Downloaded from genome.cshlp.org on November 7, 2026 . Published by Cold Spring Harbor Laboratory Press



Introgression maintains the genetic integrity of the sex-determining chromosome of the fungus *Neurospora tetrasperma*

Padraic Corcoran, Jennifer L Anderson, David J Jacobson, et al.

Genome Res. published online February 18, 2016
Access the most recent version at doi:10.1101/gr.197244.115

P<P Published online February 18, 2016 in advance of the print journal.

Accepted Manuscript Peer-reviewed and accepted for publication but not copyedited or typeset; accepted manuscript is likely to differ from the final, published version.

Open Access Freely available online through the *Genome Research* Open Access option.

Creative Commons License This manuscript is Open Access.This article, published in *Genome Research*, is available under a Creative Commons License (Attribution 4.0 International license), as described at http://creativecommons.org/licenses/by/4.0/.

Email Alerting Service Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or **click here.**



To subscribe to *Genome Research* go to:
https://genome.cshlp.org/subscriptions

Published by Cold Spring Harbor Laboratory Press

Downloaded from genome.cshlp.org on November 7, 2026 . Published by Cold Spring Harbor Laboratory Press

Introgression maintains the genetic integrity of the mating-type determining chromosome of the fungus *Neurospora tetrasperma*

Pádraic Corcoran[1,2], Jennifer L Anderson[1], David J Jacobson[1], Yu Sun[3], Peixiang Ni[4], Martin Lascoux[5] and Hanna Johannesson[1*]

[1] Department of Organismal Biology, Uppsala University, Sweden.

[2] Department of Animal and Plant Sciences, University of Sheffield, Sheffield S10 2TN, United Kingdom

[3] Department of Cell- and Molecular Biology, Uppsala University, Sweden

[4] Beijing Genome Institute, Hong Kong.

[5] Department of Ecology and Genetics, Science for Life laboratory, Uppsala University, Sweden.

*Corresponding author: Hanna Johannesson, Department of Organismal Biology, Norbyvägen 18D, SE-752 36 Uppsala, Phone: +46 18 4716479, fax: +46 18 4716310, e-mail: hanna.johannesson@ebc.uu.se

Running title: Introgression and genetic integrity in Neurospora

Key words: introgression, population genomics, recombination suppression, mating-type chromosome, Neurospora

Downloaded from genome.cshlp.org on November 7, 2026 . Published by Cold Spring Harbor Laboratory Press

Abstract

Genome evolution is driven by a complex interplay of factors, including selection, recombination, and introgression. The regions determining sexual identity are particularly dynamic parts of eukaryotic genomes that are prone to molecular degeneration associated with suppressed recombination. In the fungus *Neurospora tetrasperma*, it has been proposed that this molecular degeneration is counteracted by the introgression of non-degenerated DNA from closely related species. In this study, we used comparative and population genomic analyses of variation among 92 genomes from eight phylogenetically and reproductively isolated lineages of *N. tetrasperma*, and its three closest relatives, to investigate the factors shaping the evolutionary history of the genomes. We found that suppressed recombination extends across at least 6 Mbp (~63%) of the mating-type (*mat*) chromosome in *N. tetrasperma*, and is associated with decreased genetic diversity, which is likely the result primarily of selection at linked sites. Furthermore, analyses of molecular evolution revealed an increased mutational load in this region, relative to recombining regions. However, comparative genomic and phylogenetic analyses indicate that the *mat* chromosomes are temporarily regenerated via introgression from sister species; six out of eight lineages show introgression into one of their *mat* chromosomes, with at least three other *Neurospora* species acting as donors. The introgressed tracts have been driven to fixation in lineages, suggesting that they confer an adaptive advantage in natural populations, and our analyses support the presence of selective sweeps in at least one of the lineages. Thus, these data strongly support the previously hypothesized role of introgression as a mechanism for the maintenance of mating-type determining chromosomal regions.

Downloaded from genome.cshlp.org on November 7, 2026 . Published by Cold Spring Harbor Laboratory Press

Introduction

The content, structure and organization of eukaryote genomes change over time in response to complex interactions between selection, mutation, recombination, introgression and other factors. Chromosomal regions conferring sexual identity (*i.e.*, sex or mating type) are particularly dynamic parts of eukaryote genomes, evolving independently and divergently from a formerly homologous state (Bull 1983). This divergent evolution is associated with suppressed recombination between the chromosomes that effectively preserves their unique identities. Low, or suppressed, recombination leads to a reduction in effective population size (N_e) and introduces selection at linked sites; both factors reduce the efficacy of selection and are expected to result in an increased mutational load (Bachtrog and Charlesworth 2002; Charlesworth and Charlesworth 2010). The link between recombination suppression and molecular degeneration has been observed in the sex- and mating-type determining genomic regions of a number of taxa across all eukaryote kingdoms (*e.g.*, Bachtrog 2003; Hood et al. 2004; Liu et al. 2004; Marais et al. 2008; Whittle and Johannesson 2011; Whittle et al. 2011a; Fontanillas et al. 2015), and is one of the main factors expected to favor either rare recombination events on sex chromosomes (Malcom et al. 2014) or a high turnover of chromosomes harboring the sex-determining loci (Blaser et al. 2013).

In the filamentous ascomycete *Neurospora tetrasperma*, it has been proposed that introgression, the integration of genetic material from one species into the genome of another, serves to renew and maintain the integrity of the mating-type determining chromosomal regions (Sun et al. 2012). This species has independently evolved pseudohomothallism, a mating system in which self-fertility is achieved through the production of sexual spores that contain nuclei of both mating types (*i.e.*,

Downloaded from genome.cshlp.org on November 7, 2026 . Published by Cold Spring Harbor Laboratory Press

it is heterokaryotic for *mat A* and *mat a,* see Supplementary Fig. 1) (Raju and Perkins 1994). Pseudohomothallism in *N. tetrasperma* is thought to have evolved from a heterothallic ancestor, for which sexual spores are of single mating type (Supplementary Fig. 1), about one million years ago (Corcoran et al. 2014). To accomplish correct nuclear packaging in the spores of *N. tetrasperma*, crossing over is suppressed between the *mat* locus and the centromere, ensuring that *mat A* and *mat a* will segregate at the first division of meiosis. Indeed, in *N. tetrasperma* recombination is suppressed across most of the mating-type (*mat*) chromosome in all strains examined to date (Howe and Haysman 1966; Merino et al. 1996; Gallegos et al. 2000; Menkis et al. 2008; Ellison et al. 2011b). In the *Neurospora* genus, this recombination suppression is unique to *N. tetrasperma* and is not found in its heterothallic sister taxa, for which the *mat* chromosomes freely recombine except in a very short region (3-5 kb) surrounding the *mat* locus (Glass et al. 1990; Staben and Yanofsky 1990). Furthermore, in *N. tetrasperma* suppressed recombination is accompanied by an accumulation of non-beneficial mutations (Ellison et al. 2011b; Whittle and Johannesson 2011; Whittle et al. 2011a; Sun et al. 2012). Sun et al. (2012) used comparative genomics of six *N. tetrasperma* strains to show that introgression of the *mat* chromosomes from other freely recombining *Neurospora* species may have reduced degeneration on the *mat a* chromosomes. Here, we use a large-scale genomic sampling of 92 genomes, and a population and comparative genomic approach to evaluate the roles of selection, recombination, and introgression in shaping the *mat* chromosomes of *N. tetrasperma* populations.

Downloaded from genome.cshlp.org on November 7, 2026 . Published by Cold Spring Harbor Laboratory Press

Results

Global pattern of variation in *N. tetrasperma*

Genome sequencing (to mean coverage of 25-45x) and reference assembly of 92 strains of *N. tetrasperma* from across the globe (Supplementary Table 1) resulted in the discovery of 1,693,770 bi-allelic single nucleotide polymorphisms (SNPs) within this clade. After filtering of heterokaryotic strains and clones (Supplementary Fig. 2, Supplementary Table 1), we analyzed the autosomes (the set of six chromosomes corresponding to linkage groups II to VII in *N. crassa*) to reveal the global pattern of variation in *N. tetrasperma*. The largest chromosome, the *mat* chromosome (linkage group I in *N. crassa*), was excluded due to the large regions of suppressed recombination on this chromosome in *N. tetrasperma*. All strains of *N. tetrasperma* form a monophyletic group, as confirmed by both Maximum Likelihood phylogenomic analysis of variable sites, and a species tree inference of autosomal gene trees (Fig. 1A and Supplementary Fig. 3). Furthermore, for the first time we show with strong phylogenetic support that *N. sitophila* is the sister taxon of *N. tetrasperma* (Supplementary Fig. 3)(*c.f.,* Dettman et al. 2003; Corcoran et al. 2014). Additionally, phylogenomic and principal component analyses confirms the previously defined lineages of *N. tetrasperma* (Fig. 1A, B)(Corcoran et al. 2014), henceforth referred to as L1 to L10. In accordance with previous studies (Saenz et al. 2003; Menkis et al. 2009; Corcoran et al. 2014), lineages primarily correlate with geographical region, although this pattern is not universal, for example, the strains of *N. tetrasperma* from Louisiana (LA) belong to three genetically divergent lineages (L1, L7 and L8; Fig. 1A, Fig. 1B). Despite lineages of the *N. tetrasperma* clade constituting well-supported phylogenetic groups (Fig. 1A), Bayesian clustering

Downloaded from genome.cshlp.org on November 7, 2026 . Published by Cold Spring Harbor Laboratory Press

126 analysis on a randomly chosen subset of 9000 autosomal SNPs indicates that the
127 genomic ancestry of only five of the lineages belongs to one population (Fig. 1C). L4,
128 L9 and L10 show mosaic ancestries, which may result from past hybridization
129 between lineages, or in the case of L4 and L9, may reflect the inability to assign their
130 ancestry to a single population given the small sample sizes for these lineages (Fig.
131 1C, Supplementary Fig. 4).
132
133 **A history of selfing and admixture in *N. tetrasperma***
134 Our analyses strengthen the view of *N. tetrasperma* as a predominantly selfing
135 species. First, linkage disequilibrium (LD) is much more extensive in all *N.*
136 *tetrasperma* lineages than previously determined for populations of the heterothallic
137 close relative *N. crassa* (Ellison et al. 2011a)(Supplementary Table 2). The levels of
138 LD observed in L5 and L8 extend for 11Kb and 31Kb, respectively, and in L10, LD
139 extends for hundreds of Kb on some chromosomes (Supplementary Figs 5 and 6).
140 Second, few differences were found across the autosomes of *mat A* and *mat a*
141 homokaryons (*i.e.*, strains containing nuclei of a single mating type) isolated from the
142 same natural heterokaryon (Fig. 2, Supplementary Fig. 7), and the vast majority of
143 such paired mating-type homokaryons (*e.g.*, CJ57 *A* and CJ58 *a* from L8) group
144 together in the phylogeny of Figure 1A.
145 However, the allelic distribution of two autosomal heterokaryon
146 incompatibility (*het*) genes supports a history of occasional outcrossing in *N.*
147 *tetrasperma*. These genes govern self-nonself recognition in natural fungal
148 populations, and typically evolve under balancing selection and therefore maintain
149 ancestral polymorphism through speciation (*e.g.*, Powell et al. 2007). We found
150 different alleles of *het* genes among closely related strains of *N. tetrasperma* and

Downloaded from genome.cshlp.org on November 7, 2026 . Published by Cold Spring Harbor Laboratory Press

151 shared alleles in distantly related lineages (Supplementary Figs 8 and 9), a pattern
152 inconsistent with obligate selfing in the history of *N. tetrasperma* (*c.f.*, Powell et al.
153 2001; Menkis et al. 2009).

