Population genomic processes in Brassicaceae in the light of mating system variation

Tiina M. Mattila1 & Tanja Slotte2

1 Ecology and Genetics Research Unit, University of Oulu, PO Box 3000, 90014, Oulu, Finland

2 Department of Ecology, Environment and Plant Science, Stockholm University, Stockholm, Sweden

ORCiDs: 0000-0002-1298-7370 (Tiina M. Mattila), 0000-0001-6020-5102 (Tanja Slotte)

Contact email: tiina.mattila@oulu.fi

Abstract

Mating system in plants is a highly variable trait ranging from obligatory outcrossing to full self-fertilization, i.e. selfing. In nature, there are various mechanisms to prevent self-fertilization, indicating evolutionary advantages of outcrossing. Still, self-fertilization is very common in plants. In addition, selfing is a useful property in plant breeding and research since it allows replication and maintenance of the same genotype over generations. Hence, it is of great interest to understand the evolutionary drivers of mating system shifts, as well as population genomic consequences across different time scales. In recent years, advances in population genomics methodologies, as well as large-scale population sequencing projects, have provided multiple insights into plant mating system variation, particularly in Brassicaceae model systems. Utilizing information from Brassicaceae systems, we review the effects of self-fertilization and genetically controlled single locus self-incompatibility (SI) on population genetic and genomic variation. The empirical results are focused on model systems from the Brassicaceae family, where in the ancestral form, high levels of outcrossing are ensured by genetically controlled sporophytic SI. The standard

form of SI in Brassicaceae is controlled by a single tightly linked genetic locus, the S-locus, containing two genes, the male specificity *SCR*/*SP11* and the female specificity *SRK* gene, involved in the self-recognition reaction. Loss-of-function at these S-locus genes or in the downstream cascade leading to the SI reaction leads to self-compatibility (SC) allowing self-fertility. Breakdown of SI and associated selfing can be favorable under specific circumstances, such as if mate availability is low, for example, in colonizing or range edge populations. Population investigations have characterized multiple independent losses of SI across the Brassicaceae, which has provided a framework to investigate the genome-wide diversity patterns associated with different mating systems. By summarizing the current knowledge on the population genomic properties of selfing and SI populations in Brassicaceae model systems, we aim to improve our understanding of the evolutionary forces shaping mating system variation.

1 Introduction

Even though hermaphroditism in plants is common, many flowering plants harbor independently evolved self-incompatibility (SI) mechanisms that prevent self-fertilization [1]. However, there is high variation in rates of self-fertilization, and many species are even fully self-fertilizing. This indicates that there are contrasting evolutionary forces shaping the mating system of species and populations. The mating system is also a key parameter affecting plant diversification [2] and evolutionary trajectories. For example, a phylogenetic study in Solanaceae found that speciation rates were higher in self-compatible (SC) lineages, which are capable of selfing [3]. However, the estimated total extinction rate was also higher in SC lineages, and the overall net diversification rates were lower for SC lineages [3] shaping the balance of species numbers of different mating systems. This study thus highlights the macroevolutionary importance of mating system variation.

Contrasting patterns of genetic variation in populations with variable mating systems has provided insights into the underlying processes associated with different mating systems. Based on theoretical work, the balance between inbreeding depression and the genetic transmission advantage of selfing, as well as reproductive assurance [4], are central

54 determinants of the optimal mating strategy [5]. These factors strongly depend on the
55 population history of the species, which is also affected by the mating system. The classical
56 population genetic consequences of elevated self-fertilization include smaller effective
57 population size, reduced effective recombination rate, and reduced heterozygosity, which
58 together in turn predict decreased levels of efficacy of selection and accumulation of
59 deleterious mutations (reviewed in [6–9]), as well as increased linked selection [10, 11].
60 Furthermore, self-fertilization also restricts gene flow from sister lineages, increasing
61 genetic differentiation via exacerbated genetic drift, which further influences the
62 accumulation of genetic incompatibilities and affects the speciation process [12], although
63 the mechanisms and direction are not fully understood [13]. Empirical evidence on the
64 population genetics effects has, however, not been as clear as predicted. This can be due to
65 the confounding effects of multiple co-occurring processes affecting the genome-wide
66 pattern of variation, such as demographic perturbations [14] and associated changes in
67 genome evolution, such as polyploidization [15]. In addition, recent theoretical and
68 simulation-based approaches have indicated that selfing has complex consequences on
69 various population genetic processes that are also dependent on the underlying genetic
70 properties. For example, local recombination rate and the level of dominance have been
71 shown to affect purifying selection strength and fitness in selfing populations versus
72 outcrossing populations [16]. In more detail, in low recombination regions, purifying
73 selection can also be efficient in selfing populations relative to outcrossers, especially in the
74 presence of recessive alleles. In turn, in high recombination regions with a low proportion of
75 recessive alleles, the outcrossing populations have clearly more efficient purifying selection
76 [16]. Furthermore, the fixation probability of deleterious alleles is affected by the interaction
77 of the selfing rate and whether it is expressed in the haploid reproductive (gametophytic) or
78 diploid asexual (sporophytic) stage [17]; recessive deleterious alleles exclusively expressed
79 in the gametophytic stage are more easily fixed in selfing populations, while the opposite
80 holds for deleterious variants exclusively expressed in the sporophytic stage. Finally, selfing
81 has been shown to have an effect on the allelic expression in the gametic phase and hence
82 has an effect on the selection landscape in the presence of pollen or sperm competition

[18]. This prediction also has empirical support from a contrast of distributions of fitness effects in pollen-specific genes in outcrossing and partially selfing populations of *Arabis alpina* [19]. When investigated in empirical datasets, such processes and their magnitude can also be affected by the timing of transition to selfing and selfing rate, and hence it is important to understand these properties in any selfing population under investigation. However, many modelling-based predictions obtained via scaled forward simulations can be biased due to unintended background selection [20]. In addition, the estimation of the distribution of fitness effects itself can be complicated by Hill-Robertson interference (a specific type of linked selection) in selfing populations [21].

