The evolutionary trajectory of the mating-type (mat) genes in Neurospora relates to reproductive behavior of taxa

Lotta Wik¹, Magnus Karlsson² and Hanna Johannesson*³

Address: ¹Swedish University of Agricultural Sciences, Department of Biomedical Sciences and Veterinary Public Health, Husargatan 3, SE-75123 Uppsala, Sweden, ²Swedish University of Agricultural Sciences, Department of Forest Mycology and Pathology, Ulls väg 26, SE-750 07 Uppsala, Sweden and ³Uppsala University, Department of Evolutionary Biology, Norbyvägen 18D, SE-752 36 Uppsala, Sweden

Email: Lotta Wik - lotta.wik@bvf.slu.se; Magnus Karlsson - Magnus.Karlsson@mykopat.slu.se; Hanna Johannesson* - hanna.johannesson@ebc.uu.se

* Corresponding author

Abstract

Background: Comparative sequencing studies among a wide range of taxonomic groups, including fungi, have led to the discovery that reproductive genes evolve more rapidly than other genes. However, for fungal reproductive genes the question has remained whether the rapid evolution is a result of stochastic or deterministic processes. The mating-type (mat) genes constitute the master regulators of sexual reproduction in filamentous ascomycetes and here we present a study of the molecular evolution of the four mat-genes (mat a-1, mat A-1, mat A-2 and mat A-3) of 20 Neurospora taxa.

Results: We estimated nonsynonymous and synonymous substitution rates of genes to infer their evolutionary rate, and confirmed that the mat-genes evolve rapidly. Furthermore, the evolutionary trajectories are related to the reproductive modes of the taxa; likelihood methods revealed that positive selection acting on specific codons drives the diversity in heterothallic taxa, while among homothallic taxa the rapid evolution is due to a lack of selective constraint. The latter finding is supported by presence of stop codons and frame shift mutations disrupting the open reading frames of mat a-1, mat A-2 and mat A-3 in homothallic taxa. Lower selective constraints of mat-genes was found among homothallic than heterothallic taxa, and comparisons with non-reproductive genes argue that this disparity is not a nonspecific, genome-wide phenomenon.

Conclusion: Our data show that the mat-genes evolve rapidly in Neurospora. The rapid divergence is due to either adaptive evolution or lack of selective constraints, depending on the reproductive mode of the taxa. This is the first instance of positive selection acting on reproductive genes in the fungal kingdom, and illustrates how the evolutionary trajectory of reproductive genes can change after a switch in reproductive behaviour of an organism.

Background

Rapidly evolving genes are those with a higher than average percentage of amino-acid substitutions between species. Rapid evolution may be due to adaptive evolution, which occurs when natural selection promotes amino-acid divergence. Alternatively, rapid evolution may be due to a lack of functional constraint; for example, a pseudog-
ene rapidly accumulates mutations because of an absence of purifying selection.

Comparative sequencing studies among a wide range of taxonomic groups have led to the discovery that reproductive genes evolve more rapidly than other genes. In animals, essentially all steps in fertilization where the genes have been identified there is evidence of rapid divergence, and numerous reports of positively selected animal reproductive genes exist [1]. In contrast, in plants and fungi the selective forces driving reproductive gene diversity, and the functional consequences of reproductive gene evolution, are just beginning to be understood. In filamentous ascomycetes, comparatively low levels of between species nucleotide identity have been observed for mating-type genes, which constitute the master-regulators of sexual reproduction in this group of fungi [2]. Consequently these genes have been suggested to evolve at a faster rate than genes coding for e.g. metabolic enzymes [2-4], although the question has remained whether the rapid evolution is a result of stochastic or deterministic processes.

Many fungi have both asexual and sexual stages as a part of their life cycle. Their sexual mating systems are classified as either heterothallic or homothallic. In heterothallic fungi, strains must be of opposite mating type for sexual reproduction and morphogenesis, while homothallic species, generally thought to evolve from heterothallic ancestors [5] and references therein, reproduce sexually without a mating partner. In all heterothallic filamentous ascomycetes examined to date, the mat locus is the sole determinant of mating type, and sexual development in heterothallic taxa is regulated by alternative mat A or mat a genes at this locus. Alternate sequences at mat, denoted idiomorphs [6], lack significant sequence similarity and encode different transcriptional regulators [7-10]. Genetic analyses of the model species Neurospora crassa have identified one open reading frame (ORF) in the mat a idiomorph (mat a-1) as the major mating regulator in mat a strains [8]. The mat A idiomorph contains three genes, one of which is the main regulator of sexual development in mat A strains (mat A-1) [7,11]. While MAT A-1 and MAT a-1 constitute the critical factors for both mating and sexual development in N. crassa, MAT A-2 and MAT A-3 are suggested to increase the efficiency of the process [12].

In contrast to many yeast species [13], filamentous ascomycetes of either mating type are unable to switch to the opposite mating type. Nevertheless, it has been suggested that conversion from heterothallic to homothallic mode in filamentous ascomycetes resides within the mat locus, and that homothallic species carry both idiomorphs in a single thallus [14]. In so called pseudohomothallic species, represented in this study by Neurospora tetrasperma, two haploid nuclei of opposite mating-type (mat a and mat A) are maintained in each heterokaryotic ascospore progeny and vegetative cell, while in true homothallic species mat-genes from both idiomorphs are found within a single haploid genome. The structure of the mating type genes in homothallic Neurospora has not been determined in detail; however, hybridizations using cloned portions of the idiomorphs of N. crassa to the genome of related homothallic taxa have revealed three classes of homothallic taxa (Figure 1). The haploid genomes of one group contain both mat a and mat A sequences [15]. Another distinct group consists of taxa containing the entire mat A sequence, while the mat a-1 gene is absent [16,17]. Finally, the hybridization patterns of the homothallic N. tetricola provide evidence for the presence of mat a-1, mat A-1 and mat A-2 sequences, but it appears to lack parts of the A-idimorphic region corresponding to the mat A-3 region [15].

