

ORIGINAL ARTICLE

DNA polymorphism in recombining and non-recombining mating-type-specific loci of the smut fungus *Microbotryum*

AA Votintseva and DA Filatov

Department of Plant Sciences, University of Oxford, Oxford, UK

The population-genetic processes leading to the genetic degeneration of non-recombining regions have mainly been studied in animal and plant sex chromosomes. Here, we report population genetic analysis of the processes in the non-recombining mating-type-specific regions of the smut fungus *Microbotryum violaceum*. *M. violaceum* has A1 and A2 mating types, determined by mating-type-specific ‘sex chromosomes’ that contain 1–2 Mb long non-recombining regions. If genetic degeneration were occurring, then one would expect reduced DNA polymorphism in the non-recombining regions of this fungus. The analysis of DNA diversity among 19 *M. violaceum* strains, collected across Europe from *Silene latifolia* flowers, revealed that (i) DNA polymorphism is relatively low in all 20 studied loci ($\pi \sim 0.15\%$), (ii) it is not significantly different between the two mating-type-specific

chromosomes nor between the non-recombining and recombining regions, (iii) there is substantial population structure in *M. violaceum* populations, which resembles that of its host species, *S. latifolia*, and (iv) there is significant linkage disequilibrium, suggesting that widespread selfing in this species results in a reduction of the effective recombination rate across the genome. We hypothesise that selfing-related reduction of recombination across the *M. violaceum* genome negates the difference in the level of DNA polymorphism between the recombining and non-recombining regions, and may possibly lead to similar levels of genetic degeneration in the mating-type-specific regions of the non-recombining ‘sex chromosomes’ and elsewhere in the genome.

Heredity (2011) **106**, 936–944; doi:10.1038/hdy.2010.140; published online 17 November 2010

Keywords: *Microbotryum*; recombination; selfing; DNA polymorphism

Introduction

Recombination is a major determinant in genome evolution (Gaut *et al.*, 2007; Larracuente *et al.*, 2008). Regions that lack recombination tend to lose functional genes and accumulate junk DNA because the evolutionary fates of mutations in non-recombining regions are not independent and natural selection cannot effectively eliminate deleterious and fix advantageous mutations (Hill and Robertson, 1966; Felsenstein, 1974). The best studied cases of such genetic degeneration include animal and plant sex chromosomes (Charlesworth and Charlesworth, 2005). Despite their independent evolution in different groups of animals and plants (Bull, 1983), sex chromosomes have similar properties: recombination is restricted between the X and Y chromosomes, and the male-specific non-recombining Y chromosome exhibits genetic degeneration—the loss of functional genes and accumulation of repetitive DNA. Similar processes are expected to occur in other non-recombining regions, such as plant self-incompatibility loci and fungal mating-type loci. Indeed, it has been suggested that mating-type-determining regions of

fungi, such as *Cryptococcus neoformans* (Lengeler *et al.*, 2002), *Neurospora tetrasperma* (Menkis *et al.*, 2008) and *Microbotryum violaceum* (Hood, 2002; Votintseva and Filatov, 2009), represent fungal ‘sex chromosomes’. However, at present there is no solid evidence for genetic degeneration in fungal mating-type-determining regions. For example, the mating-type locus of the fungal pathogen *Cryptococcus neoformans* contains ~20 genes in about 100 kb of non-recombining DNA and does not show any obvious signs of genetic degeneration (Lengeler *et al.*, 2002; Fraser *et al.*, 2004). Here, we focus on the analysis of population-genetic processes dominating evolution in the non-recombining mating-type-specific regions of the smut fungus *M. violaceum*.

Microbotryum is a basidiomycete that causes anther-smut disease in many species in the plant family Caryophyllaceae. Until recently, the same species name, *M. violaceum*, has been used for strains parasitising different host species. However, multiple sources of evidence (sequence divergence, teliospore colour, cross-mating experiments) support the view that strains on different hosts represent separate species (Le Gac *et al.*, 2007; De Vienne *et al.*, 2009). The systematics of the genus is being revised to assign species names to strains with different host specificities. In particular, the *Microbotryum* strains from *Silene latifolia* analysed below are now called *Microbotryum lynchmidis-dioicae*, but the classification is not yet settled (Denchev *et al.*, 2009), and so here, we will use the traditional species complex name *M. violaceum* to

Correspondence: Dr DA Filatov, Department of Plant Sciences, University of Oxford, South Parks Road, Oxford OX1 3RB, UK.

E-mail: dmitry.filatov@plants.ox.ac.uk

Received 6 December 2009; revised 21 July 2010; accepted 6 September 2010; published online 17 November 2010

refer to strains collected from *S. latifolia*. For strains collected from other host species, we will specify the species name of the host plant.

The life cycle of *M. violaceum* has been reviewed recently (Giraud *et al.*, 2008). Dispersal of the fungus occurs via teliospores that form in the anthers of infected plants and are spread by pollinators. Interestingly, in female plants of dioecious *Silene*, the smut can cause masculinisation of the phenotype and the formation of anther-like structures (Ruddat *et al.*, 1991). Teliospores undergo meiosis on the host plant, and the resulting haploid cells bud off sporidia that eventually conjugate to form infectious dikaryons. Dikaryotic hyphae biotrophically proliferate through host plant tissues. The fungus has a bipolar heterothallic breeding system, and conjugation occurs only between sporidia of opposite mating types. The two mating types, A1 and A2, are determined by a mating-type-specific region, which is similar to the Y chromosomes of animals and plants, as recombination is suppressed in a large region around the mating-type-determining genes (Hood, 2002; Votintseva and Filatov, 2009).

The A1 and A2 mating-type-specific chromosomes of *M. violaceum* are distinguishable by size in pulse-gel electrophoresis (Hood, 2002), and multiple fragments from the A1, A2 and other chromosomes have been isolated (Hood *et al.*, 2004). The analysis of these fragments suggested a lower gene density on the mating-type-specific chromosomes, compared with the rest of the genome in this species (Hood *et al.*, 2004). However, a lower gene density may be due to the addition of junk DNA to the mating-type-specific chromosomes rather than the loss of functional genes. Recent segregation analysis of the mating-type specificity of 86 fragments isolated from the A1 and A2 chromosomes revealed that only about a quarter are located in the non-recombining mating-type-specific regions, whereas the rest mapped to recombining regions and are not expected to undergo genetic degeneration because of a lack of recombination (Votintseva and Filatov, 2009). Based on this analysis, the size of the non-recombining regions was estimated to be about 1 Mb long on each of the mating-type-specific chromosomes; however, this estimate is very approximate. Thus, it remains unclear whether or not the non-recombining mating-type-specific chromosomes in *M. violaceum* are undergoing genetic degeneration via gene loss, as is typical for Y chromosomes in animals and plants.

