Understanding Adaptation, Coevolution, Host Specialization, and Mating System in Castrating Anther-Smut Fungi by Combining Population and Comparative Genomics

Fanny E Hartmann, Ricardo Rodriguez de La Vega, Fantin E Carpentier, Pierre Gladieux, Amandine Cornille, Michael E Hood, Tatiana Giraud

Understanding Adaptation, Coevolution, Host Specialization and Mating System by combining Population and Comparative Genomics in Castrating Anther-Smut Fungi

Fanny E. Hartmann[1], Ricardo C. Rodríguez de la Vega[1], Fantin Carpentier[1], Pierre Gladieux[2], Amandine Cornille[3], Michael E. Hood[4], Tatiana Giraud[1]

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

[2] INRA, UMR BGPI, Bâtiment K; Campus International de Baillarguet, F-34398, Montpellier, France

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

[4] Biology Department, Science Center Building, Amherst College, Rts 9 & 116, Amherst, MA USA 01002-5000

Corresponding author: fanny.hartmann@u-psud.fr.

Abstract

Anther-smut fungi constitute a powerful system to study host-pathogen specialization and coevolution, with hundreds of *Microbotryum* species specialized on diverse Caryophyllaceae plants, castrating their hosts through particular manipulation of hosts' reproductive organs that facilitates disease transmission. *Microbotryum* fungi also have exceptional genomic traits, including dimorphic mating-type chromosomes, that make this genus also an excellent model for the evolution of mating systems and their influence on population-genetic structure and adaptive potential. Important insights into the adaptation, coevolution, host specialization and mating system evolution have been gained using anther-smut fungi, in particular with the recent advent of genomic approaches. We argue and illustrate based on the *Microbotryum* case studies that using a combination of genomic analyses is a powerful approach, where comparative genomics, population genomics and transcriptomics data allow the integration of different evolutionary perspectives and across timescales. We also highlight current challenges and future studies that will contribute to advance our understanding of mechanisms involved in adaptive processes in fungal pathogen populations.

Keywords : comparative genomics, population genomics, transcriptomics, adaptation, positive selection, selective sweeps, divergence, gene flow, rearrangements, suppressed recombination

Introduction

Pathogens thrive using living organisms as nutritional resources, which reduces their host fitness. This leads to coevolutionary arms races, in which pathogens are selected for increased abilities of host infection and exploitation, while hosts are selected for mechanisms of resistance to particular diseases. Such coevolution occurs on short evolutionary scales, as a never-ending process of adaptation and counter-adaptation (113). Across macro-evolutionary scales, some pathogens also may undergo host shifts, forming new species by specialization in combination with new hosts (42). Coevolution is a very different evolutionary process from host specialization, despite the terms often being used interchangeably, and may involve different genomic mechanisms and/or molecular interactions that have yet to be well resolved (42).

An integrated understanding of the ecological and genetic/genomic mechanisms underlying both coevolution and host specialization by pathogens is of fundamental importance. These phenomena indeed represent cases of rapid adaptation, diversification and long-term species interactions, shedding light on the processes generating and maintaining biodiversity and ecosystem dynamics. Furthermore, knowledge on the genomic mechanisms involved in coevolution and host specialization in fungal pathogens is important for controlling crop and animal diseases and preventing emerging diseases that are a rising threat in domestic and wild populations (52, 70). Fungi are the most important plant pathogens, causing dramatic crop diseases, including many devastating diseased that are newly emergent following host shifts (9, 45, 51).

Fungal pathogens also have to cope with their abiotic environment, such as temperature and humidity (2, 35, 44, 47, 129). Understanding the mechanisms of adaptation to climatic variables is thus similarly of fundamental and applied interest. Adaptation ability is however impacted by genetic diversity and gene flow, which are themselves influenced by dispersal rates and mating systems, that are therefore important life history traits to study for an integrated understanding of evolution, adaptation, population subdivision and speciation (18, 60, 63).

From the advent of modern genetics a century ago, anther-smut fungi (*Microbotryum violaceum* species complex, previously *Ustilago violacea*) have served as useful models for the molecular controls of mating and adaptations to abiotic conditions (1, 17, 23, 29, 61, 62, 73, 93, 123, 124). With advances in population genetics and genomics, emphasis has grown with regard to the natural diversity within this pathogen group, the dynamics of diseases, the mating systems and genetic differentiation in relation to host plants in natural ecosystems (2, 27, 32, 38, 39, 58, 65, 67, 97, 105, 111, 131, 132). The anther-smut fungi belong to the *Microbotryum* genus (basidiomycetes), which castrate plants of the Caryophyllaceae family, replacing the pollen with their spores and aborting ovaries (Figure 1B). They constitute an excellent model pathosystem, with hundreds of closely related fungal species specialized on different host plants, resulting from numerous host shifts, with conspicuous symptoms, a rich scientific history and occurring in natural ecosystems (Figure 1A) (83, 92, 97, 111, 118). Furthermore, they are phylogenetically close to the rust fungi as damaging crop pathogens (127). Most *Microbotryum* species are highly host-specific, but a few are more generalist, parasitizing closely related host species (Figure 1A) (98, 105). Other *Microbotryum* species, while distantly related, co-occur on the same host species, representing cases of convergence (Figure 1A) (2, 97).

Host-pathogen coevolution in the *Microbotryum* model systems has been suggested based on patterns of plant local adaptation (43, 49) and congruent plant-pathogen genetic structure (49). *Microbotryum* species show little pre-zygotic isolation and increasing post-zygotic isolation strength with phylogenetic distance (28, 41, 98), which may allow gene flow among closely related species. Moreover, abiotic factors have been shown to play a role with the disease interactions in important ways (2), and *Microbotryum* fungi display an interesting mating system, with predominant automixis (i.e., intra-tetrad selfing), which has fostered multiple chromosomal rearrangements across the genus linking the mating-type loci controlling gamete compatibility (25, 26). A consideration of these features altogether allows the studying adaptation, coevolution, host specialization, differentiation and mating systems with unique power.

For tackling this complex suite of questions, comparative genomics and population genomics constitute highly relevant and complementary approaches, addressing different time scales of evolution. In contrast to life history traits, ecology and population structure, which have been extensively studied (7, 10, 22, 90), the genetic basis of interactions between *Microbotryum* fungi and their hosts is still little known; genomic approaches can elucidate the mechanisms and the functions involved in adaptation, coevolution, host specialization and speciation in this pathosystem. Analyses of gene expression between different stages of the life cycle can also inform on these processes. In particular, the pathogen's mating systems also influence adaptation, coevolution and host specialization (63, 66), especially in anther-smut fungi that are obligately completing the sexual cycle upon every disease transmission. Genomics can further help to understand the evolution of mating systems by studying the changes at the mating-type loci.

In this review, we discuss the recent insights into our understanding of adaptation, coevolution and host specialization in anther-smut fungi gained from gene expression data and comparative genomics (part 1) and from population genomics (parts 2 & 3). We then discuss insights gained from genomics on mating system evolution (part 4). We illustrate that the combination of multiple genomic approaches is needed for a full understanding of evolution, as comparative genomics, population genomics and transcriptomics address different timescales and have power for detecting different footprints of adaptive events. Finally, future challenges to be addressed using genomics tools are discussed (part 5).

1- COMPARATIVE GENOMICS AND TRANSCRIPTOMICS APPROACHES TO UNDERSTAND ADAPTATION AND HOST SPECIALIZATION IN ANTHER-SMUT FUNGI

The sequencing of genomes and transcriptomes of *Microbotryum* species sheds light on pathogenicity, adaptation and specialization mechanisms across long evolutionary timescales; speciation events in castrating *Microbotryum* fungi have been dated from 0.4 to 11 MYA (26, 72) (Figure 2A). Phylogenomics enables obtaining an accurate understanding of the lineage histories, and comparative genomics is highly suitable to identify genetic changes associated with diversification at such large evolutionary scales.

Genome architecture and identification of candidate genes involved in pathogenicity using expression data

One of the best studied anther-smut species is *M. lychnidis-dioicae*, parasitizing the white campion *Silene latifolia* (Figure 1B). The diploid genomes of the Lamole *M. lychnidis-dioicae* strain was the first eukaryote genome to be assembled with new sequencing technologies (16).

Comparative analysis of the Lamole strain of *M. lychnidis-dioicae* with other basidiomycetes genomes revealed specific gene content features such as the absence of plant cell wall degrading enzymes and expanded repertoires of major facilitator superfamily transporters, secretory lipases, glycosyltransferases and enzymes that could manipulate host development (104). Such features are likely related to the castrating and biotrophic lifestyle of anther-smut fungi (104), where the fungus takes up a largely symptomless residence between the host cells in the plant's growing points/meristems until the host initiates flower development. Additionally, this pathogen has a remarkable ability to developmentally transform female host plants to take on a male-like floral structure, with the growth of stamens that then bare spores in place of pollen and the abortion of the ovary early in its development (13). Apart from the accumulation of transposable elements (TEs) in the non-recombining regions of the mating-type chromosomes, there was no genome compartmentalization into more or less repeat-rich regions on autosomes (16, 104), in contrast to some other fungal pathogens with isochore genomic architecture and localization of effector genes in repeat-rich regions (74). Nevertheless, transposable elements were locally associated across the Lamole *M. lychnidis-dioicae* genome with gene clusters of small secreted proteins and genes affected by within species presence-absence polymorphism, suggesting a role of transposable elements in genome rearrangements and duplications of genes putatively involved in host adaptation (80, 104). Although footprints typical of genome defense mechanisms against TEs, similar to repeat-induced point mutation (RIP), were identified in anther-smut genomes, a massive burst-like expansion of Gypsy-like retrotransposons in a *Microbotryum* strain suggested that persistent transposable elements activity and expansion can occur (86, 87).

Transcriptomics conducted at several *in vitro* stages allow detecting genes upregulated in certain conditions and thus likely involved in important functions at a given life stage.

Transcriptomic analyses using the Lamole *M. lychnidis-dioicae* strain enabled identifying genes likely associated with nutrient uptake, the mating program and the dikaryotic switch (54, 104, 125, 126, 136). *In silico* effector gene prediction combining *in planta* expression data, sequence conservation and predicted localization, allowed identifying small secreted proteins genes as candidate effectors, i.e. involved in pathogenicity, in *M. lychnidis-dioicae*, *M. silenes-dioicae* and *M. violaceum var paradoxa* (20, 96). Eight genes in *M. silenes-dioicae* and three genes in *M. violaceum var paradoxa* predicted to encode secreted proteins were further confirmed to be secreted using yeast secretion trap (20, 96). Compared expression data in male and female *S. latifolia* individuals during fungal infection revealed pathogen-mediated changes in sex-biased gene expression and altered sexual dimorphism in the host (137). Another transcriptome analysis of the early development stages of infected flowers detailed changes in gene expression in *M. lychnidis-dioicae,* identifying gene categories likely to manipulate the host development and reproductive system, such as potential effectors and virulence factors (125). Further coupling experiments of host and pathogen gene expression changes, and in further paired host-*Microbotryum* fungi, should help deciphering the major components of the tight host-pathogen interactions described in the system.

