbuffer (0.5 M EDTA, 10 mM Tris–HCl pH 8.0, 1% N laurylsarcosine and 1 mg ml^{-1} proteinase K), washed twice with PMSF (10 mM

Tris–HCl pH 8.0, 10 mM EDTA, 1 mM PMSF), once with buffer (10 mM Tris–HCl pH 8.0, 10 mM EDTA) and stored in 0.5 M EDTA.

Partial restriction digests were performed by titration of decreasing units of *Hind*III to the agarose plugs using an incubation time of 30 min. Subsequently, the plugs were loaded in a 1% LMP agarose CHEF gel and DNA in the size range of 80–150 kb was excised from the gel. The DNA was compressed by pulse-field gel electrophoresis (10 h, 3 V/cm, 5 s switchtime) and stored in 0.5 M EDTA pH 8.0 until ligation. An agar plug containing 100 μg partially digested DNA was dialyzed in TE buffer, melted at 70 °C for 3 min, and the agarose was digested using GELase (Epicentre) according to the instructions of the manufacturer. The DNA concentration was estimated by loading an aliquot of the DNA in a 0.8% agarose gel. Forty ng of partially digested DNA was ligated into pBeloBAC11 (Kim et al., 1996) and the ligation mixture was transformed into *Escherichia coli* DH10B ElectroMAX (Invitrogen). Transformants were subsequently plated on LB agar with 12.5 μg ml^{-1} chloramphenicol. Approximately 8000 BAC clones were picked by a Flexys™ Picker (Applied Biosystems) into 384-well microtiter plates containing LB freeze (LB broth with 36 mM K_2HPO_4, 13.2 mM KH_2PO_4, 1.7 mM sodium citrate, 400 μM $MgSO_4$, 6.8 mM $(NH_4)_2SO_4$, 4.4% glycerol), grown for 16 h at 37 °C and stored at −80 °C. For the determination of the insert size, DNA of 18 random clones was isolated and their insert size was calculated after restriction digestion and gel electrophoresis.

Screening of the BAC library

A 3D-screening of the library was performed by stacking 384-well plates and pooling colonies from plates ($n = 21$), from rows ($n = 16$) as well as from columns ($n = 24$) into 61 pools from which DNA was isolated; 10 ng of pooled DNA was used in PCR reactions. To identify clones containing mating type *mat1-2-1,* primers *Fp*HMG-F (5′-GCGATCAAGGCTCAACGCC-3′) and *Fp*HMG-R (5′-CTGATCCGCCATCTGCTTG-3′) were used for 2′ at 94 °C, followed by 40 cycles of 1′ 95 °C, 30″ at 65 °C and 1′ at 72 °C. The fumonisin cluster was identified using primers PKS-exon5-F (5′-TGGAAATGGATCTCTTCGAGGC-3′) and PKS-exon6-R (5′-GCACACGCGCTTCCCAATCTCAT-3′) for 2′ at 94 °C, followed by 40 cycles of 1′ 95 °C, 30″ at 60 °C and 1′ at 72 °C. Clones at the intersections of the three coordinates were verified by PCR and overlap between positive clones was determined by restriction profiles as well as by PCR using primers designed from BAC-end sequences.

Sequencing of chromosomal regions

DNA from selected BAC clones was isolated using the Large-Construct Kit from Qiagen. The DNA was sheared by sonification to obtain a size range of 1.0–1.5 kb. These fragments were cloned into pBluescript SK^+ and transformants were picked by a Flexys™ Picker (Applied Biosystems) into 384-well microtiter plates containing LB freeze. Plates were incubated for 16 h at 37 °C and stored at −80 °C. Plasmid inserts were amplified by PCR and sequenced using BigDye technology. Primers designed from the BAC-ends sequences of the fumonisin region of two overlapping clones were applied in a long ranch PCR, using Herculase *Taq* polymerase (Roche) to generate an additional 8 kb amplicon that was sheared and cloned as described above.

Sequence processing and annotation

BACs were sequenced by a shot-gun approach until an approximately sixfold coverage was achieved. The sequences were clipped for quality as well as vector sequence and assembled using GAP4 (Bonfield et al., 1995). Specific primers were developed to close gaps and to improve the sequence quality in regions of low quality. Identification of ORFs was performed by using BLASTX comparisons to the non-redundant protein database from GenBank and fungal databases. Splicing of introns was performed manually using the intron consensus splice sites (Bruchez et al., 1993) and, whenever possible, using the homology to known DNA and/or protein sequences.

Comparative studies

For comparative studies, the assembled genome sequences generated by the Whitehead institute

with the following URLs were used:
http://www.broad.mit.edu/annotation/fungi/aspergillus
http://www.broad.mit.edu/annotation/fungi/fusarium.
http://www.broad.mit.edu/annotation/fungi/magnaporthe.
http://www.broad.mit.edu/annotation/fungi/neurospora.

Comparative studies were performed using the BLAST algorithms, BLASTN, BLASTX and BLAST 2 SEQUENCES (Tatusova and Madden, 1999) as well as the Megalign module of DNAstar (Lasergene). Initial studies were performed using the non-redundant protein database. Selected contigs from assembled fungal genomes were used in BLAST 2 SEQUENCES comparisons at the nucleotide as well as the protein level. Finally, all predicted proteins were blasted against the corresponding proteins in homologous regions. The level of similarity between proteins was expressed as the percentage identity as it was retrieved from the BLAST output.

Results

A BAC library for Fusarium proliferatum

A total of approximately 8000 clones of *F. proliferatum* isolate ITEM 2287 were picked and ordered in 384-well microtiter plates. Eighteen random clones were tested for their insert size, which ranged from 20 to 70 kb with an average insert length of 36 kb. With the expected genome size of 36 Mb, the library contains approximately eight genome equivalents. The library was screened by PCR with primers homologous to five single copy genes. All five genes were recovered from the library albeit with lower numbers than the expected average of eight copies.