154 A notable exception to the pattern of phylogenetic grouping of paired
155 autosomes derives from natural heterokaryons is L10, where homokaryon pairs
156 originating from natural heterokaryons do not cluster together (Fig. 1 A), and strains
157 are highly divergent (Fig. 2, Supplementary Fig. 7). Many of the chromosome pairs
158 within heterokaryons of L10 are extensively differentiated from each other (Fig. 2,
159 Supplementary Fig. 7), a pattern which is particularly striking on LG IV, with a
160 chromosomal divergence of over 2% (Fig. 2, Supplementary Fig. 7). High genetic
161 variation (Supplementary Table 2), an excess of intermediate frequency variants
162 across the genome (Supplementary Fig. 10) and extensive linkage disequilibrium
163 (Table 1, Supplementary Figs 5 and 6) are consistent with a history of recent
164 admixture in L10.
165

166 **Recombination suppression and reduced diversity of the *mat* chromosomes**
167 In *N. tetrasperma*, unlike its heterothallic sister taxa, suppression of recombination
168 between the *mat* locus and the centromere is found. Without recombination, genetic
169 material in regions with suppressed recombination (SR) on homologous chromosomes
170 is expected to diverge. Indeed, for all strains of *N. tetrasperma* studied here, pairs of
171 *mat A* and *mat a* chromosomes originating from the same heterokaryons harbor large
172 regions of elevated divergence, a pattern in stark contrast to most of the autosomes
173 (Fig. 2, Supplementary Fig. 7). Furthermore, all SNPs within this region in L5, L8
174 and L10 (lineages with sample sizes allowing for population level analyses), were

Downloaded from genome.cshlp.org on November 7, 2026 . Published by Cold Spring Harbor Laboratory Press

found to be in near complete LD (Fig. 3A), indicating that recombination is absent in this region.

Our results show that the extent of divergence between *mat A* and *mat a* chromosomes vary significantly across lineages (Kruskal-Wallis test, $p < 0.001$) (Fig. 2, Supplementary Figs 7 and 11, Table 1). Divergence between the SR regions of *mat A* and *mat a* is highest in L6 at 3.2 %, which is more than two-fold greater than the lowest divergence of 1.4% observed in L8 (Supplementary Fig. 11, Table 1). The region of the reference genome corresponding to the SR regions also vary in size, ranging from 6 Mbp for L8, to 8.1 Mbp for L4, which translates into 63% to 86% of the entire *mat* chromosome, the largest chromosome in this species (Table 1). Note that the given sizes of the SR regions are relative to the reference genome assembly; actual sizes of the SR regions may differ in cases where an individual genome differs from the reference, but are not possible to assess with these data.

As a direct effect of recombination suppression, we expect N_e of the *mat* chromosomes to be reduced to at least half of that of the autosomes (*c.f.* Kimura 1983; Charlesworth and Charlesworth 2000), and selection at linked sites is expected to further reduce N_e in this region of the genome. Accordingly, the diversity of all investigated *N. tetrasperma* lineages is greatly reduced in the SR region, as compared to recombining chromosomal (R) regions. Specifically, when analyzing the synonymous nucleotide diversity (π_s) in the SR and R regions we found that they differ by over 75-fold in L10, up to ~24-fold in L8, and up to ~5-fold in L5 (Supplementary Fig. 12, Supplementary Table 3).

Widespread occurrence of *mat* chromosome introgressions in *N. tetrasperma*

Downloaded from genome.cshlp.org on November 7, 2026 . Published by Cold Spring Harbor Laboratory Press

Hybridization of *N. tetrasperma* with other heterothallic species is predicted to leave long tracts of introgression in the SR regions of the *N. tetrasperma mat* chromosomes, due to the lack of recombination to break them up over time (Fig 4 A). Using genomic scans of divergence between *N. tetrasperma* lineages and heterothallic species, and phylogenetic analysis of *mat* chromosome genes, we found evidence that introgression into the *mat* chromosomes has occurred in six out of the eight investigated lineages of *N. tetrasperma* (Table 2). Moreover, introgressions have originated from multiple species of *Neurospora*: at least three heterothallic species appear to have been donors to *N. tetrasperma* (Fig. 5, Table 2, Supplementary Table 4).

Typically, divergence in the SR regions is higher within than between lineages of *N. tetrasperma* (Supplementary Fig. 13), suggesting that these regions of the genome have not diverged through mutations alone. When comparing the genomes of *N. tetrasperma* strains to genomes of the heterothallic species *N. crassa*, *N. hispaniola* and *N. sitophila*, we observed cases where one of the SR regions (in either the *A* or *a* homokaryon) is significantly more similar to a heterothallic species than the opposite *mat* SR and R regions for that strain (Fig. 4B, Supplementary Fig. 14) suggesting that these regions have been introgressed from other species. This pattern is notable in comparisons between L10A (*i.e.*, the *mat A* chromosome of lineage 10), L4a and L7a and *N. hispaniola*, and between L9a and *N. crassa* (Fig 4B, Supplementary Fig 13, Supplementary Table 4). The SR regions of these chromosomes are visible as long tracts of low divergence to one or the other of the investigated heterothallic species (Fig. 4C, Supplementary Fig. 14). Also, in L10, for which population level analyses are possible, we found that *mat A* chromosomes share a large excess of derived alleles with *N. hispaniola* within the SR region, further supporting introgression from *N.*

Downloaded from genome.cshlp.org on November 7, 2026 . Published by Cold Spring Harbor Laboratory Press

hispaniola in this lineage (Supplementary Fig. 16). In the *mat A* SR regions of L5 and L6, our analyses show that the divergence from *N. hispaniola* and *N. sitophila* are higher than the other strains of *N. tetrasperma* (Fig. 4C), suggesting that they have been introgressed from more distantly related species.

To test whether the regions of low sequence divergence across the *mat* chromosome (Fig 4 C) can be explained by variance in coalescent times across the genome due to drift in the ancestral population of the species being compared, we simulated sequence data for species pairs under a multispecies coalescent model without gene flow, using the MCcoal program (Rannala and Yang 2003). If the regions of low divergence are due to introgression, they should be extreme outliers compared to the divergence calculated from the simulated data. Using this approach, we found that the simulations could not reproduce the low levels of sequence divergence that we observe between L10A, L4a and L7a and *N. hispaniola*, providing additional for the hypothesis that these SR regions have been introgressed (Supplementary Table 4), while a pure isolation model, with no introgression, can explain the levels of divergence observed across comparisons for the other lineages (Supplementary Table 4). Interestingly, for L9a the divergence to *N. crassa* genome within the SR region is not significantly different from that expected under a model without gene flow ($p = 0.25$, Supplementary Table 4), a result that may be due to the donor species being distant from *N. crassa*, as supported by the phylogenetic analysis outlined below.

Phylogenetic analysis of six genes in the SR region and two genes from the pseudoautosomal regions (PAR: the recombining flanking regions of the *mat* chromosome), show that L10A, L4a and L7a all group closely with the *N. hispaniola* strains, further supporting that these regions have been introgressed from *N.*

Downloaded from genome.cshlp.org on November 7, 2026 . Published by Cold Spring Harbor Laboratory Press

hispaniola. Furthermore, L9a is closest to *N. perkinsii* for the SR region (Fig. 5) (0.5% divergence at the investigated genes), a species close to *N. crassa* but not included in the comparative genomic analysis shown in Fig. 4, while the heterothallic species closest to L5A and L6A could not be identified with certainty.

The PAR has a pattern similar to the autosomal phylogenetic tree with *A* and *a* strains from the same lineage forming clades. However, *N. tetrasperma* lineages do not form a monophyletic group in the PAR. This may be due to introgression or incomplete lineage sorting, but few conclusive explanations can drawn from the analysis of two genes.

Evidence of autosomal introgression supports interspecific hybridization in *Neurospora*

Genome-wide analyses of sequence divergence to the heterothallic *Neurospora* species show that large tracts of introgression are mainly restricted to the *mat* chromosome of *N. tetrasperma* (Supplementary Fig. 16). However, we found evidence that the autosomes of *N. tetrasperma* lineages have patterns consistent with a history of gene flow with their heterothallic relatives. We observe long tracts of low divergence to *N. hispaniola* in the center of LGIV (Supplementary Fig. 16) in four strains of L10. These tracts correspond to large regions of elevated divergence between the homokaryons isolated from those heterokaryons (Fig. 2 and Supplementary Fig. 7). The diverged regions are much larger than the expected centromeric regions in *Neurospora* and the specific tracts mentioned above do not include the centromere (see Sun et al. 2012), and thus can not be explained by maintenance of heterozygosity of centromeric regions under automixis (Hood and

Downloaded from genome.cshlp.org on November 7, 2026 . Published by Cold Spring Harbor Laboratory Press

Antonovics 2000). This long tract on LGIV may instead be due to intogression into L10 being more recent than for the other lineages of *N. tetrasperma*.

To specifically test for a history of introgression on the *N. tetrasperma* autosomes we used three approaches: Patterson's D-statistics (Green et al. 2010), TreeMix analysis (Pickrell and Pritchard 2012) and a likelihood ratio test of post divergence gene flow with the program 3S (Yang 2010). The D-statistics were calculated for the autosomes using 'ABBA-BABA' site patterns using the genomic data available from heterothallic species. Using this approach, we found that all L10 strains show signatures of autosomal introgression from *N. hispaniola* (significantly positive D-statistic for all strains (Supplementary Table 5) with L1 strain 9033 as P1, L10 as P2, *N. hispaniola* as P3 and *N. discreta* as the outgroup), while no such signal was detected when using *N. sitophila* or *N. crassa* as P3 (Supplementary Tables 6 and 7, respectively). Lineages 4, 7, and 8 also show signatures of autosomal introgression from *N. hispaniola* (Supplementary Table 5), consistent with the patterns of introgression found on the *mat* chromosomes of L4 and L7 (Table 2).

We used a Treemix analysis to reconstruct evolutionary relationships of *N. tetrasperma* from autosomal SNPs and to see whether a history of migration between lineages gives a better fit to the data than a tree-like history based on visual inspection of the plot of residuals (Supplementary Fig. 17). The addition of migration events to the model improves the fit to the data (Supplementary Fig. 17 B and C). With the addition of five migration events, we found support for a history where L10 is an admixed lineage with past migration from L4 and *N. hispaniola* (Supplementary Fig. 17C) and for gene flow from L7 to *N. hispaniola*.

Finally, the likelihood ratio test of gene flow implemented with the 3S program indicated that the divergence model including gene flow between the

Downloaded from genome.cshlp.org on November 7, 2026 . Published by Cold Spring Harbor Laboratory Press

heterothallic species and *N. tetrasperma* lineages fits the data better than a pure isolation model for all comparisons involving *N. hispaniola* and *N. sitophila* for the autosomes. A model without gene flow shows a better fit to the data for all comparisons with *N. crassa*, with the exception of L9, L1 and L4 where the model with gene flow is preferred (Supplementary Table 8).

Introgression reduces mutational load in regions of suppressed recombination

Suppression of recombination and reduced N_e are expected to lead to a decrease in the efficacy of natural selection and to contribute to the degeneration of regions determining sexual identity (Charlesworth and Charlesworth 2000). As outlined below, we observe a pattern of mutation accumulation consistent with a reduced efficacy of selection in the SR region within each lineage of *N. tetrasperma*. Our analyses also show population level support for the hypothesis that introgression from species that do not suffer from SR related mutational load acts to reduce the mutational load on *N. tetrasperma mat* chromosomes (Table 2).

We assayed mutational load in the *N. tetrasperma* genomes by estimating three parameters: i) the ratio of non-synonymous polymorphisms per non-synonymous site to synonymous polymorphisms per synonymous site (P_N/P_S), ii) the non-synonymous/synonymous substitution rate (d_N/d_S), and iii) the switches from optimal to non-optimal synonymous codons. The accumulation of non-synonymous polymorphisms/substitutions and non-optimal codons are typically interpreted as molecular degeneration in regions with reduced recombination (Betancourt et al. 2009; Whittle et al. 2011a). In L5, L8, and L10 the P_N/P_S ratio is significantly higher in all SR than R regions (Supplementary Table 9; Fisher's exact test (FET), $p < 0.001$ for all SR v R comparisons), a result consistent with a reduced N_e. In L10 and L5 the

Downloaded from genome.cshlp.org on November 7, 2026 . Published by Cold Spring Harbor Laboratory Press

introgressed *mat A* regions have a significantly lower P_N/P_S than the *mat a* (Supplementary Table 9; FET, L10 : $p < 0.001$ and L5 $p < 0.05$), while in L8 (not introgressed) P_N/P_S between the *mat* chromosomes is not significantly different (Table 2, Supplementary Fig. 18).