Comparisons of DNA polymorphism in recombining and non-recombining regions provide a way to test whether the population-genetic processes causing genetic degeneration are acting in the region of interest. The complete linkage of genes in non-recombining regions, such as Y chromosomes, makes the evolutionary fates of mutations in such regions non-independent, which leads to genetic hitchhiking (the fixation of deleterious mutations due to linkage to favourable mutations spreading in the population, (Maynard Smith and Haigh, 1974; Rice, 1987)) and background selection (selection against deleterious mutations at linked genes, resulting in a reduced effective population size and stochastic accumulation of mildly deleterious mutations, (Charlesworth *et al.*, 1993, 1995)). These processes reduce the effective population size and lead to the accumulation of deleterious mutations and a gradual genetic degeneration.

Indeed, multiple studies have reported reduced genetic diversity on Y chromosomes, compared with the X-linked or autosomal genes, in animal and plant species (Filatov *et al.*, 2000; Yi and Charlesworth, 2000).

Another factor affecting genetic diversity throughout *M. violaceum* genome is its tendency to mate between the products of the same meiosis (Hood and Antonovics, 2004). This leads to high degree of selfing, which is expected to reduce diversity throughout the genome. Indeed, the analysis of microsatellite diversity revealed high homozygosity (Bucheli *et al.*, 2001; Giraud, 2004; Granberg *et al.*, 2008). However, DNA diversity in *M. violaceum* populations has not been studied. Here, we analyse DNA polymorphism in a set of mating-type-specific loci in *M. violaceum* and compare it with DNA polymorphism in loci not linked to the mating-type locus. If non-recombining mating-type-specific regions do undergo genetic degeneration in *M. violaceum*, similar to the Y chromosomes of diploid species, then loci in these regions are expected to have reduced genetic diversity compared with non-mating-type-specific loci; however, it is not clear whether the reduction of DNA diversity in the non-recombining regions is detectable if the overall DNA diversity is reduced because of selfing.

Materials and methods

Samples

The samples of *M. violaceum* used in this study were collected across Europe and are listed in Table 1. Teliospores were collected from flowers of *S. latifolia* plants, and cultures were grown in petri dishes on 3.6% potato dextrose agar medium. DNA was extracted from these mixed mating-type cultures using Invitrogen Charge Switch magnetic beads-based DNA extraction

Table 1 *M. violaceum* samples used in the study

| ID | Host plant | Location | Country |
|---------|-----------------------|--------------|-------------|
| SI-769 | <i>S. latifolia</i> | Galicia | Spain |
| SI-920 | <i>S. latifolia</i> | Valencia | Spain |
| SI-562 | <i>S. latifolia</i> | nr Krakow | Poland |
| SI-563 | <i>S. latifolia</i> | nr Krakow | Poland |
| SI-405 | <i>S. latifolia</i> | nr Stockholm | Sweden |
| SI-406 | <i>S. latifolia</i> | nr Stockholm | Sweden |
| SI-432 | <i>S. latifolia</i> | nr Stockholm | Sweden |
| SI-433 | <i>S. latifolia</i> | nr Stockholm | Sweden |
| SI-1088 | <i>S. latifolia</i> | nr Lausanne | Switzerland |
| SI-1089 | <i>S. latifolia</i> | nr Lausanne | Switzerland |
| SI-127 | <i>S. latifolia</i> | Pyrenees | France |
| SI-320 | <i>S. latifolia</i> | Central | France |
| SI-321 | <i>S. latifolia</i> | North | France |
| SI-1099 | <i>S. latifolia</i> | nr Guéret | France |
| SI-144 | <i>S. latifolia</i> | nr Chernobyl | Ukraine |
| SI-1069 | <i>S. latifolia</i> | Cotswolds | England |
| SI-1103 | <i>S. latifolia</i> | nr Woodstock | England |
| SI-1075 | <i>S. latifolia</i> | Sardinia | Italy |
| SI-1076 | <i>S. latifolia</i> | Sardinia | Italy |
| Sd-1060 | <i>S. dioica</i> | East | Scotland |
| Sd-1066 | <i>S. dioica</i> | West | Scotland |
| Sn-87 | <i>S. nutans</i> | Alps | France |
| Sv-1 | <i>S. vulgaris</i> | nr La Sage | Switzerland |
| Lfc-109 | <i>L. flos-cuculi</i> | Alps | France |

Abbreviations: *L. flos-cuculi*, *Lychmis flos-cuculi*; *M. violaceum*, *Microbotryum violaceum*; nr, near; *S. dioica*, *Silene dioica*; *S. latifolia*, *Silene latifolia*; *S. nutans*, *Silene nutans*; *S. vulgaris*, *Silene vulgaris*.

kit (Invitrogen Ltd, Paisley, UK). Two samples of *M. violaceum* were also collected from *Silene dioica* plants, which were used as an outgroup, mainly for HKA (Hudson, Kreitman & Aguade) analysis (Hudson *et al.*, 1987), see below. We also attempted to obtain more divergent outgroup sequences from *M. violaceum* strains collected from *S. vulgaris*, *S. nutans* and *Lychmis flos-cuculi*. Total nucleotide divergence between these outgroups and the sequences from *M. violaceum* strains collected from *S. latifolia* exceeded 5%, and in some cases, we could not amplify the loci from these more diverged strains. Using these more diverged sequences changes neither the outcome of the HKA test nor the conclusions of the work; hence, throughout the paper, we use *M. violaceum* from *S. dioica* as an outgroup, as it is available for all loci in this study.

Loci

Using pulse-gel electrophoresis, Hood *et al.* (2004) isolated and sequenced random fragments from the A1, A2 and other chromosomes of *M. violaceum*. In an earlier study, we demonstrated that only a fraction of loci isolated from the A1 and A2 chromosomes are linked to mating type, whereas the rest are located in recombining regions and are not mating-type-specific (Votintseva and Filatov, 2009). In this study, we use seven loci that are located in the non-recombining region specific to the A1 mating type and five loci specific to the A2 mating type. We also used eight non-mating-type-specific loci from recombining regions (Table 2) for comparisons with the loci in the non-recombining regions. Two of the mating-type-specific loci corresponded to pheromone receptor genes from the A1 and A2 chromosomes (Yockteng *et al.*,

2007), whereas the rest of the recombining and non-recombining loci were anonymous fragments originally isolated by Hood *et al.* (2004).

DNA sequencing

PCR primers for amplification and sequencing are listed in Table 2. PCR amplification was conducted using Bioline BioMixRed PCR kits (BioLine inc., Taunton, MA, USA). The PCR fragments were checked by standard agarose gel electrophoresis and purified using Qiagen PCR purification or Gel extraction kits (Qiagen Ltd, Crawley, UK). Direct DNA sequencing of PCR-amplified fragments was conducted using ABI BigDye v3.1 on an ABI 3730 automated DNA sequencer (Applied Biosystems, Foster City, CA, USA). Sequence chromatograms were corrected, and contigs were assembled and checked using ProSeq3 (Filatov, 2009). The sequences were submitted to GenBank under accession numbers HM176794-HM177244.

