Causes and consequences of variability in peptide mating pheromones of ascomycete fungi

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, 0002, South Africa

[2] Department of Microbiology and Plant Pathology, Forestry and Agricultural Biotechnology Institute, University of Pretoria, 0002, South Africa

*Corresponding author

Department of Genetics

University of Pretoria

Hatfield

Pretoria, 0002

South Africa

Email: brenda.wingfield@fabi.up.ac.za

Keywords: a-factor; α-factor; Pheromone precursor; *Gibberella fujikuroi*; Ascomycota

Running head: Pheromone diversity in ascomycete fungi

Abstract

The reproductive genes of fungi, like those of many other organisms, are thought to diversify rapidly. This phenomenon could be associated with the formation of reproductive barriers and speciation. Ascomycetes produce two classes of mating-type-specific peptide pheromones. These are required for recognition between the mating types of heterothallic species. Little is known regarding the diversity or the extent of species-specificity in pheromone peptides among these fungi. We compared the putative protein-coding DNA sequences of the two pheromone classes from 70 species of Ascomycetes. The dataset included previously-described pheromones and putative pheromones identified from genomic sequences. In addition, pheromone genes from twelve *Fusarium* species in the *Gibberella fujikuro*i complex were amplified and sequenced. Pheromones were largely conserved among species in this complex and, therefore, cannot alone account for the reproductive barriers observed between these species. In contrast, pheromone peptides were highly diverse among many other Ascomycetes, with evidence for both positive diversifying selection and relaxed selective constraint. Repeats of the α-factor-like pheromone, which occur in tandem arrays of variable copy number, were found to be conserved through purifying selection and not concerted evolution. This implies that sequence-specificity may be important for pheromone reception and that inter-specific differences may indeed be associated with functional divergence. Our findings also suggest that frequent duplication and loss causes the tandem-repeats to experience "birth-and-death" evolution, which could in fact facilitate inter-specific divergence of pheromone peptide sequences.

Introduction

Sexual reproduction in ascomycete fungi (phylum Ascomycota) commences with the interaction between compatible cells, which is facilitated by pheromones and receptors (Kurjan 1993; Kim and Borkovich 2004, 2006). In heterothallic Ascomycetes sexual reproduction can occur only between two individuals of different "mating type". Pheromones are essential for successful mating and are produced in a mating type-specific manner in heterothallic species (Bender and Sprague-Jr 1989; Bobrowicz et al. 2002; Kim and Borkovich 2004; Coppin et al. 2005; Kim and Borkovich 2006; Mayrhofer et al. 2006). In homothallic species, which can reproduce sexually through self-fertilization, mating pheromones might be somewhat dispensable (Mayrhofer and Pöggeler 2005; Kim et al. 2008; Lee et al. 2008).

Ascomycete fungi produce two classes of mating pheromones. These were first comprehensively studied in the yeast *Saccharomyces cerevisiae* and characterized as mating factors a and α (or a-factor and α-factor) (Stötzler and Duntze 1976; Betz et al. 1977; Betz et al. 1987). Both pheromone classes are cleaved from larger precursor proteins (Kurjan and Herskowitz 1982; Singh et al. 1983; Brake et al. 1985). Homologous pheromones and precursor genes have been described from numerous yeasts and filamentous Ascomycetes (all references and GenBank accession numbers are provided as supplementary material).

In *S. cerevisiae*, two genes (*Mfa1* and *Mfa2*) each encode a Prepro a-factor precursor polypeptide, from which the eleven amino acid mature a-factor pheromone is cleaved. The precursor is characterized by the presence of a carboxyl (C)-terminal prenylation signal or so-called CaaX

motif, typical of fungal pheromone precursors. Prenylation is required for transport of the precursor to the plasma membrane where cleavage occurs (Brake et al. 1985). There are also two genes (*MFα1* and *MFα2*) that encode Prepro α-factor polypeptides, which respectively contain four and two repeats of the thirteen-residue mature α-factor (Singh et al. 1983). Each mature peptide repeat is bordered by signals for cleavage by Ste13p and Kex2p on its N and C-terminal borders, respectively (Julius et al. 1983; Julius et al. 1984). It is thought that all the copies of the mature α-factor are cleaved and released by the cell (Caplan and Kurjan 1991).

Observations in a wide range of organisms have demonstrated that reproductive proteins, particularly those under sex-biased expression such as gamete recognition proteins, evolve rapidly, often under positive, diversifying selection (reviewed by Swanson and Vacquier 2002; Clark et al. 2006). Whether ascomyete pheromones display this trend has not been ascertained, although this appears to be the case for fungal mating-type (*MAT*) genes (Turgeon 1998; Brown and Casselton 2001; Wik et al. 2008). The *MAT* loci encode putative transcription factors that determine mating type and control sexual development (Coppin et al. 1997) (including pheromone production [Bobrowicz et al. 2002]). A number of processes have been proposed to explain the phenomenon of rapid evolution in reproductive proteins. These include reinforcement (Howard 1993), and sexual selection (Palumbi 1999; Galindo et al. 2003). Regardless of its source, a fundamental consequence of the rapid evolution of interacting reproductive proteins is the potential to generate reproductive isolation between populations or species (Clark et al. 2006; Palumbi 2008).

The importance of pheromone peptide sequence for successful mate recognition is unclear. The distinct pheromones of certain yeast species appear to offer some level some species-specificity, although this may be weak when pheromones are only slightly different, such as between members of the same genus (McCullough and Herskowitz 1979; Burke et al. 1980; Hisatomi et al. 1988). Naider and Becker (2004) have demonstrated that only certain residues in the α-factor are essential for receptor stimulation while others might be less important. For the pheromones of Basidiomycetes, alteration of certain residues can drastically affect the success of pheromone reception (Olesnicky et al. 2000; Fowler et al. 2001). It is, therefore, conceivable that reproductive barriers between some ascomycete species could stem from differences in their pheromone peptide sequences.

The potential for reproductive isolation as a result of species-specific mating cues, including chemical signals, is well known (Coyne and Orr 2004, p. 214-215). However, in fungi, there has only recently been exploration into the potential role of the pheromone/receptor system in the development of species boundaries (e.g., Karlsson et al. 2008). In this study we considered the diversity and evolution of ascomycete pheromones, with particular focus on the multiple tandem repeat α-factor-like pheromone, and whether these peptides could potentially play a role in the generation of reproductive isolation and speciation.

Nucleotide sequences of previously described pheromones, as well as putative pheromones identified from genomic sequences available in public domain databases were used. In addition, we determined the pheromone gene sequences for twelve *Fusarium* spp. in the *Gibberella fujikuroi* complex. Species in this group display varying levels of sexual inter-compatibility,

ranging from complete reproductive isolation to complete inter-fertility (Desjardins et al. 2000; Leslie et al. 2004; Leslie and Summerall 2006; Kvas et al. 2009). To understand diversity in ascomycete pheromones, the rate of pheromone evolution was compared with that of house-keeping genes and tests for positive diversifying selection were performed. Furthermore, to test the hypothesis that concerted evolution could act to homogenize multiple tandem repeats of the α-factor-like pheromone, evolutionary relationships among repeats within and between species were examined. The structural organization of the α-class precursor gene was also compared among the twelve *G. fujikuroi* complex species.

Materials and Methods

Fungal isolates and DNA isolation

A total of 24 *Fusarium* isolates were included in this study (see supplementary material). These isolates represent the standard MAT-1 and MAT-2 mating type tester strains for the nine biological species or reproductively-isolated mating populations (MP-A to MP-I) of the *G. fujikuroi* complex (reviewed by Kvas et al. 2009). All but one of these mating populations corresponds to a single phylogenetic species (O'Donnell et al. 1998; Kvas et al. 2009). The exception is MP-E, which consists of two distinct phylogenetic species that are inter-fertile (referred to here as *F. subglutinans* groups '1' and '2') (Steenkamp et al. 2002). Mating populations C and D (*F. fujikuroi* and *F. proliferatum*, respectively) have been shown to share a moderate level of inter-fertility (Leslie et al., 2004). A single inter-specific cross has also been observed between a pair of isolates from MP-H (*F. circinatum*) and MP-E (*F. subglutinans* group

1) (Desjardins et al. 2000). We also two isolates from each of two species (*F. mangiferae* and *F. sterilihyphosum*) for which no sexual stage is known (Britz et al. 2002). DNA was extracted from all isolates following a protocol based on that of Steenkamp et al. (1999).

Sequences and organisation of pheromone precursor genes

Putative pheromone precursor genes were identified by BLASTp and tBLASTn using, as queries, all currently known homologues of these genes in Ascomycetes (see supplementary material for references and GenBank accession numbers). In addition, a-class precursors were recognized by the presence of the characteristic C-terminal 'CaaX' motif (Brake et al. 1985). Likewise, α-class precursors were recognized by the presence of multiple repeats of an α-factor-like peptide bordered by the characteristic cleavage signals for homologues of Ste13p (Julius et al. 1983; Pöggeler 2000; Bobrowicz et al. 2002) and Kex2p (Julius et al. 1984; Darby and Smyth 1990).

To amplify the putative a- and α-class precursor genes (designated *ppg1* and *ppg2*, respectively) from species of the *G. fujikuroi* complex, Primer Designer 4.20 (Sci Ed Central, Cary, NC) was used to design primers in the conserved regions flanking the genes in *Fusarium verticillioides* and *Fusarium oxysporum* (GenBank accession numbers AAIM02000073 and AAXH01000548). To amplify *ppg2*, primers ppg2-2 (5'-TGTCTGGCAGCAACACCATC-3') and ppg2-3 (5'-CCGTCACTCAGAGCCAGGTA-3') were used (Fig.1). The *ppg1* gene was amplified in three overlapping fragments (Fig. 1) with primer set ppg1-7c (5'-ATATCACCGACGTACTGTAA-3') and ppg1-8c (5'-TACGAGTACCACTCACTT-3'), primer set ppg1-S1 (5'CTGCAACCTCGAYTAYAA-3') and ppg1-T1 (5'-ACCAGTAGCACCRTCRTC-3'), and

primer set ppg1-S1 and ppg1-3 (5'-ATGGAGCGCTTGGCCTTGTG-3'). PCR reaction mixtures were 25µl in volume and contained 4ng/µl template DNA, 1.5mM $MgCl_2$, 0.4µM of each primer, 1mM deoxynucleotide triphosphates (0.25 mM of each), and 0.05U/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°C for 60s; thirty cycles of denaturation at 94°C for 30s, annealing for 30s and extension at 70°C for 90s; followed by a final extension step at 70°C for 10 min. An annealing temperature of 60°C was used for all primer combinations, except ppg1-7c and ppg1-8c, where an annealing temperature of 58°C was used.

For sequencing, PCR products were precipitated overnight in ethanol containing 0.1M Sodium Acetate (pH 3.8) at 4°C followed by centrifugation at 16,000 *rcf* for 30 min at 4°C. The pellets were washed in 70% ethanol, air dried following centrifugation (16,000 *rcf*; 10 min; 4°C) and removal of the supernatant, and resuspended in 50µl sterile distilled water. These purified products were then subjected to automated Sanger sequencing in both directions using the original PCR primers, the BigDye® terminator v3.1 cycle sequencing kit (Applied Biosystems, Foster City, CA) and an ABI PRISM® 377 DNA sequencer (Applied Biosystems). Sequence assembly, annotation and *in silico* translation was performed using Vector NTI 9.0 (Invitrogen Life Technologies, Carlsbad, CA, USA) and CLC Bio Main Workbench (CLC Bio, Aarhus, Denmark).

PCR products that did not yield clean chromatograms when sequenced were purified by cloning 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 25µl and contained 1mM deoxynucleotide triphosphates (0.25 mM of each), 0.4µM of each primer, 0.05U/µl FastStart *Taq* DNA Polymerase and FastStart reaction buffer with $MgCl_2$ (Roche Diagnostics, Mannheim, Germany). The PCR reaction conditions were as follows: denaturation at 94°C for 5 min; followed by thirty cycles of denaturation at 94°C for 30s, annealing at 50°C for 30s and extension for 90s at 72°C; followed by a final extension step at 72°C for 10 min. Sequencing was performed as described above, using the SP-6 and T-7 primers.