Mating type locus

Primers homologous to the conserved HMG-box of *mat1-2-1* were tested on *F. proliferatum* ITEM 2287 and the resulting 163 bp fragment was sequenced to verify the amplification of the conserved part of the *mat1-2-1* gene. After this test, these primers were used for a three-dimensional screening of the BAC library. Only a single positive clone, BAC 17D1, was identified and sequenced. After assembly a contig of 31,941 bp was obtained in which 11 ORFs were identified. To obtain the DNA fragment corresponding to the idiomorph of the opposite mating type, various primers described in the literature (Arie et al., 1997) were tested using different PCR conditions on several *F. proliferatum* isolates from both mating types. However, a fragment of the expected size could not be amplified. Finally, primers with homology to the flanking regions of the *mat1-2* idiomorph on BAC 17D1 were designed, a strategy that was successful for cloning the *mat1-1* idiomorph of *Mycosphaerella graminicola* (Waalwijk et al., 2002). Using these primers a 4.5 kb amplicon was obtained in *F. proliferatum* isolate ITEM 2400 and this PCR product was sequenced by primer walking from both ends. In this sequence three ORFs were identified with similarity to various *mat1-1-1*, *mat1-*

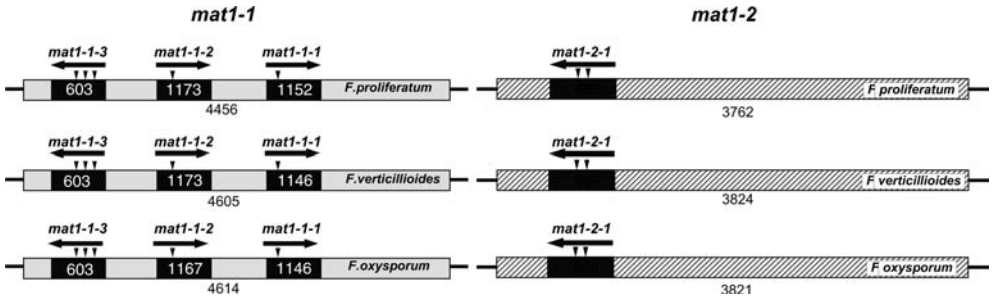

Figure 1. Organization of the genes in the mating type idiomorphs of *F. proliferatum*, *F. verticillioides* and *F. oxysporum*. The idiomorphs *mat1-1* (left) and *mat1-2* (right) are boxed and the position of the ORFs is indicated in black. The arrows indicate the direction of transcription and the position of the introns are specified by filled triangles. Sizes of the genes are shown in boxes and the size of the idiomorph is listed below the boxes.

1-2 and *mat1-1-3* genes (Figure 1). The sequences of the two mating type idiomorphs of *F. proliferatum* were found to be highly dissimilar except for the ends. Both sequences were compared with the published mating type idiomorphs of *F. verticillioides* and *F. oxysporum* as well as the corresponding ORFs (Table 1). However, not all the proteins described by Yun et al. (2000) are available in GenBank, while for others, dissimilarities were found between the predicted protein in the article and the predicted protein present in GenBank. The coding regions of the *F. proliferatum* genes were compared with the published proteins rather than the proteins in GenBank with the following modifications. We noticed that the translation of the *F. verticillioides mat1-1-2* gene did not result in the predicted protein described by Yun et al., (2000). The sequence of the *mat1-1-2* genes in *F. proliferatum* region was identical to that in *F. verticillioides* but in *F. oxysporum* a stretch of four C-residues was found instead of three in *F. verticillioides* as well as *F. proliferatum*. Upon removal of this C-residue the predicted MAT1-1-2 proteins of the three species were much more similar. Secondly, we observed several alternative starting positions for the *mat1-1-3* gene. Comparison of the *F. proliferatum*, *F. verticillioides*, *F. oxysporum* and *F. graminearum* sequences showed that the only ATG start codon for *mat1-1-3* available in all species is upstream of the starting codon for *mat1-1-3* predicted by Yun et al. (2000). This alternative starting codon was used to predict the proteins in all species for comparative analyses. After these adjustments the genomic structure in these different *Fusarium* spp. was highly similar as is shown in Figure 1. All three species have the same number of genes and the order and orientation is identical. In each homolog the same number of introns was observed.

Table 1. Similarity among the mating type proteins of different *Fusarium* spp.

	F. verticillioides	*F. oxysporum*	*F. graminearum*
F. proliferatum mat1-1-1	93	89	50
F. proliferatum mat1-1-2	91	83	45
F. proliferatum mat1-1-3	94	93	64
F. proliferatum mat1-2-1	94	93	53

Fumonisin region

Primers PKS-exon5F and PKS-exon6-R have been designed on the basis of the sequence of *fum5,* the polyketide synthase gene of the fumonisin cluster of *F. verticillioides* (Proctor et al., 1999). These primers amplified a fragment of 1.3 kb in *F. proliferatum* isolate ITEM 2287. The sequence of this fragment demonstrated strong homology to *fum5*. Subsequent three-dimensional screening of the BAC library with these primers resulted in three positive BAC clones: 3O10, 5C1 and 7B6. Based on restriction fragment profiles and results from PCR reactions using primers derived from BAC-end sequences, clone 7B6 was selected for sequencing. The sequenced region was extended by the amplified 8 kb fragment and the assembly in the fumonisin cluster eventually resulted in a contig of 77,606 bp that contained 25 ORFs.

