Open access • Posted Content • DOI:10.1101/2020.01.04.894774

Congruent population genetic structures and divergence histories in anther-smut fungi and their host plants Silene italica and the S. nutans species complex

Fanny E. Hartmann, Alodie Snirc, Amandine Cornille, Cécile Godé ...+6 more authors

Institutions: Université Paris-Saclay, university of lille, Botanic Garden Meise, SupAgro

Published on: 06 Jan 2020 - bioRxiv (Cold Spring Harbor Laboratory)

Topics: Genetic structure, Silene nutans, Population, Microbotryum and Species complex

Congruent population genetic structures and divergence histories in anther-smut fungi and their host plants Silene italica and the S. nutans species complex

Fanny Hartmann, Alodie Snirc, Amandine Cornille, Cécile Godé, Pascal Touzet, Fabienne van Rossum, Elisabeth Fournier, Stéphanie Prieur, Jacqui Shykoff, Tatiana Giraud

Congruent population genetic structures and divergence histories in anther-smut fungi and their host plants *Silene italica* and the *S. nutans* species complex

Running title:

Co-structure of anther-smuts and their hosts

Fanny E. Hartmann[1], Alodie Snirc[1], Amandine Cornille[2], Cécile Godé[3], Pascal Touzet[3], Fabienne Van Rossum[4,5], Elisabeth Fournier[6], Stéphanie Le Prieur[1], Jacqui Shykoff[1], Tatiana Giraud[1]

[1] Ecologie Systématique Evolution, Bâtiment 360, Univ. Paris-Sud, AgroParisTech, CNRS, Université Paris-Saclay, 91400 Orsay, France;

[2] Génétique Quantitative et Evolution–Le Moulon, INRA; Univ. Paris-Sud, CNRS, AgroParisTech, Université Paris-Saclay, 91198 Gif-sur-Yvette, France;

[3] Univ. Lille, CNRS, UMR 8198 - Evo-Eco-Paleo, F-59000 Lille, France;

[4] Meise Botanic Garden, B-1860 Meise, Belgium;

[5] Fédération Wallonie–Bruxelles, B-1080 Brussels, Belgium;

[6] UMR BGPI, Univ. Montpellier, INRA, CIRAD, Montpellier SupAgro, Montpellier, France;

[*] Author for Correspondence: Fanny E. Hartmann, Ecologie Systematique Evolution, Bâtiment 360, Univ. Paris-Sud, AgroParisTech, CNRS, Université Paris-Saclay, 91400 Orsay, France, phone number : 01 69 15 72 82, e-mail address: fanny.hartmann@u-psud.fr

Abstract

The study of population genetic structure congruence between hosts and pathogens gives important insights into their shared phylogeographic and coevolutionary histories. We studied the population genetic structure of castrating anther-smut fungi (*Microbotryum* genus) and of their host plants, the *Silene nutans* species complex, and the morphologically and genetically close *S. italica,* which can be found in sympatry. Phylogeographic population genetic structure related to persistence in separate glacial refugia has been recently revealed in the *S. nutans* plant species complex across Western Europe, identifying several distinct lineages. We genotyped 171 associated plant-pathogen pairs of anther-smut fungi and their host plant individuals using microsatellite markers and plant chloroplastic SNPs. We found clear differentiation between fungal populations parasitizing *S. nutans* and *S. italica* plants. The population genetic structure of fungal strains parasitizing the *S. nutans* plant species complex mirrored the host plant genetic structure, suggesting that the pathogen was isolated in glacial refugia together with its host and/or that it has specialized on the plant genetic lineages. Using random forest approximate Bayesian computation (ABC-RF), we found that the divergence history of the fungal lineages on *S. nutans* was congruent with the one previously inferred for the host plant and likely occurred with ancient but no recent gene flow. Genome sequences confirmed the genetic structure and the absence of recent gene flow between fungal genetic lineages. Our analyses of host-pathogen individual pairs contribute to a better understanding of co-evolutionary histories between hosts and pathogens in natural ecosystems, in which such studies are still scarce.

Key words: population genetic structure, host-pathogen interaction, cryptic speciation, genetic divergence, approximate Bayesian computation, coevolution

Introduction

Host-pathogen interactions are pervasive in natural ecosystems, with many important ecological and evolutionary consequences (Poulin, 2005; Thompson, 2005). Pairs of tightly interacting hosts and pathogens may share common evolutionary histories due to co-evolution and/or shared geographic and climatic constraints, or may follow distinct evolutionary trajectories due to differences in reproductive systems, dispersal ranges, population sizes or contingent histories (Tellier, de Vienne, Giraud, Hood, & Refrégier, 2010; Thompson, 2005). The comparison of genetic divergence histories between host and pathogen populations gives important insights into their shared phylogeographic history and possible local adaptation or host specialization (Croll & Laine, 2016; Feurtey et al., 2016). Studies of phylogeny congruence of hosts and pathogens can for example allow identifying host shifts or co-speciation events (de Vienne et al., 2013; Hafner & Page, 1995; Wilson, Falush, & McVean, 2005). Several cases of host shift or co-speciation events between hosts and pathogens have led to damaging diseases in plants, animals and humans (Fisher, Gow, & Gurr, 2016; McDonald & Stukenbrock, 2016; Wolfe, Dunavan, & Diamond, 2007). The study of population genetic structure congruence within species or species complexes can help identifying and understanding patterns of local adaptation (Gandon, Capowiez, Dubois, Michalakis, & Olivieri, 1996), which play an important role in the dynamics of pathogen and host communities (Gandon & Michalakis, 2002; Laine, 2005; Laine, 2008). Studies investigating the congruence of population genetic structure and divergence histories between hosts and pathogens are yet surprisingly still scarce despite their importance for understanding the evolutionary mechanisms and histories leading to host specialization and local adaptation (Barrett, Thrall, Burdon, & Linde, 2008; Croll & Laine, 2016), but see (Dybdahl & Lively, 1996; Feurtey et al., 2016; McCoy, Boulinier, & Tirard, 2005; Michalakis, Sheppard, Noel, & Olivieri, 1993; Tsai &

Manos, 2010).

Anther-smut fungi (*Microbotryum* genus) are generally highly specialized on their host plant species of the Caryophyllaceae family (Hartmann et al., 2019; Kemler, Göker, Oberwinkler, & Begerow, 2006; Le Gac, Hood, Fournier, & Giraud, 2007; Refrégier et al., 2008) and patterns of local adaptation have been reported in the system (Feurtey et al., 2016; Kaltz, Gandon, Michalakis, & Shykoff, 1999). Therefore, they constitute a highly suitable system to study the congruence of population genetic structure and divergence history between hosts and pathogens at different evolutionary scales. Comparisons of phylogenies at the genus level suggested a prevalence of host shifts at large evolutionary scales (Refrégier et al., 2008). The plant *Silene latifolia* and its anther-smut fungus *M. lychnidis-dioicae* display strong congruence of population genetic structures and plant local adaptation at regional and continental scales, suggesting the existence of co-evolution in the system (Delmotte, Bucheli, & Shykoff, 1999; Feurtey et al., 2016; Kaltz et al., 1999). Population genetic structures of both the host and the fungal pathogen likely resulted from past climatic events, showing hallmarks of recolonization from former glacial refugia in Europe (Badouin et al., 2017; Gladieux, Devier, Aguileta, Cruaud, & Giraud, 2013; Gladieux et al., 2011; Taylor & Keller, 2007; Vercken et al., 2010). The congruence of host and pathogen population genetic structures has not been investigated in other anther-smut fungi-*Silene* pairs despite their importance as models of pathosystems in natural ecosystems (Bernasconi et al., 2009; Toh & Perlin, 2016) and the importance of assessing whether congruence in population genetic subdivision is a general pattern (Croll & Laine, 2016). Furthermore, the evolutionary history of population genetic divergence and of gene flow levels occurring between pathogens infecting closely related hosts with overlapping distribution ranges still remain poorly studied at the genome-wide-scale (but see Badouin et al., 2017).

Silene nutans is an assemblage of strongly differentiated cryptic genetic plant lineages, corresponding to at least seven "evolutionary significant units", which can be grouped into two main phylogeographic (eastern and western) genetic clusters in Europe (Van Rossum et al., 2018). The *S. nutans* genetic lineages display several morphological and ecological differences, as well as nuclear and plastid genetic differentiation (Martin et al., 2016; Van Rossum et al., 2018), and strong postzygotic reproductive isolation has been shown between the eastern and western clusters in Western Europe (Martin et al., 2017). Geographic distribution patterns suggest that *S. nutans* lineages have diverged in allopatry during the Quaternary climate oscillations, and then recolonized northwards, without admixture (Martin et al., 2016; Martin et al., 2017; Van Rossum et al., 2018). The anther-smut fungal species *Microbotryum violaceum sensu stricto* is specialized on *S. nutans* (Kemler et al., 2006; Lutz et al., 2005). *Microbotryum* fungi are pollinator-borne pathogens castrating plants by replacing the pollen by their own spores and aborting ovaries, and they are usually highly specific on their *Silene* host plant species (Le Gac, Hood, & Giraud, 2007; Refrégier et al., 2008). The existence of cryptic lineages within *S. nutans* therefore raises the question of whether cryptic lineages also exist in the fungus and whether their genetic divergence history mirrors that of the host. In Southern Europe, *S. nutans* can be found in sympatry with *S italica*, a closely related species, in particular in the Cévennes (Lafuma & Maurice, 2006). *Silene nutans* and *S. italica* can be difficult to distinguish, being very similar morphologically, mainly differing in petal shape, in the length of the gynophore and with rather nodding or erect flowers, respectively (Rameau, Mansion, & Dumé, 1989, 2008; Tison & de Foucault, 2014). Their flowering time, pollinator guilds and ecological niches overlap, although they slightly differ in their ecological requirements, with *S. italica* being strictly calcicolous, more xero-thermophilous and heliophilous than *S. nutans* (Rameau et al., 1989, 2008; Tutin et al., 2001). *Microbotryum* fungi have been found on the two plant species and appeared differentiated

based on a small sample and a few genetic markers (Bucheli, Gautschi, & Shykoff, 2000); however, spillover, i.e., non-sustainable cross-species disease transmission, has been shown to occur in *Microbotryum* fungi (Antonovics, Hood, & Partain, 2002; Gladieux et al., 2011). The study of the population genetic structure of *M. violaceum s. s.* on *S. nutans* therefore requires molecular typing to check plant and fungal species identities.

In this study, we analyzed associated plant-pathogen samples of *S. nutans* and *S. italica* populations from Europe in order to address the following questions: 1) Is there genetic differentiation between anther-smut fungi parasitizing the closely related plant species *S. nutans* and *S. italica*? Are there hybrids and/or spill-overs in anther-smut fungi? 2) Is the population genetic structure of the anther-smut fungi parasitizing *S. nutans* and *S. italica* congruent with those of their hosts? 3) What is the divergence history of these anther-smut fungi on their hosts? Did genetic divergence occur with gene flow? Does the divergence history of the fungi mirror that of their hosts? In order to address these questions, we used genetic markers (nuclear microsatellite markers and/or chloroplastic SNPs) in both plants and anther-smut fungi to analyse diseased material (171 plant-pathogen pairs) collected across Western Europe and infer population genetic structures. We used approximate Bayesian computation (ABC) to compare genetic divergence scenarios in anther-smut fungi in order to assess whether the genetic divergence history in anther-smut fungi mirrored that inferred previously for their host plants. We also sequenced 53 genomes of *Microbotryum* fungi and analysed them together with 46 available *Microbotryum* genomes to assess whether the inferences on population genetic structure and gene flow based on microsatellite markers hold at the genome-wide levels and whether there were spill-overs or gene flow from other sympatric *Microbotryum* species than those analysed with microsatellite markers.

140

141 # Materials and Methods

142 ## Fungal and plant materials

143 We analyzed 171 anther-smut fungi collected from diseased plant individuals of *S. nutans* and *S.*
144 *italica* that were sampled in 55 distinct geographic sites across Europe, including the United
145 Kingdom, France, Belgium, Switzerland, Italy, Germany and Norway (Table S1, Fig. 1A).
146 Diseased plant individuals were stored in individual paper envelops kept in plastic bags filled
147 with silica gel, in dark conditions at 8°C. Small pieces of leaves were used as raw material for
148 DNA extraction of host plant individuals. Diploid spores of *Microbotryum* anther-smut fungi
149 were collected in buds of infected flowers and grown as yeasts on potato dextrose agar (PDA)
150 medium, and then stored at -20°C. Spores did not grow for a few samples older than two years.
151 For DNA extraction of these strains, we used the dried spores directly collected from anthers of
152 diseased flowers. In most cases, spores from anthers of a single flower were used and therefore
153 corresponded to a single *Microbotryum* genotype (López-Villavicencio et al., 2007). When spore
154 material was limited, we used multiple anthers from several flowers of the same plant. We never
155 observed more than two alleles per strain and found low levels of heterozygosity per strain,
156 indicating that we unlikely genotyped more than one fungal genotype.

157

158 ## Genotyping and identification of species and genetic lineages of the *Silene* plants based on
159 nuclear microsatellite markers and chloroplastic SNPs

160 We extracted plant DNA of at least one diseased plant individual per site using the NucleoSpin®
161 96 Plant II kit (Macherey-Nagel, Germany). We obtained DNA for 134 out of the 171 diseased

plant individuals. At least one host individual was genotyped per site, except for six sites due to the lack of plant material (Table S1). In addition, we genotyped three plant individuals for which no anther-smut fungus was isolated but strains were isolated at the same sites (Table S1). Using morphological criteria of the host plant species, sample collectors initially identified 153 *S. nutans* and 18 *S. italica* diseased individuals. However, as the morphologies of *S. nutans* and *S. italica* are very similar, misidentification could occur. Therefore, we sequenced four chloroplast fragments (*psbA*, *LF*, *MATK*, *GS*; Lahiani et al., 2013) to check species identity of the diseased plants collected in southeastern France and in Italy, where both *S. nutans* and *S. italica* can be found. We thus reassigned one individual to *S. nutans* and seven individuals to *S. italica,* while one host individual could not be assigned to either *S. nutans* or *S. italica* species (from the pair # 1436). We excluded this host-pathogen individual pair from our analyses. We thus had 145 diseased *S. nutans* plants and 25 diseased *S. italica* plant (Table S1).