Inter-specific variation and tests for positive selection

For evolutionary analyses, only the regions of the precursor genes encoding the predicted mature peptides were considered. As these peptides were divergent over large phylogenetic distances, species were only compared with other members of the same fungal Class (Fig. 2, 3 and 4). Coding DNA sequences were aligned using the codon-based algorithm implemented in SQUINT (Goode and Rodrigo 2007). As a measure of the rate of protein sequence evolution, pair-wise d_N/d_S ratios were used, where d_N is the number of non-synonymous substitutions per non-synonymous site and d_S the number of synonymous substitutions per synonymous site. This is an effective measure of evolutionary rate as it controls for differential phylogenetic distances and differential mutation rates (Ina 1996). To calculate d_N/d_S the Nei and Gojobori (1986) method was used as implemented by the PAML package, version 4.3, (Yang 1997). All sequence pairs that were inferred by the program to be saturated for synonymous differences were excluded. For

the multiple-copy peptides, all possible pairs between two species were averaged to give a single pair-wise value. Pair-wise d_N/d_S values were also determined for house-keeping (HK) genes using coding sequences for the β-tubulin and translation elongation factor 1-α genes (when available on GenBank).

Statistical analyses were performed using MYSTAT (Wilkinson 1987) version 12. The data were non-normally distributed according to the Shapiro-Wilk test (Shapiro and Wilk 1965). Since the data were paired (pheromone and HK value for each species pair) the Wilcoxon Signed Ranks test (Wilcoxon 1945) was used to determine whether d_N/d_S values differed significantly between pheromone and HK genes.

To test for positive diversifying selection acting at specific sites, two "site-models" (Yang et al. 2000a) were tested using the maximum-likelihood approach implemented by the CODEML program in the PAML package. Model M7 (the beta model) assumes a β distribution of ω with $0 \leq \omega \leq 1$, while model M8 (the beta&ω model) adds a class of codons with $\omega \geq 1$. The two models were compared with respect to their fit to the data using a likelihood ratio test (LRT), as described by Yang et al. (2000a). The Bayes empirical Bayes (BEB) output of CODEML was used to identify specific codons likely to fall into codon class $\omega \geq 1$ in model M8. For the multiple-repeat α-factor pheromone, the first repeat in each module was used to represent each species. This was done to avoid the risk that purifying selection acting at the within-species level could mask diversifying selection at the between-species level.

Relationships among tandemly-repeated α-factor-like pheromone sequences

To test the hypothesis that concerted evolution acts to homogenize DNA sequences among tandemly repeated copies of the α-factor-like pheromone, the relationships among DNA sequences encoding the mature pheromone peptides were examined. All repeats from all species in the Saccharomycetes and Sordariomycetes datasets (see supplementary material for sequences) were considered, as these two datasets contained a large number of species including both close and distant relatives. DNA sequences were aligned using CLUSTAL W (Thompson et al. 1994). The proportion of nucleotide sites at which each pair of sequences differed (*i.e.*, p-distances) (Nei and Kumar 2000) was used to construct neighbor-joining trees using MEGA version 4 (Tamura et al. 2007). This simple distance-based model was selected because the mode of evolution of repeats was unclear, and repeats experiencing concerted evolution would not necessarily follow any particular nucleotide-substitution model. Hence, the focus was not on evolutionary relationships *per se* but simply on whether the DNA sequence of each repeat shared more similarity with adjacent tandem repeats or with a corresponding repeat from a related species.

Another means, used by Swanson and Vacquier (1998), to distinguish between concerted evolution and purifying selection was applied by comparing synonymous distance (d_S) with non-synonymous distance (d_N). Purifying selection should theoretically act only at the protein level, consequently maintaining a low d_N while d_S is allowed to increase. In contrast, concerted evolution acts at the DNA level, causing both d_N and d_S to be low among tandem repeats. These values were calculated using the Yang and Nielsen method, as implemented in the program yn00 in the PAML package. Pair-wise d_S and d_N values were generated for every possible peptide pair

and then averaged such that each species had a single within-species value and each species pair had a single between-species value. Pairs saturated for synonymous or non-synonymous substitutions were discarded. When plotted, the gradient (m) of the of each point represents the d_N/d_S ratio. The between-species and within-species samples were independent and different in size, therefore the Mann-Whitney *U* test (Mann and Whitney, 1947) was implemented to determine whether m differed significantly between the samples.

To investigate the structural events that give rise to gain or loss of repeats, the *ppg1* gene sequence from an isolate of each of the twelve *Fusarium* species in the *G. fujikuroi* complex were used. The programme GATA (graphic alignment tool for comparative sequence analysis) (Nix and Eisen 2005) was used to visualise the DNA sequence relationships between each pair of species. Given two input sequences, this application uses the NCBI BLASTn (Altschul et al. 1990) and Bl2seq (Tatusova and Madden 1999) algorithms to generate all possible local alignments between two sequences. In this way the application is able to identify duplications, deletions and inversions. A word size of 20 was used, along with a lower cut-off bit-score of 24.9 to ensure that only regions of high homology were considered.

Results

Sequences and organisation of pheromone precursor genes

For the a-class precursors, published sequences were available for nine species (three yeast and six filamentous ascomycetes). For the α-class precursors, published sequences were available for

seventeen species (nine yeast and eight filamentous ascomycetes) (see supplementary material for references and GenBank accession numbers). Putative a-class and α-class precursor genes in the genomes of another 20 and 42 species respectively, including *F. verticillioides* were identified. In addition, the putative a-class and α-class precursor genes (*ppg2* and *ppg1*, respectively) from the 24 *Fusarium* isolates, representing 12 species of the *G. fujikuroi* complex, including *F. verticillioides* were sequenced (see supplementary material isolate information, sequences and GenBank accession numbers). Sequences were obtained for both genes from all twelve species, although one isolate of *F. sterilihyphosum* (MRC2802), failed to yield a fragment corresponding to *ppg1*. Thus a-class precursor sequences were obtained from 40 species, covering one filamentous class: Sordariomycetes, and two yeast classes: Saccharomycetes and Taphrinomycetes (Hibbett et al. 2007) (Fig. 2). For the α-class precursor, sequences were obtained from 69 species, covering four filamentous classes: Sordariomycetes, Eurotiomycetes, Leotiomycetes and Dothidiomycetes (Fig. 3), and two yeast classes: Saccharomycetes, and Taphrinomycetes (Fig. 4) (Hibbett et al. 2007). Most of the yeasts carried two or more copies of one or both precursor genes, while among the filamentous species, only *Cryphonectria parasitica* carried two copies of the a-class gene (*Mf2/1* and *Mf2/2*), as previously reported by Zhang et al. (1998) (Fig. 2). There was only one species, *Ashbya gosypii*, for which the a-class but no α-class precursor was obtained. In contrast, there were 31 species for which only the α-class precursor was obtained.

The putative a-class precursors were all small, ranging from 21 to 66 amino acids in length (Fig. 2). All carried a C-terminal CaaX motif (containing a Cystein [C] residue, followed by two aliphatic (a) residues and terminating in an arbitrary [X] residue). However, in the *Fusarium*

(*Gibberella*), *Verticillium* and *Trichoderma* (*Hypocrea*) species, the first aliphatic residue was replaced by a polar residue (Serine, Threonine or Asparagine), thus creating the pattern "CpaX", as was recognized by Schmoll et al. (2010). In some of the species in these three genera, one or more additional putative CpaX motifs were present in a repeated nature, and were usually preceded by a conserved peptide sequence, thus creating a precursor similar in structure to the α-class (Schmoll et al. 2010). Analysis by Schmoll et al. (2010) revealed that these genes are required for male fertility and are thus likely to encode true pheromone precursors. However, unlike those authors, we do not consider these to be a new or "hybrid" class of pheromone precursor but rather a variation of the a-class. In all twelve species of the *G. fujikuroi* complex, the a-class gene, *ppg2*, carried three repeats, except for *F. nygamai*, which carried two (Fig. 2). Similar repeats were found in the a-class precursors of *F. oxysporum* and both *Verticillium* species studied. Multiple CpaX motifs were also identified in two of the three *Trichoderma* species, although these were not preceded by a conserved peptide motif. *Gibberella zeae* (*Fusarium graminearum*) and *Trichoderma atroviridae* both had a single, C-terminal CpaX motif, and therefore resembled the typical a-class precursors (Fig. 2).

Based on sequence similarity to described a-factor-like peptides (Brake et al. 1985; Davey 1992), it was possible to identify predicted mature a-factor-like peptides from the remaining Saccharomycetes species (Fig. 2). Similar inferences could not be made for the 25 Sordariomycetes species examined because the exact sequence of the mature a-factor-like peptide has not been described from any filamentous Ascomycetes and the processing signals have not been identified. Further analysis of a-factor peptide evolution was, therefore, based only on the yeast a-factors.

In the *ppg2* gene of the twelve *Fusarium* species representing the *G. fujikuroi* complex, there was a conserved eight-residue peptide located immediately upstream of each CpaX motif (Fig. 2). These are not bordered by known Kex2p-like cleavage signals, and further investigation will be necessary to determine whether these represent repeats of a mature peptide that are each cleaved from the precursor. If so it is notable that there was complete identity in this amino acid sequence across all twelve *Fusarium* species. Even outside of these eight-residue motifs, there is no significant variation in the *ppg2* precursor among the *G. fujikuroi* species.

Among all the fungal sequences examined, the putative α-class precursor proteins were larger, ranging from 102 to 711 amino acids in length. All α-class precursors carried copies of a motif bordered by the Ste13p and Kex2p protease-like cleavage signals: an N-terminal run of X-Alanine or X-Proline dipeptides (Julius et al. 1983; Pöggeler 2000; Bobrowicz et al. 2002) and a C-terminal KR (Lysine-Arginine) or RR dipeptide (Julius et al. 1984; Darby and Smyth 1990). The motifs occurred in a tandem arrangement of between one (e.g. *Saccharomyces castellii*) and sixteen (e.g. *F. subglutinans* group 1) repeats, with an average of five (Fig. 3 and 4).

The predicted sequences of the mature α-factor-like peptides were inferred on the basis of sequence similarity to previously described homologues, and the presence of the Ste13p and Kex2p protease-like cleavage signals. Some predicted mature peptides, such as the last peptide repeat in all of the *Fusarium* spp., had Kex2p protease-like cleavage sites at both ends of the peptide. In several species, one or more copies of the predicted mature peptide were not bordered by one or either of the signature cleavage motifs. For example, in 20% of the putative peptides

among the 12 *Fusarium* species examined, the Kex2p protease-like cleavage site (usually KR or RR) was replaced by TR. However, these were still regarded as pheromones for evolutionary analyses, unless they were highly dissimilar from other repeats in the precursor. In 29 species, there was complete identity in amino acid sequence among all encoded repeats of the mature peptide, while the remaining species each encoded between two and six similar yet distinct "variants" of the pheromone (Fig. 3 and 4). The number of repeats in the precursor gene varied considerably among the species and sometimes between the two isolates of the same species; from a minimum of one to a maximum of sixteen.

In each of the *G. fujikuroi* complex isolates, the α-class gene, *ppg1*, encoded between four and eleven copies (average = 8) of a peptide with the sequence WCTWRGQPCW. Each isolate also encoded between two and five copies (average = 4) of a second peptide, in which the Threonine residue was substituted with Methionine. A single repeat of two additional variants of the peptide (one in which the Threonine was replaced by Leucine, and one in which the third Tryptophan residue was replaced with Cysteine) were each found in both isolates of *F. konzum*. A single copy of a fifth variant of the peptide in which the second Tryptophan residue is replaced with an Arginine residue, occurred in one of the two *F. circinatum* isolates (Fig. 3).