Here we present the first, to our knowledge, extensive study of molecular evolution of fungal mating-type genes in relation to reproductive strategy. We sequenced the four mat-genes of Neurospora and Gelasinospora taxa, which we for simplicity refer to as Neurospora in this study. In addition, for comparative purposes we sequenced parts of three genes with function assumed to be independent of reproductive mode of fungi; encoding actin (act), translational elongation factor EF-1 alpha (tuf-1) and glyceraldehyde 3-phosphate dehydrogenase (gcg-7). We confirm that the mat-genes evolve rapidly in this group of fungi and show that the evolutionary trajectory of each of the mat-genes is specifically related to the reproductive modes of the taxa; positive selection drives mat-gene diversity in heterothallic taxa, while the rapid evolution is due to a lack of selective constraint among homothallic taxa.

**Results**

**Sequence analysis of mat a-1**

Amplification products and sequences of mat a-1, originally described by Staben and Yanofsky [8], were successfully obtained from all investigated heterothallic strains of mating type a (mat a strains), and from all homothallic strains except N. africana, N. galapagosensis, N. dodgei and N. lineolata, thus confirming a previous report [17], suggesting that these four homothallic taxa lack the mat a-1 ORF. Taken together, mat a-1 was sequenced from twelve heterothallic and five homothallic strains, and the alignments of nucleotide and amino acid sequences of the gene are shown in Supplemental Figure S1 (Additional file 1). Several differences were found when comparing genomic and cDNA sequences of strain FGSC #4200 with the originally published N. crassa sequence of the same strain [8]. The most striking difference was an additional cytosine found downstream from the conserved, DNA binding domain of the high mobility group (HMG), at nucleotide...
Evolutionary relationship of the included taxa

Figure 1
Evolutionary relationship of the included taxa. Topology is constructed from previously reported topologies based on rDNA loci (28S, ITS), nuclear gene loci (Bml, ccg-7, mat a-1, mat A-1), and four anonymous nuclear loci [36-38, 46]. A cross indicates lineages where any of the ORFs is disrupted. A, B and C indicates intraspecific subgroups and PS1 through 3 phylogenetic species [46].
The size of the ORF of porting our revised description of #4200 was consistently found in all strains analyzed, sup-
previously reported MAT a-1 protein. This feature of FGSC C-terminal part of the protein is different from that of the
suggested by Staben and Yanofsky [8]. Consequently, the
homothallic taxa,
tions of amino acids were found. In three of the
cially in the C-terminal part in which insertions and dele-
was not the case among the six homothallic taxa, espe-
few amino acid substitutions were found. However, this
modified.

Sequence analysis of mat A-1
amplification products and sequences of mat A-1 were
mously obtained from the 12 heterothallic mat A and
the nine homothallic strains investigated here. The
lignments of nucleotide and amino acid sequences of mat A-1 are shown in Supplemental Figure S1 (Additional file 1). We confirmed the 293 amino-acid mat A-1 sequence of
ref erence strain FGSC# 2489 of N. crassa, originally published by Glass et al. 1990 [7], including a DNA-binding α1 domain showing similarity to the α1 transcription factor of Saccharomyces cerevisiae [7]. All other strains included here, however, differed from strain FGSC# 2489 in that amino acids at positions 32 and 33 were found to be lacking. With the exception of this difference, and a premature stop codon found in the heterothallic taxon
phylogenetic species 1 (PS1) resulting in a truncated protein lacking the last three amino acids, the length of the
ORF of mat A-1 was identical among all heterothallic and homothallic strains investigated here.

Sequence analysis of mat A-2
The mat A-2 ORF was successfully amplified and
sequences from all investigated heterothallic mat A strains and the homothallic strains. The nucleotide and amino acid alignments of the gene are shown in Supplemental
Figure S1 (Additional file 1), and confirmed the presence of a DNA-binding, acidic amphipathic α-helix (HPG-
domain) [11]. A number of differences were found when
comparing the sequence of genomic DNA and cDNA from
strain FGSC #2489 with the originally published sequence of the same strain [11], and these were also found in all
the other strains analyzed, indicating sequencing errors in
the original submission.

N. intermedia (both subgroups NiA and NiB) and N. crassa (NcC) were found to lack the start codon of mat A-2. No
alternative start-codon was found when investigating 40
bp of the upstream sequence, but another methionine was
found at aa-position 6 in both these taxa, which may serve
as an alternative start-codon. This pattern was confirmed
for additional strains of N. intermedia (NiA: FGSC# 8844)
and N. crassa (NcC: FGSC#8865) (data not shown). With
these exceptions, the size of the mat A-2 ORF is identical
among the heterothallic taxa. However, among the
homothallic taxa, the size of the mat A-2 ORF is not
conserved. Deletions of up to 6 amino acids are found in G.
cerealis, G. calospora and G. sublineolata, and, more importantly, in N. dodgei, N. africana, and N. galapagosensis the
mat A-2 ORF is disrupted by a stop codon (Figure 1). In N.
dodgei, the stop occurred at codon 39, while in N. africana
and N. galapagosensis, the stop was found at codon 288. By
sequencing mat A-2 from cDNA of N. dodgei and N. gala-
pagosensis we confirmed the structure of the ORF in these
strains. However, in spite of several independent trials, we
failed to amplify mat A-2 from cDNA of N. africana (data
not shown).