To further genotype the plant individuals for studying the population genetic structure, we used a combination of plastid (chloroplast) SNPs and nuclear microsatellite markers as previously described (Godé et al., 2014; Martin et al., 2016; Martin et al., 2017). For plastid markers, using the KASPAR® protocol we genotyped six SNPs, named Cp42, Cp397, Cp540, Cp656, Cp730 and Cp804, and polymorphic for [T/G], [A/C], [C/T], [G/T], [C/T], and [T/G], respectively. Individual haplotypes were defined as combinations of allelic states for all six SNPs. For nuclear markers, we used six multiplexes genotyping 24 microsatellite markers (Table S2). We followed the previously published protocols (Godé et al., 2014; Martin et al., 2016; Martin et al., 2017), except that we used different dye colors, bought from Eurofins Genomics. We used the Multiplex PCR Kit (Qiagen) following manufacturer instructions for PCR reactions, performed separately for each multiplex in 15 µL volume containing 3 µL of DNA, 3.4 µL of H2O, 7.1 µL of multiplex

PCR Kit (Buffer 2X, Qiagen, USA), and 1.4 µL of the primer mix. The primer mix included 2 µM of unlabelled forward and reverse primers and 0.5 µM or 0.75 µM of the labelled forward primer depending on the dye label. We used the same PCR cycling program as described in (Godé et al., 2014) with some modifications of the final elongation for multiplex 1 (Table S3A-C). We checked successful PCR amplifications on 2% (w/v) agarose gel electrophoresis. We outsourced genotyping at the Gentyane Genotyping Platform (INRA, Clermont, France) and scored alleles with GENEMAPPER v.4.0 (Applied Biosystems). We excluded three markers (SIL18, SIL26, SIL42) for which we had less than 50% of the individuals successfully genotyped. Given the material available, we could genotype 136 host plant individuals using both plastid SNPs and nuclear microsatellite markers. All 136 genotyped host individuals had a determined plastid haplotype and a determined genotype for at least 50% of the 21 remaining nuclear microsatellite markers. The genotypes are deposited on Dryad (accession available upon acceptance).

Genotyping and species identification of the *Microbotryum* fungal strains: microsatellite markers and ITS sequences

We extracted fungal DNA using the Chelex protocol (Biorad, USA) following (Giraud, 2004). We extracted DNA from the 170 strains of anther-smut fungi, either from mixes, stored at -20°C, of haploid sporidia resulting from clonal growth after meiosis or from diploid teliospores directly collected in anthers of diseased flowers and stored at 10°C. Diploid genomes were therefore genotyped. DNA was diluted half-fold for PCR amplification. We used the internal transcribed spacer (ITS) to check that the genus of the fungal strains belonged to the *Microbotryum* genus. To genotype the fungal strains, we used 22 microsatellite markers arranged into multiplex (Table S4 ; (Fortuna et al., 2016; Giraud et al., 2008)). We used the Multiplex PCR Kit (Qiagen)

following manufacturer instructions for PCR reactions which were performed separately for each multiplex as described above. We used the PCR cycling programs as in previous studies (Fortuna et al., 2016; Giraud et al., 2008; Table S3C-D). We checked the success of PCR amplifications on 2 % agarose gel electrophoresis. For genotyping, we pooled multiplexes 7 and 8 as a single multiplex. We outsourced genotyping at the Gentyane Genotyping Platform (INRA, Clermont, France) and scored alleles with GENEMAPPER v.4.0 (Applied Biosystems). We could determine the genotypes for all strains for at least 50% of the 22 markers. We identified 21 fungal strains that were likely siblings of other fungal strains based on null genetic distances and we excluded them from the STRUCTURE software analyses (see below).

Population genetic structure based on microsatellite markers

To analyze and compare host and pathogen population genetic structures, we used a combination of three complementary approaches using microsatellite nuclear markers on both the *Silene* host plants and anther-smut fungal datasets. First, we used the model-based Bayesian clustering approach implemented in the software STRUCTURE version 2.3.4 (Pritchard, Stephens, & Donnelly, 2000). The program performs partitions of multilocus genotypes into genetic clusters and assigns individuals to genetic clusters, minimizing the departure from expected frequencies and linkage equilibrium among loci. We tested an admixture model with correlated frequencies and no prior information for K = 2 to K = 10 clusters. A total of 10 repetitions were run for each K value. We used 50,000 samples as a burn-in period and 100,000 samples per run for the Monte Carlo Markov Chain (MCMC) replicates. Cluster assignment probabilities were computed using the CLUMPP program (Jakobsson & Rosenberg, 2007) implemented in the R package {Pophelper}. We used the R package {Pophelper} (https://github.com/royfrancis/pophelper) to build the barplots. We choose as the biologically most relevant K value the finer population

234 structure, as the highest K value for which a new cluster could be identified with individuals
235 highly assigned to it, the new cluster at {K+1} having only admixed individuals (i.e. mean
236 membership coefficient <0.80 to the given cluster). For the anther-smut fungus dataset, genetic
237 data were haploidized as individuals were highly homozygous. High homozygosity levels might
238 bias inferences in the software STRUCTURE as these are based on Hardy-Weinberg expectations
239 in a diploid setting (Pritchard et al., 2000). Furthermore, 21 strains were removed, appearing as
240 siblings of other strains, with identical genotypes. *Microbotryum* fungi have one obligatory
241 sexual event before plant infection, so that clonemates cannot be found in different plants, but
242 high selfing rates may allow the same genotype to be found in neighbor plants if they are
243 parasitized by the same diploid spores or offspring (Giraud, 2004). For the host plant dataset,
244 genetic data were kept as diploid and five individuals were removed as they had identical
245 genotypes as other neighbor plants and may be clonemates. We identified clonemates and
246 siblings in both datasets using the dist(X) function in the R package {Ape} (Paradis, Claude, &
247 Strimmer, 2004) and considered two individuals to be siblings if their distance was equal to 0.
248 We then used two methods for assessing population genetic structure that do not assume
249 outcrossing or a lack of linkage disequilibrium. We performed a discriminant analysis of
250 principal components (DAPC) using the R package {ADEGENET} (Jombart, 2008; Jombart &
251 Ahmed, 2011) and used a principal component analysis (PCA) with the dudi.pca function using
252 the R package {ade4} (Dray & Dufour, 2007) on the entire set of individuals in both datasets.
253 Maps showing genotypes per locality were drawn using the R package {maps} (Becker, Wilks,
254 Brownrigg, Minka, & Deckmyn, 2017) and {mapplots} (Gerritsen, 2013). Scatter plots were
255 performed using the R package {ggplot2} (Wickham, 2009).
256
257 **Population statistics of genetic diversity and structure based on microsatellite markers**

We computed, using the R package {diveRsity} (Keenan, McGinnity, Cross, Crozier, & Prodöhl, 2016), the following estimates of genetic diversity per locus, site, genetic cluster identified in the STRUCTURE analysis and/or species for the host plants and for the anther-smut fungi datasets: the number of alleles, allelic richness (A_R), observed heterozygosity (H_O), expected heterozygosity (H_E), the fixation indexes (F_{ST}, F_{IS} and F_{IT}), and the Jost's D statistics corresponding to the fraction of allelic variation found among genetic clusters. We tested whether genotype frequencies fitted the Hardy-Weinberg expectations using a standard Chi² goodness of fit method and assessed the significance of F_{ST}, F_{IS} and F_{IT} values using a bootstrap procedure with 1000 iterations and calculating 95% confidence intervals. Sites with less than three sampled individuals were pooled with other sites of the same genetic lineage when they were closer than 1/10th of latitude or longitude or were otherwise excluded from the analyses. We excluded nine sites for the plants (site # 1014, 303, 1532, 719, 1546, 1547, 1437, 429, 1548, 1438) and 13 sites for the fungal pathogens (site # 1068, 1014, 303, 1532, 719, 1546, 940, 1547, 1437, 333, 1249, 6809, 1548). In total, we considered 20 groups of sites for the host and 19 groups of sites for the pathogen. We included sibling individuals but excluded individuals with admixed membership between genetic clusters inferred from the STRUCTURE analysis (i.e. individuals with mean membership coefficient <0.80 to the given cluster), which may be due to low assignment power or admixture. To take into account differences of sample size between sites, we also estimated allelic richness using ADZE (Szpiech, Jakobsson, & Rosenberg, 2008) which corrects for sample size difference. Calculations were performed using a standardized sample size of N=3, corresponding to the smallest number of observations per site. We tested for mean differences in diversity statistics between species and genetic clusters using a Wilcoxon rank sum test and a Kruskal-Wallis rank sum test in the R software v3.5.3, respectively, considering values across groups of sites. To study isolation-by-distance patterns in the *S. nutans* and fungi parasitizing *S.*

nutans datasets, we computed correlations between matrices of genetic and geographic distances of plant and fungal populations using a Mantel test. Genetic distances between populations at each group of sites were calculated as the Nei's distance (Nei, 1972) using the dist.genpop() function of the R package {ADEGENET} (Jombart, 2008; Jombart & Ahmed, 2011). Geographic distances were calculated using the distm() function of the R package{geosphere}. We computed Mantel tests using the mantel.rtest() function of the R package {ade4} with 1,000,000 resamples for the null distribution. We also studied the correlations between matrices of genetic distances of plant and fungal populations using a Mantel test as above. To remove the effect of the correlation between genetic and geographic distances in the correlation between matrices of genetic distances, we also performed a partial Mantel test using the partial.mantel.test() function of the R package {ncf} with 1,000,000 resamples for the null distribution (Feurtey et al., 2016). We considered 16 group of sites, the same both for the host and the pathogen, corresponding to *S. nutans* host plants and fungal strains belonging to the genetic cluster parasitizing S. *nutans.* We excluded sites of *S. italica* host plants and fungal strains belonging to the genetic cluster parasitizing *S. italica,* and one site studied in the host but not studied in the pathogen (site # 333). We investigated the occurrence of recent events of effective population size reduction (i.e. bottlenecks) within the identified fungal genetic clusters using the program BOTTLENECK version 1.2.02 (Piry, Luikart, & Cornuet, 1999).

Inference of anther-smut fungi divergence history based on microsatellite markers

We studied the divergence history of the anther-smut fungi sampled on the *S. nutans* lineage complex using an approximate Bayesian computation-random forest (ABC-RF) procedure that performs ABC inferences based on the machine learning tool named "random forest" (Breiman,

2001; Pudlo et al., 2016; Raynal et al., 2019). The divergence history of the host had already been previously inferred (Martin et al., 2016; Martin et al., 2017; Van Rossum et al., 2018). To increase our power to disentangle between different evolutionary scenarios, we performed model choice and parameter estimation by comparing scenarios or groups of scenarios in sequential rounds, each round testing a particular type of evolutionary event, either divergence time, order of divergence or presence of gene flow (Estoup, Raynal, Verdu, & Marin, 2018; Liu et al., 2019); Table 1; Table S5). We built scenarios based on the genetic clusters obtained using microsatellite markers and on previous analyses of divergence of anther-smut fungi and their host species (Badouin et al., 2017; Gladieux et al., 2013; Martin et al., 2016; Martin et al., 2017; Van Rossum et al., 2018). The tested scenarios varied regarding the time of divergence, the relative order of divergence of fungal genetic clusters and the occurrence of gene flow among genetic fungal clusters (Table S5). We did not include scenarios with variation in effective population sizes as the BOTTLENECK analysis did not identify any signature of effective population size reduction (see 'Results' section). To test if the order of sequential rounds had an effect on the outcome of the analysis, we tested the time of divergence and the relative order of divergence as either the first or the second round (Note S1). We used as populations the fungal genetic clusters identified through population structure analyses and removed fungal strains that were likely siblings of other fungal strains in each genetic cluster.

We ran the ABC procedure modified from (Liu et al., 2019). Briefly, we simulated datasets with 22 fungal microsatellite markers using the ABCtoolbox program (Wegmann, Leuenberger, Neuenschwander, & Excoffier, 2010) and used fastsimcoal 2.5 for coalescent-based simulations (Excoffier & Foll, 2011). We simulated 10,000 genetic datasets per scenario using coalescent simulations with model parameters drawn from prior distributions (Table S6A). We set prior

distributions based on previous analyses of divergence of anther-smut fungi and of their host species (Badouin et al., 2017; Branco et al., 2018; Gladieux et al., 2013; Martin et al., 2016; Martin et al., 2017; Van Rossum et al., 2018). We set a generation time of one year for the pathogen based on its life cycle (Thrall, Biere, & Antonovics, 1993). We estimated the following parameters: effective size of each genetic cluster (*N*), divergence time and migration rate per generation between two genetic clusters *x* and *y* (*Txy* and *mxy*, respectively). We computed 16 summary statistics for the observed and simulated datasets with the program arlsumstats (Table S6B; (Excoffier & Lischer, 2010). We assumed a generalized stepwise model of microsatellite evolution (Estoup, Jarne, & Cornuet, 2002) and allowed the mutation rate to vary across microsatellite markers. We drew locus-specific mutation rates from a gamma distribution (α,α/μ) in which μ is the mutation rate per generation and α is a shape parameter (Cornille et al., 2012; Liu et al., 2019). We used the R package {abcrf} v1.7.0 (Pudlo et al., 2016) to compute a classification vote through ABC-RF, representing the number of times a scenario was selected among classification *n* trees of the constructed random forest. We chose n=500 trees. For each round, we selected the scenario or group of scenarios with the highest number of classification votes. We computed the posterior probability and the prior error rates over 10 replicated analyses. We performed a linear discriminant analysis (LDA) in the R package {abcrf} v1.7.0 (Pudlo et al., 2016) on the simulated and observed datasets to visually check the fit of the model to the observed data. Finally, we performed parameter inferences using the group of models eventually selected.