Homology and synteny of the F. proliferatum *mating type region with related species*

The sequence of the *mat1-2* mating type region of *F. proliferatum* ITEM 2287 contained 11 ORFs that were blasted against the non-redundant database using BLASTX. Based on their similarity, nine out of these 11 ORFs could be annotated (Table 2). The two remaining ORFs were homologous to predicted genes in *Neurospora crassa* with unknown functions. For *F. verticillioides* and *F. oxysporum* only limited sequence information is available outside the idiomorph. Since the available genome sequences of *F. graminearum* and *N. crassa* are derived from isolates with the *mat1-1* genotype, the *mat1-2* idiomorph sequence of ITEM 2287 was substituted *in silico* by the *mat1-1* idiomorph of ITEM 2400 to allow more appropriate comparison with these fungi.

The gene order, gene orientation and level of similarity of this synthetic *mat1-1* region were compared with the gene order in *F. graminearum*, *N. crassa*, *Magnaporthe grisea* and *Aspergillus nidulans*. In all cases strong similarity was found with regions in these genomes. In *F. graminearum* and *N. crassa* the similarity was found to reside on two consecutive contigs, spanning a region of 36 kb in *F. graminearum* and 116 kb in *N. crassa* (Figure 2). In *M. grisea*, the similarity with this region was found to reside on two contigs, separated by three small contigs, covering

Table 2. Open Reading Frames in the vicinity of the mating type locus of *F. proliferatum* and the homologs in the GenBank non-redundant database or fully sequenced genomes of related ascomycetes

F. prolif eratum	*GenBank*[a]	*N. crassa*				*M. grisea*				*A. nidulans*				*F. graminearum*		
Gene	Gene	Gene[b]	Contig[c]	Identity[d]	*e*-value	Gene[b]	Contig[c]	Identity[d]	*e*-value	Gene[b]	Contig[c]	Identity[d]	*e*-value	Gene	Contig[c]	Identity[d]
ORF1	DNA-poly-merase	NCU01951.1	3.86	53	0	MG02986.3	2.600	58	0	AN4789.2	1.81	52	0	FG08898.1	1.359	83
ORF2	*slu 7*	NCU01950.1	3.86	61	*e*-134	MG02985.3	2.600	67	0	AN4788.2	1.81	59	*e*-161	FG08897.1	1.359	87
ORF3	S9	NCU01949.1	3.86	91	*e*-93	MG02952.3	2.596	89	*e*-92	AN4803.2	1.81	90	*e*-92	FG08896.1	1.359	99
ORF4	L21	NCU01948.1	3.86	76	*e*-65	MG02953.3	2.596	70	*e*-38	AN4802.2	1.81	70	*e*-61	FG08895.1	1.359	90
ORF5	*sla2*	NCU01956.1	3.86	84	0	MG02949.3	2.596	83	0	AN2756.2	1.81	71	0	–[f]	1.359/1358	96
ORF6a	*mat1-2-1*	–	–	–	–	MG02978.3	2.600	56	*e*-22	AN4734.2	149	41	*e*-14	FG08893.1	1.358	53
ORF6	*mat1-1-1*	NCU01958.1	3.86	33	*e*-12	–	–	–	–	AN2755.2	149	34	*e*-18	FG08892.1	1.358	50
ORF7	*mat1-1-2*	NCU01959.1	3.87	34	4.3	–	–	–	–	–	–	–	–	FG08891.1	1.358	45
ORF8	*mat1-1-3*	NCU01960.1	3.87	24	0.007	–	–	–	–	–	–	–	–	FG08890.1	1.358	64
ORF9	DNA lyase	NCU01961.1	3.87	52	0	MG02980.3	2.600	57	0	AN4736.2	1.80	48	*e*-146	FG08889.1	1.358	74
ORF10	ORF 1	NCU01962.1	3.87	61	*e*-35	MG02983.3	2.600	61	*e*-35	AN4737.2	1.80	59	*e*-33	FG08888.1	1.358	85
ORF11	APC	NCU01963.1	3.87	55	0	MG02981.3	2.600	49	0	AN4735.2	1.80	40	*e*-147	FG08887.1	1.358	89
ORF12	CIA30	NCU01975.1	3.87	56	*e*-81	MG03723.3	2.723	58	*e*-84	AN4740.2	1.80	41	*e*-28	FG08886.1	1.358	86
ORF13	ORF 2	NCU01983.1	3.87	28	*e*-19	MG03720.3	2.723	29	*e*-16	no hit	–	–	–	FG08885.1	1.358	76

Shaded areas contain the idiomorphic genes, *mat1-1-1*, *mat1-1-2* and *mat1-1-3* present in the *mat1-1* allele and *mat1-2-1* in the *mat1-2* allele.

[a]Best descriptive hit in the non-redundant database: *slu7*, putatively involved in preRNA splicing; S9 and L21, are ribosomal proteins; *sla2*, cytoskeleton assembly protein; APC, anaphase promoting complex, putatively involved in the regulation of meiotic division and CIA30, complex I intermediate-associated protein. The predicted ORFs have no assigned function, yet their presence in *F. proliferatum* and homologs in the other fungi warrant functional analyses of these genes.

[b,c]Annotated genes in *N. crassa*, *M. grisea*, *A. nidulans* and *F. graminearum* are given as NCUxxxxxx.1, MGxxxxx.3, ANxxxx.2 and FGxxxxxxx.1 respectively, with their physical localization.

[d]The identity of the homologous proteins to the ORF in *F. proliferatum* is given in percentages.

[e]Not applicable, due to the absence of the homolog.

[f]Not applicable, because this gene was not annotated.