Comparative genomics studies within the *Microbotryum* genus

Comparative genomics among *Microbotryum* fungi, and with other plant pathogens, has provided insights into the specificity of castrating biotrophic pathogens growing intracellularly (115) relative to other forms of parasitic nutritional ecology. Comparative genomics among anther-smut fungi specialized on different hosts can help unravel the genomic determinants of host specificity as well as the shared pathogenicity mechanisms. Indeed, while substantial insights has been gained by the study of individual genomes of *Microbotryum* species, whether features such as the conspicuous lack of cell-wall degrading

187 enzymes in the *M. lychnidis-dioicae* Lamole genome are common to the genus cannot be
188 known without a comparative genomics analysis that addresses both distantly and closely-
189 related species (77).
190
191 Early comparative studies focused on orthologous genes across single pass Sanger-sequenced
192 cDNA libraries, i.e. expressed sequence tags, from four *Microbotryum* species. The primary
193 focus was looking for signals of positive selection in terms of frequent amino-acid changes
194 (4). A subset of the genes evolving under positive selection between species was further
195 shown to be under strong purifying selection within two closely-related *Microbotryum*
196 species, *M. lychnidis-dioicae* and *M. silenes-dioicae*, suggesting that adaptive changes
197 concomitant with host shifts can be later fixed due to strong functional constraints within
198 species (69). Although the inferred function of some of the orthologous groups with signals of
199 positive selection could be associated with aspects of virulence or speciation, none of these
200 displayed features of effectors (such as secretory signals), likely because the expressed
201 sequence tags did not exhaustively cover the genomes. Indeed, only 53 clusters of orthologs
202 shared by at least three species and at least 300 nucleotides long could be retrieved (4).
203 Therefore, even though these analyses demonstrated the utility of comparative genomics to
204 identify candidate genes for diversifying selection in non-model organisms, the lack of whole
205 genome sequences prevented any insight about presence-absence polymorphisms or
206 substitutions both known to be important for adaptation to new hosts.
207
208 The number of high-quality genomes assemblies or shotgun sequencing from *Microbotryum*
209 species/strains has exploded recently, reaching nearly a hundred as by late 2018 (Table 1;
210 Figure 1A; (15, 16, 25, 26, 30, 56, 112, 134)). In comparative genomics, near-complete gene
211 lists can be clustered to obtain groups of homologous sequences that can be then used to build

phylogenetic profiles of gene content. Such comparisons allow the identification of gene families that are species-specific and those that have been expanded or reduced in particular lineages (8, 76). Species- or population-specific genes are either derived from within-group innovation, a rather uncommon phenomenon (31), the result of differential losses or gene duplications (78), or due to the non-vertical acquisition of gene-coding genome fragments, for instance horizontal gene transfer (50). Expanded gene families require the escape from the rampant pseudogenization (non-functionalization) of duplicated genes (99), whereas reduced or complete losses of gene families is often related to ecological shifts (117), rendering the product of those genes no longer needed for survival (5). Understanding these processes is fundamental to the study of evolutionary ecology as they help to explain the genomic architecture underlying the phenomenon of adaptive divergence. *In silico* annotations and comparative analyses have identified hundreds of candidate effectors across multiple *Microbotryum* species, enriched in gene families showing presence-absence polymorphism across species (Figure 2B) (112), along with orthologous genes with landmarks of positive selection between species and purifying selection within species (20), thus generalizing and expanding previous findings. High-quality genome assemblies revealed little genomic rearrangements in autosomes (26).

Studies of positive selection based on the comparisons of non-synonymous and synonymous substitution rates (dN/dS; (135)) and on the comparisons of the proportions of non-synonymous and synonymous polymorphisms within species and differences between species (McDonald and Kreitman test; (100)) revealed no signature of diversifying selection between sister *Microbotryum* species specialized on two closely related host species (15), but detected a dozen of genes encoding secreted proteins with signs of positive selection between more distantly related *Microbotryum* species specialized on more distant host species (Figure 2C)

(20). Future comparative genomics studies encompassing all currently sequenced genomes will likely have high power to detect genes involved in host specialization by allowing further disentangling the effects of pathogen and host phylogenetic distances. In particular, combining population and comparative analyses should be very powerful to identify genes under diversifying selection between species and purifying selection within species as well as species-specific gene gains and losses. Building gene genealogies based on whole genomes also allowed to resolve previously ambiguous relationships among some anther-smut species (Figure 1A). The comparison of repeat contents and genomic rearrangements between genomes will be a further key step to understand the role of genome dynamics in adaptive processes in anther-smut fungi.

2-POPULATION GENOMICS TO IDENTIFY ADAPTIVE GENETIC VARIATION IN NATURAL PATHOGEN POPULATIONS

Population genomics is a complementary approach to comparative genomics for understanding adaptation in pathogen populations. Population genomics indeed address more recent adaptive events, and on a broader range of evolutionary genetic phenomena, not only gene gains/losses and recurrent changes in amino-acids. Selective sweeps can be detected using population genomics, which can reveal positive selection on a single amino-acid change or basepair substitutions in non-coding regions. Furthermore, population genomics can address the questions of the genomic bases of host-pathogen coevolution and local adaptation, that constitute more recent selection compared to the long-term selection underlying host specialization, and possibly differential selection among geographically distant populations (Figure 2A) (36, 75, 107). In contrast to major fungal-plant pathosystems, no gene-for-gene relationship has been reported for anther-smut fungi. Instead, the probability of infection shows quantitative variation (6, 7, 33), which suggests a rather complex genetic basis of co-

evolution and host local adaptation. Genome-wide population genomics approaches in anther-smut fungi allowed identification of the complex genetic basis of recent adaptive events through genome scans of selective sweeps and gene-presence absence polymorphism (3, 15, 80).

Selective sweep analyses allow one to identify loci that have recently been under positive selection within populations and thus likely underlying coevolution and local adaptation, whereas genes involved in host specialization are likely under purifying selection within species after the initial adaptive events following host shifts. Analyses of whole genome sequences of 53 genomes of the anther-smut sister species *M. lychnidis-dioicae* and *M. silenes-dioicae* identified selective sweeps (Figure 2D) (15), likely resulting from dynamic co-evolutionary arm race of the fungus with its hosts. The overlap between genes differentially expressed *in planta* and *in vitro* and those lying within selective sweeps, together with functional annotations, provided clues to genes and functions involved in plant-pathogen interaction in the *Microbotryum-Silene* system. Candidate genes included glycoside hydrolases, pectin lyases and an extracellular membrane protein with CFEM domain (15). The pectin lyase function seems relevant in that *Microbotryum* fungi grow between cells of the meristem (115), which is a pectin-occupied space. Extracellular membrane proteins with a cysteine-rich CFEM domain are present in effectors in several fungal pathogens (95). This study was also an opportunity to test for differences in intensity of coevolution between anther smut fungi on different hosts. Interestingly, differences in the number and the location of the selective sweeps were found between sister species. Footprints of positive selection affected 17 % of the genome in *M. lychnidis-dioicae* and 1 % of the genome in *M. silenes-dioicae* (15). Selective sweeps were scattered throughout the genomes. Linkage disequilibrium was found to decay relatively slowly with physical distance along chromosomes, as expected for selfing species, but still indicated effective recombination.

289 Polymorphism in each fungal species was negatively correlated with the recombination rates
290 along chromosomes, consistent with recurrent positive and/or background selection erasing
291 diversity on larger genomic regions when recombination is less frequent (15).
292
293 Population genomics can also contribute to our understanding of the impact of recent
294 anthopogenic factors on the genome and subsequent adaptation. Analyses of *M. lychnidis-*
295 *dioicae* genomes along a gradient of ionizing radiation levels around Chernobyl showed no
296 evidence of deleterious mutation accumulation in the form of non-synonymous substitutions
297 (3). Lower mean values of dN/dS were even found in Chernobyl compared to other areas of
298 the same eastern genetic cluster (3), which may be due to stronger selection in contaminated
299 areas against individuals bearing mildly deleterious mutations, i.e. stronger purifying
300 selection.
301
302 In addition to genome scans looking at signatures of positive selection, other population
303 genomic approaches make use of the genetic variation in pathogen populations to identify the
304 genomic architecture of local adaptation (19, 36, 107, 114). Population genomics enables the
305 unravelling the genomic bases of adaptation to abiotic conditions by searching for correlations
306 between local population allele frequencies and local environments (genetic-environment
307 association methods) (82). Such approaches can be used in anther-smut fungi along altitudinal
308 clines in Alpine populations on *Dianthus* or *Silene* hosts. Studies on the three species
309 parasitizing *S. vulgaris* in particular could be interesting as elevation and climate has been
310 shown to impact these anther-smut fungi (1, 2). Strong population structure as found in many
311 *Microbotryum* species at European scale (2, 15, 27, 55, 105) might be a challenge to the use
312 of such methods in particular, but these methods can be utilized at small geographical scales
313 and/or in species with less population subdivision.

Gene copy number variation segregating within species is also a widespread and an important source of genetic variation and several examples of adaptive evolution through gene loss or gene gain have been identified in agricultural fungal plant pathogens (57). Population genomics allow to explore the extent and adaptive potential of such within-species variation. Gene presence/absence polymorphism was found to contribute to the genetic variation in populations of the two closely related species of castrating anther-smut fungi, *M. lychnidis-dioicae* and *M. silenes-dioicae* (80). Genes displaying presence/absence polymorphism were mostly recently acquired, in a single species, through duplications in multiple-gene families and few genes predicted to encode secreted proteins were affected, suggesting defense against host recognition by other genetic changes than gene loss or gain. Although most gene presence/absence polymorphisms were likely neutral, the putative functions of some genes affected by presence–absence polymorphism (e.g., secreted proteins) or their localization within previously identified selective sweeps suggested that some gene loss or gain events may be adaptive (80).

3-INSIGHTS INTO THE DYNAMICS OF DIVERGENCE AND GENE FLOW FROM POPULATION GENOMICS

By providing a glimpse into intra-specific genetic diversity and its variation across the genome, population genomics analyses are also highly useful to understand processes underlying species divergence and phylogeography, quantifying rates of gene flow and its heterogeneity along genomes, and providing accurate estimation of population size variations. The occurrence of multiple *Microbotryum* sister species pairs in sympatry makes the system a perfect model to study the dynamics of divergence and gene flow in fungal pathogen populations.

Contrasted patterns of interspecific gene flow in the *Microbotryum* genus: a speciation continuum?

The two pathogens *M. lychnidis-dioicae* and *M. silenes-dioicae* and their respective sister host plants, *Silene latifolia* and *S. dioica*, are ubiquitous in Europe and their geographic distributions are largely overlapping, providing an ideal system for research on the formation and maintenance of species in sympatry. Microsatellite data from samples across Europe revealed rare disease transmission events between the host species and rare pathogen hybrids (72, 132). However, these approaches using a dozen microsatellite markers may lack power. Analyses of whole genome sequences of many pathogen samples that appeared of pure ancestry based upon the microsatellite data then revealed no evidence for admixture, indicating that introgression does not persist beyond one or two generations (15). In the laboratory, both fungal species can infect both host plants (40, 64, 98). Experimental crosses showed little premating isolation by assortative mating between the two pathogen species (28, 98, 132), even at sympatric sites (110), and a lack of post-mating barriers (41, 98). Hybrids were viable and fertile at least through the F2 generation in the greenhouse (40, 98, 131). F2 hybrids produced by selfed F1s had mostly returned to homozygosity, suggesting that genomic content derived from one of the two parental species had already begun to be purged (28, 40). This latter finding, combined with the fact that introgression does not appear to persist in nature, is consistent with strong genome-wide selection by the host plant and the scattering of genes involved in host specialization across the genome, as revealed in genome scans of selective sweeps (15). F_{ST} values were found near their maximum all along their genomes (Figure 3A).