Sequencing data and genome assemblies

We performed whole-genome sequencing for 23 anther-smut fungal strains parasitizing *S. italica*

and 30 strains parasitizing *S. nutans*. Out of these 53 strains, 38 strains belonged to the 170 host-pathogen pairs that were genotyped with microsatellite markers (Table S7). Due to limited fungal material, we sequenced six additional strains that were collected at our studied sites but not genotyped for microsatellite markers and nine additional strains collected at three new sites (Table S7). We extracted DNA from fresh spores stored at -20°C using the Nucleospin Soil Kit (Macherey-Nagel, Germany). Genomes were sequenced using Illumina 150 bp paired-end sequencing technology at 46X coverage on average. We also included outgroups using available whole-genome sequences of 56 anther-smut fungal strains parasitizing closely-related host species of *S. nutans* and *S. italica* occurring in the studied geographic range (Le Gac, Hood, Fournier, et al., 2007; Refrégier et al., 2008): 33 *M. lychnidis-dioicae* strains parasitizing *S. latifolia* (Badouin et al., 2017), 19 *M. silenes-dioicae* strains parasitizing *S. dioica* (Badouin et al., 2017) and four *M. violaceum* var *paradoxa* strains parasitizing *S. paradoxa* (Branco et al., 2018) (Table S7). We downloaded raw data publicly available from the NCBI SRA under the BioProject IDs PRJNA295022 and PRJEB16741. We thus analysed a total of 109 *Microbotryum* genomes. For read mapping, we used as reference genome the high-quality haploid genome assembly of the MvSn-1249 *M. violaceum s. str.* strain corresponding to the a_2 mating type (collected on *S. nutans*) previously obtained with P6/C4 Pacific Biosciences SMRT technology and annotated for gene models (Branco et al., 2017). The MvSn-1249-A2 assembly was accessed from GenBank BioProject accession number PRJEB12080 (BioSample ID: SAMEA3706514, assembly: GCA_900014965).

Genome read mapping and variant calling procedure

We performed read mapping and SNP calling of the 109 analysed genomes against the MvSn-1249-A2 genome as previously described (Branco et al., 2018; Hartmann, Rodríguez de la Vega,

375 Brandenburg, Carpentier, & Giraud, 2018). First, we trimmed Illumina raw reads for sequence
376 quality and removed adapter sequences using the software Cutadapt v1.8.3 (Martin, 2011) with
377 the options: -q 10, 10; –minimum-length 50; -a
378 AGATCGGAAGAGCACACGTCTGAACTCCAGTCAC; -
379 AAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTAGATCTCGGTGGTCGCCGTATT.
380 We aligned trimmed reads using the short read aligner bowtie2 v2.1.0 (Langmead, Trapnell, Pop,
381 & Salzberg, 2009) with the three following software options: –very-sensitive-local; –phred33; -X
382 1000. We removed PCR duplicates using the MarkDuplicates tool of Picard tools version 2.14.1
383 (http://broadinstitute.github.io/picard) and performed local realignment of mapped reads using
384 the tools RealignerTargetCreator and IndelRealigner of the Genome Analysis Toolkit (GATK)
385 version 3.8 (McKenna et al., 2010) to improve alignment accuracy in indel regions. Mean
386 alignment rates to the reference genome ranged from 53 to 96% (Table S7). For SNP calling, we
387 used GATK version 3.7 (McKenna et al., 2010) and ran HaplotypeCaller on each strain
388 individually using a diploid mode. Then, we used GenotypeGVCFs on a merged gvcf variant file
389 to perform joint variant calls. We filtered SNP calls for quality using VariantFiltration and
390 following the GATK good practices for hard-filtering of variants (QUAL < 250; QD < 2; MQ <
391 30.0; -10.5 > MQRankSum > 10.5; -5 > ReadPosRankSum > 5; FS > 60; SOR > 3) and masked
392 SNPs for repeats. We found 41 SNPs in the strain MvSn-1249-A2 (the very same fungal strain as
393 the reference genome also sequenced with Illumina), indicating low rates of false positives in our
394 SNP calling. We only kept bi-allelic SNPs with a high genotyping rate (> 90%) on autosomes
395 (i.e., excluding SNPs on the mating-type contigs MvSn-1249-A2-R1_MC15 and MvSn-1249-A2-
396 R1_MC16). We excluded SNPs on the mating-type chromosomes for population genomics
397 analyses as they exhibit suppressed recombination on 90% of their lengths (Branco et al., 2018).
398 Removing the strain MvSn-1249-A2, we kept 108 fungal strains and 1,305,369 SNPs for

399 population genomic analyses.

400

401 **Population genomics analyses**

402 To analyze the population genetic structures of anther-smut fungi parasitizing several *Silene*
403 species using the genome-wide SNP dataset called on autosomes, we used the model-based
404 Bayesian clustering approach implemented in the software STRUCTURE version 2.3.4
405 (Pritchard et al., 2000), from K=2 to K=10 clusters, and performed a PCA, similarly as for the
406 microsatellite datasets. For the STRUCTURE analysis, we selected a set of 233 unlinked SNPs
407 randomly distributed at intervals of 100 kb along the autosomes, linkage disequilibrium decaying
408 over 50-100 kb in *Microbotryum* populations (Badouin et al., 2017). We used the R package
409 {Pophelper} (https://github.com/royfrancis/pophelper) to perform barplots. We performed the
410 PCA on all SNPs (but excluding missing data and heterozygote genotypes) using the --pca
411 command of the Plink v1.9 software (Chang et al., 2015; Purcell et al., 2007). To study footprints
412 of gene flow between fungal strains, we performed a phylogenetic network analysis implemented
413 in splitsTree (Huson, 1998; Huson & Bryant, 2006) with the neighbornet method.

414

415 **Results**

416 **Population genetic structure and diversity of *Silene nutans* and *S. italica* host plants**

417 After checking plant species identity with molecular markers, we retained 145 diseased *S. nutans*
418 plants and 25 diseased *S. italica* plants (Table S1; Fig 1A). The 21 plant nuclear microsatellite
419 markers distinguished well *S. nutans* and *S. italica* both in STRUCTURE analyses from K=2 and
420 in the principal component analysis (PCA; Fig 1B-C; Fig S1).

421 Within *S. nutans*, we identified four distinct haplotypes based on the six plastid SNPs, retrieving
422 in our sample the four major genetic lineages of *S. nutans* previously identified in Western
423 Europe (Fig 1A; (Martin et al., 2016; Martin et al., 2017; Van Rossum et al., 2018). The blue,
424 yellow, orange, and red haplotypes were found, as in previous studies, in the northeastern,
425 southeastern, western and southwestern parts of Europe, respectively. All *S. italica* samples had a
426 blue plastid haplotype (Fig 1A). Within *S. nutans,* the PCA based on the 21 plant nuclear
427 microsatellite markers mainly differentiated the orange genetic lineage from the other ones (Fig
428 1B; Fig S1B) and the STRUCTURE analysis detected four genetic clusters, corresponding to the
429 different plastid haplotypes, except for the red haplotype for which we had too few samples (Fig
430 1C; Fig S1A). The STRUCTURE analysis further identified two well-separated genetic clusters
431 within the *S. nutans* orange plastid haplotype (Fig. 1C). These two genetic clusters also appeared
432 separated on the second axis of the PCA and corresponded to plant individuals collected in the
433 Northern part of France *versus* Belgium and Western part of France (Fig S1), in agreement with
434 previous findings (Martin et al., 2016). From the STRUCTURE analyses, we identified five host
435 individuals with admixed cluster memberships (mean membership coefficient <0.80 to the given
436 cluster), which may be due to low assignment power or admixture between genetic clusters. We
437 found no further clear subdivision within *S. nutans* at higher K values and no clear population
438 subdivision within *S. italica* (Fig S1A).

439 We found no significant differences in allelic richness neither between *S. nutans* and *S. italica*
440 (Wilcoxon rank sum (WRS) test on 20 groups of sites, W = 12, p-value = 0.169; Table S8A) nor
441 between the four *S. nutans* genetic clusters (Kruskal-Wallis rank sum (KWRS) test on 16 groups
442 of sites, $Chi^2 = 4.92$, degree of freedom (df) = 3, p-value = 0.178; Table S8A). In *S. nutans,* levels
443 of observed heterozygosity were significantly lower than those of expected heterozygosity under

444 Hardy-Weinberg equilibrium for most loci, the mean F_{IS} value per marker being 0.33 (Table S2).
445 In *S. italica,* levels of observed and expected heterozygosities were not significantly different for
446 most markers (78%; Table S2), the mean F_{IS} value per locus being 0.37 (Table S2). As expected,
447 pairwise F_{ST} and Jost's D indicated higher divergence between *S. italica* and *S. nutans* (mean
448 F_{ST}=0.32; mean Jost's D=0.57; Table S9A) than between genetic clusters of *S. nutans* (mean
449 F_{ST}=0.20; mean Jost's D=0.35; Table S9A; KWRS test for both statistics, W = 0, p-value =
450 0.010). For *S. nutans,* we found on average lower pairwise F_{ST} and Jost's D values between the
451 yellow Western cluster and the two orange Western and Northwestern genetic clusters (mean
452 F_{ST}=0.17; mean Jost's D=0.28; Table S9A) than between the blue Eastern genetic cluster and the
453 Western genetic clusters (mean F_{ST}=0.24; mean Jost's D=0.42; Table S9A), although the
454 difference was not significant (KWRS test for F_{ST}, W = 8, p-value = 0.200; KWRS test for Jost's
455 D, W = 9, p-value = 0.100). The correlation between the matrices of pairwise genetic and
456 geographic distances for *S. nutans* (tested on 16 sites) was significantly positive (Mantel test: r =
457 0.54, p-value < 0.001), indicating an isolation-by-distance pattern.

458

459 **Differentiation between anther-smut fungi parasitizing *Silene nutans* and *S. italica***

460 Population genetic structure analyses separated *Microbotryum* strains into two main genetic
461 clusters, corresponding to their host plant of sampling, *S. italica* and *S. nutans*. The host of
462 collection separated strains according to the first axis of the PCA (Fig 2B). The STRUCTURE
463 analysis and the discriminant analysis of principal component (DAPC) also delimitated two clear
464 fungal genetic clusters according to the host of sampling at K=2 (Fig 2C; Fig S2). In the PCA
465 (Fig. 2B), four *Microbotryum* strains sampled on *S. nutans* clustered with strains parasitizing *S.*

italica and one *Microbotryum* strain sampled on *S. italica* clustered with strains parasitizing *S. nutans*. These strains may correspond to spill-overs, i.e. cross-species disease transmissions, between the two studied *Silene* species or from other *Silene* species. We excluded for further analyses the 11 strains that had their main cluster memberships < 80% in the STRUCTURE analysis at K=2 (Fig 2C), which can be due to low assignment power or admixture.

Anther-smut fungi parasitizing *S. italica* had on average higher levels of allelic richness (WRS test on 19 groups of sites, W = 48, p-value = 0.004) and observed heterozygosity (WRS test on 19 groups of sites, W = 30, p-value = 0.014) than anther-smut fungi parasitizing *S. nutans* (Table S8B). Levels of observed heterozygosity were significantly lower than heterozygosity expected under Hardy-Weinberg equilibrium for eight markers for *Microbotryum* fungi parasitizing *S. nutans* and six markers for *Microbotryum* fungi parasitizing *S. italica* (Table S4).

Similar genetic structure in anther-smut fungi as in their *Silene nutans* and *S. italica* host plants

The existence of additional levels of population genetic structure within fungal strains was indicated by the STRUCTURE barplots (Fig S2), the second and third principal components of the PCA that explained 27.9 % of the total variance between strains (Fig 2B). To investigate the population structure within each of the two main genetic clusters of anther-smut fungi, corresponding mainly to populations parasitizing *S. nutans* and *S. italica*, respectively, we analyzed them separately, which confirmed the existence of a strong population structure within anther-smut fungi parasitizing *S. nutans* (Fig 3A). The first two PCA axes separated fungal strains into three distinct genetic clusters, corresponding to the three most frequent plastid

haplotypes identified in the host plant (Fig 3A), which was also supported by the DAPC and the STRUCTURE analyses (Fig 3B; Fig S3). We named these three fungal genetic clusters the blue Eastern, yellow Western and orange Western *Microbotryum* clusters, in reference to their *S. nutans* host plastid haplotypes. In contrast to the host plant *S. nutans*, we found no clear subdivision in the orange Western anther-smut fungal cluster (Fig S3). We found no genetic clusters specific to strains sampled from a host with a red haplotype, perhaps due to the low strain number. We excluded for further analyses the two admixed fungal strains (cluster memberships <0.80 in a given cluster) and the three fungal strains sampled on a host with a red haplotype. We found no further population structure in the set of 28 anther-smut strains parasitizing *S. italica* (Fig S4), as for the host plant.

All three fungal genetic clusters parasitizing *S. nutans* had similar levels of allelic richness (KWRS test on 16 groups of sites, $Chi^2 = 1.33$, df = 2, p-value = 0.514; Table S8B) and observed heterozygosity (KWRS test, $Chi^2 = 0.67$, df = 2, p-value = 0.716; Table S8B). The F_{IS} varied between genetic clusters from -0.11 to 0.82 (Table S8B). We found no significant differences in differentiation between the three fungal genetic clusters (KWRS tests for F_{ST} and Jost's D, $Chi^2 = 2$, df = 2, p-value = 0.368; Table S9B). The significant correlation between the matrices of genetic and geographic distances between pairs of fungal populations parasitizing *S. nutans* (16 sites) indicated an IBD pattern (Mantel test: r = 0.62, p-value < 0.001). There was a significantly positive correlation between genetic distance matrices for the fungal and the *S. nutans* populations (16 sites; Mantel test: r = 0.80, p-value < 0.001), and it remained significant when controlling for the IBD effect (partial Mantel test: r = 0.70, p-value < 0.001). Altogether, our findings indicate strong congruence between the population genetic structures of the anther-smut fungi parasitizing *S. nutans* and its host plant *S. nutans* (Fig 3C). We found no signatures of

511 recent reduction in effective population size in either of the three fungal genetic clusters
512 parasitizing *S. nutans* nor the fungal genetic cluster parasitizing *S. italica* using the
513 BOTTLENECK software (Piry et al., 1999) under either stepwise or two-phase models of
514 microsatellite evolution (Table S10; one-tailed Wilcoxon signed rank test, all P-values>0.36).

