Maintenance of Fungal Pathogen Species That Are Specialized to Different Hosts: Allopatric Divergence and Introgression through Secondary Contact

Pierre Gladieux,*†,[1] Elodie Vercken,†,[1] Michael C. Fontaine,[1] Michael E. Hood,[2] Odile Jonot,[1] Arnaud Couloux,[3] and Tatiana Giraud[1]

[1]Laboratoire Ecologie, Systématique et Evolution, Univ. Paris Sud, UMR8079, Orsay, France; CNRS, Orsay, France; AgroParisTech, Orsay, France

[2]Department of Biology, Amherst College

[3]Centre National de Sequençage, Genoscope, CNRS, UMR8030, Evry, France

†These authors contributed equally to the work and should be considered as sharing first authorship.

***Corresponding author:** E-mail: pierre.gladieux@u-psud.fr.

Associate editor: Michael Nachman

Research article

Abstract

Sympatry of species that lack complete prezygotic isolation is ideal for the study of how species can be maintained in the face of potential gene flow. This is particularly important in the context of emerging diseases on new hosts because pathogen adaptation is facilitated by reduced gene flow from ancestral populations. Here, we investigated divergence and gene flow between two closely related fungal species, *Microbotryum lychnidis-dioicae* and *M. silenes-dioicae*, causing anther-smut disease on the wide-spread plant species *Silene latifolia* and *S. dioica*, respectively. Using model-based clustering algorithms on microsatellite data from samples across Europe, we identified rare disease transmission between the host species and rare pathogen hybrids. Using a coalescent-based approach and an isolation-with-migration model, the age of divergence between the two fungal species was estimated at approximately 4.2×10^5 years. Levels of gene flow were low and concentrated in very recent times. In addition, gene flow appeared unidirectional from *M. silenes-dioicae* to *M. lychnidis-dioicae*. Altogether, our findings are consistent with a scenario of recurrent introgressive hybridization but at a very low level and through secondary contact following initial divergence in allopatry. Asymmetry in the direction of gene flow mirrors previous findings on introgression between the two host plants. Our study highlights the consequences of bringing closely related pathogens into contact, which is increasing through modern global changes and favors cross-species disease transmission, hybridization, and introgression by pathogens.

Key words: speciation, gene flow, hybridization, fungi, pathogen, *Microbotryum violaceum*.

Introduction

Speciation in the face of gene flow in sexual populations has long been dismissed because it is expected to overwhelm selection for gene combinations that are locally adapted to difference resources (e.g., Felsensein 1981; Rice 1984). Methods have been developed recently that use DNA variation at multiple loci within and between closely related species to provide insights into the history of speciation, and in particular the degree and timing of gene flow. It has thus become possible to test whether divergence occurred in the face of gene flow or without gene flow (in allopatry) with the potential for gene flow to occur through secondary contact. This analytical framework, known as "divergence population genetics" (Kliman et al. 2000), estimates parameters of speciation history using a coalescent approach and enables comparison of the likelihood of different models of speciation (e.g., Nielsen and Wakeley 2001; Hey and Nielsen 2004; Hey et al. 2004; Hey and Nielsen 2007). Such approaches have been applied to different organisms, supporting the allopatric model in some cases (Kliman et al. 2000) and divergence with gene flow in others (e.g., Städler et al. 2005; Zhou et al. 2007; Geraldes et al. 2008; Nadachowska and Babik 2009).

Fungi, and in particular pathogenic species, are valuable models for the study of speciation through ecological divergence and specialization on different hosts (Giraud 2006; Giraud et al. 2006; Giraud, Refrégier, et al. 2008; Giraud et al. 2010). The divergence population genetics framework has been applied to a couple of plant pathogenic fungi, for which the allopatric model of speciation with no gene flow was not supported (Stukenbrock et al. 2007; Gladieux et al. 2010).

Here, we aim to elucidate speciation mechanisms in *Microbotryum violaceum* sensu lato (Basidiomycota), in particular whether sister species on different hosts were formed and/or are maintained in the face of gene flow. This fungal species complex is responsible for anther-smut disease on many plant species in the Caryophyllaceae (Thrall et al. 1993; Hood et al. 2010). Diploid teliospores of *Microbotryum* are produced in anthers of diseased plants and are

dispersed by pollinators. *Microbotryum violaceum* sensu lato is composed of several sibling species highly specialized on their host plants (Lutz et al. 2005; Kemler et al. 2006; Le Gac, Hood, Fournier, and Giraud 2007; Le Gac, Hood, and Giraud 2007; de Vienne, Hood, and Giraud 2009). We focus on the two sister species *M. lychnidis-dioicae* (Denchev et al. 2009; called MvSl in Le Gac, Hood, Fournier, and Giraud 2007) and *M. silenes-dioicae* (Denchev et al. 2009; called MvSd in Le Gac, Hood, Fournier, and Giraud 2007), parasitizing, respectively, the plants *Silene latifolia* and *S. dioica*, two closely related host plants with largely overlapping ranges.

Several aspects of pre- and postmating barriers between the *Microbotryum* sibling species have been investigated. In vitro crosses showed no evidence of assortative mating among *Microbotryum* species (Van Putten et al. 2005; Le Gac, Hood, and Giraud 2007; Refrégier et al. 2010). Artificial inoculations on plants revealed that hybrids between *M. lychnidis-dioicae* and *M. silenes-dioicae* were viable and fertile at least through the F_2 generation (Van Putten et al. 2003; Le Gac, Hood, and Giraud 2007; de Vienne, Refregier, et al. 2009). In nature, there is partial premating ecological isolation between these two fungal species, resulting from differences in the habitat preferences of the host plants and specificity of pollinators (Van Putten et al. 2007). This ecological reproductive isolation is, however, not complete (Goulson and Jerrim 1997; Van Putten et al. 2007; Minder et al. 2007; Karrenberg and Favre 2008).

The goal of the current study is to assess whether hybrids between *M. lychnidis-dioicae* and *M. silenes-dioicae* exist in the field, whether introgression can be detected, and in this case, whether gene flow appears to have occurred during the speciation process. We collected *Microbotryum* samples from natural populations of *S. dioica* and *S. latifolia*, in many locations across Europe, including some where the two host plants were growing in sympatry. The specific questions addressed in this study were the following: 1) are there hybrids between the sister *Microbotryum* species parasitizing *S. latifolia* and *S. dioica* plants, in particular where the hosts are sympatric? 2) are there cross-species transmissions in natural populations, that is, samples of the *Microbotryum* species specific to *S. latifolia* infecting *S. dioica* or vice versa? 3) when did the pathogen species start to diverge? 4) did gene flow decrease, or increase, or occur continuously during the divergence process?

Materials and Methods

Sample Collection

Analyses were performed on two different types of genetic markers. We used a microsatellite data set to detect hybrids and individuals resulting from cross-species disease transmission (i.e., "spillover" pathogens; Daszak 2000), and we generated DNA sequence data to analyze the history of divergence and gene flow. Most of microsatellite data were originally generated for a study investigating the phylogeography of the two *Microbotryum* species in Europe (Vercken et al. 2010). We used the sequence data set to validate the results on phylogeography obtained using the microsatellite data set (Vercken et al. 2010), but the full analysis of the genetic structure of the species is described in (Vercken et al. 2010).

The microsatellite data set consisted in 1,028 individuals collected across Europe on *S. latifolia* ($n = 677$) and *S. dioica* ($n = 351$). This set included the 1,011 individuals analyzed in (Vercken et al. 2010) plus 17 additional individuals obtained later (detailed sample information and microsatellite data are available in supplementary table S1, Supplementary Material online). We included samples from previously analyzed contact sites between the two plants *S. dioica* and *S. latifolia*, thanks to the sharing of material by the authors (Van Putten et al. 2005; Karrenberg and Favre 2008; Refrégier et al. 2010). We also sampled additional contact sites between the plant species (supplementary table S1; Supplementary Material online).

We analyzed population subdivision, selfing rate, divergence, and gene flow, based on DNA sequence data for a reduced set of 117 individuals collected on *S. latifolia* and 43 individuals collected on *S. dioica* (supplementary table S1, Supplementary Material online). These individuals were chosen as a random subsample of the full set without using prior knowledge on population subdivision but excluding the first-generation hybrids detected and infections resulting from cross-species disease transmission. These two categories of individuals were excluded because their presence in a data set can lead to the false inference of gene flow when species are actually isolated: Cross-species disease transmission does not constitute recombination between fungal species and first-generation hybrids do not necessarily lead to long-term introgression, for instance, if the hybrids are sterile with low viability.

Microsatellite Genotyping and Sequencing

Microsatellite genotyping is described in Vercken et al. (2010). Sequence data were obtained by direct sequencing of nine noncoding loci (eight autosomal, one mitochondrial). Polymerase chain reaction (PCR) primer pairs (see supplementary table S2, Supplementary Material online) were designed based on a shotgun sequence library of autosomal DNA (Hood 2005). DNA was extracted as in Giraud (2004) from anthers full of diploid teliospores, which will be considered to represent fungal "individuals." The PCR program was a touchdown from 58 °C down to 50 °C. Sequencing was performed at the Genoscope on ABI310 using standard techniques. All polymorphisms were checked visually from electrophoregrams using Codon Code Aligner v. 3.0.1 (CodonCode Corporation). The resulting contigs were processed with automated shell and Perl programs and aligned using MAFFT (Katoh et al. 2005) as implemented in Jalview 9.5 (Clamp et al. 2004). The high level of homozygosity in the two *Microbotryum* species (Giraud 2004) enabled direct sequencing of diploid teliospores. All indels were excluded from analyses. Sequences are available from GenBank (accession numbers HQ019166–HQ020319).

