Massive Changes in Genome Architecture Accompany the Transition to Self-Fertility in the Filamentous Fungus *Neurospora tetrasperma*

Christopher E. Ellison,*[,1] Jason E. Stajich,[†] David J. Jacobson,* Donald O. Natvig,[‡] Alla Lapidus,[§,2] Brian Foster,[§] Andrea Aerts,[§] Robert Riley,[§] Erika A. Lindquist,[§] Igor V. Grigoriev,[§] and John W. Taylor*

*Department of Plant and Microbial Biology, University of California, Berkeley, California 94720, [†]Department of Plant Pathology and Microbiology, University of California, Riverside, California 92521, [‡]Department of Biology, University of New Mexico, Albuquerque, New Mexico 87131, and [§]Department of Energy Joint Genome Institute, Walnut Creek, California 94598

ABSTRACT A large region of suppressed recombination surrounds the sex-determining locus of the self-fertile fungus *Neurospora tetrasperma*. This region encompasses nearly one-fifth of the *N. tetrasperma* genome and suppression of recombination is necessary for self-fertility. The similarity of the *N. tetrasperma* mating chromosome to plant and animal sex chromosomes and its recent origin (<5 MYA), combined with a long history of genetic and cytological research, make this fungus an ideal model for studying the evolutionary consequences of suppressed recombination. Here we compare genome sequences from two *N. tetrasperma* strains of opposite mating type to determine whether structural rearrangements are associated with the nonrecombining region and to examine the effect of suppressed recombination for the evolution of the genes within it. We find a series of three inversions encompassing the majority of the region of suppressed recombination and provide evidence for two different types of rearrangement mechanisms: the recently proposed mechanism of inversion via staggered single-strand breaks as well as ectopic recombination between transposable elements. In addition, we show that the *N. tetrasperma mat a* mating-type region appears to be accumulating deleterious substitutions at a faster rate than the other mating type (*mat A*) and thus may be in the early stages of degeneration.

THE elimination of recombination can have a dramatic effect on the evolutionary trajectory of a genomic region. Without recombination, selection acts on linked genetic complexes rather than independent genetic elements. Theory predicts the accumulation of deleterious mutations and selfish genetic elements in the absence of recombinational purging and a reduced ability to fix adaptive mutations (Charlesworth and Charlesworth 2000; Charlesworth *et al.* 2005).

The genetic consequences of suppressed recombination have been best studied in the sex chromosomes of outcrossing eukaryotes, *e.g.*, plants, insects, and mammals, because the initial step in the formation of sex chromosomes is posited to be a cessation of recombination across a genomic region that includes the sex-determining locus (Charlesworth *et al.* 2005). The suppression of recombination across such a region will be selected for if it creates linkage between the sex-determining locus and other genes that are sexually antagonistic in that their functions are beneficial to only one of the sexes. The nonrecombining region can be formed from the spread of recombinational suppressors or structural changes to the chromosome that prevent synapsis. In either case, studies in mammals, birds, and plants have shown evidence that present-day nonrecombining regions are composed of multiple discrete blockage events that occurred at different time points in the evolutionary history of the taxon in question [termed "evolutionary strata" by Lahn and Page 1999 (Charlesworth *et al.* 2005)].

Because large, nonrecombining, sex-linked regions appear to have evolved independently across a diverse array of taxonomic groups, comparing the evolution of such regions across disparate taxa will make it possible to understand the evolutionary events associated with their formation as well as the genomic consequences of suppressed recombination.

In several species of fungi, the properties of the chromosomal region surrounding the mating-type locus have been

doi: 10.1534/genetics.111.130690
Manuscript received May 16, 2011; accepted for publication June 25, 2011
Available freely online through the author-supported open access option.
Supporting information is available online at http://www.genetics.org/content/suppl/2011/07/12/genetics.111.130690.DC1.
[1]Corresponding author: Department of Plant and Microbial Biology, 321 Koshland Hall, University of California, Berkeley, CA 94720-3102. E-mail: cellison@berkeley.edu
[2]Present address: Fox Chase Cancer Center, Philadelphia, PA 19111.

studied extensively because of their similarities to the sex chromosomes of other organisms (Fraser and Heitman 2004, 2005). The mating-type locus of *Cryptococcus neoformans* occurs within an ~100-kb region where recombination is suppressed due to multiple chromosomal rearrangements (Lengeler *et al.* 2002; Fraser *et al.* 2007). The evolutionary history of this region includes the accumulation of transposable elements as well as gene conversion, gene loss, and pseudogenization (Fraser *et al.* 2004; Metin *et al.* 2010). The mating-type region of *Ustilago hordei* resides within a 500-kb region that is characterized by suppressed recombination and chromosomal rearrangements (Lee *et al.* 1999) and the mating-type chromosomes of the fungus *Microbotryum violaceum* are heteromorphic and contain a region of suppressed recombination that has been roughly estimated to be 1000 kb in size (Votintseva and Filatov 2009). DNA polymorphism within *M. violaceum* populations has been compared between the recombining and nonrecombining portions of the genome; however, nucleotide variation within the nonrecombining region did not stand out from the rest of the genome due to overall low levels of polymorphism and high linkage disequilibrium (Votintseva and Filatov 2011), consistent with previous results showing that *M. violaceum* is predominantly selfing (Giraud *et al.* 2005).

The sex-determining locus of *Neurospora tetrasperma* (termed *mat*) is surrounded by a region of suppressed recombination that is ~7-fold larger than that of *M. violaceum* and 70-fold larger than that of *C. neoformans* (Menkis *et al.* 2008). From the work presented here, we now know that it includes ~2000 genes and spans a distance of ~7.8 Mb, which represents 80% of the mating-type chromosome and approximately one-fifth of the *N. tetrasperma* genome. The formation of the region of suppressed recombination was part of a series of evolutionary events that allowed this species to become self-fertile and it is relatively young compared to most vertebrate sex chromosomes [< ~4.5 MYA compared to ~160 MYA for the XY system in marsupial and placental mammals (Veyrunes *et al.* 2008) and the ZW system in snakes (O'Meally *et al.* 2010) and >120 MYA for the ZW system in birds (Mank and Ellegren 2007)]. Due to the large size and recent origin of the nonrecombining region, *N. tetrasperma* has emerged as an important model for the study of early sex chromosome evolution (Merino *et al.* 1996; Gallegos *et al.* 2000; Jacobson 2005; Menkis *et al.* 2008, 2010) alongside other organisms with relatively young sex chromosomes such as medaka (Kondo *et al.* 2006), papaya (Liu *et al.* 2004), and several species of *Drosophila* (Charlesworth *et al.* 2005; reviewed in Fraser and Heitman 2005).

Sex in *N. tetrasperma* is controlled by two mating-type idiomorphs, *mat a* and *mat A*. The majority of isolates collected from nature are heterokaryotic: haploid nuclei of opposite mating type can be found in a single fungal individual, allowing the individual to be self-fertile (Raju 1992). The maintenance of self-fertility via the packaging of two nuclei of opposite mating type into a single sexual spore is known as pseudohomothallism and is usually accompanied by strict regulation of recombination between the mating-type locus and the centromere. In *N. tetrasperma*, a crossover in this region during meiosis can result in two nuclei of the same, rather than opposite, mating type being packaged into the sexual spore, producing progeny that are not self-fertile (Raju and Perkins 1994; Merino *et al.* 1996; Gallegos *et al.* 2000). The suppression of recombination in this region is therefore thought to have evolved to ensure the correct packaging of nuclei of opposite mating type into the sexual spore, thereby maintaining the heterokaryotic, self-fertile condition (Merino *et al.* 1996; Gallegos *et al.* 2000).

The only other pseudohomothallic ascomycete where recombination has been studied genetically is *Podospora anserina*, which, as a close relative, has mating-type genes that share homology and gene order with those of *Neurospora* (Marcou *et al.* 1979; Raju and Perkins 1994; Coppin *et al.* 1997). This fungus exhibits a contrasting approach for regulating recombination between the mating-type locus and the centromere whereby a single obligate crossover occurs and additional crossovers are suppressed (Marcou *et al.* 1979). The mechanism by which this occurs is unknown but, as in *N. tetrasperma*, it acts to ensure the correct packaging of nuclei of opposite mating type into the sexual spore (Raju and Perkins 1994).

In *N. tetrasperma*, the heterokaryon occasionally breaks down via the production of vegetative or sexual spores containing nuclei with only one of the two mating types (Raju 1992). The haploid, homokaryotic individual that grows from such a spore can mate with another homokaryotic individual of opposite mating type to restore the heterokaryotic condition (Raju 1992). Thus, the reproductive strategy of *N. tetrasperma* is likely to include repeated rounds of selfing with an occasional outcrossing event (Powell *et al.* 2001). Repeated selfing within a heterokaryon is also supported by previous work showing that allelic differences between *mat a* and *mat A* nuclei were confined to the nonrecombining region of the mating chromosome (Merino *et al.* 1996).

A considerable body of work has built upon the initial observations of Howe and Haysman (1966) that recombination was reduced on the *N. tetrasperma* mating-type chromosome. Raju (1992) and Raju and Perkins (1994) determined the steps during ascus development that are required for correct packaging of *mat a* and *mat A* nuclei into a single ascospore. Merino *et al.* (1996) and Gallegos *et al.* (2000) confirmed that recombination is suppressed across the majority of the mating-type chromosome and Gallegos *et al.* visualized the nonrecombining interval as an anomalous unpaired region visible during pachytene. Jacobson (2005) reciprocally introgressed mating-type chromosomes between *N. tetrasperma* and *N. crassa*. His results suggested that the *N. tetrasperma mat a* chromosome may be collinear with the *N. crassa mat A* chromosome while the *N. tetrasperma mat A* chromosome may be structurally rearranged. However, he also found evidence for the existence of genetic, rather than structural, modifiers of recombination and was unable to determine the relative contributions of

these two phenomena with respect to the suppression of recombination in this region. More recently, Menkis *et al.* (2008) sequenced 35 genes spanning the mating-type chromosome from each of the two *N. tetrasperma mat a* and *mat A* strains used in this study. On the basis of sequence divergence between these genes, the authors predicted the existence of two evolutionary strata: one large pericentric region encompassing most of the chromosome and another much smaller region distal to the *mat* locus. In two other studies, Menkis *et al.* (2009) showed that *N. tetrasperma* is actually a species complex composed of nine genetically isolated lineages and found evidence of differences in the size of the nonrecombining region among these lineages (Menkis *et al.* 2010). Together, these results suggest that there may be important differences in the region of suppressed recombination around the mating-type locus among the different *N. tetrasperma* lineages (Menkis *et al.* 2008, 2009, 2010). Whittle and Johannesson (2011) and Whittle *et al.* (2011b) have investigated patterns of codon usage and nonsynonymous substitution within the region of suppressed recombination in another *N. tetrasperma* lineage (strain P4492, lineage 1), one different from the strain that was the main subject of Jacobson (2005), Menkis *et al.* (2008), and the work presented here (strain P581, lineage 6). They found evidence for relaxed purifying selection within this region in the form of substitutions from preferred to nonpreferred codons and, in a branch-specific analysis using PAML, a higher d_N/d_S ratio along the *N. tetrasperma* branch for the genes in the nonrecombining region compared to their *N. crassa* and *N. discreta* orthologs. However, these analyses assume that the location and number of nonrecombining strata in their lineage of interest (lineage 1; see Menkis *et al.* 2009) are the same as those reported for lineage 6 (Menkis *et al.* 2008), despite the previous evidence suggesting that these regions may have evolved independently (Menkis *et al.* 2008, 2009, 2010). While several other recent studies have investigated the molecular evolution of the strain we study here (Nygren *et al.* 2011; Whittle *et al.* 2011c; reviewed in Whittle *et al.* 2011a), none of these have examined the evolution and structure of the region of suppressed recombination in detail.

