Structure and evolution of the *Fusarium* mating type locus: New insights from the *Gibberella fujikuroi* complex

Simon H. Martin[1], Brenda D. Wingfield[1], Michael J. Wingfield[2] and Emma T. Steenkamp[2]

(1) Department of Genetics, Forestry and Agricultural Biotechnology Institute, University of Pretoria, Pretoria 0002, South Africa.
(2) Department of Microbiology and Plant Pathology, Forestry and Agricultural Biotechnology Institute, University of Pretoria, Pretoria 0002, South Africa.

Brenda D. Wingfield
Email: brenda.wingfield@fabi.up.ac.za

Abstract

Mating type genes are central to sexual reproduction and compatibility in Ascomycete fungi. However the "*MAT*" loci experience unique evolutionary pressures that can result in rapid divergence and enhanced inter-specific gene-flow (lateral gene transfer). In this study, molecular evolution of *MAT* loci was considered using the genus *Fusarium* (Teleomorph: *Gibberella*) as a model. Both *MAT1-1* and *MAT1-2* "idiomorphs" from eleven species of the *Gibberella fujikuroi* species complex were sequenced. Molecular evolution of the *MAT* loci from these heterothallic (self-sterile) species was compared with that of the *MAT* loci from nine homothallic (self-fertile) species in the *Fusarium graminearum* species complex. Although *Fusarium* has previously been thought to have the same complement of four *MAT* genes that are found in *Neurospora*, we found evidence of a novel gene, *MAT1-2-3*, that may be specific to the Hypocreales. All MAT genes share a similar set of cis-regulatory motifs, although homothallic species might have recruited novel regulatory elements, which could potentially facilitate alternate expression of *MAT1-1-1* and *MAT1-2-1*. *Fusarium MAT* loci displayed evidence consistent with historical lateral gene-flow. Most notably, the *MAT1-1* idiomorph of *Fusarium sacchari* appears to be unrelated to those of other species in the *G. fujikuroi* complex. In general, *Fusarium MAT* genes are highly divergent. Both positive selection and relaxed selective constraint could account for this phenomenon. However, the extent of both recombination and inter-specific gene-flow in the *MAT* locus also appears to affect the rate of divergence.

Introduction

In heterothallic (self-sterile) Ascomycetes, sexual reproduction can occur only between individuals of opposite mating type. The two mating types are determined by dissimilar "idiomorphic" forms of the mating type (*MAT*) locus. The dissimilarity between the two idiomorphs is maintained because this specific portion of the genome does not pair up and undergo homologous recombination during meiosis (Coppin *et al*., 1997; Kronstad and Staben, 1997). By convention, the idiomorphs are referred to as *MAT1-1* and *MAT1-2* (Turgeon and Yoder, 2000).

The genes encoded by the *MAT* locus are putative transcription factors thought to control sexual development and to regulate the expression of downstream, mating-type-specific genes (Coppin *et al*., 1997; Shiu and Glass, 2000). The gene composition of the locus vary can dramatically among species, but there are two *MAT* genes that are consistently found in filamentous Ascomycetes.

MAT1-1 always contains a gene called *MAT1-1-1*, which encodes a protein homologous to *MATa1* of *Saccharomyces cerevisiae*. This protein has a unique motif called the a-box. *MAT1-2* always contains a gene called *MAT1-2-1* that encodes a protein with a high-mobility-group (HMG) DNA-binding domain.

In homothallic (self-fertile) Ascomycetes sexual reproduction can occur between any two individuals. These species carry homologues of both *MAT1-1-1* and *MAT1-2-1* genes in the same genome (with the possible exception of a few *Neurospora* species (Glass *et al*., 1988; Glass and Smith, 1994). Coppin *et al*. (1997) proposed that self-fertility may be achieved via alternate expression of genes for one or the other mating type in distinct cells. Homothallic species would thus require a mechanism for independent regulation of *MAT1-1-1* and *MAT1-2-1* (Coppin *et al*., 1997). In certain Ascomycete genera, it is clear that homothallic species have arisen from heterothallic ancestors (e.g. Yun *et al*., 1999; O'Donnell *et al*., 2004). However, in many other genera, it is likely that multiple independent transitions from heterothallism to homothallism and vice–versa have occurred (reviewed by Lee *et al*., 2010).

Mating type genes have been found to be highly divergent between species (Cisar *et al*., 1994; Arie *et al*., 1997; Turgeon, 1998; Wik *et al*., 2008), yet they may be strongly conserved within species (Turgeon, 1998). Rapid diversification of sex-related proteins has been observed in many plants and animals, and the forces of sexual selection, sexual conflict and reinforcement have all been proposed as possible causes (reviewed by Civetta and Singh, 1998; Swanson and Vacquier, 2002; Clark *et al*., 2006). The ability to reproduce asexually adds a layer of complexity to this situation. Selection against the accumulation of detrimental mutations in *MAT* loci could at times be lacking. In *Neurospora*, Wik *et al*. (2008) found the *MAT* genes of heterothallic species to be under positive (diversifying) selection. However, the *MAT* genes of homothallic species appeared to have diverged under a lack of selective constraint, and some contained premature stop codons. It was proposed that *MAT* genes may, therefore, be dispensable in homothallic *Neurospora* spp.

An important possibility that may have been overlooked is that the distinct genetic organization of the *MAT* loci could also shape their evolution. In heterothallic species, the effective population size of each idiomorph is half that of the rest of the genome (assuming equal mating type frequencies). This could cause reduced within species diversity, but enhanced inter-specific diversity as a result of genetic drift (Charlesworth, 2009). Unlike heterothallic *MAT* idiomorphs, homothallic *MAT* loci can presumably recombine freely during meiosis. The extent of recombination experienced by a locus could affect the efficiency of selection (Birky and Walsh, 1988) as well as the actual nucleotide substitution rate (Vicoso and Charlesworth, 2009). It is, therefore, important to distinguish between the effects of adaptive and non-adaptive forces on the patterns of diversity in *MAT* loci.

Recent research has shown that phylogenies constructed using *Neurospora MAT* genes are in conflict with species trees (Strandberg *et al*., 2010). It was proposed that reproductive genes may be more predisposed to inter-specific gene-flow than others. Transfer of sexual genes through hybridization could restore sexuality to populations in which it has been lost, as was observed in *Ophiostoma* (Paoletti *et al*., 2006). It has not been established whether homothallic species display similar characteristics (Strandberg *et al*., 2010). It could in fact be more difficult to detect lateral transfer in homothallic than heterothallic *MAT* loci. In the latter group, the lack of recombination ensures that the phylogenetic signal of hybridization or lateral transfer is not diluted by exchange of genetic material between "native" and introgressed alleles (Chaturvedi *et al*., 2002; Devier *et al*., 2010).

A useful model in which to study *MAT* locus evolution is found in the genus *Fusarium* (Teleomorph: *Gibberella*). Heterothallic *Fusarium* spp. such as those in the well-known Gibberella fujikuroi species complex, have three genes in the *MAT1-1* idiomorph (*MAT1-1-1*, *MAT1-1-2* and

MAT1-1-3) and a single gene in the *MAT1-2* idiomorph (*MAT1-2-1*) (Arie *et al*., 1999; Yun *et al*., 2000). These four *MAT* genes are homologous to the four *MAT* genes of the model species *Neurospora crassa* (Yun *et al*., 2000), where they are found in the same arrangement (Glass *et al*., 1990; Staben and Yanofsky, 1990). In the homothallic *Fusarium graminearum* complex, in which homothallism has a monophyletic origin (O'Donnell *et al*., 2004) all four *MAT* genes are adjacent within the *MAT* locus (Yun *et al*., 2000). Homothallism is thought to have arisen through an unequal crossover between *MAT* idiomorphs in a heterothallic ancestor (O'Donnell *et al*., 2004). Unlike *Neurospora*, there is no evidence that *MAT* genes are dispensable in the homothallic *Fusarium* spp. Knock-out strains of *F. graminearum* lacking either the *MAT1-1* or *MAT1-2* genes can only outcross in a heterothallic manner (Lee *et al*., 2003). A previous study of *MAT* locus evolution in the *F. graminearum* complex (O'Donnell *et al*., 2004) yielded results in conflict with the *Neurospora* model: reduced divergence in *MAT* genes and few phylogenetic inconsistencies. However, there has been no thorough, equivalent investigation of *MAT* locus evolution in a monophyletic, heterothallic, *Fusarium* lineage. Interestingly, an earlier phylogeny constructed using short *MAT* sequences from species in the heterothallic *G. fujikuroi* complex did yield an unexpected topology for *MAT1-1* (Steenkamp *et al*., 2000), but this was not investigated further.

In this study *MAT* locus structure and evolution in the *G. fujikuroi* complex has been studied in depth, allowing a thorough comparison between heterothallic and homothallic *Fusarium* species. The aims were to (1) identify and compare evolutionarily conserved regions and cis-regulatory motifs in the *MAT* loci of heterothallic and homothallic *Fusarium* spp.; (2) investigate whether phylogenetic conflicts exist between the *MAT* and species trees; and (3) compare patterns of molecular evolution in both coding and non-coding portions of heterothallic and homothallic loci. To achieve these aims, the complete *MAT1-1* and *MAT1-2* idiomorphs of eleven species in the *G. fujikuroi* complex, including nine mating populations (A–I), were sequenced.

Materials and methods

Fungal isolates

Twenty-two *Fusarium* isolates, a *MAT1-1* and a *MAT1-2* isolate for each of eleven species (Table 1), were used in this study. For the nine mating populations, the isolates used represent the standard *MAT1-1* and *MAT1-2* mating type tester strains. For two additional species, *F. subglutinans* group 1 and *F. mangiferae*, isolates characterized by Steenkamp *et al*. (2002) and Britz *et al*. (2002), respectively, were used. DNA was extracted from all isolates using the procedure described by Steenkamp *et al*. (1999).