Analyses of substitution rates confirm the prediction of a lower non-synonymous/synonymous substitution rate (d_N/d_S) within the introgressed SR regions of the *mat* chromosomes (Table 2). Specifically, in the introgressed L7a and L9a regions, we observed lower d_N/d_S ratios compared to genes in the *mat A* SR-region of the same lineages, although this was only significant in L9 (Supplementary Table 10, Mann-Whitney *U* test (MWU), $p = 0.055$ and $p < 0.001$ for L7 and L9, respectively). In the cases of L10 and L5 the introgressed *mat A* SR regions had significantly lower d_N/d_S than the *mat a* (MWU, $p < 0.05$ and $p < 0.001$ for L5 and L10 respectively), indicating less degeneration in the introgressed *mat A* chromosomes. For L6 and L4, which were excluded from this analysis due to high similarity to other lineages, the introgressed *mat* region was previously confirmed to exhibit a lower d_N/d_S ratio than its non-introgressed counterparts (Ellison et al. 2011b; Sun et al. 2012). In the two lineages for which introgression was not confirmed in this study (L1 and L8), we found no significant differences in d_N/d_S ratio between the *mat* chromosomes (Table 2). The result for L1 differs from that reported by Sun et. al (2012), who found that the *mat A* chromosome was more degenerated than the *mat a* counterpart in this lineage. The lack of a significant difference in this study may be due to the much larger number of genes and taxa included in the analysis.

The higher d_N/d_S values of the non-introgressed *mat* chromosomes in *N. tetrasperma* is unlikely to be the result of a history of positive selection acting specifically at these sites in these lineages, because the efficacy of selection is

Downloaded from genome.cshlp.org on November 7, 2026 . Published by Cold Spring Harbor Laboratory Press

reduced, as indicated by an elevated P_N/P_S, within the SR regions (see above, and Supplementary Fig. 18). This reduction in efficacy of selection may result in an elevated d_N/d_S due to the fixation of deleterious non-synonymous mutations.

Finally, analyses of codon usage also support the notion that introgressed regions have a lower level of degeneration: the introgressed L9a and L6A both exhibit a significant excess of switches from non-optimal to optimal codons relative to the non-introgressed *mat* chromosome in those lineages (Fisher's exact test (FET) L6 $p < 0.001$; L9 $p = 0.0003$). The number of switches to non-optimal codons is lower in the introgressed L4a relative to L4A (FET $p < 0.05$,Table 2, Supplementary Table 11). However, no significant differences in switches from optimal to non-optimal, and *vice versa*, were observed for the other lineages where introgression of one of the *mat* chromosomes has occurred (Table 2, Supplementary Table 11).

Models of linked selection on the *mat* chromosomes

The large SR regions on the *mat* chromosomes of *N. tetrasperma* should be susceptible to the effects of both selective sweeps and background selection (Charlesworth and Charlesworth 2010). We calculated Tajima's D and Fay and Wu's H statistics within the SR regions of L5, L8 and L10 to test for selective sweeps. Tajima's D statistic (*D*) can detect an excess of low frequency variants following a selective sweep in a region, while Fay and Wu's H statistic (*H*) can detect the excess of high frequency derived mutations that remain following a recent sweep (Fig. 3B)(Tajima 1989; Fay and Wu 2000; Zeng et al. 2006). These statistics were calculated within a lineage on the *mat A* and *mat a* SR regions separately and support the history of selective sweeps on the non-introgressed L8a region and the introgressed L10A region. Specifically, the SR region of most of the *mat*

Downloaded from genome.cshlp.org on November 7, 2026 . Published by Cold Spring Harbor Laboratory Press

chromosomes in this analysis (*i.e.*, L5A and L5a, L8a and L10A) shows a negative *D* within the SR region (Fig. 3), indicating an excess of low frequency polymorphisms in the regions. However, only L8a has a significantly negative *D* for the SR region (*D* = -2.078, $P < 0.0001$ from 10,000 coalescent simulations without recombination).

A recent selective sweep event within an SR region should leave a signature of negative *H* values in the flanking recombining regions. Hence, in order to identify regions that were subject to recent hard selective sweeps we divided the genome in non-overlapping 100kb windows to detect regions of the genome with most extreme (top 5%) negative *H*. Using this approach, we observed two adjacent windows of extreme negative *H* ~300kb outside the L8a SR boundary between position 7.8 Mb – 8.0 Mb, and a 100kb window ~100kb outside the L10A SR boundary from position 7.7Mb – 7.8 Mb.

To test further for signatures of selective sweeps, we carried out coalescent simulations for a single hard sweep model in a non-recombining region and determined that such a hard selective sweep model does not fit the data for the majority of the SR regions examined. The observed measures of neutral diversity from the SR regions fit well with a large range of values of two parameters varied in the sweep model, T_s (the time since the sweep occurred) and θ_0 (the neutral variation present in the absence of a sweep) for L10 and L5 SR regions (Supplementary Fig. 19). However, the L8a SR region fits a much narrower range of parameter values for the sweep model than the other SR regions, with a narrow range of T_s that are more recent than observed in the other SR regions (Supplementary Fig. 19). These data are also consistent with values of θ_0 in the model that overlap with the levels of neutral diversity observed on the autosomes. The observation that a selective sweep model

Downloaded from genome.cshlp.org on November 7, 2026 . Published by Cold Spring Harbor Laboratory Press

398 only fits one (L8a) of the six SR regions examined here suggests that selective sweeps
399 are difficult to detect or may be rare in the SR regions of *N. tetrasperma*.
400
401 **Discussion**
402 In this study, we extend previous genomic studies of *N. tetrasperma* (*e.g.*, Sun et al.
403 2012) to include population level analyses of genomes from multiple lineages within
404 the species, and provide novel insights into the complex interactions between
405 recombination, introgression and selection in natural eukaryote populations. In
406 particular, we extend the understanding of factors driving the evolution of non-
407 recombining genomic regions by showing that adaptive introgression contributes to
408 evolutionary transitions of chromosomal regions determining mating-type, and that
409 this has the potential to generate cycles of reinvigoration following accumulation of
410 deleterious mutations. First, we found that suppressed recombination of the *mat*
411 chromosome is associated with decreased genetic diversity. Although such low
412 diversity is commonly found for non-recombining sex chromosomes in animals and
413 plants (Bachtrog and Charlesworth 2002; Hellborg and Ellegren 2004; Laporte et al.
414 2005), this provides, to our knowledge, the first demonstration of a reduction in
415 genetic variation in non-recombining genomic regions conferring sexual identity in
416 fungi (*c.f.* Votintseva and Filatov 2011). Second, we found that introgression has been
417 a common phenomenon in the history of *N. tetrasperma* and played an important role
418 in the evolution of the *mat* chromosomes of this species. Large tracts have been
419 introgressed from multiple heterothallic species into both *mat A* and *mat a*
420 chromosomes of *N. tetrasperma*, and thus, introgression is not biased towards *mat a*
421 chromosomes as indicated by the smaller scale study by Sun et al. (2012). Structural
422 heterozygosity due to rearrangements of the *mat A* chromosome of the *N. tetrasperma*

Downloaded from genome.cshlp.org on November 7, 2026 . Published by Cold Spring Harbor Laboratory Press

reference strain (FGSC 2508) has been reported previously (Jacobson 2005; Ellison et al. 2011b). If rearrangements are the sole cause of recombination suppression in this system, and the changes found on the *mat A* chromosome in FGSC 2508 are ancestral in *N. tetrasperma*, we would expect long introgression tracts to be confined to the *mat a* chromosomes throughout the species (*c.f.*, Sun et al. 2012). However, the data presented herein does not support this hypothesis. In the absence of chromosome level *de novo* assemblies of the *N. tetrasperma* genomes analyzed in this study, we are not able to infer the gene order and architecture of the *mat* chromosomes of the strains investigated here, and thus unable to connect structural heterozygosity with recombination suppression and introgression in *N. tetrasperma*.

Two possible alternative explanations to introgression for the patterns of low divergence to heterothallic species are balancing selection and ancestral variance in coalescent times. Mating-type loci in fungi have been shown to evolve under balancing selection and alleles within these regions can be more similar to the same allele from a different species than a different allele of the same species (*e.g.*, May et al. 1999). However, since recombination is not suppressed around the *mat* locus in heterothallic *Neurospora*, this pattern should be confined to the *mat* locus and not extend to encompass most of the flanking regions of the *mat* chromosome in heterothallic species and is unlikely to produce the low levels of sequence divergence we observe. We can also discount ancestral variance in coalescent times as an explanation for the low divergence we observe in L10A, L4 a and L7a to *N. hispaniola*, and L9a to *N. perkinsii*.

The observation of large introgressed blocks from multiple heterothallic species, fixed in *N. tetrasperma* lineages, supports the view that adaptive introgression has contributed to the evolutionary history of *N. tetrasperma.* A history

Downloaded from genome.cshlp.org on November 7, 2026 . Published by Cold Spring Harbor Laboratory Press

of recombination on the *mat* chromosomes of heterothallic 'donor' species, has acted to purge deleterious mutations by an effective purifying selection. Therefore, we hypothesize that fixation of an introgressed haplotype by positive selection occurred because it contained fewer linked deleterious mutations than the other *N. tetrasperma* haplotypes already present in the recipient lineage.

It is noteworthy that we did not detect the signature of recent large scale selective sweeps on the introgressed SR regions of L5 A and L10 A; thus, our data suggest that the alternative model of background selection may be a better model for explaining the patterns of molecular evolution in *N. tetrasperma* SR regions for these lineages. This finding, however, does not necessarily argue against the adaptive introgression model, as the action of background selection occurring after the fixation of the introgressed region could have removed any signatures of older sweep events. The fact that the SR regions of *N. tetrasperma* are relatively young (~1 million years old) compared to sex chromosomes or other neo-sex chromosomes (Zhou and Bachtrog 2012; Charlesworth 2013; Cortez et al. 2014), and that the majority of the genes in this SR region are expressed (Samils et al. 2013), suggests that purifying selection may be a potent force in these regions. Also, the fact that *N. tetrasperma* is viable in a haploid homokaryotic state suggests that deleterious mutations in the SR regions, that may otherwise be masked from selection in a heterokaryon, are subject to periods of haploid selection. Taken together, the view that emerges of the history of linked selection in the *mat* SR regions of *N. tetrasperma* is one of a genomic region that is exposed to negative background selection, and shaped by occasional bouts of positive selection of introgressed tracts.

Recombination suppression on the chromosome carrying the mating-type locus has evolved independently in several fungal species distantly related to *N.*

Downloaded from genome.cshlp.org on November 7, 2026 . Published by Cold Spring Harbor Laboratory Press

tetrasperma, including *Podospora, Cryptococcus* and *Microbotryum* (Fraser and Heitman 2004; Grognet et al. 2014; Fontanillas et al. 2015), and contributes to our understanding of different aspects of eukaryote genome evolution. Signatures of molecular degeneration associated with recombination suppression have been reported in *Neurospora* and *Microbotryum* (Hood et al. 2004; Fontanillas et al. 2015). There are massive changes in architecture on the *Microbotryum* mating-type chromosome (Badouin et al. 2015), a pattern also reported from one of the *N. tetrasperma* strains (Ellison et al. 2011b), and further studies using high quality genomic data have the potential to reveal the commonality of such rearrangements and their link to recombination suppression of fungal *mat* chromosomes. In summary, this study shows the value of fungal *mat* chromosomes to provide insights into factors driving eukaryote genome evolution, by providing a new model for how to evade the inevitable cost of recombination suppression on chromosomes determining sexual identity.