515

516 **Inferred divergence history of anther-smut fungi on *Silene nutans* congruent with of their**
517 **host plant**

518 To study if the anther-smut fungi on *S. nutans* shared the same history of divergence into cryptic
519 lineages as the one previously inferred for their host plants, the *S. nutans* cryptic species complex
520 (Martin et al., 2016; Martin et al., 2017; Van Rossum et al., 2018), we used an approximate
521 Bayesian computation random forest (ABC-RF) procedure. We compared various divergence
522 scenarios for the three identified *Microbotryum* genetic clusters parasitizing *S. nutans.* We
523 removed the 21 fungal strains that were likely siblings of other fungal strains, keeping 110
524 anther-smut fungal strains belonging to the blue Eastern, yellow Western and orange Western
525 *Microbotryum* genetic clusters parasitizing *S. nutans* and we used as outgroup the 25 genotyped
526 *Microbotryum* strains parasitizing *S. italica.*

527 We performed three rounds of scenario comparison, each testing a particular evolutionary event,
528 including divergence times, divergence order and the occurrence of gene flow (Table 1; Table S5;
529 Fig S5). At each round, we retained the inferred most likely evolutionary scenario to be used as
530 backbone for the subsequent rounds. To assess our power to discriminate between scenarios, we
531 checked posterior probability (Table 1) and assessed visually whether the observed data fell
532 within the clouds of simulated data of the compared scenarios (Fig S6).

In the first round ("time of divergence"; Table 1, “round 1”), we tested four different time periods of divergence for various divergence orders (A, B or C; Fig S5A). The group of scenarios with strongest support included a first divergence occurring between 100,000-1,500,000 years ago, and a second divergence occurring between 0-20,000 years ago (posterior probability = 0.54 +-0.02, prior error rate = 34%; Table 1, “round 1”; Fig S6A). In the second round ("divergence order"; Table 1, “round 2”), we found with a high posterior probability for the scenario assuming a first divergence of the blue Eastern *Microbotryum* genetic cluster, followed by the divergence between the yellow and orange Western *Microbotryum* genetic clusters (posterior probability = 0.79 +-0.02, prior error rate = 2%; Table 1, “round 2”; Fig S6B). When testing the order of divergence between clusters as first round and the time of divergence as second round, we selected the same scenarios with a high posterior probability (Note S1; Table S11).

In the third round ("occurrence of gene flow"; Table 1, “round 3, step 1”), we compared in a first step three groups of scenarios differing in the clusters affected by gene flow. The most supported group of scenarios assumed gene flow among *Microbotryum* genetic clusters parasitizing *S. nutans* but not with *S. italica* (posterior probability = 0.60 +-0.02, prior error rate = 5%; Table 1, “round 3, step 1”; Fig S6C). In the following steps (Table 1, “round 3, steps 2 to 4”), we compared groups of scenarios of divergence with different timing of gene flow. Scenarios of divergence with ancient gene flow were the most supported, which suggested that episodes of ancient gene flow occurred during divergence or just following divergence of the three *Microbotryum* genetic clusters parasitizing *S. nutans* and then completely stopped (posterior probability = 0.76 +-0.02, prior error rate = 35%; Table 1, “round 3, step 2”; Fig S6D-F). However, we did not have enough power to differentiate between scenarios of ancient gene flow that involved different genetic clusters (either all three genetic clusters or only the two Western

556 genetic clusters) or different times of past gene flow (either over a 0-20000 year period or a
557 50,000-500,000 year period; Table S5; posterior probability = 0.49 +-0.03, prior error rate = 60%;
558 Table 1, "round 3, step 5"). We performed parameter inferences for the group of four models
559 with ancient migration ("round 3, step 2", group 3 see Table 1; Table S12). The inferred dates
560 and order of divergence between fungal genetic clusters parasitizing *S. nutans* were consistent
561 with the history of divergence previously inferred for the host plant *S. nutans* (Fig 4; Martin et
562 al., 2016; Martin et al., 2017; Van Rossum et al., 2018).

563

564 **No genome-wide signatures of gene flow between anther-smut fungi on distinct *Silene***
565 **species**

566 To check whether genome-wide data support the patterns inferred above on population genetic
567 structure in the anther-smut fungi and lack of recent gene flow, we sequenced the genomes of 53
568 *Microbotryum* strains parasitizing *S. nutans* or *S. italica* (Table S7). We also used previously
569 sequenced genomes of anther-smut fungi parasitizing species with overlapping geographical
570 ranges in order to check if some of the strains that appeared as admixed or did not cluster
571 according to their host species may actually correspond to other *Microbotryum* species. After
572 filtration, we obtained 1,305,369 SNPs for 108 *Microbotryum* strains. We performed PCA
573 analyses on the genome-wide SNPs and STRUCTURE analyses and confirmed strong population
574 differentiation between fungal strains parasitizing *S. nutans* and *S. italica* (Fig 5A-B; Fig S7A).
575 Only one strain parasitizing *S. nutans* clustered with strains parasitizing *S. italica,* similarly as
576 found from microsatellite data (strain # 1438; Fig 2C), therefore likely being a genuine spill-over
577 between *S. nutans* and *S. italica*. Two other strains parasitizing *S. nutans* clustered with strains

parasitizing *S. latifolia.* These strains had admixed cluster membership (i.e. mean membership coefficient <0.80 to the given cluster) or clustered with strains parasitizing *S. italica* from microsatellite data (Fig 2C). These strains likely represent spill-overs between *S. nutans* and *S. latifolia*.

Genome sequences retrieved the same genetic structure as microsatellites for anther-smut fungi parasitizing *S. nutans,* with a strong differentiation between *Microbotryum* strains sampled on the *S. nutans* blue plastid haplotype on the one hand and those sampled on yellow, red or orange plastid haplotypes on the other hand (Fig S7B), and confirmed the absence of strong population structure within anther-smut fungi parasitizing *S. italica* (Fig S7C). We also found strong population differentiation between fungal strains parasitizing the three closely related host species of *S. nutans* and *S. italica* (Fig 5A-B). STRUCTURE analyses showed no signatures of recent admixture between strains parasitizing different hosts (Fig 5B).

The phylogenetic network analysis implemented in SplitsTree also supported the absence of recent gene flow at the genome-wide levels between the *Microbotryum* species parasitizing different hosts and the close genetic similarity of the three genetic clusters of anther-smut fungi parasitizing the *S. nutans* complex (Fig 5C). Within *Microbotryum* strains parasitizing *S. nutans*, we found no footprint of recombination between strains of the two main genetic clusters (Fig 5C) which was consistent with a scenario of divergence with ancient migration and no contemporary gene flow inferred from ABC demographic inferences.

Discussion

We found that genetic structure and divergence history of anther-smut fungi parasitizing the *S. nutans* plant species complex closely mirrored those of their host plant genetic lineages. Such a congruence in population structure and divergence history of the host and the pathogen in the *S. nutans* complex, the distribution area of the genetic clusters in Western Europe, as well as the inferred dates of divergence suggest that the host and the pathogen differentiated in parallel following isolation in shared glacial refugia. To confirm the co-divergence events between *S. nutans* lineages and anther-smut fungi genetic clusters, we need to more accurately date the node of lineage divergence events in the host and the pathogen. Such molecular dating is essential to support co-divergence events as preferential shifts of the pathogen to closely related host species can also produce congruent phylogenies (de Vienne et al., 2013), but is currently difficult due lack of appropriate calibration points.

Nevertheless, our results rule out the hypothesis that anther-smut fungi on some of the *S. nutans* genetic lineages may result from specialization by host shifts from distant anther-smut fungi parasitizing other *Silene* species in sympatry or parapatry. This contrasts with the prevalence of host shifts between distant lineages observed in the Caryophyllaceae-*Microbotryum* system at larger evolutionary scales (Refrégier et al., 2008), but is consistent with the strong congruence reported between the population genetic structures of *S. latifolia* and its anther-smut fungi, also corresponding to glacial refugia footprints (Feurtey et al., 2016). Our study thus contributes to gain general insights into the processes of divergence in host-pathogen systems. While at large evolutionary scales, host shifts seem to be the rule (de Vienne, Hood, & Giraud, 2009; Thines, 2019), finer and more recent population subdivisions may more often result from codivergence due to shared geographic and climatic constraints, such as glacial refugia. The discrepancy between the two evolutionary scales may be due to recurrent pathogen lineage extinctions

622 followed by recolonizations through host shifts: codivergence may occur frequently, but
623 pathogen lineages would regularly go extinct over longer evolutionary scales and plant lineages
624 would be recolonized by host shifts. This novel hypothesis can be tested by studying further
625 closely related pairs of natural host-pathogen associations, which unfortunately still remain
626 scarce.

627 The significant IBD pattern in both the *S. nutans* plant and anther-smut fungi and the correlation
628 between the genetic distances between host and pathogen pairs when controlling for IBD suggest
629 that the anther-smut fungi followed similar recolonization routes as the plant and/or became
630 specialized on the host genetic lineages. However, contrary to the system *S. latifolia* - *M.*
631 *lychnidis-dioicae* (Feurtey et al., 2016) and other host-pathogen systems (Barrett et al., 2008;
632 Criscione, Poulin, & Blouin, 2005; Nieberding et al., 2008; Nieberding & Olivieri, 2007), we
633 found higher levels of subdivision in *S. nutans* than in its associated anther-smut fungi. The
634 weaker genetic structure in anther smut fungi compared to its host suggests again extinction of
635 the pathogen lineage in a plant lineage, followed by recolonization from another fungal lineage,
636 or the ability of an anther-smut lineage to remain generalist on two closely related plant lineages.

637 Future studies using cross-inoculation experiments between strains and host plants of different
638 genetic clusters within the *S. nutans* complex could assess whether the congruence of host-
639 pathogen genetic structure is associated with a pattern of local adaptation of the pathogen or of
640 the pathogen, as in the system *S. latifolia* - *M. lychnidis-dioicae* (Delmotte et al., 1999; Feurtey et
641 al., 2016; Kaltz et al., 1999). We could not obtain enough fresh material for this study to be able
642 to test local adaptation. Furthermore, strong asymmetric postzygotic reproductive isolation was
643 found between the Eastern and Western *S. nutans* plant lineages (Martin et al., 2017) and it would
644 be interesting in future studies to explore if a similar reproductive isolation pattern is present

between the Eastern and Western fungal lineages on *S. nutans*, and whether reproductive isolation can be found among anther-smut fungi.

We found clear differentiation between fungal populations parasitizing the two closely related *S. nutans* and *S. italica* species, supporting previous findings of differentiation based on a few strains and loci (Bucheli et al., 2000). No strong population structure was found within *S. italica* or within its associated anther-smut fungi. However, future studies with larger sampling distribution in the *S. italica* range in the Mediterranean Basin (Naciri, Pasquier, Lundberg, Jeanmonod, & Oxelman, 2017) may identify population differentiation. Cross-species disease transmissions, i.e. spill-overs of fungal strains, were more frequent on *S. nutans* than on *S. italica*, which may be due to unequal sampling size between the two *Silene* species or to biological differences. Some Caryophyllaceae plants indeed seem more susceptible to cross-species transmissions (Antonovics et al., 2002; de Vienne, Hood, et al., 2009; Hood et al., 2019). Putative spill-overs may actually correspond to other *Microbotryum* species, that were too rare to be identified in our population analyses using microsatellite markers, as revealed with the whole genome sequence data. In fact, two strains collected on *S. nutans* were actually assigned, based on whole genome sequences, to *M. lychnidis-dioicae,* specialized on *S. latifolia.*

The statistical comparison of demographic models and the whole-genome SNP analyses indicated the absence of recent gene flow between *Microbotryum* species parasitizing closely related *Silene* species, as well as between the Eastern and Western fungal lineages parasitizing *S. nutans*, although some spill-over cases were detected. Noteworthy, we also found no evidence of hybrid individuals between the two host plants species *S. nutans* and *S. italica* based on SNP data. The admixed barplots observed in the microsatellite STRUCTURE analyses must therefore have been due to low power of assignment rather than recent hybridization. This result was consistent with

668 previous studies on other several anther-smut fungi, in which no hybrids were detected in natural
669 populations, even between very closely related species that can hybridize in the laboratory
670 (Abbate et al., 2018; Badouin et al., 2017; Gladieux et al., 2013; Petit et al., 2017). The only case
671 where hybrids were detected was among closely related anther-smut fungi with overlapping host
672 and geographic ranges, on *Dianthus* species (Petit et al., 2017). Ecological factors, such as
673 different habitats, pollinator guilds or flowering time of the host species (Jürgens, Witt, &
674 Gottsberger, 1996; Kephart, Reynolds, Rutter, Fenster, & Dudash, 2006), may constitute pre-
675 zygotic factors favoring reproductive isolation of host-specialized anther-smut fungi in natural
676 populations. Both extrinsinc and intrinsic post-zygotic barriers can be strong in anther-smut fungi
677 (Giraud & Gourbière, 2012; Le Gac, Hood, & Giraud, 2007). Strong host specialization may
678 indeed play a role in reproductive isolation, through migrant inviability and hybrid maladaptation
679 on parental hosts, especially given the life cycle of *Microbotryum* fungi, with many spores falling
680 on a plant and competing for systemic infection, and selfing being frequent, exposing hybrids to
681 systematic competition with non-hybrids (Gibson, Hood, & Giraud, 2012). In addition,
682 comparative genomics of anther-smut fungi showed presence of large genomic rearrangements
683 and gene content variation between species (Branco et al., 2018; Hartmann et al., 2018), and
684 experimental crosses suggested high frequency of hybrid sterility and abnormal genomic contents
685 in hybrids (de Vienne, Refrégier, et al., 2009).