Direct sequencing of PCR products from mating-type-specific loci generates haploid data, whereas sequencing of the non-mating-type-specific loci generates diploid unphased sequences. Visual inspection of the sequencing chromatograms revealed that only two loci in two individuals were heterozygous for several nucleotide polymorphisms, whereas the rest were completely homozygous, which simplified the reconstruction of haplotypes for non-mating-type-specific loci. Haplotypes for these sequences were reconstructed using fastPhase (Scheet and Stephens, 2006). Alignments of the resulting sequences were loaded into a ProSeq3 database, allowing sequences to be linked to the individuals from which they were obtained. Such linking makes the handling and analysis of multigenic DNA polymorphism data sets more convenient (Filatov,

Table 2 Loci used in this study

| N | Loci | PCR primers | | T _m , °C ^a |
|--|----------|-----------------------------------|----------------------------------|----------------------------------|
| | | Forward primers | Reverse primers | |
| <i>Non-recombining A1 mating-type specific</i> | | | | |
| 1 | A1-pr | A1-prF TGGCATCCCTCAATGTTTCC | A1-prR CACCTTTTGATGAGAGGCCG | 54.6/54.6 |
| 2 | A1-077 | A1-077f CCCTCGACGATGTTCTTCTTG | A1-077r CCAGTCGAATATTGGAAACCG | 53.8/53.7 |
| 3 | A1-116 | A1-116f TGAGTGATGGTCGACAGGTCA | A1-116r TGACATCAGCTGAGTTCACAGT | 53.5/53.6 |
| 4 | A1-202 | A1-202f ACAGTTCATTCACAACGGCT | A1-202r TCCTCGACCGTATTATCGTG | 51.8/51.9 |
| 5 | A1-210 | A1-210f GGAATGCAGTGGGTCATGGT | A1-210r CGTTGCGATGACGAGGTATCT | 54.3/54.2 |
| 6 | A1-276 | A1-276f AGAACAGAGGCCAAGGGAGC | A1-276r CCTGAACAACCCCATCTACCC | 54.8/54.7 |
| 7 | A1-281 | A1-281f ACATGTTCTGATAGCGATTCC | A1-281r GTAGAGTAAGATGGCGGCCG | 52.7/53.9 |
| <i>Non-recombining A2 mating-type specific</i> | | | | |
| 8 | A2-pr | A2-prF TGACGAGAGCATTCTACCG | A2prR GAAGCGGAACCTTGCCTTCT | 52.7/52.8 |
| 9 | A2-400 | A2-400f TGGATGAGATGAAGTTGCGC | A2-400r CATACTTTTGGCGGGTTCA | 53.5/53.3 |
| 10 | A2-441 | A2-441f GNCTTGAAGCGAAGGCAAAC | A2-441r CGAAACTTACGGCTGCCATAG | 54.1/54.2 |
| 11 | A2-500 | A2-500f GAAGAGGGGGTTTTGACGAC | A2-500r CCCAACATGACCTCTTTTCA | 52.5/52.3 |
| 12 | A2-530 | A2-530f CAAGATTGAGGATAGAAGATGGATG | A2-530r GAAGTATTCGAGGTTCTTGTGGTT | 53.0/52.6 |
| <i>Recombining</i> | | | | |
| 13 | paA1-134 | A1-134f GGGCTTCGGAAGAAATGATC | A1-134r TCTTCCTTCCCTCAGATGGC | 53.1/53.0 |
| 14 | paA1-211 | A1-211f ACAGATCTCACAAAGTTCGAGA | A1-211r CACAGTGGGCTATATTACACAAAT | 53.2/53.2 |
| 15 | paA2-433 | A2-433f GTCTACCGAGTCAGACTGCACA | A2-433r GGATCTATGCAGCCTTCGAC | 53.5/53.4 |
| 16 | paA2-461 | A2-461f CAGAAATTGGAAGCGATTGA | A2-461r ATTCCTTGCAAGTCTCGACC | 53.3/53.2 |
| 17 | paA2-547 | A2-547f TGGCTGCTTCTTTATGACCA | A2-547r CAAACAATTTCAATCGATGCAG | 50.9/52.5 |
| 18 | Aut-932 | Aut-932f ATTCGACTGCTCCCACTCTG | Aut-932r AGGAGGAGAAAGTCGATCGC | 51.6/52.5 |
| 19 | Aut-935 | Aut-935f TGTGGTCAGTAGAGAATGGCG | Aut-935r TCATCGTTGATCCGACCTGT | 52.7/52.5 |
| 20 | Aut-948 | Aut-948f GCTTGTGTCAATCATTCATCA | Aut-948r TATGCTAATGCGGGAACAAG | 50.1/50.6 |

Abbreviation: A1-pr and A2-pr, A1- and A2-pheromone receptor genes.

^aAnnealing temperature for each primer; PCR was conducted at the T_m intermediate for the two primers.

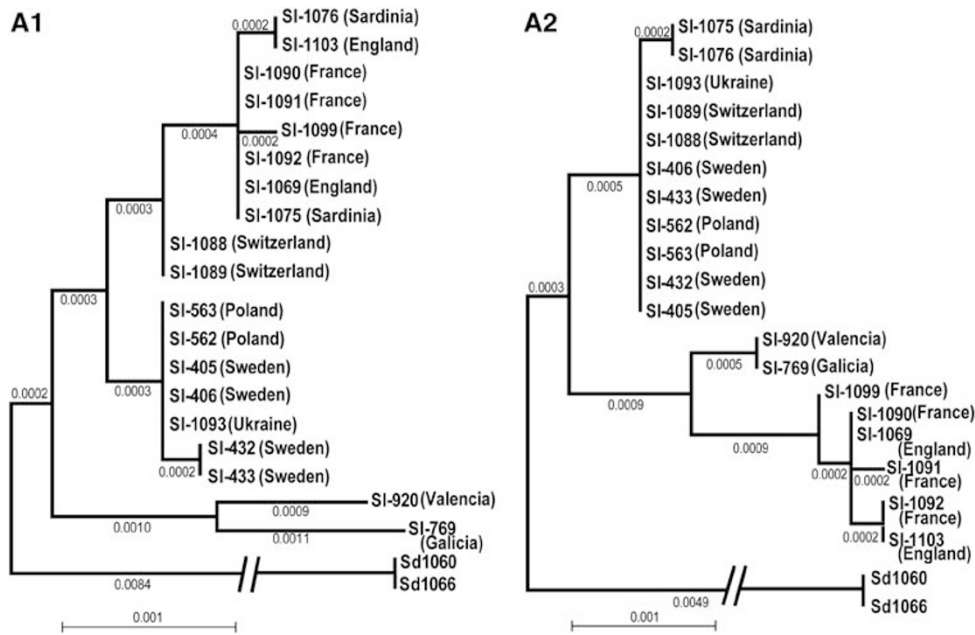


Figure 1 Neighbour-joining phylogenies of *M. violaceum* constructed from concatenated A1- and A2-linked genes. The branch lengths are proportional to the maximum-composite likelihood distance from MEGA4 (Tamura *et al.*, 2007).