In this review, we will highlight empirical and methodological advances in understanding population genomic processes in selfing and SI outcrossing populations from multiple Brassicaceae species. The most well-studied model systems are *Arabidopsis* and *Capsella,* while *Brassica* has mostly been used in molecular characterizations of the S-locus. In *Arabidopsis*, the most studied groups are the diploid and predominantly SI diploid *A. lyrata* and *A. halleri*, the autopolyploid SI *A. arenosa*, the diploid SC and selfing *A. thaliana,* as well as the autopolyploid selfers *A. suecica* and *A. kamchatica*. In *Capsella,* there are at least 3 species of which only one, *C. grandiflora,* is SI (see [6] for more details). In recent years, new, more variable model systems particularly suitable for population genetic and genomic investigations have also emerged, such as *Arabis alpina*. Reference genomes and genome-wide datasets available for SC and SI strains, such as for instance SC *A. thaliana* [22], SI *A. halleri* [23], SI/SC, *A. lyrata* ssp. *lyrata* [24], SI/SC *A. lyrata* ssp. *petraea* [25], SI *A. arenosa* [26], SC *C. rubella* [27], SC *C. orientalis* [28], SC allopolyploid *C. bursa-pastoris* [29], *C.* SI/SC *A. alpina* [30, 31], SC *Cardamine hirsuta* [32], SC *Leavenworthia alabamica* [33], SC *Sisymbrium irio* [33], SC *Thellungiella parvula* (aka *Schrenkiella parvula* [34]) [35], SC *Eutrema salsugineum* [36], SC *Aethionema arabicum* [33] have allowed investigations of multiple aspects of SI and its breakdown [34] across the Brassicaceae phylogeny (Fig. 1). These data also provide an excellent framework and resources for further studies investigating population genomics of mating system evolution in this plant family. Here, by compiling empirical results from Brassicaceae model systems, we first describe the

functioning of the S-locus from a population genetics perspective. We also discuss how advances in S-locus sequencing and methodologies in genomics have eased the population genetic study of variation in this multiallelic locus. Second, we review population genetics and genomics approaches and empirical studies to estimate selfing rates and reconstruct the paths to selfing. Finally, we focus on genetic and other population processes associated with the transition to selfing and discuss how these link to biogeographic patterns of species distributions.

2 Properties and evolution of the Brassicaceae S-locus

SI in many Brassicaceae is controlled by a molecular mechanism that contains two major actors involved in the SI reaction: the female specificity *S-locus receptor kinase* (*SRK*, Arabidopsis Genome Initiative/AGI locus code [42]: AT4G21370) expressed on the stigma surface and the male specificity *S-locus cysteine-rich protein/S-locus protein 11* (*SCR*/*SP11*) that is deposited on the pollen coat [43–45]. To recognize self-pollen, these genes are arranged in a tightly linked non-recombining [45, 46] multiallelic genomic region known as the self-incompatibility locus (S-locus). Suppression of recombination in this genomic region is important to keep the incompatible gene forms in the same allele, preventing incompatibility between unrelated alleles. The self-recognition reaction prevents pollen tube growth on the stigma surface and hence prevents self-fertilization [43, 47].

In SI *A. lyrata* and *A. halleri*, functional gene copies of well-studied SCR alleles contain two exons and code for proteins of approximately 85 amino acids, whereas SRK genes comprise seven exons and encode proteins of about 850 amino acids [34]. In many Brassicaceae, the S-locus is located between the genes *PUB8/U box* (AGI locus code: AT4G21350) coding for U-box domain-containing protein 8 and ARK3 (AGI locus code: AT4G21380) [32, 48]. In *A. lyrata* [25], *Cardamine hirsuta* [32] and *C. orientalis* [49] genomes, which all have 8 chromosomes, the S-locus is located on chromosome 7 (see Fig. 2 for schematic illustration). In *A. thaliana*, where the ancestral 8 chromosome number was reduced to 5 due to loss of genetic material, translocations, and chromosome fusions [24], the S-locus is

located on chromosome 4. The size of the S-alleles varies considerably between alleles and species, due to rearrangements and repeat element accumulation in non-coding regions (see Fig. 2 for examples). For example, in the SC *C. hirsuta, the* S-locus size of the allele in the sequenced individual was 18 kb, while in SI *A. lyrata,* the sizes of the different S-alleles range from 31 kb to more than 109 kb with high divergence between different alleles and transposable element accumulation in non-coding regions [32, 48, 50].

The exact origin of the Brassicaceae type SI is still not known, but the patterns of variation have been used to reconstruct the evolution of the S-locus. One model investigating eight *Brassicaceae* species indicated that the gain of SI happened before the split of Camelinodae (represented by *Arabidopsis* and *Capsella* sp.) and Brassicodae (represented by *Eutrema salsugineum*, *Schrenkiella parvula,* and *Sisymbrium irio*) [34]. This split has been dated to ca. 17–25 million years ago (MYA) [37, 38], although much older estimates have also been obtained, ca. 43 MYA [51]. These estimates serve only as a lower bound for the gain of SI since it has so far been impossible to separate between non-existence and ancient loss of SI and S-locus in the SC *Aethionema arabicum*, which represents the most basal lineage of Brassicaeae [34]. Age estimates based on S-locus variation have shown that some alleles have diverged at least 40 MYA [52], which also serves as a lower bound for the gain of SI.