Sequence analysis of mat A-3
We successfully amplified and sequenced mat A-3 from all
investigated mat A and homothallic strains, except N. ter-
ricola, in which this gene previously has been indicated to
be absent [15]. The nucleotide and amino acid alignment of
mat A-3 is shown in Supplemental Figure S1 (Additional
de file 1), and shows the presence of an HMG-box. When
comparing our DNA sequence obtained from
 genomic and cDNA of strain FGSC# 2489 of N. crassa with
the original published mat A-3 sequence of this strain
[11], we found only one silent nucleotide difference. This
difference was found in all other strains analyzed here,
supporting our revised description of mat A-3.

The length of the ORF was found to be intact in all taxa
studied here, except for the homothallic taxa N. africana,
N. galapagosensis, N. dodgei and N. lineolata. Sequencing
of both genomic and cDNA from these four taxa revealed
that amino acids 100–177 have been deleted in *N. africana* and *N. galapagensis*, a region spanning the first half of the HMG-box. In *N. dodgei*, no start codon was found, although a methionine was found after 30 amino acids, which might serve as a start codon for *mat A-3* in this taxon. Finally, *N. lineolata* has a stop-codon at position 215 (Supplemental Figure S1, Additional file 1).

**Evolutionary pattern among sites**

Figure 2 shows the synonymous substitution rate (*d*~S~) for *mat*- and non-reproductive genes of the heterothallic and homothallic datasets, obtained through pairwise comparisons between taxa in all possible combinations. We found that *d*~S~ did not differ significantly between *mat*- and non-reproductive genes within each reproductive class. Furthermore, we found that the synonymous substitution rate of the *mat*-genes did not differ between heterothallic and homothallic taxa, while for the non-reproductive genes we found a significantly higher *d*~S~ in the homothallic as compared to the heterothallic dataset (*P* = 0.05, Mann-Whitney U-test).

Six models of variable ratios of nonsynonymous (*d*~N~) to synonymous substitution rates (*d*~S~), *ω*, among gene codons (site models outlined by Nielsen and Yang [18] and implemented in PAML version 3.14 [19]) were used to investigate the evolutionary constraints acting on each gene. Detailed descriptions of the models are given in the Methods section. Supplemental Table S1 (Additional file 1) lists likelihood ratio statistics and parameter estimates for the four *mat*-genes, and three genes which function is assumed to be independent of reproductive mode of a fungus; *act*, *tef-1* and *ccg-7* [20-22]. The Bayes Empirical Bayes (BEB) probabilities for each codon of the *mat*-genes to fit into one of the three classes of sites obtained by running the positive-selection model are shown in Figure 3.

We found a higher *ω* among *mat*-genes than among non-reproductive genes (*P* < 0.05, Fisher exact test). Given an equal synonymous substitution rate between *mat* and non-reproductive genes within each reproductive class (Figure 2), this finding indicates that *mat*-genes evolve faster than the non-reproductive genes. Furthermore, all four *mat*-genes were found to evolve faster among the homothallic than the heterothallic taxa; although the synonymous substitution rate of the *mat*-genes was equal among these groups, the averaged value of *ω* for each *mat*-gene, obtained using the one-ratio model implemented in PAML version 3.14, was higher in the homothallic compared to the heterothallic dataset (0.385 versus 0.144 for *mat a-1*, 0.487 versus 0.233 for *mat A-1*, 0.581 versus 0.386 for *mat A-2* and 0.427 versus 0.316 for *mat A-3*; *P* < 0.001, Fisher exact tests). This pattern was also found for the *tef-1* gene, for which we found a *ω* of 0.133 in the homothallic dataset as compared to 0.017 for the heterothallic. In contrast, the *ω* between homothallic and heterothallic dataset did not differ for the *act* or the *ccg-7* genes. For the *act*-gene we estimated *ω* of 0 for both datasets and for *ccg-7* we found *ω* of 0.102 versus 0.097 for the heterothallic and homothallic datasets, respectively.

*Mat a-1* was found to be the most conserved *mat*-gene in both datasets, with lowest average *ω* ratios. Although *ω* was smaller in the heterothallic than the homothallic dataset, *mat a-1* was found to evolve under positive selection among the heterothallic taxa (i.e. both models allowing for sites with *ω* > 1 fitted the data significantly better than the corresponding neutral model, *P* < 0.05). Bayes empirical Bayes (BEB) calculations for posterior probabilities of sites belonging to site classes revealed two codons, codon 7 and 43, with a posterior probability > 0.95 to belong to the class with *ω* > 1, using both the positive-selection and the beta*ω* model (Figure 3a).

The sites of *mat a-1* showing an elevated *ω* were not identical between the heterothallic and homothallic datasets (Figure 3a, b). The second exon was found to be extremely conserved among the heterothallic taxa, and codes for a DNA binding HMG-domain. The PRPPNAYILYRK motif suggested by Staben & Yanofsky [8] to be functionally important for mating, based on conservation between *Schizosaccharomyces pombe* and *N. crassa*, was found to be completely conserved in both heterothallic and homothallic taxa possessing this gene, but variation was found
Figure 3
Probabilities that sites of the mat-genes belong to site classes with different selective pressures. Bar height represents the BEB probabilities for each site to belong to one of the three site classes obtained by running model 2a in PAML version 3.14. Asterisks indicate a probability of > 95% that the site belongs to a class with $\omega > 1$ in the positive-selection and/or B&\(\omega\)-models implemented in codeml-package, with site number as in amino acid alignment in Supplemental Figure S1. Arrow indicate intron splicing positions and pins indicate disrupted ORFs in N.a. (N. africana), N.g. (N. galapogosensis), N.t. (N. terricola), N.l. (N. lineolata), G.c. (G. calospora), N. p. (N. pannonica). Functional domains of each gene, as defined by Debuchy and Turgeon [14], are indicated.
within the other part of the HMG-box for both groups of taxa (Figure 3a, b).