Inter-specific variation and tests for positive selection

Four datasets (a-factor-like pheromones of the Saccharomycetes and α-factor-like pheromones of the Saccharomycetes, Sordariomycetes and Eurotiomycetes) included a sufficient number of species pairs for sound statistical analyses. For most species pairs, pheromones displayed a far

greater evolutionary rate than house-keeping genes, although a few species had pheromones that were strongly conserved, with even slower evolutionary divergence than the house-keeping genes. However, average pairwise d_N/d_S values were found to be significantly greater for pheromones than house-keeping genes in all four datasets ($P \leq 0.05$) (Fig. 5).

Results of the CODEML analyses and LRTs are displayed in Table 1. In three of the four datasets, there was no significant difference in the likelihood values of each pair of models. Hence, the α-factor-like pheromone of Sordariomycetes and both the a- and α-factor-like pheromones of Saccharomycetes appear not to have diversified under positive selection. In contrast, the α-factor-like pheromone of the Eurotiomycetes showed evidence of having diversified under positive selection. Only the codon encoding the third residue of this nine-residue peptide, was found to have a probability greater than 0.95 of falling into the "positive selection" class of codons with $\omega > 1$. Indeed, this residue is highly variable among the nine Eurotiomycete species (see Fig. 3). Whether the identity of this amino acid affects pheromone function or specificity will require future investigations such as site-directed mutagenesis. To investigate whether the inclusion of homothallic and asexual species could have skewed the results, the Sordariomycete dataset was re-analysed retaining only species that are known to be sexual and heterothallic. This did not yield a significantly different result.

Relationships among tandemly-repeated α-factor-like pheromone sequences

Distance analysis of multiple tandem repeats of the α-factor-like pheromone yielded two conflicting patterns (Fig. 6, 7 and 8). In some species, the multiple tandem repeats were more

closely related in DNA sequence to one another than to repeats from other related species. In contrast, some closely related species, including the members of the *G. fujikuroi* complex and *Saccharomyces sensu stricto* clade, displayed a different pattern in which repeats from several different species grouped together. In these instances there was some positional bias in the clustering, where clusters usually consisted of repeats from a similar region in the precursor protein. For example, repeat 5 of *F. sacchari* clustered with repeat 5 of *F. mangiferae* while repeats 13 and 14 of the same two species clustered together. However, this pattern only appeared in species that shared close relatives within the dataset.

Broken down into synonymous and non-synonymous distances, within-species comparisons among peptide repeats had, on average, a notably lower d_N than between-species comparisons. However d_S values were not vastly different, and some within-species d_S values were greater than many between-species values (Fig. 9). The gradient of the d_N vs d_S plots (Fig. 9), which corresponds to the d_N/d_S ratio, varied widely, but was significantly lower in within-species than between-species values, implying stronger purifying selection acting at the within-species level.

Results of the GATA analysis were compiled (Fig. 10) so that all changes in repeat copy number in the *G. fujikuroi* complex could be considered simultaneously. The results demonstrated that multiple independent changes in repeat copy number have occurred within and among these species. In particular, it appeared that the more central repeats have been more frequently involved in these structural events.

Discussion

This study has substantially increased the number of predicted a- and α-class pheromone precursor genes that have been described, providing new insights into the distribution, sequence diversity and structural organisation of these genes. Some acomycete pheromones are strongly conserved between species but most are highly divergent. The rich diversity in peptide sequence has probably resulted from both adaptive and non-adaptive evolutionary forces. Both pheromone precursor classes can have a modular nature, a characteristic that could contribute to the inter-specific divergence of the pheromone peptides. Much of this variation might also have functional relevance, implying a potential role of pheromones in species-specific mate recognition. The rapid evolution of pheromones could, therefore, contribute to speciation in Ascomycetes.

The α-class precursor was distributed across species from all four filamentous classes (Sordariomycetes, Eurotiomycetes, Dothidiomycetes and Leotiomycetes) as well as both yeast classes (Saccharomycetes and Taphrinomycetes) included in this study. By contrast, the a-class precursor was found in far fewer species, spanning the two yeast classes and a single filamentous class, Sordariomycetes. Several previous studies have been unable to identify the a-class precursor in various genomes (Pöggeler 2002; Dyer et al. 2003; Hoff et al. 2008; Butler et al. 2009). This small gene might have been altered to a point where it is undetectable by BLAST analysis in these fungi. If it is indeed absent, this would require that some other element has assumed the role of the a-factor-like pheromone, at least in heterothallic species. In some homothallic species, only one pheromone-receptor pair might be necessary, while the other could eventually be lost (Mayrhofer et al. 2006; Kim et al. 2008; Lee et al. 2008).

All the pheromone precursor genes that were found in this study appeared to have intact open reading frames. These include those in species for which no sexual stage is known such as *F. mangiferae*, *F. sterilihyphosum* and *F. oxysporum*. While this could be interpreted as evidence that these species might have a cryptic sexual cycle, these genes could be retained in asexual species if they perform additional functions outside of sexual reproduction. For example, work on *N. crassa* pheromone precursor *Mfa-1* has suggested a possible role in "conglutination", the cementing of hyphae which occurs, for example, during sclerotium formation (Kim et al. 2002). In *Candida* spp., the α-factor-like pheromone appears to be necessary for inter-cellular signalling during biofilm formation (Daniels et al. 2006; Sahni et al. 2010). Further work is, therefore, necessary to distinguish between the sexual and non-sexual functions of these pheromones.

This study is the first to scrutinize similarities and differences in the pheromones of a group of closely related filamentous Ascomycetes with varying degrees of sexual compatibility. Most species pairs in the *G. fujikuroi* complex are inter-sterile. However, we have found a lack of significant differences among the putative pheromone peptides in species of this complex. We hypothesise that the repeated eight-residue motif encoded by *ppg2* represents the mature a-factor like pheromone. If so, there is complete identity in this pheromone among all twelve species. In *ppg1* each species carried multiple copies of two or more distinct variants of the pheromone that differed slightly in amino acid sequence. Nevertheless, only one species (*F. konzum*) carried variants that were unique and potentially species-specific. Unless these pheromones undergo differential post-translational modifications (an avenue that has not been explored in ascomycete pheromones), our results imply that these species should recognise one-another as potential mates. A mechanism further downstream must, therefore, be responsible for the observed

reproductive barriers. Indeed, inter-specific mating experiments in *Fusarium* can lead to the formation of barren perithecia (Leslie et al. 2004), suggesting that hyphal fusion could have occurred. Such post-mating barriers are common in the genus *Neurospora*, which led Turner et al. (2010) to propose that pheromone divergence could be prevented by strong selective constraint.

On the whole, strong conservation of pheromones was not the norm; in fact pheromones were significantly more divergent than house-keeping genes in all four datasets considered. These findings are in agreement with the rapid evolution of reproductive proteins observed in the plant and animal kingdoms (as reviewed by Swanson and Vacquier 2002; Clark et al. 2006). This phenomenon has been proposed as a driving force in speciation, particularly when associated with mate-recognition proteins. A comparable example is the rapid evolution of interacting sperm and egg surface proteins in Abalone (*Haliotis rufescens*), which could cause reproductive isolation between populations (Yang et al. 2000b; Galindo et al. 2003). If pheromone peptide sequence is indeed crucial for successful reception, the observed inter-specific differences and rapid evolution of pheromones in this study might have major implications for reproductive isolation and speciation in Ascomycetes. An understanding of the forces that drive pheromone diversification could shed light on the importance of these modifications for mate recognition. Alterations that come about through relaxed selective constraint are probably inconsequential, while those driven by positive selection are likely to have functional relevance.

Our analyses identified evidence for positive diversifying selection acting in the dataset of α-factor-like pheromones from Eurotiomycetes, but not in the larger datasets of the

Sordariomycetes and Saccharomycetes. This implies that the variation among species in the latter groups might have accumulated over time through relaxed constraint. However, the test performed to detect positive selection in this study addresses each dataset as a whole and is most sensitive when positive selection acts in all lineages. Yang and Nielsen (2002) demonstrated that selection acting only along certain lineages of the phylogeny might go undetected by the "site models". It is plausible that pheromone peptides follow this sort of evolutionary pattern given the observation that the pheromones of some species pairs are highly divergent, while those of a few remain even more conserved than house-keeping genes. The ability of Ascomycetes to reproduce efficiently through asexual means could further complicate the issue. In periods of preferentially asexual reproduction, sex-related genes become obsolete and may diverge under relaxed selective constraint. Therefore the detection of relaxed constraint cannot necessarily be construed as evidence that peptide sequence is not important for successful pheromone reception.

In the α-class precursors described here and also in those previously described, amino-acid sequences of the multiple tandem repeats of the mature peptide were usually identical or nearly-identical. We found that this similarity was largely reflected at the DNA level, in that tandem repeats were often more closely related to one another in DNA sequence than to repeats from any other species. Such a pattern has been interpreted as the signature of concerted evolution, (e.g., Swanson and Vacquier 1998). However, the repeats of more closely-related species did not form independent clusters. For example, all repeats in each *G. fujikuroi* complex species had their closest relative in another species of this complex. This pattern has been termed trans-specific-polymorphism (Klein 1987), and occurs when duplication predates speciation. Thus, paralogues of a certain sequence are more similar to their corresponding copy in a related species than to the

sister paralog in the same genome. The inclusion of many closely-related species in this study has therefore demonstrated that concerted evolution in the conventional sense, which is thought to rapidly homogenise DNA sequences among tandem repeats (Elder and Turner 1995), appears not to occur in the α-class precursor.

The trans-specific polymorphism was absent in all the species that did not have close relatives in our dataset, indicating that it degrades over larger evolutionary distances. Rather than concerted evolution, this pattern is probably an inevitable result of continuous duplication and deletion of repeats, as evidenced by the huge variation in repeat copy number. Evolution through duplication and deletion has been termed "birth-and-death" evolution and is common in tandem repeat sequences (Nei and Rooney 2005). Characteristic examples of genes experiencing birth-and-death are the major histocompatibility complex (MHC) and imunoglobin (Ig) genes of mammals (Nei et al. 1997) as well as the ribosomal RNA genes in fungi (Rooney and Ward 2005). Concerted evolution has been rejected in these cases because, while some members of the multigene families are highly similar, others have their closest relatives in distantly related species. Fig. 11 illustrates how continuous duplications and deletions acting on a set of repeats could create such a phylogenetic scenario, starting with complete trans-specific polymorphism and progressing, eventually, to exclusive monophyletic clustering.

Birth-and death is not the only mechanism acting to maintain similarity among peptide repeats. We have shown that while repeats from the same species always have low non-synonymous distances, some have high synonymous distances, and that d_N/d_S is significantly lower at the within-species level.. This indicates that not all repeats are similar only due to recent duplication,

but that the sequences of the encoded peptides are also conserved by purifying selection (Ina 1996). We, therefore, propose that the similarity observed among repeats of the mature peptide is adaptive, and reflects a requirement for sequence specificity in pheromone reception. Differences between species could therefore indeed be associated with functional divergence.

The unique structure and evolutionary system of the α-class precursor could further contribute to the rapid evolution of the α-factor-like pheromones. Random events of duplication and loss could cause the "fixation" of a certain pheromone variant in one species and the simultaneous fixation of a different variant in another species. This could be facilitated by positive selection, but might even occur by chance in the absence of selection (Nei and Rooney 2005). The gradual process of succession by which a new repeat variant can spread through the precursor, replacing the old one, might also facilitate the co-evolution of pheromones and receptors. A new compatible pheromone/receptor pair could spread in a population without a transition through an "adaptive valley". Individuals carrying repeats of both the new and the wild-type pheromone maintain the ability to be recognized by the "wild-type" receptor, and would therefore not suffer a dramatic fitness cost. Lastly, the ability to harbor multiple repeat variants could increase the capacity to evolve by providing a constant source of variation, the raw material for natural selection.