Despite the orthology of the S-locus in many Brassicaceae species, the S-locus region between *PUB8* and *ARK3* is void of genes in Leavenworthia [32, 53, 54]. A diallel crossing experiment, however, indicated the existence of a single locus sporophytic SI in *Leavenworthia alabamica* [55]. SRK amplification attempts resulted in the finding of multiple genes of sequence homology with *SRK* [55], one of which, *Lal2*, harbored high genetic variation and the alleles co-segregated with the crossing-based compatibility S-alleles. A phylogenetic analysis indicated that this gene diverged much earlier than the *SRK* alleles between species and was linked to a *SCR-like* (aka *SCRL*) gene [53]. The suggested evolutionary model includes the loss-of-function of the *SRK*/*SCR* mediated SI and the neofunctionalization of *Lal2* and *SCRL*, which were rearranged to form the new S-locus. This event likely took place after the lineage splits between *Cardamine* and *Leavenworthia* (ca. 8 MYA [38]), since the *Cardamine hirsuta* genome contains the syntenic S-locus with

Arabidopsis and *Capsella*, but not the *Leavenworthia*-type S-locus [32]. This is an intriguing case of very recent gain-of-function S-locus evolution that could have involved transition via SC intermediate [53].

One of the most striking properties of the S-locus is its high number of very diverged alleles. This is due to negative frequency-dependent selection [56] (a form of balancing selection), where rare alleles are favored by selection due to higher mate availability. This increases the probability that multiple alleles are maintained in a population for a long time. It has been suggested that the most favorable path to new alleles is through an SC form where the pollen specificity component is first mutated, while the second mutation in the stigma component completes the evolution of the new fully SI allele [57]. An alternative way to obtain new variation in the S-locus is trans-specific gene flow. For example, in *Arabidopsis* species, patterns of variation in the S-locus suggested introgression of alleles between species [58]. Once established, the S-alleles can persist for millions of years, resulting in very deep divergence of S-alleles with trans-specific polymorphism [59].

The Brassicaceae self-recognition mechanism is sporophytic, meaning that the expressed incompatibility type is determined by the parental genotype [60]. The expressed allele depends on the dominance hierarchy between the S-haplotypes. In pollen, this allele silencing is determined by small RNAs in the dominant alleles [61–63]. In the stigma, the molecular understanding of the determination of dominance is much more limited [44]. In outcrossing Arabidopsis species, S-alleles are classified into four dominance classes [64], commonly referred to as Roman numerals I (most recessive), II, III, and IV (most dominant) (see, e.g., [48]). This type of expression dominance hierarchy with negative frequency-dependent selection is expected to cause unequal equilibrium allele frequencies [65], with recessive alleles being more common than dominant alleles.

Due to the functioning of the S-locus, maintenance of high polymorphism is important for population viability. To understand this diversity, it can be of interest to genotype the S-locus at the population level. However, due to long independent evolution of S-alleles, the low sequence similarity between alleles and accumulation of transposable elements poses

196 challenges in genotyping the S-locus [66]. The sequencing of BAC cloned fragments has
197 been used in some instances [67], and now, facilitated by the development of long-read
198 sequencing methods [25], investigating the evolution of the entire S-locus has become more
199 feasible. In addition, knowledge of sequence variation of different S-alleles can also be used
200 to genotype S-alleles based on coding regions of *SCR* and *SRK* [68]. Particularly for
201 *Arabidopsis* species, more than 20 years of work in molecular characterization of the S-locus
202 has produced sequence information on a large number of S-haplotypes that can be utilized
203 in S-locus genotyping using a mapping-based approach or in *de novo* assembly of S-locus
204 genes [25, 69]. A genotyping tool using such an approach is implemented in the tool
205 NGSgenotyp [69], which is specifically designed for this purpose. The original publication
206 includes a database of the extracellular domain of SRK (eSRK) for allele assignments. In the
207 mapping approach, genotypes are assigned based on reads mapping to different S-alleles,
208 and if the mismatch rate between the matching allele and the query reads is low, this results
209 in high allele assignment certainty. The read count ratio with paralogous loci can be used to
210 evaluate if all alleles were identified. If the allelic database does not match the studied
211 population very well, which may be an issue, particularly in less well-investigated species,
212 the assembly function of NGSgenotyp can be used to investigate unknown S-alleles [68, 69].

213 The structure and functioning of the S-locus also influence the surrounding genetic diversity.
214 The genetic signals of balancing selection, such as increased nucleotide diversity and
215 distorted allelic phylogenies, were observed in a ca. 10-30 kb region around the S-locus in
216 *A. lyrata* and *A. halleri* [70, 71]. Furthermore, suppression of recombination and high
217 heterozygosity compromise the efficacy of deleterious allele removal. This phenomenon is
218 known as *sheltered load*. Population genetic theory predicts that the rarer alleles (lower
219 frequency expected for the alleles in the higher dominance category [65]) accumulate more
220 deleterious mutations due to lower chances for recombination [72]. This prediction was
221 tested in *A. lyrata* and *A. halleri* [70] using a combination of population genomics
222 investigation of load along S-haplotypes of different dominance classes, as well as
223 phenotypic investigation of individuals with homozygous S-genotypes. In the crossing
224 experiment, homozygosity in three different alleles with different dominance classes was

