Unidirectional Evolutionary Transitions in Fungal Mating Systems and the Role of Transposable Elements

Anastasia Gioti,*1 Alexandra A. Mushegian,1 Rebecka Strandberg,1 Jason E. Stajich,2 and Hanna Johannesson1

1Department of Evolutionary Biology, Uppsala University, Uppsala, Sweden
2Department of Plant Pathology & Microbiology and Institute for Integrative Genome Biology, University of California

*Corresponding author: E-mail: anastasia.gioti@ebc.uu.se.
Associate Editor: Naoki Takebayashi

Abstract

In the fungal kingdom, the evolution of mating systems is highly dynamic, varying even among closely related species. Rearrangements in the mating-type (mat) locus, which contains the major regulators of sexual development, are expected to underlie the transitions between self-sterility (heterothallism) and self-fertility (homothallism). However, both the genetic mechanisms and the direction of evolutionary transitions in fungal mating systems are under debate. Here, we present new sequences of the mat locus of four homothallic and one heterothallic species of the model genus Neurospora (Ascomycota). By examining the patterns of synteny among these sequences and previously published data, we show that the locus is conserved among heterothallic species belonging to distinct phylogenetic clades, while different gene arrangements characterize the four homothallic species. These results allowed us to ascertain a heterothallic ancestor for the genus, confirming the prediction of the dead-end theory on unidirectional transitions toward selfing. We show that at least four shifts from heterothallism to homothallism have occurred in Neurospora, three of which involve the acquisition of sequences of both mating types into the same haploid genome. We present evidence for two genetic mechanisms allowing these shifts: translocation and unequal crossover. Finally, we identified two novel retrotransposons and suggest that these have played a major role in mating-system transitions, by facilitating multiple rearrangements of the mat locus.

Key words: mating system, evolutionary transition, transposable element, mating-type locus, unequal crossover.

Introduction

The fungal kingdom exhibits an astonishing diversity in reproductive strategies. Most importantly, fungi show a high versatility in mating systems, which often vary between closely related species (Heitman et al. 2007). However, the direction of transitions in mating systems remains understudied and under debate (Lin and Heitman 2007). Sexual reproduction is tightly linked with survival in fungi (Aanen and Hoekstra 2007) and understanding the genetic mechanisms behind transitions in sexual reproductive mode is important for assessing the costs of such shifts and thus, their generality across the tree of life (Lee et al. 2010). Assessing the ancestry of mating systems is also fundamental from a theoretical point of view, as one-way transitions from outcrossing to self-fertility are predicted to lead to an evolutionary dead-end, i.e., the extinction of species in the long term (Takebayashi and Morrell 2001).

There are two ways of assessing the ancestry of mating systems: ancestral state reconstruction analyses and structural studies of the genomic regions that contain genes controlling sexual compatibility, such as the self-compatibility loci in plants (e.g., Kusaba et al. 2001; Guo et al. 2009), and the mating-type loci in fungi (reviewed in Kronstad and Staben 1997; Shiu and Glass 2000; Debuchy and Turgeon 2006). In fungi, ancestral state reconstruction analyses of mating systems have rarely been used (Buschbom and Barker 2006; Nygren et al. 2011). This is not only due to limited sampling of fungal species, but also to the general scarcity of available methods that take into account differences in the mechanistic cost of the transitions from one state to the other across a phylogeny (discussed in Rydholm et al. 2007; Ekman, Andersen and Wedin 2008; Skinner 2010). In contrast, many studies have made inferences about mating-system transitions based on comparisons of the physical organization of mat genes from heterothallic (self-sterile) and homothallic (self-fertile) species of fungi. These studies commonly compare a limited number of phylogenetically distant species, which may be the reason for inconclusive results on the evolution of fungal mating systems (e.g., in Aspergilli; Galagan et al. 2005; Rydholm et al. 2007). Moreover, although a few studies so far have indicated a polyphyletic origin of self-fertility in fungi (e.g., Inderbitzin et al. 2005; Nygren et al. 2011), no study to date has delved into the genetic basis of this reproductive versatility.

The filamentous ascomycete genus Neurospora provides a suitable model system for the study of mating-system evolution. A primary reason is that the genus contains species with multiple distinct mating systems; heterothallism, homothallism, and pseudohomothallism (partial self-fertility) (Davis 2000). Heterothallism involves fusion between haploid...
nuclei of opposite mating-types (determined by alternative idiomorphs at the mating-type locus [Glass et al. 1988]) followed by meiosis. Homothallic species do not have a matingtype and can go through meiosis without a partner. A second reason for the study of mating system evolution in *Neurospora* is that homothallic species in this genus allow one to test the assumption of the dead-end theory (Takebayashi and Morrell 2001) without having to consider the effects of occasional outcrossing. We argue that homothallic *Neurospora* can be assumed to reproduce by obligate intra-haploid mating since these species appear to lack the morphological structures important for outcrossing, that is, the "male" fertilizing agents, the conidia, and the "female" receptive hyphae, the trichogynes (Howe and Page 1963; Perkins 1987). Third, a study on the evolution of mating types in *Neurospora* is now facilitated by the recent availability of a comprehensive phylogenetic framework for the genus. Notably, ancestral state reconstruction analysis based on this phylogeny indicated a homothallic ancestor for *Neurospora* (Nygren et al. 2011), although the authors argue that this is not likely the case. Mating-type genes have been sequenced from a number of homothallic species (e.g., Wik et al. 2008), but functional studies have mainly focused on *N. africana* (Glass et al. 1990; Glass and Smith 1994). In certain species, such as *N. terricola*, alternative mating-type genes have been found in the same haploid nucleus, but their close physical linkage was not proven (Beatty et al. 1994). Overall, with the exception of the heterothallic *N. crassa*, which represents a long-established model in fungal biology (Davis 2000; Selker 2011), the physical organization of the *mat* locus among members of the *Neurospora* genus remains unknown.