More than 40 years of genetic and cytological analyses involving *N. tetrasperma*, combined with the open question of the degree to which changes in chromosome structure are involved in the suppression of recombination, made this organism an ideal candidate for a genome sequencing project. This sequencing project was undertaken by the Joint Genome Institute and involved the sequencing and assembly of two genomes: the haploid *mat A* and *mat a* strains derived from a single heterokaryotic *N. tetrasperma* isolate. These are the same strains studied in Jacobson (2005) and Menkis *et al.* (2008) and here we use their genome assemblies to investigate the mechanisms of recombination suppression as well as the evolutionary consequences for the genes residing within the nonrecombining region. We find that the majority of the region of suppressed recombination is covered by several large chromosomal rearrangements. Additionally, we show that the young evolutionary stratum identified by Menkis *et al.* (2008) is located in a region where the two mating chromosomes are collinear and identify an additional stratum, created by an inversion, on the opposite end of the chromosome. We propose a model for the sequence of events and mechanisms of rearrangement that produced the current orientation and show that the region of suppressed recombination within the *mat a* strain appears to be in the early stages of degeneration.

Materials and Methods

Genome sequencing and assembly

Both *N. tetrasperma* strains (FGSC 2508 *mat A* and FGSC 2509 *mat a*) were sequenced using a hybrid approach on Roche 454 pyrosequencing and Sanger platforms (Supporting Information, Table S3 and Table S4) and assembled with Newbler. The sequencing projects have been deposited in GenBank under accession nos. AFBT00000000 (*mat A*) and AFCY00000000 (*mat a*). The *N. tetrasperma mat A* assembly was postprocessed to close gaps *in silico* using JGI gapResolution software for the entire genome and targeted finishing for the mating-type chromosome. Statistics of both assemblies are summarized in Table S2.

Genome annotation

Both *N. tetrasperma* assemblies were annotated using the JGI annotation pipeline with results deposited to the integrated fungal resource MycoCosm (http://jgi.doe.gov/fungi) for further analysis. Genome assembly scaffolds were masked using RepeatMasker (Smit *et al.* 1996–2010) and tRNAs were predicted using tRNAscan-SE (Lowe and Eddy 1997). Several gene predictors were used on the repeat-masked assembly: (i) *ab initio* FGENESH (Salamov and Solovyev 2000) and GeneMark (Isono *et al.* 1994), (ii) homology-based FGENESH+ and Genewise (Birney and Durbin 2000) seeded by BLASTx alignments against GenBank's database of nonredundant proteins, and (iii) direct mapping of expressed sequence tag (EST)-derived full-length genes to genome assembly. Genewise models were extended where possible, using scaffold data to find start and stop codons. EST BLAT alignments (Kent 2002) were used to extend, verify, and complete the predicted gene models. From the resulting set of models, a nonredundant representative set of best models was selected (Table S6).

All predicted gene models were functionally annotated using SignalP (Nielsen *et al.* 1997); TMHMM (Melen *et al.* 2003); InterProScan (Zdobnov and Apweiler 2001); BLASTp (Altschul *et al.* 1990) against nr; and hardware-accelerated double-affine Smith–Waterman alignments (deCypherSW; http://www.timelogic.com/catalog/758/decyphersw) against SwissProt (Boeckmann *et al.* 2003), KEGG (Kanehisa *et al.* 2008), and KOG (Koonin *et al.* 2004). KEGG hits were used to assign EC numbers (Gasteiger *et al.* 2003), and Interpro and SwissProt hits were used to map GO terms (Ashburner

et al. 2000). Multigene families were predicted with the Markov clustering algorithm (MCL) (Enright *et al.* 2002) to cluster the proteins, using BLASTp alignment scores between proteins as a similarity metric.

ESTs

N. tetrasperma FGSC 2508 was grown in Vogel's liquid media (Vogel 1956) and total RNA was extracted by bead-beating in TRIzol (Invitrogen Life Science Technologies, Carlsbad, CA) with zirconia/silica beads (0.2 g, 0.5-mm diameter; Biospec Products). ESTs were sequenced using 454 pyrosequencing. The sequencing library protocol and sequence processing procedure are described in Swarbreck *et al.* (2011).

Genome synteny

Whole-genome synteny was assessed between the two *N. tetrasperma* strains as well as between each *N. tetrasperma* strain and the outgroup *N. crassa*. Dotplots were created using the program MUMMER (Kurtz *et al.* 2004) to visualize synteny while more precise identification of rearrangement breakpoints was achieved using a combination of whole-genome orthology map construction with Mercator and whole-genome alignment with MAVID (Dewey 2007).

Identification of orthologs

Orthologs between the two *N. tetrasperma* strains as well as those between each *N. tetrasperma* strain and *N. crassa* were initially identified on the basis of best-reciprocal-BLAST hits and further verified using synteny information from Mercator. This procedure resulted in the identification of a total of 7693 single-copy orthologs between the three species.

Calculation of nonsynonymous and synonymous substitutions per site (K_a and K_s)

K_a/K_s ratios were calculated for each pair of orthologs between the two *N. tetrasperma* strains as well as between each *N. tetrasperma* strain and the outgroup *N. crassa*, using the modified Yang–Nielsen method in the program KaKs_Calculator (Zhang *et al.* 2006). It has previously been shown that the way in which the total number of synonymous sites are counted can bias the calculation of K_s (Bierne and Eyre-Walker 2003). For this reason, we also calculated the K_a and K_s values for all pairwise orthologs as in Bierne and Eyre-Walker (2003) using only fourfold degenerate sites for K_s and the physical site definition for both K_a and K_s and found similar results (Figure S2). For consistency, all K_s results reported here are from the modified Yang–Nielsen method in the program KaKs_Calculator.

Identification of pseudogenes

We used the tfasty program within the FASTA sequence comparison package (Pearson *et al.* 1997) to perform translated searches (allowing for frameshifts and premature stop codons) against the *N. tetrasperma* genome sequence. As queries, we used the subset of *N. crassa* proteins from within the genomic region of suppressed recombination for which we were unable to find *N. tetrasperma* orthologs. We used a custom Perl script to parse the results to identify matches that showed frameshifts and/or nonsense mutations. We additionally extracted the genomic sequence for candidate pseudogenes and used the program Exonerate (Slater and Birney 2005) to create genomic DNA/protein sequence alignments to confirm the presence of nonsense mutations or frameshifts.

Codon adaptation index

We used the *N. tetrasperma mat A* ESTs and *N. crassa* Illumina RNA-Seq data (Ellison *et al.* 2011) to identify the top 100 most highly expressed genes in each species. We used these genes to compute codon usage tables for each species using the *cusp* application in the EMBOSS package (Rice *et al.* 2000). We then used the EMBOSS application *cai* to calculate the codon adaptation index for each *N. tetrasperma* and *N. crassa* gene. All codon adaptation index (CAI) analyses were performed twice, once with the *N. tetrasperma* codon usage table and once with the *N. crassa* codon usage table. The results were equivalent in all cases and the reported *P*-values are those from the *N. crassa* codon usage table. We defined substitutions to preferred codons as a change to a synonymous codon whose frequency of usage in the top 100 most highly expressed genes was at least 10-fold greater than that of the ancestral codon. Similarly, we defined substitutions to unpreferred codons as a change to a synonymous codon whose frequency of usage in the top 100 most highly expressed genes was at least 10-fold less than that of the ancestral codon.

Repetitive elements

Repetitive elements were identified *de novo* using the program RepeatModeler (Smit and Hubley 2008–2010). The results of RepeatModeler were added to a fungal-specific repeat library downloaded from RepBase (Jurka 2000) and repetitive elements were identified on the basis of sequence homology using this library and the program RepeatMasker (Smit *et al.* 1996–2010). Centromeric regions in Neurospora can be easily identified because they are composed almost entirely of transposable element remnants. The numbers of repetitive elements per kilobase for the chromosomal segment within the *N. tetrasperma* nonrecombining region and in the homologous region in *N. crassa* were calculated after excluding centromeric regions so that these measures would not be confounded by differences between assemblies in the number of gaps in these repeat-dense regions.

Repeat-induced point mutation index

Repeat-induced point mutation (RIP) indexes were calculated for 500-bp sliding windows across the genome for both *N. tetrasperma* strains, using a custom Perl script and the composite index described in Lewis *et al.* (2009).

Permutation tests

CAI and K_a: This test was performed on both *N. tetrasperma* strains separately. Twenty-six genes (the number of genes within the relocated region) were drawn randomly from the set of all genes within the nonrecombining region. Each

gene's CAI was subtracted from that of its ortholog in *N. crassa* and the median of the differences was calculated for the random sample and compared to that of the real data (*i.e.*, the genes within the relocated region). This procedure was repeated 10,000 times and the *P*-value was calculated as the proportion of random samples (of the 10,000 permutations) that had a value greater than that of the real data.

K_a permutation tests were performed similarly except the median K_a of randomly sampled *N. tetrasperma mat A* and *mat a* ortholog pairs was compared to that of the genes within the relocated region.

Transposon enrichment

We compared each *N. tetrasperma* strain separately to *N. crassa*. We took the total number of transposons within each genome and shuffled their locations. We then counted the number of shuffled transposons that landed within the boundaries of the nonrecombining region in *N. tetrasperma* and within the homologous region in *N. crassa*. We subtracted the *N. crassa* sum from the *N. tetrasperma* sum and compared this difference to the true difference. The *P*-value was calculated as the proportion of permutations (of 10,000 total) where the permuted difference was larger than the true difference.

All permutation tests were performed using custom Perl scripts.

Results

Chromosomal rearrangements

To determine whether chromosomal rearrangements could be responsible for the suppression of recombination, we created a whole-genome alignment between the *mat a* and *mat A* strains of *N. tetrasperma* as well as between each *N. tetrasperma* strain and the outgroup species *N. crassa*. We visualized large-scale synteny between these pairs using dotplots (Figure 1). Synteny is strongly conserved across all genome pairs, with the exception of those involving the *N. tetrasperma mat A* mating-type chromosome. Knowing that the *N. tetrasperma mat a* mating-type chromosome is collinear with that of *N. crassa*, we concluded that a series of chromosomal rearrangements had occurred on the *N. tetrasperma mat A* chromosome.

Focusing on the mating-type chromosome and the comparison between the two *N. tetrasperma* strains, we found that there have been two large inversions (~5.3 and 1.2 Mb), a smaller inversion (68 kb), and an apparent translocation of an ~143-kb segment from one end of the chromosome to the other (Table 1 and Figure 2). Consistent with previous cytological data, the mating-type chromosome ends are collinear and the rearrangements encompass contiguous regions within the central portion of the chromosome; however, none of the rearrangement events include the *mat* locus itself. We propose that the evolution of the *mat A* chromosome can be explained by two overlapping inversions that, together, resulted in the movement of the 143-kb segment from one end of the chromosome to the other (Figure 2). We therefore hereafter refer to this segment as the "relocated" region. The alternative explanation for this movement, that a 143-kb genomic segment was excised from one end of the *mat* chromosome and reinserted into the other, appears less likely for reasons explained below.

Mechanisms of rearrangement

Ectopic recombination between transposable elements is generally believed to be a common mechanism for the generation of chromosomal inversions (Casals and Navarro 2007) and its role in creating chromosomal rearrangements has been experimentally verified in yeast (Argueso *et al.* 2008). However, a study by Ranz *et al.* (2007) found no evidence for ectopic recombination in the majority of fixed inversions between species within the *Drosophila melanogaster* group. Instead, the authors found short duplications of nonrepetitive sequence at the breakpoints of most of the inversions they identified and proposed a novel mechanism for the generation of inversions involving staggered single-strand breaks followed by nonhomologous end joining.

We compared the two breakpoints associated with each inversion shown in our model to determine whether there was evidence for either ectopic recombination via transposable elements or the staggered break model of Ranz *et al.* (2007). Interestingly, we find evidence for both types of rearrangement mechanisms, adding credence to the novel mechanism proposed by Ranz *et al.* while also supporting the notion that ectopic recombination is a common mechanism underlying chromosomal rearrangements. It is important to note that there is another class of rearrangement mechanisms that occur in slipped, stalled, or collapsed DNA replication forks. These mechanisms include serial replication slippage (Chen *et al.* 2005), microhomology-mediated break-induced replication (MMBIR) (Hastings *et al.* 2009), and fork stalling and template switching (FoSTeS) (Lee *et al.* 2007). We cannot eliminate the possibility that such mechanisms created the rearrangements we observe in *N. tetrasperma*. However, previously described rearrangements that have been associated with these mechanisms showed a complex combination of double or triple deletions, insertions, and inversions (Mani and Chinnaiyan 2010), which are not seen here.