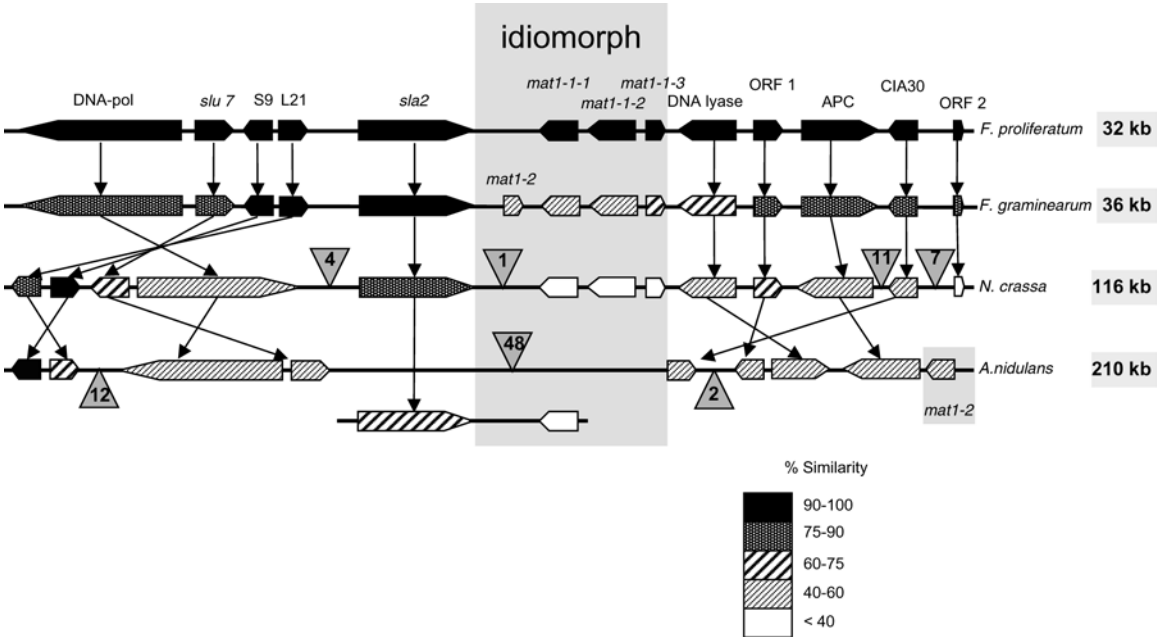

Figure 2. Alignment of the genes in the mating type regions of *F. proliferatum* with *F. graminearum*, *N. crassa* and *A. nidulans* respectively. Consecutive contigs are depicted as a single line on which the genes with similarity to the *F. proliferatum* sequence are shown as arrows that also indicate the direction of transcription. The best descriptive hit in the non-redundant database is listed above each of the genes. Predicted genes present in the contigs of *N. crassa* and *A. nidulans* that do not show similarity to *F. proliferatum* genes in the mating type region are marked by triangles in which the number of predicted genes that do not show similarity is listed. The size of the corresponding regions in these fungi is given in kb at the right. The level of similarity by BLASTP comparison is visualized by different shades of gray as indicated in the box below.

approximately 130 kb. In *A. nidulans*, the genes homologous to the DNA polymerase, *slu7*, S9, L21, DNA lyase, predicted ORF 1, APC and CIA30 were located on two consecutive contigs, spanning 210 kb, while the gene homologous to *sla2* gene was found adjacent to a *mat1-1-1* homolog elsewhere in the genome. In *A. nidulans* no homolog of the predicted ORF2 was found (Figure 2). In all cases the hits described here were the best hit in the genome sequence indicating that these genes are orthologs.

The order and orientation of the entire set of genes is completely conserved between the heterothallic species *F. proliferatum* and homothallic species *F. graminearum*, except for the obvious presence of the *mat1-2-1* gene, situated between *sla2* and *mat1-1-1* in *F. graminearum*, but missing in the *mat1-1* locus of *F. proliferatum*. In *N. crassa* the order of genes is similar but not completely identical and several insertions were found. A block of four genes (DNA polymerase, *slu7*, S9 and L21) seems to be inverted. Four putative genes occur between the DNA polymerase and *sla2*, one ORF between *sla2* and *mat1-1-1*, eleven ORFs between APC and CIA30 and seven putative genes between the CIA30 gene and predicted gene 2. When the gene order and gene orientation of *F. proliferatum* were compared to that of *A. nidulans*, most genes were found on two adjacent contigs (Figure 2). However, several rearrangements were found and blocks of genes appear to have inserted. The *sla2* homolog was found on a completely unrelated contig and this homolog is flanked by the homolog of *mat1-1-1*. Finally, in *M. grisea* more small inversions and insertions have occurred. In the overall comparison of these fungal species it appears that in particular neighboring genes with opposite directions of transcription tend to be linked among highly diverged ascomycetes.

When the predicted proteins are compared by BLAST 2 SEQUENCES using the BLASTP algorithm or, in case of the comparison with the *sla2* homolog of *F. graminearum*, with the tBLASTn algorithm, the percentage of amino acid identity varied from 24% to 99% indicating that different

genes in this region have different levels of conservation. The homologs of the ribosomal proteins S9 and L21 as well as the *sla2* protein are particularly highly conserved (70–99%). On the other hand, the sequence conservation is very low in the mating type genes (24–64%). The similarity of the *F. proliferatum* proteins is always highest to *F. graminearum*, somewhat less to *N. crassa* and *M. grisea* and lowest similarity was found with *A. nidulans* (Table 2). The steep drop in sequence similarity between the mating type proteins of *F. proliferatum* and *F. graminearum* is remarkable, since the percentage of sequence conservation remained similar with the homologs of *N. crassa*, *M. grisea* and *A. nidulans*.