Whereas strict host specialization is often the rule on *Silene* species (15, 97, 133), on *Dianthus* hosts in contrast population genetics approaches revealed four *Microbotryum*

lineages with broader and overlapping host specificities (Figure 1A) (97, 105). One *Microbotryum* lineage was found only on *D. pavonius* while the others occurred spread across several host species, some of them being shared among *Microbotryum* lineages. The sympatry of *Microbotryum* lineages within populations, in particular in the Alps, led to hybridization (105). The individuals with mixed ancestry based on clustering analyses of microsatellite data suffered from significant meiotic sterility, which confirmed they were hybrids between species (105). The larger host ranges of *Microbotryum* lineages on *Dianthus* hosts may be explained by the recent divergence of their host plants. The *Dianthus* genus has indeed undergone a recent radiation in Europe with morphologically diverse European *Dianthus* species restricted to small geographically restricted ranges (130). The full extent and evolutionary consequences of the hybridization on pathogen dynamics and evolution remains to be explored. Along this line, the *Dianthus*-*Microbotryum* system may become, in the coming years, a tractable model to investigate the impact of gene flow during divergence, and whether selection due to local/host adaptation can make some genomic regions more or less permeable to gene flow, which represents a current debate in evolutionary biology (37). These questions could not be addressed so far based on the population genomics analyses of *M. lychnidis-dioicae* and *M. silenes-dioicae* as no genomic introgression could be detected in natural populations (15). In contrast, the hybrids detected in natural populations on *Dianthus* hosts with significant sterility suggest the occurrence of introgressions (105). Other pairs of *Microbotryum* species might also be suitable to address these questions of the impact and heterogeneity of gene flow along the genome. For example, anther-smut fungi on the closely related and sympatric native American species *S. virginica* and *S. caroliniana* (11, 12) could not be separated into host-specialized species based on a few gene genealogies (58, 92, 111). In this system, population genomics should allow elucidating whether anther-smut fungi on these American *Silene* species show host differentiation or genome-wide gene flow, or

389 introgression only in genomic regions not involved in host specialization. Based on the few
390 genomes available so far (26), we find F_{ST} values between *Microbotryum* populations on the
391 two hosts, *S. virginica* and *S. caroliniana*, that are lower and more heterogeneous along the
392 genomes than between *M. lychnidis-dioicae* and *M. silenes-dioicae* referenced above (Figure
393 3B). These initial results suggests the occurrence of gene flow in some genomic regions. The
394 situation of anther-smut fungi on *S. vulgaris*, with three distant lineages with convergent
395 specialization on this same host species (2), would also be worth exploring using population
396 genomics to determine the extent of introgression and its genomic localization, and whether
397 the interspecific exchange of alleles has been deleterious or adaptive.
398
399 Another promising approach in anther-smut fungi for identifying genomic regions involved in
400 host adaptation will be to perform genome scans of differentiation between closely related
401 species or host races, if possible to avoid the potential pitfalls of such approaches (37). This
402 could contribute to our understanding of the role of gene flow in the early stages of
403 divergence and to identifying genomic regions less permeable to gene flow because of
404 selection for host adaptation and/or genetic incompatibilities between lineages (24, 37). More
405 generally, such population genomics approaches would be valuable to use in plant pathogen
406 fungi.
407
408 **Phylogeography and demographic history inferences**
409 *Microbotryum lychnidis-dioicae* and *M. silenes-dioicae* also constituted case studies in
410 providing one of the most clear-cut examples of phylogeographic structure in pathogens,
411 thanks to a collection of samples whose density and geographical scale was unprecedented for
412 a disease association in natural populations. In *M. lychnidis-dioicae*, clustering analyses based
413 on microsatellite markers (133), as well as nuclear gene sequences (69, 72), revealed the

414 existence of three genetically distinct clusters, reflecting recolonization from well-recognized
415 southern refugia after glaciation. Little admixture has been found between clusters based on
416 microsatellites (49, 133), and this has later been confirmed by whole genome sequences (15).
417 Indeed, SNPs (single nucleotide polymorphisms) revealed few shared polymorphisms and
418 many fixed differences among the clusters, and pairwise F_{ST} values between them were high
419 (0.56–0.74; Figures 3C and D), supporting low levels of inter-cluster gene flow (15). Whole
420 genome sequences provided further insights into the age of divergence between the three *M.*
421 *lychnidis-dioicae* lineages (Southern, Western and Eastern clusters), sequential size changes
422 in the population size of derived lineages and also supported low levels of gene flow (15).
423 Most notably, the pathogen genetic structure closely matched with the genetic structure of the
424 host species *S. latifolia* with the same regionally defined Southern, Western and Eastern
425 clusters, indicating that the anther-smut pathogen remained during the last glaciation in the
426 same three distinct refugia as its host (i.e. in the Iberian, Italian and Balkan peninsulas) (49).
427 The congruence of population structures between M. *lychnidis-dioicae* and its host appeared
428 even stronger than what could be expected because of isolation by distance alone, suggesting
429 that coevolution has played a significant role in the congruence of the population structures
430 (49). Genome-wide gene presence-absence polymorphism recovered the same population
431 structure (80). Inoculation experiments, indicating plant local adaptation for resistance to
432 pathogens (49, 89, 91), were consistent with a contribution of adaptive factors to the observed
433 congruence between pathogen and host population structures.
434
435 Microsatellite markers and genome-wide SNPs indicated that *M. silenes-dioicae* also
436 exhibited a genetic structure, albeit with biogeographic patterns more difficult to interpret (15,
437 133) and very low F_{ST} values genome-wide (Figure 3E). Genome-wide gene presence/absence
438 polymorphism revealed two different clusters with a more obvious east/west separation (80),

that may correspond to local adaptation of *S. dioica* clusters (81, 109), although this remains to be assessed. This case study shows the power of various kinds of population genomic studies to unravel weak and/or adaptive population subdivision.

4-UNRAVELLING MATING SYSTEM AND GAMETE COMPATIBILITY SYSTEMS USING BOTH COMPARATIVE GENOMICS AND POPULATION GENOMICS

The combination of comparative genomics and population genomics also can reveal remarkable transitions in mating systems by elucidating the changes in genomic mechanisms controlling mating compatibility. For the broad group of basidiomycete fungi, gamete compatibility is controlled by two loci acting at the haploid stage, mating being successful only between haploid cells carrying different alleles at both mating-type loci (34). The two mating-type loci are i) the PR locus which encodes pheromone genes and a pheromone receptor gene implicated in gamete recognition and fusion, and ii) the HD locus which encodes homeodomain protein-coding genes allowing, after fusion, for the maintenance of the dikaryon and hyphal growth (48, 94). Most basidiomycetes are outcrossing and have these two loci unlinked, although some fungi in this group have linked mating-type loci (103). Linkage of the two mating type loci is considered to be favored due to increased odds of gamete compatibility under selfing when mating-type loci are linked (103). Interestingly, most *Microbotryum* species are highly selfing and were long known to segregate only two mating type phenotypes, but it remained uncertain whether this was due to mating-type loci linkage or to the loss of role in mating-type determinism for one of the two mating-type loci, as both cases occurred in basidiomycetes (85, 103). Comparative genomics of well-assembled genomes allowed to resolve the complex genome architecture and long-term evolutionary history of the repeat-rich and rearranged mating type chromosomes in anther-smut fungi, and

464 population genomics datasets were essential for identifying young events of recombination
465 suppression.
466
467 Population genomics confirmed high rates of selfing in all studied *Microbotryum* species, by
468 showing high levels of genome-wide homozygosity (15, 25, 26) and confirmed massive
469 recombination suppression on mating-type chromosomes (84, 134). High-quality genomes
470 assemblies allowed reconstructing the history of genomic events underlying the shift in
471 gamete compatibility system (25, 26). The long-read sequencing technology allowed
472 assembling the two repeat-rich mating-type chromosomes of the Lamole *M. lychnidis-dioicae*
473 strain, which confirmed linkage between the two mating-type loci HD and PR (Figure 4A)
474 (16). Genome comparisons between multiple *Microbotryum* species showed that the ancestral
475 state had unlinked mating-type loci on two distinct chromosomes, and that independent
476 rearrangements and chromosome fusions occurred in multiple species, convergently linking
477 the two mating-type loci by large regions without recombination (Figure 4B) (25, 26). This
478 shows that natural selection can repeatedly lead to similar phenotypes through multiple
479 different evolutionary pathways.
480
481 Following recombination suppression, a chaos of rearrangements occurred on mating-type
482 chromosomes (16), as well as TE and non-synonymous substitution accumulation (16, 54), as
483 is typical in non-recombining regions, and in particular on sex chromosomes (14). The high-
484 quality assemblies allowed the detailed characterization of extensive rearrangements and
485 repeat accumulations on the two mating-type chromosomes (16, 25, 26). Another important
486 characteristic feature of sex chromosomes was observed on the mating-type chromosomes of
487 multiple *Microbotryum* species, i.e., the stepwise extension of the regions with recombination
488 suppression. The progressive extension of the regions without recombination revealed a

pattern of clear “evolutionary strata”, i.e., decreasing divergence between alleles on the alternative mating-type chromosomes farther from the mating-type loci (Figure 4C). Population genomics was essential for providing evidence of early events of recombination suppression in several species, by showing the segregation of alleles according to their associated mating-type, decreased levels of diversity as expected under lower population effective sizes and that high divergence between alleles associated with the alternative mating types was due to balancing selection on mating types rather than elevated substitution rates (25, 26). Indeed, as soon as recombination ceased, alleles on the two mating-type chromosomes diverged gradually with time (Figure 4D). Finding such evolutionary strata in fungi, which lack male and female roles, challenged the classical view for the evolution of sex chromosomes. Indeed stepwise recombination suppression in sex chromosomes was thought to be due primarily to sexual antagonism, i.e., the selection to link genes with alleles beneficial in one sex, and deleterious in the other, to the sex determining gene (21). The finding of evolutionary strata in fungi without sexual antagonism indicates that alternative hypotheses should be explored to explain the progressive spread of recombination suppression, such as overdominance, epigenetic modifications associated with transposable elements or neutral rearrangements (88, 108).

5-CONCLUDING REMARKS AND PERSPECTIVES

A thorough understanding of the major roles played by pathogens requires the integrative study of both ongoing processes of coevolution and dynamics of specialization that impact the emergence new diseases. Investigations of the anther-smut fungi utilizing comparative approaches to genomics and gene expression profiles, combined with population-level studies, illustrate the strength of combining different genomic approaches addressing different

514 scales of evolution (Figure 2). The availability of genomic data for multiple sister species and
515 multiple populations within species makes the anther-smut system quite exceptional for
516 identifying the genetic mechanisms involved in adaptation, coevolution, host specialization
517 and mating system at different evolutionary times (Table 1; Figure 1). Comparative genomics
518 has long been the predominant approach for studying adaptation in plant fungal pathogens
519 (46, 53, 59, 102, 107) and has provided important insights into the mechanisms of adaptation,
520 e.g., through horizontal gene transfers, gene gains/losses, hybridization or recurrent amino-
521 acid changes (71). Comparative genomics by definition does not consider population-level
522 variation, such that population genomics is a complementary approach for insights into
523 evolutionary processes acting at the local and regional scales. For example, several recent
524 studies have revealed gene presence/absence polymorphism within species (78, 80, 119, 120).
525 In addition, comparative genomics can only detect a specific type of positive selection,
526 involving frequent changes of amino-acids. Positive selection of a single amino-acid change
527 or of regulatory regions can only be detected by looking for selective sweeps using population
528 genomics. Some recent studies based on population genomics have in fact revealed important
529 aspects of adaptation in fungal plant pathogens, showing footprints of introgression, selective
530 sweeps and amino acid-changes (68, 79, 101, 116, 121, 122).
531
532 Furthermore, cross referencing candidate genes that are highlighted by multiple indications of
533 being subject to natural selection during parasitism as outlined here (e.g. genes found within a
534 selective sweep, upregulated in the plant and having experienced gene family expansion
535 compared to other fungal pathogens) can strengthen their putative roles as pathogen effector
536 that are central in the specificity of fungal-plant combinations. Functional studies can help
537 understanding the role of the candidate genes. Promising transformation protocol have been
538 developed in *M. lychnidis-dioicae* (128) and will likely facilitate the characterization of key

genes involved in the interaction between the anther-smut fungi and their Caryophyllaceae host plants, an important challenge for the coming years. Transcriptomes and epigenomes of multiple *Microbotryum* species and multiple strains within species will likely be complementary to the current available genomic ressources to identify the role of regulatory and epigenetic mechanisms in the adaptation of anther-smut fungi to their hosts and environment, their divergence and their mating-type chromosome organisation, contributing to further understanding the mechanisms involved in adaptive processes in fungal pathogen populations. It will also be interesting to investigate similar levels of among and within species sampling and genome sequencing using pathogens of different levels of obligate parasitism, including facultative and opportunistic pathogens, as well as hemibiotrophy and necrotrophy.

ACKNOWLEDGMENTS

This work was supported by a Marie Curie European grant (PRESTIGE-2016-4-0013 to FEH). FEH received the Young Biological Researcher Prize from the Fondation des Treilles, created by Anne Gruner Schlumberger, which supports research in Science and Art (http://www.les-treilles.com).