At the breakpoints of the 1.2-Mb inversion, we find in *mat A* a short duplication of a 50-bp segment that is unique in the *N. tetrasperma mat a* genome. The orientation of this duplication is consistent with those found in *Drosophila* (Ranz *et al.* 2007) and with our model of the rearrangement events. In *N. tetrasperma*, after the 1.2-Mb inversion, the duplicated segments would have been in an inverted orientation as in *Drosophila* (Ranz *et al.* 2007). The subsequent 5.3-Mb inversion, which contained the left duplication, would have then returned the leftmost duplication to the same orientation as the right duplication (Figure 2).

The 5.3-Mb inversion is flanked by inverted *Mariner* transposable elements. The amino acid sequences of the

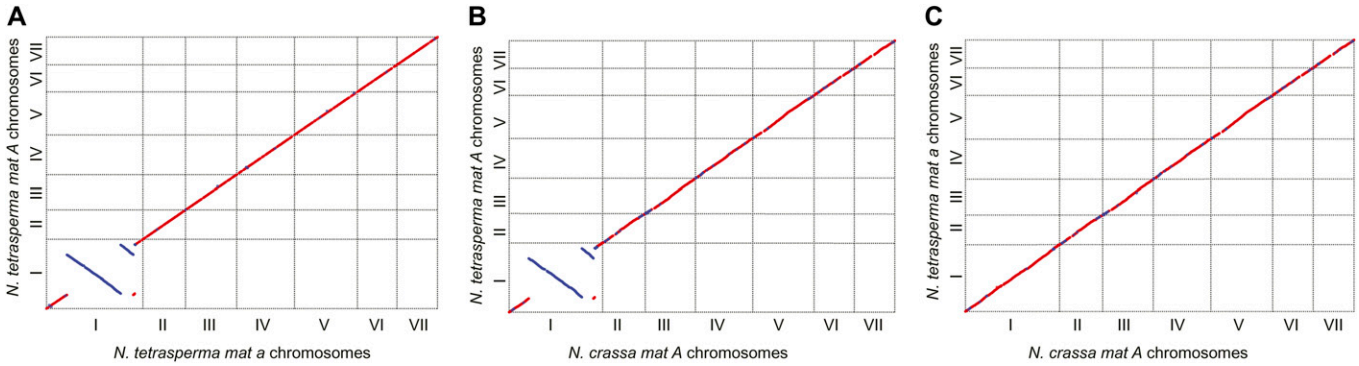

Figure 1 Whole-genome synteny between *N. tetrasperma* strains and *N. crassa*. (A) *N. tetrasperma mat A* compared to *N. tetrasperma mat a*; (B) *N. tetrasperma mat A* compared to *N. crassa*; (C) *N. tetrasperma mat a* compared to *N. crassa*. The mating-type chromosome (chromosome I) is rearranged in the *N. tetrasperma mat A* strain compared to both *N. tetrasperma mat a* and *N. crassa*. Alignments occurring between positive strands are in red while those occurring between opposite strands are in blue and indicate inversions.

transposase ORFs of the two elements are ~28% identical and ~59% similar to that of the *Pogo* family Mariner-3_AN from *Aspergillus nidulans* (Kapitonov and Jurka 2003). Both transposase ORFs have multiple stop codons, suggesting that they are no longer active.

These sequence features also support the inclusion of the relocated region within both inversions, providing additional support for our model of how this region moved from one end of the chromosome to the other (Figure 2). We found no such sequence features supporting the alternative model, in which the inversions occurred independently from the relocation.

The small 68-kb inversion could also have resulted from ectopic recombination. Although the sequences located at the breakpoints of this inversion have no homology to any known transposable elements, both flanks contain several microsatellite regions and regions of low sequence complexity, creating several short blocks of microhomology that could facilitate ectopic recombination.

Relative ages of rearrangement events

We used the whole-genome alignment between the two *N. tetrasperma* strains to locate 192,225 nucleotide differences between them, >99% (190,728) of which are located within the boundaries of the nonrecombining region on the mating-type chromosome. These two strains are homokaryons that were derived from a single heterokaryotic strain. The lack of nucleotide differences across most of the genome is most likely due to repeated inbreeding of the self-fertile heterokaryotic strain (Merino *et al.* 1996).

If nucleotide differences have accumulated on the mating-type chromosome because of recombination suppression, the number of neutral differences should be proportional to the amount of time since the formation of the nonrecombining region. If the rearrangement events that we observed occurred at different evolutionary time points, they should have different levels of neutral nucleotide divergence with the oldest event being the most divergent.

As a measure of neutral divergence between the two *N. tetrasperma* strains, we calculated the number of synonymous substitutions per site (K_s) for each gene within each rearrangement event (Figure 3). The large 5.3-Mb inversion was the rearrangement event that allowed *N. tetrasperma* to become self-fertile by suppressing recombination along the majority of the chromosomal region between the *mat* locus and the centromere. We have therefore compared the distribution of K_s values for genes within the large inversion to those distributions for each of the other rearrangement events.

We found no significant difference between K_s values for the 5.3-Mb inversion and the 1.2-Mb inversion, nor is there a significant difference between K_s values for the 5.3-Mb inversion and the *mat* proximal block (the 218-kb chromosomal region between the *mat* locus and the leftmost rearrangement event; see Figure 3), suggesting that the suppression of recombination in the 5.3-Mb and 1.2-Mb inversions occurred at the same evolutionary time and that these inversions also suppressed recombination in the *mat* proximal block (Bonferroni-corrected Mann-Whitney U (MWU) tests: $P = 0.6$ and $P = 1$, respectively; Figure 3).

We found two chromosomal regions whose K_s values were much smaller compared to those from the 5.3-Mb inversion: the *mat* distal block (the 850-kb chromosomal region between the recombining left chromosome arm and the *mat* locus; see Figure 3) and the 68-kb inversion (Bonferroni-corrected MWU tests: $P = 1.1e\text{-}15$ and $P = 4.4e\text{-}05$,

Table 1 Summary of the *N. tetrasperma* mating chromosome regions

Chromosomal region	Size (kb)	Single-copy orthologs
Recombining left arm	930	198
mat distal block	850	166
mat proximal block	218	29
Relocated region	143	26
5.3-Mb inversion	5300	972
1.2-Mb inversion	1200	245
68-kb inversion	68	11
Recombining right arm	728	128

The approximate size of each chromosomal region and number of single-copy orthologs are listed for the eight segments of the *N. tetrasperma* mating chromosome that are discussed in this study.

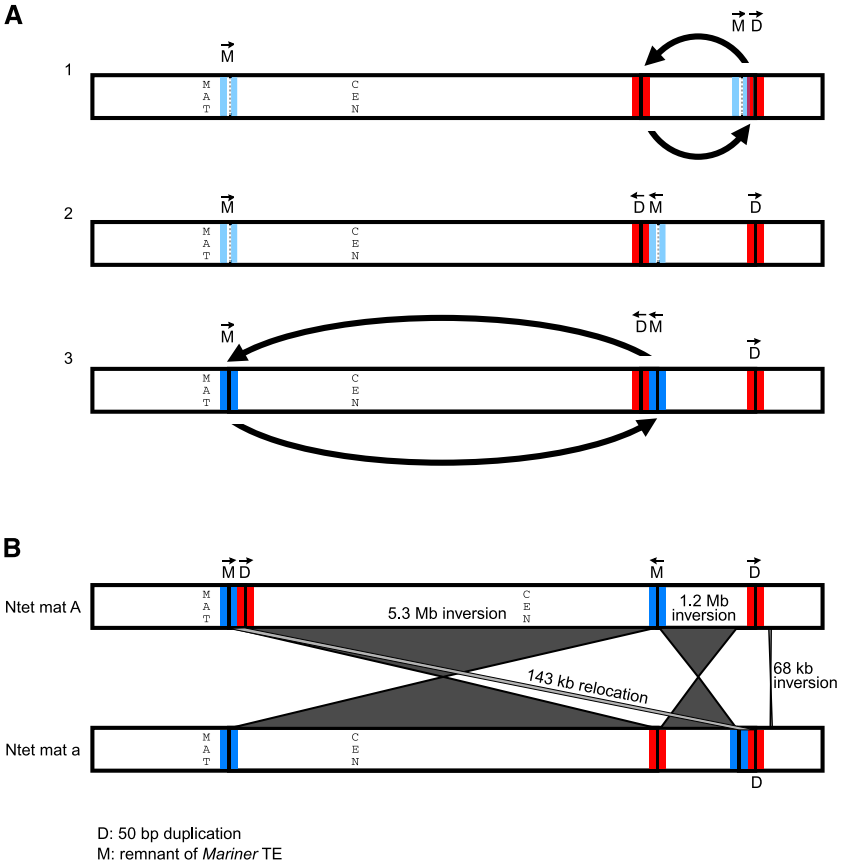

Figure 2 Model of the evolution of the *N. tetrasperma mat A* mating-type chromosome. The order of rearrangement events is shown in A and begins with the ancestral *mat A* chromosome (1) which was collinear with *mat a* and the mating-type chromosome of *N. crassa*. The 1.2-Mb inversion occurred first and produced the orientation in 2. This event was followed relatively quickly by the 5.3-Mb inversion (3). The 68-kb inversion, shown as the line at the far right of B, occurred much later to produce the current arrangement of the *mat A* chromosome (B). The 1.2-Mb inversion (breakpoints shown in red) is flanked by unique 50-bp duplications (D) that would have been in an inverted orientation before the occurrence of the large inversion, consistent with rearrangement via staggered single-strand breaks. The 5.3-Mb inversion (breakpoints shown in blue) is flanked by *Mariner* transposable elements (M), consistent with rearrangement via ectopic recombination. *Mariner* remnants were not present in either of the homologous regions in the *mat a* chromosome. The overlapping nature of these two inversions explains the relocated genomic region. The 68-kb inversion is flanked by a microsatellite containing, low-complexity sequence and may have occurred due to ectopic recombination between blocks of microhomology. MAT denotes the location of the mating-type locus while CEN shows the location of the centromere.

respectively; Figure 3). The suppression of recombination in these regions most likely occurred after the rearrangement event that created the large inversion.

Surprisingly and counter to our model of the order of rearrangement events, we found that the K_s values for genes within the relocated region were much larger than those for genes within the 5.3-Mb inversion (Bonferroni-corrected Mann-Whitney U (MWU) test: $P = 4.33e\text{-}06$; Figure 3). One interpretation of this result is that recombination was suppressed in this region much earlier than that in the 5.3-Mb inversion. The median K_s values between the *N. tetrasperma mat A* and *mat a* strains for the genes within the relocated region are ~2.7 times greater than those for the next most divergent regions. If the substitution rate was the same between these chromosomal regions, the suppression of recombination within the relocated region would have had to occur at an evolutionary time point that was more than twofold earlier than the other rearrangement events. However, the divergence between *N. tetrasperma mat A* and its close relative *N. crassa* is only ~1.3-fold greater than the divergence between the two *N. tetrasperma* strains (median K_s between *N. tetrasperma* and *N. crassa* for genes within the 5.3- and 1.2-Mb inversions, 0.070; median K_s for the same genes between the *N. tetrasperma* strains, 0.054). This sequence of events would place the time point of the relocation well before the divergence of the species. Given that only one of the species now exhibits the relocation, this scenario is very unlikely.