Identification of Hybrids and Cross-species Disease Transmission

Hybrids and samples of infections resulting from cross-species disease transmission were identified using a model-based Bayesian clustering algorithm. Because *Microbotryum* species show high rates of selfing (Hood and Antonovics 2000; Giraud 2004; Giraud et al. 2005), we used a method implemented in INSTRUCT (Gao et al. 2007) that relaxes the assumption of Hardy–Weinberg equilibrium and uses a Markov chain Monte Carlo (MCMC) to simultaneously infer population structure and selfing rates. We estimated the proportion q of individual genomes originating in $K = 2$ clusters (that correspond to the two *Microbotryum* species) using five independent chains, each with 5×10^5 steps, 2.5×10^5 burn-in, a thinning interval of 100 steps, and assuming different starting points. Infections resulting from cross-species disease transmission were identified as being collected from one plant species but clustering with samples collected on the other plant species: Individuals with $q > 0.70$ to the other species' cluster were considered as spillover individuals. Hybrids were identified based on inferred mixed ancestry: Individuals with $0.30 < q < 0.70$ to their own species' cluster were considered as hybrids.

Population Subdivision and Mating System

We used the INSTRUCT program to infer population subdivision and selfing rates for the two *Microbotryum* species, using only the sequence data set (similar analyses were performed using the microsatellite data set by Vercken et al. 2010). The mitochondrial locus *mt168* was not included as the method implemented in INSTRUCT requires diploid data. We ran ten independent chains assuming different starting points, each chain with 5×10^5 steps following 1.5×10^6 burn-in and a thinning interval of 10 steps. Selfing rates were updated using the adaptive independence sampler. The program was run for $K = 1$ to $K = 8$ clusters for each species, and 30 independent runs were performed for each value of K. Results were analyzed with CLUMPP v. 1.1.2 (Jakobsson and Rosenberg 2007) using the "Greedy" algorithm with random input order and 10,000 permutations. Distinct modes among runs were identified by finding sets of runs with less 85% similarity in the G′ pairwise similarity matrix ("modes" refer to distinct clustering solutions represented within the set of replicate cluster analyses). CLUMPP was used again to align outputs of the runs with the same clustering mode and to provide average cluster membership coefficients across aligned runs. In cases where several modes were found, only results of the major modes are presented.

The extent of multilocus linkage disequilibrium in each cluster was analyzed using the index of association statistic $\bar{r}_d$ (Maynard-Smith et al. 1993; Agapow and Burt 2001). The statistical significance of $\bar{r}_d$ was established using the program MULTILOCUS by comparing the observed value of the statistics with the distribution obtained from data sets for which alleles at each locus are resampled without replacement to simulate the effect of linkage equilibrium (Agapow and Burt 2001).

Polymorphism and Divergence

DNA polymorphism statistics, measures of divergence, and tests of neutrality based on the site frequency spectrum were computed using the libsequence C++ library (Thornton 2003). Haplotype networks were constructed using the statistical parsimony method implemented in TCS 1.21 (Clement et al. 2000). Intraspecific variation was estimated using haplotypic diversity (H_d), and two estimators of the population mutation parameter $4N_e\mu$ (where N_e is the effective population size and μ the neutral mutation rate): π, which uses the average number of pairwise differences (Tajima 1983) and θ_W, which uses the number of polymorphic sites (Watterson 1975). The standard neutral model was tested using Tajima's D (Tajima 1989), Fu and Li's D^* (Fu and Li 1993), and the HKA framework (Hudson et al. 1987). We performed coalescent simulations using MS (Hudson 2002) to test for significance of Tajima's D and Fu and Li's D^*. We used a maximum likelihood ratio method for analyzing polymorphism and divergence in a HKA framework, as implemented in the MLHKA program (Wright and Charlesworth 2004). For each model, the MCMC sampler was run for 10^5 steps, and the starting value of the divergence time parameter was determined using a standard HKA test performed in the HKA program (available from Jody Hey's website, Rutgers University). We did not use sequences from outgroups to perform HKA tests as sequences could not be obtained for all loci (supplementary table S1, Supplementary Material online). Analyses of recombination based on the four-gamete test were performed using SITES 1.1 (available from Jody Hey's website, Rutgers University). Genetic differentiation between species was estimated in DNAsp, using locus-specific F_{ST} indices calculated from average number of differences within and between species (Hudson et al. 1992). Significance was evaluated based on 1,000 permutations with the S_{nn} statistic, which measures how often the nearest neighbor of a sequence (in sequence space) is from the same species (Hudson 2000). Genetic divergence between species was estimated using the net divergence (D_a) (Nei and Li 1979).

Speciation Models

We used the computer program IMA to estimate parameters of an isolation-with-migration model (Hey and Nielsen 2004). Because IMA assumes that there is no recombination within loci and free recombination between loci, we first determined the longest region without recombination for each locus using the program IMgc (Woerner et al. 2007). This resulted in the removal of a single haplotype in *nu1942* and a single segregating site for both *nu1976* and *nu1994* (with 177 and 6 bp of flanking monomorphic sequence, respectively). We assigned bounds to prior distributions of the parameters on the basis of preliminary trial runs. We used Metropolis coupling among 30 chains with a geometric heating (increment parameters: 0.7

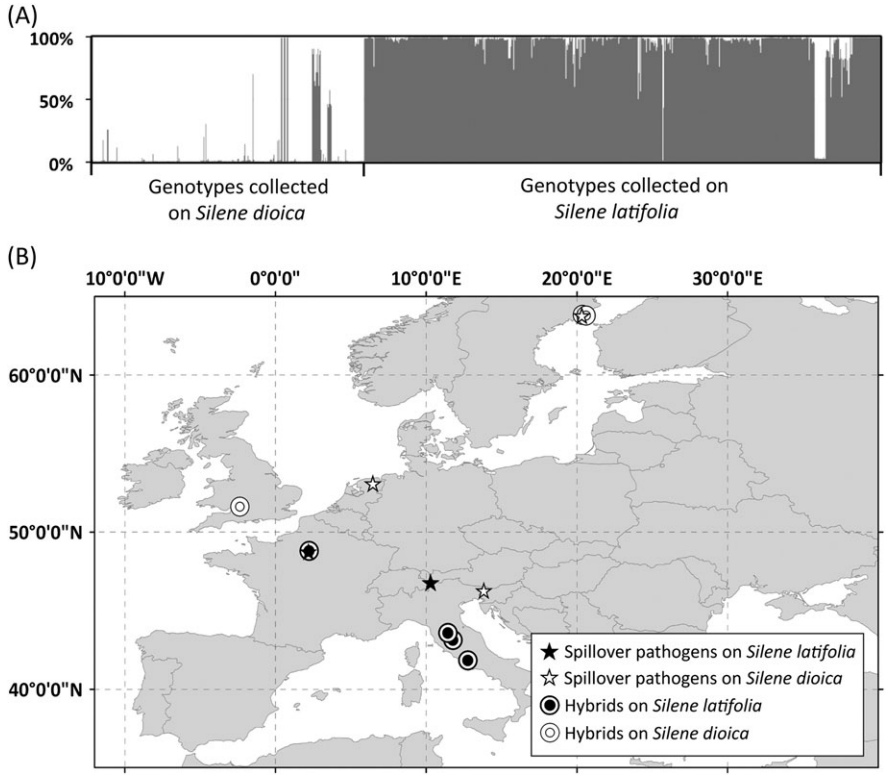

Fig. 1 Hybrids and strains resulting from cross-species disease transmission identified using the model-based Bayesian clustering algorithm implemented in the INSTRUCT program. (*A*) Membership proportions (q) in $K = 2$ clusters were inferred for 1028 multilocus microsatellite genotype of *M. lychnidis-dioica* (MvSl) and *M. silenes-dioicae* (MvSd) collected on *Silene latifolia* and *S. dioica*. Each genotype is represented by a thin bar, partitioned into segments representing membership proportions in each cluster. Hybrids and strains resulting from cross-species disease transmission (spillover strains) were identified as follows: genotypes with $q > 0.70$ in the cluster found on the other host species were identified as spillover genotypes, and genotypes with $0.3 < q < 0.70$ in the cluster found on the same host species were identified as hybrids. (*B*) Sites where genotypes were identified as hybrids and strains resulting from cross-species disease transmission are represented on a map.

for g1 and 0.95 for g2) and a burn-in period of 4×10^5 steps. We applied the infinite-site model of sequence evolution for all loci except *nuB15*, for which we used the Hasegawa–Kishino–Yano model due to the occurrence of three different states at one position. To check for convergence of the MCMC, we ran the program six times with different random seed numbers. For all runs, MCMCs mixed well, with swap rates and effective sample size values above 25% and 8×10^3, respectively. The average numbers of genealogies saved during each run was 3.5×10^4 (runs of 3.5×10^6 steps) and 6.5×10^4 (runs of 6.5×10^6 million steps). To convert parameter estimates to biologically more meaningful units, we used an averaged substitution rate of 8.8×10^{-9} per bp per year corresponding to the median of estimates for nuclear coding loci obtained by Kasuga et al. (2002). Per locus substitution rates were obtained by multiplying the averaged per-site rate by the length of each locus, and the geometric mean of per locus substitution rates was used for parameter conversion. The mutation rate is expected to be higher in the mitochondrial genome than in the nuclear genome (Lynch et al. 2006). Therefore, only nuclear loci were used in parameter conversion, making use of the estimates of the mutation rate scalars for those same loci. Because mutation rates are expected to be lower in coding regions, this rate should therefore be viewed as a minimal possible substitution rate for our set of noncoding regions. *Microbotryum* species undergo a sexual generation with every disease transmission event (Giraud, Yockteng, et al. 2008). We used a generation time of 1 year based on field data (Hood ME, unpublished data).