560

561 **REFERENCES :**
562

563 1. Abbate JL, Antonovics J. 2014. Elevational disease distribution in a natural plant–
564 pathogen system: insights from changes across host populations and climate. *Oikos*.
565 123(9):1126–36

566 2. Abbate JL, Gladieux P, Hood ME, de Vienne DM, Antonovics J, et al. 2018. Co-
567 occurrence among three divergent plant-castrating fungi in the same *Silene* host
568 species. *Mol. Ecol.* 27(16):3357–70

569 3. Aguileta G, Badouin H, Hood ME, Møller AP, Prieur SL, et al. 2016. Lower
570 prevalence but similar fitness in a parasitic fungus at higher radiation levels near
571 Chernobyl. *Mol. Ecol.* 25(14):3370–83

572 4. Aguileta G, Lengelle J, Marthey S, Chiapello H, Rodolphe F, et al. 2010. Finding
573 candidate genes under positive selection in non-model species: examples of genes
574 involved in host specialization in pathogens. *Mol. Ecol.* 19(2):292–306

575 5. Albalat R, Cañestro C. 2016. Evolution by gene loss. *Nat. Rev. Genet.* 17(7):379–91

576 6. Alexander HM, Antonovics J. 1995. Spread of anther-smut disease (*Ustilago violacea*)
577 and character correlations in a genetically variable experimental population of *Silene*
578 *alba*. *J. Ecol.* 83(5):783–94

579 7. Alexander HM, Thrall PH, Antonovics J, Jarosz AM, Oudemans PV. 1996. Population
580 dynamics and genetics of plant disease: a case study of anther- smut disease. *Ecology*.
581 77(4):990–96

582 8. Ames RM, Money D, Ghatge VP, Whelan S, Lovell SC. 2012. Determining the
583 evolutionary history of gene families. *Bioinformatics*. 28(1):48–55

9. Anderson PK, Cunningham AA, Patel NG, Morales FJ, Epstein PR, Daszak P. 2004. Emerging infectious diseases of plants: pathogen pollution, climate change and agrotechnology drivers. *Trends Ecol. Evol.* 19(10):535–44

10. Antonovics J, Alexander HM. 1992. Epidemiology of anther-smut infection of *Silene alba* (= *S. latifolia*) caused by *Ustilago violacea* : patterns of spore deposition in experimental populations. *Proc R Soc Lond B*. 250(1328):157–63

11. Antonovics J, Hood ME, Thrall PH, Abrams JY, Duthie GM. 2003. Herbarium studies on the distribution of anther-smut fungus (*Microbotryum violaceum*) and *Silene* species (Caryophyllaceae) in the eastern United States. *Am. J. Bot.* 90(10):1522–31

12. Antonovics J, Stratton D, Thrall PH, Jarosz AM. 1996. An anther-smut disease (*Ustilago violacea*) of fire-pink (*Silene virginica*): Its biology and relationship to the anther-smut disease of white campion (*Silene alba*). *Am. Midl. Nat.* 135(1):130–43

13. Audran J-C, Batcho M. 1981. Microsporogenesis and pollen grains in *Silene dioica* (L.) Cl. and alterations in its anthers parasitised by *Ustilago violacea* (Pers.) Rouss. *Acta Soc Bot Pol*, pp. 29–32

14. Bachtrog D. 2013. Y-chromosome evolution: emerging insights into processes of Y-chromosome degeneration. *Nat. Rev. Genet.* 14(2):113–24

15. Badouin H, Gladieux P, Gouzy J, Siguenza S, Aguileta G, et al. 2017. Widespread selective sweeps throughout the genome of model plant pathogenic fungi and identification of effector candidates. *Mol. Ecol.* 26(7):2041–62

16. Badouin H, Hood ME, Gouzy J, Aguileta G, Siguenza S, et al. 2015. Chaos of rearrangements in the mating-type chromosomes of the anther-smut fungus *Microbotryum lychnidis-dioicae*. *Genetics*. 200(4):1275–84

17. Baker HG. 1947. Infection of species of *Melandrium* by *Ustilago violacea* (Pers.) Fuckel and the transmission of the resultant disease. *Ann. Bot.* 11(43):333–48

18. Barrett LG, Thrall PH, Burdon JJ, Linde CC. 2008. Life history determines genetic structure and evolutionary potential of host–parasite interactions. *Trends Ecol. Evol.* 23(12):678–85

19. Bartoli C, Roux F. 2017. Genome-wide association studies in plant pathosystems: toward an ecological genomics approach. *Front. Plant Sci.* 8:763

20. Beckerson WC, Rodríguez de la Vega RC, Hartmann FE, Giraud T, Perlin Michael H. 2018. Cause and effectors: Secretome comparison of members from the anther-smut pathogen species complex, *Microbotryum violaceum*. *In prep.*

21. Bergero R, Charlesworth D. 2009. The evolution of restricted recombination in sex chromosomes. *Trends Ecol. Evol.* 24(2):94–102

22. Bernasconi G, Antonovics J, Biere A, Charlesworth D, Delph LF, et al. 2009. *Silene* as a model system in ecology and evolution. *Heredity*. 103(1):5–14

23. Biere A, Honders SC. 1998. Anther smut transmission in *Silene latifolia* and *Silene dioica*: impact of host traits, disease frequency, and host density. *Int. J. Plant Sci.* 159(2):228–35

24. Bierne N, Welch J, Loire E, Bonhomme F, David P. 2011. The coupling hypothesis: why genome scans may fail to map local adaptation genes. *Mol. Ecol.* 20(10):2044–72

25. Branco S, Badouin H, Rodríguez de la Vega RC, Gouzy J, Carpentier F, et al. 2017. Evolutionary strata on young mating-type chromosomes despite the lack of sexual antagonism. *Proc. Natl. Acad. Sci.* 114(27):7067–72

26. Branco S, Carpentier F, Rodríguez de la Vega RC, Badouin H, Snirc A, et al. 2018. Multiple convergent supergene evolution events in mating-type chromosomes. *Nat. Commun.* 9(1):2000

27. Bueker B, Eberlein C, Gladieux P, Schaefer A, Snirc A, et al. 2016. Distribution and population structure of the anther smut *Microbotryum silenes-acaulis* parasitizing an arctic–alpine plant. *Mol. Ecol.* 25(3):811–24

28. Büker B, Petit E, Begerow D, Hood ME. 2013. Experimental hybridization and backcrossing reveal forces of reproductive isolation in *Microbotryum*. *BMC Evol. Biol.* 13(1):224

29. Carlsson U, Elmqvist T. 1992. Epidemiology of anther-smut disease (*Microbotryum violaceum*) and numeric regulation of populations of *Silene dioica*. *Oecologia*. 90(4):509–17

30. Carpentier F, Rodríguez de la Vega RC, Branco S, Snirc A, Coelho MA, et al. 2018. Convergent recombination cessation between mating-type genes and centromeres in selfing anther-smut fungi. *In prep.*

31. Casola C. 2018. From de novo to "de nono": The majority of novel protein coding genes identified with phylostratigraphy are old genes or recent duplicates. *Genome Biol. Evol.* 10(11):2906–18

32. Cattrall ME, Baird ML, Garber ED. 1978. Genetics of *Ustilago violacea*. III. Crossing over and nondisjunction. *Bot. Gaz.* 139(2):266–70

33. Chung E, Petit E, Antonovics J, Pedersen AB, Hood ME. 2012. Variation in resistance to multiple pathogen species: anther smuts of *Silene uniflora*. *Ecol. Evol.* 2(9):2304–14

34. Coelho MA, Bakkeren G, Sun S, Hood ME, Giraud T. 2017. Fungal sex: The Basidiomycota. *Microbiol. Spectr.* 5(3):FUNK-0046-2016

35. Cordier T, Robin C, Capdevielle X, Fabreguettes O, Desprez-Loustau M-L, Vacher C. 2012. The composition of phyllosphere fungal assemblages of European beech (*Fagus sylvatica*) varies significantly along an elevation gradient. *New Phytol.* 196(2):510–19

36. Croll D, McDonald BA. 2017. The genetic basis of local adaptation for pathogenic fungi in agricultural ecosystems. *Mol. Ecol.* 26(7):2027–40

37. Cruickshank TE, Hahn MW. 2014. Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. *Mol. Ecol.* 23(13):3133–57

38. Day AW, Garber ED. 1988. 30 - *Ustilago violacea*, anther smut of the Caryophyllaceae. In *Advances in Plant Pathology*, ed GS Sidhu. 6:457–82. Academic Press

39. Day AW, Jones JK. 1968. The production and characteristics of diploids in *Ustilago violacea*. *Genet. Res.* 11(1):63–81

40. de Vienne DM, Hood ME, Giraud T. 2009a. Phylogenetic determinants of potential host shifts in fungal pathogens. *J. Evol. Biol.* 22(12):2532–41

41. de Vienne DM, Refrégier G, Hood ME, Guigue A, Devier B, et al. 2009b. Hybrid sterility and inviability in the parasitic fungal species complex *Microbotryum*. *J. Evol. Biol.* 22(4):683–98

42. de Vienne DM, Refrégier G, López-Villavicencio M, Tellier A, Hood ME, Giraud T. 2013. Cospeciation vs host-shift speciation: methods for testing, evidence from natural associations and relation to coevolution. *New Phytol.* 198(2):347–85

43. Delmotte F, Bucheli E, Shykoff JA. 1999. Host and parasite population structure in a natural plant–pathogen system. *Heredity*. 82(3):300–308

44. Desprez-Loustau M-L, Capron G, Dupuis F. 1998. Relating germination dynamics of *Melampsora pinitorqua* teliospores to temperature and rainfall during overwintering. *Eur. J. For. Pathol.* 28(5):335–47

45. Desprez-Loustau M-L, Robin C, Buée M, Courtecuisse R, Garbaye J, et al. 2007. The fungal dimension of biological invasions. *Trends Ecol. Evol.* 22(9):472–80

681 46. Duplessis S, Cuomo CA, Lin Y-C, Aerts A, Tisserant E, et al. 2011. Obligate biotrophy
682 features unraveled by the genomic analysis of rust fungi. *Proc. Natl. Acad. Sci.*
683 108(22):9166–71
684 47. Enjalbert J, Duan X, Leconte M, HovmøLler MS, De Vallavieille-Pope C. 2005.
685 Genetic evidence of local adaptation of wheat yellow rust (*Puccinia striiformis* f. sp.
686 *tritici*) within France. *Mol. Ecol.* 14(7):2065–73
687 48. Feldbrügge M, Kämper J, Steinberg G, Kahmann R. 2004. Regulation of mating and
688 pathogenic development in *Ustilago maydis*. *Curr. Opin. Microbiol.* 7(6):666–72
689 49. Feurtey A, Gladieux P, Hood ME, Snirc A, Cornille A, et al. 2016. Strong
690 phylogeographic co-structure between the anther-smut fungus and its white campion
691 host. *New Phytol.* 212(3):668–79
692 50. Feurtey A, Stukenbrock EH. 2018. Interspecific gene exchange as a driver of adaptive
693 evolution in fungi. *Annu. Rev. Microbiol.* 72(1):377–98
694 51. Fisher MC, Hawkins NJ, Sanglard D, Gurr SJ. 2018. Worldwide emergence of
695 resistance to antifungal drugs challenges human health and food security. *Science*.
696 360(6390):739–42
697 52. Fisher MC, Henk DA, Briggs CJ, Brownstein JS, Madoff LC, et al. 2012. Emerging
698 fungal threats to animal, plant and ecosystem health. *Nature*. 484(7393):186–94
699 53. Floudas D, Binder M, Riley R, Barry K, Blanchette RA, et al. 2012. The paleozoic
700 origin of enzymatic lignin decomposition reconstructed from 31 fungal genomes.
701 *Science*. 336(6089):1715–19
702 54. Fontanillas E, Hood ME, Badouin H, Petit E, Barbe V, et al. 2014. Degeneration of the
703 non-recombining regions in the mating-type chromosomes of the anther-smut fungi.
704 *Mol. Biol. Evol.* 32(4):928–43

55. Fortuna TM, Namias A, Snirc A, Branca A, Hood ME, et al. 2018. Multiple infections, relatedness and virulence in the anther-smut fungus castrating *Saponaria* plants. *Mol. Ecol.* doi:10.1111/mec.14911:

56. Fortuna TM, Snirc A, Badouin H, Gouzy J, Siguenza S, et al. 2016. Polymorphic microsatellite markers for the tetrapolar anther-smut fungus *Microbotryum saponariae* based on genome sequencing. *PLOS ONE*. 11(11):e0165656

57. Fouché S, Plissonneau C, Croll D. 2018. The birth and death of effectors in rapidly evolving filamentous pathogen genomes. *Curr. Opin. Microbiol.* 46:34–42

58. Freeman AB, Kellye Duong K, Shi T-L, Hughes CF, Perlin MH. 2002. Isolates of *Microbotryum violaceum* from North American host species are phylogenetically distinct from their European host-derived counterparts. *Mol. Phylogenet. Evol.* 23(2):158–70

59. Friesen TL, Stukenbrock EH, Liu Z, Meinhardt S, Ling H, et al. 2006. Emergence of a new disease as a result of interspecific virulence gene transfer. *Nat. Genet.* 38(8):953–56