DNA sequencing

Using published *MAT1-1* and *MAT1-2* idiomorphs of *F. oxysporum* and *F. verticillioides* (Yun *et al*., 2000) and their genome sequences (GenBank accession numbers: AAXH01000548 and AAIM02000073, respectively), PCR primers were designed in conserved regions to amplify overlapping fragments covering the length of both idiomorphs (Fig. 1 and Table A1). Reaction mixtures were 25 µl in volume and contained 4 ng/µl template DNA, 1.5 mM $MgCl_2$, 0.4 µM of each primer, 1 mM deoxynucleotide triphosphates (0.25 mM of each), and 0.05 u (units)/µl Super-Therm DNA Polymerase and reaction buffer (Southern Cross biotechnology [Pty.] Ltd., Cape Town, South Africa). The PCR cycling conditions consisted of an initial denaturation at 94 ºC for 60 s; thirty cycles of denaturation at 94 ºC for 30 s, annealing for 30 s and extension at 70 ºC; followed by a final extension step at 70 ºC for 10 min. All PCR amplifications from genomic DNA were performed using the same protocol, except for annealing temperatures that depended on the primers used (Table A1) and extension times that depended on the expected length of the amplicons. Extension times were calculated by adding 1 min for each kb of sequence amplified.

Table 1. *Fusarium* isolates in the *Gibberella fujikuroi* species complex used in this study.

Fusarium spp.	Isolate[a]			
	Mating population	Mating type	MRC no.[b]	KSU no.[c]
F. verticillioides	A	MAT-1	8559	A-00149
		MAT-2	8560	A-00999
F. sacchari	B	MAT-1	8552	B-03853
		MAT-2	8551	B-03852
F. fujikuroi	C	MAT-1	8532	C-01993
		MAT-2	8534	C-01995
F. proliferatum	D	MAT-1	8549	D-04854
		MAT-2	8550	D-04853
F. subglutinans (group 2)	E	MAT-1	8553/6483	E-00990
		MAT-2	8554/6512	E-02192
F. subglutinans (group 1)		MAT-1	1084	
		MAT-2	7828	
F. thapsinum	F	MAT-1	8558	F-04094
		MAT-2	8557	F-04093
F. nygamai	G	MAT-1	8546	G-05111
		MAT-2	8547	G-05112
F. circinatum	H	MAT-1	7488	H-10847
		MAT-2	6213	H-10850
F. konzum	I	MAT-1	8545	I-11616
		MAT-2	8544	I-11615
F. mangiferae	N/A	MAT-1	8092/8093	X4382
		MAT-2	7559	11,781

[a] Mating populations are represented by the standard *MAT-1* and *MAT-2* "tester" strains (Kuhlman, 1982; Leslie, 1991, 1995; Klittich and Leslie, 1992; Klaasen and Nelson, 1996; Klittich et al., 1997; Britz et al., 1999; Britz et al., 2002; Zeller et al., 2003; Leslie et al., 2005).

[b] Medical Research Council (MRC), Tygerberg, South Africa.

[c] Department of Plant Pathology, Kansas State University (KSU), Manhattan, Kansas, USA.

The complete idiomorphs were amplified using primers MFC-R1 and MFC-L1, located in the 5' and 3' flanking regions respectively. In cases, where this PCR failed, MFC-L1 was substituted with a primer just inside the 30 end of the idiomorph (M1C-2 for *MAT1-1* and M2C-4 For *MAT1-2*). Products were analyzed by electrophoresis in a 1% agarose gel at 20 V/cm to verify purity and fragment size. *MAT1-1* and *MAT1-2* products were pooled to be sequenced together. Pools consisted of products from distinct species to avoid confusion in assembly arising from homology in the flanks. 454 sequencing was performed by Inqaba Biotech (Pretoria, South Africa). The pools were prepared separately using adapters carrying distinct identifiers, and sequenced together in one lane on a Roche GS-FLX Platinum sequencer.

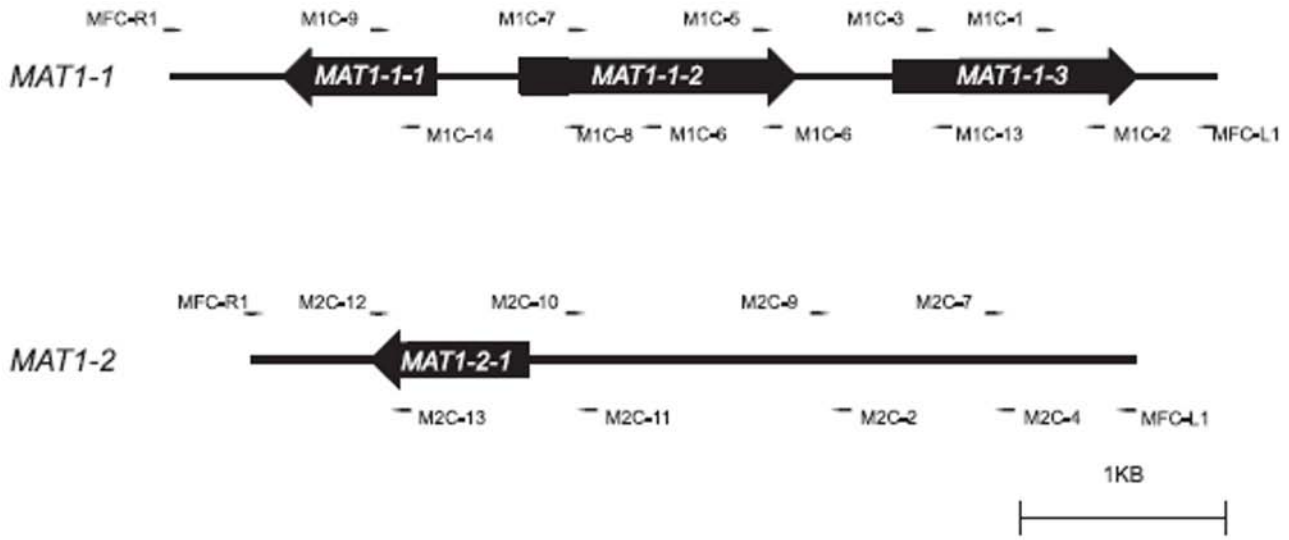


Figure 1. Binding sites of primers used to amplify *MAT1-1* and *MAT1-2* from *G. fujikuroi* species. Small arrows represent primer binding sites. Large block arrows represent *MAT* genes.

To fill in gaps as well as verify areas of low pyro-sequence coverage, PCR products were amplified using the appropriate primer pairs (Fig. 1). For *F. proliferatum* in which all PCRs incorporating the 30 MFC-L1 primer were unsuccessful, an additional primer, MFC-R2, was designed based on the 30 flank region of close relatives *F. fujikuroi* and *F. mangiferae* as a substitute for MFC-L1. Products were cloned in *Escherichia coli* using the pGEM®-T Easy Vector cloning system (Promega Corporation, Madison, WI). Cloned inserts were then amplified directly from colonies using the primers SP-6 (5'-ATTTAGGTGACACTATAG-3') and T-7 (5'-TAATACGACTCACTATAGGG-3'). This PCR reaction mixture was 25 µl and contained 0.05 l/µl FastStart *Taq* DNA Polymerase (Roche Diagnostics, Mannheim, Germany), 1 × Faststart buffer with $MgCl_2$, 1 mM deoxynucleotide triphosphates (0.25 mM of each), 0.4 µM of each primer and one bacterial colony. The colony PCR reaction conditions were as follows: Denaturation at 94 ºC for 5 min; followed by thirty cycles of denaturation at 94 ºC for 30 s, annealing at 50 ºC for 30 s and extension for 90 s at 70 ºC; followed by a final extension step at 70 ºC for 10 min. Following purification with sodium acetate (pH 3.8; 0.1 M) (Sambrook and Russell, 2001) the PCR products were sequenced in both directions using the SP-6 and T-7 primers, the BigDye® terminator v3.1 cycle sequencing kit (Applied Biosystems, Foster City, CA) and an ABI PRISM 377 automated DNA sequencer (Perkin Elmer, Warrington, United Kingdom).

Structural organization and conserved regions in the MAT loci

Sequence assembly, annotation and in silico translation were performed using CLC Bio Genomics Workbench (CLC Bio, Aarhus, Denmark). Sequences were aligned using ClustalW (Thompson *et al*., 1994), except for coding sequences, which were aligned using the codon-based algorithm of SQUINT (Goode and Rodrigo, 2007). The start of the flanking regions (i.e. the point at which homology is restored between *MAT-1* and *MAT-2* genomes) were identified manually by aligning the *MAT1-1* and *MAT1-2* sequences from the same species. Structural relationships were visualized with the graphical alignment program GATA (Nix and Eisen, 2005), based on the Bl2Seq algorithm (Tatusova and Madden, 1999), which enables identification of putative duplication and inversion events.

Evolutionarily conserved regions were identified using dnaSP version 5 (Librado and Rozas, 2009) by determining nucleotide diversity (π) over a sliding window of 100 bp and sliding in increments of 25 bp. Gene identification was performed using the online tools FGENESH (Salamov and Solovyev, 2000) and AUGUSTUS (Stanke *et al*., 2004). The 500 bp region upstream of each start codon was investigated to identify putative cis-regulatory transcription factor binding sites using

MOTIFSEARCH (http://www.motif.genome.jp), based on the TRANSFAC database (Wingender *et al*., 1996), using the default cut-off value of 85% similarity. Only 500 bp was used because this is roughly the size of most intergenic regions in the MAT locus.