Methods

Genome sequencing and assembly of *Neurospora* strains used in the study

We selected 94 strains of *Neurospora* for genome sequencing: 92 strains of *N. tetrasperma* and one each of the two species *N. sitophila* and *N. hispaniola* (Supplementary Table 1). These taxa differ in mating system; while *N. tetrasperma* is pseudohomothallic, *N. sitophila* and *N. hispaniola* are heterothallic (see Supplementary Fig. 1 for details). All strains were obtained from, or have been deposited to, the Fungal Genetics Stock Center (FGSC), University of Missouri, Kansas City, USA. The majority of the sequenced strains were single mating-type

Downloaded from genome.cshlp.org on November 7, 2026 . Published by Cold Spring Harbor Laboratory Press

homokaryons (*mat A* and *mat a*) originating from mating-type heterokaryons sampled in nature (Supplementary Fig. 1; Supplementary Table 1). We sequenced strains identified previously as homokaryotic for mating type by laboratory crosses and PCR screens (Corcoran et al. 2012). We then verified mating type and homokaryosis as outlined in Supplementary methods, and used only verified strains in further analyses.

Genomic DNA was extracted from mycelial tissue using the Easy-DNA Kit (Invitrogen). Paired end library preparation and whole genome sequencing was carried out at BGI Hong Kong using Illumina HiSeq 2000, which generated an average of 1.5 Gbp of paired 90bp reads per strain (an average of 1.4 Gbp of sequenced reads after quality filtering) and an average coverage per strain of between 25 and 45x (Supplementary Table 1). The quality filtering of the FASTQ files was performed by BGI Hong Kong as described in Sun et al. (2015). We performed both reference and *de novo* genome assemblies of the reads. Detailed procedures for read mapping to two available reference genomes of *N. tetrasperma* (2508 and 2509; Ellison et al. 2011b), variant calling and *de novo* assembly is outlined in Supplementary methods.

Genome alignments of outgroup heterothallic species

Whole genome alignments of the *N. tetrasperma*, *N. crassa* and *N. discreta* reference genomes were performed using Mauve v2.3.1 (Darling et al. 2010) (see Supplementary methods).

Phylogenomic analyses

We used two approaches to resolve the phylogenetic relationships of the *Neurospora* species and lineages included in this study. In the first analysis, we carried out

Downloaded from genome.cshlp.org on November 7, 2026 . Published by Cold Spring Harbor Laboratory Press

Maximum Likelihood phylogenomic analysis using an alignment of only variable sites from all autosomes using RAxML v7.3.1 (Stamatakis 2006). The second analysis used the STAR method (Liu et al. 2009) for estimating species trees from a collection of rooted gene trees from genes on the autosomes. Gene trees from 5,723 genes were used as input for STAR (see Supplementary Methods for details)

Population structure analyses

The Bayesian clustering program InStruct (Gao et al. 2007) was used to analyze the population structure within the *N. tetrasperma* clade. We used 1,500 randomly chosen SNPs from each of the autosomes for which we had complete genotype information (*i.e.*, no missing data) across all strains, resulting in a total set of 9000 SNPs. InStruct was run with K=1 to 12 and 5 replicate runs for each K, where K is the number of ancestral clusters to which an individual can be assigned (See Supplementary methods for details) .

A principal component analysis (PCA) was performed to further investigate the genetic relationships among our global sample of *N. tetrasperma* strains. We performed the PCA using Adegenet (Jombart and Ahmed 2011) on 509,119 bi-allelic SNPs from autosomes where we had data for every strain.

Determining the region of suppressed recombination (SR)

The region of suppressed recombination (SR) on the *mat* chromosome of *N. tetrasperma* was determined by two approaches. The first used comparisons between the level of linkage disequilibrium (LD) on the *mat* chromosome and the autosomes to demarcate the SR region. This method was used for L5, L8 and L10, the lineages for which we had the largest number of strains (>8). LD between SNPs was calculated

Downloaded from genome.cshlp.org on November 7, 2026 . Published by Cold Spring Harbor Laboratory Press

548 using the square of Pearson's correlation coefficient (r^2) as implemented in the python
549 module egglib v2.1.5 (De Mita and Siol, 2012). The r^2 for all chromosomes was
550 calculated by sliding across the chromosomes in 100kb windows with a step size of
551 20kb. The mean r^2 value for each window was estimated by calculating the mean of
552 the r^2 for all pairwise combinations of SNPs within a window. We used the 95%
553 quantile for r^2 across the autosomes as a cut-off to identify the SR region on the *mat*
554 chromosome (i.e. windows with r^2 above this cut-off on the *mat* chromosome were
555 assigned to the SR).
556 The second approach to determine the SR region on the *mat* chromosome was
557 used for lineages where the sample size was small (≤8) and was the same as used by
558 Sun et al., (2012) (see supplementary methods for details).
559
560 **Patterns of genetic variation within *N. tetrasperma* lineages**
561 To quantify the levels of variation within *N. tetrasperma* lineages we calculated the
562 nucleotide diversity (π) (Tajima 1983) across the genome. We also analyzed the
563 frequency spectrum of mutations across the genomes using Tajima's D statistic
564 (Tajima 1989) and Fay and Wu's H statistic (Fay and Wu 2000; Zeng et al. 2006)
565 using egglib v2.1.7 (De Mita and Siol 2012). For the analysis of the autosomes of L5
566 and L8 we randomly chose a single homokaryon from each heterokaryon, as the vast
567 majority of heterokaryons had few heteroallelic differences on the autosomes. L10
568 was excluded from these analyses because high levels of diversity were observed
569 within heterokaryons.
570 The decay of linkage disequilibrium in the genome was examined in lineages
571 L5, L8 and L10. Linkage disequilibrium between pairs of SNPs as measured by r^2 was
572 calculated using egglib. The mean r^2 was calculated for all SNPs within a particular

Downloaded from genome.cshlp.org on November 7, 2026 . Published by Cold Spring Harbor Laboratory Press

physical distance at increasing 1kb increments (*i.e.*, the mean r^2 between pairs of SNPs within 1kb, 2kb, 3kb and so on until 500kb).

Genetic distance between pairs of *N. tetrasperma* homokaryons

We calculated genetic distance as the proportion of nucleotide differences per nucleotide site between the *mat A* and *mat a* genomes of each heterokaryon in sliding windows of 100kb with a 20kb step. The observed values of genetic distance were corrected for multiple hits using the Juke-Cantor correction (Jukes and Cantor 1969). Windows that had less than 10000 called sites (10%) were excluded.

Divergence between *N. tetrasperma* lineages

The level of divergence between lineages was measured using the D_{xy} statistic (Nei and Li 1979), the average number of pairwise differences between populations. The divergence was measured in 25kb non-overlapping windows across the genomes. Sites with missing data were excluded in the calculation of D_{xy}. Windows with less than 2500 (10%) called sites were excluded.

Divergence between *N. tetrasperma* lineages and heterothallic species

Here we calculated the divergence between each strain, or each lineage, of *N. tetrasperma* and the genomes of heterothallic species: *N. crassa*, *N. hispaniola* and *N. sitophila*. The Jukes-Cantor corrected divergence between each *N. tetrasperma* strain and each heterothallic species was calculated across the genome in 25kb non-overlapping windows. Windows with less than 2500 (10%) called sites were excluded.

Downloaded from genome.cshlp.org on November 7, 2026 . Published by Cold Spring Harbor Laboratory Press

598 **Tests for introgression between species using D-statistics**

599 To test for introgression from heterothallic species into the autosomes of the *N.*
600 *tetrasperma* lineages, we calculated the D-statistic (aka the 'ABBA-BABA' test) as
601 outlined by Green at al., (2010) (See Supplementary methods for details). Block-
602 jackknife resampling with a block size of 250kb was used to estimate the standard
603 error of the D-statistic and to calculate a Z-score. The Z-score was then converted to a
604 p-value to determine statistical significance of the D-statistic.

605 To visualize the pattern of ABBA and BABA sites on the *mat* chromosome,
606 we took a sliding window approach that considered only sites fixed between the *mat A*
607 and *mat a* mating-type chromosome from the same lineage (*e.g.*, L10A and L10a)
608 (See Supplementary methods for details).

609

610 **Treemix analysis**

611 We performed an analysis on autosomal SNPs with the program Treemix (Pickrell
612 and Pritchard 2012) to reconstruct the population history of the *N. tetrasperma*
613 lineage incorporating any history of migration between lineages or any migration
614 from *N. sitophila* or *N. hispaniola* using the following parameters *-k 10000 -noss -*
615 *root N. hispaniola*. The program was run with different numbers of migration events
616 (-m0 to -m10).

617

618 **Likelihood ratio test of divergence with gene flow in autosomes**

619 We used a likelihood-based approach to test for gene flow between a heterothallic
620 species and lineages of *N. tetrasperma* using the program 3S (Yang 2010). The model
621 implemented in the 3S program requires at least one genome per species. We chose *N.*
622 *discreta* as the outgroup for all triplets of species. We chose the *N. tetrasperma*

Downloaded from genome.cshlp.org on November 7, 2026 . Published by Cold Spring Harbor Laboratory Press

genome with the highest coverage within a lineages as the representative genome for that lineage. We examined all combinations of *N. tetrasperma* lineages and heterothallic species pairs. We used a likelihood ratio test, as described by Yang (2010), to test whether we could reject the M_0 isolation model in favor of the M_1 model with post divergence gene flow (see Supplementary methods).

Simulations of sequence divergence between species

To determine if the regions of low divergence that we observed in the SR regions of some lineages is due to introgression, or can be better explained by variance in coalescent times due to genetic drift in the ancestral population of a pair species, we performed simulations using the MCcoal program. The ancestral population sizes (θ_4 and θ_5) and speciation times (τ_0 and τ_1), The parameter values used were taken from the results for the M0 model in Supplementary Table S8. The parameters estimated for the different *N. tetrasperma* lineages were similar when the same heterothallic was used. Therefore, we calculated the median from the range of estimates as our input parameter value for MCcoal (*e.g.* For θ_5 in (*N. tetrasperma*, *N. hispaniola), N. discreta*) the values ranges from 0.016997 - 0.019756. We used the median from this range of values as our choice for θ_5). We used the same divergence time to *N. discreta* in all simulations (see Supplementary methods for parameter values used).

We simulated 50000 25kb regions without migration between species and calculated the divergence between Species A and B (See supplementary methods) and compared them to our observed median divergence value within a *mat* SR for a lineage. The p-value was calculated as the proportion of simulated divergence values that were equal to or more extreme (greater or less than, depending on whether the SR region being considered showed higher or lower divergence to Species B than the R

Downloaded from genome.cshlp.org on November 7, 2026 . Published by Cold Spring Harbor Laboratory Press

region of the same lineage) than the observed median divergence of a lineage's SR region to a heterothallic species.

Phylogenetic analysis of the genes on the *mat* chromosome

We performed Maximum likelihood phylogenetic analysis of genes within the SR region and the PAR region of the *mat* chromosome using RAxML v7.3.1. We selected six genes from the SR region and two genes from the PAR common to all lineages of *N. tetrasperma,* and included representatives from all heterothallic species (Supplementary Table 12; see Supplementary methods).

Analyses of molecular evolution in coding regions

We investigated the levels of polymorphism and divergence in coding regions of genes listed in *N. tetrasperma* 2509 annotation file that had an ortholog in *N. crassa* and *N. discreta* (See Supplementary methods for details**)**. For polymorphism analysis of coding regions we retained only the genes that had a minimum length of 70% of the full-length coding sequence, as specified in the *N. tetrasperma* 2509 genome annotation file. We calculated the number of polymorphisms at non-synonymous (P_N) and synonymous sites (P_S) for each gene within lineages L5, L8 and L10. We also calculated π for both synonymous and non-synonymous polymorphisms. All counts of coding sites and calculation of polymorphism statistics were carried out using methods from the bio++ library (Gueguen et al. 2013) through egglib (De Mita and Siol 2012).