Results

Species Identification, Hybrids, and Cross-species Disease Transmission

Clustering analysis of 1,028 pathogen individuals using microsatellite data revealed clear partitioning of genotypes in two separate clusters corresponding to their host of origin: Genotypes found on *S. latifolia* had a high membership in one cluster (average: 0.95; standard deviation [SD]: 0.15) and genotypes found on *S. dioica* in the other (average: 0.94; SD: 0.21) (fig. 1A). This dichotomy is consistent with the previous description of *Microbotryum* populations found on *S. latifolia* and *S. dioica* as two distinct species, respectively, *M. lychnidis-dioicae* (MvSl) and *M. silenes-dioicae* (MvSd).

Cutoff values on membership proportions revealed the existence of 15 hybrid individuals in the full data set. Moreover, 32 genotypes were identified as resulting from cross-species disease transmission, that is, individuals being

Table 1. Description of Individuals Identified as Resulting from Cross-species Disease Transmission (i.e., spillover pathogens) or as Hybrids.

Host Plant/ Country	Coordinate (N)	Coordinate (E)	Description (for host plants)	Description (for pathogens)	Sample Size	Number of Spillover Strains	Number of Hybrids
Silene latifolia							
Italy	41.796561	12.752523	Site with mixed genuine and putative hybrid *S. latifolia* (light pink flowers)	Pure site	6	0	1
Italy	43.090687	11.754122	Pure site	Pure site	2	0	1
Italy	43.543972	11.42175	Pure site	Pure site	5	0	2
Switzerland	46.79888916	10.31638908	Contact site PRA in Karrenberg and Favre (2008)	Pure site	15	15	0
France	48.76252365	2.219386101	Pure site	Contact site	3	1	1
Silene dioica							
Slovenia	46.288244	13.857418	Pure site	Pure site	1	1	0
England	51.5866127	−2.372364044	Pure site	Pure site	6	0	1
Netherlands	53.1016655	6.498610973	Contact site Ng in Van Putten et al. (2005)	Contact site	12	6	0
Sweden	63.76833344	20.60861206	Pure site	Pure site	15	0	6
Sweden	63.80583191	20.35138893	Pure site	Pure site	12	9	3

assigned to the cluster corresponding to *M. lychnidis-dioicae* (MvSl) while having been collected on *S. doica* or reciprocally. Hybrids or spillover pathogens were identified in 10 different locations distributed throughout the two species ranges (fig. 1*B*; supplementary table S1, Supplementary Material online). Some, but not all, corresponded to identified contact sites between the two host plants *S. latifolia* and *S. dioica* (table 1). The counts of hybrids and spillover genotypes were not independent of the host species (χ^2 tests: $P < 0.01$ and $P < 0.05$, respectively), with a higher frequency of both categories on *S. dioica* (hybrids: 2.8%; spillover genotypes: 4.6%; $n = 351$) than on *S. latifolia* (hybrids: 0.7%; spillover genotypes: 2.4%; $n = 677$).

Population Subdivision and Mating System

In the following analyses, we used only DNA sequence data for a reduced set of 160 individuals from which spillover genotypes and hybrids were excluded. The INSTRUCT program was used to infer population subdivision and selfing rates of each species. For *M. lychnidis-dioicae* (MvSl), the major modes contained the 30 replicates for $K \leq 4$ and at least 26 replicates for $K \geq 5$. A clear East/West partition was observed with $K = 2$, and setting $K = 3$ subdivided the Eastern cluster into Southern and Northern parts (fig. 2*A*). Increasing K produced no obvious geographical pattern and introduced some heterogeneity in individual membership proportions (supplementary fig. S1, Supplementary Material online). For *M. silenes-dioicae* (MvSd), a single mode was found for all models, except $K = 3$ for which two modes were found, with 24 replicates in the major mode. A North/South partition was observed for *M. silenes-dioicae* (MvSd) at $K = 2$ (fig. 2*B*) but with less obvious correspondence between the geographic origins of samples and cluster membership than for *M. lychnidis-dioicae* (MvSl). Increasing K yielded no obvious geographical pattern and introduced some heterogeneity in individual membership proportions (supplementary fig. S2, Supplementary Material online). Based on these results, in close agreement with those of Vercken et al. (2010) based on microsatellite data, we assumed the existence of three clusters in *M. lychnidis-dioicae* (MvSl) and of two clusters in *M. silenes-dioicae* (MvSd) for all subsequent analyses taking into account population structure. Selfing rates inferred for each cluster were high in both species (fig. 2). Multilocus linkage disequilibrium, as measured by the index of association, was close to zero and not statistically significant in all clusters ($\bar{r}_d$ values: 0.067 and −0.258 in *M. lychnidis-dioicae* [MvSl]; 0.003, −0.004, and −0.026 in *M. silenes-dioicae* [MvSd]).

Polymorphism within Species

We observed a very low level of heterozygosity in autosomal sequence loci of both species ($H_0 = 0.022$ for *M. lychnidis-dioicae* [MvSl], $H_0 = 0.003$ for *M. silenes-dioicae* [MvSd]), and we therefore used a haploid setting to analyze data (i.e., for the few heterozygous loci, one allele was drawn randomly). Summaries of variation are given in table 2 (see supplementary fig. S3 for tables of polymorphisms, Supplementary Material online). Averaged over all autosomal loci, *M. lychnidis-dioicae* (MvSl) had a higher genetic variation than *M. silenes-dioicae* (MvSd) (MvSl: $H_d = 0.553$, $\pi = 0.279\%$, $\theta_W = 0.370\%$; MvSd: $H_d = 0.158$; $\pi = 0.045\%$, $\theta_W = 0.072\%$), the differences being significant for H_d and π but not θ_W (Wilcoxon signed-rank test, $P > 0.007$, at $\alpha_{max} = 0.05$ level after Bonferroni correction for multiple tests). These differences in levels of variation were also observable in intraspecific comparisons among clusters, though differences were not significant (Wilcoxon signed-rank test, $P > 0.007$, at $\alpha_{max} = 0.05$ level after Bonferroni correction for multiple tests; supplementary table S3, Supplementary Material online). Nucleotide variation at the mitochondrial locus showed the same trend as autosomal loci both for interspecific and intraspecific comparisons (table 2). Taken together, these observations are consistent with *M. lychnidis-dioicae* (MvSl) having a higher effective population size than *M. silenes-dioicae*

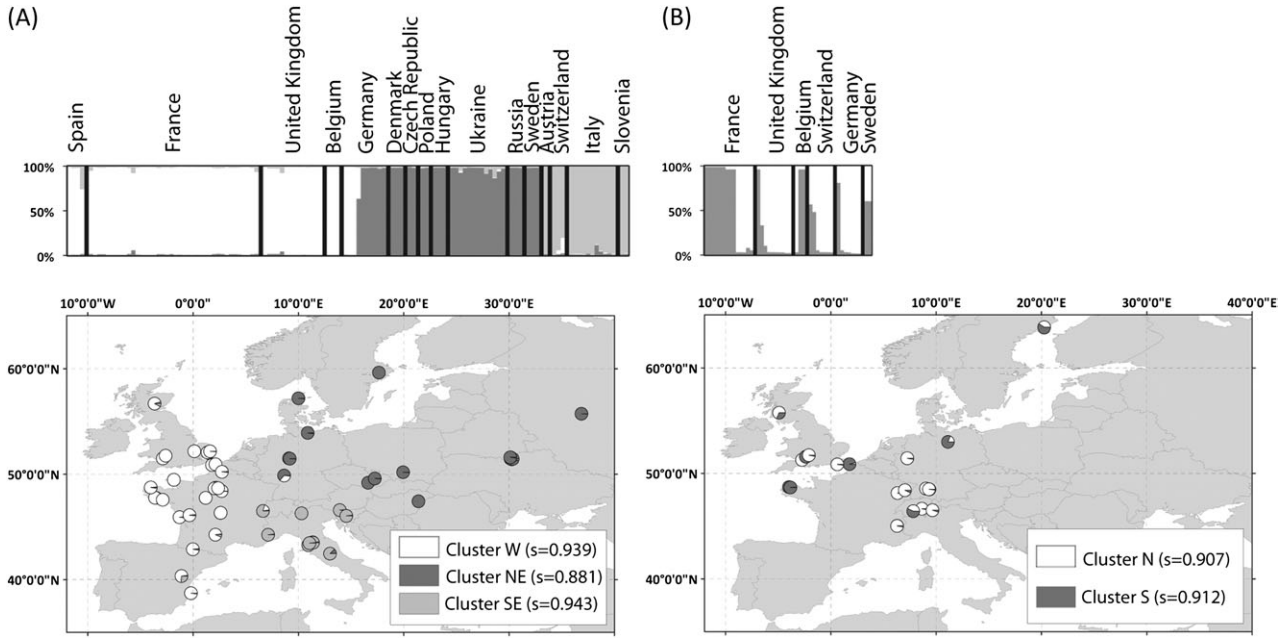

FIG. 2 Population subdivision inferred using the model-based Bayesian clustering algorithm implemented in the INSTRUCT program for 117 and 43 multilocus sequence genotypes of *Microbotryum lychnidis-dioica* (MvSl) (*A*) and *M. silenes-dioicae* (MvSd) (*B*). Each genotype is represented by a pie chart or by a thin bar, representing membership proportions in $K = 3$ and $K = 2$ clusters.