For mat A-1, the neutral models of evolution fitted the data equally well as the models of positive selection. The rapidly evolving sites in the heterothallic dataset were clustered into domains, e.g. in the C-terminal end of mat A-1 we found a stretch of adjacent rapidly evolving sites, including one site with a BEB posterior probability of $P > 0.95$ to belong to a class with $\omega > 1$, while the a1-box was found to be conserved (Figure 3c). This was in contrast to the homothallic dataset in which the rapidly evolving sites are more evenly distributed along the gene (Figure 3d).

In mat A-2 we found the most pronounced difference between the heterothallic and homothallic dataset. In the heterothallic dataset, mat A-2 was found to evolve under positive selection. BEB calculations of posterior probabilities revealed seven sites with $P > 0.95$ to belong to the class with $\omega > 1$ using the positive-selection and the beta model (Figure 3e). This was very different from the homothallics in which a large proportion of sites were found to evolve neutrally; more than half of the codons were found in a class with $\omega = 1.011$ using the discrete model. Noteworthy, the discrete model fitted the data significantly better than the one-ratio model for all datasets except the mat A-2 of the homothallics, for which the one-ratio model with a $\omega$ of all codons of 0.581 fitted equally well. The HPG-domain that was previously suggested to be conserved among euascomycetes [14] was found to be conserved in our heterothallic dataset, except for three rapidly evolving sites. However, the HPG-domain is not conserved in the homothallic dataset (Figure 3e, f).

In mat A-3, the smallest difference in average $\omega$ between the heterothallics and the homothallics was observed. Similarly to mat a-1, and in contrast to the observations at mat A-1 and mat A-2, when using the discrete model no codons were found in a class with $\omega \approx 1$ in the homothallic dataset for mat A-3. The HMG domain was more conserved in the heterothallic than in the homothallic taxa. However, the one site found to be in a class with $P > 0.95$ of positive selection (site 205) in the heterothallics was located within this domain.

The investigated region of the act-gene was found to be extremely conserved in both datasets. This was also the case for all codons of tef-1 for the heterothallic dataset, while a small proportion (3.3%) of the sites of this gene was found to evolve under positive selection among homothallic taxa (i.e. found in a class with $\omega$ of 6.062 using both the discrete and positive selection model: Supplemental Table S1 (Additional file 1)). For both of these genes, no neutrally evolving codons were identified in any of the datasets. For the ceg-7 gene the majority (> 90%) of the sequenced codons are conserved in both datasets, however, a small proportion of the codons was found in a class with $\omega$ of 1.6 in the heterothallic dataset and 1.175 in the homothallic dataset, under the discrete model (Supplemental Table S1, Additional file 1).

**Discussion**

In this study we confirm that the mat-genes evolve rapidly in *Neurospora*, and show that the rapid divergence is due to either adaptive evolution or lack of selective constraint, depending on the reproductive mode of taxa.

A lower overall value of $\omega$ was found for the mat-genes among heterothallic as compared to homothallic taxa. Nevertheless, the two mating type genes mat a-1 and mat A-2 were found to evolve under positive selection in heterothallic taxa. Although a high between-taxon nucleotide divergence of reproductive genes in fungi has been observed previously [2-4], this is the first evidence of positive selection of reproductive genes in this kingdom. The finding fits directly with patterns observed for animal reproductive genes, in which essentially all identified genes are characterized by high levels of nonsynonymous divergence and positive selection [1,23,24]. None of the positively selected sites were found within the regions encoding known functional domains of the genes, and the biological significance of alterations at the positively selected sites is unknown. MAT a-1 and its orthologs in other filamentous ascomycetes are characterized by a conserved region similar to motifs found in several DNA binding proteins of the high mobility group (HMG), and has been shown to bind to DNA in _in vitro_ assays [25]. MAT a-1 is one of the master genes regulating fertilization in _mat a_ strains of _Neurospora_, as well as in other species [14]. During the initial phase of mating in heterothallic _Neurospora_, the trichogynes (female receptive hyphae) grow toward and fuse with male cells of the opposite mating type, and the directed growth is mediated by chemotropic communication between _mat_-regulated pheromones diffused from male cells and their corresponding receptors on the trichogynes. Competition between individuals could potentially drive adaptation of _mat_-genes, as indirectly involved in localization of, binding to and fusion between trichocyte and conidia of opposite mating types. An alternative hypothesis would be that the signature of positive selection observed in heterothallic _mat a-1_ and _mat A-2_ could be related to mating type associated vegetative incompatibility, rather than sexual development. Vegetative incompatibility genes are highly variable, and some have been shown to evolve to diversifying manner [26].

Our result of _mat a-1_ in heterothallic taxa confirms previous reports suggesting that an intact HMG domain and...
the presence of a C-terminal tail, although non-specific, are required for mating [25,27]; at least the first part of the HMG-domain is under strong purifying selection among heterothallic taxa, while lower levels of conservation are found in the C-terminal tail (Figure 3a). The finding that a group of closely related homothallic taxa lack mat a-1 indicate that this gene is not essential for completion of the homothallic sexual cycle, and it is possible that this function of mat a-1 is redundant with another gene in the genomes of this group of taxa. The premature stop codons in mat a-1 in the homothalic N. pannonica and N. terricola indicate that this gene may not be required for completing the sexual cycle in a wider range of homothallic taxa. However, in these two taxa the stop codons are found distal to the HMG box, and functional assays have to be performed to verify this hypothesis.