generated using crosses between individuals carrying the same haplotypes masked by dominant alleles in the pollen donor. The proportion of homozygotes followed Mendelian expectations for the most recessive and most dominant allele crosses. In the intermediate dominance allele cross, there was a significant deficit of homozygotes, suggesting recessive load in this allele. The phenotypic comparison of homozygous and heterozygous individuals did not show large differences, except that the homozygous individuals had longer flowering stem length [70]. Furthermore, accumulation of deleterious alleles in different S-haplotypes were studied by investigating the number of non-synonymous (amino acid changing) and its ratio to synonymous (amino acid not changed) mutations to quantify population genetic signatures of genetic load, assuming that most amino acid changing mutations are deleterious. When quantifying the total number of such changes in the S-locus region, no overall difference was observed between the S-haplotypes of different dominance classes This unexpected finding led the authors to investigate the expectations in more detail, and they found that the number of fixed deleterious alleles was higher in the dominant category, but the opposite was found for the level of segregating load, which results in no overall difference in genetic load between the alleles of different levels of dominance [70]. These findings advance our understanding of the detailed population genetic and genome-wide effects SI.

3 Identification of self-compatible lineages

Despite the widespread presence of sophisticated molecular mechanisms preventing self-fertilization, transitions to SC and selfing from an ancestral state of SI are extremely common in Brassicaceae. To understand the evolutionary pathways of these transitions, a combination of functional and population genetics, as well as genomics approaches, can be used to reconstruct the evolutionary histories of selfing populations and species. In the case of breakdown of SI in Brassicaceae, an increasing number of independent losses have been identified. These instances of SI loss have provided a possibility to also investigate the timeframe of repeated mating system transitions. Before such analyses can be initiated, however, it is important to determine the mating system of the study populations.

To do so in the first place, experimental crosses can be performed to investigate SI and SC. It is worth noting that SC does not necessarily lead to selfing but only allows it, potentially leading to inbreeding. Different population genetics methods can be utilized to estimate the level of inbreeding (e.g., [73]); a very common and simple method measures genotype frequency deviation from Hardy-Weinberg expectation within a population using the F_{IS} from Wright's F-statistics [74], calculated as follows

F_{IS} = 1- observed heterozygosity/expected heterozygosity

Where expected heterozygosity is $2pq$ from the Hardy-Weinberg equation [75], and p and q are the allele frequencies for biallelic loci. In predominantly selfing populations, heterozygosity is expected to be lost very quickly, and if purely selfing, F_{IS} is close to 1, while in purely outcrossing large populations, F_{IS} is close to 0. Negative F_{IS} values indicate an excess of observed heterozygotes that can be due to, for example, balancing selection acting on the investigated or linked genetic loci, as for example in the case of S-locus genotypes. Alternatively, negative values can also indicate genotyping errors due to, for example, comparison of paralogous loci. Under equilibrium, F_{IS} can be directly used to estimate selfing rates. As an empirical example, F_{IS}-based selfing estimates in *A. alpina* indicated that the selfing rate in this species ranges from almost fully selfing to mixed-mating and fully outcrossing populations [19, 76–78]. The form and variability in SI, outcrossing, selfing, and SC were also verified by genotyping maternal plants and offspring progeny, controlled crossing experiment, S-locus genotyping, and pollen tube growth visualization [77].

Accurate estimation of selfing rates can, however, be difficult as different types of population processes (e.g., mating between close relatives, also known as biparental inbreeding, and demographic processes) can affect levels of heterozygosity [79]. Long stretches of homozygous genotypes in diploid genomes, i.e., runs of homozygosity (ROHs), can differentiate low heterozygosity due to recent inbreeding and long-term background inbreeding due to demographic history. This can be separated based on the number of ROHs (NROH) and the total size of ROHs (SROH) [80]. Furthermore, ROHs also have the potential to distinguish between different forms of inbreeding (selfing vs. biparental inbreeding).

Zeitler and Gilbert [81] leveraged this property and generated a machine learning algorithm to estimate individual-based self-fertilization rates from whole-genome data. This method was used to estimate ROH-based selfing rates from presumably highly selfing *A. thaliana,* multiple *A. alpina* populations that are previously estimated to be self-incompatible or mixed mating, and a Siberian self-compatible lineage of *A. lyrata* [25]. These results confirmed the high selfing in *A. thaliana* [82] and mixed mating in *A. alpina* [78]. The two populations from the self-compatible lineage of Siberian *A. lyrata* were found to be mixed-mating, with a relatively large difference in selfing rates: 0.39 for the more southern NT1 population and 0.82 for two individuals further to the north [25, 81].

Another method for selfing rate estimation utilizes the Sequential Markovian Coalescent (SMC) to infer selfing rates in concert with demographic history and level of dormancy [83]. Multiple different SMC methods have been developed for demographic inference, but the previous tools do not consider these additional parameters that violate the standard population genetics model assumptions of non-overlapping generations and random mating. The new method, eSMC (see Chapter 7), performs much better in demographic inference, particularly in scenarios with selfing and dormancy/seed-banks, in comparison with routinely used PSMC (pairwise sequentially Markovian coalescent) [84], MSMC [85], and MSMC2 [86] (multiple sequentially Markovian coalescent) in a demographic scenario of fluctuating population size. In both cases, the most recent population size peak tends to be heavily overestimated by the other methods, which do not account for selfing and dormancy [83]. Furthermore, in the seed banking scenario, the entire demographic model was shifted from the true model if not simultaneously modelled. Applying eSMC to German and Swedish *A. thaliana* populations resulted in population average selfing rate estimates of 0.86 and 0.87, respectively, and no seed banking was observed. It was noted that in this method, the accurate simultaneous estimation of selfing and dormancy required setting priors for the estimated parameters based on external ecological knowledge [83]. For example, the original publication used uniform priors ranging from 0.5 to 1 for both the germination rate and the selfing rate in *A. thaliana* [83].