Homology and synteny of the fumonisin gene cluster and surrounding genes of F. proliferatum *with related species*

The contig containing the fumonisin gene cluster and flanking sequences is composed of two distinct regions. The first half of the contig contains ORFs also found in other, fully sequenced, ascomycetes, whereas the second half only showed homology to the fumonisin gene cluster of *F. verticillioides*. This second half was completely syntenic over the entire gene cluster, ranging from the polyketide synthase on one end of the cluster to the ABC transporter as the other end (Proctor et al., 2003). The genes and non-coding intergenic regions in *F. proliferatum* and *F. verticillioides* showed a high degree of similarity and the identity among the encoded proteins varied from 73% for Fum17p and Fum18p to 89% from the putative dioxygenase Fum9p (Table 3). In sharp contrast to the similarity within the cluster, sequences in the flanking regions are completely unrelated between *F. proliferatum* and *F. verticillioides* (Figure 3).

In the case of *F. proliferatum*, as well as *F. verticillioides,* the sequences flanking the fumonisin gene cluster showed similarity to *F. graminearum*, *N. crassa*, *M. grisea* and *A. nidulans*. Comparison to the *F. graminearum* genome sequence revealed that the upstream region in *F. proliferatum* showed similarity to two different regions in the *F. grami-*

Table 3. Open Reading Frames in the sequenced BAC of *F. proliferatum*

ORF	Position	Geneproduct	*e*-value	Organism	% identity
1	0.4–2.8	Cellobiase	*e*-105	*Neurospora crassa*	
2	6.1–7.6	P450 monooxygenase	*e*-40	*Burkholderia fungorum*	
3	11.0–12.6	Putative transporter	*e*-35	*Neurospora crassa*	
4	18.3–19.8	Dioxygenase	*e*-22	*Escherichia coli*	
5	20.2–21.9	P450 monooxygenase	*e*-97	*Aspergillus oryzae*	
6	23.6–25.4	MFS transporter	*e*-65	*Neurospora crassa*	
7	25.9–27.2	Epoxide hydrolase	*e*-28	*Xanthophyllomyces dendrorhous*	
8	28.5–29.5	Maltase	*e*-175	*Aspergillus oryzae*	
9	30.5–32.0	Maltose permease	*e*-50	*Aspergillus oryzae*	
10	33–35.5	AmyR transcription regulator	*e*-59	*Aspergillus oryzae*	
11	39.9–48.1	Polyketide synthase (FUM5)	0.0	*F. verticillioides*	85
12	48.5–52.1	P450 monooxygenase (FUM6)	0.0	*F. verticillioides*	85
13	52.7–54.0	Alcohol dehydrogenase (FUM7)	0.0	*F. verticillioides*	82
14	54.0–57.0	Amino transferase (FUM8)	0.0	*F. verticillioides*	77
15	57.7–58.6	Dioxygenase (FUM9)	0.0	*F. verticillioides*	89
16	58.9–60.4	Fatty acid-CoA synthetase (FUM10)	0	*F. verticillioides*	80
17	61.1–62.3	Tricarboxylate transporter (FUM11)	*e*-121	*F. verticillioides*	78
18	62.8–64.5	P450 monooxygenase (FUM12)	0	*F. verticillioides*	88
19	65.2–66.3	Short-chain dehydrogenase (FUM13)	*e*-159	*F. verticillioides*	77
20	66.6–68.6	Peptide synthase (FUM14)	0	*F. verticillioides*	81
21	69.1–70.9	P 450 monooxygenase (FUM15)	0	*F. verticillioides*	79
22	71.1–73.3	Fatty acid-CoA synthetase (FUM16)	0	*F. verticillioides*	83
23	73.7–75.0	Longevity assurance factor (FUM 17)	*e*-156	*F. verticillioides*	73
24	75.1–76.6	Longevity assurance factor FUM 18)	*e*-160	*F. verticillioides*	73
25	76.9–end	ABC transporter (FUM19)[a]	*e*-68	*F. verticillioides*	81

Shaded ORFs belong to the fumonisin biosynthesis gene cluster.

[a]The contig of *F. proliferatum* contained only part of the ABC transporter gene, encoding the first 175 amino acids.

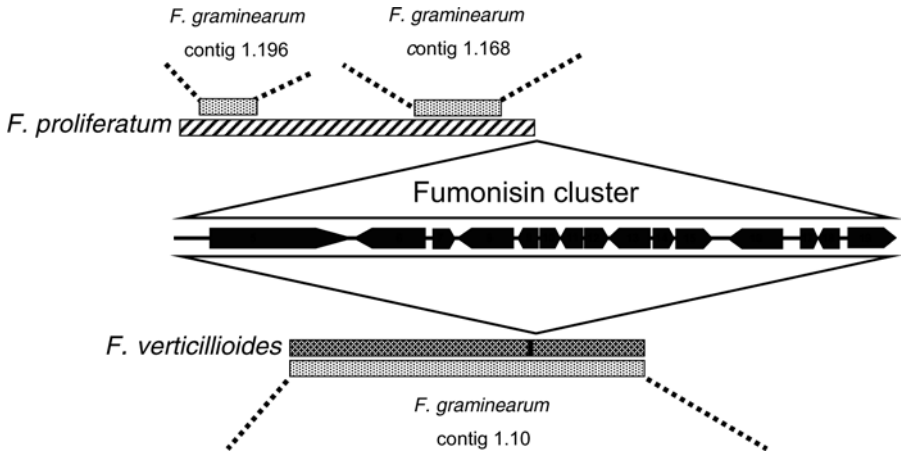

Figure 3. Alignment of the genes in the fumonisin cluster of *F. proliferatum* and *F. verticillioides*. The fumonisin gene cluster is depicted as the numbered arrows 5–19, where the arrows indicate the direction of transcription. Similarity of the flanking regions of both species with contigs of *F. graminearum* is given as parallel bars. *F. verticillioides* showed similarity to contig 1.10 of *F. graminearum* and in *F. proliferatum* similarity with two unrelated contigs was found. (contigs 1.168 and 1.196).

nearum genome (contigs 1.168 and 1.196; Figure 3). These contigs are members of different supercontigs (http://www-genome.wi.mit.edu/annotation/fungi/fusarium). In *F. verticillioides* on the other hand, both the upstream as well as the downstream regions showed high similarity to a contiguous region on the *F. graminearum* contig 1.10.