2009), as any action is applied once to all individuals rather than 20 times to each of the 20 alignments in the data set.

DNA polymorphism analyses

Neighbour-joining trees for concatenated non-recombining regions (Figure 1) were reconstructed with MEGA4 (Tamura *et al.*, 2007). The basic descriptive statistics of DNA polymorphism, average number of differences per nucleotide, π (Tajima, 1983), Watterson's θ (Watterson, 1975) and Tajima's D (Tajima, 1989) were calculated using ProSeq3 (Filatov, 2009). The HKA test (Hudson *et al.*, 1987) was conducted using Jody Hey's HKA programme (<http://genfaculty.rutgers.edu/hey/software>) on concatenations of all A1-linked, A2-linked and loci unlinked to mating type. Linkage disequilibrium was measured using the D' statistic (Lewontin, 1964) and its significance was assessed with χ^2 , as implemented in DnaSP5 (Librado and Rozas, 2009). The multilocus LD statistic ZnS (Kelly, 1997) was calculated using ProSeq3. The minimum number of recombination events, R_{\min} (Hudson and Kaplan, 1985) and the population recombination rate, R (Hudson, 1987) were estimated using DnaSP5. Bayesian clustering analysis of recombining loci was conducted with Structure2 software (Pritchard *et al.*, 2000).

The power analysis to detect reduction of DNA polymorphism in the non-recombining regions, compared with recombining loci was conducted using coalescent simulations (Hudson, 2002) in two-deme model with no migration. In the simulation, one sequence corresponding to the outgroup was sampled from one deme, whereas the samples corresponding to the number of individuals sequenced from *M. violaceum* from *S. latifolia* plants were sampled from the other deme. The simulation parameters were adjusted to fit the observed level of polymorphism and divergence from the outgroup. While simulating non-recombining loci,

we reduced polymorphism 4- and 10-fold (2- and 5-fold after adjustment for 2-fold ploidy difference), compared with recombining loci, whereas keeping divergence from the outgroup unchanged. The simulated 10 000 data sets were run through J Hey's HKA programme using Perl script and the significance of rejecting neutrality was recorded.

Results and discussion

DNA polymorphism

To estimate DNA polymorphism in the non-recombining mating-type-specific regions, we sequenced seven A1-specific and five A2-specific loci in 19 *M. violaceum* strains collected from *S. latifolia* populations across Europe (Tables 1 and 2). The absence of recombination in these loci was checked in a previously described genetic cross (Votintseva and Filatov, 2009). To compare the diversity in recombining and non-recombining regions, we also sequenced eight non-sex-specific loci in the same sample of 19 *M. violaceum* strains. Three of these loci were originally isolated by Hood *et al.*, (2004) from chromosomes other than the A1 and A2, and five of the non-mating-type-specific loci come from fragments originally isolated from the A1 or A2 pulse-gel bands (Hood *et al.*, 2004); however, they amplify from both mating types, and there is no divergence between the sequences of the fragments amplified from the A1 and A2 cultures (Votintseva and Filatov, 2009). We refer to all these fragments as 'recombining loci'.

The same regions were also sequenced from two outgroup samples of *M. violaceum* that parasitise *S. dioica*. DNA divergence between the strains parasitising *S. latifolia* and *S. dioica* was much higher than the polymorphism between the strains from *S. latifolia* (Figure 1). This supports the view that *M. violaceum* strains living on different *Silene* species are in fact

Table 3 DNA polymorphism in non-recombining mating-type-specific (A1 and A2), and recombining genes for the entire sample of *M. violaceum* from *S. latifolia* plants

| | N ^a | L ^b | S ^c | π (%) ^d | s.d. (π) ^e | TajD ^f | ZnS ^g | R ^h | R _{min} ⁱ |
|-------------|----------------|----------------|----------------|---------------------------|--------------------------------|-------------------|------------------|----------------|-------------------------------|
| A1 | 19 | 4652 | 23 | 0.11 | 0.071 | -0.828 | 0.2738 | 0.0002 | 0 |
| A2 | 19 | 3136 | 16 | 0.2 | 0.07 | 1.352 | 0.4168 | 0.0006 | 0 |
| A1+A2 | 19 | 7791 | 39 | 0.15 | 0.071 | 0.088 | 0.24 | 0.0007 | 1 |
| Recombining | 38 | 5835 | 53 | 0.22 | 0.105 | 0.0228 | 0.1078 | 0.0066 | 9 |

Abbreviations: *M. violaceum*, *Microbotryum violaceum*; *S. latifolia*, *Silene latifolia*.

^aSample size.

^bNumber of positions analysed (after excluding indels).

^cNumber of polymorphic sites.

^dAverage heterozygosity (per 100 sites).

^es.d. of average heterozygosity (per 100 sites).

^fTajima's *D* (Tajima, 1989).

^gMultilocus linkage disequilibrium statistic *ZnS* (Kelly, 1997).

^hPer-nucleotide population recombination rate (Hudson, 1987).

ⁱMinimal number of recombination events (Hudson and Kaplan, 1985).

separate species (Le Gac *et al.*, 2007) and that the *M. violaceum* samples collected from *S. latifolia* belong to the same biological species, whereas *M. violaceum* from different host species (for example, *S. dioica*) belong to different species.

The total alignment length was 14 044 bp, with 4896 bp in the concatenated A1-linked loci, 3150 bp in the A2-linked loci and 5998 bp in the non-mating-type-specific loci. The analysis of these alignments identified 23, 16 and 53 single-nucleotide polymorphisms between *M. violaceum* strains for the A1-, A2- and non-mating-type-specific loci, respectively (Table 3). As most loci analysed in this paper originated as random fragments isolated from mating-type-specific chromosomes, no information with regard to coding regions is available; thus, we do not distinguish between silent and non-silent sites, and report DNA polymorphism and divergence for all sites only. Average heterozygosity (π) per 100 sites in the A1- and A2-linked non-recombining regions was relatively low, 0.11 and 0.20%, respectively. Polymorphism in the recombining regions was similar, 0.22%, providing no evidence that the lack of recombination in the mating-type-specific regions has led to a reduction in DNA polymorphism in *M. violaceum* (Table 3). Indeed, an HKA test that takes into account the twice lower ploidy for mating-type-specific loci as well as possible differences in mutation rates did not reject the null hypothesis ($P=0.887$) that the differences in DNA polymorphism between the compared loci are due to chance. Coalescent simulations (see methods) demonstrate that with our data set we have 48 and 73% power to reject neutrality with two- and five-fold (after adjusting for ploidy difference) reduction of polymorphism in the non-recombining region. Thus, the non-significant HKA for our data set suggests that if there is any reduction of DNA diversity in the non-recombining region of *M. violaceum* genome, such reduction is very modest.