4 Investigating molecular paths and timing of transition to selfing

Once selfing populations or species have been identified, an evolutionary investigation of the transition to selfing can be initiated. Three aspects are of particular interest: 1. What are the mutations behind SC? 2. Does the transition have a single origin? 3. When and where did the transition occur? [87] When in search of the underlying mutation causing selfing, the molecular understanding of the functioning of the S-locus helps to direct the research, while identifying the exact mutation that initially allowed selfing may be difficult. For example, the number of S-alleles in the selfing population can be used to draw scenarios of how selfing has evolved [34]. *A. thaliana*, for instance, which is predominantly selfing throughout its distribution range, carries multiple highly divergent non-functional S-haplotypes that likely derive from the ancestral functional S-haplotypes. This indicates that the transition to selfing was not due to the fixation of a single loss-of-function mutation at the S-locus [34, 50]. However, the presence of all known S-haplotypes in the Moroccan *A. thaliana* population suggests that selfing originated in a single geographical region [88].

The underlying mutation that caused the breakdown of SI is not always straightforward to pinpoint, especially when the transition happened hundreds of thousands of years ago and there was time for other mutations to accumulate. In *A. thaliana,* transformation of S-locus genes with a functional *A. lyrata* allele restores self-incompatibility in the Col-0 background, indicating that the causal mutation is located in this region [89]. The likely causal mutation is the *SCR* frame-disruptive 213 bp inversion, which is found in the majority of accessions across Europe [90]. In other selfing lineages, there is less information on the causal mutation, but in some cases, the genomic locus has been identified. For example, using long-read sequencing of the Northeastern Siberian selfing *A. lyrata* [25] lineage (NT1), in the mostly SI Siberian *A. lyrata* ssp. *petraea* complex, it was shown that this lineage lacks *SRK* entirely. However, it was stated that this was not likely the original causal mutation behind selfing, but that the transition is S-locus linked. This differs from the North American selfing *A. lyrata*, where in at least one of the selfing populations [91], the results of large-scale

controlled crosses were in line with a model where selfing is caused by a genetic modifier not linked to the S-locus [92, 93]. However, the molecular identity of this modifier is still unknown. Based on between-population crosses with contrasting breeding systems (SC × SI), it was shown that the modifier caused SC in individuals homozygous for the most recessive S-allele (S_1), whereas combinations involving dominant S-alleles yielded SI progeny. [92].

In *Capsella orientalis*, which is a SC diploid species distributed in temperate latitudes in Eastern Europe and Central Asia [40], a quantitative trait locus (QTL) mapping experiment between selfing SC *C. orientalis* and SI *C. grandiflora* revealed that SC is a dominant trait and a single major QTL peak was found in the chromosome region containing the S-locus [49]. In further investigations of the entire S-locus using BAC sequencing, it was found that the *SCR* gene has a single-base frameshift deletion in its coding sequence [49]. Taken together, studies in *Capsella* are in line with the theoretical prediction that loss of self-incompatibility is more likely to evolve via mutations at the S-locus, and especially at the *SCR* gene, than at unlinked modifiers [94], whereas the results from *Arabidopsis* demonstrate the possibility for multiple routes.

While confident identification of the causal mutation leading to SC can be challenging, other aspects of selfing species evolution can be investigated using population genetic and genomic methods. For example, several methods have been developed to estimate the timing of the transition to selfing. Such information can also be utilized to infer the geographic origins of the mating system variation. For instance, the earliest estimate for the timing of transition to selfing in *A. thaliana* was obtained using a method that assumes relaxation of selection constraint (ratio of synonymous to non-synonymous substitutions, $d_N/d_S < 1$) in the S-locus genes, after the SI is lost (with d_N/d_S then gradually approaching 1). In a simple two-species model with SI and SC species, the current d_N/d_S in the SI and SC species depends on the species split time (t_{SPLIT}), the timing since SC evolved (t_{SC}), and the background d_N/d_S. Based on this model, t_{SC} can be estimated. Using this method, Bechsgaard et al. [95] estimated t_{SC} for *A. thaliana* to be 413,000 years ago, assuming a divergence time between *A. lyrata* and *A. halleri* of 5 MYA. If using divergence estimates of 8

MYA, 13 MYA or 17.9 MYA from Beilstein et al. [51], the t_{SC} would be 660,800, 1,073,800 or 1,478,540, years ago, respectively.

Similarly, S-locus variation-based estimates have also been obtained for the selfing *Capsella* species, but using slightly different methodological approaches [49, 96]. In these studies, the times to the most recent ancestor (TMRCA) of S-haplotypes within selfing lineages (lower bound) and between selfing and outcrossing lineages (upper bound), were used to set boundaries for the evolution of selfing. In *C. rubella,* such an approach yielded a very recent selfing evolution estimate, ca. 14,000-65,000 years ago, based on variation in *SRK* [96]. *SCR* variation gave a similar estimate for the lower bound, but no estimate could be obtained for the upper bound [96]. This study used a mutation rate estimate of 1.5×10^{-8} bp^{-1} obtained based on calibration of synonymous substitution rates with pollen data [97]. However, using a slower – and now considered more realistic – mutation rate estimate of 7.1×10^{-9} bp^{-1}, based on *A. thaliana* mutation accumulation lines [98], would shift the estimate for *C. rubella* further back in time. In *C. orientalis*, such analysis indicated that selfing evolved within the time frame of ca. 50,000–2,900,000 years ago [49].

Genome-wide data can provide alternative ways to estimate the timing of the transition to selfing. Depending on the underlying split history of the study system, different approaches can be utilized. For example, the haplotype-based founding population approach in *C. rubella* resulted in a shift to a selfing estimate of ca. 106,000–211,000 years ago [99]. This analysis also suggested that *C. rubella* was established by multiple individuals instead of a single individual as proposed by Guo et al. [96].