Phylogenetic relationships

The two *MAT* idiomorphs were treated separately for phylogenetic analysis. All flanking DNA was excluded to avoid any dilution of phylogenetic signal due to recombination. Most appropriate nucleotide substitution models were chosen using jModelTest (Posada, 2008). Maximum likelihood phylogenetic trees were generated using PhyML 3.0 (Guindon *et al*., 2009) and internal branches were evaluated with 1000 bootstrap replicates. The *F. oxysporum MAT* idiomorph sequences (AB011379 and AB011378) were used as outgroups.

Competing tree topologies of *MAT* and species trees were compared with respect to their fit to the sequence data using the Approximately Unbiased (AU) (Shimodaira, 2002) and Shimodaira and Hasegawa (1999) tests, as implemented in CONSEL (Shimodaira and Hasegawa, 2001). The two *MAT* trees for the *G. fujikuroi* complex as well as the *MAT* tree for the *F. graminearum* complex generated by O'Donnell *et al*. (2004) were compared to their respective species trees, which were inferred by identifying consistently well-supported nodes in previous studies (O'Donnell *et al*., 1998, 2000, 2008; Kvas *et al*., 2009; Yli-Mattila *et al*., 2009). Only nodes supported by bootstrap-values higher than 70% were retained for these analyses. Site likelihood values required by CONSEL were generated for each competing tree using Tree Puzzle (Schmidt *et al*., 2002).

Sequence evolution in coding and non-coding regions

To compare the nucleotide substitution rate in heterothallic and homothallic *MAT* loci, nucleotide diversity (π) and its variance were measured using dnaSP. *Fusarium sacchari* was excluded from this analysis to avoid any skewing of the result by the highly divergent *MAT1-1* sequence. Coding and non-coding portions of the locus were considered as separate datasets. To eliminate variation due to different evolutionary time-scales, each measurement of π was divided by that generated for a sequence consisting of partial segments of house-keeping genes Translation Elongation Factor 1-α and Tubulin β-1 chain from the same set of species (accession numbers are listed in Table A2). This scaling procedure was based on the assumption that these house-keeping genes should experience comparable evolutionary rates in both heterothallic and homothallic species.

Using the available *Fusarium* genomes, variability in the *MAT* proteins was compared with an estimated genome-wide average. For this purpose, a set of amino acid sequences was generated using 100 randomly-selected protein-coding genes from the genomes of *F. verticillioides* (AAIM02000073), *F. graminearum* (NZ_AACM00000000), *F. oxysporum* (AAXH01000548) and *F. circinatum*. To avoid variation due to differences in gene annotation, alignments for genes with large gaps or non-homologous regions were not included in the analysis. The percentage identity values for the 100 proteins among the three species were compared with those for the five *MAT* proteins.

In order to consider the selective pressures affecting each *MAT* gene, the likelihoods of various models of codon evolution were calculated using the CODEML program of the PAML package, version 4.3 (Yang, 2007). Each model imposed different constraints on ω, which is the ratio of non-synonymous substitutions per non-synonymous site to synonymous substitutions per synonymous site. We first investigated whether genes have experienced selective constraint significantly stronger than what could be expected under neutral evolution. A value of $\omega = 1$ indicates an equal rate of synonymous and non-synonymous substitutions, implying neutrality. A neutral model in which $\omega = 1$ was fixed, was compared to model M0 which allowed an estimated ω value. We then investigated

whether positive selection drives *MAT* gene evolution. Positive selection is expected to cause more non-synonymous than synonymous substitutions, resulting in $\omega > 1$. However, since most proteins probably have some codons consistently under selective constraint, models that allowed codons to fall into various classes with different x values were compared. Model M7 (the beta model) allowed a β distribution of codons in the range $0 \leq \omega \leq 1$, while model M8 (the β and ω model) allowed an additional class of $\omega \geq 1$. The likelihoods of these two pairs of models were compared using the likelihood ratio test (LRT), as described by Yang *et al*. (2000).

Results

Structural organization and conserved regions in the *MAT* loci

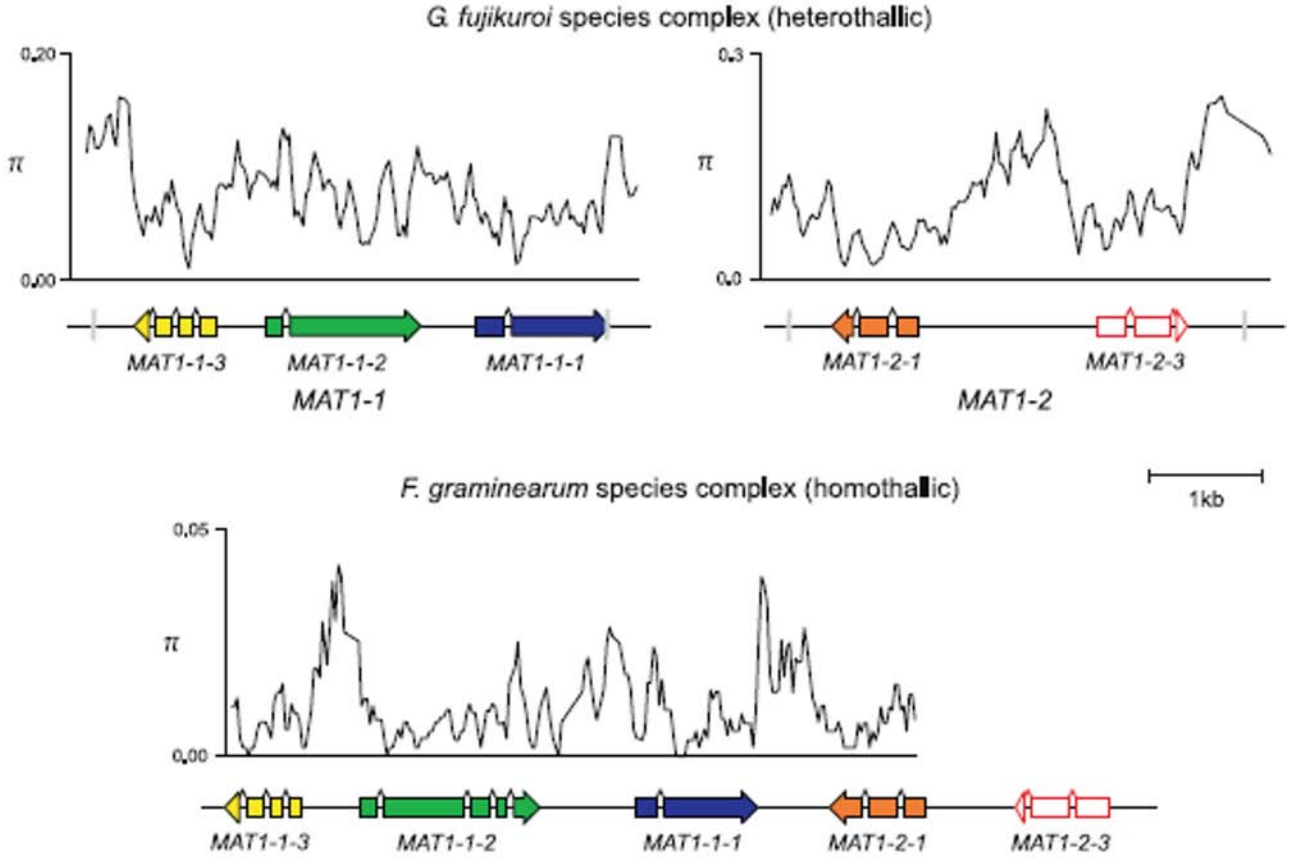
Complete idiomorph sequences, including portions of both 50 and 30 flanks, were obtained for all 22 *Fusarium* isolates, except for the region at the 30 end of the *F. sacchari MAT1-2* idiomorph. Because the rest of the derived idiomorph sequence for the same isolate was identical to that obtained by Arie *et al*. (1999), the missing sequence was also assumed to be identical. The length of the idiomorph (non-homologous region) ranged from 4.3 kb to 4.6 kb for *MAT1-1* and 3.5 kb to 3.8 kb for *MAT1-2*. GenBank accession numbers for *MAT* sequences are pending.

In general, non-coding portions of the *MAT* loci were found to be more variable between species than coding portions (Fig. 2). However, an area of low diversity was observed in the heterothallic *MAT1-2* idiomorph, where no gene had previously been described (Fig. 2). Gene prediction using both FGENESH and AUGUSTUS revealed a predicted gene in this region with three exons. Using the predicted protein product for tBLASTn analysis (protein query against translated nucleotide database) against the complete genomes of *F. verticillioides* and *F. oxysporum*, both of which represent *MAT-1* isolates, yielded no hits. In the homothallic *F. graminearum* genome however, there was a single significant hit to a predicted gene (FGSG 08894) found in an homologous location, immediately adjacent to *MAT1-2-1* at the *MAT* locus, but in an inverted orientation. GATA analysis provided evidence of an inversion of this region in *F. graminearum* (Fig. 3). Gene expression data from *F. graminearum* (Güldener *et al*., 2006a), available via the *F. graminearum* database (FGDB) (Guldener *et al*., 2006b) verified that FGSG 08894 is transcribed. Furthermore, these results show a noticeable increase in transcript levels during the stages "*in vitro sexual development*" and "*in vitro sexual development of Fusarium Cch1 calcium channel deletion mutant*". This pattern was also observed in all of the four known *MAT* genes (Güldener et al., 2006a). The predicted MAT gene was designated *MAT1-2-3*. The name *MAT1-2-2* has already been used to describe a different gene (Kanamori *et al*., 2007), which is not homologous to *MAT1-2-3*.