A recurrent statement in fungal literature is that genetic diversity around the mating-type loci is expected to be higher than elsewhere in the genome (Glass *et al.*, 1988; Zambino *et al.*, 1997; James *et al.*, 2004). This is only the case if the divergence between the mating-type-specific alleles is included in the estimate of polymorphism.

As mating-type-specific regions do not recombine, they gradually diverge from each other, which inflates overall genetic diversity in the region. A similar effect can be seen around self-incompatibility loci in plants and close to any locus under balancing selection (Uyenoyama, 2005; Charlesworth, 2006). However, this inflation of genetic diversity occurs only between the alleles and does not affect diversity within the allelic classes. In the case of *M. violaceum*, there are two allelic classes in the mating-type-specific region, A1- and A2-specific alleles. If one includes both A1- and A2-specific alleles in the calculation of genetic diversity, it will be elevated around the mating-type locus (because A1 and A2 are diverged from each other), compared with the rest of the genome. On the other hand, taking into account only A1- or only A2-specific alleles, as was the case in our calculation of DNA diversity in *Microbotryum*, genetic diversity is expected to be the same as elsewhere in the genome or even reduced because of background selection and selective sweeps in the non-recombining region.

Population structure

Similar levels of DNA polymorphism in the recombining and non-recombining loci may perhaps be because of population structure, inflating intraspecific DNA polymorphism in the non-recombining regions due to divergence between populations (Charlesworth *et al.*, 1997; Pannell and Charlesworth, 2000). It is evident from Figure 1 that geographically close samples cluster together in the gene trees, suggesting that the phylogenies based on A1- and A2-linked loci reflect geographic structuring of *M. violaceum* populations. However, these phylogenies are based on the non-recombining loci from the mating-type-linked loci, and the recombining loci may not reveal the same geographic structure. We used Bayesian clustering of samples as implemented in the software Structure2 (Pritchard *et al.*, 2000) to analyse the signal of population structure present in the recombining loci. The posterior probability of the data, $\text{LnP}(D)$, given the number of clusters (K), which is often used as a criterion of the true number of clusters in the data, was highest at $K=7$, but did not form a clear peak (Figure 2a). Instead it plateaus after reaching $K=7$, and the variance between the runs starts increasing. However, if we use the ΔK criterion (Evanno *et al.*, 2005), the most likely number of clusters is either 2 or 4, as ΔK forms two clear peaks (Figure 2b). The cluster memberships identified by Structure2 at $K=2, 4$ and 7 are shown in Figure 3. With $K=2$, one cluster is composed of Northern and Eastern European samples, whereas the other includes all other samples collected across Southern and Western Europe. Surprisingly, one of the Iberian samples, sample 920, from Valencia appears closer to the Eastern European samples, whereas the other Iberian sample (796 from Galicia) is closer to the Western samples. With $K=4$, the East/West subdivision remains (groups 3/1+2), but the two Iberian samples fall into a cluster of their own (group 4) and the Sardinian and Swiss samples (group 1) split up from the rest of the Western European cluster (group 2). With seven clusters, the picture is more complicated and difficult to interpret (Figure 3c). Given that there is no support for $K=7$ using ΔK and $K=2$ has much lower posterior probability of the

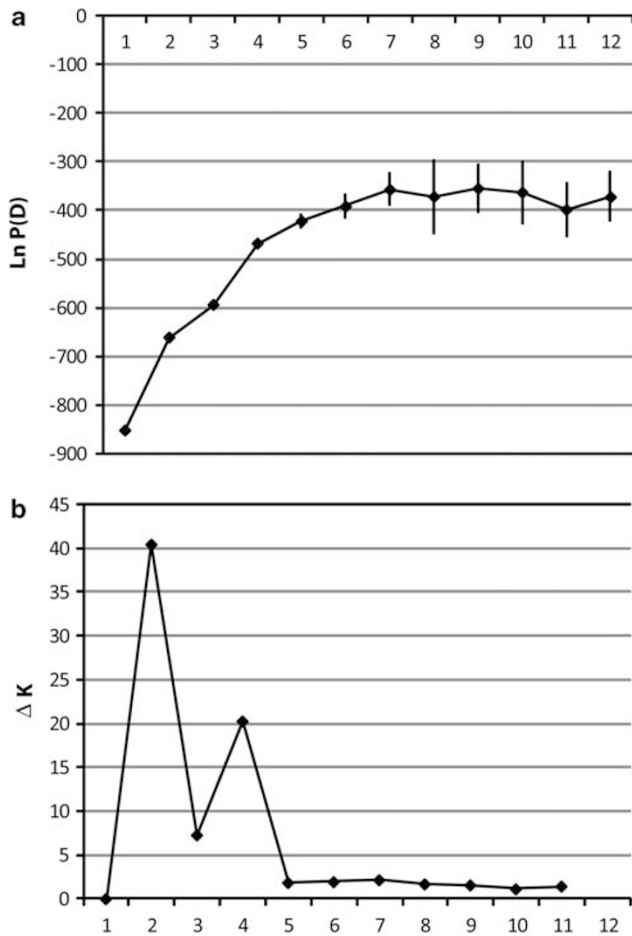


Figure 2 The results of Bayesian clustering analysis of recombining genes using Structure2 (Pritchard *et al.*, 2000). The horizontal axis shows the number of clusters (K). (a) The posterior probability of the data, given the number of clusters, averaged across 10 runs with s.d. between runs. (b) The second order rate of change of the likelihood function (ΔK) with respect to K (Evanno *et al.*, 2005).

data than $K=4$, four clusters seem to be the optimal number that reflects the population structure present in our sample.

To test whether the population structure affects the comparisons of DNA polymorphism between the recombining and non-recombining regions of the *M. violaceum* genome, we compared the diversity in these regions within the groups identified by Structure analysis. As the clustering of the samples was done with recombining loci only, this approach could underestimate the DNA polymorphism in the recombining regions, and thus the comparison between recombining and non-recombining regions is conservative, if used to test whether the non-recombining regions lack DNA polymorphism compared with recombining regions. This intra-population analysis revealed that the non-recombining A1- and A2-linked loci contain less polymorphism than the recombining loci (Figure 4). However, once the two-fold difference in ploidy between the mating-type-specific and non-mating-type-specific loci is taken into account, this difference in DNA polymorphism is not significant (HKA test, $P > 0.05$).