Another genome-wide method to roughly estimate the timing of the transition to selfing uses patterns of linkage disequilibrium (LD) decay [50]. Based on a simulation approach under simple demographic scenarios of constant size and population growth, it was shown that in selfing populations, LD is expected to increase in relatively recent transitions to selfing (40,000, 80,000, and 150,000 years ago) [50]. The increase of LD due to selfing has been empirically demonstrated, for example, in the North American SC and SI *A. lyrata* ssp *lyrata* populations. In this population complex, it was found that short-range LD was increased in

selfing populations, but demographic processes also had an effect on the LD decay [100]. In *A. thaliana,* the LD patterns were compared to coalescent simulations, and the observed LD decay resembled simulations where transition to selfing occurred either a very long time ago (more than a million years ago) or extremely recently [50]. The latter scenario was considered unlikely, and hence a very old transition was concluded.

More precise population split modeling can help to indirectly set boundaries to the timing of the transition to selfing, if there is sufficient population structure and if it is correlated with the mating system variation within species. Several demographic modelling methods are available that can be used in such an approach; many of which utilize the patterns of site frequency spectrum or structured coalescent in the demographic inference, such as dadi [101], fastsimcoal2 [102], momi [103], MSMC, MSMC2, SMC++[104]. As a note on these methods, both MSMC methods require phased genotype data (see Chapter 6), which can be difficult to obtain, particularly for non-model species.

Such approaches have been utilized in multiple instances. In *A. thaliana,* split modelling was able to provide information on the lower bound for selfing evolution assuming a single geographical origin [88]. By putting together multiple lines of evidence, supported by various population split modelling, it was suggested that the transition to selfing initiated approximately 800,000–1,200,000 years ago in multiple S-allele backgrounds [88].

In Siberian *A. lyrata*, a SC diploid lineage was identified in the Northeastern part of the distribution range [25]. The SC and the SI lineages have overlapping distributions, but all SC individuals cluster genetically [25], indicating a single origin of the selfing lineage with some level of reproductive isolation between the lineages. At the same time, gene flow was identified between the lineages, which was stronger from the SI to the SC lineage. The lineage split took place after the separation of the pan-Siberian lineage from Central European linages, which provided a framework where lineage split modeling was very efficient to study the evolution of SC. Using a simple two-population demographic model, it was estimated that this SC lineage was established relatively recently, 87,756 (36,002–

89,150) years ago, serving as an upper bound estimate for the emergence of self-compatibility and possibly also for selfing/mixed mating [25].

In the North American *A. lyrata* ssp. *lyrata* complex, where multiple – likely independent – transitions have happened [105], a collection of lineage split times can be used to bracket the timeframe of selfing evolution. For example, this population cluster contains the selfing species *A. arenicola*, which derives from a selfing sister lineage of the North American *A. lyrata* [39, 106]. It was estimated that *A. arenicola* separated from the *A. lyrata* ssp. *lyrata* lineage approximately 6,000 years ago (5,500–10,000) or 12,000 (11,000–20,000), assuming a generation time of one or two years, respectively [106]. This provides likely a lower-bound estimate for the evolution of SC. The upper bound for the timing of the transition to selfing in North American *A. lyrata* can be roughly inferred from the estimated split time between the European and North American lineages, which is approximately 200,000–320,000 years ago, assuming a generation time of two years [107]. Population genetic structure and the geographical distribution of selfing populations in combination with knowledge on past glaciation in this region, however, suggest much younger selfing origins, ca. less than 10,000 years ago [112], but precise model-based estimates are still missing.

In cases where recent population events hinder reconstruction of population history, improved demographic inference can be obtained by utilizing single-methylation polymorphism (SMPs, see Chapter 1) data in combination with single-nucleotide polymorphism approaches [108, 109]. For example, if the evolutionary transitions are very recent, SMPs can be utilized instead of SNP variants to give more resolution in demographic inference. Methylation variants are common in angiosperms [110] and they have much higher mutation rate in comparison to SNP variants, making them helpful in resolving very recent population processes.

The aforementioned timing methods are limited in their ability to accurately estimate the actual transition to selfing, since the estimation is mostly indirect. Another method can jointly estimate demographic history and transition to selfing based on the comparison of population mutation and recombination rates. Based on population genetic theory, it is

expected that there is a difference in the change in the ratio of population mutation rate ($4N_e\mu$) and population recombination rate ($4N_er$) in the presence of selfing relative to outcrossing [111]. This prediction was utilized to separate and jointly estimate the purely demographic and selfing evolution-related events from each other, implemented in two new tools, teSMC (see Chapter 7) and tsABC. The application of these methods to three different accessions of *A. thaliana* very consistently estimated the timing of the transition to selfing, the point estimates ranging from ca. 600,000 to 760,000 years ago.

It is worth noting that the divergence models often use a single constant mutation rate and recombination rate [112]. Commonly used mutation rates are 1.5×10^{-8} bp^{-1} based on lineage substitution rates and fossil record from Arabideae [97], and 7.1×10^{-9} [98] or 6.95×10^{-9} bp^{-1} [113] based on *A. thaliana* mutation accumulation lines. In turn, a genome-wide recombination rate estimate of 3.6 cM/Mb has been obtained for *A. thaliana* [114], while slightly lower estimates have been obtained, for example, for the diploid *A. arenosa* [115]. These parameters also vary across the genome [113], which can have an effect on the population genetic inference in general [116], and single-locus-based estimates in particular. Whether mutation rates can be affected by the environment and selection is also debated, which in turn questions the often-accepted assumption of a constant mutation rate assumption [117–119].