Investigating the large sequence divergence of the relocated region

The alternative interpretation for the large sequence divergence of the relocated region is that it has been experiencing an elevated substitution rate relative to the rest of the nonrecombining region. However, because we used only synonymous codon positions to calculate divergence, the elevated rate of substitution would have to pertain to nucleotide substitutions that did not change the amino acid, rather than an elevated rate of protein evolution. One possibility is that the elevated synonymous substitution rate is due to relaxed selection for codon usage. Another is that some of the annotated genes within this region are actually pseudogenes and thus, because they are completely unconstrained by purifying selection, are accumulating nucleotide substitutions at a faster rate than synonymous positions. To address the possibility that the divergence measures for this region are biased by pseudogenes, we repeated the analysis on a subset of genes from the region for which we had evidence of expression from the *mat A* ESTs. The increased divergence was also present in the subset of genes showing evidence of expression, suggesting that the potential inclusion of pseudogenes in our analysis was not responsible for the increased divergence we observed. Obviously, such a filter is not infallible because a minority of pseudogenes have been shown to be transcribed on the basis of profiling

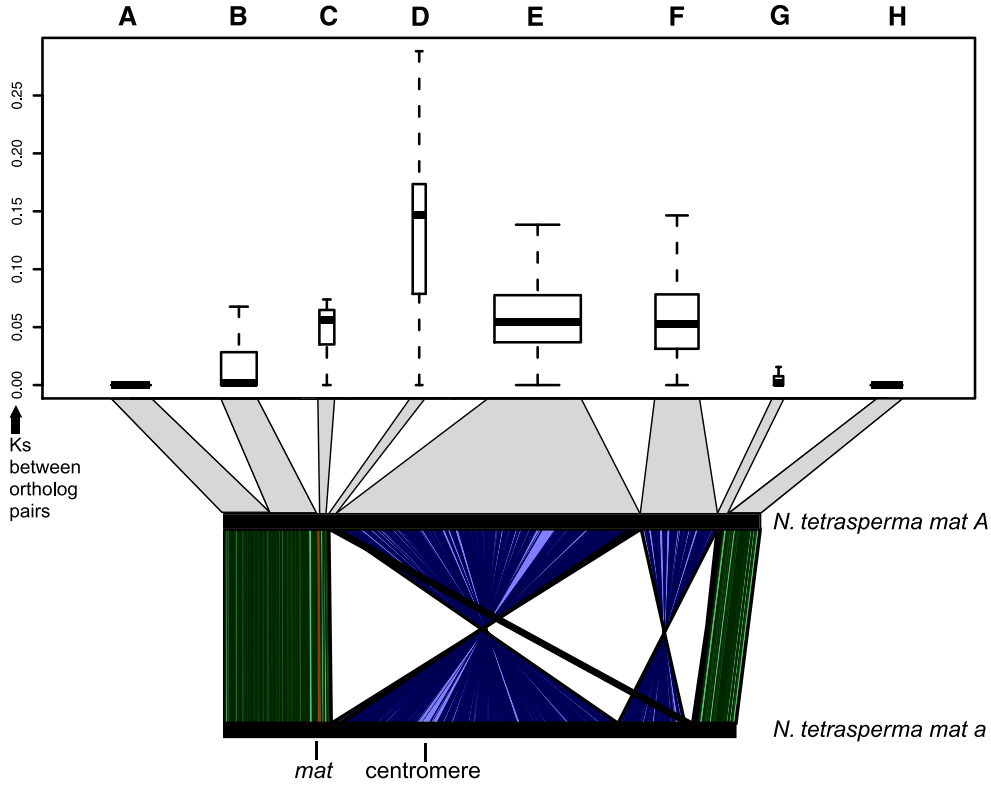

Figure 3 Relative ages of rearrangement events. The bottom panel shows the two *N. tetrasperma* mating-type chromosomes with lines connecting pairs of orthologous genes. The top panel shows the distribution of synonymous substitutions per site (K_s) between ortholog pairs for chromosomal regions that have been rearranged and for two other chromosomal segments that are collinear between the two *N. tetrasperma* strains but have sequence divergence: an 850-kb segment distal to the *mat* locus and a 218-kb segment proximal to the *mat* locus. The amount of synonymous sequence divergence between orthologs within a given rearrangement event will be proportional to the amount of time since the event occurred. Using the 5.3-Mb inversion as a point of reference, the genes within the relocated region have a distribution of K_s values that is significantly larger, while the genes within the *mat* distal unknown block and small 68-kb inversion both have K_s distributions that are significantly smaller (Bonferroni-corrected MWU test: $P = 4.33e\text{-}06$, $P = 1.1e\text{-}15$, and $P = 4.4e\text{-}05$, respectively). The letters A–H denote the chromosomal regions that are referred to in the main text: A, recombining left arm; B, *mat* distal block; C, *mat* proximal block; D, 143-kb relocated region; E, 5.3-Mb inversion; F, 1.2-Mb inversion; G, 68-kb inversion; H, recombining right arm.

of mRNA/ESTs from a variety of organisms [between 2% and 15% of pseudogenes studied in rice (Zou *et al.* 2009), *Arabidopsis* (Zou *et al.* 2009), humans (Zheng *et al.* 2007), and yeast (Lafontaine and Dujon 2010)]. Assuming Neurospora also follows this general pattern, such a filter, although incomplete, would nevertheless eliminate the majority of pseudogenes from this region.

To test the hypothesis that relaxed selection for codon usage is occurring in the relocated region, we calculated the codon usage in the 100 most highly expressed genes in *N. crassa*. We used these values to calculate the CAI for each one-to-one ortholog between *N. crassa* and the two *N. tetrasperma* strains. CAI values range from 0 to 1 with higher values indicating more codon usage bias (Sharp and Li 1987). To control for any bias resulting from using *N. crassa* highly expressed genes to identify favored codons, we repeated this analysis with codon usage information gleaned from the 100 most highly expressed genes in *N. tetrasperma mat A* (as determined by EST coverage) and obtained similar results (Figure S1).

We compared the CAI values for genes from each *N. tetrasperma* strain to their ortholog in *N. crassa*. We found that the genes within the chromosomal region that was relocated in *N. tetrasperma mat A* have a median CAI that is lower than that of their *N. crassa* orthologs (one-sided permutation test: *N. tetrasperma mat A*, $P = 0.0016$; using only the subset of genes with ESTs, $P = 0.0023$) (Figure 4). This situation is not seen for the other rearrangement events; to the contrary, the genes within the 68-kb inversion appear to be evolving higher codon usage bias in both *N. tetrasperma* strains [one-sided permutation test: $P = 0.026$ (*mat A*), $P = 0.003$ (*mat a*); Figure 4]. Across the 7693 ortholog pairs that we identified in the *N. tetrasperma* and *N. crassa* genome comparisons, we found slightly more pairs where the *N. crassa* ortholog had a larger CAI value, for both *N. tetrasperma* mating types [one-sided binomial test: $P = 0.0001$ (*mat A*), $P = 4.871e\text{-}08$ (*mat a*); Table 2].

Given that most amino acid-changing mutations are likely to be deleterious, it is reasonable to expect that relaxed selection across a genomic region would result in an increase in the number of nonsynonymous substitutions per site (K_a). To minimize inflation of K_a due to the inclusion of pseudogenes or misannotated genes in the analysis, we used only genes that had evidence of expression in the form of ESTs.

We found that the genes located within the relocated region of the *N. tetrasperma mat A* strain as well as their orthologs in *N. tetrasperma mat a* both have a median K_a that is greater than that of the nonrecombining region as a whole (one-sided permutation test: *N. tetrasperma mat A*, $P = 0.005$; *N. tetrasperma mat a*, $P = 0.013$; Figure 5). Together, the elevated K_a and the reduced CAI of the genes in the relocated region suggest that it has experienced reduced purifying selection in the *N. tetrasperma* lineage since its divergence from *N. crassa*.

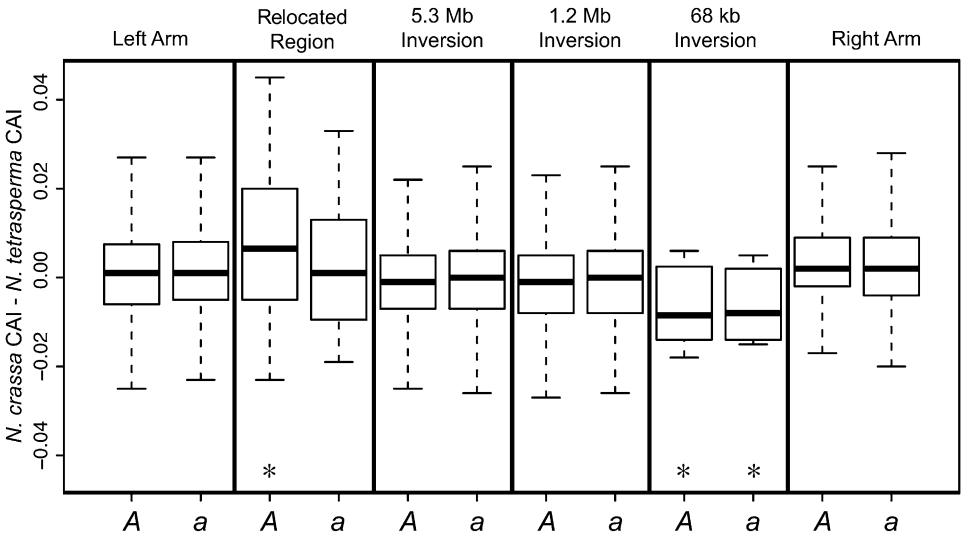

Figure 4 Relaxed selection for codon usage in the *N. tetrasperma mat A* genes within the relocated region. The CAI is a measure of codon usage bias and the difference between the *N. crassa* CAI and those from each *N. tetrasperma* strain was calculated for every set of three-way orthologs. The genes within the *N. tetrasperma mat A* relocated region appear to be evolving reduced codon usage bias (one-sided permutation test: $P = 0.0016$) while the genes within the small inversion appear to be evolving increased codon usage bias [one-sided permutation test: $P = 0.026$ (*mat A*), $P = 0.003$ (*mat a*)]. Outliers are not shown.

Evidence for asymmetrical degeneration within the nonrecombining region

The origin of the *N. tetrasperma* region of suppressed recombination is ~37 times younger than that of the XY sex chromosomes of marsupial and placental mammals (~4.5 MYA *vs.* ~166 MYA) (Veyrunes *et al.* 2008). Unlike mammals, it is possible for a functionally diploid (heterokaryotic) *N. tetrasperma* individual to become haploid (homokaryotic). These haploid individuals grow and reproduce via mitotic spores in the laboratory and there is evidence of outcrossing between them in the wild (Menkis *et al.* 2009; Powell *et al.* 2001). These observations suggest that there should be selection to maintain the function of both copies of the sex-linked genes and one would therefore expect the nonrecombining regions to be maintained intact. However, the two strains that we examine here are almost identical across their genomes except for the nonrecombining region, implying a long history of inbreeding, and previous studies have observed high instances of sexual dysfunction when *N. tetrasperma* strains are outcrossed in the laboratory (Jacobson 1995; Saenz *et al.* 2001). These results suggest that the degree of outcrossing in nature may be limited, in which case selection against degeneration within the nonrecombining regions of both mating types may be reduced.

One signal of degeneration that has been found previously in *N. tetrasperma* (Whittle and Johannesson 2011; Whittle *et al.* 2011b), as well as in regions of reduced recombination in many other organisms (Kliman and Hey 1993; Betancourt and Presgraves 2002; Bachtrog 2003; Liu *et al.* 2004; Peichel *et al.* 2004; Nicolas *et al.* 2005; Presgraves 2005; Haddrill *et al.* 2007; Marais *et al.* 2008; Zhou *et al.* 2008; Betancourt *et al.* 2009), is the accumulation of deleterious alleles. The elimination of recombination across a genomic region will, in essence, lower the effective population size of that region, making fixation more likely for slightly deleterious mutations and less likely for slightly beneficial ones. Given that most amino acid-changing substitutions are deleterious, suppression of recombination should result in an increased proportion of nonsynonymous substitutions compared to synonymous substitutions (K_a/K_s) across the nonrecombining region (Charlesworth and Charlesworth 2000).