Supplementary Material

Information on the *Fusarium* isolated used in this study is provided as supplementary material. Also provided are all nucleotide sequences of pheromone precursor genes used in this study,

along with translation, GenBank accession numbers and references. The supplementary material is presented in a single HTML, which is available online at http://mbe.oxfordjournals.org.

Acknowledgements

This work was funded by The National Research Foundation (NRF), the DST/NRF Centre of Excellence in Tree Health Biotechnology (CTHB), the University of Pretoria and the Tree Protection Cooperative Programme (TPCP).

Literature Cited

Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. 1990. Basic local alignment search tool. J Mol Biol **215**: 403-410.

Bender, A. & Sprague-Jr, G. F. 1989. Pheromones and pheromone receptors are the primary determinants of mating specificity in the yeast *Saccharomyces* cerevisiae. Genetics **121**: 463-476.

Betz, R., Crabb, J. W., Meyer, H. E., Wittig, R. & Duntze, W. 1987. Amino acid sequences of a-factor mating peptides from *Saccharomyces cerevisiae*. J Biol Chem **262**: 546-548.

Betz, R., MacKay, V. L. & Duntze, W. 1977. a-Factor from *Saccharomyces cerevisiae*: partial characterization of a mating hormone produced by cells of mating type a. J Bacteriol **132**: 462-472.

Bobrowicz, P., Pawlak, R., Correa, A., Bell-Pedersen, D. & Ebbole, D. J. 2002. The *Neurospora crassa* pheromone precursor genes are regulated by the mating type locus and the circadian clock. Mol Microbiol **45**: 795-804.

Brake, A. J., Brenner, C., Najarian, R., Laybourn, P. & Merrywheather, J. (1985) Structure of genes encoding precursors of the yeast peptide mating pheromone a-factor. In: Gething, M. J., editor. Protein Transport and Secretion. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

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.

Brown, A. J. & Casselton, L. A. 2001. Mating in mushrooms: increasing the chances but prolonging the affair. Trends Genet **17**: 393-400.

Burke, D., Mendonça-Previato, L. & Ballou, C. E. 1980. Cell-cell recognition in yeast: purification of *Hansenula wingei* 21-cell sexual agglutination factor and comparison of the factors from three genera. P Natl Acad Sci USA **77**: 318-322.

Butler, G., Rasmussen, M. D., Lin, M. F. et al. (51 co-authors). Evolution of pathogenicity and sexual reproduction in eight *Candida* genomes. Nature **459**: 657-662.

Caplan, S. & Kurjan, J. 1991. Role of α-factor and the MFα1 α-factor precursor in mating in yeast. Genetics **127**: 299-307.

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 R **61**: 411-428.

Coppin, E., de Renty, C. & Debuchy, R. 2005. The function of the coding sequences for the putative pheromone precursors in *Podospora anserina* is restricted to fertilization. Eukaryot Cell **4**: 407-420.

Coyne, J. A. & Orr, H. A. 2004. *Speciation*. Sinauer, Sunderland, Massachusetts.

Daniels, K. J., Srikantha, T., Lockhart, S. R., Pujol, C. & Soll, D. R. 2006. Opaque cells signal white cells to form biofilms in *Candida albicans*. EMBO J **25**: 2240-2252.

Darby, N. J. & Smyth, D. G. 1990. Endopeptidases and prohormone processing. Bioscience Rep **10**: 1-13.

Davey, J. 1992. Mating pheromones of the fission yeast *Schizosaccharomyces pombe*: purification and structural characterization of M-factor and isolation and analysis of two genes encoding the pheromone. EMBO J **11**: 951-960.

Desjardins, A. E., Plattner, R. D. & Gordon, T. R. 2000. *Gibberella fujikuroi* mating population A and *Fusarium subglutinans* from teosinte species and maize from Mexico and Central America. Mycol Res **104**: 865-872.

Dyer, P. S., Paoletti, M. & Archer, D. B. 2003. Genomics reveals sexual secrets of *Aspergillus*. Microbiology **149**: 2301-2303.

Dyer, P. S., Ingram, D. S. & Johnstone, K. J. 1992. The control of sexual morphogenesis in the Ascomycotina. Biol Rev **67**: 421-458.

Elder, J. F., Jr. & Turner, B. J. 1995. Concerted evolution of repetitive DNA sequences in eukaryotes. Q Rev of Biol **70**: 297-320.

Fowler, T. J., Mitton, M. F., Vaillancourt, L. J. & Raper, C. A. 2001. Changes in mate recognition through alterations of pheromones and receptors in the multisexual mushroom fungus *Schizophyllum commune*. Genetics **158**: 1491-1503.

Galindo, B. E., Vacquier, V. D. & Swanson, W. J. 2003. Positive selection in the egg receptor for abalone sperm lysin. P Natl Acad Sci USA **100**: 4639-4643.

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

Hall, T. A. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for windows 95/98/NT. Nucleic Acids Symp Ser **41**: 95 - 98.

Hibbett, D. S., Binder, M., Bischoff, J. F. et al. (67 co-authors). 2007. A higher-level phylogenetic classification of the Fungi. Mycol Res **111**: 509-547.

Hisatomi, T., Yanagishima, N., Sakurai, A. & Kobayashi, H. 1988. Interspecific actions of α mating pheromones on the a mating-type cells of three *Saccharomyces* yeasts. Curr Genet **13**: 25-27.

Hoff, B., Pöggeler, S. & Kuck, U. 2008. Eighty years after its discovery, Fleming's *Penicillium* strain discloses the secret of its sex. Eukaryot Cell **7**: 465-470.

Howard, D. J. (1993) Reinforcement: origin, dynamics and fate of an evolutionary hypothesis. In: Harrison, R. G., editor. Hybrid Zones and the Evolutionary Process. Oxford University Press, England.

Ina, Y. 1996. Pattern of synonymous and nonsynonymous substitutions: An indicator of mechanisms of molecular evolution. J Genet **75**: 91-115.

Julius, D., Blair, L., Brake, A., Sprague, G. & Thorner, J. 1983. Yeast α factor is processed from a larger precursor polypeptide: The essential role of a membrane-bound dipeptidyl aminopeptidase. Cell **32**: 839-852.

Julius, D., Schekman, R. & Thorner, J. 1984. Glycosylation and processing of prepro-α-factor through the yeast secretory pathway. Cell **36**: 309-318.

Karlsson, M., Nygren, K. & Johannesson, H. 2008. The evolution of the pheromonal signal system and its potential role for reproductive isolation in heterothallic *Neurospora*. Mol Biol Evol **25**: 168-178.

Kim, H.-K., Lee, T. & Yun, S.-H. 2008. A putative pheromone signaling pathway is dispensable for self-fertility in the homothallic ascomycete *Gibberella zeae*. Fungal Genet Biol **45**: 1188-1196.

Kim, H. & Borkovich, K. A. 2004. A pheromone receptor gene, *pre-1*, is essential for mating type-specific directional growth and fusion of trichogynes and female fertility in *Neurospora crassa*. Mol Microbiol **52**: 1781-1798.

Kim, H. & Borkovich, K. A. 2006. Pheromones are essential for male fertility and sufficient to direct chemotropic polarized growth of trichogynes during mating in *Neurospora crassa*. Eukaryot Cell **5**: 544-554.

Kim, H., Metzenberg, R. L. & Nelson, M. A. 2002. Multiple functions of *mfa-1*, a putative pheromone precursor gene of *Neurospora crassa*. Eukaryot Cell **1**: 987-999.

Klein, J. 1987. Origin of major histocompatibility complex polymorphism: The trans-species hypothesis. Hum Immunol **19**: 155-162.

Kurjan, J. 1993. The pheromone response pathway in *Saccharomyces cerevisiae*. Annu Rev of Genet **27**: 147-179.

Kurjan, J. & Herskowitz, I. 1982. Structure of a yeast pheromone gene (MFα): A putative α-factor precursor contains four tandem copies of mature α-factor. Cell **30**: 933-943.

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., Leslie, J. F. & Bowden, R. L. 2008. Expression and function of sex pheromones and receptors in the homothallic ascomycete *Gibberella zeae*. Eukaryot Cell **7**: 1211-1221.

Leslie, J., Zeller, K., Wohler, M. & Summerell, B. 2004. Interfertility of Two Mating Populations in the Gibberella Fujikuroi Species Complex. Eur J Plant Pathol **110**: 611-618.

Leslie, J. F. & Summerall, B. A. 2006. The *Fusarium* laboratory manual. Blackwell Professional, Ames, Iowa.

Mann, H. B. & Whitney, D. R. 1947. On a Test of Whether one of Two Random Variables is Stochastically Larger than the Other. Ann Math Stat **18**: 50-60.

Mayrhofer, S. & Pöggeler, S. 2005. Functional characterization of an α-factor-like *Sordaria macrospora* peptide pheromone and analysis of its interaction with its cognate receptor in *Saccharomyces cerevisiae*. Eukaryot Cell **4**: 661-672. Mayrhofer, S., J. M. Weber, and S. Pöggeler. 2006. Pheromones and pheromone receptors are required for proper sexual development in the homothallic ascomycete *Sordaria macrospora*. Genetics **172**:1521-1533.

McCullough, J. & Herskowitz, I. 1979. Mating pheromones of *Saccharomyces kluyveri*: pheromone interactions between *Saccharomyces kluyveri* and *Saccharomyces cerevisiae*. J Bacteriol **138**: 146-154.

Naider, F. & Becker, J. M. 2004. The α-factor mating pheromone of *Saccharomyces cerevisiae*: a model for studying the interaction of peptide hormones and G protein-coupled receptors. Peptides **25**: 1441-1463.

Nei, M. & Gojobori, T. 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol Biol Evol **3**: 418-426.

Nei, M., Gu, X. & Sitnikova, T. 1997. Evolution by the birth-and-death process in multigene families of the vertebrate immune-system. P Natl Acad Sci USA **94**: 7799-7806.

Nei, M. & Kumar, S. 2000. Molecular evolution and phylogenetics. Oxford University Press, New York, New York.

Nei, M. & Rooney, A. P. 2005. Concerted and birth-and-death evolution of multigene families. Annu Rev of Genet **39**: 121-152.

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.

Olesnicky, N. S., Brown, A. J., Honda, Y., Dyos, S. L., Dowell, S. J. & Casselton, L. A. 2000. Self-compatible B mutants in *coprinus* with altered pheromone-receptor specificities. Genetics **156**: 1025-1033.

Palumbi, S. R. 1999. All males are not created equal: Fertility differences depend on gamete recognition polymorphisms in sea urchins. P Nat Acad Sci USA **96**: 12632-12637.

Palumbi, S. R. 2008. Speciation and the evolution of gamete recognition genes: pattern and process. Heredity **102**: 66-76.

Pöggeler, S. 2000. Two pheromone precursor genes are transcriptionally expressed in the homothallic ascomycete *Sordaria macrospora*. Curr Genet **37**: 403-411.

Pöggeler, S. 2002. Genomic evidence for mating abilities in the asexual pathogen *Aspergillus fumigatus*. Curr Genet **42**: 153-160.

Rooney, A. P. & Ward, T. J. 2005. Evolution of a large ribosomal RNA multigene family in filamentous fungi: Birth and death of a concerted evolution paradigm. P Nat Acad Sci USA **102**: 5084-5089.

Sahni, N., Yi, S., Daniels, K. J., Huang, G., Srikantha, T. & Soll, D. R. 2010. Tec1 mediates the pheromone response of the white phenotype of *Candida albicans*: insights into the evolution of new signal transduction pathways. PLoS Biol **8**.

Shapiro, S. S. & Wilk, M. B. 1965. An Analysis of Variance Test for Normality (Complete Samples). Biometrika **52**: 591-611.

Singh, A., Chen, E. Y., Lugovoy, J. M., Chang, C. N., Hitzeman, R. A. & Seeburg, P. H. 1983. *Saccharomyces cerevisiae* contains two discrete genes coding for the α-factor pheromone. Nucleic Acids Res **11**: 4049-4063.