Discussion

A BAC library was generated from *F. proliferatum* isolate ITEM 2287. Based on the sequenced genomes of *F. graminearum*, *N. crassa*, *M. grisea* and *A. nidulans*, *F. proliferatum* was estimated to have a genome size of around 36 Mb. Initial calculations suggested that the size of the genomic library would exceed a coverage of eight genome equivalents. However, all single copy fragments yielded lower numbers of clones. Possibly, we overestimated the insert size based on a relatively small number of clones tested. Alternatively, the BAC library may be biased for specific chromosomal regions or the genome size of *F. proliferatum* may be larger. Nevertheless, enough BACs were identified to cover the regions of interest. Presently, most fungal genomes are sequenced by shot-gun sequencing of entire genomes but BAC libraries remain efficient for studies of targeted genome regions as shown in this study. A three dimensional PCR screen was used to identify BAC clones corresponding to the mating type and the fumonisin regions, which are of major importance to plant pathologists.

The level of synteny in the mating type region between distantly related ascomycetes is extraordinary. The region is not only completely syntenic between *F. proliferatum* and *F. graminearum*, but the level of synteny is also high between *F. proliferatum* and *N. crassa* or *M. grisea*, species that are supposed to have diverged approximately 175 million years ago (Berbee and Taylor, 1993). Indeed, according to the same authors, *F. proliferatum* and *A. nidulans*, species that separated about 280 million years ago, share regions that are still partly syntenic. For other fungi, little sequence information of regions flanking the mating type is available, making comparisons less meaningful. However, when genes of the flanking region are available they support the high level of synteny described in this study. Homologs of the DNA lyase and the anaphase promoting complex in the vicinity of the idiomorph in *Mycosphaerella graminicola* were identified (Waalwijk et al., 2002) and in *Leptosphaeria maculans* a homolog the DNA lyase was found adjacent to the mating type locus (Cozijnsen and Howlett, 2003).

Interestingly, *A. nidulans* also showed homology for the mating type genes. A gene homologous to the *mat1-1-1* genes (*e*-value 8.00 *e*-18, 34% identity at the amino acid level) was found on contig 1.49, whereas a gene with similarity to the

HMG box was found on contig 1.80 (*e*-value 1.00 *e*-14, 41% identity at the amino acid level). Contigs 1.80 and 1.81 correspond to linkage group III and contig 1.49 is located in linkage group IV of *A. nidulans*, demonstrating that in this case, genes with similarity to the mating type genes are located on different parts of the genome, although analyses support the view that they were once at the same position. Both mating type homologs carry part of the mating type region present in *N. crassa* and *F. graminearum*. The *sla2* gene is close to the *mat1-1-1* homolog and the HMG box protein resides on two adjacent contigs containing eight other genes also found in the vicinity of the mating type idiomorphs in *N. crassa* and *F. graminearum*. The long evolutionary history of five different species makes a comparison of this region particularly relevant. The changes observed are predominantly caused by insertions and inversions. Genes in opposite orientation share the same upstream region of the gene, such as DNA-polymerase and *slu7*, the ribosomal protein genes S9 and L21, as well as DNA-lyase and ORF1 remain couples in all the genomes analyzed. The upstream regions of genes are known to contain the promoter sequences and in these cases the promoter sequences may be twinned.

The level of synteny found at the mating type locus might be exceptional. Additional BACs of *F. proliferatum* have been sequenced and in one case the synteny with *F. graminearum* was high but only fragmented synteny was found with *N. crassa*, *M. grisea* and *A. nidulans* (data not shown). As described, the syntenies in the regions flanking the fumonisin clusters of either *F. proliferatum* or *F. verticillioides* to *N. crassa* and *M. grisea* were also less evident. A reason for exceptional conservation of the gene order in the mating type region may be the presence of the two highly different idiomorphs, which may block genetic recombination, that may be a possible mechanism for gene dispersion. The X chromosome in mammals, which is also limited in its possibility for genetic recombination is also the chromosome that shows the highest synteny between mouse and man. (Waterston et al., 2002)

Another striking feature is the variation in the level of similarity between neighboring genes. Whereas some genes are strongly conserved, e.g., the ribosomal proteins L21 and S9 and the SLA2 protein, others such as the mating type genes are only weakly conserved. Particularly noteworthy is the drop in the similarity level of the mating type genes between *F. proliferatum* and *F. graminearum;* where other genes in the region have a similarity ranging from 74% to 99%, the mating type genes have a similarity of between 45% and 65%. This is consistent with the level of similarity of the mating type genes among *Fusarium* species and also explains the difficulty in identifying the mating type genes in various filamentous fungi (Waalwijk et al., 2002; Goodwin et al., 2003). The mating type genes may be involved in self–nonself recognition and lower levels of similarity may prevent mating and thus fix diverged genotypes and drive the generation of different species. It is possible that parts of the mating type genes are under diversifying selection. To test this, the rate of non-synonymous substitutions was compared with the synonymous substitutions on the nucleotide level. No significant bias for non-synonymous substitutions was found (data not shown).