686 The low levels of gene flow among anther-smut fungi parasitizing different hosts found in
687 *Microbotryum* fungi stand in high contrast with frequent reports of signatures of introgression in
688 other fungal pathogens, such as crop pathogens or human disease-associated pathogens (Feurtey
689 & Stukenbrock, 2018). Few studies have focused on fungal pathogens diversification in natural
690 host communities while several evolutionary processes, such as time scale of divergence, host

density and heterogeneity, are likely very different than those occurring on human-modified environment pathogens (Laine, 2005; Laine, Barrès, Numminen, & Siren, 2019; Stukenbrock & McDonald, 2008). To understand how biodiversity arises and what the mechanisms of host-pathogen evolution are over large evolutionary scales, we need more studies on co-evolutionary histories of parasites and their hosts in natural ecosystems. Studies of population genetic structure and divergence in plant fungal pathogens indeed remain so far mostly focused on crop pathogens (Barrès et al., 2008; Enjalbert, Duan, Leconte, HovmøLler, & De Vallavieille-Pope, 2005; Fournier & Giraud, 2008; Linde, Zhan, & McDonald, 2002; Saleh, Milazzo, Adreit, Fournier, & Tharreau, 2014; Stukenbrock, Banke, & McDonald, 2006; Zaffarano, McDonald, & Linde, 2008), in which patterns are heavily impacted by host genetic homogeneity and high abundance, as well as by human-mediated plant and pathogen dispersal. We found here strong population structure congruence between hosts and pathogens, in agreement with a previous study on other anther-smut lineages (Feurtey et al., 2016). Further studies on other natural systems are needed to assess whether this represents a general pattern in natural pathogen-host communities.

Acknowledgements

We thank all the collectors cited in Table S1 and S7 for their contribution to plant material sampling, the Département de la Nature et des Forêts (Service Public de Wallonie, Belgium), B. Clesse (Centre Marie-Victorin, Cercles des Naturalistes de Belgique asbl), Ardenne & Gaume and Natagora for access to natural sites and for the authorization to collect plant material. We are grateful to the INRA MIGALE bioinformatics platform (http://migale.jouy.inra.fr) for providing computational resources and the Genotoul platform for sequencing. We thank the Plateforme de Genotypage GENTYANE INRA UMR1095 for help with microsatellite genotyping. This work

714 was supported by the European Research Council (ERC) (starting grant GenomeFun 309403), the
715 French National Research Agency (ANR) (Gandalf ANR Grant ANR-12-ADAP-0009), the Louis
716 D. Foundation (Institut de France) to TG, and a Marie Curie European grant (PRESTIGE-2016-4-
717 0013) to FEH. PT and CG thank the Région Hauts-de-France, and the Ministère de
718 l'Enseignement Supérieur et de la Recherche (CPER Climibio), and the European Fund for
719 Regional Economic Development for their financial support.

720

721 **Datasets**

722 dataset 1: Hartmann F.E., Snirc A., Cornille A., Godé C., Touzet P., Van Rossum F., Fournier E.,
723 Le Prieur S., Shykoff J., Giraud T. 2019. Plant microsatellite and chloroplastic genotypes. Dryad
724 doi (to be completed upon acceptance).

725 dataset 2: Hartmann F.E., Snirc A., Cornille A., Godé C., Touzet P., Van Rossum F., Fournier E.,
726 Le Prieur S., Shykoff J., Giraud T. 2019. Fungal microsatellite genotypes. Dryad doi (to be
727 completed upon acceptance).

728 dataset 3: Hartmann F.E., Snirc A., Cornille A., Godé C., Touzet P., Van Rossum F., Fournier E.,
729 Le Prieur S., Shykoff J., Giraud T. 2019. Whole genome raw data of *Microbotryum* fungi
730 infecting *S. nutans* and *S. italica*. NCBI Sequence Read Archive (SRA). BioProject xxx (ID
731 available upon acceptance).

732

733

734 **References**

735 Abbate, J. L., Gladieux, P., Hood, M. E., de Vienne, D. M., Antonovics, J., Snirc, A., & Giraud,
736 T. (2018). Co-occurrence among three divergent plant-castrating fungi in the same *Silene*
737 host species. *Molecular Ecology*, *27*(16), 3357–3370. doi: 10.1111/mec.14805
738 Antonovics, J., Hood, M. E., & Partain, J. (2002). The ecology and genetics of a host shift:
739 *Microbotryum* as a model system. *The American Naturalist*, *160*(S4), S40–S53. doi:
740 10.1086/342143
741 Badouin, H., Gladieux, P., Gouzy, J., Siguenza, S., Aguileta, G., Snirc, A., … Giraud, T. (2017).
742 Widespread selective sweeps throughout the genome of model plant pathogenic fungi and

identification of effector candidates. *Molecular Ecology*, *26*(7), 2041–2062. doi: 10.1111/mec.13976

Barrès, B., Halkett, F., Dutech, C., Andrieux, A., Pinon, J., & Frey, P. (2008). Genetic structure of the poplar rust fungus *Melampsora larici-populina*: Evidence for isolation by distance in Europe and recent founder effects overseas. *Infection, Genetics and Evolution*, *8*(5), 577–587. doi: 10.1016/j.meegid.2008.04.005

Barrett, L. G., Thrall, P. H., Burdon, J. J., & Linde, C. C. (2008). Life history determines genetic structure and evolutionary potential of host–parasite interactions. *Trends in Ecology & Evolution*, *23*(12), 678–685. doi: 10.1016/j.tree.2008.06.017

Becker, R., Wilks, A., Brownrigg, R., Minka, T., & Deckmyn, D. (2017). *maps: draw geographical maps. R package version 3.2.0.* (https://cran.r-project.org/web/packages/maps/index.html).

Bernasconi, G., Antonovics, J., Biere, A., Charlesworth, D., Delph, L. F., Filatov, D., … Widmer, A. (2009). *Silene* as a model system in ecology and evolution. *Heredity*, *103*(1), 5–14. doi: 10.1038/hdy.2009.34

Branco, S., Badouin, H., Rodríguez de la Vega, R. C., Gouzy, J., Carpentier, F., Aguileta, G., … Giraud, T. (2017). Evolutionary strata on young mating-type chromosomes despite the lack of sexual antagonism. *Proceedings of the National Academy of Sciences*, *114*(27), 7067–7072. doi: 10.1073/pnas.1701658114

Branco, S., Carpentier, F., Rodríguez de la Vega, R. C., Badouin, H., Snirc, A., Prieur, S. L., … Giraud, T. (2018). Multiple convergent supergene evolution events in mating-type chromosomes. *Nature Communications*, *9*(1), 2000. doi: 10.1038/s41467-018-04380-9

Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5–32. doi: 10.1023/A:1010933404324

Bucheli, E., Gautschi, B., & Shykoff, J. A. (2000). Host-specific differentiation in the anther smut fungus *Microbotryum violaceum* as revealed by microsatellites. *J. EVOL. BIOL.*, 11.

Chang, C. C., Chow, C. C., Tellier, L. C., Vattikuti, S., Purcell, S. M., & Lee, J. J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience*, *4*(1). doi: 10.1186/s13742-015-0047-8

Cornille, A., Gladieux, P., Smulders, M. J. M., Roldán-Ruiz, I., Laurens, F., Le Cam, B., … Giraud, T. (2012). New insight into the history of domesticated apple: Secondary contribution of the European wild apple to the genome of cultivated varieties. *PLOS Genetics*, *8*(5), e1002703. doi: 10.1371/journal.pgen.1002703

Criscione, C. D., Poulin, R., & Blouin, M. S. (2005). Molecular ecology of parasites: elucidating ecological and microevolutionary processes. *Molecular Ecology*, *14*(8), 2247–2257. doi: 10.1111/j.1365-294X.2005.02587.x

Croll, D., & Laine, A.-L. (2016). What the population genetic structures of host and pathogen tell us about disease evolution. *New Phytologist*, *212*(3), 537–539. doi: 10.1111/nph.14203

de Vienne, D. M., Hood, M. E., & Giraud, T. (2009). Phylogenetic determinants of potential host shifts in fungal pathogens. *Journal of Evolutionary Biology*, *22*(12), 2532–2541. doi: 10.1111/j.1420-9101.2009.01878.x

de Vienne, D. M., Refrégier, G., Hood, M. E., Guigue, A., Devier, B., Vercken, E., … Giraud, T. (2009). Hybrid sterility and inviability in the parasitic fungal species complex *Microbotryum*. *Journal of Evolutionary Biology*, *22*(4), 683–698. doi: 10.1111/j.1420-9101.2009.01702.x

de Vienne, D. M., Refrégier, G., López-Villavicencio, M., Tellier, A., Hood, M. E., & Giraud, T. (2013). Cospeciation vs host-shift speciation: methods for testing, evidence from natural

associations and relation to coevolution. *New Phytologist*, *198*(2), 347–385. doi: 10.1111/nph.12150

Delmotte, F., Bucheli, E., & Shykoff, J. A. (1999). Host and parasite population structure in a natural plant–pathogen system. *Heredity*, *82*(3), 300–308. doi: 10.1046/j.1365-2540.1999.00485.x

Dray, S., & Dufour, A.-B. (2007). The ade4 Package: Implementing the Duality Diagram for Ecologists. *Journal of Statistical Software*, *22*(1), 1–20. doi: 10.18637/jss.v022.i04

Dybdahl, M. F., & Lively, C. M. (1996). The geography of coevolution: comparative population structures for a snail and its trematode parasite. *Evolution*, *50*(6), 2264–2275. doi: 10.1111/j.1558-5646.1996.tb03615.x

Enjalbert, J., Duan, X., Leconte, M., HovmøLler, M. S., & De Vallavieille-Pope, C. (2005). Genetic evidence of local adaptation of wheat yellow rust (*Puccinia striiformis* f. sp. *tritici*) within France. *Molecular Ecology*, *14*(7), 2065–2073. doi: 10.1111/j.1365-294X.2005.02566.x

Estoup, A., Jarne, P., & Cornuet, J.-M. (2002). Homoplasy and mutation model at microsatellite loci and their consequences for population genetics analysis. *Molecular Ecology*, *11*(9), 1591–1604. doi: 10.1046/j.1365-294X.2002.01576.x

Estoup, A., Raynal, L., Verdu, P., & Marin, J.-M. (2018). Model choice using Approximate Bayesian Computation and Random Forests: analyses based on model grouping to make inferences about the genetic history of Pygmy human populations | Journal de la Société Française de Statistique. *Special Issue on Models and Inference in Population Genetics*, *159*(3). Retrieved from http://journal-sfds.fr/article/view/709

Excoffier, L., & Foll, M. (2011). fastsimcoal: a continuous-time coalescent simulator of genomic diversity under arbitrarily complex evolutionary scenarios. *Bioinformatics*, *27*(9), 1332–1334. doi: 10.1093/bioinformatics/btr124

Excoffier, L., & Lischer, H. E. L. (2010). Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, *10*(3), 564–567. doi: 10.1111/j.1755-0998.2010.02847.x

Feurtey, A., Gladieux, P., Hood, M. E., Snirc, A., Cornille, A., Rosenthal, L., & Giraud, T. (2016). Strong phylogeographic co-structure between the anther-smut fungus and its white campion host. *New Phytologist*, *212*(3), 668–679. doi: 10.1111/nph.14125

Feurtey, A., & Stukenbrock, E. H. (2018). Interspecific gene exchange as a driver of adaptive evolution in fungi. *Annual Review of Microbiology*, *72*(1), 377–398. doi: 10.1146/annurev-micro-090817-062753

Fisher, M. C., Gow, N. A. R., & Gurr, S. J. (2016). Tackling emerging fungal threats to animal health, food security and ecosystem resilience. *Phil. Trans. R. Soc. B*, *371*(1709), 20160332. doi: 10.1098/rstb.2016.0332

Fortuna, T. M., Snirc, A., Badouin, H., Gouzy, J., Siguenza, S., Esquerre, D., … Giraud, T. (2016). Polymorphic microsatellite markers for the tetrapolar anther-smut fungus *Microbotryum saponariae* based on genome sequencing. *PLOS ONE*, *11*(11), e0165656. doi: 10.1371/journal.pone.0165656

Fournier, E., & Giraud, T. (2008). Sympatric genetic differentiation of a generalist pathogenic fungus, Botrytis cinerea, on two different host plants, grapevine and bramble. *Journal of Evolutionary Biology*, *21*(1), 122–132. doi: 10.1111/j.1420-9101.2007.01462.x

Gandon, S., & Michalakis, Y. (2002). Local adaptation, evolutionary potential and host–parasite coevolution: interactions between migration, mutation, population size and generation time. *Journal of Evolutionary Biology*, *15*(3), 451–462. doi: 10.1046/j.1420-

9101.2002.00402.x

Gandon, S., Capowiez, Y., Dubois, Y., Michalakis, Y., & Olivieri, I. (1996). Local adaptation and gene-for-gene coevolution in a metapopulation model. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *263*(1373), 1003–1009. doi: 10.1098/rspb.1996.0148

Gerritsen, H. (2013). *Mapplots, r package version 1.5*. https://CRAN.R-project.org/package=mapplots.