To gain insight into the range of species in which *MAT1-2-3* may be present, a tBLASTn analysis against the GenBank database was performed. Apart from the *Fusarium* species mentioned herein, there were weak hits in two additional species in order Hypocreales; *Cordyceps militaris* (accession, AB084257) ($E = 0.026$) and *Paecilomyces tenuipes* (AB084921) ($E = 0.31$). Both hits corresponded to a position immediately adjacent to the described *MAT1-2-1* gene, suggesting that these may be true orthologs.

All five *MAT* genes (*MAT1-1-1*, *MAT1-1-2*, *MAT1-1-3*, *MAT1-2-1*, and *MAT1-2-3*) had intact open reading frames (ORFs) of near-identical length among the eleven heterothallic *Fusarium* species, with two exceptions. In *F. sacchari*, *MAT1-1-1* had a stop codon 11 codons upstream of that in the other species. This stop codon was present in both the sequence produced in this study as well as that published by Arie *et al*. (1999). This species probably encodes a slightly truncated *MAT1-1-1* protein. In *F. circinatum*, the first exon of the newly-identified *MAT1-2-3* gene was predicted to begin 126nt downstream of the start codon predicted in the other species, thus reducing the predicted protein by 42 amino acids. The upstream start codon is present in *F. circinatum* but a stop

codon follows downstream. The presence of this stop codon was confirmed by PCR and Sanger sequencing. It is, therefore, unclear whether the *F. circinatum* prediction represents the true start codonfor all of these species or whether it has a different start codon to the others.



	coding regions		non-coding regions	
	π	variance	π	variance
G. fujikuroi complex	0.051	1.91×10^{-5}	0.108	8.6×10^{-6}
F. graminearum complex	0.007	5×10^{-7}	0.022	5.9×10^{-6}

Figure 2. Structures and nucleotide diversity of *Fusarium MAT* loci. Boxes represent exons with the direction of transcription indicated. Gray lines separate idiomorphs from flanking regions, although these points varied among species. Plots indicate nucleotide diversity (π), measured in a sliding window of 100 bp. For species in the *G. fujikuroi* complex, p was measured using sequences generated in this study. For the *F. graminearum* species complex, sequences generated by O'Donnell *et al*. (2004) were used, hence the plot only covers the region previously thought to constitute *MAT* locus.

In the upstream 500 bp of the *MAT* genes, multiple motifs that appear to be homologous to the binding sites of eleven previously described transcription factors (most described in *S. cerevisiae*) were identified (Table A3). Multiple predicted binding sites of a heat shock factor (HSF), an alcohol dehydrogenase gene regulator (ADR1) and an activator of nitrogen-regulated genes (NIT2) were identified upstream of all genes, although the number of motifs sometimes differed among species. Several predicted bindings sites were identified upstream of a subset of the genes. For example, a stress response element (STRE) binding site was predicted upstream of all genes except *MAT1-2-3*. The most notable finding was the identification motifs similar to the binding sites of two yeast *MAT* genes: *MAT-Mc* (*Schizosaccharomyces pombe*) and *MATa2* (*S. cerevisiae*). In heterothallic species, all *MAT* genes carried motifs similar to the *MAT1-Mc* binding site. In homothallic species only

MAT1-1-1 carried this motif, the remaining four genes instead carried a motif similar to the *MATa2* binding site (Table A3).

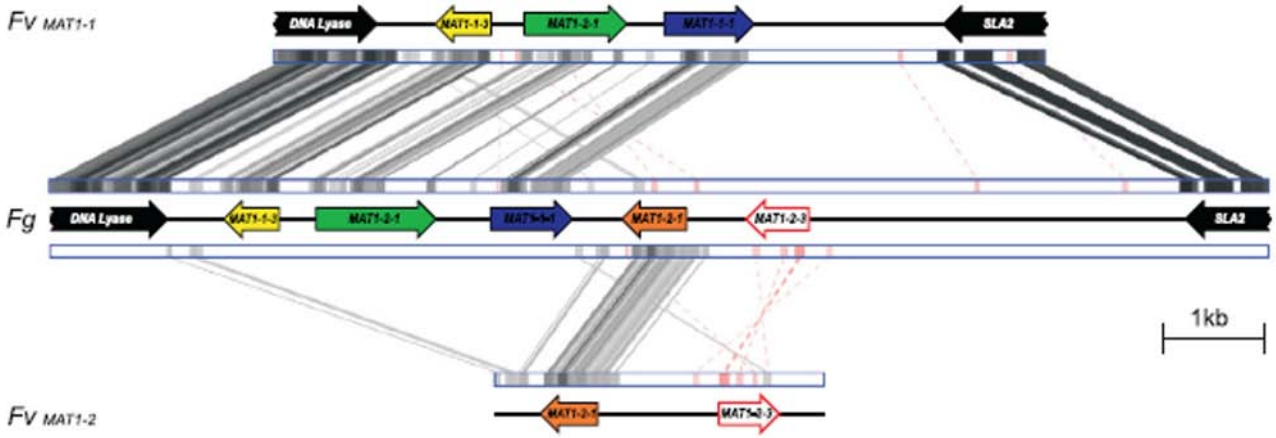


Figure 3. Homology between the *F. graminearum* and *F. verticillioides MAT* loci. The diagram was prepared from the output of GATA. Regions of strong homology are shaded and connected by lines. The intensity of shading indicates the strength of homology. Red shading with dashed lines indicates inverted homology. Genes are represented by box arrows. The *F. graminearum* (*Fg*) and *F. verticillioides* (*Fv*) *MAT1-1* sequences were obtained from the genome sequences, while the *F. verticillioides MAT1-2* sequence was obtained in this study.

Phylogenetic relationships

Phylogenetic analysis of both *MAT1-1* and *MAT1-2* sequences revealed three major clades for the *G. fujikuroi* isolates (Fig. 4). These correspond to the African, American and Asian clades described by O'Donnell *et al*. (1998). However the *MAT1-1* idiomorph of *F. sacchari* did not fall into any of the three major clades, while the *MAT1-2* idiomorph of this species grouped within the Asian clade as expected. The ML trees for *MAT1-1* and *MAT1-2* displayed several topological conflicts with the known species tree (Fig. 4). Comparison of competing topologies using the AU and SH tests revealed that the *MAT* loci did not support the known species trees over the topology of the ML tree (Table 2). Furthermore, neither *MAT1-1* nor *MAT1-2* supported the phylogeny representing the opposing idiomorph (Table 2). In contrast, for the homothallic *F. graminearum* complex, the *MAT* tree generated by O'Donnell *et al*. (2004) did not have a significantly better fit to the data than the species tree (Table 2).

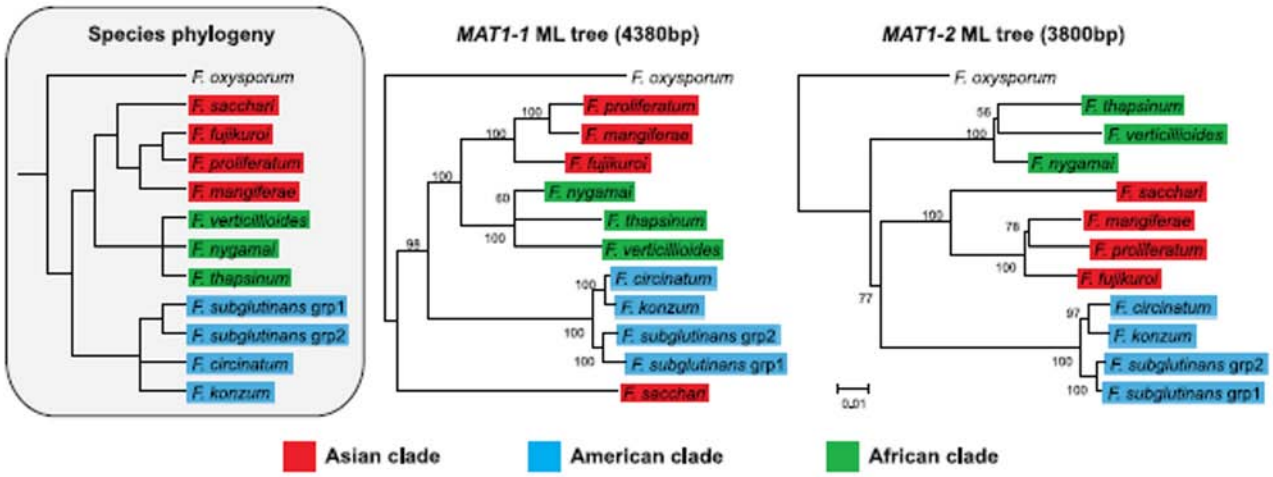


Figure 4. Maximum likelihood phylograms for *G. fujikuroi* complex *MAT* idiomorphs. The major clades of O'Donnell *et al*. (1998) are represented by colors. Numbers on branches indicate bootstrap values. The *MAT1-1* and *MAT1-2* phylogenies were inferred using, respectively, the general time

reversible (GTR) model (Tavaré, 1986) and three-parameter (Kimura, 1981) model with unequal base frequencies (TPM1uf; Posada 2008), both with gamma correction for among site rate variation.

Table 2. Results of the Approximately Unbiassed (AU) test of competing tree topologies.

Trees Tested[a]	*P* [b]		Rejected[c]	
	MAT-1	*MAT-2*	*MAT-1*	*MAT-2*
G. fujikuroi complex species vs. *MAT* trees				
Species	0.000	0.034	Yes	Yes
MAT-1 ML tree	1.000	0.000		Yes
MAT-2 ML tree	0.000	0.962	Yes	
Placement of *F. sacchari*				
Within Asian clade	0.009	1.000	Yes	
Outside Asian clade	0.016	0.001	Yes	Yes
Outside *G. fujikuroi* complex	0.993	0.000		Yes
F. graminearum complex species vs. *MAT* trees				
Species tree	0.157		Yes	
MAT tree	0.843			

[a] Alternative topologies being compared. Conflicts between trees are illustrated in Fig. 6.