We used the codeml program in PAML v4.3 (Yang 2007) to test hypotheses on the ratio of non-synonymous to synonymous substitution rates within the SR region of each *N. tetrasperma* lineage. We carried out the codeml analysis on

Downloaded from genome.cshlp.org on November 7, 2026 . Published by Cold Spring Harbor Laboratory Press

673 individual gene alignments of genes within the SR region common to all lineages of
674 *N. tetrasperma*. We removed codons from gene alignments where there were
675 polymorphisms segregating within a lineage. These were removed because for the
676 codeml analysis we are only interested in investigating the rates of non-synonymous
677 and synonymous substitutions that have occurred in the SR regions of each lineage.
678 Following the filtering of polymorphic sites, we retained only a single sequence from
679 each lineage in our gene alignments. We required that the gene alignment length
680 covers at least 70% of the complete CDS sequence for a gene as defined in the *N.*
681 *tetrasperma* 2509 annotation file. These steps resulted in 1133 genes being used for
682 the codeml analysis. Because estimates of d_N/d_S are unreliable when branch lengths
683 are short, we excluded L4, L5a and L6 from the analyses: these lineages showed very
684 low synonymous sequence divergence with other lineages, and were analyzed for
685 asymmetrical substitution rates in previous studies (Ellison et al. 2011b; Whittle and
686 Johannesson 2011; Whittle et al. 2011a; Sun et al. 2012). The free ratio model was
687 implemented in codeml to estimate d_N/d_S for each branch in the phylogeny.
688
689 **Codon usage analysis**
690 We further investigated the signal of molecular degeneration in the SR regions of
691 each lineage of *N. tetrasperma* by examining patterns of synonymous substitutions.
692 For this analysis we also used the polymorphism filtered gene set (described above)
693 for each lineage. We used a perl script to run through all codon positions for each
694 three-way alignment, including the coding sequence of *N. discreta*, *N. tetrasperma*
695 *mat A* and *mat a*. We took an optimal (OP) and non-optimal (NOP) codon list from
696 previous published studies of codon usage study in *N. tetrasperma* and *N. discreta*
697 (Whittle et al. 2011a; Whittle et al. 2011b). We assumed the codons from *N. discreta*

Downloaded from genome.cshlp.org on November 7, 2026 . Published by Cold Spring Harbor Laboratory Press

represent the ancestral state, thus synonymous codon changes in one of the two *N. tetrasperma* genomes represent allele specific derived changes. We further divided the allele specific codon changes into four categories for *N. tetrasperma mat A* and *mat a*: OP to OP, OP to NOP, NOP to OP and NOP to NOP. The changes of OP to OP and NOP to NOP were used as neutral changes.

Fit of selective sweep model to SR regions

We tested if a recent hard selective sweep could explain the patterns of reduced diversity of the *mat* SR regions of *N. tetrasperma* by fitting a hard selective sweep model to neutral variation from SR regions of L5, L8 and L10. We followed the approach outlined in Jensen et al., (2002) and used in a number of studies (e.g, Bachtrog, 2004; Betancourt et al., 2009; Campos et al., 2014) to model a hard selective sweep in a genomic region without recombination. This approach to modeling a selective sweep is accomplished by performing neutral coalescent simulations without recombination that proceed backward in time until a time T_s, when all the lineages in the sample are forced to coalesce to a single node, which represents the effect of a selective sweep. Two parameters were varied in the model, 1) the amount of neutral variation (θ_0) that would be present in the absence of a sweep and 2) the time (T_s) since the sweep took place. We performed 25,000 replicate simulations for each parameter pair combination. We considered θ_0 and T_s pairs compatible with the data if the simulated number of segregating sites (S) was within ± (0.05 x S) of the observed S and the simulated average number of pairwise differences (k) was within ± (0.05 x k) of the observed k. The likelihood of the data was calculated for each parameter pair as described by Jensen et al. (2002). For our observed measures of neutral variation within SR regions of L5, L8 and L10 we used

Downloaded from genome.cshlp.org on November 7, 2026 . Published by Cold Spring Harbor Laboratory Press

723 only four fold degenerate sites and calculated *S* and K the average number of pairwise
724 differences between alleles using polydNdS (Thornton 2003).
725
726 **Phylogenetic analysis of heterokaryon incompatibility (*het*) genes**
727 We extracted the *het-6* and *het-c* gene sequences from the *de novo* assemblies of the
728 *N. tetrasperma*, *N. sitophila* and *N. hispaniola* genomes (see Supplementary
729 methods). We aligned the *het-c* sequences with the *het-c* sequences from Hall et al.,
730 (2010), and *het-6* sequenced from Powell et al., (2007). We constructed phylogenetic
731 trees from the resulting *het-c* and *het-6* alignments using PhyML v3.0.
732

Downloaded from genome.cshlp.org on November 7, 2026 . Published by Cold Spring Harbor Laboratory Press

Data access

FASTQ files generated in this study have been submitted to the SRA: SRA accession number SRP040006. The VCF files, consensus sequences for the *N. tetrasperma* strains, *N. hispaniola* 8817 and *N. sitophila* 1135, whole genome alignments and all gene alignments used for the phylogenetic analysis are available from the Dryad Digital Repository: https://datadryad.org/resource/doi:10.5061/dryad.162mh

Acknowledgements

We are grateful for financial support provided by the Swedish Research Council, Helge Ax:son Johnsons stiftelse, Uppsala Centre for Evolution and Genomics and Nilsson-Ehle-Donationerna from Kungl. Fysiografiska Sällskapet i Lund. Computations were performed on resources provided by the Swedish National Infrastructure for Computing (SNIC) through Uppsala Multidisciplinary Center for Advanced Computational Science (UPPMAX).

Disclosure declaration

All author's declare no conflict of interest in the submission of this manuscript.

Figure legends

Figure 1. The global pattern of variation in *N. tetrasperma.* **(A)** The phylogenetic relationships of all *N. tetrasperma* strains used in this study, inferred from 2,259,433 variable sites on the autosomes. A subtree excluding *N. discreta*, *N. crassa* and *N. hispaniola* is shown. Numbers on the branches indicate the bootstrap support for that relationship expressed as a proportion. **(B)** Principal component analysis (PCA) of genetic variation (509,199 biallelic autosomal SNPs) across the global sample of *N.*

Downloaded from genome.cshlp.org on November 7, 2026 . Published by Cold Spring Harbor Laboratory Press

759 *tetrasperma* strains. The first two principal components are shown. (**C**) Population
760 structure of *N. tetrasperma* inferred from 9000 SNPs (1,500 from each of the six
761 autosomes) using InStruct at K = 6. Lineages color coded in A, B, and C according to
762 the legend in B.
763
764 **Figure 2.** Pair-wise divergences between the *mat A* and *mat a* homokaryons sampled
765 from the same heterokaryon for the *mat* chromosomes (linkage group I) and six
766 autosomes (II-VII) of representatives from all *N. tetrasperma* lineages. Each linkage
767 group is shown in a separate row (labeled on the right). The pair-wise divergences
768 were calculated as the fraction of differences (in bp) between the sequences, using a
769 100kb sliding window (step size 20kb).
770
771 **Figure 3** Patterns of genetic variation across the *mat* chromosome for *N. tetrasperma*
772 lineages L5, L8 and L10. **(A)** Linkage disequilibrium given as the mean Pearson's
773 correlation coefficient (r^2). The vertical black line shows the position of the *mat a*
774 locus in the 2509 reference genome. (**B)** Nucleotide diversity (π), Tajima's *D* and Fay
775 and Wu's *H*. For all variables, we used a100kb window size (step size 20kb). The
776 values for each window are represented by the grey points and smooth lines were
777 plotted with stat_smooth in the ggplot2 R package using the gam method with a span
778 of 0.2. Dashed vertical lines indicate lineage specific limits of the SR-region.
779
780 **Figure 4.** Introgression of large chromosomal regions from heterothallic *Neurospora*
781 to *N. tetrasperma*. **(A)** Diagrams showing the genealogy of the *mat* chromosome SR
782 region in a *N. tetrasperma* lineage when suppression of recombination had begun at
783 SR Time. The shaded regions show that a lineage is in SR. The models from left to

Downloaded from genome.cshlp.org on November 7, 2026 . Published by Cold Spring Harbor Laboratory Press

right show expected relationships in the absence of introgression, SR A introgression and SR a introgression. **(B)** Boxplots of mean divergence between L10, L4 and L7 in the SR region. Asterisks above the horizontal black lines are the p-values for the Mann-Whitney *U* test between SR regions within a lineage (***:$p < 0.001$, n.s :non-significant. **(C)** Pair-wise divergences between the *mat* chromosomes of *N. tetrasperma* strains and the heterothallic species *N. crassa, N. hispaniola* and *N. sitophila*. Each row in the figure shows the sequence divergence between a strain of *N. tetrasperma* and the heterothallic species indicated in the heading of the column, using a non-overlapping sliding window of 25kb. *N. tetrasperma* strains are sorted by lineage and according to mating-type. Regions lacking sufficient number of sites (2500 sites) are colored grey. A maximum divergence of up to 0.06 is plotted, windows exceeding this are colored grey.

Figure 5. The phylogenetic relationships of *Neurospora* strains in the PAR and SR regions common to all *N. tetrasperma* lineages. The SR tree was reconstructed from a concatenated dataset of six genes and the PAR tree from two genes (see Supplementary Table 12). The heterothallic species are colored in blue and *N. tetrasperma* lineages in red. Numbers on the branches are bootstrap support values expressed as a percentage; values below 70 are not shown.

Downloaded from genome.cshlp.org on November 7, 2026 . Published by Cold Spring Harbor Laboratory Press

Table 1. The sizes and sequence divergence of the regions of suppressed recombination (SR) in each lineage of *N. tetrasperma*. Percent of chromosome covered by the SR-region is given, and Start (Left) and Start (Right) give the location (in Mbp) of SR boundaries. Sequence divergence was measured by average pairwise differences (D_{xy}) between *mat A* and *mat a* chromosomes within the SR region of each lineage. Coordinates given are those of the *N. tetrasperma* 2509 *a* reference genome.

Lineage	Size (Mbp)	%	Start (Right)	Start (Left)	D_{xy}
L1 (LA)	7.14	75	1.15	8.29	0.014
L4 (MX)	8.10	86	0.90	9.00	0.017
L5 (NZ)	7.24	76	0.74	7.98	0.025
L6 (HI)	7.92	84	0.68	8.60	0.032
L7 (LA)	7.00	74	1.10	8.10	0.014
L8 (LA)	6.0	63	1.54	7.54	0.014
L9 (LB)	7.90	83	1.00	8.90	0.027
L10 (UK)	6.2	66	1.38	7.58	0.02

Downloaded from genome.cshlp.org on November 7, 2026 . Published by Cold Spring Harbor Laboratory Press

821 **Table 2.** Summary of the *mat* chromosome introgressions and the asymmetry of
822 molecular degeneration.

Lineage	Introgressed *mat* chromosome	Source of Introgression[1]	Molecular degeneration[2]		
			P_N/P_S	**d_N/d_S**	**Codon Usage**
L1	none	n/a	n/a	=	=
L4	*mat a*	*N. hisp*	n/a	*mat A* > ***mat a***[4]	*mat A* > ***mat a***
L5[3]	*mat A*	?	*mat a* > ***mat A***	*mat a* > ***mat A***	=
L6	*mat A*	?	n/a	*mat a* > ***mat A***[5]	*mat a* > ***mat A***[6]
L7	*mat a*	*N. hisp*	n/a	=	=
L8[3]	none	n/a	=	=	=
L9	*mat a*	*N. crassa*	n/a	*mat A* > ***mat a***	*mat A* > ***mat a***
L10[3]	*mat A*	*N. hisp*	*mat a* > ***mat A***	*mat a* > ***mat A***	=

823

824 [1] The species of *Neurospora* that show(s) the highest genetic similarity to the
825 introgressed region (*N. hisp* : *N. hispaniola*; ? : Uncertain; n/a is not applicable.
826 [2] Lack of a difference in degeneration between *mat A* and *mat a* is indicated by '='.
827 The introgressed *mat* chromosome for each lineage indicated in bold. Higher P_N/P_S
828 and d_N/d_S ratios are interpreted as increased degeneration. For optimal codon usage,
829 increased degeneration indicates significantly more synonymous changes from
830 optimal to non-optimal codons (*OP → NOP)* or less changes from non-optimal to
831 optimal codons (*NOP → OP)* as compared to the neutral standard.