In this study, we address the question of the ancestry of fungal mating systems, and of the genetic basis of self-fertility, using the *Neurospora* phylogenetic framework. We sequenced the *mat* locus from four homothallic species belonging to distinct phylogenetic clades, as well as from a heterothallic species belonging to a distinct clade than *N. crassa*. We explored different models for the transitions between sexual reproductive modes, so as to understand their relative mechanistic ease and shed more light on the high versatility in reproductive modes observed in fungi.

**Materials and Methods**

**Strains and Culture Conditions**

All strains used in this study were obtained from the Fungal Genetics Stock Center (FGSC, University of Missouri, Kansas City). The strains belong to the homothallic species *N. africana* (strain ID: FGSC 1740), *N. sublineolata* (FGSC 5508), *N. pannonica* (FGSC 7221), and *N. terricola* (FGSC 1889). In addition, we used the heterothallic strains FGSC 8238 (mating type A) and FGSC 8239 (mating type a). The species status of these strains has never been investigated, and they were deposited to FGSC under the name *Gelasinospora uncertaina*. Based on observations of a high sequence similarity between them (Johannesson H, unpublished data), we assume they belong to the same species. Furthermore, without making any taxonomic judgment, we followed the rationale by Garcia and colleagues (2004) to merge the genera *Neurospora* and *Gelasinospora*, and thus, for simplicity, refer to FGSC 8238 and FGSC 8239 herein as belonging to one heterothallic species of *Neurospora: Neurospora uncertaina* (FGSC 8238: *N. uncertaina* A, FGSC 8239: *N. uncertaina* a). All strains were maintained in SC solid medium (Westergaard and Mitchell 1947) plates at 25° under constant dark. For DNA extractions, strains were cultured for 2–3 days in liquid 3% malt extract (Sigma Aldrich). For RNA extractions, strains grown on SC solid medium plates were harvested during three developmental stages: “early mating” (formation of immature protoperithecia), “late mating” (after shooting of mature ascospores), and “vegetative growth” (mycelia grown for 4–6 days in liquid SC medium).

**Nucleic Acid Isolation, PCR, and Sequencing**

Genomic DNA was extracted from 2-days-old fungal mycelium using the Easy-DNA Kit (Invitrogen). We followed the protocol for “small scale isolation of DNA from yeast cells” (provided by the kit) with the following modifications: for the preparation of cells we homogenized the tissue with a plastic pestle in 200 μl 2% CTAB solution (Johannesson and Stenlid 1999) and incubated at 65°C for 1 h, whereafter we followed the manufacturer’s instructions for DNA isolation. RNA was extracted using the RNeasy Plant mini kit (Qiagen) according to the manufacturer’s instructions for filamentous fungi. The RNA was treated with DNase (Fermentas) and was translated into cDNA using the iScriptTM Select cDNA synthesis kit (BioRad, Hercules, CA), following manufacturer’s instructions. PCR was performed using the Expand High Fidelity PCR system or Expand Long-Range dNTPack system (Roche) for amplification of regions shorter or >7 kb, respectively, following the manufacturer’s specifications. The primers used to perform PCR and sequencing were designed using Primer 3 (Rozen and Skaletsky 2000) and are listed in supplementary table S1, Supplementary Material online. Annealing temperatures and elongation times used for reactions are specified in supplementary table S2, Supplementary Material online. Reactions were performed on either an Eppendorf Mastercycler or a BioRad MyCycler. PCR products were cleaned using the ExoSAP-IT protocol (Affymetrix). Single-coverage sequencing via primer walks was performed on an ABI 31730 sequencer (Applied Biosystems) using the BigDye Terminator v3.1 cycle sequencing kit, either at Uppsala University or at Iowa State University (Ames, Iowa). Sequences were assembled and edited using Sequencer v 4.7 (Gene Codes) or Lasergene (DNASTar).

**Characterization of the *mat* Locus in Neurospora Strains**

The *mat* locus in *Neurospora crassa* has two alternative allelic sequences, denoted idiomorphs (Glass et al. 1988): *mat A* and *mat a*. The *mat A* idiomorph contains three genes; *mat A-1, mat A-2, and mat A-3*, while the *mat a* idiomorph contains one gene; *mat a-1*. The *mat* locus is flanked by the genes *sla2* and *apn2* (Butler 2007). To characterize the *mat* locus in the
four homothallic species N. africana, N. sublineolata, N. pannonica, and N. terricola, we first searched corresponding draft genome assemblies (Gioti A, Stajich JE, Johannesson H, unpublished data) for published mat genes (Beatty et al. 1994; Poggeler 1999; Dettman et al. 2001) and genes sla2 and apn2 (GenBank accession numbers: XP_964240, XP_964325.2, respectively) using BlastN (Altschul et al. 1997). In cases where mat genes were found in different scaffolds of the genome assemblies, linkage between scaffolds was determined with PCR and sequencing using the primers and PCR conditions given in supplementary tables S1 and S2, Supplementary Material online. To identify the mat locus in the two strains of the heterothallic N. uncertaina, for which genome assemblies are not available, we undertook a PCR and sequencing approach: briefly, we designed primers for the amplification of the sla2 - apn2 region (supplementary tables S1, Supplementary Material online) based on the alignment of the genes’ orthologs in Neurospora species, and sequenced the corresponding fragments of ∼10 kb using primer walking.