60. Gandon S, Michalakis Y. 2002. Local adaptation, evolutionary potential and host–parasite coevolution: interactions between migration, mutation, population size and generation time. *J. Evol. Biol.* 15(3):451–62

61. Garber ED, Day AW. 1985. Genetic mapping of a phytopathogenic Basidiomycete, *Ustilago violacea*. *Bot. Gaz.* 146(4):449–59

62. Garber ED, Ruddat M. 2002. Transmission genetics of *Microbotryum violaceum* (*Ustilago violacea*): a case study. *Advances in Applied Microbiology*, pp. 107–127

63. Gibson AK, Hood ME, 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):1917–30

64. Gibson AK, Refrégier G, Hood ME, Giraud T. 2014. Performance of a hybrid fungal pathogen on pure-species and hybrid host plants. *Int. J. Plant Sci.* 175(6):724–30

65. Giraud T. 2004. Patterns of within population dispersal and mating of the fungus *Microbotryum violaceum* parasitising the plant *Silene latifolia*. *Heredity*. 93(6):559–65

66. Giraud T, Gladieux P, Gavrilets S. 2010. Linking the emergence of fungal plant diseases with ecological speciation. *Trends Ecol. Evol.* 25(7):387–95

67. Giraud T, Jonot O, Shykoff JA. 2005. Selfing propensity under choice conditions in a parasitic fungus, *Microbotryum violaceum*, and parameters influencing infection success in artificial inoculations. *Int. J. Plant Sci.* 166(4):649–57

68. Gladieux P, Condon B, Ravel S, Soanes D, Maciel JLN, et al. 2018. Gene flow between divergent cereal- and grass-specific lineages of the rice blast fungus *Magnaporthe oryzae*. *mBio*. 9(1):e01219-17

69. Gladieux P, Devier B, Aguileta G, Cruaud C, Giraud T. 2013. Purifying selection after episodes of recurrent adaptive diversification in fungal pathogens. *Infect. Genet. Evol. J. Mol. Epidemiol. Evol. Genet. Infect. Dis.* 17:123–31

70. Gladieux P, Feurtey A, Hood ME, Snirc A, Clavel J, et al. 2015. The population biology of fungal invasions. *Mol. Ecol.* 24(9):1969–86

71. Gladieux P, Ropars J, Badouin H, Branca A, Aguileta G, et al. 2014. Fungal evolutionary genomics provides insight into the mechanisms of adaptive divergence in eukaryotes. *Mol. Ecol.* 23(4):753–73

72. Gladieux P, Vercken E, Fontaine MC, Hood ME, Jonot O, et al. 2011. Maintenance of fungal pathogen species that are specialized to different hosts: allopatric divergence and introgression through secondary contact. *Mol. Biol. Evol.* 28(1):459–71

73. Goldschmidt, V. 1928. Vererbungsversuche mit den biologischen arten den Antherenbrandes (*Ustilago violacea* Pers.). *Z. Bot.* 21(1):1–90

74. Gout L, Fudal I, Kuhn M-L, Blaise F, Eckert M, et al. 2006. Lost in the middle of nowhere: the *AvrLm1* avirulence gene of the Dothideomycete *Leptosphaeria maculans*. *Mol. Microbiol.* 60(1):67–80

75. Grünwald NJ, McDonald BA, Milgroom MG. 2016. Population genomics of fungal and oomycete pathogens. *Annu. Rev. Phytopathol.* 54(1):323–46

76. Han MV, Thomas GWC, Lugo-Martinez J, Hahn MW. 2013. Estimating gene gain and loss rates in the presence of error in genome assembly and annotation using CAFE 3. *Mol. Biol. Evol.* 30(8):1987–97

77. Hardison RC. 2003. Comparative Genomics. *PLOS Biol.* 1(2):e58

78. Hartmann FE, Croll D. 2017. Distinct trajectories of massive recent gene gains and losses in populations of a microbial eukaryotic pathogen. *Mol. Biol. Evol.* 34(11):2808–22

79. Hartmann FE, McDonald BA, Croll D. 2018. Genome-wide evidence for divergent selection between populations of a major agricultural pathogen. *Mol. Ecol.* 27(12):2725–41

80. Hartmann FE, Rodríguez de la Vega RC, Brandenburg J-T, Carpentier F, Giraud T. 2018. Gene presence–absence polymorphism in castrating anther-smut fungi: recent gene gains and phylogeographic structure. *Genome Biol. Evol.* 10(5):1298–1314

81. Hathaway L, Malm JU, Prentice HC. 2009. Geographically congruent large-scale patterns of plastid haplotype variation in the European herbs *Silene dioica* and *S. latifolia* (Caryophyllaceae). *Bot. J. Linn. Soc.* 161(2):153–70

82. Hoban S, Kelley JL, Lotterhos KE, Antolin MF, Bradburd G, et al. 2016. Finding the genomic basis of local adaptation: Pitfalls, practical solutions, and future directions. *Am. Nat.* 188(4):379–97

779 83. Hood ME, Mena-Alí JI, Gibson AK, Oxelman B, Giraud T, et al. 2010. Distribution of
780 the anther-smut pathogen *Microbotryum* on species of the Caryophyllaceae. *New*
781 *Phytol.* 187(1):217–29
782 84. Hood ME, Petit E, Giraud T. 2013. Extensive divergence between mating-type
783 chromosomes of the anther-smut fungus. *Genetics*. 193(1):309–15
784 85. Hood ME, Scott M, Hwang M. 2015. Breaking linkage between mating compatibility
785 factors: Tetrapolarity in *Microbotryum*. *Evolution*. 69(10):2561–72
786 86. Horns F, Petit E, Hood ME. 2017. Massive expansion of Gypsy-Like retrotransposons
787 in *Microbotryum* fungi. *Genome Biol. Evol.* 9(2):363–71
788 87. Horns F, Petit E, Yockteng R, Hood ME. 2012. Patterns of repeat-induced point
789 mutation in transposable elements of Basidiomycete fungi. *Genome Biol. Evol.*
790 4(3):240–47
791 88. Ironside JE. 2010. No amicable divorce? Challenging the notion that sexual antagonism
792 drives sex chromosome evolution. *BioEssays*. 32(8):718–26
793 89. Kaltz O, Gandon S, Michalakis Y, Shykoff JA. 1999. Local maladaptation in the
794 anther-smut fungus *Microbotryum violaceum* to its host plant *Silene latifolia*: evidence
795 from a cross-inoculation experiment. *Evol. Int. J. Org. Evol.* 53(2):395–407
796 90. Kaltz O, Shykoff JA. 2001. Male and female *Silene latifolia* plants differ in per-contact
797 risk of infection by a sexually transmitted disease. *J. Ecol.* 89(1):99–109
798 91. Kaltz O, Shykoff JA. 2002. Within- and among-population variation in infectivity,
799 latency and spore production in a host–pathogen system. *J. Evol. Biol.* 15(5):850–60
800 92. Kemler M, Göker M, Oberwinkler F, Begerow D. 2006. Implications of molecular
801 characters for the phylogeny of the Microbotryaceae (Basidiomycota:
802 Urediniomycetes). *BMC Evol. Biol.* 6:35

803 93. Kneip H. 1919. Untersuchungen uber den Antherenbrand (*Ustilago violacea* Pers.): Ein
804 Beitrag zum Sexualitätproblem. *Zeitschrt Bot*. 11:275–84
805 94. Kües U. 2000. Life history and developmental processes in the basidiomycete *Coprinus*
806 *cinereus*. *Microbiol Mol Biol Rev*. 64(2):316–53
807 95. Kulkarni RD, Thon MR, Pan H, Dean RA. 2005. Novel G-protein-coupled receptor-
808 like proteins in the plant pathogenic fungus *Magnaporthe grisea*. *Genome Biol.*
809 6(3):R24
810 96. Kuppireddy V, Uversky V, Toh S, Tsai M-C, Beckerson W, et al. 2017. Identification
811 and initial characterization of the effectors of an anther smut fungus and potential host
812 target proteins. *Int. J. Mol. Sci.* 18(11):2489
813 97. Le Gac M, Hood ME, Fournier E, Giraud T. 2007a. Phylogenetic evidence of host-
814 specific cryptic species in the anther smut fungus. *Evol. Int. J. Org. Evol.* 61(1):15–26
815 98. Le Gac M, Hood ME, Giraud T. 2007b. Evolution of reproductive isolation within a
816 parasitic fungal species complex. *Evolution*. 61(7):1781–87
817 99. Levasseur A, Pontarotti P. 2011. The role of duplications in the evolution of genomes
818 highlights the need for evolutionary-based approaches in comparative genomics. *Biol.*
819 *Direct*. 6:11
820 100. McDonald JH, Kreitman M. 1991. Adaptive protein evolution at the *Adh* locus in
821 Drosophila. *Nature*. 351(6328):652–54
822 101. Mohd-Assaad N, McDonald BA, Croll D. 2018. Genome-wide detection of genes
823 under positive selection in worldwide populations of the barley scald pathogen.
824 *Genome Biol. Evol.* 10(5):1315–32
825 102. Nagy LG, Riley R, Tritt A, Adam C, Daum C, et al. 2016. Comparative genomics of
826 early-diverging mushroom-forming fungi provides insights into the origins of
827 lignocellulose decay capabilities. *Mol. Biol. Evol.* 33(4):959–70

828 103. Nieuwenhuis BPS, Billiard S, Vuilleumier S, Petit E, Hood ME, Giraud T. 2013.
829 Evolution of uni- and bifactorial sexual compatibility systems in fungi. *Heredity*.
830 111(6):445–55
831 104. Perlin MH, Amselem J, Fontanillas E, Toh SS, Chen Z, et al. 2015. Sex and parasites:
832 genomic and transcriptomic analysis of *Microbotryum lychnidis-dioicae*, the biotrophic
833 and plant-castrating anther smut fungus. *BMC Genomics*. 16:461
834 105. Petit E, Silver C, Cornille A, Gladieux P, Rosenthal L, et al. 2017. Co-occurrence and
835 hybridization of anther-smut pathogens specialized on *Dianthus* hosts. *Mol. Ecol.*
836 26(7):1877–90
837 106. Pfeifer B, Wittelsbürger U, Ramos-Onsins SE, Lercher MJ. 2014. PopGenome: an
838 efficient Swiss army knife for population genomic analyses in R. *Mol. Biol. Evol.*
839 31(7):1929–36
840 107. Plissonneau C, Benevenuto J, Mohd-Assaad N, Fouché S, Hartmann FE, Croll D. 2017.
841 Using population and comparative genomics to understand the genetic basis of
842 effector-driven fungal pathogen evolution. *Front. Plant Sci.* 8(119)
843 108. Ponnikas S, Sigeman H, Abbott JK, Hansson B. 2018. Why do sex chromosomes stop
844 recombining? *Trends Genet.* 34(7):492–503
845 109. Rautenberg A, Hathaway L, Oxelman B, Prentice HC. 2010. Geographic and
846 phylogenetic patterns in *Silene* section *Melandrium* (Caryophyllaceae) as inferred from
847 chloroplast and nuclear DNA sequences. *Mol. Phylogenet. Evol.* 57(3):978–91
848 110. Refrégier G, Hood ME, Giraud T. 2010. No evidence of reproductive character
849 displacement between two sister fungal species causing anther smut disease in *Silene*.
850 *Int. J. Plant Sci.* 171(8):847–59
851 111. Refrégier G, Le Gac M, Jabbour F, Widmer A, Shykoff JA, et al. 2008. Cophylogeny
852 of the anther smut fungi and their caryophyllaceous hosts: prevalence of host shifts and

importance of delimiting parasite species for inferring cospeciation. *BMC Evol. Biol.* 8:100

112. Rodríguez de la Vega RC, Rodriguez P, Hoitinga P, Aguileta G, Badouin H, Gouzy J, Siguenza S, Le Prieur S, Esquerre D, Hood ME, Giraud T. 2018. Genomic mechanisms of host specialization in plant pathogens: the case of the castrating anther-smut fungi. *In prep.*

113. Salvaudon L, Giraud T, Shykoff JA. 2008. Genetic diversity in natural populations: a fundamental component of plant–microbe interactions. *Curr. Opin. Plant Biol.* 11(2):135–43

114. Sánchez-Vallet A, Hartmann FE, Marcel TC, Croll D. 2018. Nature's genetic screens: using genome-wide association studies for effector discovery. *Mol. Plant Pathol.* 19(1):3–6