In structured populations, background selection reduces the DNA polymorphism within populations,

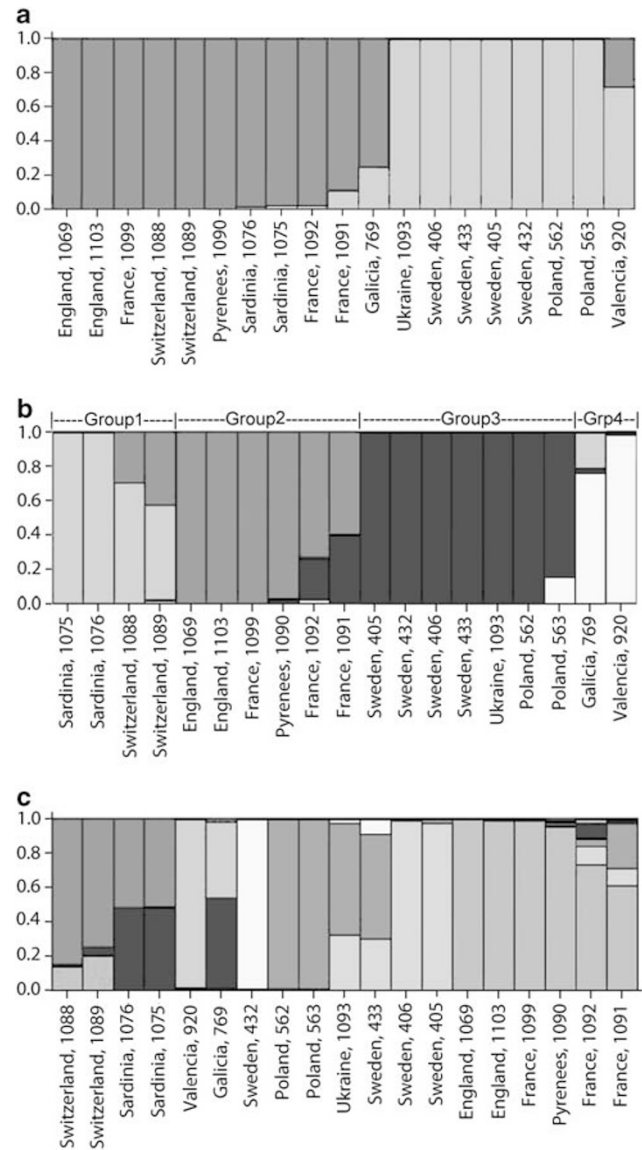


Figure 3 Cluster membership for $K=2$ (a), 4 (b) and 7 (c) based on the recombining regions in the data set, identified by Structure (Pritchard *et al.*, 2000).

but does not reduce between population differentiation (Charlesworth *et al.*, 1997), leading to stronger population subdivision in non-recombining regions (for example, Ironside and Filatov, 2005). This could negate the difference in the overall DNA polymorphism between recombining and non-recombining regions. Local (intra-population) selective sweeps are expected to have a similar effect, whereas global sweeps (across the species range) should lead to a loss of genetic diversity and population structure preferentially in non-recombining regions. Thus, the presence of population subdivision argues against global sweeps in the non-recombining mating-type-specific regions of *M. violaceum*.

DNA polymorphism and selfing

The high degree of selfing in *M. violaceum* (Giraud *et al.*, 2005) is another factor that may affect the overall level of DNA polymorphism across the genome of this species. In accordance with published microsatellite-based results (Bucheli *et al.*, 2001; Giraud, 2004; Granberg

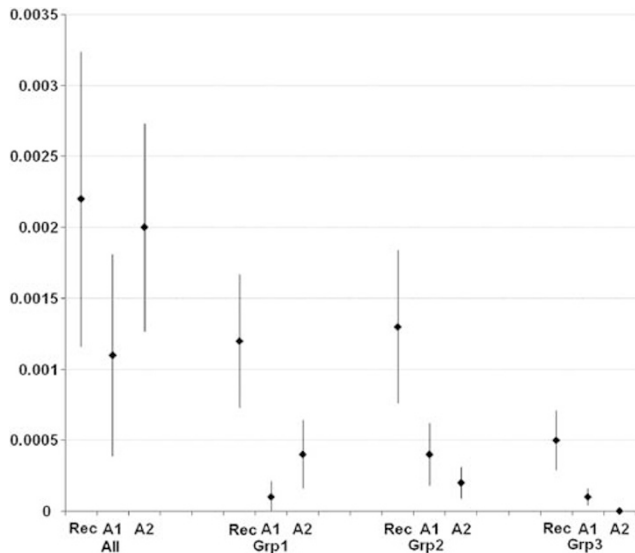


Figure 4 Per-nucleotide DNA polymorphism (π) in recombining (Rec) and non-recombining (A1 and A2) regions of the *M. violaceum* genome for the entire sample (three leftmost data points) and for three sub-populations (groups 1–3, as defined in Figure 3b). Group 4 consists only of two Iberian individuals and is not shown. The error bars represent s.d. of the DNA diversity estimates.

et al., 2008), we observed a very low proportion of heterozygotes in our sequence data in *M. violaceum*: only two out of the nineteen individuals in our study had any heterozygous sites in any of the non-mating-type-linked loci studied. A high degree of selfing inflates homozygosity, which leads to a low effective recombination rate in the recombining regions and reduces the difference between the recombining and non-recombining regions. Although for most loci in the recombining regions in our data set, the minimum number of recombination events (Hudson and Kaplan, 1985) is greater than zero, the overall effective recombination rate is still low (Table 3). There is significant linkage disequilibrium (LD) within as well as between most loci in the data set (Figure 5), suggesting that LD in *M. violaceum* genome extends for megabases, which is similar to LD in *Caenorhabditis elegans* (Rockman and Kruglyak, 2009), but much wider than in other studied selfers, such as *Arabidopsis thaliana* (Kim *et al.*, 2007). Genetic mapping in *M. violaceum* revealed crossing-over rates of 0–15% (Garber *et al.*, 1987; Votintseva and Filatov, 2009). Thus, substantial LD in *M. violaceum* genome is probably due to high degree of homozygosity in this selfing species rather than to lack of crossing over.