(MvSd). Only two pairs of polymorphisms incompatible with a lack of recombination were detected in the loci surveyed (four-gamete tests, Hudson and Kaplan 1985; table 2), suggesting a low level of effective recombination.

The allele frequency spectrum, as measured by Tajima's *D* (Tajima 1983) and Fu and Li's *D** (Fu and Li 1993), generally conformed to expectations under a standard neutral model of molecular evolution (table 2). Both at the species

Table 2. Summary of Nucleotide Variation within *Microbotryum lychnidis-dioicae* (MvSl) and *M. silenes-dioicae* (MvSd).

Locus	Species	n[a]	L[b]	S[c]	h[d]	H_d[e]	π[f] (%)	θ_W[f] (%)	R_m[g]	*mhits*[h]	D[i]	D^*[i]
nuB15	**MvSl**	**96**	**601**	**5**	**7**	**0.543**	**0.113**	**0.162**	**0**	**1**	**−0.947**	**−0.780**
	MvSd	**36**	**644**	**2**	**3**	**0.367**	**0.059**	**0.075**	**0**	**0**	**−0.421**	**−0.796**
nu1942	**MvSl**	**106**	**300**	**8**	**7**	**0.568**	**0.243**	**0.509**	**1**	**0**	**−1.259**	**−2.000**
	MvSd	**38**	**289**	**3**	**3**	**0.479**	**0.190**	**0.247**	**0**	**0**	**−0.510**	**−1.592**
nu1976	**MvSl**	**83**	**499**	**5**	**6**	**0.630**	**0.242**	**0.210**	**0**	**0**	**0.455**	**0.015**
	MvSd	**28**	**502**	**0**	**1**	**0.000**	**0.000**	**0.000**	**0**	**0**	**NA**[j]	**NA**[j]
nu1983	**MvSl**	**74**	**229**	**3**	**3**	**0.608**	**0.643**	**0.270**	**0**	**0**	**2.627**	**0.853**
	MvSd	**14**	**230**	**0**	**1**	**0.000**	**0.000**	**0.000**	**0**	**0**	**NA**[j]	**NA**[j]
nu1994	**MvSl**	**106**	**378**	**12**	**11**	**0.626**	**0.314**	**0.606**	**1**	**0**	**−1.270**	**−3.602***
	MvSd	**42**	**385**	**2**	**3**	**0.180**	**0.048**	**0.121**	**0**	**0**	**−1.128**	**−0.843**
nu1998	**MvSl**	**108**	**379**	**5**	**6**	**0.595**	**0.185**	**0.251**	**0**	**0**	**−0.557**	**−0.595**
	MvSd	**43**	**381**	**1**	**2**	**0.133**	**0.035**	**0.061**	**0**	**0**	**−0.042**	**0.557**
nu2008	**MvSl**	**114**	**304**	**9**	**8**	**0.696**	**0.390**	**0.563**	**0**	**0**	**−0.751**	**−0.826**
	MvSd	**38**	**321**	**1**	**2**	**0.102**	**0.032**	**0.074**	**0**	**0**	**−0.222**	**0.569**
nu2032	**MvSl**	**60**	**221**	**4**	**3**	**0.158**	**0.103**	**0.388**	**0**	**0**	**−1.606**	**−2.407**
	MvSd	**26**	**221**	**0**	**1**	**0.000**	**0.000**	**0.000**	**0**	**0**	**NA**[j]	**NA**[j]
Average of autosomal loci	**MvSl**	**93**	**364**	**6**	**6**	**0.553**	**0.279**	**0.370**	**0**	**0**		
	MvSd	**33**	**372**	**1**	**2**	**0.158**	**0.045**	**0.072**	**0**	**0**		
***mt168* (mitochondrial locus)**	**MvSl**	**103**	**534**	**2**	**2**	**0.177**	**0.066**	**0.072**	**0**	**0**	**0.119**	**0.683**
	MvSd	**39**	**534**	**2**	**2**	**0.100**	**0.037**	**0.089**	**0**	**0**	**−1.098**	**0.774**

[a] Sample size.
[b] Locus length in base pair (bp).
[c] Number of segregating sites.
[d] Number of haplotypes.
[e] Haplotypic diversity.
[f] Estimators of the population mutation parameter based on the average number of pairwise differences (π; Tajima 1983) and the number of polymorphic sites (θ_W; Watterson 1975), respectively.
[g] Minimum number of recombination events (Hudson and Kaplan 1985).
[h] Number of sites with more than two states.
[i] Tajima's *D* and Fu and Li's *D** statistics to test for the standard neutral model (Tajima 1989; Fu and Li 1993)
[j] Not applicable due to a lack of polymorphism.
*$P < 0.0125$, significant at $\alpha_{max} = 0.05$ level after Bonferroni correction for multiple tests.

Table 3. Divergence and Differentiation between *M. lychnidis-dioicae* (MvSl) and *M. silenes-dioicae* (MvSd).

Locus	D_a[a] (%)	S_f[b]	S_s[c]	S_{x1}[d]	S_{x2}[d]	F_{ST}[e]	S_{nn}[e]
Mt168	2.068	11	0	2	2	0.975	***
nuB15	0.226	0	0	6	2	0.720	***
Nu1942	2.158	3	0	8	3	0.907	***
Nu1976	0.415	1	0	5	0	0.774	***
Nu1983	0.277	0	0	3	0	0.463	***
Nu1994	0.914	0	1	11	1	0.834	***
Nu1998	0.292	1	0	5	1	0.727	***
Nu2008	1.356	3	0	9	1	0.865	***
Nu2032	0.002	0	0	4	0	0.029	ns
Average	0.856	2	0	6	1	0.699	

[a] Net divergence (Nei and Li 1979).
[b] Polymorphisms fixed between species.
[c] Polymorphisms shared between species.
[d] Polymorphisms exclusive to *M. lychnidis-dioicae* (S_{x1}) and *M. silenes-dioicae* (S_{x2}).
[e] Population differentiation, measured by F_{ST} (Hudson et al. 1992), and its significance, evaluated with the S_{nn} statistic (Hudson 2000).
***$P < 0.001$, [ns] $P > 0.05$.

scale or when intraspecific clusters were analyzed separately, a significant deviation from neutral expectation was observed in a single case (table 2; supplementary table S3, Supplementary Material online). The standard neutral model was also tested by comparing ratios of polymorphism within species with divergence between species across loci using a maximum likelihood implementation of the HKA test (Hudson et al. 1987; Wright and Charlesworth 2004). Likelihood ratio tests revealed that models with selection did not fit the data significantly better than the standard neutral model (supplementary table S4, Supplementary Material online). Altogether, our analyses do not reveal strong evidence for selection, though the reduced levels of polymorphism and divergence may have compromised the statistical power of the tests (Zhai et al. 2009).

Divergence between Species

Interspecific divergence and differentiation indexes are summarized in table 3. Net divergence, corresponding to the number of nucleotide substitutions that occurred in each species after splitting, was low between *M. lychnidis-dioicae* and *M. silenes-dioicae* (on average $D_a = 0.856\%$), which is consistent with a recent separation of the two species. However, net divergence displayed extensive variation across loci, with values ranging from 0.002% to 2.158%. The same pattern is reflected in F_{ST} estimates (average among loci: 0.699; range: 0.029–0.975).

Table 3 also shows the number of shared polymorphisms, exclusive polymorphisms, and fixed differences between *M. lychnidis-dioicae* and *M. silenes-dioicae*. Fixed differences were present between species at five loci, shared polymorphisms at one locus, and no locus showed both kinds of polymorphisms. The low numbers of fixed differences are consistent with the estimates of divergence and differentiation in suggesting that the split between the two species is recent. The low numbers of shared polymorphisms indicate that extensive gene flow has not occurred between the two species. The largest count of fixed differences was observed at the mitochondrial locus *mt168* consistent with the lower effective population size and thus faster sorting of polymorphisms among lineages of the mitochondrial genome and possibly also with a higher mutation rate relative to the nuclear genome (Lynch et al. 2006). The number of exclusive polymorphisms was significantly higher in *M. lychnidis-dioicae* (MvSl) than in *M. silenes-dioicae* (MvSd) (Wilcoxon signed-rank test, $P < 0.01$), consistent with *M. lychnidis-dioicae* (MvSl) having a higher effective population size than *M. silenes-dioicae* (MvSd).

Beyond sharing polymorphism, the two species shared entire haplotypes at four loci (fig. 3). For the two loci with the highest number of haplotypes (*nuB15*, *nu1994*), this pattern likely reflects the occurrence of contemporary interspecific gene flow as it seems unlikely that identical haplotypes would be preserved intact between species since their splitting or since early events of gene flow, even if the divergence is presumed relatively recent. For the two other loci, which have very few distinct haplotypes, the sharing of haplotypes may equally result from a recent gene flow, from a slow pace of molecular evolution, or from a short sequence length (<230 bp).