The previous results from a different lineage within the *N. tetrasperma* species complex found a higher K_a/K_s ratio for a set of *N. tetrasperma* genes located within the nonrecombining region but not for a set of genes from outside of this region, consistent with the genes within the nonrecombining region being under relaxed purifying selection (Whittle and Johannesson 2011; Whittle *et al.* 2011b). To determine whether there is evidence supporting this prediction for the genes within the nonrecombining region of this *N. tetrasperma* lineage, we calculated K_a and K_s for all ortholog pairs between each *N. tetrasperma* strain and *N. crassa* and compared the K_a/K_s ratios for the genes within the nonrecombining region to those for the genes outside it.

Consistent with the prediction of Charlesworth and Charlesworth (2000) and the previous results from another *N. tetrasperma* lineage (Whittle and Johannesson 2011; Whittle *et al.* 2011b), we found that the genes within the nonrecombining region of the *mat a* strain have a significantly higher median K_a/K_s ratio than the genes outside of the nonrecombining region, but this pattern did not hold for

Table 2 Genome-wide comparison of codon usage between *N. crassa* and *N. tetrasperma*

N. tetrasperma	Nc CAI > Nt CAI	Nc CAI < Nt CAI	Nc CAI = Nt CAI
mat A	3844	3531	318
mat a	3905	3447	341

The codon adaptation index (CAI) was compared for ortholog pairs between each *N. tetrasperma* mating type and *N. crassa*. The *N. crassa* gene has a larger CAI value than its *N. tetrasperma* ortholog in more comparisons than expected by chance [one-sided binomial test: $P = 0.0001$ (*mat A*), $P = 4.871e\text{-}08$ (*mat a*)]. Nc, *N. crassa*; Nt, *N. tetrasperma*.

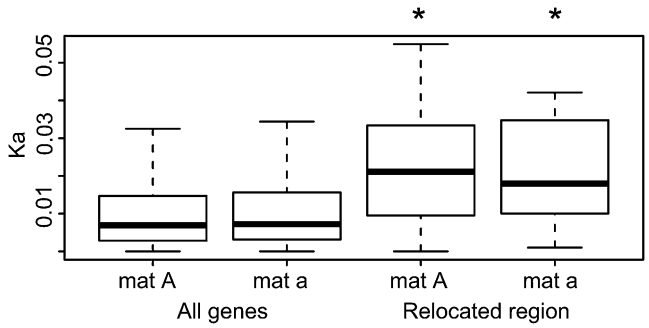

Figure 5 Increased nonsynonymous divergence in the *N. tetrasperma mat A* and *mat a* genes within the relocated region. The numbers of nonsynonymous substitutions per site (K_a) between *N. tetrasperma mat A* and *mat a* orthologs within the relocated region are significantly larger than those values for the entire nonrecombining region [one-sided permutation test: $P = 0.005$ (*mat A*), $P = 0.013$ (*mat a*)]. Although K_a is larger for these genes, no gene has a K_a/K_s ratio significantly greater than one, implying that the increased K_a is due to the accumulation of slightly deleterious amino acid substitutions rather than adaptive evolution. Outliers are not shown.

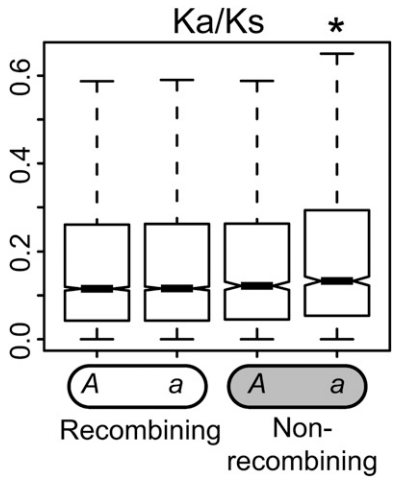

Figure 6 Reduced efficiency of selection in the *mat a* nonrecombining region. The ratio of nonsynonymous substitutions per site (K_a) to synonymous substitutions per site (K_s) was calculated for every pair of orthologs between each *N. tetrasperma* strain and *N. crassa*. The distribution of K_a/K_s ratios was compared for genes inside of and outside of the nonrecombining region. Consistent with the evidence in Fig. 7, genes within the region of suppressed recombination in the *mat a* strain, but not in the *mat A* strain, have significantly larger K_a/K_s ratios than those outside of the nonrecombining region (MWU test: $P = 0.001$ and $P = 0.6$, respectively). Outliers are not shown.

the genes from the *mat A* strain (MWU test: $P = 0.001$ and $P = 0.6$, respectively; Figure 6 and Figure S2). One explanation for this is that the *mat a* region may be accumulating deleterious alleles at a faster rate than the *mat A* region. To further explore this possibility, we compared the accumulation of deleterious mutations within the nonrecombining region in the form of pseudogenes, codon usage, and nonsynonymous substitutions between the two *N. tetrasperma* mating types. For the pseudogene analysis, we identified candidate pseudogenes from the set of *N. crassa* predicted proteins for which we were unable to find an ortholog in the *N. tetrasperma* set of predicted peptides. Confidently identifying pseudogenes can be difficult because misannotations in the *N. crassa* genome that incorrectly identify start sites or incorrectly predict ORFs that are not actually transcribed can make the homologous region in *N. tetrasperma* appear to be a pseudogene. For these reasons, we required a candidate pseudogene in *N. tetrasperma* to have either a frameshift or a nonsense mutation, full-length homology to functional genes in both *N. crassa* and *N. discreta*, and evidence of expression in *N. crassa*. Applying these conservative criteria, we found a total of 10 candidate pseudogenes: 2 with nonsense and/or frameshift mutations in both *N. tetrasperma* genomes, 5 with such mutations only in the *N. tetrasperma mat a* genome, and 3 appearing only in the *N. tetrasperma mat A* genome (Table S1). While it is notable that we observed more pseudogenes on the *mat a* chromosome compared to *mat A*, the sample size (10 pseudogenes in total) is not large enough to confidently conclude that this represents evidence of asymmetrical degeneration.

As an additional approach to assess evidence of asymmetrical degeneration in one of the two *N. tetrasperma* sex-linked regions, we used *N. crassa* as an outgroup to assign the nucleotide substitutions that we identified within the nonrecombining region to one of the two *N. tetrasperma* lineages. After normalizing by the total number of nucleotide substitutions, we compared the frequencies of each nonsynonymous substitution (at the codon level) between the two *N. tetrasperma* strains and found that nonsynonymous substitutions have occurred at higher frequencies on the *N. tetrasperma mat a* lineage (one-sided, paired MWU test: $P = 2.749e\text{-}15$; Figure 7A), suggesting that the *N. tetrasperma mat a* strain may be accumulating deleterious substitutions at a higher rate than the *mat A* strain.

We also used the codon usage table mentioned previously to identify synonymous changes involving the substitution of a more preferred codon to a less preferred codon and vice versa. After normalizing by the total number of synonymous substitutions within each lineage, we found a tendency for substitutions in *N. tetrasperma mat a* that involve a change to an unpreferred codon to have occurred at higher frequencies, although this difference is not significant at $\alpha = 0.05$ (one-sided, paired MWU test: $P = 0.072$; Figure 7B). We also found that substitutions involving a change to a preferred codon have occurred at lower frequencies compared to those in *N. tetrasperma mat A* (one-sided, paired MWU test: $P = 0.039$; Figure 7B). Together, these results suggest that the *N. tetrasperma mat a* nonrecombining region may be in the early stages of degeneration and are consistent with observations that, within a heterokaryotic *N. tetrasperma* individual, *mat A* nuclei outnumber *mat a* nuclei during growth and early sexual development (H. Johannesson, personal communication).

Suppression of recombination in other systems has often been accompanied by the accumulation of transposable

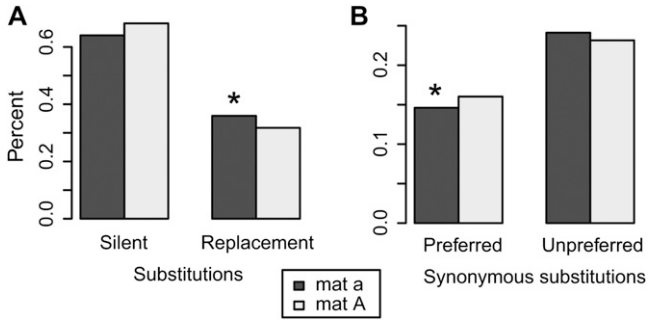

Figure 7 Evidence that the *N. tetrasperma mat a* mating-type chromosome may be in the early stages of degeneration. Nucleotide substitutions occurring within the nonrecombining region were assigned to either the *N. tetrasperma mat A* or *mat a* lineage depending upon which allele was present in *N. crassa*. Examination of this set of polarized substitutions showed that nonsynonymous substitutions have occurred at higher frequencies in the *N. tetrasperma mat a* lineage compared to *mat A* (one-sided, paired MWU test: $P = 2.749e\text{-}15$). In addition, examination of the set of polarized synonymous substitutions showed that, in the *N. tetrasperma mat a* lineage, preferred substitutions have occurred at significantly lower frequencies and there is a trend toward significance with respect to unpreferred substitutions having occurred at higher frequencies, compared to the *mat A* strain (one-sided, paired MWU test: preferred, $P = 0.039$; unpreferred, $P = 0.072$). Preferred substitutions were defined as a change to a synonymous codon whose frequency of usage in the top 100 most highly expressed genes in *N. crassa* is at least 10-fold larger than that of the ancestral codon while unpreferred substitutions were defined as the opposite (the usage frequency of the derived codon is at least 10-fold smaller than that of the ancestral codon).

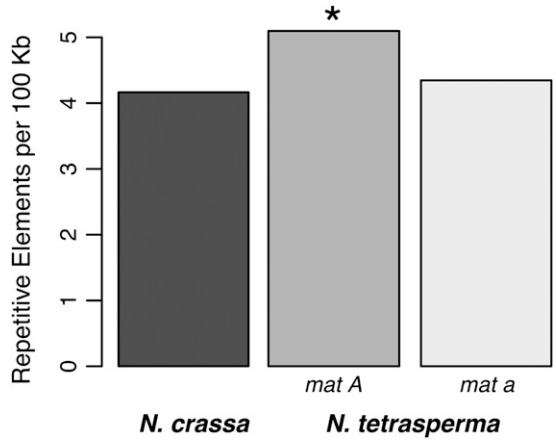

Figure 8 The *N. tetrasperma mat A* region of suppressed recombination is enriched for repetitive elements. Repetitive elements were identified *de novo* and on the basis of homology to known elements. Significance was assessed using a permutation test ($P = 0.0004$). Using the same test, the homologous region from *N. tetrasperma mat a* did not have significantly more repetitive elements than *N. crassa* ($P = 0.30$).

elements. We identified *de novo* repetitive elements as well as those with homology to known fungal elements in both *N. tetrasperma* strains and used a permutation test to determine whether the nonrecombining region is enriched for transposons relative to the rest of the genome. Interestingly, the *mat A* strain has significantly more transposons in the genomic region where recombination is suppressed compared to the rest of the genome ($P = 0.0004$), but the *mat a* strain does not ($P = 0.30$) (Figure 8).

Discussion

In this study we have used two high-quality genome assemblies, representing the nuclei of opposite mating type derived from a single heterokaryotic strain, to discover a series of three inversions within the *N. tetrasperma* region of suppressed recombination. The location of these rearrangements and the collinearity of the chromosome ends are consistent with the cytology, genetic map data, and sequence divergence data from previous studies that have investigated the region of suppressed recombination. The identification of these rearrangements answers the question raised by Jacobson (2005) of the relative influence of structural *vs.* genetic modifiers in maintaining this nonrecombining region: while we show that structural rearrangements encompass most of the nonrecombining region, the *mat a* and *mat A* chromosomes are collinear in the chromosomal region surrounding the *mat* locus (Figure 3). This region includes the more recent evolutionary stratum identified in Menkis *et al.* (2008) (the ~850-kb *mat* distal block) as well as the *mat* proximal block (the ~218-kb chromosomal segment between the mating-type locus and the first chromosomal rearrangement). Genetic evidence indicates that this segment is within the nonrecombining region, and the presence of sequence divergence within this region (Figure 3) implies that it has long remained so, but it is unclear why recombination is suppressed. However, the *mat a* and *mat A* loci do not share sequence homology and, in addition, we found that within each strain, the *mat* locus is flanked by regions that have been subject to Repeat Induced Point mutation (RIP) (Galagan and Selker 2004).