Gene and Repeat Element Annotations

We predicted gene structures using MAKER (Cantarel et al. 2008) based on a combination of protein-based gene predictions, SNAP (Korf 2004), Augustus (Stanke and Waack 2003), and GeneMark.hmm (Ter-Hovhannisyan et al. 2008). We manually curated these annotations for mat genes using Artemis (Rutherford et al. 2000). Curations on exon/intron junctions, start and stop codons were based on previously published data from RNA sequencing of mat transcripts (Wik et al. 2008). RepeatModeler version open-1.0.5 (http://www.repeatmasker.org/Repeat_Modeler.html) and LTRharvest (Ellinghaus, Kurtz and Willhoft 2008) were used for de novo identification of repeats on the draft assembly scaffolds of homothallic genomes (Gioti A, Stajich JE, Johannesson H, unpublished data) and on published heterothallic genomes of N. crassa (Galagan et al. 2003), N. discreta (http://genome.jgi-psf.org/Neudi1/) and N. tetrasperma (http://genome.jgi-psf.org/Neute1/). We used these novel, species-specific repeats (988 in total, available at http://fungalgenomes.org/public/neurospora/data/repeatlib/Gioti_neurospora_repeats.renamed.lib) and the 430 existing fungal repeat families from RepBase release 15.02 (Jurka et al. 2005) within RepeatMasker (version open-3.3.0) to identify repetitive sequences on the mat loci of all studied species. Inspection of the MAKER predicted genes also revealed the nsubGypsy element. To assign this element in the long-terminal repeat (LTR)/Gypsy class, we used the translated ORF for a BlastP (Altschul et al. 1997) search against the conserved domain database (CDD) and identified similarities from 5' to 3', to a reverse transcriptase RT_LTR, a RNase_HL_RT_Ty3, an integrase and a chromodomain (supplementary fig. S4, Supplementary Material online). The sequences of all genes and repeats of the mat locus are publicly available (accession numbers HE600064, HE600065, HE600066, HE600067, HE600068, HE600069, HE600070, for N. uncertaina A, N. uncertaina a, N. africana, N. pannonica, N. sublineolata a, N. sublineolata A, and N. terricola, respectively).

We calculated the mutational silencing due to repeat induced point mutation (RIP) with the Composite Index as defined in (Lewis et al. 2009); the index was calculated on 200 bp windows and summarized over all windows using the Perl scripts RIP_index_calculation.pl and RIP_genome_summary.pl available at: https://github.com/hyphaltip/fungaltools/.

mat Locus Alignments and Unequal Crossover Analysis

For alignments of the mat locus, we used sequence data generated in this study and the following published sequences: (1) the Neurospora crassa (FGSC OR74) sequence of the mat A strain, identified within the chromosome I contig7.3 (GenBank accession: AABX02000003); (2) the Neurospora crassa (FGSC 4200) partial sequence of the mat a strain (GenBank accession number MS4787); (3) the Neurospora discreta (FGSC 8579) sequence of the mat A strain, identified using BlastN in the JGI genome database (http://genome.jgi-psf.org/Neudi1/); and (4) the Sordaria macrospora (k-hell strain) sequence of the mat locus, identified using BlastN (Altschul et al. 1997) against a local database of the genome (Nowrousian et al. 2010). To identify the genomic boundaries of the idiomorphs in N. uncertaina and the four homothallic species using the program progressiveMAUVE (Darling et al. 2010) with default parameters. This program allows for the identification of conserved blocks of synteny in coding and noncoding areas. Local identity between intergenic regions of mat idiomorphs of N. uncertaina was identified using pair-wise BlastN alignments with default parameters, except for word size, which was set to 7.

Results

Sequencing of the mat Locus in Neurospora

We identified all four mat genes (mat A-1, mat A-2, mat A-3, mat a-1) and the mat-a locus flanking genes (sla2 and apn2) in the draft genomes of the homothallic species N. sublineolata and N. pannonica. We failed to identify mat A-3 in the genome of N. terricola and mat a-1 in the genome of N. africana. This result is consistent with previous reports (Beatty et al. 1994; Glass and Smith 1994; Wik et al. 2008). In the N. africana genome, the entire mat locus (mat A genes, sla2 and apn2) was contained in a single scaffold. In genome assemblies of the other homothallic species, mat genes were found in different scaffolds and linkage between them was determined with PCR and sequencing (supplementary tables S1 and S2, Supplementary Material online), confirming the structures shown in fig. 1. In sum, our data show that the structure of the mat locus in N. africana is highly similar to that of the mat A idiomorph of the heterothallic N. crassa, whereas N. pannonica and N. terricola share in common the juxtaposition of mat A and mat a genes (fig. 1), similar to the closely related homothallic species Sordaria macrospora (fig. 1). An unexpected organization
Fig. 1. Gene and transposable element organization within the mat locus of seven Neurospora species and the homothallic Sordaria macrospora. The species phylogeny is adapted from Nygren et al. (2011) and assumes a heterothallic ancestor of Neurospora based on the result presented in this study. Red and blue branches delineate homothallic and heterothallic lineages of Neurospora respectively, whereas gray color indicates branches of unknown mating system within the Sordaria genus. For each of the homothallic species, and the heterothallic N. discreta, one haploid strain per species was investigated, while for the heterothallic N. crassa and N. uncertaina, two strains per species were investigated, one of each mating type. Each haploid genome is indicated with a box, and mating type is indicated in parentheses. Note that in N. sublineolata, mat genes are found in two different chromosomes within the same haploid nucleus. The size of the mat locus, defined as the distance between the stop codon of the left flank gene (sla2/NCU05534) and the start codon of the right flank gene (apn2/NCU05535), is noted next to each species. White and colored boxes along the solid gray line represent genes, with arrows (< and >) indicating the orientation of transcription. Asterisks indicate incompletely sequenced genes or loci. All genes are indicated according to their *N. crassa* orthologs, with mat A-1, mat A-2, mat A-3, and mat a-1 noted as A1, A2, A3, and a1, respectively. Predicted transposable elements are depicted with colored boxes: green boxes represent distinct fragments of the *npmlTR* element, except for the green box noted as *npmlTR* in N. pannonica, which represents the full copy of *npmlTR*, including the target site duplication (TSD) sites on both flanks. Note that fragments of *npmlTR* that overlap with mat a-1 gene are not shown. Brown boxes represent unknown repetitive elements and the fuchsia box represents the *nsubGypsy* element. Only repeat elements of size larger than 300 bp are shown, unless they form part of a larger element, as in *npmlTR*. Except for the size of the flanking genes (sla2/NCU05534 and apn2/NCU05535), the figure is drawn to scale.