Speciation Models

Although qualitative analyses of patterns of polymorphism and divergence and clustering results suggested recent introgressive hybridization, it remained possible that historical gene flow or ancestral variation had contributed to the low level of fixed differences between *M. lychnidis-dioicae* and *M. silenes-dioicae*. Parameter estimates using an isolation-with-migration model in the IMA program were consistent with a relatively recent speciation and with low interspecific gene flow (fig. 4 and table 4; supplementary table S5, Supplementary Material online). The most likely estimate of divergence time was 4.2×10^5 years, though the marginal posterior probability curve showed a nonzero tail that prevented a reliable estimation of credibility intervals (90% highest posterior density [HPD] interval: 1.3×10^5–1.2×10^6). This situation could correspond to a case where large values of divergence time are more difficult to exclude due to the low number of shared polymorphism between species (Storchová et al. 2010). The significant population subdivision observed in each species is a clear violation of the assumptions of the isolation-with-migration model, but simulations showed that this violation has little effect on parameter estimates even for high levels of population structure (Strasburg and Rieseberg 2009). In fact, very close estimates of divergence times were obtained when analyses were restricted to interspecific pairs of clusters (not shown). A higher effective population size was estimated in *M. lychnidis-dioicae* (MvSl, $N_2 = 1.4 \times 10^5$; 90% HPD: 1.0×10^5–1.9×10^5) than in *M. silenes-dioicae* (MvSd, $N_1 = 2.9 \times 10^4$; 90% HPD: 1.5×10^4–5.3×10^4). Little information was available to estimate the ancestral population size, as illustrated by the flat curve in figure 4, although a maximal probability was observed at $N_A = 1.5 \times 10^3$ (90% HPD: 5.1×10^2–7.6×10^5). Using this estimate of the ancestral size and a value of $D_{xy} = 0.87\%$ for average

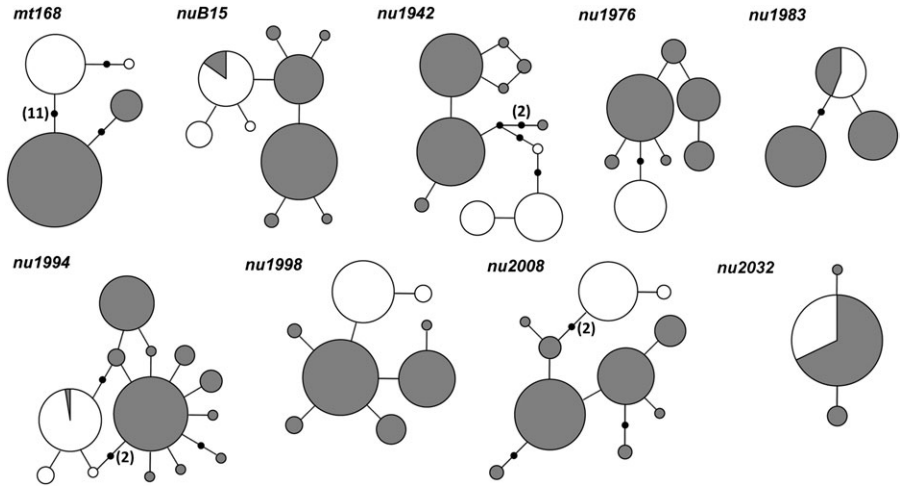

Fig. 3 Statistical parsimony networks of haplotypes of *Microbotryum lychnidis-dioica* (MvSl; gray) and *M. silenes-dioicae* (MvSd; white). Each haplotype is represented as a pie chart with surface proportional to the corresponding number of individuals.

interspecific divergence at autosomes, we estimated that *M. lychnidis-dioicae* and *M. silenes-dioicae* began to diverge approximately 4.7×10^5 years ago (using the relationship $t = (D_{xy} - 2N_A\mu)/(2\mu)$; $\mu = 8.8 \times 10^{-9}$ being the mutation rate per base pair; t the divergence time; and assuming no gene flow between species; Gillespie and Langley 1979; Takahata and Nei 1985). This value is close to the divergence time estimated using previously published sequence for the ß-tubulin, γ-tubulin, elongation factor 1α (4.4×10^5 years before present [ybp] based on data from Le Gac, Hood, Fournier, and Giraud 2007).

Estimates of gene flow revealed an asymmetric pattern, with no gene flow detected from *M. lychnidis-dioicae* (MvSl) into *M. silenes-dioicae* (MvSd) and significant, though moderate, gene flow from *M. silenes-dioicae* (MvSd) into *M. lychnidis-dioicae* (MvSl) (per generation effective number of migrant gene copies: $2Nm = 0.099$). Consistent with the overall low levels of gene flow and small differences in levels of gene flow in both directions, comparison among nested models using likelihood ratio tests (Hey and Nielsen 2007) revealed that various models with unidirectional gene flow, without gene flow, or with symmetric gene flow were a significantly better fit to the data than the full isolation-with-migration model (table 5). The likelihood ratio test was, however, only marginally significant for the model with no gene flow into *M. silenes-dioicae* (MvSd, $P = 0.019$; table 5), which is consistent with the pattern of unidirectional gene flow inferred from parameter values estimated under the full model. To gain deeper insight into the history of gene flow between the two species, the posterior densities of the numbers and times of migration events were estimated for each locus from the genealogies recorded over the course of the MCMC sampler at stationarity (Won and Hey 2005; note that here we refer to "migration" as the rate at which genes are introgressed into species). The modal number of migration events was nonzero only for migrations into *M. lychnidis-dioicae* (MvSl) and only for those loci that shared full haplotypes and

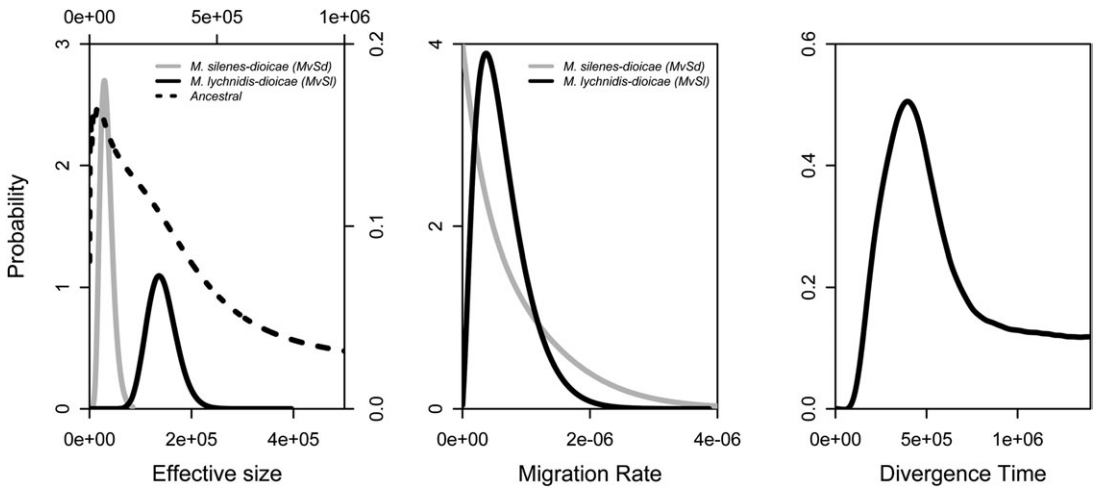

Fig. 4 Marginal posterior probability distributions of parameters of an isolation-with-migration model. Effective population size is expressed in number of individuals, migration rate in gene copies per year, and divergence time in years. For ancestral population sizes, scales are at the top and right of the graphic.

Table 4. Maximum Likelihood Estimates (MLE) and Bounds of the 90% Highest Posterior Density (HPD) Intervals for Parameters[a] of the Isolation with Migration Model.

	N_1 ($\times 10^3$)	N_2 ($\times 10^3$)	N_A ($\times 10^2$)	m_1 ($\times 10^{-6}$)	m_2 ($\times 10^{-6}$)	$2N_1m_1$	$2N_2m_2$	t ($\times 10^3$)
MLE	29	140	15	0[c]	0.35	0	0.099	419
Lower 90% HPD	15	98	5[b]	0[c]	0.06			135[b]
Higher 90% HPD	53	193	7,601[b]	1.84	1.15			1,218[b]

[a] N_1, N_2, and N_A are the effective population size of species *M. silenes-dioicae* (MvSd), *M. lychnidis-dioicae* (MvSl), and the ancestral population, respectively; m_1 is the migration rate from *M. lychnidis-dioicae* (MvSl) to *M. silenes-dioicae* (MvSd) per gene copy per generation and reciprocally for m_2; $2N_1m_1$ is the effective number of migrant gene copies per generation from *M. lychnidis-dioicae* (MvSl) to *M. silenes-dioicae* (MvSd) and reciprocally for $2N_2m_2$; t is the divergence time in years.
[b] Unreliable estimate or limit due to flat or incomplete posterior probability distribution sampled.
[c] Corresponds to the lowest bin on the parameter space and therefore represents zero.

lacked fixed differences between species (fig. 5 and table 3). This pinpointed these loci that shared full haplotypes as the cause of the signal of interspecific unidirectional gene flow and provided further support for the existence of semipermeable barriers to gene flow between the two species. Excluding these loci from the data set removed the signal of gene flow (not shown). The posterior densities of migration times suggested that introgressive hybridization at those loci was recent, as indicated by the clear peaks observed at the lowest bin of the parameter space (780 generations ago). This suggests that the rate of gene exchange may have actually been recently higher than indicated by our estimate of the population migration rate $2Nm$ that assumes constant migration over the course of the divergence process. Density curves are clearly consistent with a model in which the two species have diverged without gene flow and exchanged genes only recently, following a secondary contact.