The MAT A-1 protein is characterized by the presence of an α1 domain showing similarity to the α1 transcription factor of Saccharomyces cerevisiae [7]. Mat A-1 is present in all homothallic taxa investigated here, and the only gene in which the ORF is intact in all homothallic taxa, suggesting that this protein does not function correctly in this taxon, and we have no indications of directional selection in the N. terricola mat a-1 gene to change function. Furthermore, two of the four taxa lacking the mat a-1 gene, N. africana and N. galapagosensis, also lack the first half of the MAT A-3 HMG-domain. Thus, we are not able to provide any support for the hypothesis that the essential functions of the lost mat A-3 or mat a-1 are compensated by the remaining HMG encoding gene in Neurospora.

In addition to the mat-genes, tef-1 also shows a higher ω among homothallic taxa as compared to heterothallic, and this result implies that the evolution of this gene is also dependent of reproductive mode of the fungi. Interestingly, this elevated ω is due to a small proportion of positively selected sites among the homothallic taxa. At this point, we have no explanation for the observation, but it implies that tef-1 is involved in sexual development in Neurospora, or that it is affected by genetic linkage through a close physical proximity to the mat locus [31].

In contrast to the heterothallic taxa, rapid evolution of mat-genes in homothallic Neurospora is caused by genetic decay. Instead of being conserved, the mat-genes in homothallic Neurospora evolve under relaxed selective constraints. This is evident from a higher proportion of neutrally evolving codons among the homothallics as well as observations of a disrupted ORF in three of the four mat-genes in homothallic taxa. Comparisons with non-reproductive genes argue that this disparity is mat-specific, and not a nonspecific, genome-wide, phenomenon. The observations here provide a parallel to the observations of ecotypes with additional loss-of-function mutations at one or more SI-modifier genes after the evolutionary switch to autogamy in the Arabidopsis thaliana lineage [32], but is the first time low selective constraints of reproductive proteins associated with the switch to homothallism has been reported from fungi.

Our results indicate that we observe an early stage of degeneration of mat-genes in homothallic Neurospora. With the exception of mat A-2 in N. africana, we found that all genes are expressed in homothallic taxa despite frame shifts and stop codons, and the DNA sequences downstream of the disruptions are relatively conserved. This short degeneration time of the mat-genes in homothallic taxa might explain the results where the alterations in the N. africana mat A-1 are not sufficient to abolish its functions when assayed in a heterothallic relative [29]. Furthermore, the observation that they have not yet become pseudogenes, in the sense that they do not exhibit degenerative features that prevent their expression [33], indicates a low cost associated with their expression.
The results presented here indicate that once a Neurospora switches reproductive mode from heterothallic to homothallic, it expresses low selective pressure for maintenance of functional mat-genes. This finding, together with the observation that both heterothallic counterparts of the mat-genes are not required for sexual development in certain taxa [15,17], implies that within homothallic Neurospora, activation of sexual reproduction can be independent of mat a/A interaction. In other filamentous ascomycetes conversion from heterothallic to homothallic mode evidently resides within the mat locus [5] and both idiomorphic counterparts of heterothallic mat-genes are required for homothallic sexual development. In these taxa, the function of mat-genes has been conserved [34]. This study indicates that the transition to homothallism in Neurospora may represent a different paradigm. The exact nature of this difference is not known, but it would involve differences downstream of mat a/A interaction.

If functional mating-type genes are essential for outcrossing in the homothallics, the observed genetic decay may lead to a progressive loss of outcrossing functions and ultimately, to asexuality. However, alternative ways of outcrossing, independent of mat-genes may be possible, as expression of pheromones and receptors can be detected in some strains deleted for the mat locus [35]. Furthermore, it is possible for strains to bypass the need for fertilization by the formation of heterokaryons, which may lead to outcrossing, as done routinely in the laboratory with the homothallic Aspergillus nidulans [36].

Conclusion
In this study we show that the mat-genes evolve rapidly in Neurospora, and that the rapid divergence is due to either adaptive evolution or lack of selective constraint, depending on the reproductive mode of taxa. In heterothallic Neurospora, the mat-genes evolve in an adaptive manner while once a Neurospora switches reproductive mode from heterothallic to homothallic there is a low selective advantage for maintenance of functional mat-genes. To our knowledge, this is the first example showing that the evolutionary trajectory of reproductive proteins changes after a switch in reproductive behaviour of an organism.

Methods

Fungal material
Thirty-four strains belonging to 21 taxa of Neurospora and Gelasinospora were used in this study (Table 1). Recent studies show that these two genera are closely related and not reciprocally monophyletic [37-39]. Therefore we include homothallic taxa of both genera in this study, and for simplicity refer to all of them as Neurospora. The strains were obtained from the Fungal Genetics Stock Center (University of Kansas Medical Center, Kansas City, KS).

Nucleic acid extraction and complementary DNA (cDNA) construction
Strains were grown in minimal medium broth [40] with 1% sucrose for 3 days at 37°C, after which nucleic acids were extracted as described previously [41]. Total RNA was treated with DNase I according to the manufacturer’s protocol (Fermentas, Burlington, Canada). cDNA was synthesized from total RNA by reverse transcription using an oligo(dT) primer (5’-T25VN-3’) and Moloney Murine Leukemia Virus (M-MLV) reverse transcriptase (Invitrogen), according to the manufacturer’s instructions.