RIP is a genome-defense mechanism that subjects recently duplicated sequences to numerous G:C to A:T point mutations and is thought to have evolved to suppress the amplification of transposable elements. A side effect of this defense mechanism is the elimination of young, highly similar, gene paralogs (Galagan and Selker 2004). In the sex chromosomes of mammals and *Drosophila*, intrachromosomal gene conversion between such young duplicates has been hypothesized to play a role in rescuing Y-linked genes from degeneration (Rozen *et al.* 2003; Connallon and Clark 2010). Because RIP constrains gene duplication, we do not expect a similar phenomenon to be occurring in *N. tetrasperma*.

It is possible that the RIPed regions, together with the idiomorphic mat loci, create islands of sequence divergence that disrupt synapsis in this region and thus cause the suppression of recombination to extend past the distal portion of the *mat* locus. Other nonstructural mechanisms that have been shown to suppress recombination include genetic modifiers such as the *rec* genes of *N. crassa* (Catcheside 1975), overall high levels of sequence divergence (rather than several discrete islands) (Hunter *et al.* 1996), and DNA methylation (Maloisel and Rossignol 1998). Additionally, because

this region is located at the end of the nonrecombining interval, it may simply be that there have been fluctuations in the exact location of the boundary of the nonrecombining region over evolutionary time. At this time, we are unable to differentiate between these possibilities.

By combining our analysis of the relative timing of the rearrangement events with the analysis of sequence features associated with the boundaries of these events, we have formulated a cohesive model for the structural evolution of the *N. tetrasperma* mating-type chromosome (Figure 2). On the basis of sequence divergence, the last rearrangement event (a 68-kb inversion) occurred more recently than the first two and therefore represents a second evolutionary stratum. These results, along with those from other fungal phyla (Fraser *et al.* 2004; Idnurm *et al.* 2008), suggest that inversion-mediated suppression of recombination on mating chromosomes and the expansion of such nonrecombining regions as discrete strata have occurred independently in many fungal lineages and may be a common property of the evolution of sex in fungi, similar to the sex chromosomes of other taxa (Charlesworth *et al.* 2005; Fraser and Heitman 2005).

An additional similarity between the *N. tetrasperma* nonrecombining region and the sex chromosomes from other systems is that it may be under reduced purifying selection. The nonrecombining region in the *N. tetrasperma mat A* strain has accumulated an excess of transposons compared to the rest of genome and there is evidence that the *N. tetrasperma mat a* strain may be in the early stages of degeneration (Figures 6–8). Our finding that, across the genome, there is a slight but significant tendency for *N. crassa* genes to have a higher CAI value compared to the orthologs in both *N. tetrasperma* mating types (Table 2) is consistent with what would be expected given the evidence for extensive selfing within the *N. tetrasperma* heterokaryon. However, this result contrasts with that reported by Whittle *et al.* (2011c). In a comparison of genome-wide codon usage between the same *N. tetrasperma mat A* strain we study here and the outcrossing species *N. discreta*, the authors found that *N. tetrasperma* has a higher frequency of optimal codon usage. Whittle *et al.* hypothesize that this result may be due to *N. discreta* having an effective population size that is smaller than that of *N. tetrasperma*, despite the fact that *N. discreta* is not self-fertile. While estimates of effective population size have been inferred for two populations of *N. crassa* (Ellison *et al.* 2011), none exist for *N. discreta*. Further population-level work involving the comparison of effective population size between these three species would shed light on this inconsistency.

It is unclear why we observe an accumulation of transposons only on the *mat A* chromosome. One explanation is that we identified fewer transposons in the *mat a* strain because its genome assembly has a higher proportion of sequence bases in gaps compared to the *mat A* strain (Table S2). It is possible that the additional gaps in the *mat a* assembly are caused by a reduced ability to assemble repetitive elements due to differences in sequence coverage (Sanger *vs.* 454 as well as single end *vs.* paired end; Table S3 and Table S4) between the two genome assemblies. Alternatively, we speculate that the additional transposable elements in the *N. tetrasperma mat A* strain could be the result of "genome shock" (McClintock 1984): a single burst of activity specific to the *mat A* nucleus due to the release of transposon suppression during the reorganization of the mating-type chromosome. Of course, it is equally plausible that the relative timing of the chromosome reorganization and the transpositions was reversed, such that increased transposon activity enabled the 5.3-Mb inversion that led to self-fertility in *N. tetrasperma*, thereby linking the additional transposon copies to the new chromosome orientation.

Our finding that the *mat a* chromosome is accumulating deleterious alleles at a faster rate than the *mat A* chromosome is also somewhat unexpected and stands in contrast to previous theory (Bull 1978). *N. tetrasperma* individuals can grow as haploid homokaryons that are capable of outcrossing and acting as a maternal or paternal parent. Recessive deleterious alleles would not be sheltered under these conditions and, if *N. tetrasperma* exists often as a homokaryon in nature, there should be selection to maintain both copies of the sex-linked genes. However, there is evidence that outcrossing may be limited in nature (Jacobson 1995), in which case degeneration would be reasonable because most individuals would not leave their heterokaryotic, sheltered state. Additionally, asymmetric evolution between sex-determining chromosomes, at least with respect to differences in chromosome size, has been observed in other haploid organisms such as the liverwort *Marchantia polymorpha* (Yamato *et al.* 2007) and the fungus *M. violaceum* (Hood 2002). In neither of these systems, unfortunately, have the chromosomes of both mating types been sequenced, making it impossible to determine whether the size difference is due to asymmetric gain or loss (*i.e.*, degeneration) of genetic elements.

While Whittle *et al.* (2011b) report an asymmetry between the *mat a* and *mat A* chromosomes in terms of the frequency of substitutions to preferred codons in lineage 1 of the *N. tetrasperma* species complex, a different lineage from that studied here, they do not report whether the deviation is statistically significant and a reanalysis of their results shows that it is not (Fisher's exact test: $P = 0.285$; Table S5). Additionally, there does not appear to be a similar asymmetry with respect to nonsynonymous substitutions in lineage 1 (Whittle and Johannesson 2011). That we are able to observe such an asymmetry in both nonsynonymous substitutions and codon usage between the sex-linked regions in this *N. tetrasperma* lineage may be due to the larger sample of genes we use here (~1300 genes from within the nonrecombining region compared to 168 in Whittle and Johannesson 2011 and 228 in Whittle *et al.* 2011b) or to differences between the lineage studied in Whittle and Johannesson (2011; Whittle *et al.* 2011b) (lineage 1) and that studied here (lineage 6).

Future work comparing the structural rearrangements identified here to the other *N. tetrasperma* lineages would

address the hypothesis of independent origins of this region of suppressed recombination and lead to further insight into its evolutionary history. The examination of sequence data from *N. tetrasperma* populations within these lineages would be useful for assessing the effect of reduced recombination on nucleotide diversity. Additionally, allele-specific gene expression experiments would be an ideal approach to determine whether there is evidence that the genes within the *mat A* nonrecombining region are evolving increased expression to compensate for the increase in deleterious substitutions that are occurring within the *mat a* nonrecombining region.

Acknowledgments

We thank Deborah Charlesworth for comments on a previous version of this manuscript and Brian Charlesworth for helpful discussion. This work was supported by National Science Foundation grant DEB-0516511 (to J.W.T.), National Institutes of Health–National Institute of General Medical Sciences grant R01RGM081597 (to J.W.T.), and the Chang-Lin Tien Graduate Fellowship (to C.E.E.). The work conducted by the U.S. Department of Energy Joint Genome Institute is supported by the Office of Science of the U.S. Department of Energy under contract no. DE-AC02-05CH11231.

Literature Cited

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

Argueso, J. L., J. Westmoreland, P. A. Mieczkowski, M. Gawel, T. D. Petes *et al.*, 2008 Double-strand breaks associated with repetitive DNA can reshape the genome. Proc. Natl. Acad. Sci. USA 105: 11845–11850.

Ashburner, M., C. A. Ball, J. A. Blake, D. Botstein, H. Butler *et al.*, 2000 Gene Ontology: tool for the unification of biology. Nat. Genet. 25: 25–29.

Bachtrog, D., 2003 Protein evolution and codon usage bias on the neo-sex chromosomes of *Drosophila miranda*. Genetics 165: 1221–1232.

Betancourt, A. J., and D. C. Presgraves, 2002 Linkage limits the power of natural selection in *Drosophila*. Proc. Natl. Acad. Sci. USA 99: 13616–13620.

Betancourt, A. J., J. J. Welch, and B. Charlesworth, 2009 Reduced effectiveness of selection caused by a lack of recombination. Curr. Biol. 19: 655–660.

Bierne, N., and A. Eyre-Walker, 2003 The problem of counting sites in the estimation of the synonymous and nonsynonymous substitution rates: implications for the correlation between the synonymous substitution rate and codon usage bias. Genetics 165: 1587–1597.

Birney, E., and R. Durbin, 2000 Using GeneWise in the *Drosophila* annotation experiment. Genome Res. 10: 547–548.

Boeckmann, B., A. Bairoch, R. Apweiler, M. C. Blatter, A. Estreicher *et al.*, 2003 The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. Nucleic Acids Res. 31: 365–370.

Bull, J. J., 1978 Sex chromosomes in haploid dioecy: unique contrast to Muller's theory for diploid dioecy. Am. Nat. 112: 245–250.

Casals, F., and A. Navarro, 2007 Inversions: The chicken or the egg? Heredity 99: 479–480.

Catcheside, D. G., 1975 Occurrence in wild strains of *Neurospora crassa* of genes controlling genetic recombination. Aust. J. Biol. Sci. 28: 213–225.

Charlesworth, B., and D. Charlesworth, 2000 The degeneration of Y chromosomes. Philos. Trans. R. Soc. Lond. B Biol. Sci. 355: 1563–1572.

Charlesworth, D., B. Charlesworth, and G. Marais, 2005 Steps in the evolution of heteromorphic sex chromosomes. Heredity 95: 118–128.

Chen, J. M., N. Chuzhanova, P. D. Stenson, C. Ferec, and D. N. Cooper, 2005 Complex gene rearrangements caused by serial replication slippage. Hum. Mutat. 26: 125–134.

Connallon, T., and A. G. Clark, 2010 Gene duplication, gene conversion and the evolution of the Y chromosome. Genetics 186: 277–286.

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

Dewey, C. N., 2007 Aligning multiple whole genomes with Mercator and MAVID. Methods Mol. Biol. 395: 221–236.

Ellison, C. E., C. Hall, D. Kowbel, J. Welch, R. B. Brem *et al.*, 2011 Population genomics and local adaptation in wild isolates of a model microbial eukaryote. Proc. Natl. Acad. Sci. USA 108: 2831–2836.

Enright, A. J., S. Van Dongen, and C. A. Ouzounis, 2002 An efficient algorithm for large-scale detection of protein families. Nucleic Acids Res. 30: 1575–1584.

Fraser, J. A., and J. Heitman, 2004 Evolution of fungal sex chromosomes. Mol. Microbiol. 51: 299–306.

Fraser, J. A., and J. Heitman, 2005 Chromosomal sex-determining regions in animals, plants and fungi. Curr. Opin. Genet. Dev. 15: 645–651.

Fraser, J. A., S. Diezmann, R. L. Subaran, A. Allen, K. B. Lengeler *et al.*, 2004 Convergent evolution of chromosomal sex-determining regions in the animal and fungal kingdoms. PLoS Biol. 2: 2243–2255.