Discussion

Analyses of nucleotide and microsatellite variation in *M. lychnidis-dioicae* and *M. silenes-dioicae* yielded insights into divergence and isolation of this pathogen species pair in nature: 1) they are well-separated species, largely confined to their respective host species; 2) cross-species disease transmission and hybrids, however, occur occasionally and are found with higher frequencies on *S. dioica* than on *S. latifolia* at contact sites for the host plants but also in sites where a single host species grows; 3) both fungal species exhibit low levels of recombination and nucleotide variation, with an approximately four times higher effective population size for *M. lychnidis-dioicae* (MvSl) than for *M. silenes-dioicae* (MvSd); 4) the two species undergo high rates of self-fertilization in natural populations (ca. 0.9 for both); 5) the two species displayed strong population subdivision, with respectively three and two clusters in *M. lychnidis-dioicae* (MvSl) and *M. silenes-dioicae* (MvSd); for more detailed phylogeographic analyses see Vercken et al. (2010); 6) the time of divergence between the two species was estimated at approximately 4.2×10^5 ybp; 7) the two species exchanged genes only very recently, probably after secondary contact and at a higher rate from *M. silenes-dioicae* (MvSd) into *M. lychnidis-dioicae* (MvSl) than in the reverse direction.

Table 5. Likelihood Ratio Test of Nested Models with Different Patterns of Gene Flow Compared with the Full Model.

Model[a]	Ln(P)[b]	2LLR (df)[c]	P[d]
$m_1\ m_2$	4.1756		
$m_1 = 0\ m_2$	2.0057	4.3398 (1)	0.019
$m_1\ m_2 = 0$	−7.9523	24.2558 (1)	<0.001
$M_1 = 0\ m_2 = 0$	−9.6329	27.6171 (2)	<0.001
$M_1 = m_2$	0.8237	6.7038 (1)	0.009

[a] m_1 is migration into *M. silenes-dioicae* (MvSd); m_2 is migration into *M. lychnidis-dioicae* (MvSl); effective population size were the same as for the full model.
[b] Value obtained from posterior density function estimated under the full model when maximized over just those parameter that were free to vary under the model considered.
[c] ln-likelihood ratio statistics calculated as the difference between the highest posterior probability for the full model and the highest posterior probability for the nested model (df is the number of degrees of freedom).
[d] Probability of likelihood ratio test, calculated from a χ^2 distribution or a χ^2 mixture distribution for those models that had a parameter fixed at the boundary of the parameter space (see Hey and Nielsen 2007).

Importance of Considering Sister Species When Inferring Phylogeographical Patterns

The joint analyses of the close sister species *M. lychnidis-dioicae* and *M. silenes-dioicae* identified hybrids and cross-species disease transmission, and these two categories of individuals were removed from the data set for analyses of population structure (Vercken et al. 2010). The inclusion of these hybrids and spillover genotypes may have led to the artifactual suggestion of additional clusters within each species. Our study highlights the importance of considering sister species able to hybridize and cross-infect their respective host species when inferring phylogeographical patterns. This may represent a frequent and important concern in phylogeographic studies of pathogens given the high number of sibling species specialized on different hosts and the observation that host specificity is often not absolute (Giraud et al. 2006).

Clustering analyses based on the sequenced loci revealed a clear phylogeographical pattern in each species. The population structure inferred here based on DNA sequences alone confirmed the results obtained by Vercken et al. (2010) based on microsatellite data: 1) for *M. lychnidis-dioicae* (MvSl), the inferred population structure was consistent with recolonization from three distinct refugia in Italian, Balkan, and Iberian peninsulas, 2) for *M. silenes-dioicae* (MvSd), results suggested possible recolonization from higher latitude refugia. Unlike microsatellite data (Vercken et al. 2010), nucleotide variation did not, however, provide support for further subdivision within each of the three *M. lychnidis-dioicae* (MvSl) clusters. This may simply be owing to a difference in the resolving power of the two kinds of markers regarding population structure.

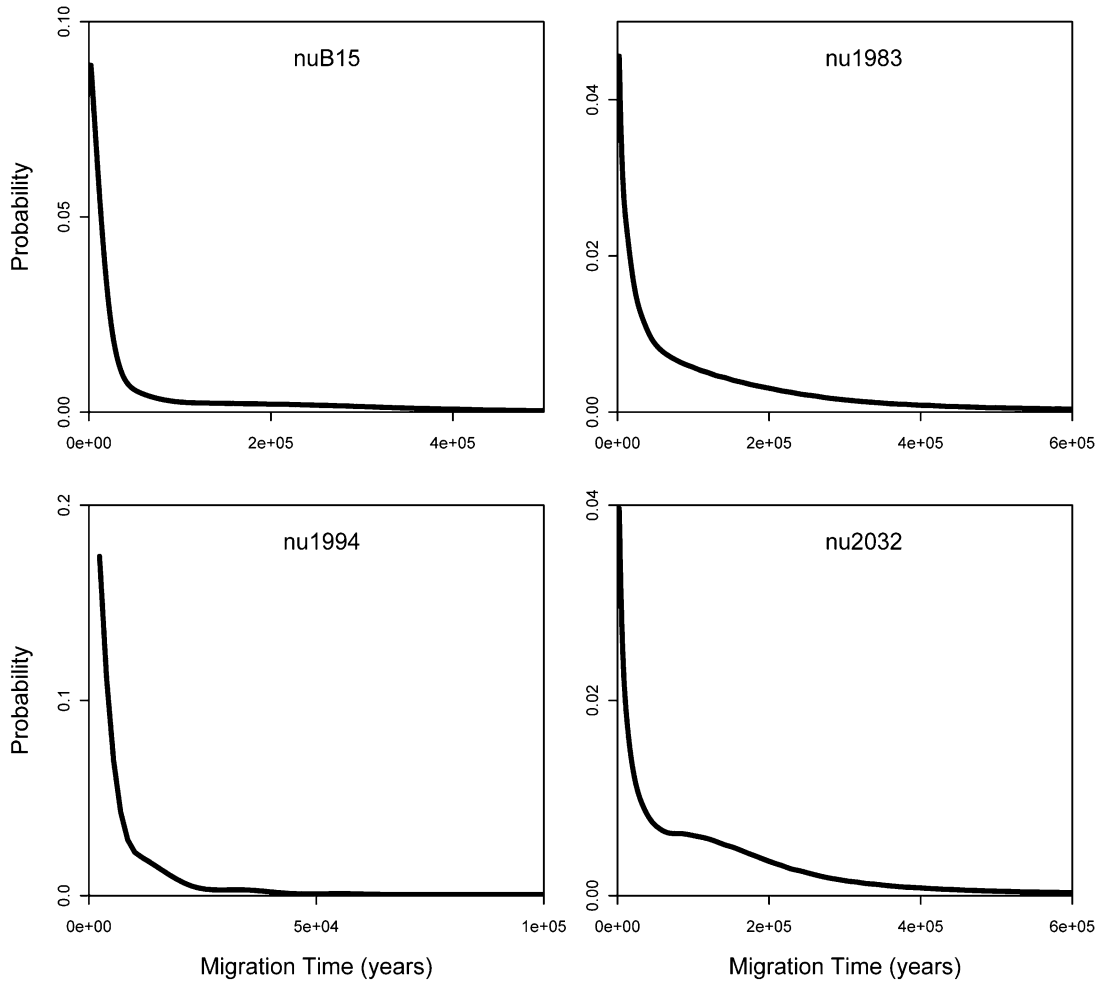

Fig. 5 Marginal posterior probability distribution of the time of migration events from *M. silenes-dioicae* (MvSd) into *M. lychnidis-dioicae* (MvSl). Results are only shown for the four loci for which numbers of migration events were nonzero.

Low Effective Population Size and Limited Recombination

Cluster-wide and species-wide nucleotide variation in the two *Microbotryum* species was low, as expected for self-fertilizing species (Charlesworth and Wright 2001). The variation measured in *M. lychnidis-dioicae* (MvSl) was indeed comparable with that of other self-fertilizing organisms, such as *Caenorhabditis elegans* (Cutter 2006), *C. briggsae* (Cutter et al. 2006), or *Arabidopsis thaliana* (Nordborg et al. 2005). Also consistent with high rates of selfing in *Microbotryum* was the detection of only two pairs of polymorphisms in the surveyed loci that were incompatible with lack of recombination. Low levels of effective recombination are characteristic of highly inbred populations (Charlesworth and Wright 2001).

Estimates of the population mutation parameter indicated an approximately four-fold lower effective population size in *M. silenes-dioicae* (MvSd) than in *M. lychnidis-dioicae* (MvSl). A previous study found a similar ratio in effective population size between the two host plant species (Filatov et al. 2001), suggesting that the same factors may be responsible for this pattern in both the plant and the pathogen. Unidirectional introgressive hybridization and higher habitat availability are other factors that may have contributed to the higher effective population size of *M. lychnidis-dioicae* (MvSl). Although *S. dioica* grows in forest margin habitats and naturally maintains small population size, *S. latifolia* is typically a ruderal or agricultural weed with large populations (Baker 1948). Assuming equal disease frequencies on both hosts, the lower effective size of the pathogen on *S. dioica* may result from the smaller number of available plants of this species. The development of agriculture may also have increased differences in effective population size by raising the surface of habitats available for *S. latifolia* and its anther-smut pathogen while causing a decline in *S. dioica* and its anther-smut pathogen through the conversion of forests into agricultural fields.