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., Desjardins, A. E., Marasas, W. F. O. & Wingfield, M. J. 2002. Cryptic speciation in Fusarium subglutinans. Mycologia **94**: 1032-1043.

Stötzler, D. & Duntze, W. 1976. Isolation and characterization of four related peptides exhibiting alpha factor activity from *Saccharomyces cerevisiae*. Eur J Biochem **65**: 257-262.

Swanson, W. J. & Vacquier, V. D. 2002. The rapid evolution of reproductive proteins. Nature R Genet **3**: 137-144.

Swanson, W. J. & Vacquier, V. D. 1998. Concerted evolution in an egg receptor for a rapidly evolving abalone sperm protein. Science **281**: 710-712.

Tamura, K., Dudley, J., Nei, M. & Kumar, S. 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) Software Version 4.0. Mol Biol Evol **24**: 1596-1599.

Tatusova, T. A. & Madden, T. L. 1999. BLAST 2 Sequences, a new tool for comparing protein and nucleotide sequences. FEMS Microbiol Lett **174**: 247-50.

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. Nucl. Acids Res. **22**: 4673-4680.

Turgeon, B. G. 1998. Application of mating type gene technology to problems in fungal biology. Ann R Phytopathol **36**: 115 - 137.

Turner, E., Jacobson, D. J. & Taylor, J. W. 2010. Reinforced postmating reproductive isolation barriers in *Neurospora*, an Ascomycete microfungus. J Evol Biol **23**: 1642-1656.

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.

Wilcoxon, F. 1945. Individual comparisons by ranking methods. Biometrics **1**: 80-83.

Wilkinson, L. 1987. MYSTAT: A Personal Version of SYSTAT. Am Stat **41**: 334.

Yang, Z. 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. Comput Appl Biosci **13**: 555 - 556.

Yang, Z. & Nielsen, R. 2002. Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. Mol Biol and Evol **19**: 908-917.

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

Yang, Z. H., Swanson, W. J. & Vacquier, V. D. 2000b. Maximum-likelihood analysis of molecular adaptation in abalone sperm lysin reveals variable selective pressures among lineages and sites. Mol Biol Evol **17**: 1446 - 1455.

Zhang, L., Baasiri, R. A. & Van Alfen, N. K. 1998. Viral repression of fungal pheromone precursor gene expression. Mol Cell Biol **18**: 953-959.

Table 1. Results of tests for positive selection

Dataset	Model [a]	Parameter Estimates [b]	lnL [c]	2δ[d]
Saccharomycetes α	M7	p=0.839 q=2.764	-566.745	0.046
	M8	p_0=0.964 p=0.906 q=3.339 ω=1.000	-566.722	
Saccharomycetes a	M7	p=0.371 q=3.837	-368.243	0.000
	M8	p_0=1.000 p=0.372 q=3.837 ω=1.000	-368.243	
Sordariomycetes α	M7	p=0.332 q=22.689	-273.156	0.000
	M8	p_0=0.999 p=0.332 q=22.689 ω=1.000	-273.156	
Eurotiomycetes α	M7	p=0.164 q=1.281	-148.346	6.230*
	M8	p_0=0.933 p=0.394 q=11.128 ω=9.923	-145.231	

[a] "Site" models in CODEML (Yang *et al.*, 2000a) M7, the beta model and M8 the beta&ω model

[b] Parameter estimates generated by CODEML

[c] ln likelihood score for each model, as calculated by CODEML

[d] Likelihood Ratio Test (LRT) statistic, δ is the differences between the ln likelihood values

[*] Indicates a significant difference at a value of $P \leq 0.05$

Fig. 1. Predicted *ppg1* and *ppg2* genes from *F. verticillioides*.

(GenBank accession number AAIM02000073). Clear block arrows indicate coding regions. Grey block arrows indicate repeats of the putative mature α-factor-like peptide. Black arrows indicate primer binding sites and direction.

Fig. 2. a-class pheromone precursor proteins and predicted mature peptides.

See supplementary material for complete DNA sequences, GenBank accession numbers and references. CaaX and CpaX motifs are shaded black. Mature a-factor peptides from yeasts as well as hypothetical mature peptides from *G. fujikuroi* species are shaded grey. Multiple genes encoded in the same genome are indicated by numbers in brackets. The default amino acid coloration of BioEdit (PAM 250) (Hall, 1999) is used.

Fig. 3. α-class precursor protein structures and mature pheromone sequences from filamentous Ascomycetes.

See supplementary material for complete DNA sequences, GenBank accession numbers and references. Lines with blocks schematically represent the number of repeats in each precursor protein. Repeat colors indicate different amino acid sequence "variants" of the pheromone within a single species. For each species, the sequence of all putative pheromone repeats are displayed on the right, with dots indicating residues that are identical to those in the first sequence. The default amino acid coloration of BioEdit (PAM 250) (Hall, 1999) is used.

Fig. 4. α-class precursor protein structures and mature pheromone sequences from yeasts.

See supplementary material for complete DNA sequences, GenBank accession numbers and references. Lines with blocks schematically represent the number of repeats in each precursor protein. Multiple genes present within the same genome are displayed beneath one another and numbered. Repeat colors indicate different amino acid sequence "variants" of the pheromone within a single species. For each species, the sequence of all putative pheromone repeats are displayed on the right, with dots indicating residues that are identical to those in the first sequence. The default amino acid coloration of BioEdit (PAM 250) (Hall, 1999) is used.

Fig. 5. Dot plots of pairwise d_N/d_S values for pheromones versus house-keeping genes.
Each point represents a pairwise d_N/d_S value between a pair of species. Datasets are displayed independently, with values for house-keeping (HK) genes on the left and pheromones on the right. "*P*" values represent significance values from the Wilcoxon signed ranks test. $P \leq 0.05$ indicates a significant difference between the two samples. Sample sizes for the four datasets are 156, 15, 145, and 55 species pairs, respectively.

Fig. 6. Neighbour joining distance trees of DNA sequences encoding mature α-factor-like peptide repeats from the Sacharomycetes.
Colors indicate repeats from the same species. Some Saccharomycetes carry multiple α-class precursor genes, therefore taxon names are followed by two numbers. The first number represents the gene, the second represents position of the particular repeat within the precursor, starting at the N-terminal. Bootstrap values greater than 50 are displayed on branches.

Fig. 7. Neighbour joining distance trees of DNA sequences encoding mature α-factor-like peptide repeats from the Sordariomycetes (excluding *Fusarium*).

Colors indicate repeats from the same species. The number after each taxon names represent the position of the particular repeat within the precursor, starting at the N-terminal. Bootstrap values greater than 50 are displayed on branches.

Fig. 8. Neighbour joining distance trees of DNA sequences encoding mature α-factor-like peptide repeats from *Fusarium* spp.

Colors indicate repeats from the same species. The number after each taxon names represent the position of the particular repeat within the precursor, starting at the N-terminal. Bootstrap values greater than 50 are displayed on branches.

Fig. 9. Scatter plots of average pairwise d_N values for α-factor-like pheromones plotted against the corresponding d_S values.

Each "between-species" point represents the averaged d_N and d_S values for pair-wise comparisons between all possible pairs of repeats for a single pair of species. Each dark "within-specie" point represents averaged d_N and d_S values for pair-wise comparisons between all possible pairs of tandem repeats from within the same precursor gene. Between-species points outnumber within-species points because the former represent all possible species pairs while the latter each represent a single species. Shaded areas represent one standard deviation above and below the mean for the gradient (m) which corresponds to d_N/d_S. "*P*" values represent significance values for the difference in gradient between the samples from the Mann-Whitney *U* test, with $P \leq 0.05$ taken as significant.

Fig. 10. GATA alignments of *ppg1* among species in the *Giberella fujikuroi* species complex.

Gray boxes indicate positions of the predicted α-factor-like mature peptides (see Fig. 3 for amino acid sequences). Lines connect portions of sequence with high homology (lower cut-off bit-score of 24.9) and the darker the shading the stronger the homology. The accompanying reference phylogeny represents a consensus that was inferred from the results of previous studies (O'Donnell et al., 1998; Kvas et al., 2009).

Fig. 11. Theoretical model showing birth and death evolution of tandem pheromone repeats.

Blocks indicate repeats of the mature α-factor-like peptide. Fill patterns indicate distinct amino acid sequence variants. A constant rate of evolution is assumed, with one substitution (indicated by a small dot) occurring at each step. Accompanying phylogenies represent relationships among underlying DNA sequences at each step.

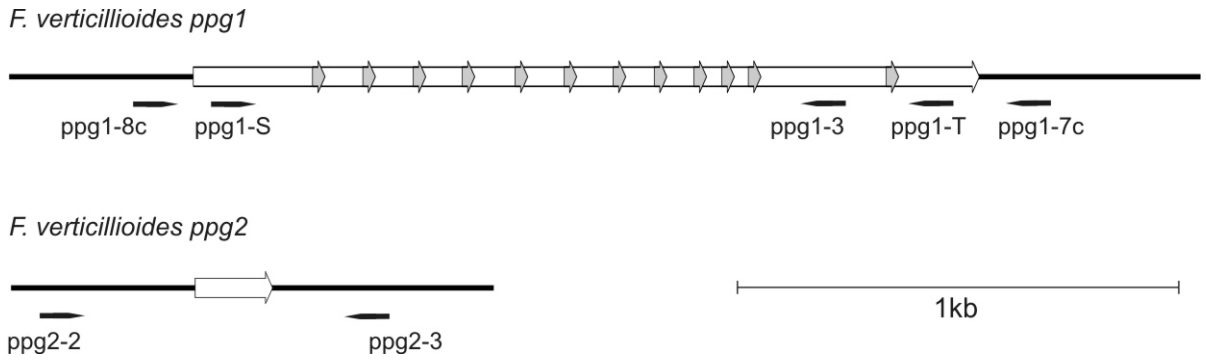

Fig. 1. Predicted *ppg1* and *ppg2* genes from *F. verticillioides*.

(GenBank accession number AAIM02000073). Clear block arrows indicate coding regions. Grey block arrows indicate repeats of the putative mature á -factor-like peptide. Black arr ows indicate primer binding sites and direction.