[b] Approximately Unbiased (AU) test statistic. Results of the Shimodaira–Hasegawa (SH) test are not shown as they echoed the AU results, but sometimes with lower significance.

[c] Likelihoods are significantly different at values of $P \leqslant 0.05$.

The unexpected grouping of the *F. sacchari* in the *G. fujikuroi MAT1-1* tree was also addressed independently by comparing three trees that differed only in the position of *F. sacchari*. Placement of *F. sacchari* within the Asian clade could be rejected for the *MAT1-1* idiomorph but not *MAT1-2*. Placement of *F. sacchari* outside of the Asian clade but within the *G. fujikuroi* complex could be rejected for both idiomorphs. However, placement of *F. sacchari* outside the complex could be rejected for *MAT1-2* but not for *MAT1-1*. This implies that the most likely phylogenetic placement of the *F. sacchari MAT1-1* idiomorph is outside of the *G. fujikuroi* complex.

Sequence evolution in coding and non-coding regions

Among the four species for which genome data was available (*F. oxysporum*, *F. verticillioides*, *F. circinatum*, and *F. graminearum*) all five *MAT* genes displayed levels of percentage identity that were well below the average for 100 randomly-selected nuclear genes (Fig. 5). The *MAT* genes shared most similarity between the two *G. fujikuroi* complex species *F. circinatum* and *F. verticillioides*, and were most divergent when the three heterothallic species were compared with *F. graminearum*.

The results of the CODEML analyses indicated that all five genes in heterothallic species deviated significantly from the neutral model of evolution. By contrast, in homothallic species, *MAT1-1-1* and *MAT1-1-3*, did not demonstrate a significant deviation from the neutral model. However, *MAT1-1-1* displayed evidence for the presence of certain codons under positive selection in both heterothallic and homothallic species (Table A4).

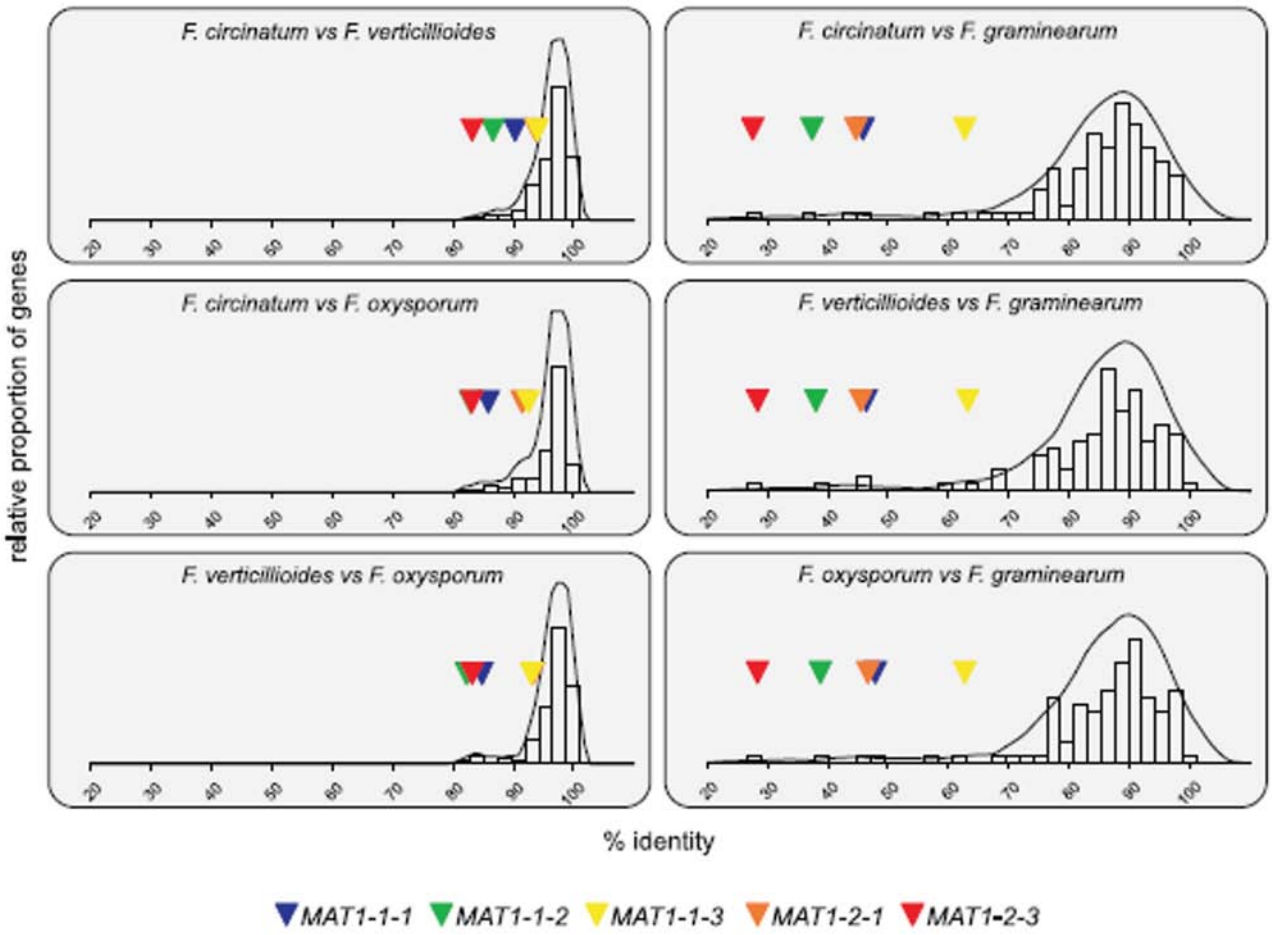


Figure 5. Scaled nucleotide diversity (π) in coding and non-coding portions of *MAT* loci. Values of π for from *MAT* loci were scaled based on the π values of EF1-α and BT. Bars indicate the length of one standard deviation from the mean.

Scaled values of nucleotide diversity (π) demonstrated that the coding regions of the *MAT* loci were comparatively less divergent among homothallic species (Fig. 6). However, the ratio of π at non-synonymous sites to π at synonymous sites was similar in the heterothallic and homothallic species (0.253 and 0.243, respectively). This indicated that the lower diversity in homothallic *MAT* genes was reflected at both non-synonymous and synonymous codon positions. The non-coding portion of the *MAT* locus was generally more divergent, but the relationship between the species complexes was reversed: non-coding portions were less divergent among heterothallic species (Fig. 6).

Discussion

To date, *Fusarium* spp. have been thought to carry the same complement of *MAT* genes as the model species *Neurospora crassa* (Arie *et al*., 1999; Yun *et al*., 2000; O'Donnell *et al*., 2004). This study has identified evidence of a previously unknown *MAT* gene in what was previously thought to be a non-coding portion of the *MAT1-2* idiomorph. The putative gene, tentatively named *MAT1-2-3*, appears to be homologous to FGSG 08894, which located adjacent to *MAT1-2-1* in the *F. graminearum* genome. The predicted protein sequence of FGSG 08894 contains no known functional domains but the gene does display an expression profile in *F. graminearum* (Güldener *et al*., 2006a) that is similar to that of the other four *MAT* genes, with significant up-regulation during the process of sexual development. Whether *MAT1-2-3* is expressed in heterothallic *Fusarium* spp. remains to be verified. If so, the term "*MAT* gene" would be appropriate, because it is found only within the *MAT1-2* idiomorph and is therefore absent from the genomes of *MAT-1* individuals. The corresponding locations of *MAT1-2-3* in heterothallic and homothallic species would be consistent