Gene Flow and Mode of Speciation

Few hybrids were detected in natural populations, confirming that *M. lychnidis-dioicae* and *M. silenes-dioicae* are two well-isolated species despite largely overlapping geographic ranges and frequent sympatry (Le Gac, Hood, Fournier, and Giraud 2007; Denchev et al. 2009). Some gene flow was, however, identified, as previously suggested (Van Putten et al. 2005). Posterior densities of migration times clearly

supported very recent gene flow events, whereas no evidence for ancient gene flow was found. The finding of cross-species disease transmission and hybrids on both host plant species and full haplotype sharing at variable loci are also consistent with ongoing gene flow between the two species. Altogether, these results support a scenario in which the two species experienced introgressive hybridization through secondary contact following initial divergence in allopatry.

The time of splitting between the two species was estimated at approximately 4.2×10^5 ybp, corresponding to the Ionian Stage (Middle Pleistocene). Results from previous studies suggest that *S. dioica* and *S. latifolia* likely diverged well before their associated pathogens (9.9×10^5 to 1.1×10^7 ybp, using lowest and highest maximum likelihood estimates of N_e from Taylor and Keller (2007) and divergence time scaled by N_e from Filatov et al. (2001)). One hypothesis for the time lag between speciation in plants and in pathogens is a host shift. A single ancestral species of the fungus may have been initially infecting a single host species and shifted to the other, for instance, during an interglacial period. Repeated retreats of *S. dioica* and *S. latifolia* in separate glacial refugia during the last hundred thousand years may have facilitated divergence in the pathogen, by allowing extended periods of allopatry. A study based on host–parasite cophylogenies in fact showed that *Microbotryum* evolved through frequent host-shifts, and preferentially to closely related hosts (Refrégier et al. 2008). An alternative hypothesis is that a single species of *Microbotryum* infected the two host species since their divergence without pathogen speciation until 4.2×10^5 ybp.

Despite largely overlapping ranges and a lack of intrinsic premating barriers between *M. lychnidis-dioicae* and *M. silenes-dioicae*, we found a low frequency of hybrids and low levels of gene flow. This suggest that a combination of high levels of self-fertilization, differences in habitats, preferences in host plant pollinators and postzygotic isolation may be sufficient to impose strong limit on interspecific gene flow since secondary contact (Giraud, Yockteng, et al. 2008). In addition, experiments of hybridization between *M. lychnidis-dioicae* and *M. silenes-dioicae* suggested that a strong selection for homozygous genotypes in F_2 and subsequent generations might prevent long-term introgression (de Vienne, Refregier, et al. 2009).

We postulate that the beginning of hybridization and introgression between the two species of the pathogen have coincided with the onset of agriculture. Previous studies on hybridization between *S. dioica* and *S. latifolia* indeed suggested that contact sites between the two plant species were associated with habitat disturbance and that human activities probably significantly increased the incidence of secondary contacts between the two *Silene* taxa (Baker 1948; Karrenberg and Favre 2008). In addition, the two species likely recolonized Europe at different times, delaying opportunities for secondary contact. *Silene latifolia* presumably spread northwards as a follower of agriculture and therefore considerably later than *S. dioica* that spread together with deciduous forests relatively early during the postglacial period (Taylor and Keller 2007; Prentice et al. 2008). The hypothesis that there would not have been ample opportunities for hybridization and introgression until recent times provide an additional explanation to the low level of gene flow inferred. Human disturbance of habitats may also have influenced the direction of interspecific gene flow. Mirroring our findings of unidirectional gene flow from *M. silenes-dioicae* (MvSd) into *M. lychnidis-dioicae* (MvSl), Minder et al. (2007) found unidirectional introgression of chloroplastic DNA haplotypes from *S. dioica* into *S. latifolia*. Asymmetrical gene flow may be associated with the fact that the two taxa of plants and pathogens mate in disturbed sites that resemble more to typical *S. latifolia* sites (Karrenberg and Favre 2008). Another possibility might be that pathogen hybrids have a lower fitness on *S. dioica* than on *S. latifolia*, which is supported by experimental studies (Le Gac, Hood, and Giraud 2007).

Concluding Remarks

Altogether, our results on the level and timing of gene flow between the anther-smut pathogens infecting *S. latifolia* and *S. dioica* are consistent with a scenario of recurrent introgressive hybridization but at a very low level and through secondary contact following initial divergence in allopatry. Such a scenario could be general among *Microbotryum* species and would be consistent with the lack of intrinsic preference for intraspecific crosses and for gradual increase in postzygotic isolation with genetic distances between the species (Le Gac, Hood, and Giraud 2007; de Vienne, Refregier, et al. 2009).

Supplementary Material

Supplementary tables S1–S5 and figures S1–S3 are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

We thank Odylle Cudelou for expert laboratory work, and Annie Tillier and Angélique Gautier for help with sequencing. T.G. acknowledges the grants ANR 06-BLAN-0201 and ANR 07-BDIV-003 (Emerfundis project), M.C.F. a postdoctoral grant from the Ile de France Région. M.E.H. acknowledges the grant NSF-DEB 0747222. We are grateful to Camille Roux for help with preliminary analyses, to Franz Lang for sharing unpublished mitochondrial sequences, and to the INRA MIGALE bioinformatics platform (http://migale.jouy.inra.fr) and Centre des Ressources Informatiques at Université Paris Sud for computational resources. Sequencing was funded by the "Consortium National de Recherche en Génomique" and the "service de systématique moléculaire" of the Muséum National d'Histoire Naturelle (CNRS IFR 101). It is part of the agreement n°2005/67 between the Genoscope and the Muséum National d'Histoire Naturelle on the project "Macrophylogeny of life" directed by Guillaume Lecointre. We are greatly indebted to all the people that helped with sample collection, and in particular Janis Antonovics. We thank anonymous

reviewers for their comments on earlier versions of the manuscript.

References

Agapow P-M, Burt A. 2001. Indices of multilocus linkage disequilibrium. *Mol Ecol Notes*. 1:101–102.

Baker HG. 1948. Stages in invasion and replacement demonstrated by species of *Melandrium*. *J Ecol*. 36:96–119.

Charlesworth D, Wright SI. 2001. Breeding systems and genome evolution. *Curr Opin Genet Dev*. 11:685–690.

Clamp M, Cuff J, Searle SM, Barton GJ. 2004. The Jalview Java alignment editor. *Bioinformatics* 20:426–427.

Clement M, Posada D, Crandall KA. 2000. TCS: a computer program to estimate gene genealogies. *Mol Ecol*. 9:1657–1659.

Cutter AD. 2006. Nucleotide polymorphism and linkage disequilibrium in wild populations of the partial selfer *Caenorhabditis elegans*. *Genetics* 172:171–184.

Cutter AD, Felix MA, Barriere A, Charlesworth D. 2006. Patterns of nucleotide polymorphism distinguish temperate and tropical wild isolates of *Caenorhabditis briggsae*. *Genetics* 173: 2021–2031.

Daszak P. 2000. Emerging infectious diseases of wildlife—threats to biodiversity and human health. *Science* 287:1756–1756.

de Vienne DM, Hood ME, Giraud T. 2009. Phylogenetic determinants of potential host shifts in fungal pathogens. *J Evol Biol*. 22:2532–2541.

de Vienne DM, Refrégier G, Hood M, Guigue A, Devier B, Vercken E, Smadja C, Deseille A, Giraud T. 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. *Mycol Balc*. 6:79–84.

Felsensein J. 1981. Skepticism towards Santa Rosalia, or why are there so few kinds of animals? *Evolution* 35:124–138.

Filatov DA, Laporte V, Vitte C, Chartesworth D. 2001. DNA diversity in sex-linked and autosomal genes of the plant species *Silene latifolia* and *Silene dioica*. *Mol Biol Evol*. 18:1442–1454.

Fu YX, Li WH. 1993. Statistical tests of neutrality of mutations. *Genetics* 133:693–709.

Gao H, Williamson S, Bustamante CD. 2007. A Markov chain Monte Carlo approach for joint inference of population structure and inbreeding rates from multilocus genotype data. *Genetics* 176:1635–1651.

Geraldes A, Basset P, Gibson B, Smith K, Harr B, Yu HT, Bulatova N, Ziv Y, Nachman M. 2008. Inferring the history of speciation in house mice from autosomal, X-linked, Y-linked and mitochondrial genes. *Mol Ecol*. 17:5349–5363.

Gillespie JH, Langley CH. 1979. Are evolutionary rates really variable? *J Mol Evol*. 13:27–34.

Giraud T. 2004. Patterns of within population dispersion and mating of the fungus *Microbotryum violaceum* parasitising the plant *Silene latifolia*. *Heredity*. 93:559–565.

Giraud T. 2006. Selection against migrant pathogens: the immigrant inviability barrier in pathogens. *Heredity*. 97:316–318.

Giraud T, Gladieux P, Gavrilets S. 2010. Linking the emergence of fungal plant diseases with ecological speciation. *Trends Ecol Evol*. 25:387–395.

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, Refregier G, de Vienne DM, Le Gac Ma, Hood ME. 2008. Speciation in fungi. *Fung Genet Biol*. 45:791–802.

Giraud T, Villaréal L, Austerlitz F, Le Gac M, Lavigne C. 2006. Importance of the life cycle in host race formation and sympatric speciation in parasites. *Phytopathology* 96:280–287.