Selection of loci
The entire coding regions of the four mat-genes, and partial coding regions of three additional genes; encoding actin (act: 767 bp) [22], translational elongation factor EF-1 alpha (tef-1: 834 bp) [20] and glyceraldehyde-3-phosphate dehydrogenase (ccg-7: 753 bp) [21], were investigated in this study. Act, tef-1 and ccg-7 were included for comparative purposes, and selected to represent genes which function is independent of reproductive mode of a fungus. As both physical proximity and expression levels can have an effect on evolutionary rate [42,43], these genes were selected to be located on different chromosomes, and expression levels of tef-1 and ccg-7 have been shown to vary between different treatments [20,21].

PCR amplification
All included loci were amplified from one selected strain per taxa. Published sequences for mat a, mat A, act, tef-1 and ccg-7 of N. crassa were used as templates for each gene when designing primers. For isolates where amplification of mat A-3 was not possible using primers based on the N. crassa gene sequence, a primer walking sequencing strategy was used for searching the N. crassa genome sequence [31] for conserved region upstream from mat A-3. All primer sequences and GenBank accession numbers of reference sequences are given as Supplemental Table S2 (Additional file 1). PCR was performed with 1.55 U Thermocycler Taq DNA polymerase (Saveen & Werner, Limhamn, Sweden) on a GeneAmp PCR System 9700 (Applied Biosystems, Foster City, CA) according to the manufacturer’s instructions.

DNA sequencing
The PCR products were purified using a PCR purification kit (QIAquick, Qiagen) and sequencing was performed by Macrogen Inc., Seoul, Korea, utilizing ABI 3730 XL automated sequencers (Applied Biosystems, Foster City, CA). Raw sequence data was analyzed using the Sequencher 4.1.4 software (Gene Codes Corporation, Ann Arbor, MI) and BioEdit version 7.0.5.2 [44].
Codon-based likelihood analyses

Likelihood-based tests were used to investigate the nature of evolutionary pressure acting on the mat-genes in *Neurospora*, using the CODEML program of the PAML version 3.14 package [19,45]. All analyses were carried out using separate datasets consisting of the heterothallic or homothallic taxa, as well as a dataset containing both hetero- and homothallics. For each model, equilibrium codon frequencies were estimated from the average nucleotide frequencies at each codon position, amino acid distances were assumed to be equal, and the transition/transversion ratio (κ) was estimated from the data. For all other parameters, we use the default settings provided by Yang et al. [46]. We assumed linkage between collinear sites. The tree structure file used as input file in PAML was:

```
(((G_calospora, N_sublineolata), G_tetrasperma),((N_africana, N_sublineolata), G_cerealis, N_terricola, N_pannonica,((((((N_crassaA, N_crassaB), N_crassaC), N_lineolata), G_cerealis, N_sublineolata)), N_tetrasperma)), (N_intermediaA, N_intermediaB), PS2), N_discreta));
```

shown in Figure 1 and constructed based on the topology obtained from previous reports where neighbor-joining, maximum parsimony and bayesian methods were used to construct species-phyllogenies based on multiple loci [37-39,47]. We assume the true evolutionary history of the genes under study is the same as the history of the species. When only a subset of the taxa was used in the analyses, the corresponding subset of the topology was used. Sequences downstream of stop codons and

Table 1: Fungal strains used in this study

| Taxa                               | FGSC number | Mating system | Mating type | mat-gene analyzed† |
|------------------------------------|-------------|---------------|-------------|---------------------|
| **Heterothallic taxa:**            |             |               |             |                     |
| *N. crassa* Shear & Dodge          | 2489        | heteroth.     | A           | -                   |
| *N. crassa* Shear & Dodge (NcA1)   | 4200        | heteroth.     | a           |                     |
| *N. crassa* Shear & Dodge (NcA1)  | 8900        | heteroth.     | A           | A1, A2, A3          |
| *N. crassa* Shear & Dodge (NcB1)  | 8848        | heteroth.     | a           | a                   |
| *N. crassa* Shear & Dodge (NcB1)  | 8830        | heteroth.     | A           | A1, A2, A3          |
| *N. crassa* Shear & Dodge (NcC1)  | 8772        | heteroth.     | a           | a                   |
| *N. crassa* Shear & Dodge (NcC1)  | 8858        | heteroth.     | A           | A1, A2, A3          |
| *N. crassa* Shear & Dodge PS3³²    | 8860        | heteroth.     | a           | a                   |
| *N. sitophila* Shear & Dodge       | 8835        | heteroth.     | a           | a                   |
| *N. tetrasperma* Shear & Dodge     | 8770        | heteroth.     | A           | A1, A2, A3          |
| *N. tetrasperma* Shear & Dodge     | 412         | heteroth.     | a           | a                   |
| *N. tetrasperma* Shear & Dodge     | 8774 pseudohomoth.³ | heteroth. | A | A1, A2, A3 |
| *N. tetrasperma* Shear & Dodge     | 8775 pseudohomoth.³ | heteroth. | A  | A1, A2, A3 |
| *PS1³²*                            | 8817        | heteroth.     | A           | A1, A2, A3          |
| *PS1³²*                            | 8815        | heteroth.     | a           | a                   |
| *N. intermedia* Tai (NiA¹)         | 8901        | heteroth.     | A           | A1, A2, A3          |
| *N. intermedia* Tai (NiA¹)         | 8841        | heteroth.     | a           | a                   |
| *N. intermedia* Tai (NiB¹)         | 8844        | heteroth.     | A           | A1, A2, A3          |
| *N. intermedia* Tai (NiB¹)         | 8768        | heteroth.     | a           | a                   |
| *PS2³²*                            | 8847        | heteroth.     | A           | A1, A2, A3          |
| *PS2³²*                            | 8853        | heteroth.     | a           | a                   |
| *N. discreto* Perkins & Raju       | 8780        | heteroth.     | A           | A1, A2, A3          |
| *N. discreto* Perkins & Raju       | 8827        | heteroth.     | a           | a                   |
| **Homothallic taxa containing all mat-genes:** | | | | |
| *G. calospora* (Mouton) Moreau & Moreau | 958 | homoth. | - | a1, A1, A2, A3 |
| *G. cerealis* Dowding              | 959         | homoth.       | -           | a1, A1, A2, A3       |
| *N. pannonica* Krug & Khan         | 7221        | homoth.       | -           | a1, A1, A2, A3       |
| *N. sublineolata* (Furuya & Udagawa von Arx) | 5308 | homoth. | - | a1, A1, A2, A3 |
| **Homothallic taxon lacking mat A-3:** | | | | |
| *N. terricola* Gochenaur & Backus  | 1889        | homoth.       | -           | a1, A1, A2          |
| **Homothallic taxa lacking mat a-1:** | | | | |
| *N. africana* Huang & Backus       | 1740        | homoth.       | -           | A1, A2, A3          |
| *N. galapagosensis* Mahoney & Backus | 1739 | homoth. | - | A1, A2, A3 |
| *N. dodgei* Nelson & Novak         | 1692        | homoth.       | -           | A1, A2, A3          |
| *N. lineolata* Frederick & Uecker  | 1910        | homoth.       | -           | A1, A2, A3          |