In the heterothallic species *N. uncertaina*, we identified orthologs of mat A-1, mat A-2 and mat A-3 flanked by orthologs of genes sla2 and apn2 in the mating type A strain, and the ortholog of mat a-1 flanked by sla2 and apn2 orthologs in the mating type a strain (fig. 1). The sequences of mat genes from the heterothallic species *N. uncertaina* were aligned with previously published sequences of mat genes from heterothallic species, which all belong to the same clade as *N. crassa* (Wiket al. 2008): We identified very few substitutions specifically occurring in the ORFs of mat genes from *N. uncertaina*, and we did not identify any in-frame stop codons or indels introducing frame-shifts (data not shown).

Conservation of the mat Locus among Neurospora Heterothallic Species

With the aim to resolve the question of the ancestral mating system of *Neurospora* species, we compared the architecture of the mat locus among species with different mating systems that belong to distinct phylogenetic groups. The mat locus

<table>
<thead>
<tr>
<th>Taxon</th>
<th>Size (kb)</th>
<th>Structure</th>
</tr>
</thead>
<tbody>
<tr>
<td>N. crassa (A)</td>
<td>13.8</td>
<td>sla2&gt;</td>
</tr>
<tr>
<td>N. crassa (a)</td>
<td>6</td>
<td>a1&gt;</td>
</tr>
<tr>
<td>N. discreta (A)</td>
<td>10.8</td>
<td>sla2&gt;</td>
</tr>
<tr>
<td>N. africana</td>
<td>9.8</td>
<td>sla2&gt;</td>
</tr>
<tr>
<td>N. sublineolata</td>
<td>13.3</td>
<td>sla2&gt;</td>
</tr>
<tr>
<td>N. pannonica</td>
<td>16.5</td>
<td>sla2&gt;</td>
</tr>
<tr>
<td>N. uncertaina (A)</td>
<td>8.5</td>
<td>sla2&gt;</td>
</tr>
<tr>
<td>N. uncertaina (a)</td>
<td>9.5</td>
<td>sla2&gt;</td>
</tr>
<tr>
<td>N. terricola</td>
<td>15.3</td>
<td>sla2&gt;</td>
</tr>
<tr>
<td>S. macrospora</td>
<td>11.4</td>
<td>sla2&gt;</td>
</tr>
</tbody>
</table>

The mat locus was identified in *N. sublineolata*, previously only reported in Eurotiomycetes and one Dothideomycete species (Yun et al. 1999; Galagan et al. 2005; Rydhom et al. 2007; Poggeler et al. 2011): Genome assemblies and PCR confirmations show that the mat A and mat a genes are not linked; mat a-1 is flanked by sla2 and apn2, while the mat A genes are flanked by the orthologs of *N. crassa* genes NCU05534 and NCU05535 (fig. 1). Although the draft *N. sublineolata* assembly does not provide chromosome-level resolution, our assembly data suggests that if mat A and mat a genes were linked, the distance between them is a minimum of 50–220 kb, depending on orientation (data not shown). Because this disagrees with previous Southern blot results suggesting that the mat a-1 of *N. sublineolata* is to be found ~12 kb upstream of the mat A-1 (Beatty et al. 1994), we used successive PCRs upstream of the NCU05534 ortholog to further confirm this region of our assembly (supplementary table S1, Supplementary Material online).
was defined in all species as the genomic region between flanking genes sla2 and apn2 (Butler 2007). We included in this comparison the published mat locus from N. crassa (partial sequence for the mat a strain (Staben, Yanofsky 1990), from N. discreta (mat A only available) and from S. macrospora, a homothallic species closely related to the Neurospora genus (fig. 1). All studied heterothallic species showed strong conservation with respect to the presence of coding and non-coding elements, their orientation of transcription and their relative positions. The only exception to this is N. crassa, where the mat locus contains an additional gene, AR2, closely linked to a short unknown repeat. The gene AR2 (AAC78559.1) encodes for a protein that is necessary for sexual development, with homology at its 5’ to an ATPase-IIIa-H domain, found in plasma-membrane proton-efflux P-type ATPases (Randall and Metzenberg 1998). Disregarding AR2, one consensus structure for each mating type is apparent in all heterothallic species independently of the phylogenetic clade they belong to (fig. 1). In sharp contrast, it was not possible to define one consensus structure from the four homothallic species, even among closely related species such as N. pannonica and N. sublineolata (fig. 1). The architecture of the mat locus in homothallic species varies in aspects such as presence/absence of mat genes (e.g., mat A-3 is lacking in N. terricola and mat a-1 in N. africana), co-localization of mat A and mat a genes, (N. sublineolata versus N. pannonica and N. terricola) and the presence of transposable elements (TEs; see detailed description below), N. sublineolata possessing a variety of putative TEs that were not identified in the other homothallic species (fig. 1).

The Intergenic Regions of Neurospora mat Loci Contain Two Novel Retroelements

Scanning the mat locus of all Neurospora species for the presence of repetitive sequences revealed two putative TEs, named npanLTR and nsubGypsy (fig. 1). NpanLTR and nsubGypsy have not been reported until now, and we attempted to further characterize them. The npanLTR element was predicted as an LTR element using the de novo prediction tool LTRHarvest (Ellinghaus et al. 2008), which identified two LTR flanked by 4 bp—TSD sites at both sides of the sequence. The full sequence of npanLTR was identified in the N. pannonica mat locus and found to overlap and extend past the mat a-1 gene (fig. 1), where the “right” (relative to mat a-1) LTR sequence flanked by a TSD was located. NpanLTR is 5,044 bp excluding the mat a-1 gene. Multiple truncated fragments of this sequence were found in the intergenic regions of the mat locus between the genes sla2 and mat a-1. The npanLTR element did not show any similarity to known repeat elements or protein domains; translation of the sequence in all six frames revealed numerous stop codons and the longest ORF is only 206 amino acids long. We tested whether the longest copies of npanLTR from N. uncertaina a, N. sublineolata a and N. pannonica bear a RIP signature by calculating the composite RIP index (Lewis et al. 2009): 26–27% of these copies’ sequence shows evidence for RIP. Thus, we propose that the multiple contiguous fragments of npanLTR along the mat locus of each species, as for example in N. uncertaina or N. sublineolata a (fig. 1), represent parts of the same original element, which has been gradually silenced through the action of RIP and random mutations. We also investigated whether npanLTR is expressed in N. pannonica, by RT–PCR amplifications of a 3.1 kb fragment of the element and did not find any evidence for expression during the three stages of N. pannonica growth tested, i.e., early mating, late mating and vegetative growth (supplementary fig. S3, Supplementary Material online).