115. Schäfer AM, Kemler M, Bauer R, Begerow D. 2010. The illustrated life cycle of *Microbotryum* on the host plant *Silene latifolia*. *Botany*. 88(10):875–85

116. Schirrmann MK, Zoller S, Croll D, Stukenbrock EH, Leuchtmann A, Fior S. 2018. Genomewide signatures of selection in *Epichloë* reveal candidate genes for host specialization. *Mol. Ecol.* 27(15):3070–86

117. Sharma R, Mishra B, Runge F, Thines M. 2014. Gene loss rather than gene gain is associated with a host jump from monocots to dicots in the smut fungus *Melanopsichium pennsylvanicum*. *Genome Biol. Evol.* 6(8):2034–49

118. Sloan DB, Giraud T, Hood ME. 2008. Maximized virulence in a sterilizing pathogen: the anther-smut fungus and its co-evolved hosts. *J. Evol. Biol.* 21(6):1544–54

119. Steenwyk J, Rokas A. 2017. Extensive copy number variation in fermentation-related genes among *Saccharomyces cerevisiae* wine strains. *G3 Genes Genomes Genet.* 7(5):1475–85

878 120. Steenwyk JL, Soghigian JS, Perfect JR, Gibbons JG. 2016. Copy number variation
879 contributes to cryptic genetic variation in outbreak lineages of *Cryptococcus gattii* from
880 the North American Pacific Northwest. *BMC Genomics*. 17(1):700
881 121. Stukenbrock EH, Christiansen FB, Hansen TT, Dutheil JY, Schierup MH. 2012. Fusion
882 of two divergent fungal individuals led to the recent emergence of a unique widespread
883 pathogen species. *Proc. Natl. Acad. Sci.* 109(27):10954–59
884 122. Taylor JW, Branco S, Gao C, Hann-Soden C, Montoya L, et al. 2017. Sources of fungal
885 genetic variation and associating It with phenotypic diversity. *Microbiol. Spectr.* 5(5):
886 123. Thrall PH, Biere A, Antonovics J. 1993. Plant life-history and disease susceptibility--
887 The occurrence of *Ustilago violacea* on different species within the Caryophyllaceae. *J.*
888 *Ecol.* 81(3):489–98
889 124. Thrall PH, Burdon JJ. 2004. Host–pathogen life history interactions affect biological
890 control success. *Weed Technol.* 18(sp1):1269–74
891 125. Toh SS, Chen Z, Rouchka EC, Schultz DJ, Cuomo CA, Perlin MH. 2018. Pas de deux:
892 an intricate dance of anther smut and Its host. *G3 Genes Genomes Genet.*
893 g3.300318.2017
894 126. Toh SS, Chen Z, Schultz DJ, Cuomo CA, Perlin MH. 2017. Transcriptional analysis of
895 mating and pre-infection stages of the anther smut, *Microbotryum lychnidis-dioicae*.
896 *Microbiology*. 163(3):410–20
897 127. Toh SS, Perlin MH. 2016. Resurgence of less-studied smut fungi as models of
898 phytopathogenesis in the omics age. *Phytopathology*. 106(11):1244–54
899 128. Toh SS, Treves DS, Barati MT, Perlin MH. 2016. Reliable transformation system for
900 *Microbotryum lychnidis dioicae* informed by genome and transcriptome project. *Arch.*
901 *Microbiol.* 198(8):813–25

129. Vacher C, Vile D, Helion E, Piou D, Desprez-Loustau M-L. 2008. Distribution of parasitic fungal species richness: influence of climate versus host species diversity. *Divers. Distrib.* 14(5):786–98

130. Valente LM, Savolainen V, Vargas P. 2010. Unparalleled rates of species diversification in Europe. *Proc. R. Soc. Lond. B Biol. Sci.* 277(1687):1489–96

131. van Putten WF, Biere A, Damme JMMV. 2003. Intraspecific competition and mating between fungal strains of the anther smut *Microbotryum violaceum* from the host plants *Silene latifolia* and *S. dioica*. *Evolution*. 57(4):766–76

132. van Putten WF, Biere A, van Damme JMM. 2005. Host-related genetic differentiation in the anther smut fungus *Microbotryum violaceum* in sympatric, parapatric and allopatric populations of two host species *Silene latifolia* and *S. dioica*. *J. Evol. Biol.* 18(1):203–12

133. Vercken E, Fontaine MC, Gladieux P, Hood ME, Jonot O, Giraud T. 2010. Glacial refugia in pathogens: European genetic structure of anther smut pathogens on *Silene latifolia* and *Silene dioica*. *PLOS Pathog.* 6(12):e1001229

134. Whittle CA, Votintseva A, Ridout K, Filatov DA. 2015. Recent and massive expansion of the mating-type-specific region in the smut fungus *Microbotryum*. *Genetics*. 199(3):809–16

135. Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* 24(8):1586–91

136. Yockteng R, Marthey S, Chiapello H, Gendrault A, Hood ME, et al. 2007. Expressed sequences tags of the anther smut fungus, *Microbotryum violaceum*, identify mating and pathogenicity genes. *BMC Genomics*. 8(1):272

137. Zemp N, Tavares R, Widmer A. 2015. Fungal infection induces sex-specific transcriptional changes and alters sexual dimorphism in the dioecious plant *Silene latifolia*. *PLOS Genet.* 11(10):e1005536

930 # Tables and Figures

931 **Table 1: Whole genome public resources in anther-smut fungi (*Microbotryum* genus).**

932

933

Figure 1: Specialization and co-phylogenies of anther-smut fungi and their corresponding host plants. *(A)* Cladograms representing relationships between species of anther-smut fungi (left, *Microbotryum* genus) and species of host plants (right, mainly Caryophyllaceae), that are a consensus from previous phylogenetic analyses (26, 30, 111). Availability of short reads or long reads-based genome assemblies or population genomic data for the species of anther-smut fungi as presented in Table 1 is indicated with a black square near the fungal cladogram leaves. Dashed lines indicate specialization of a fungal species on a host species, with pink lines for fungal species infecting different hosts, and orange links for fungal species infecting the same host. The sequenced strain *M. intermedium* was sampled on *Salvia pratensis*, although this fungal species is usually found on *Scabiosa* hosts. *(B)* Infected host plants by anther-smut fungi. Numbers refer to host species as in panel A. Spores of anther-smut fungi are visible in the anthers of the flowers (Photo credits: Michael E. Hood, Tatiana Giraud, Maarteen Strack van Schijndel).

Figure 2: Evolutionary processes in anther-smut fungi studied by comparative genomics and population genomics methods. *(A*) Schema highlighting differences in time scales between host specialization, species divergence, coevolution and local adaptation events in four host-specialized *Microbotryum* species. *(B)* Type of genomic variation investigated according to the evolutionary event time scales. *(C)* Examples of methods recently used to investigate various evolutionary events in anther smut fungi, focusing on between-species variation (1), between- and within-species variation (2), or only within-species variation (3). Information on gene annotation, gene expression and gene presence-absence polymorphism may be coupled to narrow down the number of candidate genes to be involved in host specialization, coevolution and local adaptation. (1) Study of gene content variation between whole genome shotgun assemblies of 19 *Microbotryum* species. Core and complementary

(species-specific) genomes were computed by sampling groups of 1 to 18 *Microbotryum* species (112). Increase in size of the complementary genome with the number of sampled genomes highlights the dynamic gene gain and loss within the *Microbotryum* genus. Genes contained in the complementary genome are putative candidate genes for host specialization. (2) Identification of genes under positive selection using polymorphism in one species and divergence with an outgroup using McDonald–Kreitman tests. An excess of the ratio between non-synonymous (D_N) and synonymous (D_S) substitutions between species compared to the ratio between synonymous (P_S) and non-synonymous (P_N) polymorphisms within species is indicative of positive selection within the focal species indicated by an asterisk. Examples are shown for the orthologous group ORTHAg06728 and ORTHAg05587 (20). (3) Genome scan to identify selective sweeps based on allele frequency spectrum in *M. lychnidis-dioicae*. A selective sweep is characterized by an excess of rare variants. Composite likelihood ratio (CLR) tests estimate the probability of the presence of a selective sweep taking into account demographic history and genome-wide allele frequency spectrum (15).

Figure 3: Distribution of divergence along genomes between species of host-specialized anther-smut fungi (*Microbotryum* genus) based on F_{ST} genome scans. F_{ST} distributions are based on the genomes of five strains in each group for comparisons, except for strains on *S. caroliniana* for which only three genomes were available. *(A)* Divergence distribution between *M. silenes-dioicae* and *M. lychnidis-dioicae. (B)* Divergence distribution between *Microbotryum* fungi on *S. caroliniana* and *S. virginica*. *(C)* (resp. *(D)*) Divergence distribution between Southern and Northern (resp. Eastern) European genetic clusters of *M. lychnidis-dioicae* parasitizing *S. latifolia. (E)* Divergence distribution between Eastern and Western European genetic clusters of *M. silenes-dioicae* parasitizing *S. dioica*. In each panel, from top to bottom: density curve of genome-wide per-gene F_{ST} values; chromosomal

984 distribution of per-gene F_{ST} values on the species largest chromosome (on 2 Mb for each to
985 ease comparisons); map showing the sampling location of sequenced strains (genomes used
986 for F_{ST} distributions are shown as squares); pictures of infected hosts by each host-specialized
987 species (Photo credits: Michael E. Hood). For each pairwise comparison, F_{ST} values were
988 calculated per gene for all genes present on autosomal contigs (i.e. not belonging to mating-
989 type chromosomes) and carrying at least 1 SNP using the PopGenome R package (106). Red
990 dashed lines correspond to median F_{ST} values. Genomic data were described in (15, 26, 134).
991 Mapping, SNP calling and F_{ST} calculations were described in (26, 80).
992
993 **Figure 4: Genomic rearrangements and evolutionary strata on mating-type**
994 **chromosomes of *Microbotryum lychnidis-dioicae* on *Silene latifolia*.** *(A)* Circos plot
995 allowing to retrieve the rearrangements events which occurred during the evolution of the *M.*
996 *lychnidis-dioicae* mating-type chromosome. The two mating-type chromosomes (PR and HD
997 mating-type chromosome) of *M. lagerheimii* are used as proxy for the ancestral state (25).
998 The outer tracks represent contigs with scale in Mb. The blue and orange lines link orthologs,
999 with inversions in orange. The blue and purple dots represent the HD and PR loci,
1000 respectively, and the yellow regions the centromeres. *(B)* Evolutionary scenario of the *M.*
1001 *lychnidis-dioicae* mating-type chromosome evolution from the two ancestral mating-type
1002 chromosomes through a centromere-to-telomere fusion event, which brought the PR and HD
1003 loci onto the same chromosome and allowed their linkage through a recombination
1004 suppression (dashed lines). *(C)* Demonstration of stepwise recombination suppression using
1005 per-gene synonymous divergence and their respective standard error (dS ± SE) between
1006 alleles from *M. lychnidis-dioicae* associated to the a_1 versus a_2 mating types along the mating-
1007 type chromosome gene order of *M. lagerheimii*, as a proxy for ancestral gene order.
1008 Evolutionary strata of different divergence levels (colored differently) shows that

recombination suppression extended stepwise from the HR and PR mating-type loci. *(D)* Examples of two gene genealogies showing contrasted clustering levels of alleles at non-mating-type genes associated with the a_1 versus a_2 mating types (dark grey and light grey squares, respectively, at the tips of the gene genealogy). The left panel shows the gene genealogy of a gene belonging to the pseudo-autosomal region (or PAR), with no trans-specific polymorphism, i.e., intermingled alleles associated with a_1 and a_2 mating types. The right panel shows the gene genealogy of a gene belonging to the non-recombining region, with completely separated alleles associated with a_1 versus a_2 mating types of both *M. lychnidis-dioicae* and *M. silenes-dioicae,* and thus trans-specific polymorphism. The branch length scale is indicated at the bottom of each gene genealogy.

Table 1 : Whole genome public resources in anther-smut fungi (Microbotryum genus).