¹Intraspecific subgroups and ²phylogenetic species designated by Dettman et al. 2003 [47].
³Analyzed as heterothallic since it has been shown to occasionally outcross in nature [49].
⁴ORF included in evolutionary analyses.
frameshift mutations were omitted from the analyses. We verified that including the sequence downstream of the disruption did not affect the result by performing parallel runs, in which stop codons and frameshift mutations were omitted form the sequences.

We estimated the synonymous substitution rate (dS) for each gene by comparing rates between taxa in all possible combinations (runmode = -2 in PAML 3.14 [19]). A Mann-Whitney U-test (Statistica 7.1, StatSoft, Tulsa, OK) was utilized to examine whether the synonymous substitution pattern differs between mat- and non-reproductive genes in heterothallic and homothallic taxa.

Models of variable ratios of nonsynonymous (dN) to synonymous substitution rates (dS), ω, among sites were used to test for the presence of sites under different evolutionary constraints, and to identify them. We used six models outlined by Nielsen and Yang [18] and implemented in PAML version 3.14 [19]. The one-ratio model (Nssites 0) assumes one ω for all sites, while the discrete model (Nssites 3) uses a general discrete distribution with three site classes estimated from the data. The nearly-neutral model (Nssites 1a) assumes two classes of sites in the protein, with ω < 1 and ω = 1. The positive-selection model (Nssites 2a) adds a third class of sites to the neutral model, in which ω > 1. The beta model (Nssites 7) uses a β distribution of ω over sites: β (p, q), which, depending on the parameters p and q, can take various shapes in the interval (0, 1). The beta model (Nssites 8) adds an extra class of sites to the beta model, with a proportion of ω estimated from the data.

To verify which of the models best fits the data, likelihood ratio tests (LRTs) were performed by comparing twice the difference in log-likelihood values (-2lnA) between two models using a χ² distribution, with the number of degrees of freedom equal to the differences in the number of parameters between the models. For each dataset, we made three comparisons: first, the one-ratio model (Nssites 0) was compared to the discrete model (Nssites 3), to test for a variation of ω among codons within the gene. To test whether a gene evolve under positive selection, the two models including a class of codons with positively selected sites (i.e. ω > 1; models Nssites 2a and Nssites 8) were compared to their corresponding neutral models (Nssites 1a and Nssites 7, respectively) using 4, 2 and 2 degrees of freedom, respectively [45].

We identified particular sites in the genes that were likely to have evolved under positive selection by using the Bayes empirical Bayes (BEB) calculation of posterior probabilities for site classes implemented in the positive-selection- and betaω models [48].

Authors' contributions
Lotta Wik carried out the majority of the molecular work, participated in sequence alignment and helped to draft the manuscript. Magnus Karlsson participated in the in the design of the study, the molecular work, the statistical analyses and helped to draft the manuscript. Hanna Johannesson initiated the study, performed the statistical analyses and drafted the manuscript. All authors read and approved the final manuscript.

Additional material

Additional File 1
Supplemental Table S1. Likelihood ratio statistics and parameter estimates for the two datasets as inferred under six models of ω over codons. Supplemental Table S2. Primers used in the current study. Supplemental Figure S1. Nucleotide and amino acid alignments of all genes and strains investigated in this study, and the originally published sequence of the mat-genes (Genbank accession numbers M54787, M53876). For the mat-genes, the entire coding sequence is aligned, while for the non-reproductive genes the alignment includes the part of the coding region amplified with primers given in the Supplemental Table S2.
Click here for file [http://www.biomedcentral.com/content/supplementary/1471-2148-8-109-S1.pdf]

Acknowledgements
We wish to thank Timothy Y. James, Henrik Johannesson and Anne Pringle for helpful suggestions and critical reading of the manuscript. Four anonymous reviewers are acknowledged for valuable criticism of the manuscript. Financial support was obtained from the Swedish Research Council, Carl Tryggers Stiftelse for Vetenskaplig Forskning and the Sven and Lily Lawski Foundation.