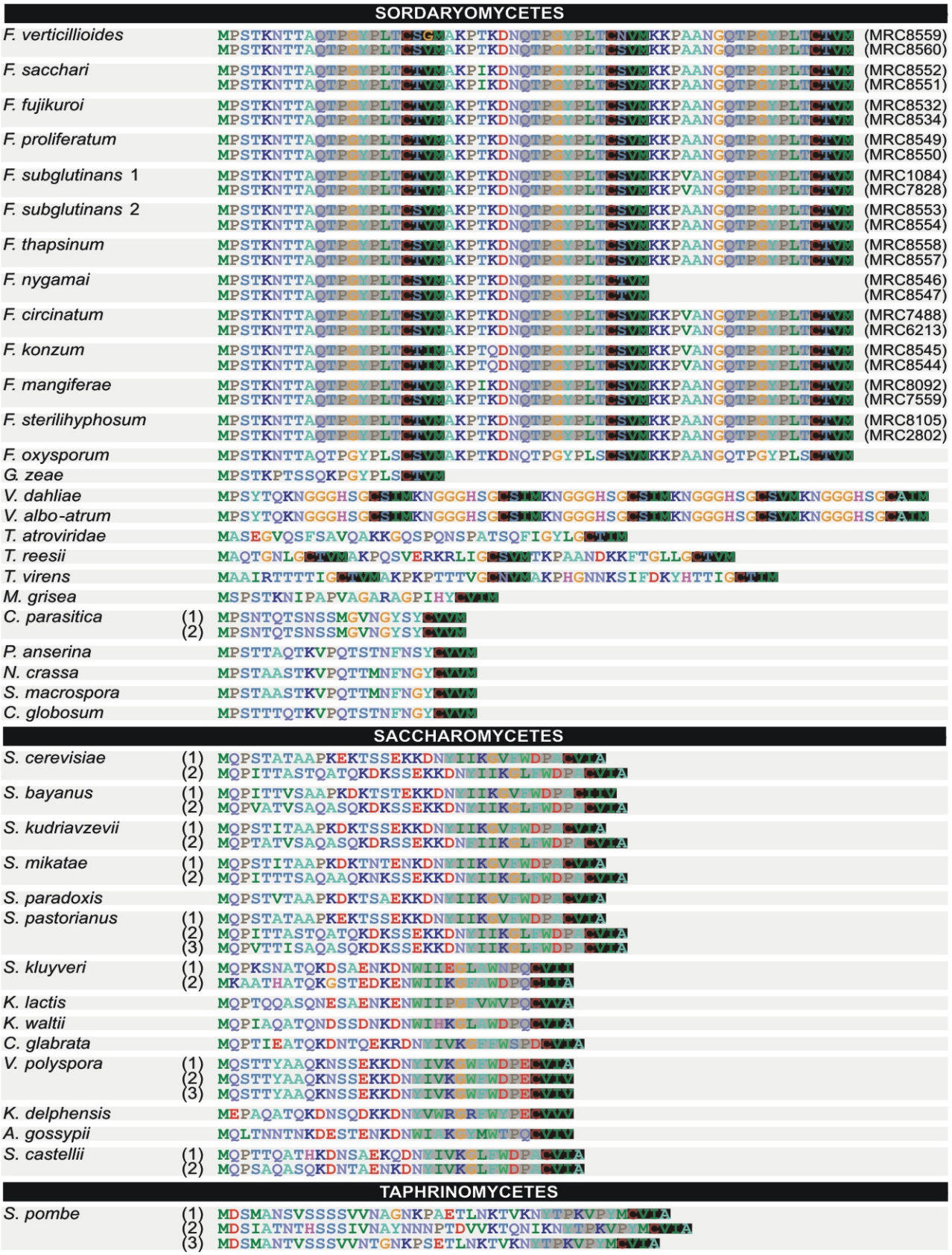

Fig. 2. a-class pheromone precursor proteins and predicted mature peptides.

See supplementary material for complete DNA sequences, GenBank accession numbers and references. CaaX and CpaX motifs are shaded black. Mature a-factor peptides from yeasts as well as hypothetical mature peptides from *G. fujikuroi* species are shaded grey. Multiple genes encoded in the same genome are indicated by numbers in brackets. The default amino acid coloration of BioEdit (PAM 250) (Hall, 1999) is used.

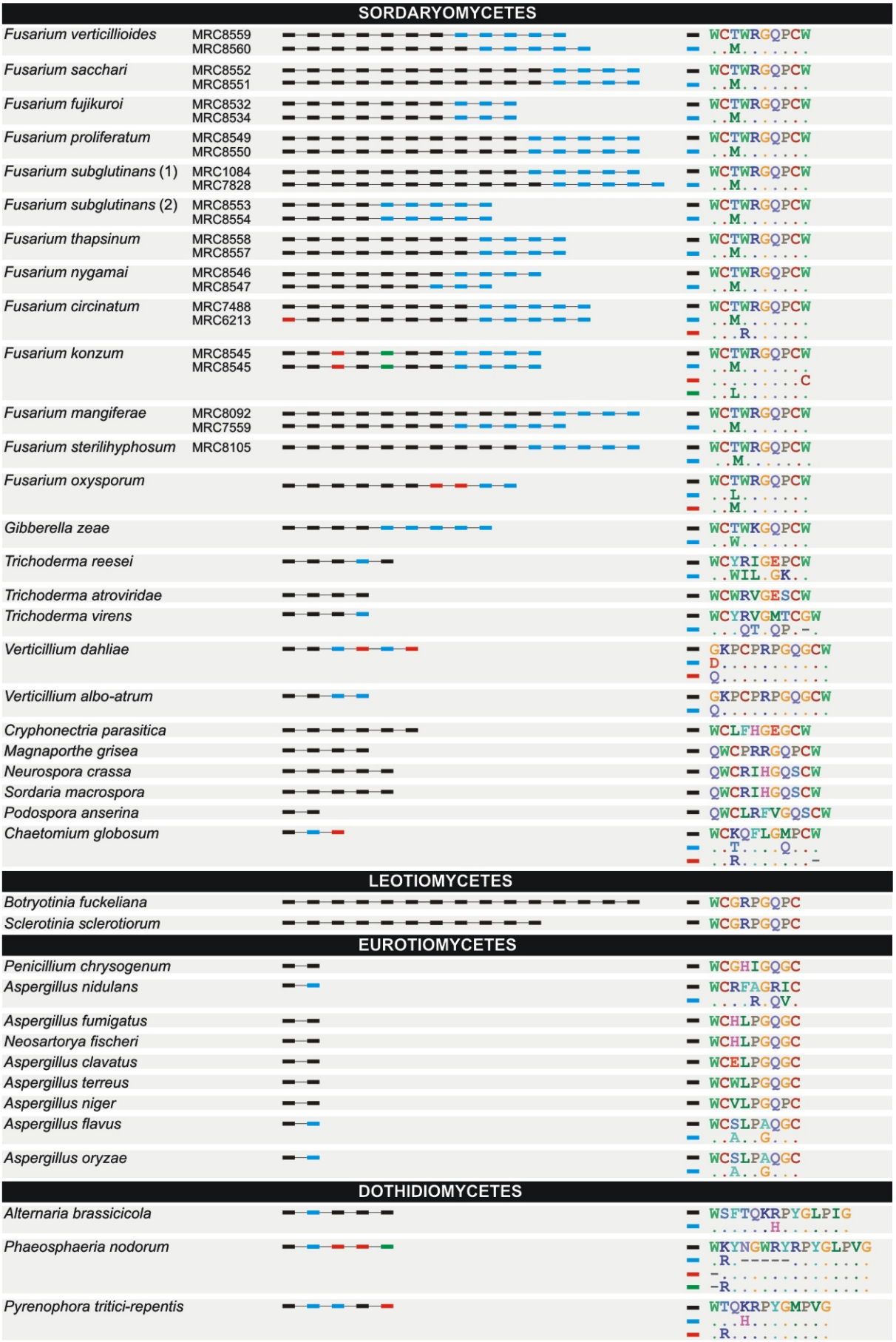

Fig. 3. α-class precursor protein structures and mature pheromone sequences from filamentous Ascomycetes.

See supplementary material for complete DNA sequences, GenBank accession numbers and references. Lines with blocks schematically represent the number of repeats in each precursor protein. Repeat colors indicate different amino acid sequence "variants" of the pheromone within a single species. For each species, the sequence of all putative pheromone repeats are displayed on the right, with dots indicating residues that are identical to those in the first sequence. The default amino acid coloration of BioEdit (PAM 250) (Hall, 1999) is used.

Fig. 4. α-class precursor protein structures and mature pheromone sequences from yeasts. See supplementary material for complete DNA sequences, GenBank accession numbers and references. Lines with blocks schematically represent the number of repeats in each precursor protein. Multiple genes present within the same genome are displayed beneath one another and numbered. Repeat colors indicate different amino acid sequence "variants" of the pheromone within a single species. For each species, the sequence of all putative pheromone repeats are displayed on the right, with dots indicating residues that are identical to those in the first sequence. The default amino acid coloration of BioEdit (PAM 250) (Hall, 1999) is used.

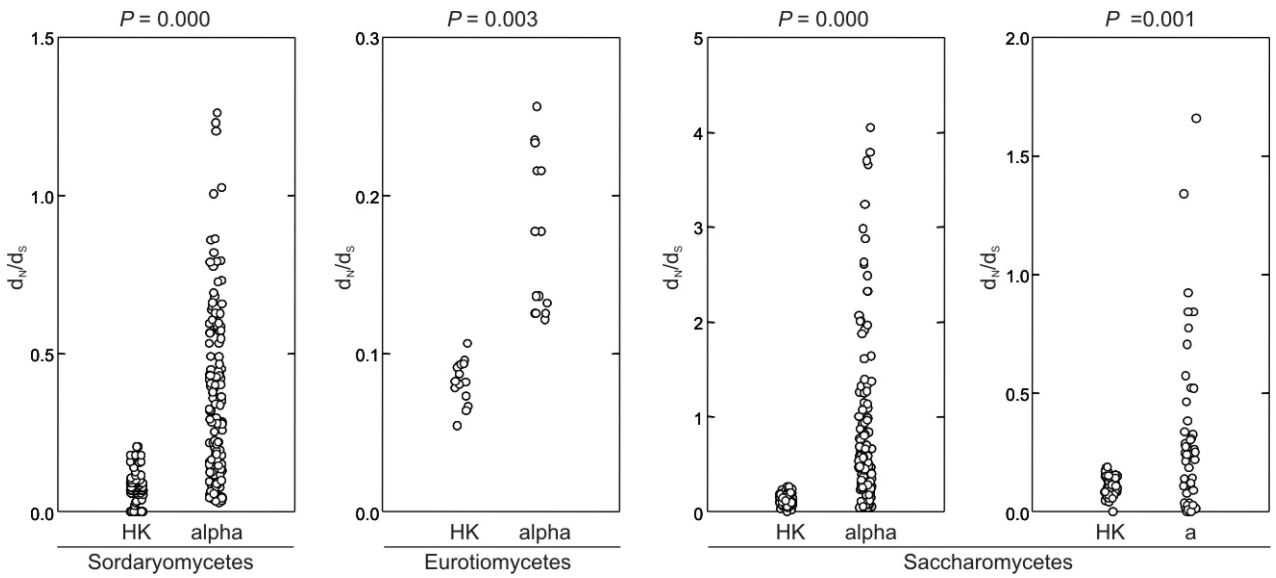

Fig. 5. Dot plots of pairwise d_N/d_S values for pheromones versus house-keeping genes.

Each point represents a pairwise d_N/d_S value between a pair of species. Datasets are displayed independently, with values for house-keeping (HK) genes on the left and pheromones on the right. "*P*" values represent significance values from the Wilcoxon signed ranks test. *P* ? 0.05 indicates a significant difference between the two samples. Sample sizes for the four datasets are 156, 15, 145, and 55 species pairs, respectively.