	Fungal species name	Host plant of sampling	Number of distinct genotypes	References	Public database	Project ID/ Strain ID / Accession ID*
Long read based assemblies	*M. intermedium*	*Salvia pratensis***	1	(25)	GenBank	PRJEB15277 : Microbotryum Intermedium Assembly (GCA_900096595)
	M. lagerheimii	*Silene vulgaris*	1	(25)	GenBank	PRJEB12080 : MvSv-1253-A1-R1 (GCA_900015505); MvSv-1253-A2-R1 (GCA_900013405)
	M. lychnidis-dioicae	*Silene latifolia*	2	(16) (80)	GenBank GenBank	PRJEB12080 : MvSl-1064-A1-R4 (GCA_900015465); MvSl-1064-A2-R4 (GCA_900015445) PRJNA437556 : MvSl-1318_A1 (GCA_003121365); MvSl-1318_A2 (GCA_003121355)
	M. violaceum sensu lato (M.v. caroliniana)	*Silene caroliniana*	1	(26)	GenBank	PRJEB12080: MvCa-1250-A1-R1 (GCA_900014965); MvCa-1250-A2-R1 (GCA_900014955)
	M. violaceum sensu lato (M.v. paradoxa)	*Silene paradoxa*	1	(26)	GenBank	PRJEB12080 : MvSp-1252-A1-R1 (GCA_900015495); MvSp-1252-A2-R1 (GCA_900015485)
	M. violaceum sensu stricto	*Silene nutans*	1	(25)	GenBank	PRJEB12080: MvSn-1249-A1-R1 (GCA_900015425); MvSn-1249-A2-R1 (GCA_900015455)
	M. saponariae	*Saponaria officinalis*	1	(30)	GenBank	PRJEB12080 : MvSof-1268-A1-R1 (GCA_900015975); MvSof-1269-A2-R1 (GCA_900015475)
	M. scabiosae	*Knautia arvensis*	1	(26)	GenBank	PRJEB12080 : MvKn-1118-A1-R1 (GCA_900008855); MvKn-1118-A2-R1 (GCA_900015415)
	M. silenes-acaulis	*Silene acaulis*	1	(26)	GenBank	PRJNA437556: ASM366583v1 (GCA_003665835); ASM366582v1 (GCA_003665825)
	M. silenes-dioicae	*Silene dioica*	1	(25)	GenBank	PRJEB16741 : MsdSdi1 (GCA_900120095); PRJNA437556 : MsdSdi2 (ID requested)
Whole genome shotgun assemblies	*M. carthusianorum*	*Dianthus superbus*	1	(112)	GenBank	PRJNA437556 : MvDC3-001-A2-G1 (ID requested)
	M. coronariae	*Lychnis flos-cuculi*	1	(112)	GenBank	PRJNA437556 : MvLf-1062-A1-G1 (ID requested)
	M. lagerheimii	*Silene vulgaris*	1	(112)	GenBank	PRJNA437556 : MvSv1-300-38-G1 (ID requested)
	M. lychnidis-dioicae	*Silene latifolia*	1	(104) (112)	GenBank GenBank	PRJNA41281 : p1A1 Lamole (GCA_000166175) PRJNA437556 : MvSlA1A2r2 (ID requested)
	M. major	*Silene otites*	1	(112)	GenBank	PRJNA437556 : MvSo-338-G1 (ID requested)
	M. silenes-acaulis	*Silene acaulis*	1	(112)	GenBank	PRJNA437556 : MvSa-10-04-A1-G1 (ID requested)
	M. silenes-dioicae	*Silene dioica*	1	(112)	GenBank	PRJNA437556 : MvSd-IT02-32-2-17A-A2-1141 (ID requested)
	M. silenes-inflatae	*Silene vulgaris*	1	(112)	GenBank	PRJNA437556 : Sv2-78-06-G1 (ID requested)
	M. stellariae	*Myosoton aquaticum*	1	(112)	GenBank	PRJNA437556 : MvMa-946-A1-G1 (ID requested)
	M. superbum	*Dianthus pavonius*	1	(112)	GenBank	PRJNA437556 : MvDp-1065-A2-G1 (ID requested)
	M. superbum	*Dianthus seguieri*	1	(112)	GenBank	PRJNA437556 : MvDC1-001-A2-G1 (ID requested)
	M. violaceum sensu lato	*Silene sp.*	1	(112)	GenBank	PRJNA437556 : MvSc-a-1127-A2-G1 (ID requested)
	M. violaceum sensu lato (M.v. caroliniana)	*Silene caroliniana*	1	(112)	GenBank	PRJNA437556 : MvCa-1131-A1-G1 (ID requested)
	M. violaceum sensu lato (M.v. italica)	*Silene italica*	1	(112)	GenBank	PRJNA437556 : MvSi-1128-A1-G1 (ID requested)
	M. violaceum sensu lato (M.v. lemmonii)	*Silene lemmonii*	1	(112)	GenBank	PRJNA437556 : MvSlm-001-A2-G1 (ID requested)
	M. violaceum sensu lato (M.v. paradoxa)	*Silene paradoxa*	1	(112)	GenBank	PRJNA437556 : MvSp-880-A1-G1 (ID requested)
	M. violaceum sensu lato	*Silene pusilla*	1	(112)	GenBank	PRJNA437556 : MvSpu-866-A1-G1 (ID requested)
	M. violaceum sensu stricto	*Silene nutans*	1	(112)	GenBank	PRJNA437556 : MvSn-1014-A1-G1 (ID requested)
	M. violaceo-irregulare	*Silene vulgaris*	1	(112)	GenBank	PRJNA437556 : MvSv3-001-G1 (ID requested)
Sequence archive (reads)	*M. lychnidis-dioicae*	*Silene latifolia*	39	(134) (15)	NCBI Short Read Archive NCBI Short Read Archive	PRJNA269361 PRJNA295022
	M. silenes-dioicae	*Silene dioica*	19	(15)	NCBI Short Read Archive	PRJNA295022
	M. violaceum sensu lato (M.v. paradoxa)	*Silene paradoxa*	4	(26)	NCBI Short Read Archive	PRJEB16741
	M. violaceum sensu lato (M.v. caroliniana)	*Silene caroliniana; Silene virginica*	11	(26)	NCBI Short Read Archive	PRJEB16741
	M. saponariae	*Saponaria officinalis*	1	(56)	GenBank	PRJEB11435

*Information were retrieved on public databases on 22.11.18. For long read based assemblies, assemblies of the two mating type a1 and a2 are indicated if available.
ID requested indicate that the genomic data were submitted to the public database and are currently processed
**the sequenced strain was sampled on *Salvia pratensis*, although the fungal species is usually found on *Scabiosa* hosts

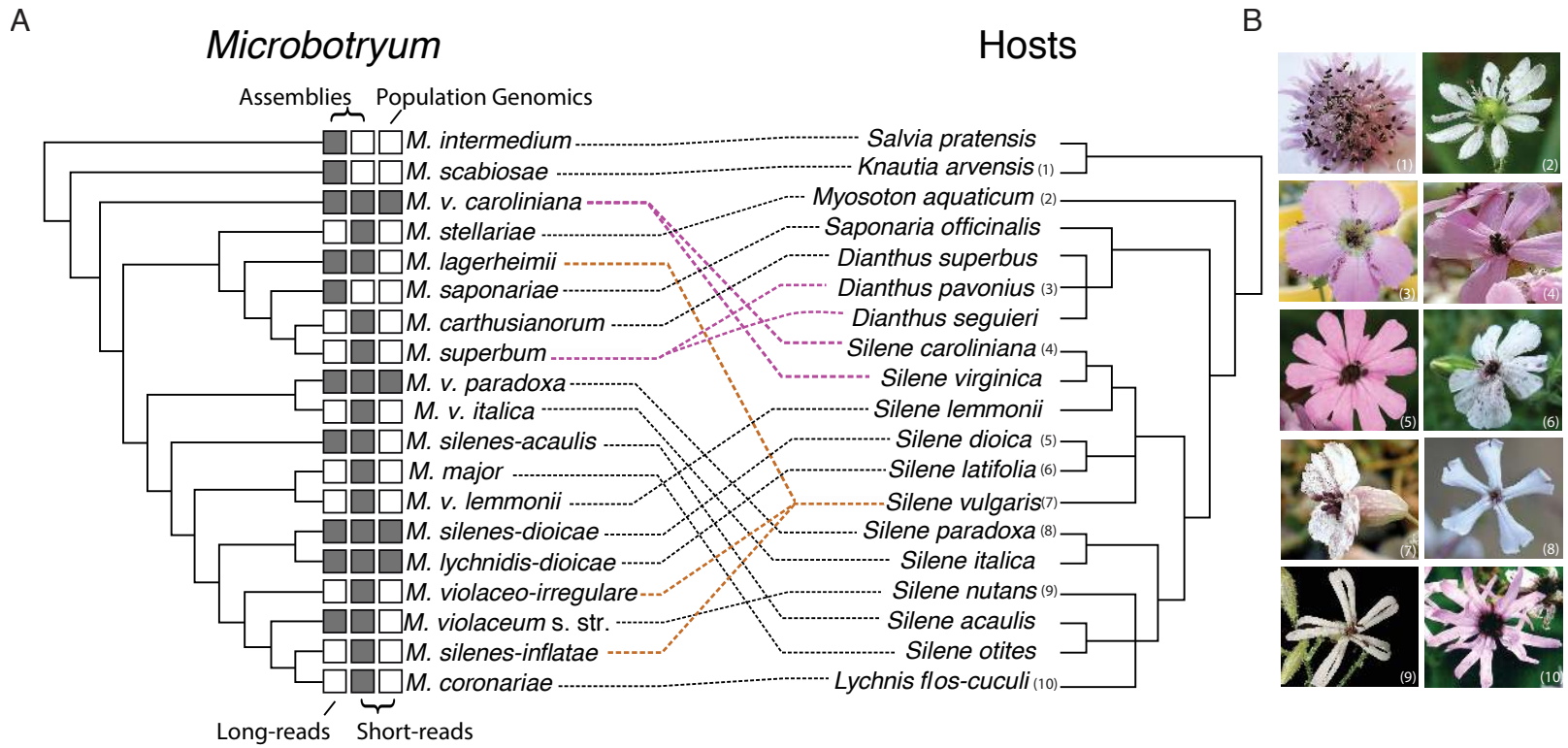

A
Microbotryum
Hosts
B
Assemblies
Population Genomics
M. intermedium
M. scabiosae
M. v. caroliniana
M. stellariae
M. lagerheimii
M. saponariae
M. carthusianorum
M. superbum
M. v. paradoxa
M. v. italica
M. silenes-acaulis
M. major
M. v. lemmonii
M. silenes-dioicae
M. lychnidis-dioicae
M. violaceo-irregulare
M. violaceum s. str.
M. silenes-inflatae
M. coronariae
Salvia pratensis
Knautia arvensis (1)
Myosoton aquaticum (2)
Saponaria officinalis
Dianthus superbus
Dianthus pavonius (3)
Dianthus seguieri
Silene caroliniana (4)
Silene virginica
Silene lemmonii
Silene dioica (5)
Silene latifolia (6)
Silene vulgaris (7)
Silene paradoxa (8)
Silene italica
Silene nutans (9)
Silene acaulis
Silene otites
Lychnis flos-cuculi (10)
Long-reads
Short-reads
(1)
(2)
(3)
(4)
(5)
(6)
(7)
(8)
(9)
(10)