The nsubGypsy element was predicted as an ORF of 3,532 bp, and is homologous with parts of two retrotransposons of the LTR—Gypsy class: Skippy-I (Anaya and Roncero 1995) on its 5’, and Pyret-I (Nakayashiki et al. 2001) on its 3’-end. Several domains characteristic of the Ty3/Gypsy superfamily (Metaviridae) were identified in nsubGypsy (supplementary fig. S4, Supplementary Material online), including a chromo domain, characteristic of a class of LTRs that are common to fungi and vertebrates and are called “chromoviruses” (Malik and Eickbush 1999; Marín and Llorens 2000). Putative active sites, histone binding sites and nucleic acid binding sites were also retrieved by sequence similarity. Furthermore, a search for the direct repeat sequences 5’-TGTTACG...CGTAACA-3’ that are highly conserved in LTRs of fungal retrotransposons of the same family (like Skippy, Boty, Cgret and Pyret: Anaya and Roncero 1995; Diolez et al. 1995; Zhu and Oudemans 2000; Nakayashiki et al. 2001) identified both of these sequences. Notably, the 5’-TGTTACG direct repeat sequence was identified in the 5’ flanking region of nsubGypsy, but overlapped the coding sequence of the ortholog of NCU05535 (fig. 1). This finding, along with the fact that we were not able to identify the characteristic gag polymerase domain (PFAM identifier: PF03732) at the 5’-end of nsubGypsy, implies that the element is no longer functional. The 3’ direct repeat sequence CGTAACA was identified at the 3’-end of nsubGypsy, at a distance of 1,628 bp from the predicted stop codon of the nsubGypsy ORF, within a region where short repeat sequences of unknown class were identified (fig. 1). Multiple stop codons were found in the translation of the sequence downstream from the predicted stop codon. Further investigation revealed that the predicted nsubGypsy element has been mutated by RIP. We estimated that 31% of the nsubGypsy sequence bears a RIP signature. RT–PCR amplifications (supplementary fig. S3, Supplementary Material online) of a 2.3 kb region corresponding to the reverse transcriptase, RNAse, and integrase domains of the element further showed that nsubGypsy is not expressed during N. sublineolata growth (for conditions tested, see Materials and Methods section).

We observed that the presence of the two TEs characterized here influences the mat locus size, i.e., the size of the region between flanking genes sla2/NCU05534 and apn2/NCU05535 genes (fig. 1). For example, the genes NCU05534 and NCU05535 are ~1,700 bp apart on the linkage group VI of N. crassa, whereas in N. sublineolata A, the distance between these genes’ orthologs is ~13.4 kb; this is due in part to the insertion of the mat A genes, but the majority of the
difference is due to the genomic fragment containing \textit{nsubGypsy} on the right flank of the gene \textit{mat A-3}. Similarly, in all the other homothallic species and in \textit{N. sublineolata} \textit{mat a}, the total size of the intergenic region between flanking genes \textit{sla2} and \textit{apn2} varies considerably, from 9.8 to 16.5 kb (fig. 1), and directly depends on the size of \textit{npanLTR} fragments. Moreover, in the herothallic species, the size of the idiomorph, defined as the region of no detectable homology between opposite mating types, is also dependent on the presence and size of \textit{npanLTR} repeats. Genomic alignments of the \textit{mat} locus (supplementary fig. S1, Supplementary Material online) suggest that the idiomorph in \textit{N. uncertaina} is almost 7 kb, in contrast to the idiomorph size of \textit{N. crassa} \textit{a} strains, which does not exceed 4 kb (Staben and Yanofsky 1990).

Models for Transition from Heterothallism to Homothallism

An ancestral state reconstruction analysis on the \textit{Neurospora} phylogeny has previously indicated that a heterothallic ancestor would be plausible for the genus if the transition rates from heterothallism to homothallism were approximately 50 times higher than the opposite direction rates (Nygren et al. 2011). This prompted us to further explore potential models for such transitions. Based on the \textit{mat} locus architecture of heterothallic and homothallic species of \textit{Neurospora} (fig. 1), we can propose two candidate mechanisms (models) for transitions from heterothallic to homothallic mating systems in this genus: TE-mediated translocations (1) and unequal crossovers (2).

A TE-mediated Translocation Explains the \textit{N. sublineolata} Mat Locus Structure

In the genomic fragment between the genes \textit{mat A-3} and the ortholog of NCU05535 in \textit{N. sublineolata} we identified the repetitive element \textit{nsubGypsy}, flanked by three short predicted repeat elements of unknown class (fig. 1). This configuration suggests that the \textit{mat A} genes were translocated from the ancestral \textit{mat} locus/LGI to a new genomic location (here termed as LGVI, as the \textit{N. crassa} linkage group) through the mobilization of a transposon (fig. 2). This translocation could have occurred by mitotic and/or meiotic events through a \textit{mat A} intermediate (fig. 2b), implying two different origins for the regions including the genes NCU05534 and NCU05535 (\textit{mat A} or \textit{mat a}: fig. 2c).