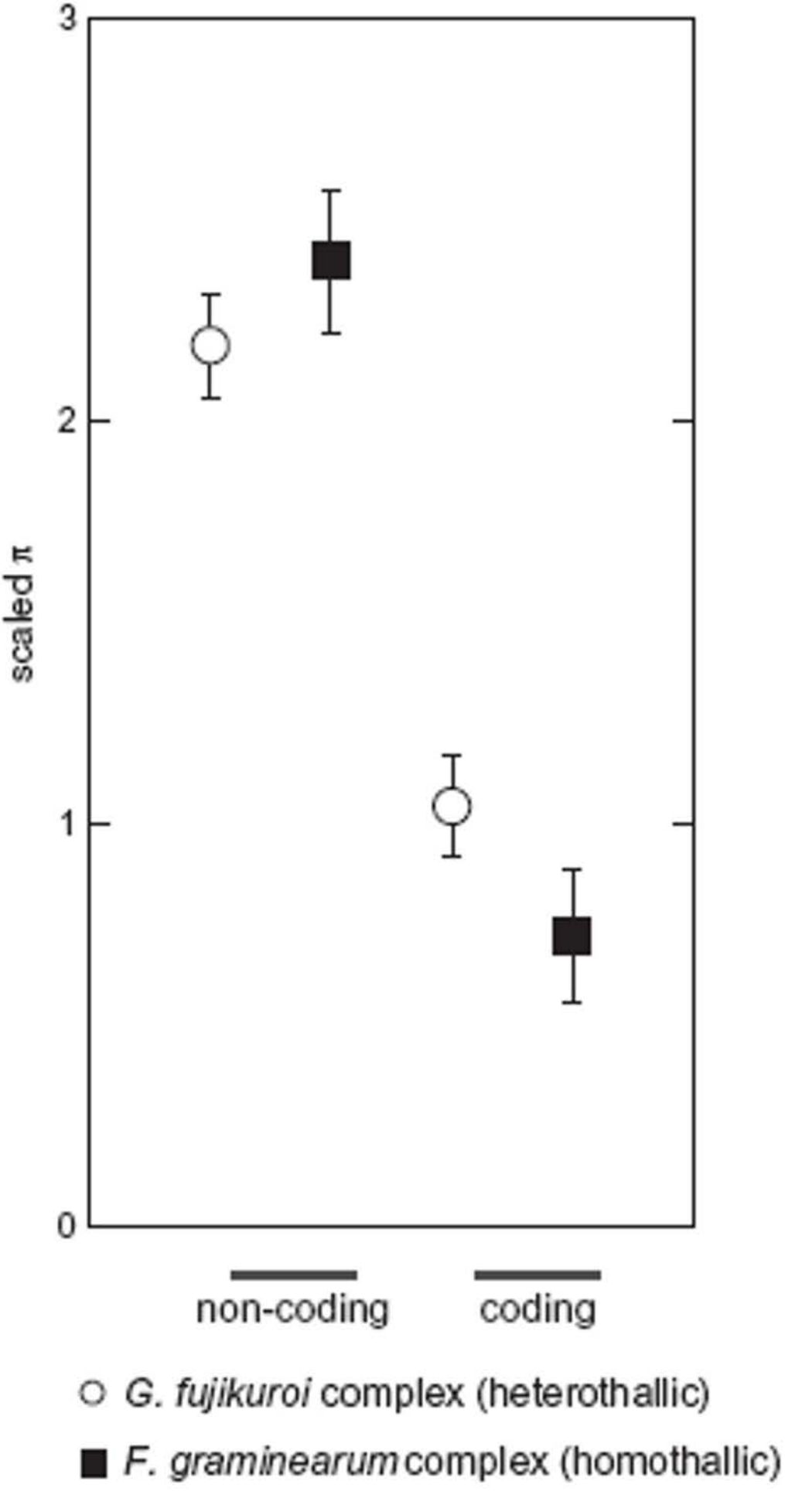

3
2
1
0
scaled π
non-coding
coding
G. fujikuroi complex (heterothallic)
F. graminearum complex (homothallic)

Figure 6. Amino acid percentage-identity of *MAT* genes as compared to 100 nuclear genes from four genomes. Histograms indicate the distribution of percent identity for all 105 genes. Colored markers indicate the positions of the four *MAT* genes. The solid line indicates an estimated non-parametric kernel curve (Gasser *et al*. 1985) of density.

with the view of O'Donnell *et al*. (2004) that homothallism in *F. graminearum* arose through the fusion of heterothallic *MAT* loci. However, the gene's orientation is reversed, suggesting a segmental inversion of this region subsequent to the fusion. Putative homologs of *MAT1-2-3* were also identified in two additional Hypocreales genomes. However, attempts to identify homologs using BLAST may be hindered by the rapid diversification of *MAT* genes as well as the fact that many Ascomycete genomes represent *MAT-1* individuals.

A search for known binding sites of transcription factors found in model Ascomycetes revealed that a similar set of predicted binding sites appears consistently upstream of all five *MAT* genes, including *MAT1-2-3*, which is further support for its authenticity. Conclusions with regard to the functions of the TFs involved should be made with caution; these TFs and their binding sites were described in species quite distantly related from *Fusarium*. Nevertheless, most TFs implicated here are involved (in their host organism) in processes consistent with a role in sexual reproduction. Most prevalent were predicted binding sites of TFs involved in response to environmental stresses such as the absence of certain nutrients or heat shock. Nutrient starvation is a well known stimulus of sexual development in Ascomycetes (Coppin *et al*., 1997; Pöggeler *et al*., 2006). Sexual development in *Fusarium* species is also known to be highly temperature-dependent (reviewed by Doohan *et al*., 2003). Whereas heat-shock per se has never been implicated in *Fusarium*; in *Podospora anserina*, a heat shock protein *mod-E* is known to be necessary for sexual reproduction (Loubradou et al., 1997).

Probably the most surprising finding in the analysis of the regulatory regions was the identification of sites that appear to be homologous to the binding sites of certain yeast *MAT* genes. In heterothallic species, all five *MAT* genes carried motifs similar to the binding site of *MAT1-Mc*. This is an HMG-domain protein found in the *S. pombe MAT* locus (Kjaerulff et al., 1997) that may be orthologous to *MAT1-2-1* of filamentous species (Staben and Yanofsky, 1990). If this is the case, it is possible that these sites are in fact binding sites of the *MAT1-2-1* HMG-domain protein, which would imply a form of self-regulation. In homothallic species, only one gene, *MAT1-1-1*, carried similar predicted *MAT1-Mc* binding sites. The other four *MAT* genes instead carried motifs similar to the binding sites of *MATa2*. This is a homeodomain protein encoded by the *S. cerevisiae MAT-a* idiomorph (Johnson, 1995), of which there is no known ortholog in filamentous species. Whether these conserved motifs indeed correspond to actual TF binding sites will be an important focus for future research. Nevertheless, the fact that the two species complexes have many similarities but a few differences in the conserved motifs upstream of the various *MAT* genes, implies that these regulatory regions have evolved, potentially reflecting adaptive changes in gene regulation.

Both *MAT* idiomorphs of species in the *G. fujikuroi* complex displayed significant support for phylogenies in conflict with the recognized species tree. It is possible that such incongruities could be the result of incomplete lineage sorting. However, this process requires the existence of polymorphism prior to speciation. Intraspecific polymorphisms in heterothallic *MAT* loci are thought to be rare and short-lived (Turgeon, 1998), possibly as a result of the lower effective population size. Another possible cause of the phylogenetic irregularities is inter-specific gene-flow. Similar findings in *Neurospora* led Strandberg *et al*. (2010) to propose that reproductive genes might be more prone to lateral gene-flow than others. This pattern could be driven by selection if the alleles that introgress from other species can restore sexuality to populations in which it has been lost. The potential to reproduce sexually can be lost through the rise of single-mating-type populations (in heterothallic species only) (e.g. Paoletti *et al*., 2006) or through genetic decay of

reproductive genes (Strandberg *et al*., 2010; Wik *et al*., 2008). Indeed in *Fusarium*, many isolates display limited or no fertility (Kuhlman, 1982), the causes of which remain unknown. It is, therefore, possible that selection may have favored lateral transfer of functional *MAT* genes among *Fusarium* spp at certain points in their evolution.

The most intriguing phylogenetic aberration arising in this study was the grouping of the *F. sacchari MAT1-1* idiomorph outside the *G. fujikuroi* complex, while the *MAT1-2* idiomorph grouped in the Asian clade as expected. Further investigation using a larger dataset might improve phylogenetic resolution and rule out potential artifacts such as long-branch-attraction. The lateral transfer hypothesis appears doubtful because this would require a hybridization event between *F. sacchari* and a somewhat distant relative. Even within the *G. fujikuroi* complex, distinct mating populations (of which *F. sacchari* constitutes MP-B) are thought to be strongly reproductively isolated (Leslie, 1991; Leslie *et al*., 2004). Interestingly, this opinion did not prevail originally; early studies on *F. sacchari* suggested that it may indeed be cross-fertile with several other mating populations (Kuhlman, 1982). However, Britz *et al*. (1999) demonstrated that the progeny produced in such crosses, rather than being hybrids, were uniparental. This rather implies that *F. sacchari* is capable of homothallic reproduction. In light of the knowledge that *F. sacchari* has a heterothallic *MAT* locus and a *MAT1-1* idiomorph that is highly dissimilar from others in the complex, these earlier claims should be re-visited.

Unlike the phylogenies of heterothallic *MAT* loci presented here and by Strandberg *et al*. (2010), the phylogeny of *MAT* loci in the homothallic *F. graminearum* complex generated by O'Donnell *et al*. (2004) contained few well-supported deviations from the species tree, and did not have a significantly better fit to the data. This could be taken as evidence that lateral gene-flow is less frequent in this homothallic lineage (despite apparently weak reproductive isolation O'Donnell *et al*. (2004)). However, in the homothallic *MAT* locus, introgressed alleles that do not become fixed might recombine with native alleles. This could make lateral transfer difficult to detect and might even cause some sequence homogenization among species. Indeed, O'Donnell *et al*. (2004) reported low inter-specific divergence and phylogenetic informativeness in *MAT* loci compared to other nuclear genes of the *F. graminearum* complex. In the current study, sequence divergence in coding portions of the *MAT* locus (relative to two house-keeping genes) was found to be lower in the homothallic than in the heterothallic complex. Nucleotide diversity was lower at both synonymous and non-synonymous sites, indicating that this was not the result of stronger purifying selection, which would only reduce diversity in the latter. These findings are therefore in agreement with the idea that homothallic *MAT* genes that experience gene flow could diverge more slowly due to recombination between native and introgressed alleles.

In contrast with the coding sequences, non-coding portions of the *MAT* loci were more divergent among homothallic than heterothallic species. This finding also fits with the expectations given the differences in recombination. Heterothallic *MAT* loci do not recombine and all parts are therefore strongly linked. As a result, noncoding regions of heterothallic *MAT* loci are not independent, and could diverge more slowly due to functional constraint acting upon linked coding regions (Hudson, 1994). Free recombination in homothallic species would mean that the non-coding portions are more independent of any constraint affecting the genes.

Of the four available *Fusarium* genomes, only one was homothallic. Therefore, at the genome-wide level, a specific comparison between the rate of divergence among homothallic vs. heterothallic species was impossible. However, in general, *MAT* proteins were highly divergent compared to the estimate for the average protein. This is in agreement with the findings for *Cochliobolus* (Turgeon, 1998) and *Neurospora* (Wik et al., 2008) and also with the trend of rapid evolution in reproductive proteins (Civetta and Singh, 1998; Swanson and Vacquier, 2002; Clark *et al*., 2006). This rapid divergence of *MAT* genes over short evolutionary distances helps to explain their dramatic lack of

conservation among genera and the difficulty experienced by researchers in identifying *MAT* loci in new species (Cisar *et al*., 1994; Arie *et al*., 1997).