Although selfing reduces the effective recombination rate (via increased homozygosity), LD is still at least

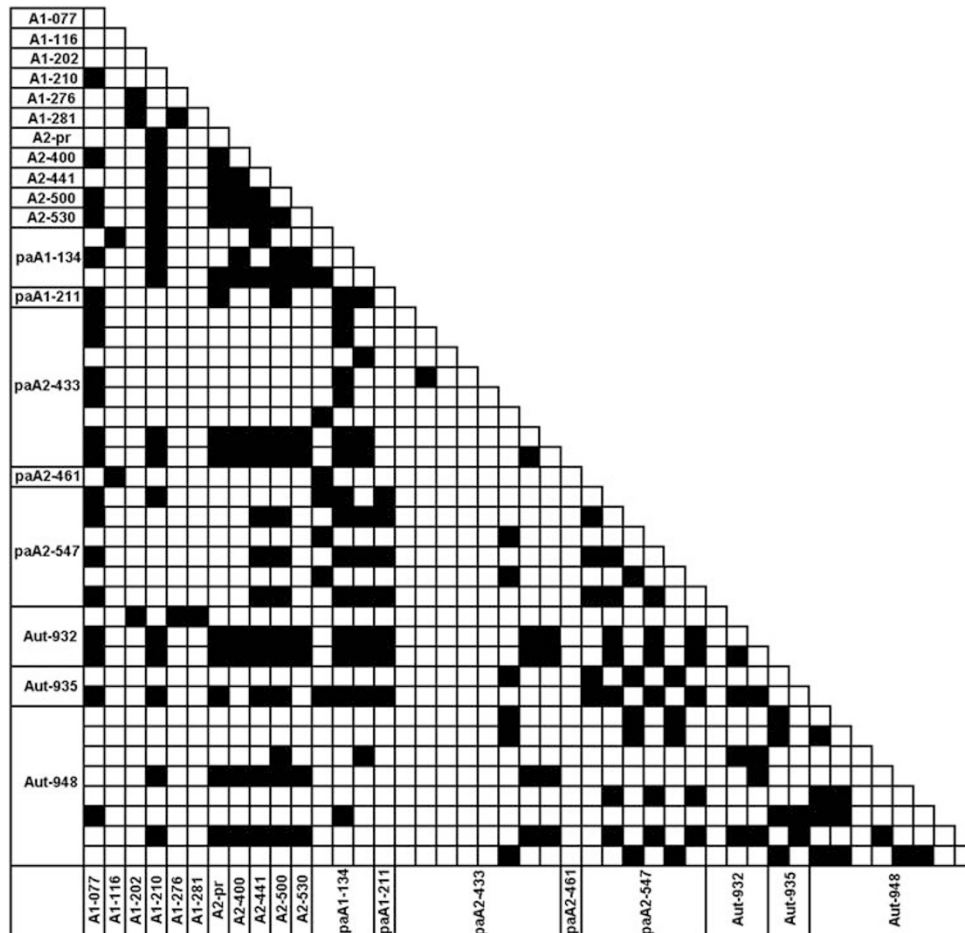


Figure 5 LD between polymorphisms in the studied genes. Significant pairwise LD values (χ^2 -test: $P < 0.05\%$) are shown as black boxes. As polymorphisms in the non-recombining regions are linked, only one randomly chosen polymorphic site is shown for the A1- and A2-linked loci to save space. For the recombining (pa- and aut-) loci, all polymorphisms are shown.

twice as high in the non-recombining A1- and A2-linked loci as in the recombining regions when measured by the ZnS statistic (Kelly, 1997); see Table 3. Thus, non-recombining regions might be more prone to various forms of hitchhiking, compared with the rest of the genome despite the selfing-related lack of recombination in the recombining regions. However, the extent to which a region is affected by hitchhiking is proportional not only to LD but also to the number of functional genes and other targets of selection that are linked together. Although LD is higher in non-recombining regions, these regions are relatively small—about 1Mb (Votintseva and Filatov, 2009)—and may not contain many genes. On the other hand, although LD in recombining regions of *M. violaceum* is lower than in A1- and A2-linked loci, the number of genes ‘linked’ together by LD is large, as LD extends over the most of the genome. Thus, the genes in the non-recombining regions may not be much affected by genetic hitchhiking (background selection and sweeps) than genes elsewhere in the genome.

Conclusion

Our comparison of DNA polymorphism in recombining and non-recombining mating-type-specific regions of *M. violaceum* revealed relatively low polymorphism in all loci and no evidence for a significant reduction of polymorphism in non-recombining regions. We also observed significant population structuring and a high level of homozygosity in our data set, as expected in a predominantly selfing species. We conclude that the selfing-related reduction of recombination across the *M. violaceum* genome negates the difference in the level of DNA polymorphism between the recombining and non-recombining regions. This conclusion is supported by extensive LD. Thus, genetic degeneration due to a lack of recombination in *M. violaceum* may not be restricted to non-recombining regions of its ‘sex chromosomes’, and genes in this species might undergo similar genetic degeneration in the mating-type-specific regions and elsewhere in the genome. It will be possible to test this hypothesis as soon as the sequence of the *M. violaceum* genome will become available.

Conflict of interest

The authors declare no conflict of interest.

Acknowledgements

We are grateful to Chris Dixon for corrections of the manuscript and to the BBSRC and John Fell fund (Oxford) for funding.

References

- Bucheli E, Gautschi B, Shykoff JA (2001). Differences in population structure of the anther smut fungus *Microbotryum violaceum* on two closely related host species, *Silene latifolia* and *S. dioica*. *Mol Ecol* **10**: 285–294.
- Charlesworth B, Morgan MT, Charlesworth D (1993). The effect of deleterious mutations on neutral molecular variation. *Genetics* **134**: 1289–1303.
- Charlesworth B, Nordborg M, Charlesworth D (1997). The effects of local selection, balanced polymorphism and background selection on equilibrium patterns of genetic diversity in subdivided populations. *Genet Res* **70**: 155–174.
- Charlesworth D (2006). Balancing selection and its effects on sequences in nearby genome regions. *PLoS Genet* **2**: e64.
- Charlesworth D, Charlesworth B (2005). Sex chromosomes: evolution of the weird and wonderful. *Curr Biol* **15**: R129–R131.
- Charlesworth D, Charlesworth B, Morgan MT (1995). The pattern of neutral molecular variation under the background selection model. *Genetics* **141**: 1619–1632.
- De Vienne DM, Refregier G, Hood ME, Guigue A, Devier B, Vercken E *et al.* (2009). Hybrid sterility and inviability in the parasitic fungal species complex *Microbotryum*. *J Evol Biol* **22**: 683–698.
- Denchev CM, Giraud T, Hood ME (2009). Three new species of anthericolous smut fungi on Caryophyllaceae. *Mycologica Balcanica* **6**: 79–84.
- Evanno G, Regnaut S, Goudet J (2005). Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol Ecol* **14**: 2611–2620.
- Felsenstein J (1974). The evolutionary advantage of recombination. *Genetics* **78**: 737–756.
- Filatov DA (2009). Processing and population genetic analysis of multigenic datasets with ProSeq3 software. *Bioinformatics* **25**: 3189–3190.
- Filatov DA, Moneger F, Negrutiu I, Charlesworth D (2000). Low variability in a Y-linked plant gene and its implications for Y-chromosome evolution. *Nature* **404**: 388–390.
- Fraser JA, Diezmann S, Subaran RL, Allen A, Lengeler KB, Dietrich FS *et al.* (2004). Convergent evolution of chromosomal sex-determining regions in the animal and fungal kingdoms. *PLoS Biol* **2**: e384.
- Garber ED, Eng C, Stevens DM (1987). Genetics of *Ustilago violacea*. XXI. Centromere-linkage values and pericentric gene clustering. **12**: 555–560.
- Gaut BS, Wright SI, Rizzon C, Dvorak J, Anderson LK (2007). Recombination: an underappreciated factor in the evolution of plant genomes. *Nat Rev Genet* **8**: 77–84.
- Giraud T (2004). Patterns of within population dispersal and mating of the fungus *Microbotryum violaceum* parasitising the plant *Silene latifolia*. *Heredity* **93**: 559–565.
- Giraud T, Jonot O, Shykoff JA (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.
- Giraud T, Yockteng R, Lopez-Villavicencio M, Refregier G, Hood ME (2008). Mating system of the anther smut fungus *Microbotryum violaceum*: selfing under heterothallism. *Eukaryot Cell* **7**: 765–775.
- Glass NL, Vollmer SJ, Staben C, Grotelueschen J, Metzberg RL, Yanofsky C (1988). DNAs of the two mating-type alleles of *Neurospora crassa* are highly dissimilar. *Science* **241**: 570–573.
- Granberg A, Carlsson-Graner U, Arnqvist P, Giles B (2008). Variation in breeding system traits within and among populations of *Microbotryum violaceum* on *Silene dioica*. *Int J Plant Sci* **169**: 293–303.
- Hill WG, Robertson A (1966). The effect of linkage on limits to artificial selection. *Genet Res* **8**: 269–294.
- Hood ME (2002). Dimorphic mating-type chromosomes in the fungus *Microbotryum violaceum*. *Genetics* **160**: 457–461.
- Hood ME, Antonovics J (2004). Mating within the meiotic tetrad and the maintenance of genomic heterozygosity. *Genetics* **166**: 1751–1759.
- Hood ME, Antonovics J, Koskella B (2004). Shared forces of sex chromosome evolution in haploid-mating and diploid-mating organisms: *Microbotryum violaceum* and other model organisms. *Genetics* **168**: 141–146.
- Hudson RR (1987). Estimating the recombination parameter of a finite population model without selection. *Genet Res* **50**: 245–250.
- Hudson RR (2002). Generating samples under a Wright-Fisher neutral model of genetic variation. *Bioinformatics* **18**: 337–338.