Unequal Crossovers Explain the Structure of the Mat Locus in \textit{N. pannonica} and \textit{N. terricola}

We explored the possibility that homothallism in \textit{N. terricola} and \textit{N. pannonica}, in which alternative \textit{mat} genes co-localize, could result from an unequal crossover in a hypothetical heterothallic ancestor (fig. 3). We used the strains of both mating types of the heterothallic \textit{N. uncertaina} as proxies for heterothallic ancestors, as they belong to the heterothallic lineage that is most closely related to \textit{N. terricola} and \textit{N. pannonica} (Nygren et al. 2011 and fig.1). Alignments of the intergenic region between \textit{mat A-2} and \textit{mat A-3} of

![Figure 2](http://Jackson-University-Serials.com/MLT/MBE/2015/3220.png)
N. uncertaina A (A2/A3-intergenic) and the intergenic region between genes sla2 and mat a-1 of N. uncertaina a (sla2/a1-intergenic) revealed the presence of a sequence (s-ter = TTGTTCTC) that could explain the mat structure of N. terricola (supplementary fig. S1, red arrow, Supplementary Material online). This sequence is also found with 100% identity in the region of N. terricola where we expected the fusion of sequences from the opposite mating type strains of N. uncertaina, i.e., the region between mat genes A-2 and a-1 (supplementary fig. S1, Supplementary Material online). We observed high similarity (96% identity along 291 bp) with the N. uncertaina A2/A3-intergenic sequence on the left flank of this fusion point, whereas the right flank of the fusion point is less conserved in the N. uncertaina sla2/a1-intergenic (supplementary fig. S2A, Supplementary Material online). The location of the putative crossover junction was also successfully identified in the N. pannonica intergenic region between mat genes A-3 and a-1 (supplementary fig. S1, blue arrow, Supplementary Material online). In this region, we identified two consecutive short sequences, one of 11 bp (s-pan1 = CACCTGCTCG), also present in the A3/apn2-intergenic of N. uncertaina A, and the second of 10 bp (s-pan2 = AAACTTAATT), also present in the sla2/a1-intergenic of N. uncertaina a (supplementary fig. S1, Supplementary Material online). Again, alignment of the left flank region of the N. pannonica-intergenic to the A3/apn2-intergenic showed strong conservation close to the fusion point (99% identity along 146 bp), whereas the right flank was less conserved (supplementary fig. S2B). However, we were not able to identify a single sequence that is present in both mating-type strains of N. uncertaina and in N. pannonica, probably because N. pannonica is less closely related to N. uncertaina compared to N. terricola (fig. 1). Thus, we conclude that s-pan1 and s-pan2 are sequences of local identity in close proximity of the sequence that directly promoted the unequal crossover event. Furthermore, we investigated the possibility that the same sequence of local identity between opposite mating-type heterothallic strains would be responsible for both N. pannonica and N. terricola transitions to homothallism through a first step of unequal crossover as in N. pannonica (fig. 3) and a subsequent loss of A-3 gene in N. terricola; we were not able to identify such a sequence.

**Discussion**

**Evidence for a Heterothallic Ancestor for Neurospora**

Structural comparisons of the genomic regions conferring sexual identity along a phylogeny represent a powerful approach to study mating-system transitions over evolutionary time (Butler et al. 2004; Metin et al. 2010; Kellner et al. 2011). In this study, we unravel the genetic architecture of mating-system loci from multiple species of Neurospora by combining high-coverage sequencing with targeted long-range PCRs and sequencing by primer walks. Our study strongly suggests that the ancestral mating system in Neurospora was heterothallic. A heterothallic ancestor is supported by a number of observations, the first being the strong conservation of coding and non-coding features of the mat locus among heterothallic species that belong to distinct phylogenetic clades (N. uncertaina versus N. crassa - N. discreta). The only exception to this is the gene AR2, only present in the N. crassa mat locus (fig. 1). However, orthologs of AR2 were identified in other genomic scaffolds of species studied here (data not shown), thus the presence of AR2 in the mat locus might represent an event specific to the N. crassa clade. Second, the presence of flanking genes sla2 and apn2 in both mating types of heterothallic species makes it difficult to imagine a scenario in which heterothallic species resulted from a homothallic ancestor with a structure sla2 - A1 - A2 - A3 - a1 - apn2, but kept both flanks intact. A final argument in favor of a heterothallic ancestor comes from the observation that the conservation of synteny in the overall genomic architecture extends to the sequence level; all mat A and mat a genes of heterothallic species are highly conserved. Furthermore, the mat genes from N. uncertaina show no signs of decay, in contrast to mat genes from closely related homothallic species, where stop codons and frameshift mutations were previously reported (Wik et al. 2008). Regarding the few non-synonymous substitutions that we identified as specifically occurring in N. uncertaina mat genes (alignments not shown), sequencing of the corresponding transcripts and directed mutagenesis experiments targeting these are needed to confirm the sequence and assess their functional significance.

**Mechanistic Ease Explains Multiple Routes to Self-fertility in Neurospora**

Inspection of the genomic architecture of the mat loci of four homothallic species revealed that at least four independent transitions to self-fertility have occurred in the evolutionary history of the genus Neurospora. Our data indicate that three of these transitions (involving N. sublineolata, N. pannonica, and N. terricola) have occurred through the acquisition of both mating types into the same haploid genome (fig. 1). This finding fully agrees with genetic manipulation studies showing that homothallism is possible once genes from compatible mating-type loci are expressed in the same individual (Yun et al. 2000; Lin and Heitman 2007). For one transition (involving N. africana), we were unable to identify the route to self-fertility. In this species, we found a conserved mat A idiomorph and no mat a-1 gene. How this species, as well as closely related species from the same clade (Nygren et al. 2011) and other Ascomycetes with similar configurations (e.g., Inderbitzin et al. 2005), are able to sexually reproduce without the need of a partner of mating type a (Glass et al. 1988; Glass et al. 1990; Glass and Smith 1994) remains thus an intriguing question (discussed in Lin and Heitman 2007). One may speculate that these particular species do not need the mat a genes for controlling cell identity.