Downloaded from genome.cshlp.org on November 7, 2026 . Published by Cold Spring Harbor Laboratory Press

832 [3] Population level analyses (P_N/P_S) was only carried out for these lineages where we

833 had more than 8 strains.

834 [4]Reported by Sun et al. (2012).

835 [5] Reported by Ellison et al. (2011b).

836 [6] Reported by Ellison et al. (2011b) and also in this study.

837

838

Downloaded from genome.cshlp.org on November 7, 2026 . Published by Cold Spring Harbor Laboratory Press

References

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

Bachtrog D, Charlesworth B. 2002. Reduced adaptation of a non-recombining neo-Y chromosome. *Nature* **416**: 323-326.

Badouin H, Hood ME, Gouzy J, Aguileta G, Siguenza S, Perlin MH, Cuomo CA, Fairhead C, Branca A, Giraud T. 2015. Chaos of Rearrangements in the Mating-Type Chromosomes of the Anther-Smut Fungus Microbotryum lychnidis-dioicae. *Genetics* **200**: 1275-1275.

Betancourt AJ, Welch JJ, Charlesworth B. 2009. Reduced Effectiveness of Selection Caused by a Lack of Recombination. *Current Biology* **19**: 655-660.

Blaser O, Grossen C, Neuenschwander S, Perrin N. 2013. Sex-chromosome turnovers induced by deleterious mutation load. *Evolution* **67**: 635-645.

Bull JJ. 1983. *Evolution of sex determination mechanisms*. Menlo Park: Benjamin Cummings.

Charlesworth B, Charlesworth D. 2000. The degeneration of Y chromosomes. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences* **355**: 1563-1572.

Charlesworth B, Charlesworth D. 2010. *Elements of Evolutionary Genetics*. Roberts and Company Publishers.

Charlesworth D. 2013. Plant sex chromosome evolution. *J Exp Bot* **64**: 405-420.

Corcoran P, Dettman JR, Sun Y, Luque EM, Corrochano LM, Taylor JW, Lascoux M, Johannesson H. 2014. A global multilocus analysis of the model fungus Neurospora reveals a single recent origin of a novel genetic system. *Mol Phylogenet Evol* **78**: 136-147.

Corcoran P, Jacobson DJ, Bidartondo MI, Hickey PC, Kerekes JF, Taylor JW, Johannesson H. 2012. Quantifying functional heterothallism in the pseudohomothallic ascomycete Neurospora tetrasperma. *Fungal biology* **116**: 962-975.

Cortez D, Marin R, Toledo-Flores D, Froidevaux L, Liechti A, Waters PD, Grutzner F, Kaessmann H. 2014. Origins and functional evolution of Y chromosomes across mammals. *Nature* **508**: 488-493.

Darling AE, Mau B, Perna NT. 2010. progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. *PLoS ONE* **5**: e11147.

De Mita S, Siol M. 2012. EggLib: processing, analysis and simulation tools for population genetics and genomics. *BMC Genet* **13**.

Dettman JR, Jacobson DJ, Taylor JW. 2003. A multilocus genealogical approach to phylogenetic species recognition in the model eukaryote *Neurospora*. *Evolution* **57**: 2703-2720.

Ellison CE, Hall C, Kowbel D, Welch J, Brem RB, Glass NL, Taylor JW. 2011a. Population genomics and local adaptation in wild isolates of a model microbial eukaryote. *Proc Natl Acad Sci U S A* **108**: 2831-2836.

Ellison CE, Stajich JE, Jacobson DJ, Natvig DO, Lapidus A, Foster B, Aerts A, Riley R, Lindquist EA, Grigoriev IV et al. 2011b. Massive changes in genome architecture accompany the transition to self-fertility in the filamentous fungus Neurospora tetrasperma. *Genetics* **189**: 55-69.

Fay JC, Wu CI. 2000. Hitchhiking under positive Darwinian selection. *Genetics* **155**: 1405-1413.

Downloaded from genome.cshlp.org on November 7, 2026 . Published by Cold Spring Harbor Laboratory Press

887 Fontanillas E, Hood ME, Badouin H, Petit E, Barbe V, Gouzy J, de Vienne DM,
888 Aguileta G, Poulain J, Wincker P et al. 2015. Degeneration of the
889 Nonrecombining Regions in the Mating-Type Chromosomes of the Anther-
890 Smut Fungi. *Mol Biol Evol* **32**: 928-943.
891 Fraser JA, Heitman J. 2004. Evolution of fungal sex chromosomes. *Mol Microbiol* **51**:
892 299-306.
893 Gallegos A, Jacobson DJ, Raju NB, Skupski MP, Natvig DO. 2000. Suppressed
894 recombination and a pairing anomaly on the mating-type chromosome of
895 Neurospora tetrasperma. *Genetics* **154**: 623-633.
896 Gao H, Williamson S, Bustamante CD. 2007. A Markov chain Monte Carlo approach
897 for joint inference of population structure and inbreeding rates from multilocus
898 genotype data. *Genetics* **176**: 1635-1651.
899 Glass NL, Grotelueschen J, Metzenberg RL. 1990. *Neurospora crassa* a mating-type
900 region. *Proc Natl Acad Sci U S A* **87**: 4912-4916.
901 Green RE, Krause J, Briggs AW, Maricic T, Stenzel U, Kircher M, Patterson N, Li H,
902 Zhai W, Fritz MH et al. 2010. A draft sequence of the Neandertal genome.
903 *Science* **328**: 710-722.
904 Grognet P, Bidard F, Kuchly C, Tong LC, Coppin E, Benkhali JA, Couloux A,
905 Wincker P, Debuchy R, Silar P. 2014. Maintaining two mating types: structure
906 of the mating type locus and its role in heterokaryosis in Podospora anserina.
907 *Genetics* **197**: 421-432.
908 Gueguen L, Gaillard S, Boussau B, Gouy M, Groussin M, Rochette NC, Bigot T,
909 Fournier D, Pouyet F, Cahais V et al. 2013. Bio plus plus : Efficient
910 Extensible Libraries and Tools for Computational Molecular Evolution. *Mol*
911 *Biol Evol* **30**: 1745-1750.
912 Hall C, Welch J, Kowbel DJ, Glass NL. 2010. Evolution and Diversity of a Fungal
913 Self/Nonself Recognition Locus. *PLoS ONE* **5**.
914 Hellborg L, Ellegren H. 2004. Low levels of nucleotide diversity in mammalian Y
915 chromosomes. *Mol Biol Evol* **21**: 158-163.
916 Hood ME, Antonovics J. 2000. Intratetrad mating, heterozygosity, and the
917 maintenance of deleterious alleles in Microbotryum violaceum (=Ustilago
918 violacea). *Heredity (Edinb)* **85 Pt 3**: 231-241.
919 Hood ME, Antonovics J, Koskella B. 2004. Shared forces of sex chromosome
920 evolution in haploid-mating and diploid-mating organisms: Microbotryum
921 violaceum and other model organisms. *Genetics* **168**: 141-146.
922 Howe HB, Jr., Haysman P. 1966. Linkage group establishment in Neurospora
923 tetrasperma by interspecific hybridization with N. crassa. *Genetics* **54**: 293-
924 302.
925 Jacobson DJ. 2005. Blocked recombination along the mating-type chromosomes of
926 *Neurospora tetrasperma* involves both structural heterozygosity and
927 autosomal genes. *Genetics* **171**: 839-843.
928 Jombart T, Ahmed I. 2011. adegenet 1.3-1: new tools for the analysis of genome-wide
929 SNP data. *Bioinformatics* **27**: 3070-3071.
930 Jukes T, Cantor C. 1969. In *Mammalian protein metabolism*, pp. 21-132. Academic
931 Press, New York.
932 Kimura M. 1983. *The neutral theory of molecular evolution*. Cambridge University
933 Press, Cambridge.
934 Laporte V, Filatov DA, Kamau E, Charlesworth D. 2005. Indirect evidence from
935 DNA sequence diversity for genetic degeneration of the Y-chromosome in

Downloaded from genome.cshlp.org on November 7, 2026 . Published by Cold Spring Harbor Laboratory Press

dioecious species of the plant Silene: the SlY4/SlX4 and DD44-X/DD44-Y gene pairs. *J Evol Biol* **18**: 337-347.

Liu L, Yu L, Pearl DK, Edwards SV. 2009. Estimating species phylogenies using coalescence times among sequences. *Systematic biology* **58**: 468-477.

Liu ZY, Moore PH, Ma H, Ackerman CM, Ragiba M, Yu QY, Pearl HM, Kim MS, Charlton JW, Stiles JI et al. 2004. A primitive Y chromosome in papaya marks incipient sex chromosome evolution. *Nature* **427**: 348-352.

Malcom JW, Kudra RS, Malone JH. 2014. The sex chromosomes of frogs: variability and tolerance offer clues to genome evolution and function. *J Genomics* **2**: 68-76.

Marais GA, Nicolas M, Bergero R, Chambrier P, Kejnovsky E, Moneger F, Hobza R, Widmer A, Charlesworth D. 2008. Evidence for degeneration of the Y chromosome in the dioecious plant Silene latifolia. *Curr Biol* **18**: 545-549.

May G, Shaw F, Badrane H, Vekemans X. 1999. The signature of balancing selection: fungal mating compatibility gene evolution. *Proc Natl Acad Sci U S A* **96**: 9172-9177.

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

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

Merino ST, Nelson MA, Jacobson DJ, Natvig DO. 1996. Pseudohomothallism and evolution of the mating-type chromosome in Neurospora tetrasperma. *Genetics* **143**: 789-799.

Nei M, Li WH. 1979. Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proc Natl Acad Sci U S A* **76**: 5269-5273.

Pickrell JK, Pritchard JK. 2012. Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS genetics* **8**: e1002967.

Powell AJ, Jacobson DJ, Natvig DO. 2001. Allelic diversity at the het-c locus in Neurospora tetrasperma confirms outcrossing in nature and reveals an evolutionary dilemma for pseudohomothallic ascomycetes. *Journal of Molecular Evolution* **52**: 94-102.

Powell AJ, Jacobson DJ, Natvig DO. 2007. Ancestral polymorphism and linkage disequilibrium at the het-6 region in pseudohomothallic Neurospora tetrasperma. *Fungal Genet Biol* **44**: 896-904.

Raju NB, Perkins DD. 1994. Diverse Programs of Ascus Development in Pseudohomothallic Species of Neurospora, Gelasinospora, and Podospora. *Developmental Genetics* **15**: 104-118.

Rannala B, Yang ZH. 2003. Bayes estimation of species divergence times and ancestral population sizes using DNA sequences from multiple loci. *Genetics* **164**: 1645-1656.

Saenz GS, Jacobson DJ, Dvorachek WH, Natvig DO. 2003. Sympatric biological and phylogenetic species among pseudohomothallic isolates identified as *Neurospora tetrasperma*. *Fungal Genetics Newsletter* **50 (Suppl):abstract #450**: 144.

Samils N, Gioti A, Karlsson M, Sun Y, Kasuga T, Bastiaans E, Wang Z, Li N, Townsend JP, Johannesson H. 2013. Sex-linked transcriptional divergence in

Downloaded from genome.cshlp.org on November 7, 2026 . Published by Cold Spring Harbor Laboratory Press

the hermaphrodite fungus Neurospora tetrasperma. *Proceedings* **280**: 20130862.

Staben C, Yanofsky C. 1990. *Neurospora crassa* A mating-type region. *Proc Natl Acad Sci U S A* **87**: 4917-4921.

Stamatakis A. 2006. RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22**: 2688-2690.

Sun Y, Corcoran P, Menkis A, Whittle CA, Andersson SG, Johannesson H. 2012. Large-scale introgression shapes the evolution of the mating-type chromosomes of the filamentous ascomycete Neurospora tetrasperma. *PLoS genetics* **8**: e1002820.

Sun Y, Whittle CA, Corcoran P, Johannesson H. 2015. Intron evolution in Neurospora: the role of mutational bias and selection. *Genome Res* **25**: 100-110.