Fraser, J. A., Y.-P. Hsueh, K. M. Findley, and J. Heitman, 2007 Evolution of the mating-type locus: the basidiomycetes, pp. 19–34 in *Sex in Fungi: Molecular Determination and Evolutionary Implications-Print*, edited by Heitman, J., J. W. Kronstad, J. W. Taylor, and L. A. Casselton. ASM Press, Washington, DC.

Galagan, J. E., and E. U. Selker, 2004 RIP: the evolutionary cost of genome defense. Trends Genet. 20: 417–423.

Gallegos, A., D. J. Jacobson, N. B. Raju, M. P. Skupski, and D. O. Natvig, 2000 Suppressed recombination and a pairing anomaly on the mating-type chromosome of *Neurospora tetrasperma*. Genetics 154: 623–633.

Gasteiger, E., A. Gattiker, C. Hoogland, I. Ivanyi, R. D. Appel *et al.*, 2003 ExPASy: the proteomics server for in-depth protein knowledge and analysis. Nucleic Acids Res. 31: 3784–3788.

Giraud, T., O. Jonot, and J. A. Shykoff, 2005 Selfing propensity under choice conditions in a parasitic fungus, *Microbotryum violaceum*, and parameters influencing infection success in artificial inoculations. Int. J. Plant Sci. 166: 649–657.

Haddrill, P. R., D. L. Halligan, D. Tomaras, and B. Charlesworth, 2007 Reduced efficacy of selection in regions of the *Drosophila* genome that lack crossing over. Genome Biol. 8: R18.

Hastings, P. J., G. Ira, and J. R. Lupski, 2009 A microhomology-mediated break-induced replication model for the origin of human copy number variation. PLoS Genet. 5: e1000327.

Hood, M. E., 2002 Dimorphic mating-type chromosomes in the fungus *Microbotryum violaceum*. Genetics 160: 457–461.

Howe, H. B., and P. Haysman, 1966 Linkage group establishment in *Neurospora tetrasperma* by interspecific hybridization with *N. crassa*. Genetics 54: 293–302.

Hunter, N., S. R. Chambers, E. J. Louis, and R. H. Borts, 1996 The mismatch repair system contributes to meiotic sterility in an interspecific yeast hybrid. EMBO J. 15: 1726–1733.

Idnurm, A., F. J. Walton, A. Floyd, and J. Heitman, 2008 Identification of the sex genes in an early diverged fungus. Nature 451: 193–196.

Isono, K., J. D. McIninch, and M. Borodovsky, 1994 Characteristic features of the nucleotide sequences of yeast mitochondrial ribosomal protein genes as analyzed by computer program GeneMark. DNA Res. 1: 263–269.

Jacobson, D. J., 1995 Sexual dysfunction associated with outcrossing in *Neurospora tetrasperma*, a pseudohomothallic Ascomycete. Mycologia 87: 604–617.

Jacobson, D. J., 2005 Blocked recombination along the mating-type chromosomes of *Neurospora tetrasperma* involves both structural heterozygosity and autosomal genes. Genetics 171: 839–843.

Jurka, J., 2000 Repbase update: a database and an electronic journal of repetitive elements. Trends Genet. 16: 418–420.

Kanehisa, M., M. Araki, S. Goto, M. Hattori, M. Hirakawa *et al.*, 2008 KEGG for linking genomes to life and the environment. Nucleic Acids Res. 36: D480–D484.

Kapitonov, V. V., and J. Jurka, 2003 Mariner-3_AN, a family of DNA transposons in the *Aspergillus nidulans* genome. Repbase Rep. 3: 196.

Kent, W. J., 2002 BLAT—the BLAST-like alignment tool. Genome Res. 12: 656–664.

Kliman, R. M., and J. Hey, 1993 Reduced natural selection associated with low recombination in *Drosophila melanogaster*. Mol. Biol. Evol. 10: 1239–1258.

Kondo, M., U. Hornung, I. Nanda, S. Imai, T. Sasaki *et al.*, 2006 Genomic organization of the sex-determining and adjacent regions of the sex chromosomes of medaka. Genome Res. 16: 815–826.

Koonin, E. V., N. D. Fedorova, J. D. Jackson, A. R. Jacobs, D. M. Krylov *et al.*, 2004 A comprehensive evolutionary classification of proteins encoded in complete eukaryotic genomes. Genome Biol. 5: R7.

Kurtz, S., A. Phillippy, A. L. Delcher, M. Smoot, M. Shumway *et al.*, 2004 Versatile and open software for comparing large genomes. Genome Biol. 5: R12.

Lafontaine, I., and B. Dujon, 2010 Origin and fate of pseudogenes in Hemiascomycetes: a comparative analysis. BMC Genomics 11: 260.

Lahn, B. T., and D. C. Page, 1999 Four evolutionary strata on the human X chromosome. Science 286: 964–967.

Lee, J. A., C. M. Carvalho, and J. R. Lupski, 2007 A DNA replication mechanism for generating nonrecurrent rearrangements associated with genomic disorders. Cell 131: 1235–1247.

Lee, N., G. Bakkeren, K. Wong, J. E. Sherwood, and J. W. Kronstad, 1999 The mating-type and pathogenicity locus of the fungus *Ustilago hordei* spans a 500-kb region. Proc. Natl. Acad. Sci. USA 96: 15026–15031.

Lengeler, K. B., D. S. Fox, J. A. Fraser, A. Allen, K. Forrester *et al.*, 2002 Mating-type locus of *Cryptococcus neoformans*: a step in the evolution of sex chromosomes. Eukaryot. Cell 1: 704–718.

Lewis, Z. A., S. Honda, T. K. Khlafallah, J. K. Jeffress, M. Freitag *et al.*, 2009 Relics of repeat-induced point mutation direct heterochromatin formation in *Neurospora crassa*. Genome Res. 19: 427–437.

Liu, Z., P. H. Moore, H. Ma, C. M. Ackerman, M. Ragiba *et al.*, 2004 A primitive Y chromosome in papaya marks incipient sex chromosome evolution. Nature 427: 348–352.

Lowe, T. M., and S. R. Eddy, 1997 TRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. 25: 955–964.

Maloisel, L., and J. L. Rossignol, 1998 Suppression of crossing-over by DNA methylation in Ascobolus. Genes Dev. 12: 1381–1389.

Mani, R. S., and A. M. Chinnaiyan, 2010 Triggers for genomic rearrangements: insights into genomic, cellular and environmental influences. Nat. Rev. Genet. 11: 819–829.

Mank, J. E., and H. Ellegren, 2007 Parallel divergence and degradation of the avian W sex chromosome. Trends Ecol. Evol. 22: 389–391.

Marais, G. A., M. Nicolas, R. Bergero, P. Chambrier, E. Kejnovsky *et al.*, 2008 Evidence for degeneration of the Y chromosome in the dioecious plant *Silene latifolia*. Curr. Biol. 18: 545–549.

Marcou, D., A. Masson, J. M. Simonet, and G. Piquepaille, 1979 Evidence for non-random spatial distribution of meiotic exchanges in *Podospora anserina*: comparison between linkage groups 1 and 6. Mol. Gen. Genet. 176: 67–79.

McClintock, B., 1984 The significance of responses of the genome to challenge. Science 226: 792–801.

Melen, K., A. Krogh, and G. von Heijne, 2003 Reliability measures for membrane protein topology prediction algorithms. J. Mol. Biol. 327: 735–744.

Menkis, A., D. J. Jacobson, T. Gustafsson, and H. Johannesson, 2008 The mating-type chromosome in the filamentous ascomycete *Neurospora tetrasperma* represents a model for early evolution of sex chromosomes. PLoS Genet. 4: e1000030.

Menkis, A., E. Bastiaans, D. J. Jacobson, and H. Johannesson, 2009 Phylogenetic and biological species diversity within the *Neurospora tetrasperma* complex. J. Evol. Biol. 22: 1923–1936.

Menkis, A., C. A. Whittle, and H. Johannesson, 2010 Gene genealogies indicates abundant gene conversions and independent evolutionary histories of the mating-type chromosomes in the evolutionary history of *Neurospora tetrasperma*. BMC Evol. Biol. 10: 234.

Merino, S. T., M. A. Nelson, D. J. Jacobson, and D. O. Natvig, 1996 Pseudohomothallism and evolution of the mating-type chromosome in *Neurospora tetrasperma*. Genetics 143: 789–799.

Metin, B., K. Findley, and J. Heitman, 2010 The mating type locus (MAT) and sexual reproduction of *Cryptococcus heveanensis*: insights into the evolution of sex and sex-determining chromosomal regions in fungi. PLoS Genet. 6: e1000961.

Nicolas, M., G. Marais, V. Hykelova, B. Janousek, V. Laporte *et al.*, 2005 A gradual process of recombination restriction in the evolutionary history of the sex chromosomes in dioecious plants. PLoS Biol. 3: e4.

Nielsen, H., J. Engelbrecht, S. Brunak, and G. Von Heijne, 1997 Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. Protein Eng. 10: 1–6.

Nygren, K., R. Strandberg, A. Wallberg, B. Nabholz, T. Gustafsson *et al.*, 2011 A comprehensive phylogeny of Neurospora reveals a link between reproductive mode and molecular evolution in fungi. Mol. Phylogenet. Evol. 59: 649–663.

O'Meally, D., H. R. Patel, R. Stiglec, S. D. Sarre, A. Georges *et al.*, 2010 Non-homologous sex chromosomes of birds and snakes share repetitive sequences. Chromosome Res. 18: 787–800.

Pearson, W. R., T. Wood, Z. Zhang, and W. Miller, 1997 Comparison of DNA sequences with protein sequences. Genomics 46: 24–36.

Peichel, C. L., J. A. Ross, C. K. Matson, M. Dickson, J. Grimwood *et al.*, 2004 The master sex-determination locus in threespine sticklebacks is on a nascent Y chromosome. Curr. Biol. 14: 1416–1424.

Powell, A. J., D. J. Jacobson, and D. O. Natvig, 2001 Allelic diversity at the het-c locus in *Neurospora tetrasperma* confirms outcrossing in nature and reveals an evolutionary dilemma for pseudohomothallic ascomycetes. J. Mol. Evol. 52: 94–102.

Presgraves, D. C., 2005 Recombination enhances protein adaptation in *Drosophila melanogaster*. Curr. Biol. 15: 1651–1656.

Raju, N. B., 1992 Functional heterothallism resulting from homokaryotic conidia and ascospores in *Neurospora tetrasperma*. Mycol. Res. 96: 103–116.

Raju, N. B., and D. D. Perkins, 1994 Diverse programs of ascus development in pseudohomothallic species of *Neurospora*, *Gelasinospora*, and *Podospora*. Dev. Genet. 15: 104–118.

Ranz, J. M., D. Maurin, Y. S. Chan, M. von Grotthuss, L. W. Hillier *et al.*, 2007 Principles of genome evolution in the *Drosophila melanogaster* species group. PLoS Biol. 5: 1366–1381.

Rice, P., I. Longden, and A. Bleasby, 2000 EMBOSS: the European Molecular Biology Open Software Suite. Trends Genet. 16: 276–277.

Rozen, S., H. Skaletsky, J. D. Marszalek, P. J. Minx, H. S. Cordum *et al.*, 2003 Abundant gene conversion between arms of palindromes in human and ape Y chromosomes. Nature 423: 873–876.

Saenz, G. S., J. G. Stam, D. J. Jacobson, and D. O. Natvig, 2001 Heteroallelism at the het-c locus contributes to sexual dysfunction in outcrossed strains of *Neurospora tetrasperma*. Fungal Genet. Biol. 34: 123–129.

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

Sharp, P. M., and W. H. Li, 1987 The Codon Adaptation Index—a measure of directional synonymous codon usage bias, and its potential applications. Nucleic Acids Res. 15: 1281–1295.

Slater, G. S., and E. Birney, 2005 Automated generation of heuristics for biological sequence comparison. BMC Bioinform. 6: 31.

Smit, A., and R. Hubley, 2008–2010 RepeatModeler Open-1.0. (http://www.repeatmasker.org).

Smit, A., R. Hubley, and P. Green, 1996–2010 RepeatMasker Open-3.0. (http://www.repeatmasker.org).