The inclusion of a closely related heterothallic species in this study, the genomic alignment of its mat locus with homothallic species and the annotation of TEs in all these loci indicated two potential genetic mechanisms for the transitions to homothallism: retrotransposition and unequal crossovers (figs. 2 and 3). We propose a retrotransposition event...
for *N. sublineolata*: Our assembly, PCR and sequencing results indicated that the genomic region containing the mat A genes in *N. sublineolata* is present in another genomic location (figs. 1 and 2). Previous cytological studies have indicated that the number, and to a less studied extent the identity, of chromosomes is conserved in homothallic species compared to *N. crassa* (Raju 1978; Beatty, Smith and Glass 1994). Thus, based on overall conservation of synteny, we may assume that mat A genes are located on a different chromosome than the mat a-1 gene in *N. sublineolata*, most probably on the equivalent of the *N. crassa* linkage group VI. To unequivocally confirm this hypothesis, separation of chromosomes by pulsed-field gel electrophoresis, followed by hybridizations with specific probes designed on the genomic sequence of the mat locus of *N. sublineolata*, would be needed.

The presence of the transposon *nsubGypsy* in the immediate proximity of the mat A-3 gene suggests that this event has occurred through the mobilization of a retrotransposon. Retrotransposon-like sequences were identified in the mat locus of other Ascomycete fungi, and proposed to be associated with the transposition event (Rydholm, Dyer and Lutzoni 2007; Poggeler et al. 2011). In all these studies, including ours, however, the repeats were found only in homothallic species or in heterothallic species representing hypothetical ancestors, but not in both. Unfortunately, *N. sublineolata* belongs to a phylogenetic clade that does not contain heterothallic species (fig. 1), so we lack the means to test the presence of *nsubGypsy* in a closely related heterothallic species, which would provide strongest evidence for this model.

**Fig. 3.** Unequal crossover models for evolutionary transitions from heterothallic ancestors to homothallic species *N. terricola* (*a*) and *N. pannonica* (*b*). Each gray box represents a haploid nucleus of a hypothetical heterothallic ancestor, of mating type A (white backbone) or a (black backbone) and white boxes depict haploid nuclei of derived homothallic species *N. terricola* (*a*) and *N. pannonica* (*b*). The organization of genes and repeat elements within the nuclei is simplified from figure 1 and is not drawn to scale. All genes are indicated according to their *N. crassa* orthologs, with mat A-1, mat A-2, mat A-3, and mat a-1 noted as A1, A2, A3, and a1, respectively. The asterisk on *npanLTR* indicates a truncated form of the element, regarding original size. The regions of the hypothetical heterothallic ancestor where the crossover occurred are noted with dotted lines; the resulting structures of *N. pannonica* and *N. terricola* are represented as chimeric sequences containing parts of both idiomorphs from this ancestor. The unequal crossover events are expected to result in two meiotic segregants, but the second, containing only *sla2* and *apn2* genes, is not presented here.
Unequal crossovers are the second potential genetic mechanism allowing self-fertility, and this mechanism has been long proposed to promote homothallism through the fusion of mating-type genes on the same haploid genome (Poggesler et al. 1997; Yun et al. 1999; Inderbitzin et al. 2005; Ramirez-Prado et al. 2008; Poggesler et al. 2011). Our study is however only the second study after Yun et al. (1999) that provides unequivocal evidence for this mechanism, through (1) identification of the sequences that are directly (s-ter) or indirectly (s-pan1, s-pan2) associated with crossovers in a heterothallic species, (2) identification of the same sequences on the homothallic “derived” species, and (3) conservation of the right and left flanking sequences at the fusion point in the homothallic species. It is noteworthy that the sequence s-pan2 harbors the CTT motif, which is the cleavage site of the eukaryotic topoisomerase I, involved in non-homologous recombination (Bullock et al. 1985). The local identity “pockets” promoting unequal crossovers in Neurospora are different in position and sequence base composition between the two closely related species N. pannonica and N. terricola, providing a good example of distinct shifts to self-fertility occurring in short evolutionary timescales. Notably, except s-ter in N. uncertaina mating-type A strain, which lies in the short intergenic region mat A2/A3 (supplementary fig. S1, Supplementary Material online), all other instances of the sequences s-ter, s-pan1 and s-pan2 share a common position on the npanLTR element and, in one case on its very close proximity, which is probably also part of the original element (s-ter in N. terricola, supplementary fig. S1, Supplementary Material online). These findings suggest that npanLTR could be associated with crossover events. The npanLTR repetitive sequence could provide a favorable environment for ectopic recombination, as was proposed for retrotransposons in yeast (Mieczkowski et al. 2006), and for fungal repeats in general (Lengeler et al. 2002; Li et al. 2010). In N. pannonica, for which our evidence for unequal crossover is less strong, we cannot exclude the additional possibility that the route to self-fertility might have implicated a transposition of the npanLTR-a-1 fragment from a hypothetical mat a ancestor to a mat A isolate. This possibility is supported by the observation that mat a-1 is flanked by the full-length copy of npanLTR in this species.

The finding that TEs on the mat locus of Neurospora have the potential to mediate transitions to self-fertility through a variety of mechanisms resolves the apparent contradiction with results from an ancestral state reconstruction analysis in Neurospora (Nygren et al. 2011). This study did not take into account a difference in rates between transitions from heterothallism to homothallism and homothallism to heterothallism. Although transitions from homothallic ancestors to heterothallic species might be more parsimonious, there is currently no conceivable mechanism proposed for such transitions. In contrast, we have shown that transitions in the opposite direction are feasible. As few as 4 bp of local identity have been associated with the occurrence of crossover events (Inderbitzin et al. 2005), and thus, the mechanistic “ease" of such events during meiosis is expected to be high. Transposition, the second mechanism proposed here for transitions to self-fertility, is a common mechanism allowing genomic rearrangements in all kingdoms, although our current data in Neurospora do not allow us to conclude with certainty that transposition has occurred.