Gibson, A. K., Hood, M. E., & Giraud, T. (2012). Sibling competition arena: selfing and a competition arena can combine to constitute a barrier to gene flow in sympatry. *Evolution*, *66*(6), 1917–1930. doi: 10.1111/j.1558-5646.2011.01563.x

Giraud, T. (2004). Patterns of within population dispersal and mating of the fungus *Microbotryum violaceum* parasitising the plant *Silene latifolia*. *Heredity*, *93*(6), 559–565. doi: 10.1038/sj.hdy.6800554

Giraud, T., & Gourbière, S. (2012). The tempo and modes of evolution of reproductive isolation in fungi. *Heredity*, *109*(4), 204–214. doi: 10.1038/hdy.2012.30

Giraud, T., Yockteng, R., Marthey, S., Chiapello, H., Jonot, O., Lopez-Villavicencio, M., … Dossat, C. (2008). Isolation of 60 polymorphic microsatellite loci in EST libraries of four sibling species of the phytopathogenic fungal complex *Microbotryum*. *Molecular Ecology Resources*, *8*(2), 387–392. doi: 10.1111/j.1471-8286.2007.01967.x

Gladieux, P., Devier, B., Aguileta, G., Cruaud, C., & Giraud, T. (2013). Purifying selection after episodes of recurrent adaptive diversification in fungal pathogens. *Infection, Genetics and Evolution: Journal of Molecular Epidemiology and Evolutionary Genetics in Infectious Diseases*, *17*, 123–131. doi: 10.1016/j.meegid.2013.03.012

Gladieux, P., Vercken, E., Fontaine, M. C., Hood, M. E., Jonot, O., Couloux, A., & Giraud, T. (2011). Maintenance of fungal pathogen species that are specialized to different hosts: allopatric divergence and introgression through secondary contact. *Molecular Biology and Evolution*, *28*(1), 459–471. doi: 10.1093/molbev/msq235

Godé, C., Touzet, P., Martin, H., Lahiani, E., Delph, L. F., & Arnaud, J.-F. (2014). Characterization of 24 polymorphic microsatellite markers for *Silene nutans*, a gynodioecious–gynomonoecious species, and cross-species amplification in other *Silene* species. *Conservation Genetics Resources*, *6*(4), 915–918. doi: 10.1007/s12686-014-0240-6

Hafner, M. S., Page, R. D. M. (1995). Molecular phylogenies and host-parasite cospeciation: gophers and lice as a model system. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *349*(1327), 77–83. doi: 10.1098/rstb.1995.0093

Hartmann, F. E., Rodríguez de la Vega, R. C., Brandenburg, J.-T., Carpentier, F., & Giraud, T. (2018). Gene presence–absence polymorphism in castrating anther-smut fungi: recent gene gains and phylogeographic structure. *Genome Biology and Evolution*, *10*(5), 1298–1314. doi: 10.1093/gbe/evy089

Hartmann, F. E., Rodríguez de la Vega, R. C., Carpentier, F., Gladieux, P., Cornille, A., Hood, M. E., & Giraud, T. (2019). Understanding adaptation, coevolution, host specialization, and mating system in castrating anther-smut fungi by combining population and comparative genomics. *Annual Review of Phytopathology*, *57*(1), 431–457. doi: 10.1146/annurev-phyto-082718-095947

Hood, M. E., Antonovics, J., Wolf, M., Stern, Z. L., Giraud, T., & Abbate, J. L. (2019). Sympatry and interference of divergent *Microbotryum* pathogen species. *Ecology and Evolution*, *0*(0). doi: 10.1002/ece3.5140

884 Huson, D. H. (1998). SplitsTree: analyzing and visualizing evolutionary data. *Bioinformatics*
885 *(Oxford, England)*, *14*(1), 68–73.
886 Huson, D. H., & Bryant, D. (2006). Application of phylogenetic networks in evolutionary studies.
887 *Molecular Biology and Evolution*, *23*(2), 254–267. doi: 10.1093/molbev/msj030
888 Jakobsson, M., & Rosenberg, N. A. (2007). CLUMPP: a cluster matching and permutation
889 program for dealing with label switching and multimodality in analysis of population
890 structure. *Bioinformatics*, *23*(14), 1801–1806. doi: 10.1093/bioinformatics/btm233
891 Jombart, T. (2008). adegenet: a R package for the multivariate analysis of genetic markers.
892 *Bioinformatics*, *24*(11), 1403–1405. doi: 10.1093/bioinformatics/btn129
893 Jombart, T., & Ahmed, I. (2011). adegenet 1.3-1: new tools for the analysis of genome-wide SNP
894 data. *Bioinformatics*, *27*(21), 3070–3071. doi: 10.1093/bioinformatics/btr521
895 Jürgens, A., Witt, T., & Gottsberger, G. (1996). Reproduction and pollination in central European
896 populations of *Silene* and *Saponaria* species. *Botanica Acta*, *109*(4), 316–324. doi:
897 10.1111/j.1438-8677.1996.tb00579.x
898 Kaltz, O., Gandon, S., Michalakis, Y., & Shykoff, J. A. (1999). Local maladaptation in the
899 anther-smut fungus *Microbotryum violaceum* to its host plant *Silene latifolia*: evidence
900 from a cross-inoculation experiment. *Evolution; International Journal of Organic*
901 *Evolution*, *53*(2), 395–407. doi: 10.1111/j.1558-5646.1999.tb03775.x
902 Keenan, K., McGinnity, P., Cross, T. F., Crozier, W. W., & Prodöhl, P. A. (2016). diveRsity: An
903 R package for the estimation and exploration of population genetics parameters and their
904 associated errors. *Methods in Ecology and Evolution*, 782–788. doi: 10.1111/2041-
905 210X.12067@10.1111/(ISSN)2041-210X.ecologyandevolutionireland
906 Kemler, M., Göker, M., Oberwinkler, F., & Begerow, D. (2006). Implications of molecular
907 characters for the phylogeny of the Microbotryaceae (Basidiomycota: Urediniomycetes).
908 *BMC Evolutionary Biology*, *6*, 35. doi: 10.1186/1471-2148-6-35
909 Kephart, S., Reynolds, R. J., Rutter, M. T., Fenster, C. B., & Dudash, M. R. (2006). Pollination
910 and seed predation by moths on *Silene* and allied Caryophyllaceae: evaluating a model
911 system to study the evolution of mutualisms. *New Phytologist*, *169*(4), 667–680. doi:
912 10.1111/j.1469-8137.2005.01619.x
913 Lafuma, L., & Maurice, S. (2006). Reproductive characters in a gynodioecious species, *Silene*
914 *italica* (Caryophyllaceae), with attention to the gynomonoecious phenotype. *Biological*
915 *Journal of the Linnean Society*, *87*(4), 583–591. doi: 10.1111/j.1095-8312.2006.00597.x
916 Lahiani, E., Dufaÿ, M., Castric, V., Le Cadre, S., Charlesworth, D., Van Rossum, F., & Touzet,
917 P. (2013). Disentangling the effects of mating systems and mutation rates on cytoplamic
918 diversity in gynodioecious *Silene nutans* and dioecious *Silene otites*. *Heredity*, *111*(2),
919 157–164. doi: 10.1038/hdy.2013.32
920 Laine, A.-L. (2005). Spatial scale of local adaptation in a plant-pathogen metapopulation. *Journal*
921 *of Evolutionary Biology*, *18*(4), 930–938. doi: 10.1111/j.1420-9101.2005.00933.x
922 Laine, A.-L. (2008). Temperature-mediated patterns of local adaptation in a natural plant–
923 pathogen metapopulation. *Ecology Letters*, *11*(4), 327–337. doi: 10.1111/j.1461-
924 0248.2007.01146.x
925 Laine, A.-L., Barrès, B., Numminen, E., & Siren, J. P. (2019). Variable opportunities for
926 outcrossing result in hotspots of novel genetic variation in a pathogen metapopulation.
927 *ELife*, *8*. doi: 10.7554/eLife.47091
928 Langmead, B., Trapnell, C., Pop, M., & Salzberg, S. L. (2009). Ultrafast and memory-efficient
929 alignment of short DNA sequences to the human genome. *Genome Biology*, *10*(3), R25.
930 doi: 10.1186/gb-2009-10-3-r25

931 Le Gac, M., Hood, M. E., Fournier, E., & Giraud, T. (2007). Phylogenetic evidence of host-
932 specific cryptic species in the anther smut fungus. *Evolution; International Journal of*
933 *Organic Evolution*, *61*(1), 15–26. doi: 10.1111/j.1558-5646.2007.00002.x
934 Le Gac, M., Hood, M. E., & Giraud, T. (2007). Evolution of reproductive isolation within a
935 parasitic fungal species complex. *Evolution*, *61*(7), 1781–1787. doi: 10.1111/j.1558-
936 5646.2007.00144.x
937 Linde, C. C., Zhan, J., & McDonald, B. A. (2002). Population structure of *Mycosphaerella*
938 *graminicola*: from lesions to continents. *Phytopathology*, *92*(9), 946–955. doi:
939 10.1094/PHYTO.2002.92.9.946
940 Liu, S., Cornille, A., Decroocq, S., Tricon, D., Chague, A., Eyquard, J.-P., … Decroocq, V.
941 (2019). From hybrid speciation of wild Armeniaca species to multiple domestication
942 events in Eastern Asia: the complex evolutionary history of apricots. *Under Review*.
943 López-Villavicencio, M., Jonot, O., Coantic, A., Hood, M. E., Enjalbert, J., & Giraud, T. (2007).
944 Multiple infections by the anther smut pathogen are frequent and involve related strains.
945 *PLOS Pathogens*, *3*(11), e176. doi: 10.1371/journal.ppat.0030176
946 Lutz, M., Göker, M., Piatek, M., Kemler, M., Begerow, D., & Oberwinkler, F. (2005). Anther
947 smuts of Caryophyllaceae: Molecular characters indicate host-dependent species
948 delimitation. *Mycological Progress*, *4*(3), 225–238. doi: 10.1007/s11557-006-0126-4
949 Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads.
950 *EMBnet.Journal*, *17*(1), 10–12. doi: 10.14806/ej.17.1.200
951 Martin, H., Touzet, P., Dufay, M., Godé, C., Schmitt, E., Lahiani, E., … Rossum, F. V. (2017).
952 Lineages of *Silene nutans* developed rapid, strong, asymmetric postzygotic reproductive
953 isolation in allopatry. *Evolution*, *71*(6), 1519–1531. doi: 10.1111/evo.13245
954 Martin, H., Touzet, P., Van Rossum, F., Delalande, D., & Arnaud, J.-F. (2016). Phylogeographic
955 pattern of range expansion provides evidence for cryptic species lineages in *Silene nutans*
956 in Western Europe. *Heredity*, *116*(3), 286–294. doi: 10.1038/hdy.2015.100
957 McCoy, K. D., Boulinier, T., & Tirard, C. (2005). Comparative host–parasite population
958 structures: disentangling prospecting and dispersal in the black-legged kittiwake Rissa
959 tridactyla. *Molecular Ecology*, *14*(9), 2825–2838. doi: 10.1111/j.1365-
960 294X.2005.02631.x
961 McDonald, B. A., & Stukenbrock, E. H. (2016). Rapid emergence of pathogens in agro-
962 ecosystems: global threats to agricultural sustainability and food security. *Phil. Trans. R.*
963 *Soc. B*, *371*(1709), 20160026. doi: 10.1098/rstb.2016.0026
964 McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., … DePristo,
965 M. A. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing
966 next-generation DNA sequencing data. *Genome Research*, *20*(9), 1297–1303. doi:
967 10.1101/gr.107524.110
968 Michalakis, Y., Sheppard, A. W., Noel, V., & Olivieri, I. (1993). Population structure of a
969 herbivorous insect and its host plant on a microgeographic scale. *Evolution*, *47*(5), 1611–
970 1616. doi: 10.2307/2410172
971 Naciri, Y., Pasquier, P.-E. D., Lundberg, M., Jeanmonod, D., & Oxelman, B. (2017). A
972 phylogenetic circumscription of *Silene* sect. *Siphonomorpha* (Caryophyllaceae) in the
973 Mediterranean Basin. *Taxon*, *66*(1), 91–108. doi: 10.12705/661.5
974 Nei, M. (1972). Genetic distance between populations. *The American Naturalist*, *106*(949), 283–
975 292. doi: 10.1086/282771
976 Nieberding, C. M., Durette-Desset, M.-C., Vanderpoorten, A., Casanova, J. C., Ribas, A.,
977 Deffontaine, V., … Michaux, J. R. (2008). Geography and host biogeography matter for

understanding the phylogeography of a parasite. *Molecular Phylogenetics and Evolution*, *47*(2), 538–554. doi: 10.1016/j.ympev.2008.01.028

Nieberding, C. M., & Olivieri, I. (2007). Parasites: proxies for host genealogy and ecology? *Trends in Ecology & Evolution*, *22*(3), 156–165. doi: 10.1016/j.tree.2006.11.012

Paradis, E., Claude, J., & Strimmer, K. (2004). APE: Analyses of Phylogenetics and Evolution in R language. *Bioinformatics*, *20*(2), 289–290. doi: 10.1093/bioinformatics/btg412

Petit, E., Silver, C., Cornille, A., Gladieux, P., Rosenthal, L., Bruns, E., … Hood, M. E. (2017). Co-occurrence and hybridization of anther-smut pathogens specialized on *Dianthus* hosts. *Molecular Ecology*, *26*(7), 1877–1890. doi: 10.1111/mec.14073

Piry, S., Luikart, G., & Cornuet, J. M. (1999). *BOTTLENECK: A computer program for detecting recent reductions in the effective population size using allele frequency data*. *90*, 502–503.

Poulin, R. (2005). *Parasite biodiversity*. Smithsonian Institution.

Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, *155*(2), 945–959.

Pudlo, P., Marin, J.-M., Estoup, A., Cornuet, J.-M., Gautier, M., & Robert, C. P. (2016). Reliable ABC model choice via random forests. *Bioinformatics*, *32*(6), 859–866. doi: 10.1093/bioinformatics/btv684

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., … Sham, P. C. (2007). PLINK: A tool set for whole-genome association and population-based linkage analyses. *The American Journal of Human Genetics*, *81*(3), 559–575. doi: 10.1086/519795

Rameau, J.-C., Mansion, D., & Dumé, G. (1989). *Flore forestière française: Plaines et collines*. Forêt privée française.

Rameau, J.-C., Mansion, D., & Dumé, G. (2008). *Flore forestière française: guide écologique illustré. Région méditerranéenne*. Forêt privée française.