Tajima F. 1983. Evolutionary Relationship of DNA-Sequences in Finite Populations. *Genetics* **105**: 437-460.

Tajima F. 1989. Statistical methods for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123**: 585-595.

Thornton K. 2003. libsequence: a C++ class library for evolutionary genetic analysis. *Bioinformatics* **19**: 2325-2327.

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

Whittle CA, Johannesson H. 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 (Edinb)* **107**: 305-314.

Whittle CA, Sun Y, Johannesson H. 2011a. Degeneration in codon usage within the region of suppressed recombination in the mating-type chromosomes of Neurospora tetrasperma. *Eukaryot Cell* **10**: 594-603.

Whittle CA, Sun Y, Johannesson H. 2011b. Evolution of synonymous codon usage in Neurospora tetrasperma and Neurospora discreta. *Genome biology and evolution* **3**: 332-343.

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

Yang Z. 2010. A likelihood ratio test of speciation with gene flow using genomic sequence data. . *Genome biology and evolution* **2**: 200-211.

Zeng K, Fu YX, Shi S, Wu CI. 2006. Statistical tests for detecting positive selection by utilizing high-frequency variants. *Genetics* **174**: 1431-1439.

Zhou Q, Bachtrog D. 2012. Sex-specific adaptation drives early sex chromosome evolution in Drosophila. *Science* **337**: 341-345.

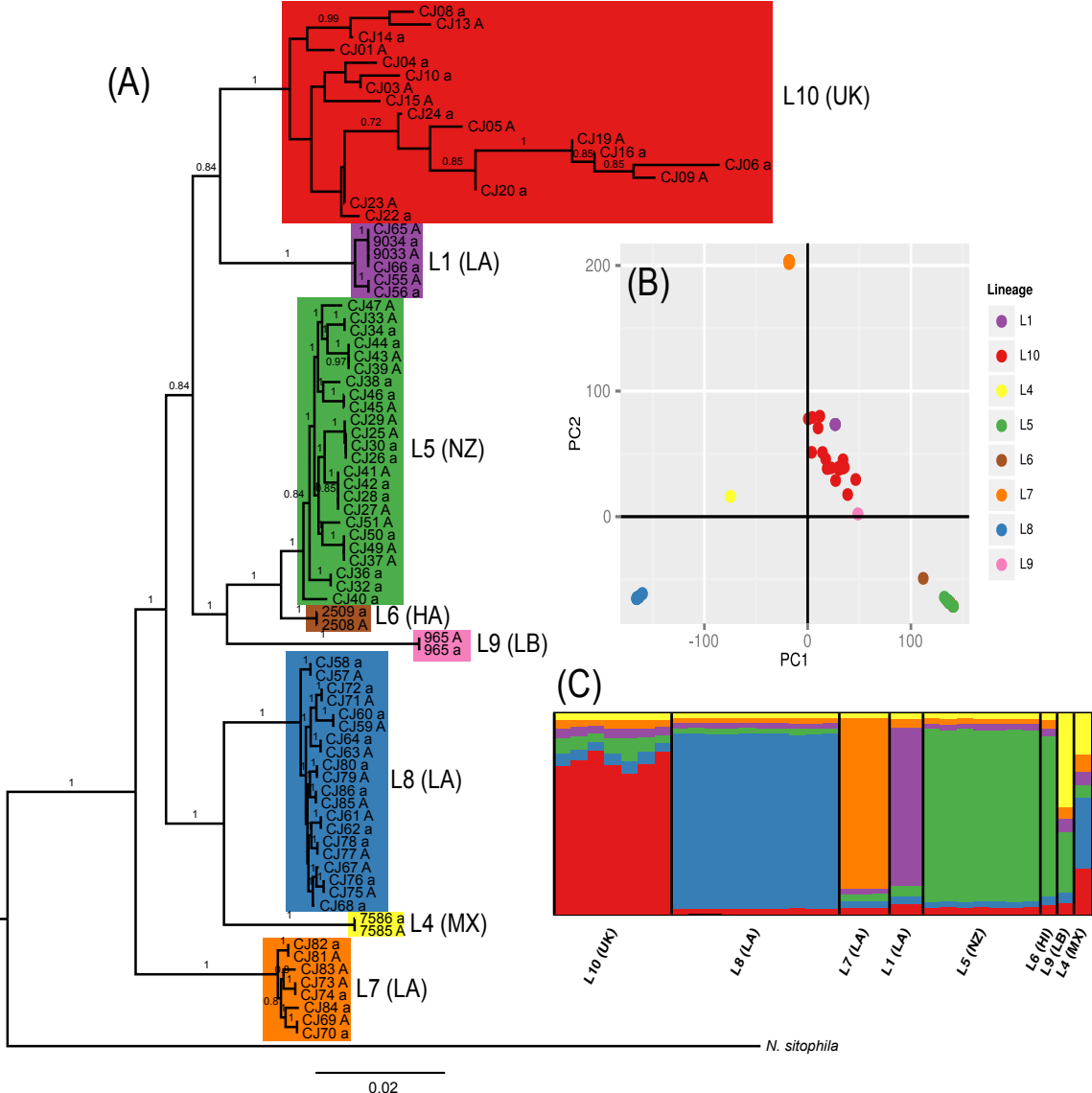
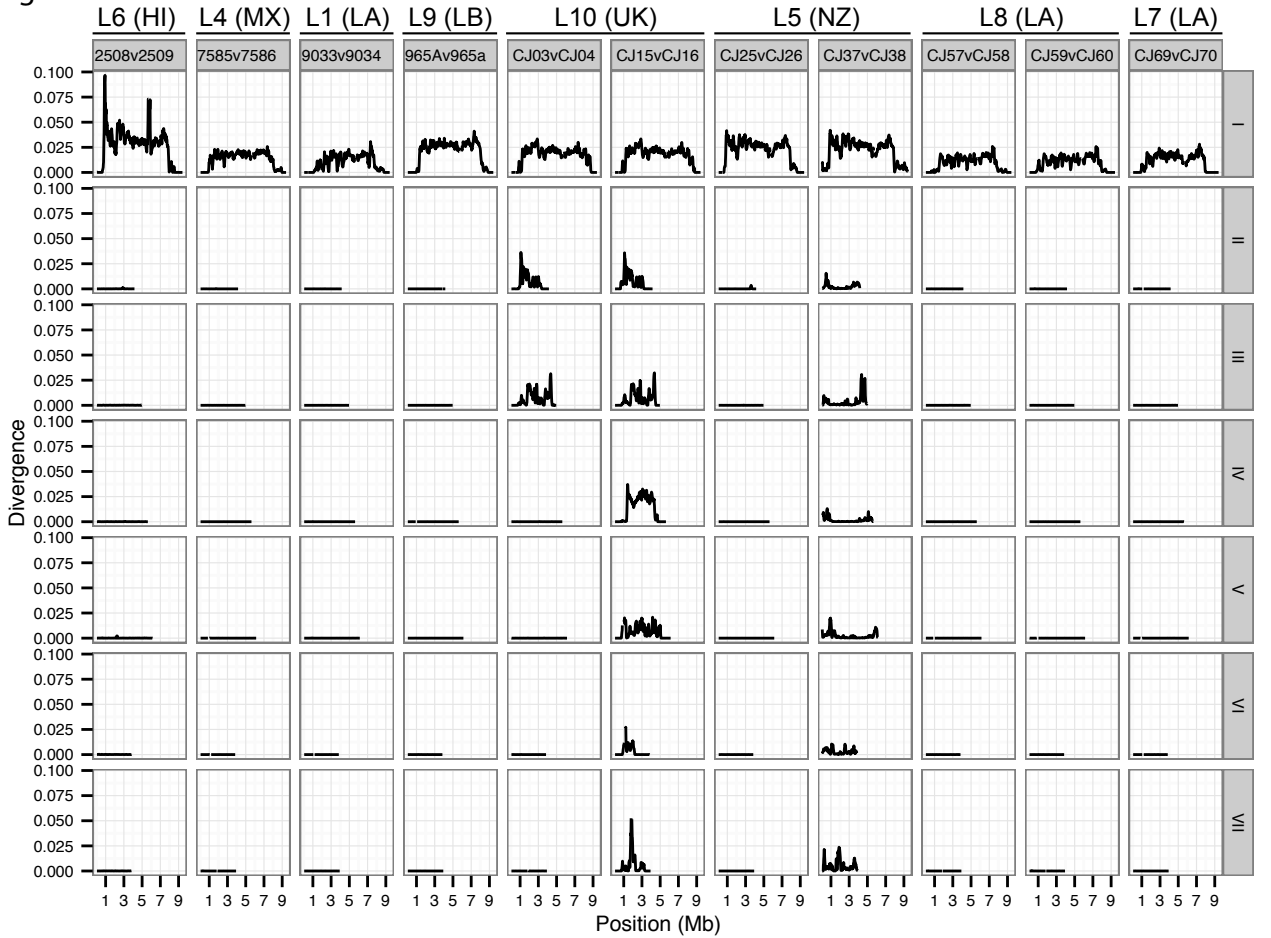
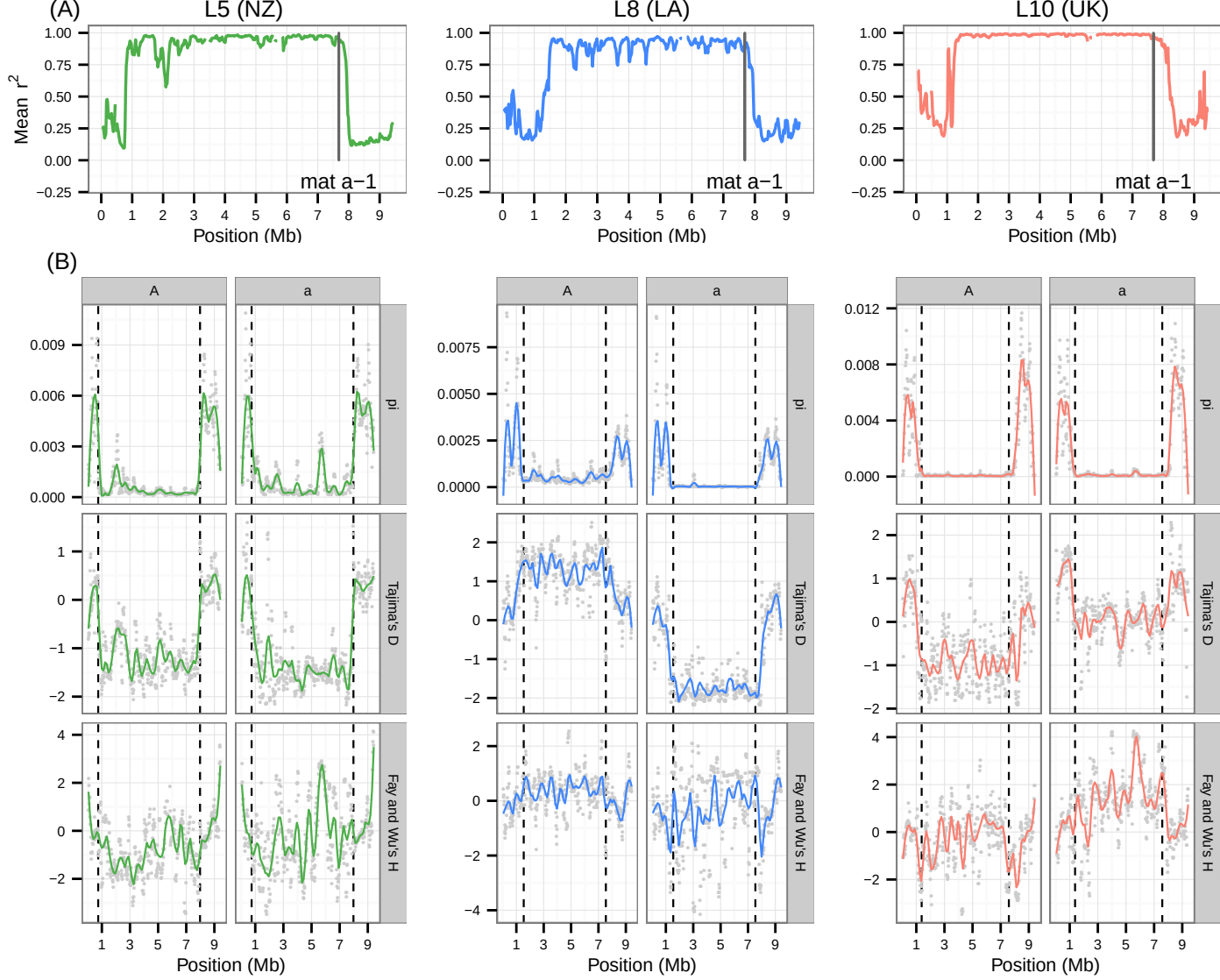
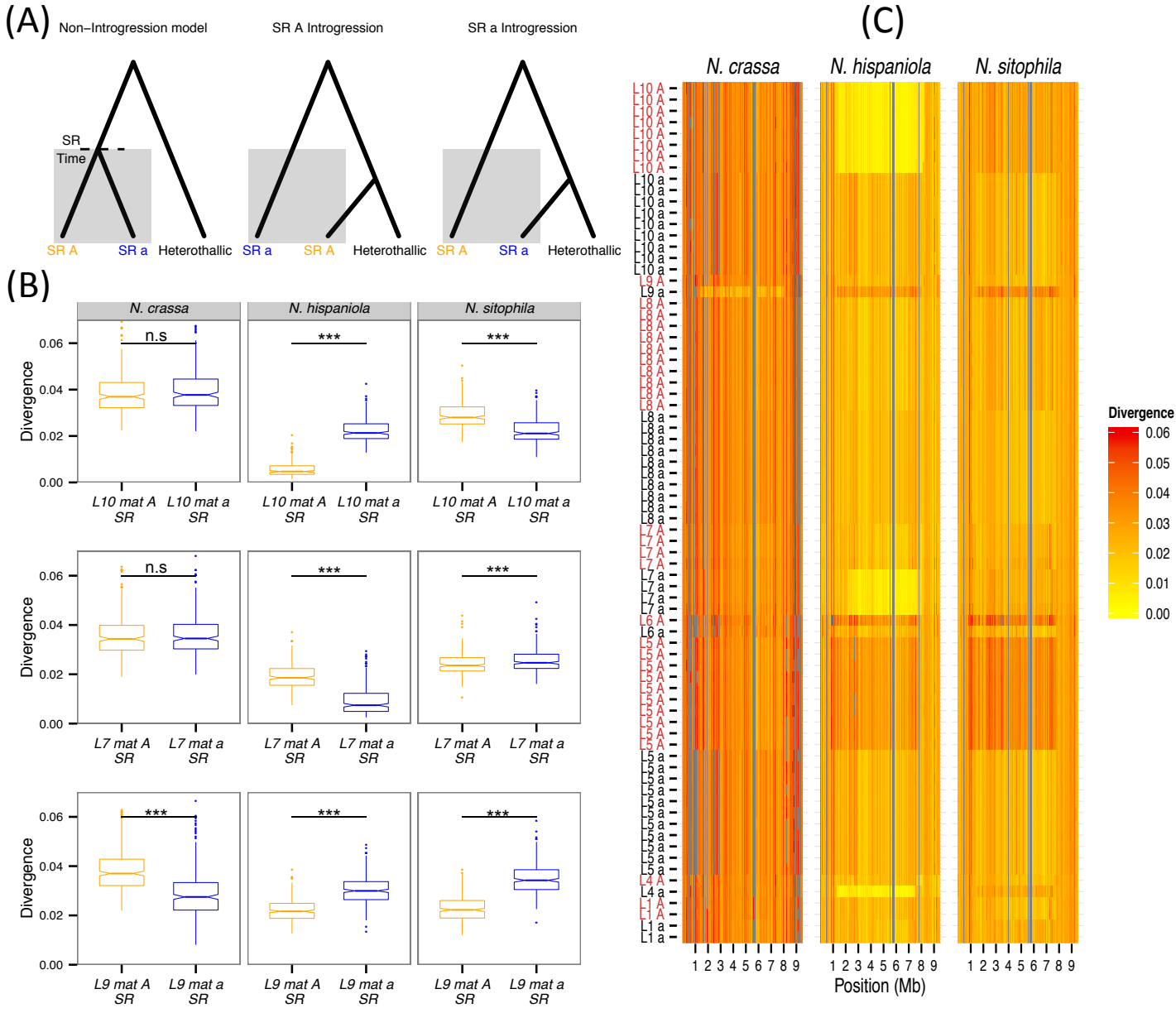
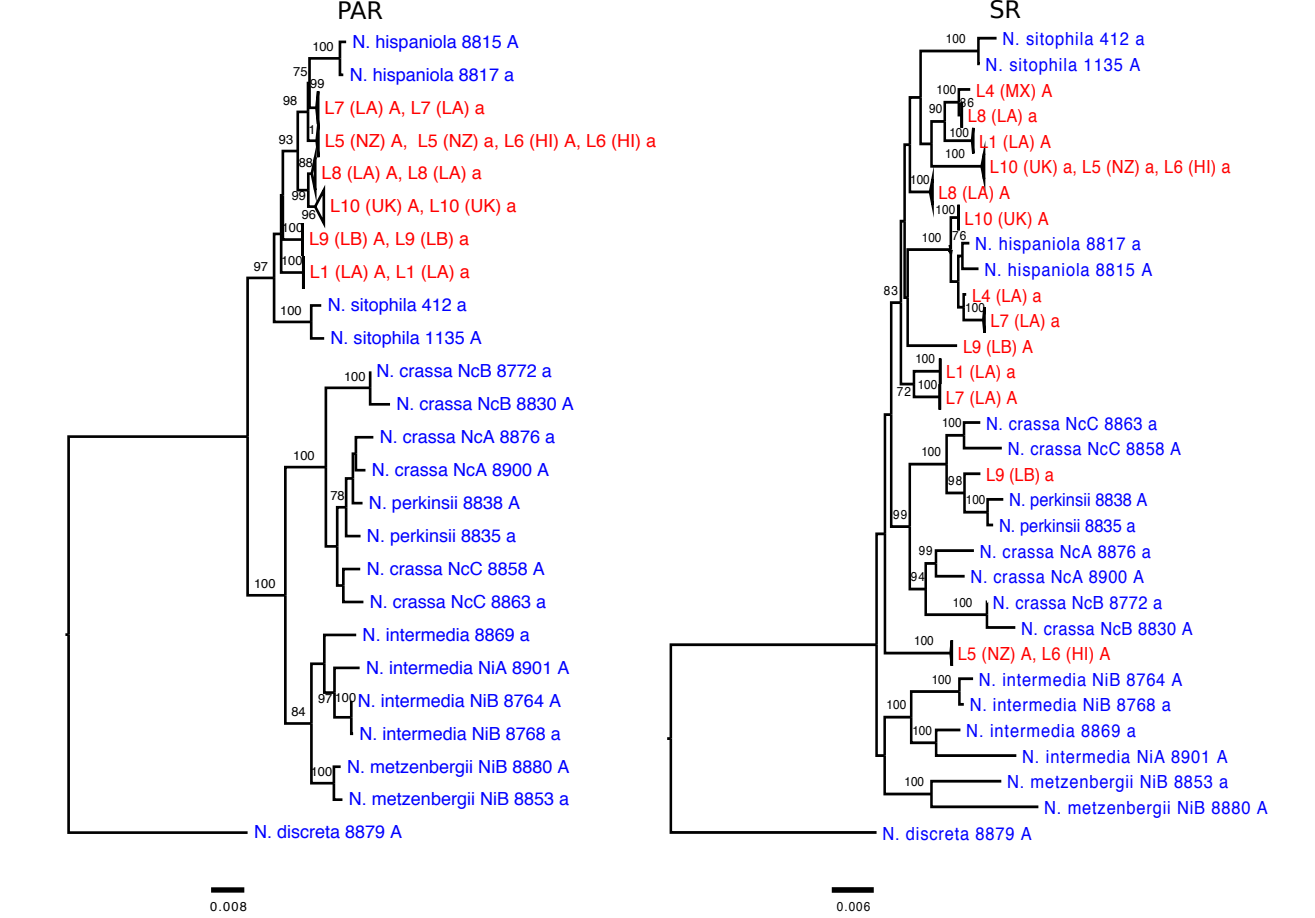