Based on the species tree of Neurospora (fig. 1), we expect to identify at least three more transitions to homothallism. The sequencing of the mat locus of additional homothallic species of different clades will further illuminate the mechanisms behind self-fertility, as we cannot exclude that the observed rearrangements on the mat locus of homothallic species may have occurred after the evolution of selfing through neutral processes, similarly to the degeneration of certain mat genes in homothallic species (Wik et al. 2008). Further studies may also reveal whether there is a selective advantage of being homothallic, for example by providing reproductive assurance (Jain 1976; Aanen and Hoekstra 2007). Such ecological features could provide an additional explanation for the numerous switches from heterothallism to homothallism in the evolutionary history of Neurospora.

Repetitive Elements Have Shaped the mat Locus Architecture

We have identified on the mat locus of Neurospora two previously uncharacterized retrotransposable elements, npan1TR and nsubGypsy. The sequence of npan1TR has no similarity to known retrotransposon protein domains, most likely due to this being a non-autonomous element that has accumulated multiple stop codons. We cannot exclude that this TE is a false positive of our detection method; however, it contains the hallmarks of an LTR with long tandem repeats and TSDs, which suggest that the element has been transposed at some point in its evolutionary history. The nsubGypsy sequence represents an LTR retrotransposon of the Ty3/Gypsy superfamily. Another element of the same superfamily, DAB1, has been identified in N. crassa (Bibbins et al. 1998), but no sequence similarity could be identified between nsubGypsy and DAB1, indicating nsubGypsy represents a novel retrotransposon of the Ty3/Gypsy superfamily. When and how this retro-element was acquired in a hypothetical heterothallic strain of mating-type A (fig. 2) remains unknown. Although nsubGypsy is present elsewhere in the genomes of all Neurospora species studied here (Gioti A et al., unpublished data), its unique presence on the mat locus of N. sublineolata suggests that the event that led to its current localization might be specific to the phylogenetic clade that contains N. sublineolata (fig. 1). In contrast, copies of the npan1TR element are present in all Neurospora and in the S. macrospora mat locus (except for N. sublineolata A, where mat A genes have been translocated to a separate chromosome). A parsimonious interpretation of this pattern implies that npan1TR was associated with the mat locus at the origin of the Sordariaceae.

RT–PCRs failed to detect expressed transcripts from either nsubGypsy or npan1TR under all tested conditions. The lack of transcripts, combined with the presence of stop codons in npan1TR, provides strong evidence that the two elements are silenced and no longer functional. We hypothesized and
confirmed that this pattern of silencing is mainly due to the action of RIP, a fungal genome defence mechanism first discovered in *Neurospora crassa* (Selker 1990) and which has had profound consequences for the evolution of its genome (Galagan et al. 2003). Alteration of the sequence composition of *npanLTR* by RIP silencing could further explain the somewhat reduced conservation of the right flanks of the crossover junction point that we observed in local alignments (supplementary fig. S2, Supplementary Material online). Regarding *nsubGypsy*, it is likely that this element has been mutated by RIP after initial inactivation during its insertion on LGVI. This could explain the absence of the *gag* domain in the *nsubGypsy* sequence.

Taken together, our study has revealed important roles for *npanLTR* and *nsubGypsy* in the history of the *mat* locus: We suggest that both retroelements have contributed to the shifts from heterothallic ancestors to homothallic species in the *Neurospora* genus, *nsubGypsy* by direct transposition of neighboring genes into a different chromosome and *npanLTR* by facilitating unequal crossovers between unrelated intergenic regions of opposite mating types. In addition to insertional mutagenesis, TEs are well known to contribute to chromosomal rearrangements by facilitating ectopic recombination (Wessler 2006). The degree of conservation of *npanLTR* further affects not only the overall size of the *mat* locus (for example, see homothallic species, fig. 1), but can also influence the patterns of recombination between opposite mating types and thus, the size of the idiomorphs in heterothallic species (supplementary fig. S1, Supplementary Material online). Moreover, we observed *npanLTR* within the immediate S’ untranslated region (UTR region) of the *mat a-1* gene in all *Neurospora* species studied here and in S. macrospora, which implies that gene expression of *mat a-1* might be affected by the presence of this element. The gene *mat a-1* is expressed during synthetic media growth of homothallic and fungi outside of Dikarya (Lengeler et al. 2002; Rydholt et al. 2007; Li et al. 2010; Zaffarano et al. 2010; Idnurm 2011; Poggeler et al. 2011), and their presence has been attributed to a neutral accumulation. This is the first time where TEs are proposed to be implicated in the mechanism of transition from heterothallic to homothallic reproductive mode. We suggest that transitions to self-fertility mediated by TEs might be more widespread than previously thought and can explain the shifts in reproductive mode commonly observed within the same genus in several fungi.

**Supplementary Material**

Supplementary tables S1 and S2 and figures S1–S4 are available at Molecular Biology and Evolution online (http://www.mbe.oxfordjournals.org/).

**Acknowledgments**

We would like to thank Kristiina Nygren for inspiring discussions, Holger Schielczeth for statistical advice, and Jerzy Jurka for assistance in curation of repetitive elements. The editor and four anonymous reviewers are acknowledged for their constructive comments on the submitted article. This work was supported by Carl Tryggers Stiftelse Foundation and the Nilsson-Ehle foundation (to A.G.); Swedish Foundation for International Cooperation in Research and Higher Education (to A.A.M.); the Swedish Research Council (to H.J.); initial complement funds from the College of Natural and Agricultural Sciences at the University of California, Riverside (to J.E.S.).

**References**


Mieczkowski PA, Lemoine FJ, Petes TD. 2006. Recombination between retrotransposons as a source of chromosome rearrangements in the yeast Saccharomyces cerevisiae. DNA Repair 5:1010–1020.

Magnaporthe grisea contains an extra domain between the nucleocapsid and protease domains. Nucleic Acids Res. 29:4106–4113.