Raynal, L., Marin, J.-M., Pudlo, P., Ribatet, M., Robert, C. P., & Estoup, A. (2019). ABC random forests for Bayesian parameter inference. *Bioinformatics*, *35*(10), 1720–1728. doi: 10.1093/bioinformatics/bty867

Refrégier, G., Le Gac, M., Jabbour, F., Widmer, A., Shykoff, J. A., Yockteng, R., … Giraud, T. (2008). Cophylogeny of the anther smut fungi and their caryophyllaceous hosts: prevalence of host shifts and importance of delimiting parasite species for inferring cospeciation. *BMC Evolutionary Biology*, *8*, 100. doi: 10.1186/1471-2148-8-100

Saleh, D., Milazzo, J., Adreit, H., Fournier, E., & Tharreau, D. (2014). South-East Asia is the center of origin, diversity and dispersion of the rice blast fungus, *Magnaporthe oryzae*. *New Phytologist*, *201*(4), 1440–1456. doi: 10.1111/nph.12627

Stukenbrock, E. H., Banke, S., & McDonald, B. A. (2006). Global migration patterns in the fungal wheat pathogen *Phaeosphaeria nodorum*. *Molecular Ecology*, *15*(10), 2895–2904. doi: 10.1111/j.1365-294X.2006.02986.x

Stukenbrock, E. H., & McDonald, B. A. (2008). The origins of plant pathogens in agro-ecosystems. *Annual Review of Phytopathology*, *46*(1), 75–100. doi: 10.1146/annurev.phyto.010708.154114

Szpiech, Z. A., Jakobsson, M., & Rosenberg, N. A. (2008). ADZE: a rarefaction approach for counting alleles private to combinations of populations. *Bioinformatics*, *24*(21), 2498–2504. doi: 10.1093/bioinformatics/btn478

Taylor, D. R., & Keller, S. R. (2007). Historical range expansion determines the phylogenetic diversity introduced during contemporary species invasion. *Evolution*, *61*(2), 334–345.

doi: 10.1111/j.1558-5646.2007.00037.x

Tellier, A., de Vienne, D. M., Giraud, T., Hood, M. E., & Refrégier, G. (2010). Theory and examples of reciprocal influence between hosts and pathogens, from short-term to long term interactions: coevolution, cospeciation and pathogen speciation following host shifts. In *Immunology and Immune System Disorders*. *Host-Pathogen Interactions: Genetics, Immunology and Physiology* (Nova Science Publishers, pp. 37–77). NY: Barton AW.

Thines, M. (2019). An evolutionary framework for host shifts – jumping ships for survival. *New Phytologist*, *224*(2), 605–617. doi: 10.1111/nph.16092

Thompson, J. N. (2005). *The geographic mosaic of coevolution*. University of Chicago Press.

Thrall, P., Biere, A., & Antonovics, J. (1993). Plant life-history and disease susceptibility - the occurrence of Ustilago-Violacea on different species within the Caryophyllaceae. *Journal of Ecology*, *81*(3), 489–498. doi: 10.2307/2261527

Tison, J., & de Foucault, B. (2014). *Flora gallica: flore de France* (Biotope).

Toh, S. S., & Perlin, M. H. (2016). Resurgence of less-studied smut fungi as models of phytopathogenesis in the omics age. *Phytopathology*, *106*(11), 1244–1254. doi: 10.1094/PHYTO-02-16-0075-RVW

Tsai, Y.-H. E., & Manos, P. S. (2010). Host density drives the postglacial migration of the tree parasite, Epifagus virginiana. *Proceedings of the National Academy of Sciences*, *107*(39), 17035–17040. doi: 10.1073/pnas.1006225107

Tutin, T., Heywood, V., Burges, N., Valentine, D., Walters, S., & Webb, D. (2001). *Flora Europaea*. Royaume-Uni: Cambridge University Press.

Van Rossum, F., Martin, H., Le Cadre, S., Brachi, B., Christenhusz, M. J. M., & Touzet, P. (2018). Phylogeography of a widely distributed species reveals a cryptic assemblage of distinct genetic lineages needing separate conservation strategies. *Perspectives in Plant Ecology, Evolution and Systematics*, *35*, 44–51. doi: 10.1016/j.ppees.2018.10.003

Vercken, E., Fontaine, M. C., Gladieux, P., Hood, M. E., Jonot, O., & Giraud, T. (2010). Glacial refugia in pathogens: European genetic structure of anther smut pathogens on *Silene latifolia* and *Silene dioica*. *PLOS Pathogens*, *6*(12), e1001229. doi: 10.1371/journal.ppat.1001229

Wegmann, D., Leuenberger, C., Neuenschwander, S., & Excoffier, L. (2010). ABCtoolbox: a versatile toolkit for approximate Bayesian computations. *BMC Bioinformatics*, *11*(1), 116. doi: 10.1186/1471-2105-11-116

Wickham, H. (2009). *ggplot2: elegant graphics for data analysis*. Springer Science & Business Media.

Wilson, D. J., Falush, D., & McVean, G. (2005). Germs, genomes and genealogies. *Trends in Ecology & Evolution*, *20*(1), 39–45. doi: 10.1016/j.tree.2004.10.009

Wolfe, N. D., Dunavan, C. P., & Diamond, J. (2007). Origins of major human infectious diseases. *Nature*, *447*(7142), 279–283. doi: 10.1038/nature05775

Zaffarano, P. L., McDonald, B. A., & Linde, C. C. (2008). Rapid speciation following recent host shifts in the plant pathogenic fungus *Rhynchosporium*. *Evolution*, *62*(6), 1418–1436. doi: 10.1111/j.1558-5646.2008.00390.x

Data accessibility

1068 The plant and fungal microsatellite and chloroplastic genotypes are available at dryad XXX (to
1069 be completed upon acceptance). We deposited the genome raw data at NCBI Sequence Read
1070 Archive (SRA) under the BioProject xxx (ID available upon acceptance).

1071

1072 **Author contributions**

1073 TG and FH conceived and designed the project with the help of FVR. TG, PT, FVR, JS and EF
1074 collected samples. AS, CG, PT, SLP and FH genotyped the samples. FH and AC performed the
1075 analyses. FH, TG and FVR wrote the manuscript. All authors read and approved the final version
1076 of the manuscript and declare no conflict of interest.

1077

1078

1079 **Supporting information**

1080 Additional Supporting Information may be found in the online version of this article.

1081

1082

Tables

Table 1: Comparisons of scenarios and group of scenarios used for approximate Bayesian computation (ABC) to reconstruct the fungi divergence history. Scenarios are described in Table S5 and Fig S5. Results of ABC random forest procedure are shown, with the percentage of vote for each scenario/group of scenarios, the posterior probability and the prior error rate.

ABC round	Tested hypothesis	Number of tested scenarios/group of scenarios	Description of tested scenarios/group of scenarios	Best scenario/group of scenarios	Percentage of vote	Posterior probability	Prior error rate in %
Round 1: "time of divergence"	divergence time T1, T2, T3 or T4	4 groups of 3 scenarios	**group 1 (time T1):** divergence-order-A-no-GF-T1, divergence-order-B-no-GF-T1, divergence-order-C-no-GF-T1		1	0.54 (+- 0.02)	34%
			group 2 (time T2): divergence-order-A-no-GF-T2, divergence-order-B-no-GF-T2, divergence-order-C-no-GF-T2		5		
			group 3 (time T3): divergence-order-A-no-GF-T3, divergence-order-B-no-GF-T3, divergence-order-C-no-GF-T3		43		
			group 4 (time T4): divergence-order-A-no-GF-T4, divergence-order-B-no-GF-T4, divergence-order-C-no-GF-T4	x	51		
Round 2: "divergence order"	Order of divergence A, B or C	3 scenarios	**scenario 1 (order A):** divergence-order-A-no-GF-T4	x	80	0.79 (+- 0.02)	2%
			scenario 2 (order B): divergence-order-B-no-GF-T4		5		
			scenario 3 (order C): divergence-order-C-no-GF-T4		15		
Round 3: "occurence of gene flow"	step1: absence of gene flow, or gene flow between nutans clusters only or gene flow with the outgroup	3 groups of 1, 6 and 6 scenarios	**group 1 (no gene flow):** divergence-order-A-no-GF-T4		33	0.60 (+- 0.02)	5%
			group 2 (gene flow between nutans clusters only): divergence-order-A-T4-GF1, divergence-order-A-T4-GF3, divergence-order-A-T4-GF5, divergence-order-A-T4-GF6, divergence-order-A-T4-GF10, divergence-order-A-T4-GF12	x	42		
			group 3 (gene flow with the outgroup): divergence-order-A-T4-GF2, divergence-order-A-T4-GF4, divergence-order-A-T4-GF7, divergence-order-A-T4-GF8, divergence-order-A-T4-GF9, divergence-order-A-T4-GF11		24		
	step2: continuous gene flow, secondary contact or ancient migration	3 groups of 3, 2 and 4 scenarios	**group 1 (continuous gene flow):** divergence-order-A-T4-GF1, divergence-order-A-T4-GF5, divergence-order-A-T4-GF6		23	0.76 (+- 0.02)	35%
			group 2 (secondary contact): divergence-order-A-T4-GF3, divergence-order-A-T4-GF3b		3		
			group 3 (ancient migration): divergence-order-A-T4-GF10, divergence-order-A-T4-GF10b, divergence-order-A-T4-GF12, divergence-order-A-T4-GF12b	x	74		
	step3: absence of gene flow vs ancient migration	2 groups of 1 and 4 scenarios	**group 1 (no gene flow):** divergence-order-A-no-GF-T4		34	0.66 (+- 0.02)	2.50%
			group 2 (ancient migration): divergence-order-A-T4-GF10, divergence-order-A-T4-GF10b, divergence-order-A-T4-GF12, divergence-order-A-T4-GF12b	x	66		
	step4: absence of gene flow vs secondary contact	2 groups of 1 and 3 scenarios	**group 1 (no gene flow):** divergence-order-A-no-GF-T4	x	99	0.99 (+- 0.004)	0.76%
			group 2 (secondary contact): divergence-order-A-T4-GF3, divergence-order-A-T4-GF3b		1		
	step5: comparison of different scenarios of ancient migration	4 scenarios	**scenario 1:** divergence-order-A-T4-GF10	x	29	0.49 (+- 0.03)	60%
			scenario 2: divergence-order-A-T4-GF10b		24		
			scenario 3: divergence-order-A-T4-GF12		26		
			scenario 4: divergence-order-A-T4-GF12b		21		

Figures

Figure 1: Population genetic structure of *Silene nutans* and *S. italica* host plants based on six plastid (chloroplast) SNPs and 21 microsatellite markers. A. Plastid haplotypes according to their geographic distribution. On the right, a zoom of the map in the northeastern region of France and Belgium is shown. The size of the symbols is proportional to the number of individuals sampled by site (1-9 plants). B. Principal component analysis (PCA), with information of plastid haplotypes. The first and second principal component axes are shown and the percentage of variance explained by each axis is indicated into brackets. C. Results of STRUCTURE on both host plant species for K=5 clusters. On panels A and B, symbol shape indicates species and color indicates plastid haplotypes. The two genetic clusters identified within the orange plastid haplotype based on the STRUCTURE analysis (panel C) are shown with light and dark orange colors, respectively.

Figure 2: Population genetic structure of anther-smut fungi (*Microbotryum*) parasitizing *Silene nutans* and *S. italica* based on 22 microsatellite markers. A. Anther-smut fungi parasitizing *S. italica* (top; photo credit M.E. Hood) and *S. nutans* (bottom; photo credit M. Strack van Schijndel). B. Principal component analysis (PCA). The first and second principal component axes are shown and the percentage of variance explained by each axis is indicated into brackets. The symbol shape indicates the sampling host species. The dark and light blue colors indicate assignment probability >0.80 to the two corresponding clusters identified in the STRUCTURE analysis for K=2 (see panel C). Pink color indicates strains for which no cluster could be assigned with a probability >0.80. Genetic variance explained by the principal component axes is shown in the right bottom corner. C. STRUCTURE barplot for K=2. Strains are ordered according to their sampling host species.

Figure 3: Congruence of population genetic structure between *Silene nutans* host plants and their anther-smut fungal (*Microbotryum*) strains based on microsatellite markers. A. Principal component analysis (PCA) on the fungal strains with information of the population genetic structure of their hosts. The first and second principal component axes are shown. Symbol shape indicates the sampling host species and color indicates the plastid haplotypes of the sampling host. B. STRUCTURE barplots for (1) K=3 in 112 *Microbotryum* strains and (2) K=4 in 110 *S. nutans* plant individuals. Fungal strains are ordered according to their sampling host plastid haplotypes. Host plant individuals are ordered according to their plastid haplotypes. C. Map of mean cluster membership proportions per site for (1) the 112 fungal strains in the STRUCTURE analysis at K=3 and (2) the 110 *S. nutans* plant individuals in the STRUCTURE analysis at K=4. The pie diameter reflects the sample size in the corresponding site (1 - 14 individuals).

Figure 4: Most likely scenario for the divergence between the fungal genetic clusters inferred with approximate Bayesian computation (ABC).

Figure 5: Absence of genome-wide signature of recent gene flow among anther-smut fungal strains parasitizing closely related *Silene* species. A. Principal component analysis (PCA) based on 1,305,369 genome-wide SNPs. B. STRUCTURE analyses based on 233 unlinked SNPs for K=5. The Y axis indicates the estimated membership proportions in the *K* clusters for each fungal strain (X axis). C. Neighbornet tree from a SplitsTree analysis based on 595,002 genome-wide SNPs with no missing data and heterozygote genotypes. The inner plot shows a zoom of the tip of the tree for strains parasitizing *S. nutans*. Information on the species or plastid haplotype of the host of sampling of the strains is given.