We studied the genomic region in *F. proliferatum* that has similarity to genes postulated to be involved in the fumonisin biosynthesis pathway in *F. verticillioides* (Seo et al., 2001; Proctor et al., 2003). The sequence generated ends within the gene encoding an ABC transporter, that delimits the fumonisin cluster in *F. verticillioides* (Proctor et al., 2003). The sequences outside this cluster are highly dissimilar except for a 2 kb region just upstream of *fum5*. Within the fumonisin gene cluster, the level of similarity between *F. proliferatum* and *F. verticillioides* is similar for all genes in the cluster (77–89% identity on protein level, Table 3), which is lower than expected based on other genes such as the elongation factor 1α and β-tubulin that are 100% conserved at the amino acid level. Also, the mating type genes *mat1-1-1*, *mat1-1-2*, *mat1-1-3* and *mat1-2-1* are more identical than the genes in the fumonisin gene cluster. This is surprising because, as discussed in the previous paragraph, the mating type genes are usually more variable than other genes (Goodwin et al., 2003). Unfortunately, not many genes of *F. proliferatum* and *F. verticillioides* can be compared because only a limited number of genes are known in both species. Genes outside the cluster cannot serve as a reference because these genes are not homologous. Analyses demonstrated that the fumonisin gene cluster was integrated at different genome positions and might indicate that this region was in-

serted in the genome in two independent events. The low similarity of the genes in the fumonisin cluster adds to this reasoning and suggests that the introductions might have come from a genetically different source. In this context it is relevant that several non-producing strains of *F. verticillioides* lack *fum5*, *fum6* and *fum8* (determined by Southern hybridization; Maite Gonzalez-Jaen, this issue), which may indicate the absence of the entire cluster. Although introduction of this cluster from a foreign source seems plausible, no obvious differences in GC content or codon usage was found between the fumonisin gene cluster and the surrounding sequences.

Fusarium proliferatum is poorly studied at the molecular level. Yet, using a comparative genomic approach, genes involved in mating and fumonisin production were identified and some light was shed on the possible evolution of these regions in other species. Our analyses provide insight into the rates of evolution of the genes in both regions. In cases where the sequence similarity is very low, synteny is helpful in identifying orthologous genes. In the future, more comprehensive comparative genomics of ascomycetes will be instrumental in identifying (toxin) pathways, and will address longstanding questions such as the evolution of species and host range, which will give plant pathology a completely new perspective.

Acknowledgements

The authors wish to thank Dr. M. Gonzalez-Jaen (Universidad Computense, Madrid, Spain) for the primers PKS-exon5-F and PKS-exon6-R. Drs. Logrieco and Moretti (ISPA, Bari, Italy) kindly provided the strains ITEM 2287 and ITEM 2400. We also like to thank Dr. R. Hoekstra for critical reading of the manuscript. This work was partly financed by the EU-DeTox Fungi project (QLRT-CT-1998-01380) and Dutch Ministry of Agriculture, Nature and Food Quality, program 397-1.

References

Arie T, Christiansen SK, Yoder OC and Turgeon BG (1997) Efficient cloning of ascomycete mating type genes by PCR amplification of the conserved MAT HMG box. Fungal Genetics and Biology 21: 118–130.

Berbee ML and Taylor JW (1993) Dating the evolutionary radiations of the true fungi. Canadian Journal of Botany 71: 1114–1127.

Bonfield JK, Smith KF and Staden R (1995) A new DNA sequence assembly program. Nucleic Acids Research 24: 4992–4999.

Bottalico A (1998) Fusarium diseases of cereals: Species complex and related mycotoxin profiles. European Journal of Plant Pathology 80: 85–103.

Bruchez JJP, Eberle J and Russo VEA (1993) Regulatory sequences in the transcription of *Neurospora crassa* genes: CAAT box, TATA box, introns, poly(A) tail formation sequences. Fungal Genetic Newsletter 40: 89–96.

Cozijnsen AJ and Howlett BJ (2003) Characterisation of the mating-type locus of the plant pathogenic ascomycete *Leptosphaeria maculans*. Current Genetics 43: 351–357.

Desjardins AE and Plattner RD (2000) Fumonisin B1-nonproducing strains of *Fusarium verticillioides* cause maize (*Zea mays*) ear infection and ear rot. Journal of Agricultural and Food Chemistry 48: 5773–5780.

Desjardins AE, Plattner RD, Nelsen TC and Leslie JF (1995) Genetic analysis of fumonisin production and virulence of *Gibberella fujikuroi* mating population A (*Fusarium moniliforme*) on maize (*Zea mays*) seedlings. Applied and Environmental Microbiology. 61: 79–86.

Desjardins AE, Proctor RH, Bai GH, McCormick SP, Shaner G, Buechley G and Hohn TM (1996) Reduced virulence of trichothecene-non producing mutants of *Gibberella zeae* in wheat field tests. Molecular Plant Microbe Interactions 9: 775–781.

Elmer WH (1995) A single mating population of *Gibberella fujikuroi* (*Fusarium proliferatum*) predominates in asparagus fields in Connecticut, Massachusetts, and Michigan. Mycologia 87: 68–71.

Gelderblom WCA, Kriek NPJ, Marasas WFO and Thiel PG (1991) Toxicity and carcinogenicity of the *Fusarium moniliforme* metabolite, fumonisin B-1, in rats. Carcinogenesis 12: 1247–1252.

Glass NL and Donaldson GC (1995) Development of primer sets designed for use with the PCR to amplify conserved genes from filamentous ascomycetes. Applied and Environmental Microbiology 61: 1323–1330.

Goodwin SB, Waalwijk C, Kema GHJ, Cavaletto JR and Zhang G (2003) Cloning and analysis of the mating-type idiomorphs from the barley pathogen *Septoria passerinii*. Molecular Genetics and Genomics 269: 1–12.

Howard PC, Eppley RM, Stack ME, Warbritton A, Voss KA, Lorentzen RJ, Kovach RM and Bucci TJ (2001) Fumonisin B-1 carcinogenicity in a two-year feeding study using F344 rats and B6C3F(1) mice. Environmental Health Perspectives 109 (Suppl. 2): 277–282.