References
1. Clark NL, Aagaard JE, Swanson WJ: Evolution of reproductive proteins from animals and plants. Reproduction 2006, 131(1):11-22.
2. Poggele S: Phylogenetic relationships between mating-type sequences from homothallic and heterothallic ascomycetes. Curr Genet 1999, 36(4):222-231.
3. Turgeon BG: Application of mating type gene technology to problems in fungal biology. Annu Rev Phytopathol 1998, 36:115-137.
4. Poggele S, Kuck U: Identification of transcriptionally expressed pheromone receptor genes in filamentous ascomycetes. Gene 2001, 280(1-2):9-17.
5. Yun SH, Berbee ML, Yoder OC, Turgeon BG: Evolution of the fungal self-fertile reproductive life style from self-sterile ancestors. Proc Natl Acad Sci USA 1999, 96(10):5592-5597.
6. Metzenberg RL, Glass NL: Mating type and mating strategies in Neurospora. Bioessays 1990, 12(2):53-59.
7. Glass NL, Groteschueschen J, Metzenberg RL: Neurospora crosa a mating-type region. Proc Natl Acad Sci USA 1990, 87(13):4912-4916.
8. Staben C, Yanofsky C: Neurospora crassa A mating-type region. Proc Natl Acad Sci USA 1990, 87(13):4917-4921.
9. Picard M, Debuchy R, Coppin E: Cloning the mating types of the heterothallic fungus Podospora anserina – developmental features of haploid transformants carrying both mating types. Genetics 1991, 128(3):539-547.
10. Turgeon BG, Bohlmann H, Ciuffetti LM, Christiansen SK, Yang G, Schafer W, Yoder OC: Cloning and analysis of the mating-type
genes from Cochliobolus heterostrophus. Mol Gen Genet 1993, 238(1-2):270-284.

11. Ferreira AVB, Saape S, Glass NL: Transcriptional analysis of the matA idiomorph of Neurospora crassa identifies two genes in addition to mat A-1. Mol Gen Genet 1996, 250(6):767-774.

12. Ferreira AVB, An ZQ, Metzenberg RL, Glass NL: Characterization of mat A-2, mat A-3 and Delta matA mating-type mutants of Neurospora crassa. Genetics 1998, 148(3):1069-1079.

13. Knop M: Evolution of the hemiascomycete yeasts: on life styles and the importance of inbreeding. Bioessays 2006, 28(7):696-708.

14. Debuchy R, Turgeon BG: Mating-type structure, evolution, and function in Euascomycetes. In: The Mycota I: Growth, Differentiation and Sexuality Edited by: Kues, Fischer. Berlin Heidelberg: Springer-Verlag; 2006:293-323.

15. Beatty NP, Smith ML, Glass NL: Molecular characterization of mating-type loci in selected homothallic species of Neurospora, Anxillia, and Aspergillus. Myc Res 1994, 98:1309-1316.

16. Glass NL, Vollmer SJ, Staben C, Grotelueschen J, Metzenberg RL, Yanofsky C: Expression of a mating-type sequence. Exp Mycol 1990, 14(4):274-289.

17. Nielsen R, Yang ZH: Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. Genetics 1998, 148(3):929-936.

18. Yang Z: PAML: a program package for phylogenetic analysis by maximum likelihood. Comput Appl Biosci 1997, 13(5):355-356.

19. Ichimura R, Modaka S, Robson J: Cloning, nucleotide sequence, and expression of the matI gene from the homothallic species of Neurospora. Jpn J Genet 1995, 70:273-287.

20. Shinohara ML, Loros JJ, Dunlap J: Glyceraldehyde-3-phosphate dehydrogenase is regulated on a daily basis by the circadian clock. J Biol Chem 1998, 273(1):464-472.

21. Tinsley J, Lee I, Minke P, Plaimann M: Analysis of actin and actin-related protein 3 (ARP3) gene expression following induction of hyphal tip formation and aporal growth in Neurospora. Mol Gen Genet 1998, 259:601-609.

22. Swanson WJ, Vacquier VD: The rapid evolution of reproductive proteins. Nat Rev Genet 2002, 3(2):137-144.

23. Philley ML, Staben C: Functional analyses of the Neurospora crassa Mat A-1 mating-type polypeptide. Genetics 1994, 137(3):715-722.

24. Wu CI, Saape S, Glass L: Evidence for balancing selection operating at the het-c heterokaryon incompatibility locus in a group of filamentous fungi. Proc Natl Acad Sci USA 1998, 95(21):12398-12403.

25. Arnaise S, Zickler D, Glass NL: Heterologous expression of mating-type genes in filamentous fungi. Proc Natl Acad Sci USA 1993, 90(14):6616-6620.

26. Lin X, Hull CH, Heitman J: Sexual reproduction between partners of the same mating type in Cryptococcus neoformans. Nature 2005, 434:1017-1021.

27. Glass NL, Smith ML: Structure and function of a mating-type gene from the homothallic species Neurospora africana. Mol Gen Genet 1994, 244(4):401-409.

28. Robertson MP, Ford DJ, Read ND: Homothallism and heterothallism in Sordaria brevicollis. Mycologia 1998, 102:1215-1223.

29. Neurospora crassa Database, Broad Institute [http://www.broad.mit.edu/annotation/genome/neurospora/Home.html].

30. Nasrallah ME, Liu P, Sherman-Broyles S, Boggs NA, Nasrallah JB: Natural variation in expression of self-incompatibility in Arabidopsis thaliana: implications for the evolution of selfing. Proc Natl Acad Sci USA 2004, 101(45):16670-16674.

31. Balakirev ES, Ayala FJ: Pseudogenes: are they “junk” or functional DNA? Ann Rev Genet 2003, 37:123-151.

32. Lee J, Lee T, Lee YW, Yun SH, Turgeon BG: Shifting fungal reproductive mode by manipulation of mating-type genes: obligatory heterothallism of Gibberella zeae. Mol Microbiol 2003, 50(1):145-152.