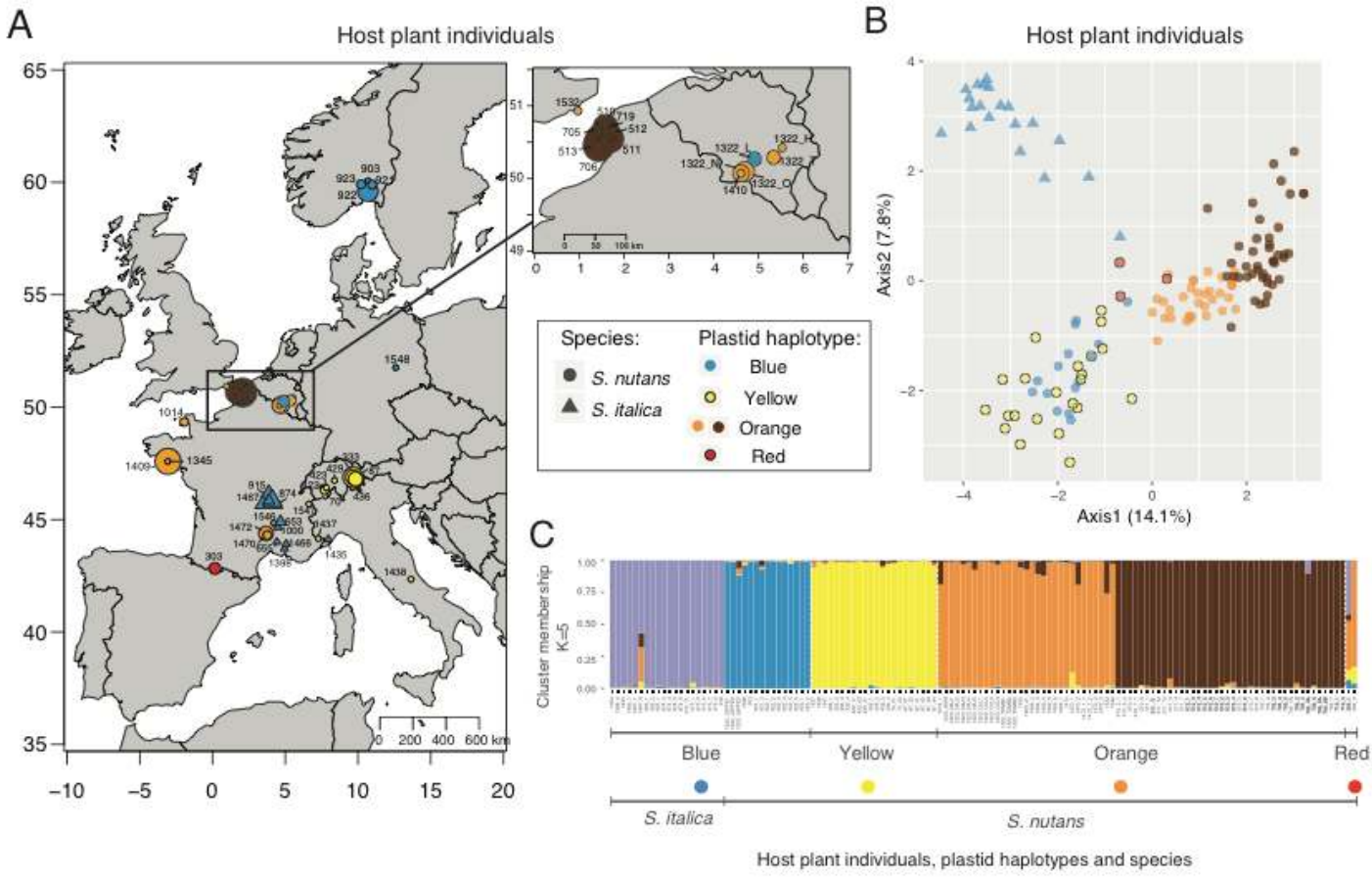

Figure 1

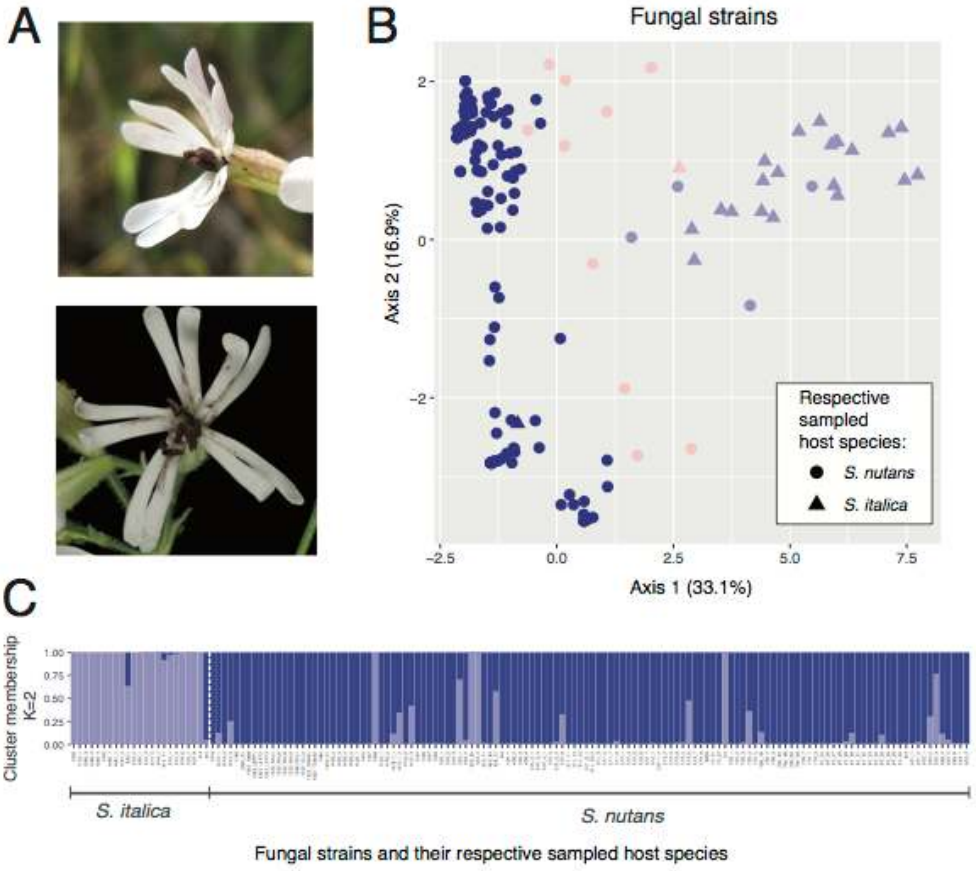

Figure 2

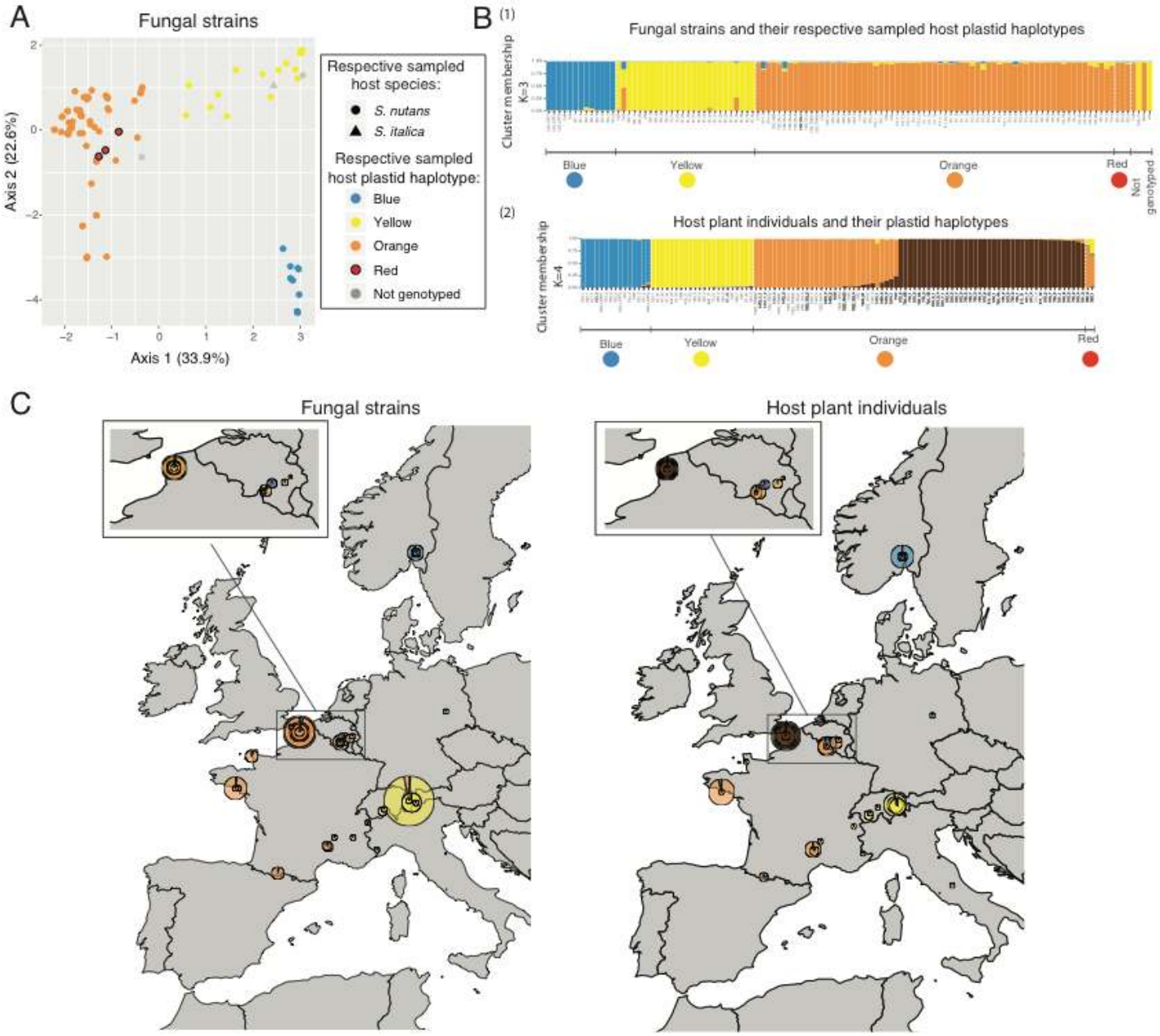

Figure 3

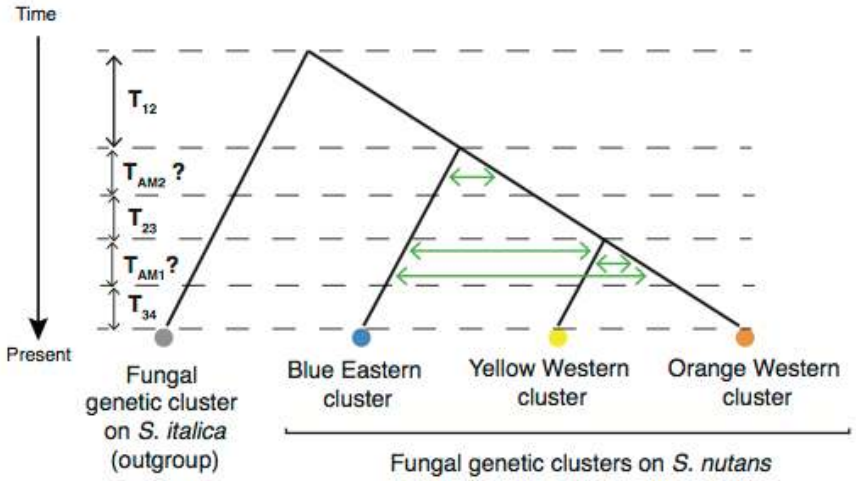

Figure 4

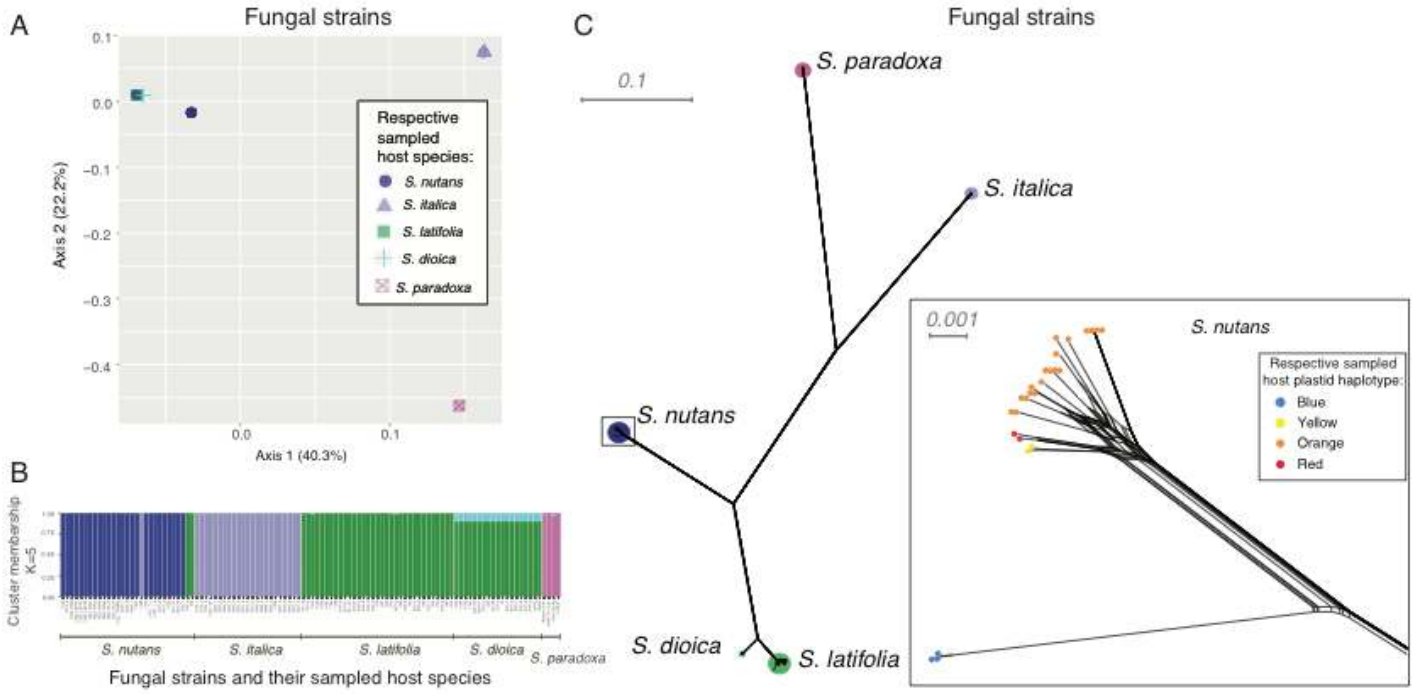

Figure 5