Figure 1

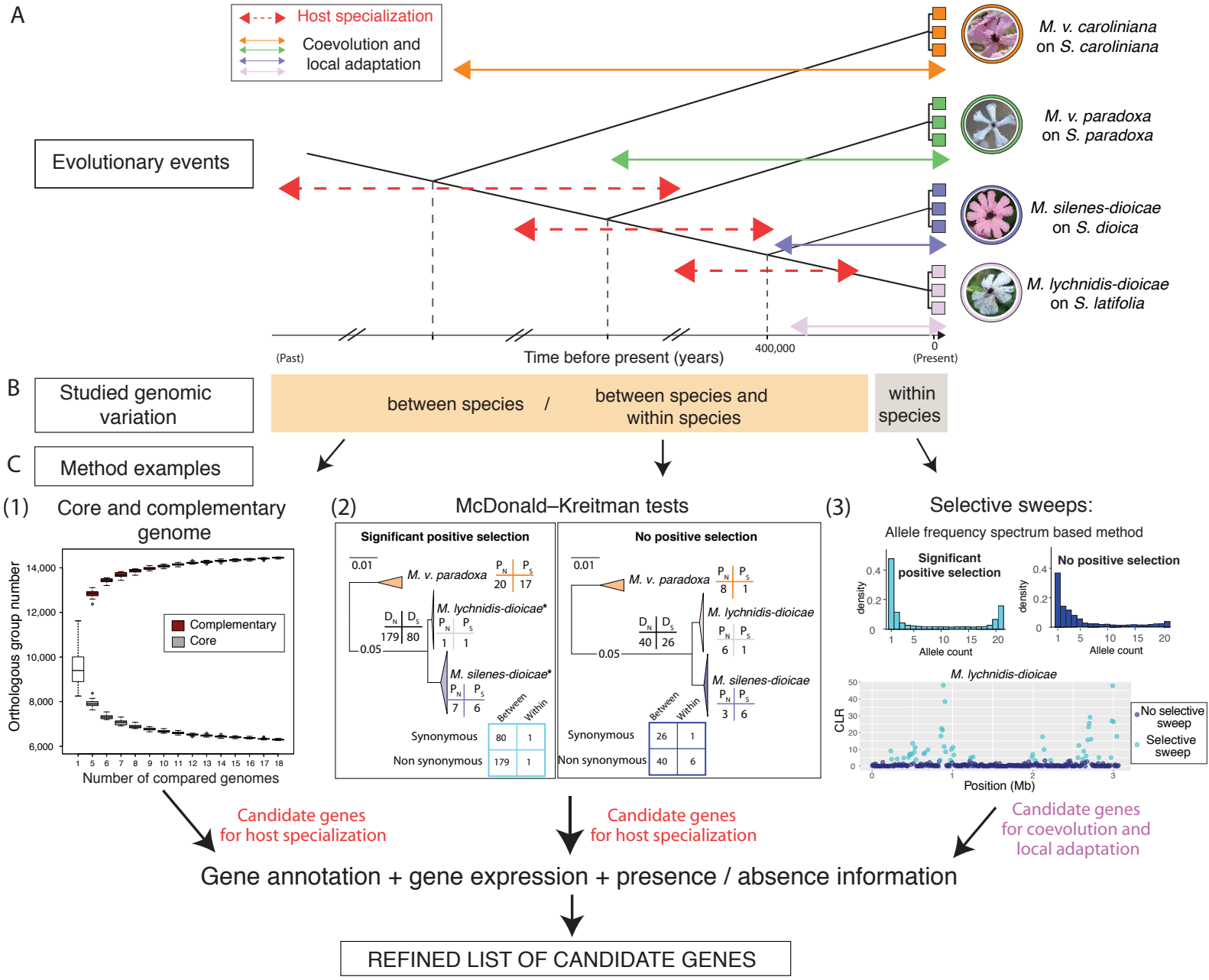
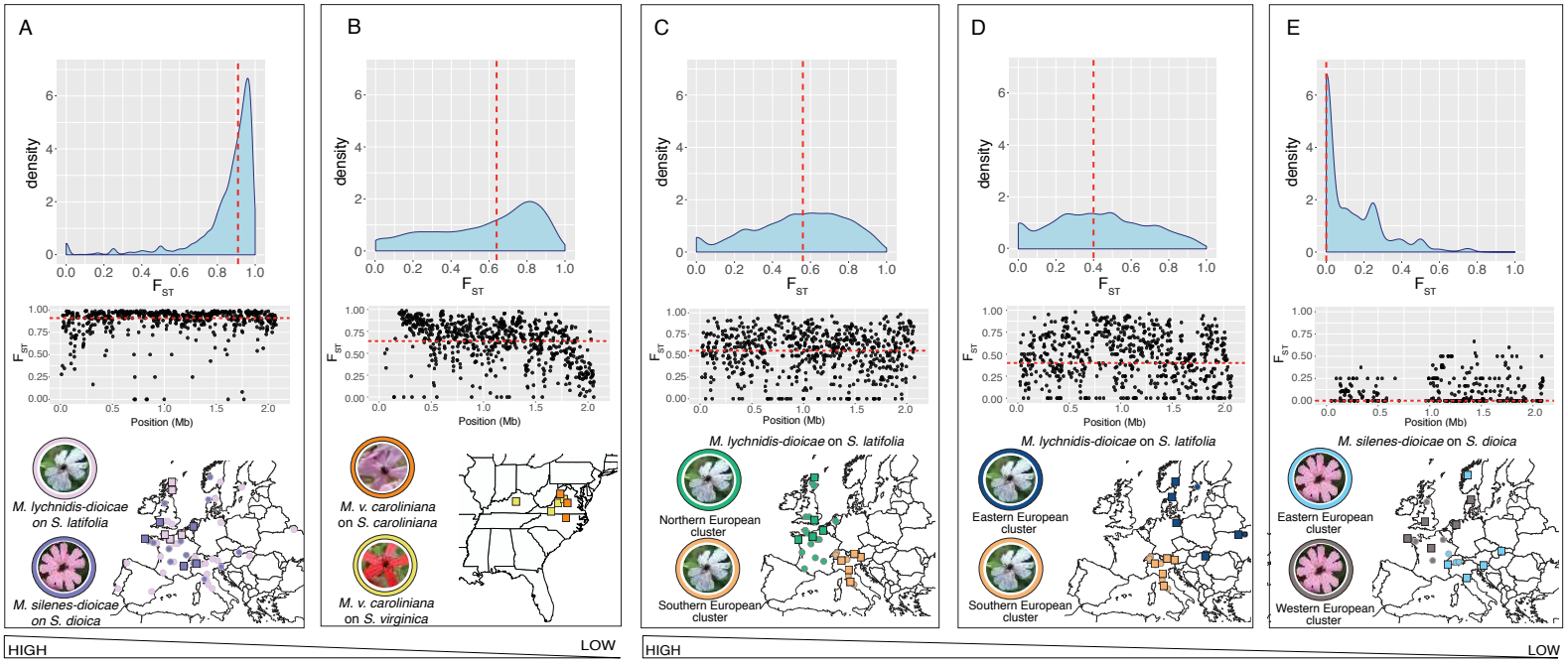

A
Host specialization
Coevolution and local adaptation
Evolutionary events
M. v. caroliniana on S. caroliniana
M. v. paradoxa on S. paradoxa
M. silenes-dioicae on S. dioica
M. lychnidis-dioicae on S. latifolia
(Past)
Time before present (years)
400,000
0
(Present)
B
Studied genomic variation
between species / between species and within species
within species
C
Method examples
(1) Core and complementary genome
Orthologous group number
Complementary
Core
Number of compared genomes
(2) McDonald–Kreitman tests
Significant positive selection
No positive selection
M. v. paradoxa
M. lychnidis-dioicae*
M. silenes-dioicae*
M. lychnidis-dioicae
M. silenes-dioicae
Between
Within
Synonymous
Non synonymous
(3) Selective sweeps:
Allele frequency spectrum based method
density
Allele count
M. lychnidis-dioicae
CLR
Position (Mb)
No selective sweep
Selective sweep
Candidate genes for host specialization
Candidate genes for host specialization
Candidate genes for coevolution and local adaptation
Gene annotation + gene expression + presence / absence information
REFINED LIST OF CANDIDATE GENES

Figure 2



Divergence between different host-specialized species or lineages

Divergence between lineages within a species specialized on one host

Figure 3

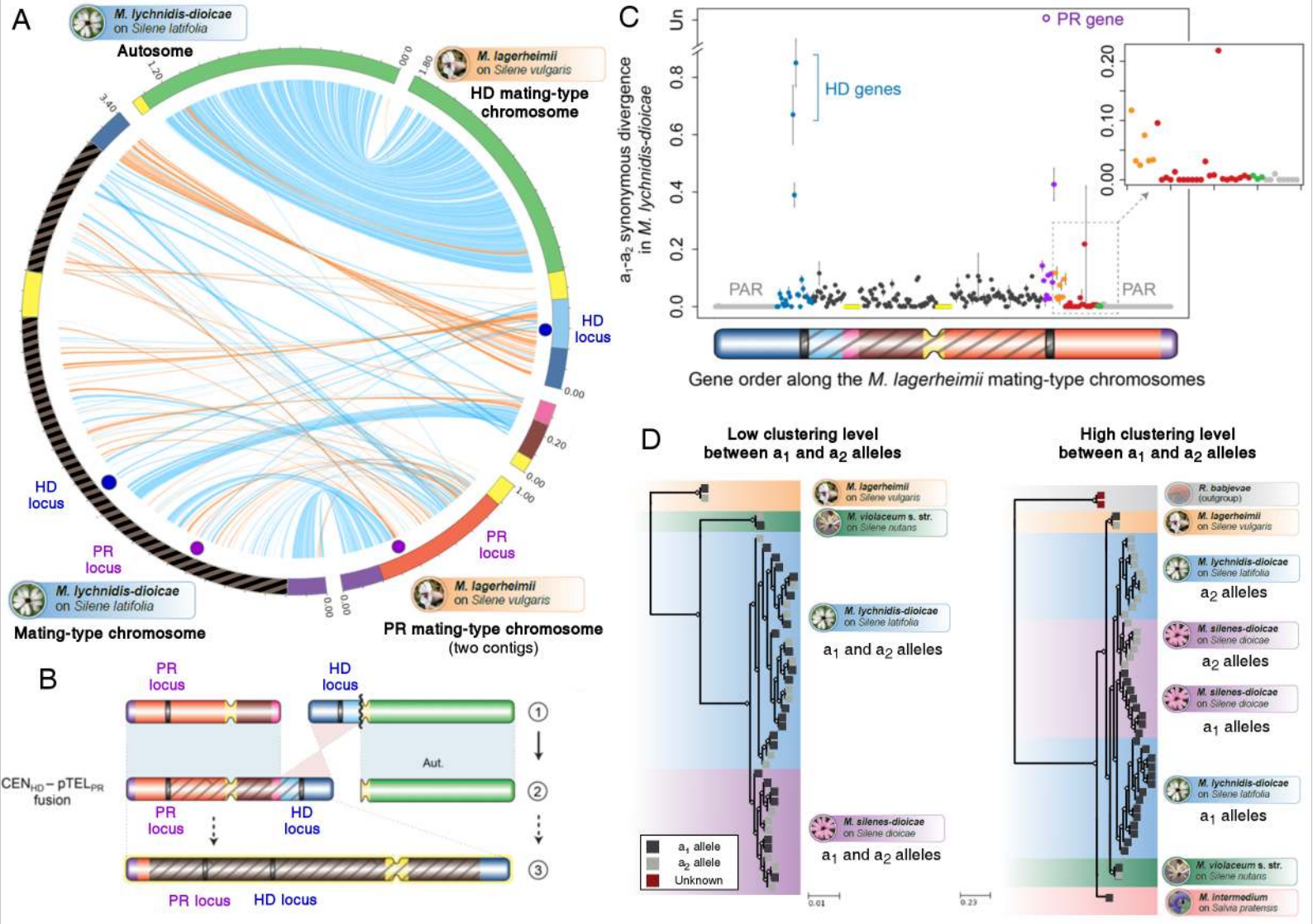

A
M. lychnidis-dioicae on Silene latifolia
Autosome
M. lagerheimii on Silene vulgaris
HD mating-type chromosome
HD locus
PR locus
HD locus
PR locus
M. lychnidis-dioicae on Silene latifolia
Mating-type chromosome
M. lagerheimii on Silene vulgaris
PR mating-type chromosome (two contigs)
B
PR locus
HD locus
Aut.
CEN_HD – pTEL_PR fusion
PR locus
HD locus
PR locus
HD locus
C
PR gene
HD genes
a1-a2 synonymous divergence in M. lychnidis-dioicae
PAR
PAR
Gene order along the M. lagerheimii mating-type chromosomes
D
Low clustering level between a1 and a2 alleles
High clustering level between a1 and a2 alleles
M. lagerheimii on Silene vulgaris
M. violaceum s. str. on Silene nutans
M. lychnidis-dioicae on Silene latifolia
a1 and a2 alleles
M. silenes-dioicae on Silene dioicae
a1 and a2 alleles
a1 allele
a2 allele
Unknown
R. babjevae (outgrowth)
M. lagerheimii on Silene vulgaris
M. lychnidis-dioicae on Silene latifolia
a2 alleles
M. silenes-dioicae on Silene dioicae
a2 alleles
M. silenes-dioicae on Silene dioicae
a1 alleles
M. lychnidis-dioicae on Silene latifolia
a1 alleles
M. violaceum s. str. on Silene nutans
M. intermedium on Salvia pratensis

Figure 4