4.1 The connection between hybridization, polyploidization, and selfing

Many of the transitions to SC are associated with polyploidization [120]. This association is also expressed in Brassicaceae, where many selfing species are polyploids. Polyploidization may evolve via the combination of genomes of two species (allopolyploidy) or the multiplication of the genome of the species itself (autopolyploidy). Polyploidization causes changes in the functioning of the genome that have a tremendous effect on the evolution of the species, and, similar to selfing, may be detrimental in the long-term evolution [121]. In Brassicaceae, polyploids are numerous, and their origins cluster with temperature minima in the recent glaciation cycles [122]. In *Capsella* and *Arabidopsis*, all allopolyploids (*A.*

suecica (♀ *A. thaliana* X ♂ *A. arenosa*) [123–125], *A. kamchatica* (♀ *A. halleri* ssp. *gemmifera* X ♂ *A. lyrata* ssp. *petraea*) [126–128], and *C. bursa-pastoris* (♀ *C. orientalis* X ♂ *C. grandiflora*) [129]) are selfing, while the autotetraploid *A. lyrata* ssp. *petraea* is SI [25], as is the successful autotetraploid colonizer *A. arenosa* [130]. Similarly, in *Brassica*, the allopolyploids are SC while the parental species have functioning SI ([131] and references therein).

In allopolyploids, mate unavailability can favor the evolution of SC immediately when the polyploid hybrid has formed via inheritance of a dominant non-functional S-allele [131]. In line with this model, analysis of whole-genome resequencing data from world-wide *A. suecica* collections revealed that it inherited the loss-of-function SCR 213-bp inversion S-allele from *A. thaliana*. Based on dominance hierarchies of S-alleles in the *A. halleri* model, the non-functional *A. suecica* allele was predicted to be dominant over the S-allele inherited from the diploid *A. arenosa* [132, 133]. Hence, it is likely that the allopolyploid *A. suecica* was self-compatible instantly when the hybrid was formed. The timing of the split suggested that the transition happened relatively recently after the last glacial maximum [132]. The most likely geographic origin was in the Baltic region based on a niche modelling approach using environmental data from the current time and at the time of the species' origin [133].

Similarly, in *Capsella*, sequencing of bacterial artificial chromosome (BAC) of the entire S-locus indicated that the dominant non-functional S-allele was inherited from SC *C. orientalis* in *C. bursa-pastoris,* indicating that this species was also likely self-compatible at the time of the hybrid formation [134]. The origin of *C. bursa-pastoris* has been estimated to ca. 128,000 (22,000–177,000) years ago using sub-genome phased genome-wide data and population split modelling [129]. Further work on synthetic polyploids between the parental species, *C. orientalis* and *C. grandiflora*, has confirmed that the dominance level of the S-alleles in the cross explains the mating system of the allopolyploid hybrid progeny [135].

In *A. kamchatica,* it was previously thought that the progenitor species were both SI [126, 128]. However, it was recently shown that the Siberian SC *A. lyrata* lineage is genetically the closest to the *A. lyrata* sub-genome of *A. kamchatica* [25]. Hence, it is likely that the *A. lyrata*

lineage contributing to *A. kamchatica* was already SC, and the instant SC model could also apply to *A. kamchatica*.

5 A path to extinction or success? On the consequences of selfing evolution

The ecological and evolutionary properties can provide further insight into the processes determining which mating system is preferred in the short and long term. For example, self-fertilizing and outcrossing species are expected to differ in their ecological strategy [136], outcrossers being stronger competitors while selfers tend to be better colonizers, invasive ruderals [137]. The possibility to reproduce without mates and pollinators makes it easier for a selfer than an outcrosser to establish a new population. Geographic analyses have indeed shown that the distribution ranges of the selfing species are larger [106] and the distribution limits reach higher latitudes [138]. Selfers also tend to have more variable habitat preferences [139]. Furthermore, the enhanced colonizing ability of selfers, a phenomenon known as Baker's law [4, 140], has, for example, been empirically verified in the nematode *C. elegans* [141]. In turn, in a common garden experiment of different *Capsella* species, the outcrossing lineages outperformed all the selfers, and they were also much stronger competitors [142]. Interestingly, however, variation in competitive ability has also been observed within different selfing lineages, for example, between the relict and highly colonizing *A. thaliana* [22, 143, 144]. In addition, in the North American *A. lyrata* ssp. *lyrata* populations [105], no differences in survival were observed between selfing and outcrossing populations [145]. Still, SC populations are enriched in the range edge colonizing populations [146]. Also, one successful colonizing lineage from the SC population complex resulted in the evolution of the selfing species *A. arenicola*, which is currently distributed in the much more northern latitudes than the other *A. lyrata* populations [106]. This also demonstrates speciation ability and the capability to escape species range limitations of selfing species.

In contrast to selfers being possibly successful colonizers, selfing species can have genetic properties that make them less fit in the long term. The reduced efficacy of selection can, for example, increase the accumulation of deleterious mutations (i.e., mutational load) in the selfing lineages. The accumulation of genetic load was, for example, contrasted in *Arabis alpina* populations with different levels of selfing [77]. Breaking it down further, the Scandinavian highly selfing populations have very low genetic diversity caused by the combined effects of a strong population bottleneck and selfing. Furthermore, these lineages have increased genetic load, while this effect was not seen in the central European mixed mating population [147]. However, it is expected that in selfing or highly inbred lineages, the effect of recessive and partially deleterious mutations becomes more visible due to increased homozygosity, and hence they can be effectively eliminated due to purging [148]. Nevertheless, simulations have shown that in colonizing selfing lineages, purging only removes large-effect deleterious alleles, whereas small-effect variants can accumulate, in line with the load patterns in the selfing *A. alpina* [149]. Surprisingly, regardless of the accumulated load, the Scandinavian *A. alpina* populations still show strong signatures of local adaptation [150], indicating that the fitness benefit obtained via adaptation outweighed the fitness loss due to accumulation of deleterious load.