- Hudson RR, Kaplan NL (1985). Statistical properties of the number of recombination events in the history of a sample of DNA sequences. *Genetics* **111**: 147–164.
- Hudson RR, Kreitman M, Aguade M (1987). A test of neutral molecular evolution based on nucleotide data. *Genetics* **116**: 153–159.
- Ironside JE, Filatov DA (2005). Extreme population structure and high interspecific divergence of the *Silene* Y chromosome. *Genetics* **171**: 705–713.
- James TY, Liou SR, Vilgalys R (2004). The genetic structure and diversity of the A and B mating-type genes from the tropical oyster mushroom, *Pleurotus djamor*. *Fungal Genet Biol* **41**: 813–825.
- Kelly JK (1997). A test of neutrality based on interlocus associations. *Genetics* **146**: 1197–1206.
- Kim S, Plagnol V, Hu TT, Toomajian C, Clark RM, Ossowski S *et al.* (2007). Recombination and linkage disequilibrium in *Arabidopsis thaliana*. *Nat Genet* **39**: 1151–1155.
- Larracuente AM, Sackton TB, Greenberg AJ, Wong A, Singh ND, Sturgill D *et al.* (2008). Evolution of protein-coding genes in *Drosophila*. *Trends Genet* **24**: 114–123.
- Le Gac M, Hood ME, Fournier E, Giraud T (2007). Phylogenetic evidence of host-specific cryptic species in the anther smut fungus. *Evolution* **61**: 15–26.
- Lengeler KB, Fox DS, Fraser JA, Allen A, Forrester K, Dietrich FS *et al.* (2002). Mating-type locus of *Cryptococcus neoformans*: a step in the evolution of sex chromosomes. *Eukaryot Cell* **1**: 704–718.
- Lewontin R (1964). The interaction of selection and linkage. I. General considerations: heterotic models. *Genetics* **49**: 49–67.
- Librado P, Rozas J (2009). DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **25**: 1451–1452.
- Maynard Smith J, Haigh J (1974). The hitchhiking effect of a favourable gene. **23**: 23–35.
- 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 Genet* **4**: e1000030.
- Pannell JR, Charlesworth B (2000). Effects of metapopulation processes on measures of genetic diversity. *Philos Trans R Soc Lond B Biol Sci* **355**: 1851–1864.
- Pritchard JK, Stephens M, Donnelly P (2000). Inference of population structure using multilocus genotype data. *Genetics* **155**: 945–959.
- Rice WR (1987). Genetic hitchhiking and the evolution of reduced genetic activity of the Y sex chromosome. *Genetics* **116**: 161–167.
- Rockman MV, Kruglyak L (2009). Recombinational landscape and population genomics of *Caenorhabditis elegans*. *PLoS Genet* **5**: e1000419.
- Ruddat M., Kokontis J, Birch L, Garber ED, Chiang K-S, Campanella J *et al.* (1991). Interactions of *Microbotryum violaceum* (*Ustilago violacea*) with its host plant *Silene alba*. *Plant Science* **80**: 157–165.
- Scheet P, Stephens M (2006). A fast and flexible statistical model for large-scale population genotype data: applications to inferring missing genotypes and haplotypic phase. *Am J Hum Genet* **78**: 629–644.
- Tajima F (1983). Evolutionary relationship of DNA sequences in finite populations. *Genetics* **105**: 437–460.
- Tajima F (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123**: 585–595.
- Tamura K, Dudley J, Nei M, Kumar S (2007). MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol Biol Evol* **24**: 1596–1599.
- Uyenoyama MK (2005). Evolution under tight linkage to mating type. *New Phytol* **165**: 63–70.
- Votintseva AA, Filatov DA (2009). Evolutionary strata in a small mating-type-specific region of the smut fungus *Microbotryum violaceum*. *Genetics* **182**: 1391–1396.
- Watterson G (1975). On the number of segregating sites in genetic models without recombination. *Theor Popul Biol* **7**: 256–276.
- Yi S, Charlesworth B (2000). Contrasting patterns of molecular evolution of the genes on the new and old sex chromosomes of *Drosophila miranda*. *Mol Biol Evol* **17**: 703–717.
- Yockteng R, Marthey S, Chiapello H, Gendrault A, Hood ME, Rodolphe F *et al.* (2007). Expressed sequences tags of the anther smut fungus, *Microbotryum violaceum*, identify mating and pathology genes. *BMC Genomics* **8**: 272.
- Zambino P, Groth JV, Lukens L, Garton JR, May G (1997). Variation at the b mating type locus of *Ustilago maydis*. *Phytopathology* **87**: 1233–1239.