Swarbreck, S. M., E. A. Lindquist, D. D. Ackerly, and G. L. Andersen, 2011 Analysis of leaf and root transcriptomes of soil-grown *Avena barbata* plants. Plant Cell Physiol. 52: 317–332.

Veyrunes, F., P. D. Waters, P. Miethke, W. Rens, D. McMillan *et al.*, 2008 Bird-like sex chromosomes of platypus imply recent origin of mammal sex chromosomes. Genome Res. 18: 965–973.

Vogel, H. J., 1956 A convenient growth medium for *Neurospora* (Medium N). Microb. Genet. Bull. 13: 42–43.

Votintseva, A. A., and D. A. Filatov, 2009 Evolutionary strata in a small mating-type-specific region of the smut fungus *Microbotryum violaceum*. Genetics 182: 1391–1396.

Votintseva, A. A., and D. A. Filatov, 2011 DNA polymorphism in recombining and non-recombining mating-type-specific loci of the smut fungus *Microbotryum*. Heredity 106: 936–944.

Whittle, C. A., and H. Johannesson, 2011 Evidence of the accumulation of allele-specific non-synonymous substitutions in the young region of recombination suppression within the mating-type chromosomes of *Neurospora tetrasperma*. Heredity (in press).

Whittle, C. A., K. Nygren, and H. Johannesson, 2011a Consequences of reproductive mode on genome evolution in fungi. Fungal Genet. Biol. 48: 661–667.

Whittle, C. A., Y. Sun, and H. Johannesson, 2011b Degeneration in codon usage within the region of suppressed recombination in the mating type chromosomes of *Neurospora tetrasperma*. Eukaryot. Cell 10: 594–603.

Whittle, C. A., Y. Sun, and H. Johannesson, 2011c Evolution of synonymous codon usage in Neurospora tetrasperma and Neurospora discreta. Genome Biol. Evol. (in press).

Yamato, K. T., K. Ishizaki, M. Fujisawa, S. Okada, S. Nakayama *et al.*, 2007 Gene organization of the liverwort Y chromosome reveals distinct sex chromosome evolution in a haploid system. Proc. Natl. Acad. Sci. USA 104: 6472–6477.

Zdobnov, E. M., and R. Apweiler, 2001 InterProScan—an integration platform for the signature-recognition methods in InterPro. Bioinformatics 17: 847–848.

Zhang, Z., J. Li, X. Q. Zhao, J. Wang, G. K. Wong *et al.*, 2006 KaKs_Calculator: calculating Ka and Ks through model selection and model averaging. Genomics Proteomics Bioinform. 4: 259–263.

Zheng, D., A. Frankish, R. Baertsch, P. Kapranov, A. Reymond *et al.*, 2007 Pseudogenes in the ENCODE regions: consensus annotation, analysis of transcription, and evolution. Genome Res. 17: 839–851.

Zhou, Q., J. Wang, L. Huang, W. Nie, Y. Liu *et al.*, 2008 Neo-sex chromosomes in the black muntjac recapitulate incipient evolution of mammalian sex chromosomes. Genome Biol. 9: R98.

Zou, C., M. D. Lehti-Shiu, F. Thibaud-Nissen, T. Prakash, C. R. Buell *et al.*, 2009 Evolutionary and expression signatures of pseudogenes in *Arabidopsis* and rice. Plant Physiol. 151: 3–15.

Communicating editor: E. Alani

GENETICS

Supporting Information
http://www.genetics.org/content/suppl/2011/07/12/genetics.111.130690.DC1

Massive Changes in Genome Architecture Accompany the Transition to Self-Fertility in the Filamentous Fungus *Neurospora tetrasperma*

Christopher E. Ellison, Jason E. Stajich, David J. Jacobson, Donald O. Natvig, Alla Lapidus, Brian Foster, Andrea Aerts, Robert Riley, Erika A. Lindquist, Igor V. Grigoriev, and John W. Taylor

DOI: 10.1534/genetics.111.130690

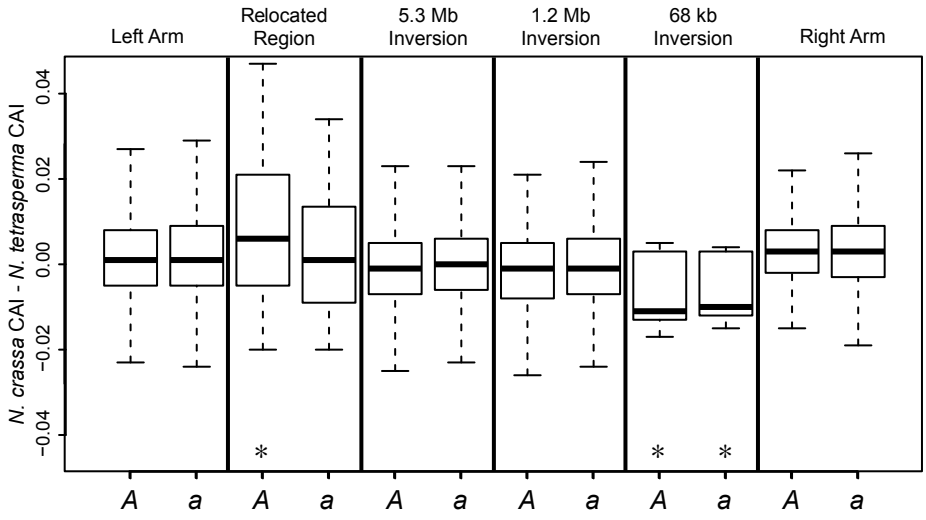

Figure S1 Re-analysis of Codon Adaptation Index (CAI) differences between *N. crassa* and *N. tetrasperma*. The analysis shown in Figure 4 was based on codon usage in *N. crassa*. To verify that the results shown in Figure 4 do not depend on the species used to obtain codon usage information, we repeated the analysis with codon usage information from *N. tetrasperma mat A* and obtained similar results. The genes within the *N. tetrasperma mat A* relocated region appear to be evolving reduced codon usage bias (one-sided permutation test: P=0.0015) while the genes within the small inversion appear to be evolving increased codon usage bias (one-sided permutation test: P=0.0006 [*mat A*], P=0.0003 [*mat a*]). Outliers are not shown.

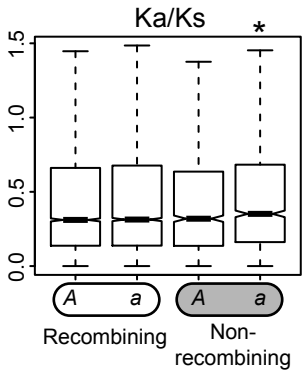

Figure S2 Re-analysis of the ratio of nonsynonymous (Ka) and synonymous (Ks) substitutions per site between recombining and non-recombining regions of the genome. The results presented in Figure 6 were reanalyzed using only four-fold degenerate sites. Genes within the region of suppressed recombination in the *mat a* strain, but not the *mat A* strain, have significantly larger Ka/Ks ratios than those outside of the non-recombining region (MWU test: *P*=0.04 and *P*=0.4, respectively).

Table S1 Pseudogenes present within the *N. tetrasperma* non-recombining region

Gene ID	Function	Strain	Mutation
NCU03134	class I alpha-mannosidase 1A	*mat A*	FS
NCU01896	Unknown	*mat A*	FS
NCU02817	Unknown	*mat A*	PS
NCU00765	amino acid permease 2	*mat a*	PS
NCU02882	Unknown	*mat a*	PS
NCU03288	Unknown	*mat a*	PS
NCU10125	Unknown	*mat a*	PS
NCU08351	Unknown	*mat a*	PS
NCU11480	Major Facilitator Superfamily	Both	*mat A*: PS; *mat a*: FS
NCU03275	Unknown	Both	*mat A*: PS; *mat a*: PS and FS

Pseudogenes were identified as genes in one or both of the *N. tetrasperma* strains where mutations or insertions/deletions resulted in premature stop codons (PS) and/or frameshifts (FS) relative to their orthologs in *N. discreta* and *N. crassa*.

Table S2 Assembly statistics

Nuclear Genome Assembly	*Neurospora tetrasperma* FGSC 2508 *mat A* v2.0	*Neurospora tetrasperma* FGSC 2509 *mat a* v1.0
Sequencing platform	Hybrid 454/Sanger	Hybrid 454/Sanger
Scaffold count	81	307
All Contig count	551	861
Scaffold sequence bases total	39.1 Mb	39.1 Mb
Scaffolded (Large) Contig sequence bases total	38.5 Mb	38.1 Mb
Estimated % sequence bases in gaps	1.7%	2.5%
Scaffold N50 / L50	3 / 5.7 Mb	3 / 5.7 Mb
Contig N50 / L50	89 / 134.9 Kb	117 / 99.8 Kb
Number of scaffolds > 50.0 Kb	7	7
% in scaffolds > 50.0 Kb	99.2%	98.2%

Table S3 Sequencing libraries for *Neurospora tetrasperma* FGSC 2508 *mat A*

Library	Library Type	Raw Reads	Raw Bases	Trimmed Bases	Assem Reads	Assem Bases	Coverage	Insert	Std Dev
FHCI	SANG	232,238	242,846,510	180,826,637	228,004	176,589,447	4.58x	2822	705
FHCN	SANG	209,117	181,677,551	156,903,722	205,454	154,099,062	3.99x	8558	2139
FHCO	SANG	52,883	44,594,073	34,316,772	49,195	32,285,178	0.84x	37937	9484
GBNH	454	701,361	269,672,184	268,587,339	679,800	261,469,578	6.78x		
GCSS	454PE	1,329,018	329,435,293	287,196,298	1,229,007	277,086,984	7.18x	17894	4473
GSHS	454PE	1,718,929	358,804,691	317,600,455	1,565,911	305,662,912	7.92x	3550	887
Total		**4,243,546**	**1,427,030,302**	**1,245,431,223**	**3,957,371**	**1,207,193,161**	**31.29x**		

Table S4 Sequencing libraries for *Neurospora tetrasperma* FGSC 2509 *mat a*

Library	Library Type	Raw Reads	Raw Bases	Trimmed Bases	Assem Reads	Assem Bases	Coverage	Insert	Std Dev
FHFB	SANG	39,341	29,329,148	23,656,065	36,529	21,613,054	0.57x	39775	9943
GCXS	454	2,224,123	824,396,125	821,323,297	2,083,410	770,085,673	20.18x		
GGUB	454PE	306,099	90,121,826	67,446,884	260,379	59,046,315	1.55x	20556	5139
GGUI	454	1,849,195	787,934,733	785,260,954	1,761,107	748,048,169	19.60x		
GHTG	454PE	968,920	257,042,191	230,156,414	874,202	215,706,069	5.65x	13704	3426
Total		**5,387,678**	**1,988,824,023**	**1,927,843,614**	**5,015,627**	**1,814,499,280**	**47.55x**		

Table S5 Assessment of results from Whittle *et al* 2011b

Substitution type	*mat a*	*mat A*
Non-preferred to preferred	134	128
Preferred to non-preferred	232	263

The numbers reported in Whittle *et al* 2011b were compared to determine if there is evidence for a statistically significant asymmetry in synonymous codon substitutions between the lineage #1 *N. tetrasperma mat a* and *mat A* non-recombining regions. There is not a statistically significant difference between these two regions in the proportion of non-preferred to preferred substitutions relative to preferred to non-preferred (Fisher's exact test: P=0.285).

Table S6 Gene model statistics

Nuclear Genome Annotation	*Neurospora tetrasperma* FGSC 2508 *mat A* v2.0	*Neurospora tetrasperma* FGSC 2509 *mat a* v2.0
# gene models	10,380	11,192
Gene density	265	286
Avg. gene length	1836	1800
Avg. protein length	468	443
Avg. exon frequency	2.72 exons/gene	2.7 exons/gene
Avg. exon length	579	548
Avg. intron length	154	152
% complete gene models (with start and stop codons)	94%	92%
% genes with homology support	91%	87%
% genes with Pfam domains	50%	46%