(A)
L10 (UK)
L1 (LA)
L5 (NZ)
L6 (HA)
L9 (LB)
L8 (LA)
L4 (MX)
L7 (LA)
N. sitophila
0.02
(B)
PC1
PC2
Lineage
L1
L10
L4
L5
L6
L7
L8
L9
(C)
L10 (UK)
L8 (LA)
L7 (LA)
L1 (LA)
L5 (NZ)
L6 (HI)
L9 (LB)
L4 (MX)

Figure 2


L6 (HI)
L4 (MX)
L1 (LA)
L9 (LB)
L10 (UK)
L5 (NZ)
L8 (LA)
L7 (LA)
2508v2509
7585v7586
9033v9034
965Av965a
CJ03vCJ04
CJ15vCJ16
CJ25vCJ26
CJ37vCJ38
CJ57vCJ58
CJ59vCJ60
CJ69vCJ70
I
II
III
IV
V
VI
VII
Divergence
0.100
0.075
0.050
0.025
0.000
1 3 5 7 9
Position (Mb)


(A)
L5 (NZ)
L8 (LA)
L10 (UK)
Mean r2
mat a−1
Position (Mb)
(B)
A
a
pi
Tajima's D
Fay and Wu's H


(A)
Non−Introgression model
SR A Introgression
SR a Introgression
SR
Time
SR A
SR a
Heterothallic
(B)
N. crassa
N. hispaniola
N. sitophila
Divergence
n.s
***
L10 mat A SR
L10 mat a SR
L7 mat A SR
L7 mat a SR
L9 mat A SR
L9 mat a SR
(C)
N. crassa
N. hispaniola
N. sitophila
Position (Mb)
Divergence


PAR
N. hispaniola 8815 A
N. hispaniola 8817 a
L7 (LA) A, L7 (LA) a
L5 (NZ) A, L5 (NZ) a, L6 (HI) A, L6 (HI) a
L8 (LA) A, L8 (LA) a
L10 (UK) A, L10 (UK) a
L9 (LB) A, L9 (LB) a
L1 (LA) A, L1 (LA) a
N. sitophila 412 a
N. sitophila 1135 A
N. crassa NcB 8772 a
N. crassa NcB 8830 A
N. crassa NcA 8876 a
N. crassa NcA 8900 A
N. perkinsii 8838 A
N. perkinsii 8835 a
N. crassa NcC 8858 A
N. crassa NcC 8863 a
N. intermedia 8869 a
N. intermedia NiA 8901 A
N. intermedia NiB 8764 A
N. intermedia NiB 8768 a
N. metzenbergii NiB 8880 A
N. metzenbergii NiB 8853 a
N. discreta 8879 A
0.008
SR
N. sitophila 412 a
N. sitophila 1135 A
L4 (MX) A
L8 (LA) a
L1 (LA) A
L10 (UK) a, L5 (NZ) a, L6 (HI) a
L8 (LA) A
L10 (UK) A
N. hispaniola 8817 a
N. hispaniola 8815 A
L4 (LA) a
L7 (LA) a
L9 (LB) A
L1 (LA) a
L7 (LA) A
N. crassa NcC 8863 a
N. crassa NcC 8858 A
L9 (LB) a
N. perkinsii 8838 A
N. perkinsii 8835 a
N. crassa NcA 8876 a
N. crassa NcA 8900 A
N. crassa NcB 8772 a
N. crassa NcB 8830 A
L5 (NZ) A, L6 (HI) A
N. intermedia NiB 8764 A
N. intermedia NiB 8768 a
N. intermedia 8869 a
N. intermedia NiA 8901 A
N. metzenbergii NiB 8853 a
N. metzenbergii NiB 8880 A
N. discreta 8879 A
0.006