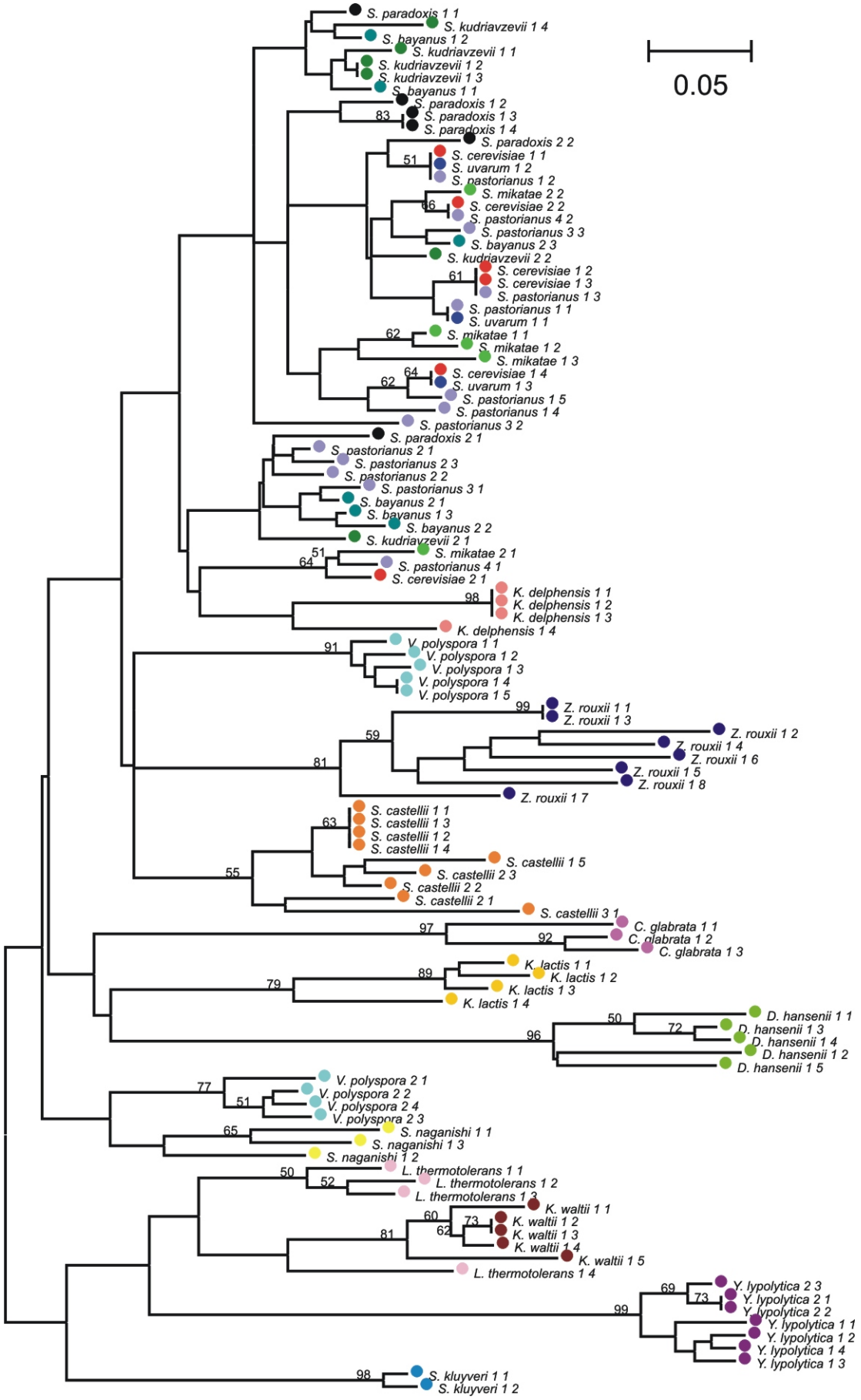

Fig. 6. Neighbour joining distance trees of DNA sequences encoding mature α-factor-like peptide repeats from the Sacharomycetes.

Colors indicate repeats from the same species. Some Saccharomycetes carry multiple α-class precursor genes, therefore taxon names are followed by two numbers. The first number represents the gene, the second represents position of the particular repeat within the precursor, starting at the N-terminal. Bootstrap values greater than 50 are displayed on branches.

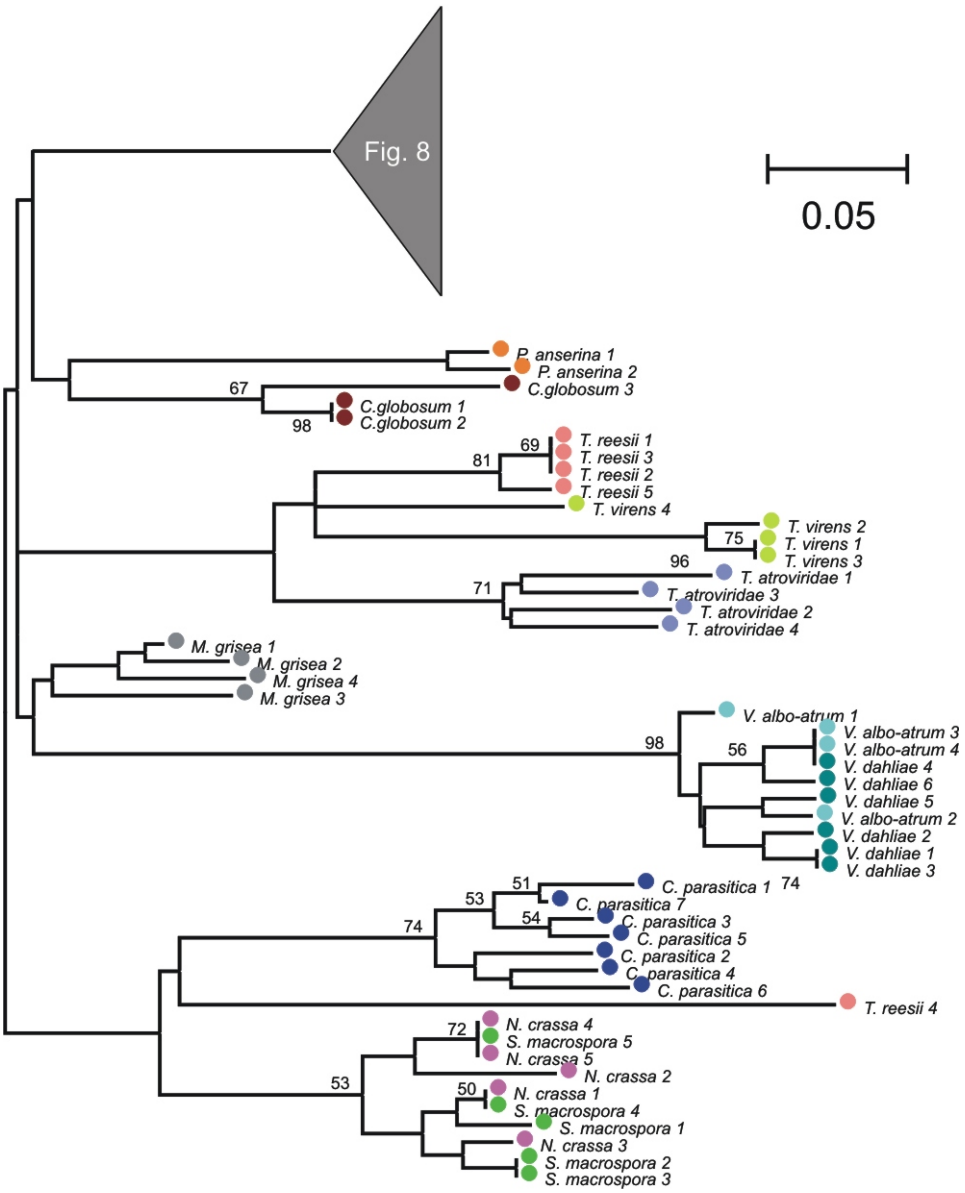

Fig. 7. Neighbour joining distance trees of DNA sequences encoding mature α-factor-like peptide repeats from the Sordariomycetes (excluding *Fusarium*).

Colors indicate repeats from the same species. The number after each taxon names represent the position of the particular repeat within the precursor, starting at the N-terminal. Bootstrap values greater than 50 are displayed on branches.

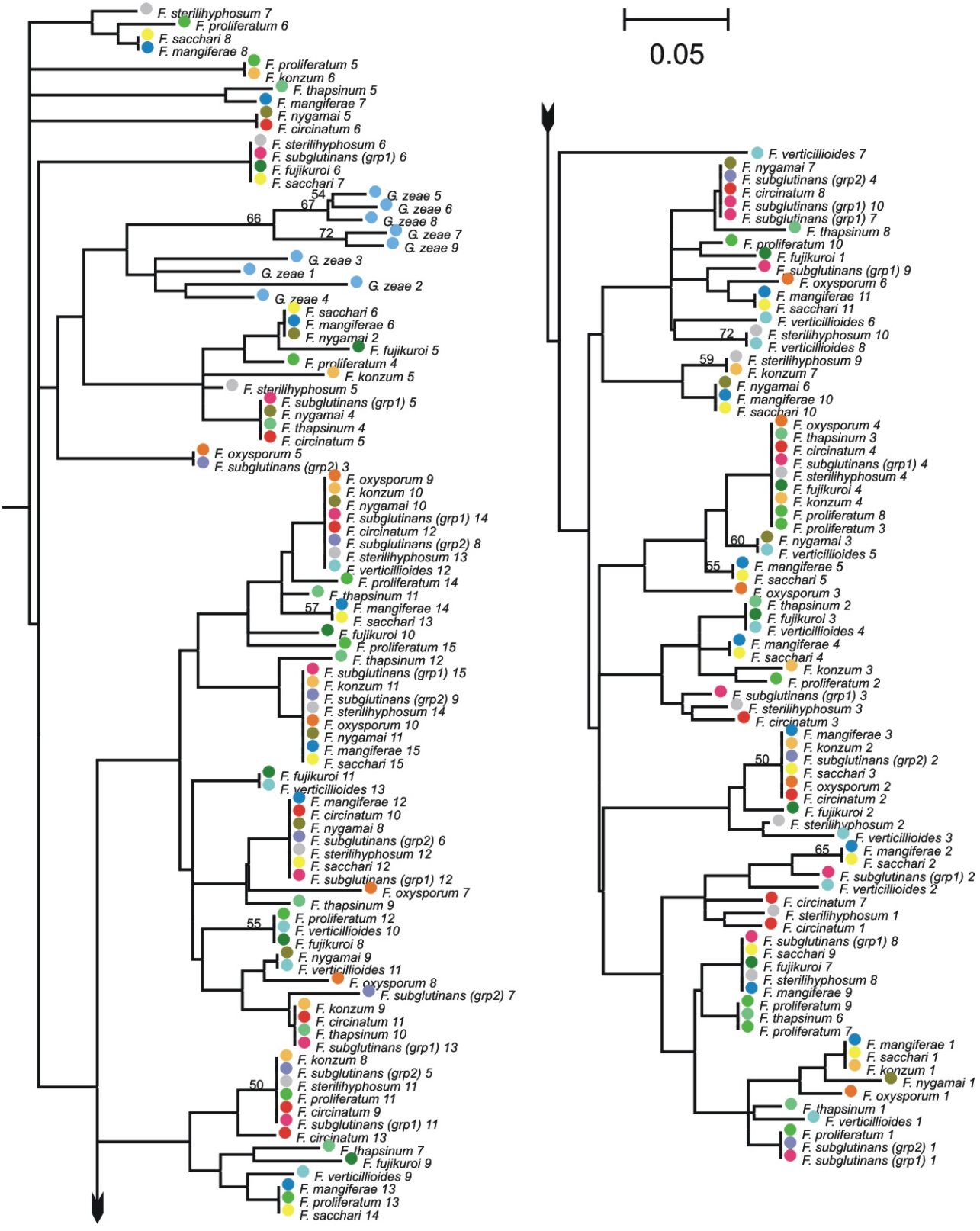

Fig. 8. Neighbour joining distance trees of DNA sequences encoding mature α-factor-like peptide repeats from *Fusarium* spp.

Colors indicate repeats from the same species. The number after each taxon names represent the position of the particular repeat within the precursor, starting at the N-terminal. Bootstrap values greater than 50 are displayed on branches.