Likelihood-based tests for selection affecting *MAT* genes, revealed evidence for the presence codons under positive diversifying selection in one gene, *MAT1-1-1*, in both species complexes. In homothallic species, *MAT1-1-1* and *MAT1-1-3* were also found to exhibit evolutionary patterns not significantly different from neutrality (lack of selective constraint). The reliability of this result is questionable, given the low frequency of inter-specific differences in the *MAT* genes in this complex. However, it does contradict O'Donnell *et al*. (2004) who reported the *MAT* genes to be under "strong purifying selection". The reason for this distinction is unclear but it might stem from the smaller dataset used here, which included only homothallic species from within the *F. graminearum* complex, with only a single sequence representing each species. Our results are, however, similar to those of Wik *et al*. (2008), who reported a lack of selective constraint affecting *MAT* genes of homothallic *Neurospora* spp. Unlike in Neurospora, there is no evidence that *MAT* genes are dispensable in homothallic *Fusarium* sp. (Lee *et al*., 2003). Furthermore, sexual reproduction cannot be considered a superfluous trait in homothallic *Fusarium* sp., in fact the sexual spores of *F. graminearum* provide an important means of dispersal and infection (Markell and Franch, 2003). However, homothallic species do retain the ability to outcross in a heterothallic manner, even in the absence of one functional *MAT* gene (Lee *et al*., 2003). This redundancy is a possible explanation for the observed relaxation of selective constraint in homothallic *MAT* genes.

Conclusion

This study has provided new insights into the structure and evolution of the *Fusarium MAT* locus. *Fusarium* spp. appear to encode a novel *MAT* gene, *MAT1-2-3*, which may be absent from many other genera. Another noteworthy finding is the possibility that novel regulatory motifs could have been recruited by certain *MAT* genes in the transition to homothallism, potentially allowing independent gene regulation. There is some phylogenetic evidence for lateral transfer of *MAT* genes in *Fusarium*, and the *MAT1-1* idiomorph of *F. sacchari* in particular will require further attention. There is little such evidence among homothallic species, although geneflow might be more difficult to detect in the face of recombination. In general *Fusarium MAT* proteins appear to evolve rapidly. One gene, *MAT1-1-1*, was found to be under positive diversifying selection. Again, recombination in homothallic *MAT* loci could stem the rapid divergence if there is gene-flow. Some homothallic *MAT* genes also display evidence for relaxed selective constraint, which may be associated with the partial functional redundancy. Overall these findings demonstrate that *Fusarium MAT* loci offer a useful model in which study many different aspects of molecular evolution.

Acknowledgements

We thank Dr. David Geiser for reading the manuscript and offering useful advice. This work was funded by The National Research Foundation (NRF), the DST (Department of Science and Technology)/NRF Centre of Excellence in Tree Health Biotechnology (CTHB), the University of Pretoria and the Tree Protection Cooperative Programme (TPCP).

Appendix A. Supplementary material

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.fgb.2011.03.005.

References

Arie, T., Christiansen, S.K., Yoder, O.C., Turgeon, B.G., 1997. Efficient cloning of ascomycete mating type genes by PCR amplification of the conserved *MAT* HMG Box. Fungal Genet. Biol. 21, 118–130.

Arie, T., Yoshida, T., Shimizu, T., Kawabe, M., Yoneyama, K., Yamaguchi, I., 1999. Assessment of *Gibberella fujikuroi* mating type by PCR. Mycoscience 40, 311–314.

Birky, C.W., Walsh, J.B., 1988. Effects of linkage on rates of molecular evolution. Proc. Nat. Acad. Sci. USA 85, 6414–6418.

Britz, H., Coutinho, T.A., Wingfield, M.J., Marasas, W.F.O., Gordon, T.R., Leslie, J.F., 1999. *Fusarium subglutinans* f. sp. *pini* represents a distinct mating population in the *Gibberella fujikuroi* species complex. Appl. Environ. Microbiol. 65, 1198–1201.

Britz, H., Steenkamp, E.T., Coutinho, T.A., Wingfield, B.D., Marasas, W.F.O., Wingfield, M.J., 2002. Two new species of *Fusarium* section *Liseola* associated with mango malformation. Mycologia 94, 722–730.

Charlesworth, B., 2009. Fundamental concepts in genetics: effective population size and patterns of molecular evolution and variation. Nat. Rev. Genet. 10, 195–205.

Chaturvedi, V., Fan, J., Stein, B., Behr, M.J., Samsonoff, W.A., Wickes, B.L., Chaturvedi, S., 2002. Molecular genetic analyses of mating pheromones reveal intervariety mating or hybridization in *Cryptococcus neoformans*. Infect. Immun. 70, 5225–5235.

Cisar, C.R., TeBeest, D.O., Spiegel, F.W., 1994. Sequence similarity of mating type idiomorphs: a method which detects similarity among the Sordariaceae fails to detect similar sequences in other filamentous ascomycetes. Mycologia 86, 540–546.

Civetta, A., Singh, R.S., 1998. Sex-related genes, directional sexual selection, and speciation. Mol. Biol. Evol. 15, 901–909.

Clark, N.L., Aagaard, J.E., Swanson, W.J., 2006. Evolution of reproductive proteins from animals and plants. Reproduction 131, 11–22.

Coppin, E., Debuchy, R., Arnaise, S., Picard, M., 1997. Mating types and sexual development in filamentous ascomycetes. Microbiol. Mol. Biol. Rev. 61, 411–428.

Devier, B., Aguileta, G., Hood, M.E., Giraud, T., 2010. Using phylogenies of pheromone receptor genes in the *Microbotryum violaceum* species complex to investigate possible speciation by hybridization. Mycologia 102, 689–696.

Doohan, F.M., Brennan, J., Cooke, B.M., 2003. Influence of climatic factors on *Fusarium* species pathogenic to cereals. Eur. J. Plant Path. 109, 755–768.

Gasser, T., Muller, H.G., Mammitzsch, V., 1985. Kernels for nonparametric curve estimation. J. Roy Stat Soc. Ser. B (Meth.) 47, 238–252.

Glass, N.L., Grotelueschen, J., Metzenberg, R.L., 1990. *Neurospora crassa* a matingtype region. Proc. Nat. Acad. Sci. USA 87, 4912–4916.

Glass, N.L., Smith, M.L., 1994. Structure and function of a mating-type gene from the homothallic species *Neurospora africana*. Mol. Gen. Genet. 244, 401–409.

Glass, N.L., Vollmer, S.J., Staben, C., Grotelueschen, J., Metzenberg, R.L., Yanofsky, C., 1988. DNAs of the two mating-type alleles of *Neurospora crassa* are highly dissimilar. Science 241, 570–573.

Goode, M.G., Rodrigo, A.G., 2007. SQUINT: a multiple alignment program and editor. Bioinformatics 23, 1553–1555.

Güldener, U., Mannhaupt, G., Münsterkötter, M., Haase, D., Oesterheld, M., Stümpflen, V., Mewes, H.-W., Adam, G., 2006a. FGDB: A comprehensive fungal genome resource on the plant pathogen *Fusarium graminearum*. Nucleic Acids Res. 34, 456–458.

Güldener, U., Seong, K., Boddu, J., Cho, S., Trail, F., Xu, J., Adam, G., Mewes, H., Muehlbauer, G., Kistler, H., 2006b. Development of a *Fusarium graminearum* affymetrix GeneChip for profiling fungal gene expression in vitro and *in planta*. Fungal Genet. Biol. 43, 316–325.

Guindon, S., Delsuc, F., Dufayard, J.F., Gascuel, O., 2009. Estimating maximum likelihood phylogenies with PhyML. Met. Mol. Cell. Biol. 537, 113–137.

Hudson, R.R., 1994. How can the low levels of DNA sequence variation in regions of the *Drosophila* genome with low recombination rates be explained? Proc. Nat. Acad. Sci. USA 91, 6815–6818.

Johnson, A.D., 1995. Molecular mechanisms of cell-type determination in budding yeast. Curr. Opin. Genet. Dev. 5, 552–558.

Kanamori, M., Kato, H., Yasuda, N., Koizumi, S., Peever, T.L., Kamakura, T., Teraoka, T., Arie, T., 2007. Novel mating type-dependent transcripts at the mating type locus in *Magnaporthe oryzae*. Gene 403, 6–17.

Kimura, M., 1981. Estimation of evolutionary distances between homologous nucleotide sequences. Proc. Nat. Acad. Sci. USA 78, 454–458.

Kjaerulff, S., Dooijes, D., Clevers, H., Nielsen, O., 1997. Cell differentiation by interaction of two HMG-box proteins: *Mat1-Mc* activates M cell-specific genes in *S. pombe* by recruiting the ubiquitous transcription factor *Ste11* to weak binding sites. EMBO J. 16, 4021–4033.

Klaasen, J.A., Nelson, P.E., 1996. Identification of a mating population, *Gibberella nygamai* sp. nov., within the *Fusarium nygamai* anamorph. Mycologia 88, 965–969.

Klittich, C.J.R., Leslie, J.F., 1992. Identification of a second mating population within the *Fusarium moniliforme* anamorph of *Gibberella fujikuroi*. Mycologia 84, 541–547.

Klittich, C.J.R., Leslie, J.F., Nelson, P.E., Marasas, W.F.O., 1997. *Fusarium thapsinum* (*Gibberella thapsina*): a new species in section *Liseola* from sorghum. Mycologia 89, 643–652.