Kim UJ, Birren BW, Slepak T, Mancino V, Boysen C, Kang HL, Simon MI and Shizuya H (1996) Construction and characterization of a human bacterial artificial chromosome library. Genomics 34: 213–218.

Larkin RP and Fravel DR (1999) Mechanisms of action and dose-response relationships governing biological control of fusarium wilt of tomato by non-pathogenic *Fusarium* spp. Phytopathology 89: 1152–1161.

Moretti A, Logrieco A, Bottalico A, Ritieni A, Fogliano V and Randazzo G (1997a) Diversity in beauvericin and fusapro-

liferin production by different populations of *Gibberella fujikuroi* (*Fusarium* section *Liseola*). Sydowia 48: 44–56.

Moretti A, Logrieco A, Doko B, Frisullo S, Visconti A and Bottalico A (1997b) *Fusarium proliferatum* from asparagus, in Italy: Occurrence, fertility and toxigenicity. Cereal Research Communications 25: 785–786.

Munkvold G, Stahr HM, Logrieco A, Moretti A and Ritieni A (1998) Occurrence of fusaproliferin and beauvericin in *Fusarium*-contaminated livestock feed in Iowa. Applied and Environmental Microbiology 64: 3923–3926.

Ocamb CM, Juzwik J and Martin FB (2002) *Fusarium* spp. and *Pinus strobus* seedlings: Root disease pathogens and taxa associated with seed. New Forests 24: 67–79.

O'Donnell K, Kistler HC, Tacke BK and Casper HH (2000a) Gene genealogies reveal global phylogeographic structure and reproductive isolation among lineages of *Fusarium graminearum*, the fungus causing wheat scab. Proceedings of the National Academy of Sciences 97: 7905–7910.

O'Donnell K, Nirenberg HI, Aoki T and Cigelnik E (2000b) A multigene phylogeny of the *Gibberella fujikuroi* species complex: Detection of additional phylogenetically distinct species. Mycoscience 41: 61–78.

Proctor RH, Brown DW, Plattner RD and Desjardins AE (2003) Co-expression of 15 contiguous genes delineates a fumonisin biosynthetic gene cluster in *Gibberella moniliformis*. Fungal Genetics and Biology 38: 237–249.

Proctor RH, Desjardins AE, Plattner RD and Hohn TM (1999) A polyketide synthase gene required for biosynthesis of fumonisin mycotoxins in *Gibberella fujikuroi* mating population A. Fungal Genetics and Biology 27: 100–112.

Proctor RH, Hohn TM and McCormick SP (1995) Reduced virulence of *Gibberella zeae* caused by disruption of a trichothecene toxin biosynthetic gene. Molecular Plant Microbe Interactions 8: 593–601.

Rheeder JP, Marasas WFO, Thiel PG, Sydenham EW, Shephard GS and Van Schalkwyk DJ (1992) *Fusarium moniliforme* and fumonisins in corn in relation to human oesophageal cancer in Transkei. Phytopathology 82: 353–357.

Seo JA, Proctor RH and Plattner RD (2001) Characterization of four clustered and co-regulated genes associated with fumonisin biosynthesis in *Fusarium verticillioides*. Fungal Genetics and Biology 34: 155–165.

Tatusova TA and Madden TL (1999) BLAST 2 SEQUENCES, a new tool for comparing protein and nucleotide sequences. FEMS Microbiology Letters 174: 247–250.

Waalwijk C, Mendes O, Verstappen ECP, de Waard MA and Kema GHJ (2002) Isolation and characterization of the mating-type idiomorphs from the wheat septoria leaf blotch fungus *Mycosphaerella graminicola*. Fungal Genetics and Biology 35: 277–286.

Wang H, Wei H, Ma J and Luo X (2000) The fumonisin B1 content in corn from North China, a high-risk area of oesophageal cancer. Journal of Environmental Pathology, Toxicology and Oncology 19: 139–141.

Waterston RH, Lindblad Toh K, Birney E, Rogers J, Abril JF, Agarwal P, Agarwala R, Ainscough R, Alexandersson M, An P, Antonarakis SE, Attwood J, Baertsch R, Bailey J, Barlow K, Beck S, Berry E, Birren B, Bloom T, Bork P, Botcherby M, Bray N, Brent MR, Brown DG, Brown SD, Bult C, Burton J, Butler J, Campbell RD, Carninci P, Cawley S, Chiaromonte F, Chinwalla AT, Church DM, Clamp M, Clee C, Collins FS, Cook LL, Copley RR, Coulson A, Couronne O, Cuff J, Curwen V, Cutts T, Daly M, David R, Davies J, Delehaunty KD, Deri J, Dermitzakis ET, Dewey C, Dickens NJ, Diekhans M, Dodge S, Dubchak I, Dunn DM, Eddy SR, Elnitski L, Emes RD, Eswara P, Eyras E, Felsenfeld A, Fewell GA, Flicek P, Foley K, Frankel WN, Fulton LA, Fulton RS, Furey TS, Gage D, Gibbs RA, Glusman G, Gnerre S, Goldman N, Goodstadt L, Grafham D, Graves TA, Green ED, Gregory S, Guigo R, Guyer M, Hardison RC, Haussler D, Hayashizaki Y, Hillier LW, Hinrichs A, Hlavina W, Holzer T, Hsu F, Hua A, Hubbard T, Hunt A, Jackson I, Jaffe DB, Johnson LS, Jones M, Jones TA, Joy A, Kamal M and Karlsson EK (2002) Initial sequencing and comparative analysis of the mouse genome. Nature 420: 520–562.

Yun SH, Arie T, Kaneko I, Yoder OC and Turgeon BG (2000) Molecular organization of mating type loci in heterothallic, homothallic, and asexual *Gibberella*/*Fusarium* species. Fungal Genetics and Biology 31: 7–20.