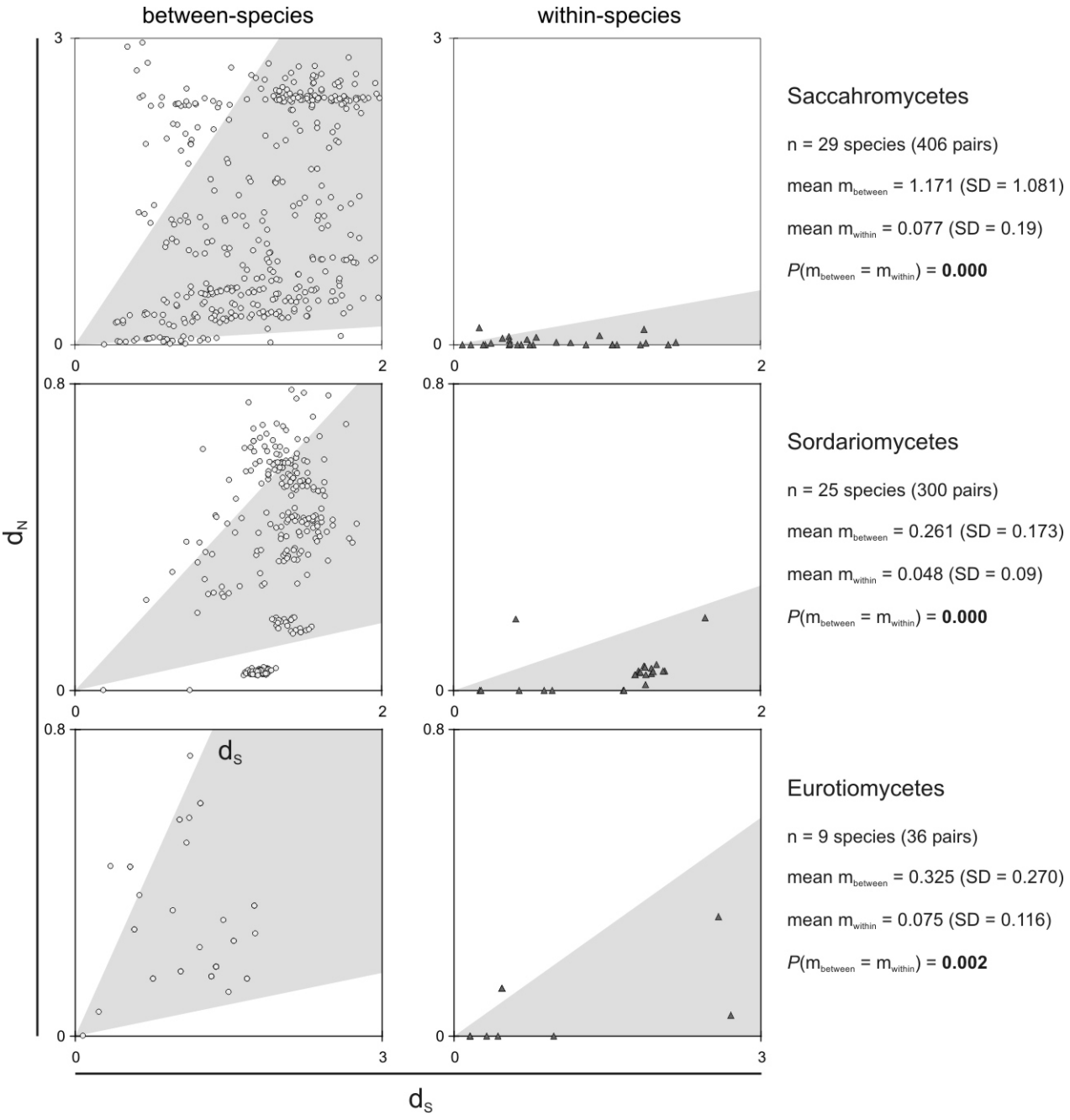

Fig. 9. Scatter plots of average pairwise d_N values for α-factor-like pheromones plotted against the corresponding d_S values.

Each "between-species" point represents the averaged d_N and d_S values for pair-wise comparisons between all possible pairs of repeats for a single pair of species. Each dark "within-specie" point represents averaged d_N and d_S values for pair-wise comparisons between all possible pairs of tandem repeats from within the same precursor gene. Between-species points outnumber within-species points because the former represent all possible species pairs while the latter each represent a single species. Shaded areas represent one standard deviation above and below the mean for the gradient (m) which corresponds to d_N/d_S. "*P*" values represent significance values for the difference in gradient between the samples from the Mann-Whitney *U* test, with *P* ? 0.05 taken as significant.

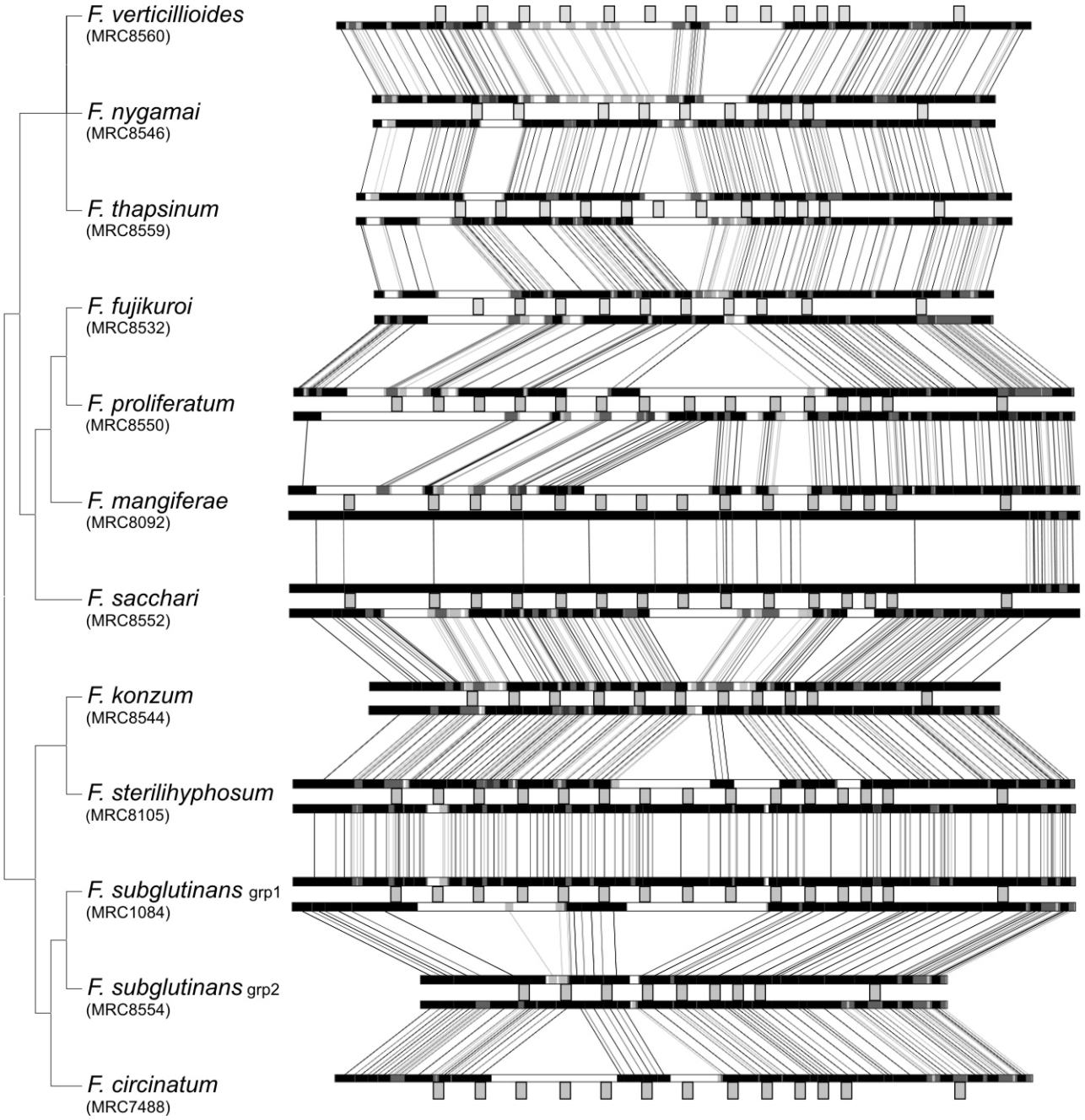

Fig. 10. GATA alignments of *ppg1* among species in the *Giberella fujikuroi* species complex. Gray boxes indicate positions of the predicted α-factor-like mature peptides (see Fig. 3 for amino acid sequences). Lines connect portions of sequence with high homology (lower cut-off bit-score of 24.9) and the darker the shading the stronger the homology. The accompanying reference phylogeny represents a consensus that was inferred from the results of previous studies (O'Donnell et al., 1998; Kvas et al., 2009).

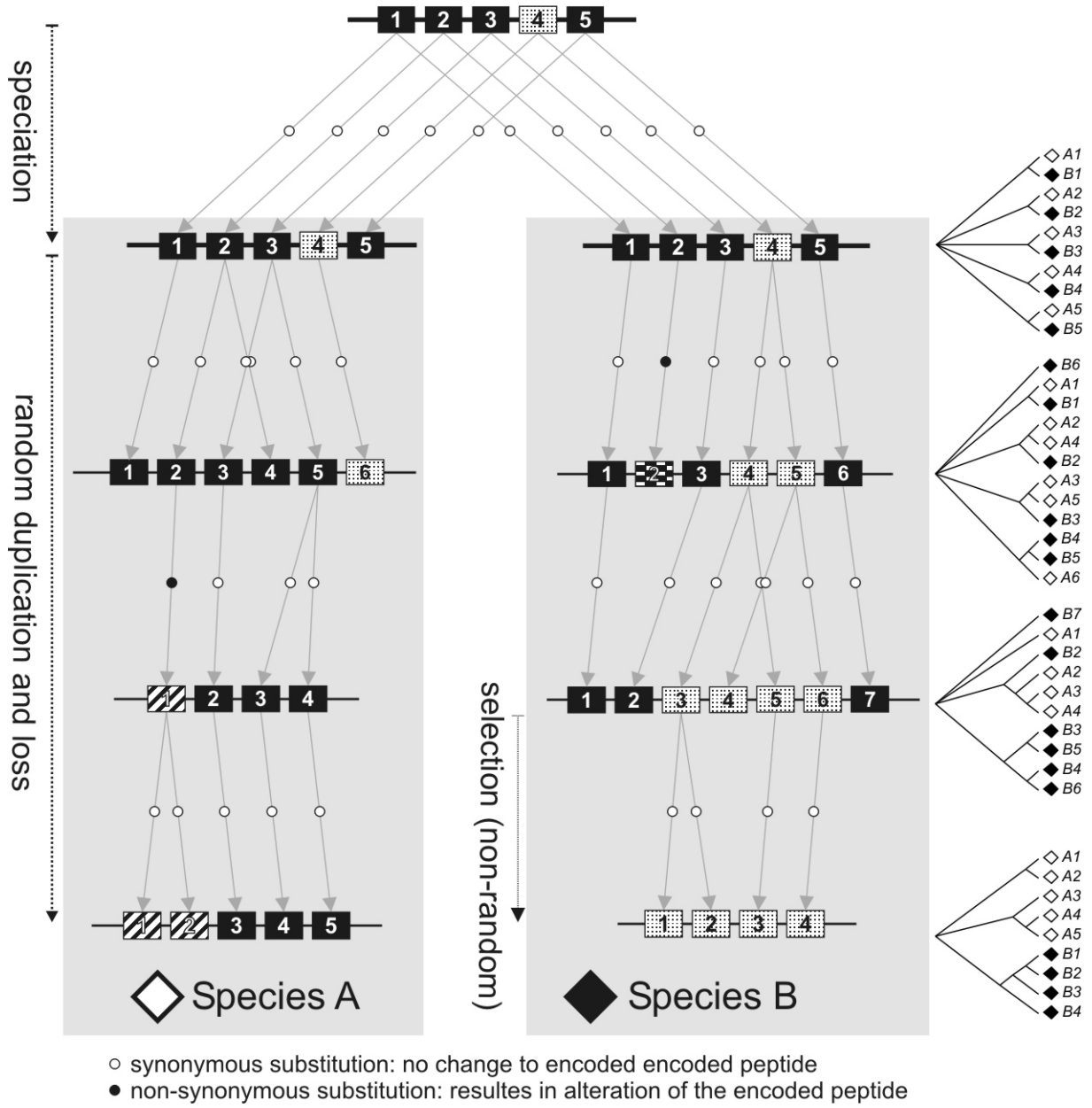

Fig. 11. Theoretical model showing birth and death evolution of tandem pheromone repeats.

Blocks indicate repeats of the mature α-factor-like peptide. Fill patterns indicate distinct amino acid sequence variants. A constant rate of evolution is assumed, with one substitution (indicated by a small dot) occurring at each step. Accompanying phylogenies represent relationships among underlying DNA sequences at each step.