Kronstad, J.W., Staben, C., 1997. Mating type in filamentous fungi. Ann. Rev. Genet. 31, 245–276.

Kuhlman, E.G., 1982. Varieties of *Gibberella fujikuroi* with Anamorphs in *Fusarium* Section. Mycologia 74, 759–768.

Kvas, M., Marasas, W.F.O., Wingfield, B.D., Wingfield, M.J., Steenkamp, E.T., 2009. Diversity and evolution of *Fusarium* species in the *Gibberella fujikuroi* complex. Fungal Div. 34, 1–21.

Lee, J., Lee, T., Lee, Y.W., Yun, S.H., Turgeon, B.G., 2003. Shifting fungal reproductive mode by manipulation of mating type genes: obligatory heterothallism of *Gibberella zeae*. Mol. Microbiol. 50, 145–152.

Lee, S.C., Ni, M., Li, W., Shertz, C., Heitman, J., 2010. The evolution of sex: a perspective from the fungal kingdom. Microbiol. Mol. Biol. Rev. 74, 298–340.

Leslie, J., Zeller, K., Wohler, M., Summerell, B., 2004. Interfertility of two mating populations in the *Gibberella fujikuroi* species complex. Eur. J. Plant Path. 110, 611–618.

Leslie, J.F., 1991. Mating populations in *Gibberella fujikuroi* (*Fusarium* section *Liseola*). Phytopathology 81, 1058–1060.

Leslie, J.F., 1995. *Gibberella fujikuroi*: available populations and variable traits. Can. J. Bot. 73, 282–291.

Leslie, J.F., Summerell, B.A., Bullock, S., Doe, F.J., 2005. Description of *Gibberella sacchari* and Neotypification of its Anamorph *Fusarium sacchari*. Mycologia 97, 718–724.

Librado, P., Rozas, J., 2009. DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25, 1451–1452.

Loubradou, G., Begueret, J., Turcq, B., 1997. A mutation in an HSP90 gene affects the sexual cycle and suppresses vegetative incompatibility in the fungus *Podospora anserina*. Genetics 147, 581–588.

Markell, S.G., Francl, L.J., 2003. Fusarium head blight inoculum: species prevalence and *Gibberella zeae* spore type. Plant Dis 87, 814–820.

Nix, D., Eisen, M., 2005. GATA: a graphic alignment tool for comparative sequence analysis. BMC Bioinf. 6, 9.

O'Donnell, K., Cigelnik, E., Nirenberg, H.I., 1998. Molecular systematic and phylogeography of the *Gibberella fujikuroi* species complex. Mycologia 90, 465–493.

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

O'Donnell, K., Ward, T.J., Geiser, D.M., Corby Kistler, H., Aoki, T., 2004. Genealogical concordance between the mating type locus and seven other nuclear genes supports formal recognition of nine phylogenetically distinct species within the *Fusarium graminearum* clade. Fungal Genet. Biol. 41, 600–623.

O'Donnell, K., Ward, T.J., Aberra, D., Kistler, H.C., Aoki, T., Orwig, N., Kimura, M., Bjørnstad, A., Klemsdal, S.S., 2008. Multilocus genotyping and molecular phylogenetics resolve a novel head blight pathogen within the *Fusarium graminearum* species complex from Ethiopia. Fungal Genet. Biol. 45, 1514–1522.

Paoletti, M., Buck, K.W., Brasier, C.M., 2006. Selective acquisition of novel mating type and vegetative incompatibility genes via interspecies gene transfer in the globally invading eukaryote Ophiostoma novo-ulmi. Mol. Ecol. 15, 249–262.

Pöggeler, S., Nowrousian, M., Kück, U., 2006. Fruiting-body development in ascomycetes. In: Kües, U., Fischer, R. (Eds.), Growth, Differentiation and Sexuality. Vol. 1 of The Mycota. Springer Berlin Heidelberg, Berlin/Heidelberg, (Chapter 16). pp. 325–355.

Posada, D., 2008. jModelTest: Phylogenetic model averaging. Mol. Biol. Evol. 25, 1253–1256.

Salamov, A.A., Solovyev, V.V., 2000. *Ab initio* gene finding in *Drosophila* genomic DNA. Genom. Res. 10, 516–522.

Sambrook, J., Russell, D., 2001. Molecular Cloning: A Laboratory Manual, third ed. Cold Spring Harbor Laboratory Press, Plainview, NY.

Schmidt, H.A., Strimmer, K., Vingron, M., von Haeseler, A., 2002. TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing. Bioinformatics 18, 502–504.

Shimodaira, H., 2002. An approximately unbiased test of phylogenetic tree selection. Syst. Biol. 51, 492–508.

Shimodaira, H., Hasegawa, M., 1999. Multiple comparisons of log-likelihoods with applications to phylogenetic inference. Mol. Biol. Evol. 16, 1114.

Shimodaira, H., Hasegawa, M., 2001. CONSEL: for assessing the confidence of phylogenetic tree selection. Bioinformatics 17, 1246–1247.

Shiu, P.K.T., Glass, N.L., 2000. Cell and nuclear recognition mechanisms mediated by mating type in filamentous ascomycetes. Curr. Opin. Microbiol. 3, 183–188.

Staben, C., Yanofsky, C., 1990. *Neurospora crassa* a mating-type region. Proc. Nat. Acad. Sci. USA 87, 4917–4921.

Stanke, M., Steinkamp, R., Waack, S., Morgenstern, B., 2004. AUGUSTUS: A web server for gene finding in eukaryotes. Nucleic Acids Res. 32, W309–W312.

Steenkamp, E.T., Wingfield, B.D., Coutinho, T.A., Wingfield, M.J., Marasas, W.F.O., 1999. Differentiation of *Fusarium subglutinans* f. sp. *pini* by histone gene sequence data. Appl. Environ. Microbiol. 65, 3401–3406.

Steenkamp, E.T., Wingfield, B.D., Coutinho, T.A., Zeller, K.A., Wingfield, M.J., Marasas, W.F.O., Leslie, J.F., 2000. PCR-based identification of *MAT-1* and *MAT-2* in the *Gibberella fujikuroi* species complex. Appl. Environ. Microbiol. 66, 4378–4382.

Steenkamp, E.T., Wingfield, B.D., Desjardins, A.E., Marasas, W.F.O., Wingfield, M.J., 2002. Cryptic speciation in *Fusarium subglutinans*. Mycologia 94, 1032–1043.

Strandberg, R., Nygren, K., Menkis, A., James, T.Y., Wik, L., Stajich, J.E., Johannesson, H. Conflict between reproductive gene trees and species phylogeny among heterothallic and pseudohomothallic members of the filamentous ascomycete genus *Neurospora*. Fungal Genet. Biol. 47: 869–878.

Swanson, W.J., Vacquier, V.D., 2002. The rapid evolution of reproductive proteins. Nat. Rev. Genet. 3, 137–144.

Tatusova, T.A., Madden, T.L., 1999. BLAST 2 Sequences, a new tool for comparing protein and nucleotide sequences. FEMS Microbiol. Lett. 174, 247–250.

Tavaré, S., 1986. Some Probabilistic and Statistical Problems in the Analysis of DNA Sequences. Amer Mathematical Society, vol. 17 , 57–86.

Thompson, J.D., Higgins, D.G., Gibson, T.J., 1994. CLUSTAL w: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 22, 4673–4680.

Turgeon, B.G., 1998. Application of mating type gene technology to problems in fungal biology. Ann. Rev. Phytopath. 36, 115–137.

Turgeon, B.G., Yoder, O.C., 2000. Proposed nomenclature for mating type genes of filamentous ascomycetes. Fungal Genet. Biol. 31, 1–5.

Vicoso, B., Charlesworth, B., 2009. Recombination rates may affect the ratio of X to autosomal non-coding polymorphism in African populations of *Drosophila melanogaster*. Genetics 181, 1699–1701.

Wik, L., Karlsson, M., Johannesson, H., 2008. The evolutionary trajectory of the mating-type (*mat*) genes in *Neurospora* relates to reproductive behavior of taxa. BMC Evol. Biol. 8, 109.

Wingender, E., Dietze, P., Karas, H., Knüppel, R., 1996. TRANSFAC: A database on transcription factors and their DNA binding sites. Nucleic Acids Res. 24, 238–241.

Yang, Z., 2007. PAML 4: phylogenetic analysis by maximum likelihood. Mol. Biol. Evol. 24, 1586–1591.

Yang, Z.H., Nielsen, R., Goldman, N., Pedersen, A.M.K., 2000. Codon-substitution models for heterogeneous selection pressure at amino acid sites. Genetics 155, 431–449.

Yli-Mattila, T., Gagkaeva, T., Ward, T.J., Aoki, T., Kistler, H.C., O'Donnell, K., 2009. A novel Asian clade within the *Fusarium graminearum* species complex includes a newly discovered cereal head blight pathogen from the Russian Far East. Mycologia 101, 841–852.

Yun, S.H., Berbee, M.L., Yoder, O.C., Turgeon, B.G., 1999. Evolution of the fungal self-fertile reproductive life style from self-sterile ancestors. Proc. Nat. Acad. Sci. USA 96, 5592–5597.

Yun, S.-H., Arie, T., Kaneko, I., Yoder, O.C., Turgeon, B.G., 2000. Molecular organization of mating type loci in heterothallic, homothallic, and asexual *Gibberella*/*Fusarium* species. Fungal Genet. Biol. 31, 7–20.

Zeller, K.A., Summerell, B.A., Bullock, S., Leslie, J.F., 2003. *Gibberella konza* (*Fusarium konzum*) sp. nov. from prairie grasses, a new species in the *Gibberella fujikuroi* species complex.. Mycologia 95, 943–954.