Giraud T, Yockteng R, Lopez-Villavicencio M, Refrégier G, Hood ME. 2008. The mating system of the anther smut fungus, *Microbotryum violaceum*: selfing under heterothallism. *Eukaryot Cell*. 7:765–775.

Gladieux P, Caffier V, Devaux M, Le Cam B. 2010. Host-specific differentiation among populations of *Venturia inaequalis* causing scab on apple, pyracantha and loquat. *Fungal Genet Biol*. 47:511–521.

Goulson D, Jerrim K. 1997. Maintenance of the species boundary between *Silene dioica* and *S. latifolia* (red and white campion). *Oikos* 79:115–126.

Hey J, Nielsen R. 2004. Multilocus methods for estimating population sizes, migration rates and divergence time, with applications to the divergence of *Drosophila pseudoobscura* and *D. persimilis*. *Genetics* 167:747–760.

Hey J, Nielsen R. 2007. Integration within the Felsenstein equation for improved Markov chain Monte Carlo methods in population genetics. *Proc Natl Acad Sci U S A*. 104:2785–2790.

Hey J, Won Y-J, Sivasundar A, Nielsen R, Markert J. 2004. Using nuclear haplotypes with microsatellites to study gene flow between recently separated Cichlid species. *Mol Ecol*. 13: 909–919.

Hood ME. 2005. Repetitive DNA in the automictic fungus *Microbotryum violaceum*. *Genetica* 124:1–10.

Hood ME, Antonovics J. 2000. Intratetrad mating, heterozygosity, and the maintenance of deleterious alleles in *Microbotryum violaceum* (=*Ustilago violacea*). *Heredity* 85:231–241.

Hood ME, Mena-Ali JI, Gibson AK, et al. (11 co-authors). 2010. Distribution of the anther-smut pathogen *Microbotryum* on species of the Caryophyllaceae. *New Phytol*. 187:217–229.

Hudson RR. 2000. A new statistic for detecting genetic differentiation. *Genetics* 155:2011–2014.

Hudson RR. 2002. Generating samples under a Wright–Fisher neutral model of genetic variation. *Bioinformatics* 18:337–338.

Hudson RR, Boos DD, Kaplan NL. 1992. A statistical test for detecting geographic subdivision. *Mol Biol Evol*. 9:138–151.

Hudson RR, Kaplan N. 1985. Inferring the number of recombination events in the history of a sample. *Biometrics* 41: 572–572.

Hudson RR, Kreitman M, Aguade M. 1987. A test of neutral molecular evolution based on nucleotide data. *Genetics*. 116:153–159.

Jakobsson M, Rosenberg NA. 2007. CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics* 23:1801–1806.

Karrenberg S, Favre A. 2008. Genetic and ecological differentiation in the hybridizing campions *Silene dioica* and *S. latifolia*. *Evolution* 62:763–773.

Kasuga T, White TJ, Taylor JW. 2002. Estimation of nucleotide substitution rates in eurotiomycete fungi. *Mol Biol Evol*. 19:2318–2324.

Katoh K, Kuma K, Toh H, Miyata T. 2005. MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic Acids Res*. 33:511–518.

Kemler M, Goker M, Oberwinkler F, Begerow D. 2006. Implications of molecular characters for the phylogeny of the Microbotryaceae (Basidiomycota: Urediniomycetes). *BMC Evol Biol*. 6:35.

Kliman RM, Andolfatto P, Coyne JA, Depaulis F, Kreitman M, Berry AJ, McCarter J, Wakeley J, Hey J. 2000. The population genetics of the origin and divergence of the *Drosophila simulans* complex species. *Genetics* 156:1913–1931.

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.

Le Gac M, Hood ME, Giraud T. 2007. Evolution of reproductive isolation within a parasitic fungal complex. *Evolution* 61:178–1–1787.

Lutz M, Goker M, Piatek M, Kemler M, Begerow D, Oberwinkler F. 2005. Anther smuts of Caryophyllaceae: molecular characters indicate host-dependent species delimitation. *Mycol Prog*. 4:225–238.

Lynch M, Koskella B, Schaack S. 2006. Mutation pressure and the evolution of organelle genomic architecture. *Science* 311:1727–1730.

Maynard-Smith J, Smith NH, O'Rourke M, Spratt BG. 1993. How clonal are bacteria? *Proc Natl Acad Sci U S A*. 90:4384–4388.

Minder AM, Rothenbuehler C, Widmer A. 2007. Genetic structure of hybrid zones between *Silene latifolia* and *Silene dioica* (Caryophyllaceae): evidence for introgressive hybridization. *Mol Ecol*. 16:2504–2516.

Nadachowska K, Babik W. 2009. Divergence in the face of gene flow: the case of two newts (Amphibia: Salamandridae). *Mol Biol Evol*. 26:829–841.

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.

Nielsen R, Wakeley J. 2001. Distinguishing migration from isolation: a Markov Chain Monte Carlo approach. *Genetics* 158:885–896.

Nordborg M, Hu TT, Ishino Y, et al. (24 co-authors). 2005. The pattern of polymorphism in Arabidopsis thaliana. *PLoS Biol*. 3:1289–1299.

Prentice HC, Malm JU, Hathaway L. 2008. Chloroplast DNA variation in the European herb *Silene dioica* (red campion): postglacial migration and interspecific introgression. *Plant Syst Evol*. 272:23–37.

Refrégier G, Le Gac M, Hood ME, Jonot O, Giraud T. Forthcoming. 2010. No evidence of reproductive character displacement between two sister fungal species causing anther smut disease in *Silene Int J Plant Sci*. 171:847–859.

Refrégier G, Le Gac M, Jabbour F, Widmer A, Hood ME, Yockteng R, Shykoff JA, Giraud T. 2008. Cophylogeny of the anther smut fungi and their Caryophyllaceous hosts: prevalence of host shifts and importance of delimiting parasite species. *BMC Evol Biol*. 8:100.

Rice WR. 1984. Disruptive selection on habitat preference and the evolution of reproductive isolation: a simulation study. *Evolution* 38:1251–1260.

Städler T, Roselius K, Stephan W. 2005. Genealogical footprints of speciation processes in wild tomatoes: demography and evidence for historical gene flow. *Evolution* 59:1268–1279.

Storchová R, Reif J, Nachman MW. 2010. Female heterogamety and speciation: reduced introgression of the Z chromosome between two species of nightingales. *Evolution* 64:456–471.

Strasburg JL, Rieseberg LH. 2009. How robust are "isolation with migration" analyses to violations of the IM model? A simulation study. *Mol Biol Evol*. 27:297–310.

Stukenbrock EH, Banke S, Javan-Nikkhah M, McDonald BA. 2007. Origin and domestication of the fungal wheat pathogen *Mycosphaerella graminicola* via sympatric speciation. *Mol Biol Evol*. 24:398–411.

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.

Takahata N, Nei M. 1985. Gene genealogy and variance of interpopulational nucleotide differences. *Genetics* 110:325–344.

Taylor DR, Keller SR. 2007. Historical range expansion determines the phylogenetic diversity introduced during contemporary species invasion. *Evolution* 61:334–345.

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

Thrall PH, Biere A, Antonovics J. 1993. Plant life-history and disease susceptibility the occurrence of *Ustilago violacea* on different species within the Caryophyllaceae. *J Ecol*. 81:489–498.

Van Putten WF, Biere A, Van Damme JMM. 2003. Intraspecific competition and mating between fungal strains of the anther smut *Microbotryum violaceum* from the host plants *Silene latifolia* and *S.dioica*. *Evolution* 57:766–776.

Van Putten WF, Biere A, Van Damme JMM. 2005. Host related genetic differentiation in the anther smut fungus *Microbotryum violaceum* in sympatric, parapatric and allopatric populations of two host species *Silene latifolia* and *S. dioica*. *J Evol Biol*. 18:203–212.

Van Putten WF, Elzinga JA, Biere A. 2007. Host fidelity of the pollinator guilds of *Silene dioica* and *Silene latifolia*: possible consequences for sympatric host race differentiation of a vectored plant disease. *Int J Plant Sci*. 168:421–434.

Vercken E, Fontaine MC, Gladieux P, Hood ME, Jonot O, Giraud T. Forthcoming 2010. Glacial refugia in pathogens: European genetic structure of anther smut pathogens on *Silene latifolia and Silene dioica*. *PLoS Pathog*.

Watterson GA. 1975. Number of segregating sites in genetic models without recombination. *Theor Popul Biol*. 7:256–276.

Woerner AE, Cox MP, Hammer MF. 2007. Recombination-filtered genomic datasets by information maximization. *Bioinformatics* 23:1851–1853.

Won YJ, Hey J. 2005. Divergence population genetics of chimpanzees. *Mol Biol Evol*. 22:297–307.

Wright SI, Charlesworth B. 2004. The HKA test revisited: a maximum-likelihood-ratio test of the standard neutral model. *Genetics* 168:1071–1076.

Zhai WW, Nielsen R, Slatkin M. 2009. An investigation of the statistical power of neutrality tests based on comparative and population genetic data. *Mol Biol Evol*. 26:273–283.

Zhou R, Zeng K, Wu W, Chen X, Yang Z, Shi S, Wu CI. 2007. Population genetics of speciation in nonmodel organisms: I. Ancestral polymorphism in mangroves. *Mol Biol Evol*. 24: 2746–2754.