Colonizing populations can have additional effects on genetic variation in combination with the consequences of selfing (if it evolved), which may be difficult to disentangle from each other. For example, they have experienced population bottlenecks, and the allele-surfing effect is stronger in such populations [151, 152], which can lead to the accumulation of deleterious variants. Empirical investigations of genome-wide analyses in SI *A. lyrata* populations with different demographic histories did not detect any difference in load patterns (e.g., the bottlenecked *A. lyrata* ssp. *petraea* population from Scandinavia compared to the more stable Central European population [153]). In contrast, a difference in the magnitude of linked selection and distribution of fitness effects was observed between diploid outcrossing but demographically differing *A. lyrata* and *C. grandiflora* [14], emphasizing the importance of careful investigation of factors affecting patterns of variation.

A change in the efficacy of selection landscape, as well as heterozygosity in selfers, is also predicted to affect the accumulation of transposable elements [154], which can in turn have an effect on the performance of selfers in the long run. Early contrasts between selfing *Arabidopsis thaliana* and multiple outcrossing *Arabidopsis lyrata* populations suggested that most of the variation in TE abundance could be assigned to neutral demographic processes [155, 156]. Using a whole-genome dataset from the two diploid selfing and one outcrossing *Capsella* species [157], it was found that TE abundance was rather decreased in the selfing *C. orientalis,* while in the selfing *C. rubella,* the level was only slightly higher than in the SI *C. grandiflora*. Other factors can, however, have a role in shaping the TE abundance that can further be associated with selfing. For example, it has recently been found that TE activity can be affected by environmental variation [158]. Furthermore, in the highly inbreeding North American *A. lyrata* populations, it was found that downstream of TEs, heterozygosity was systematically increased, suggesting balancing selection driving TE-mediated heterozygosity in the independently evolved selfing populations [159]. These findings add other layers of complexity in TE functionality and challenge the mostly deleterious model of TE evolution and their effects in selfing populations.

6 Acknowledgements

This work was supported by a grant from the Research Council of Finland to TMM (decision number 360310) and by grants from the Swedish Research Council and the European Research Council to TS, relying on resources provided by the National Academic Infrastructure for Supercomputing in Sweden (NAISS), partially funded by the Swedish Research Council through grant agreement no. 2022-06725. Uliana Kolesnikova is acknowledged for sharing knowledge regarding S-locus genotyping using NGSgenotyp.

7 References

1. Igic B, Lande R, Kohn J, et al (2008) Loss of Self-Incompatibility and Its Evolutionary Consequences. Int J Plant Sci 169:93–104. https://doi.org/10.1086/523362

2. Haghighatnia M, Machac A, Schmickl R, Lafon Placette C (2023) Darwin's 'mystery of mysteries': the role of sexual selection in plant speciation. Biol Rev 98:1928–1944.

https://doi.org/https://doi.org/10.1111/brv.12991

3. Goldberg EE, Kohn JR, Lande R, et al (2010) Species Selection Maintains Self-Incompatibility. Science 330:493–495. https://doi.org/10.1126/science.1194513

4. Baker HG (1955) Self-compatibility and establishment after “long-distance” dispersal. Evolution 9:349. https://doi.org/https://doi.org/10.1111/j.1558-5646.1955.tb01544.x

5. Schoen DJ, Morgan MT, Bataillon T (1996) How Does Self-Pollination Evolve? Inferences from Floral Ecology and Molecular Genetic Variation. Philos Trans Biol Sci 351:1281–1290

6. Mattila TM, Laenen B, Slotte T (2020) Population Genomics of Transitions to Selfing in Brassicaceae Model Systems BT - Statistical Population Genomics. In: Dutheil JY (ed). Springer US, New York, NY, pp 269–287

7. Charlesworth D, Wright SI (2001) Breeding systems and genome evolution. Curr Opin Genet Dev 11:685–690. https://doi.org/10.1016/S0959-437X(00)00254-9

8. Hartfield M, Bataillon T, Glémin S (2017) The Evolutionary Interplay between Adaptation and Self-Fertilization. Trends Genet 33:420–431

9. Burgarella C, Glémin S (2017) Population genetics and genome evolution of selfing species. Els 1–8. https://doi.org/https://doi.org/10.1002/9780470015902.a0026804

10. Slotte T (2014) The impact of linked selection on plant genomic variation. Briefings Funct Genomics Proteomics 13:268–275. https://doi.org/10.1093/bfgp/elu009

11. Burgarella C, Brémaud M-F, Von Hirschheydt G, et al (2024) Mating systems and recombination landscape strongly shape genetic diversity and selection in wheat relatives. Evol Lett 8:866–880. https://doi.org/10.1093/evlett/qrae039

12. Marie-Orleach L, Brochmann C, Glémin S (2022) Mating system and speciation I: Accumulation of genetic incompatibilities in allopatry. PLOS Genet 18:1–27. https://doi.org/10.1371/journal.pgen.1010353

13. Marie-Orleach L, Glémin S, Brandrud MK, et al (2024) How Does Selfing Affect the Pace and Process of Speciation? Cold Spring Harb Perspect Biol 16:a041426

14. Mattila TM, Laenen B, Horvath R, et al (2019) Impact of demography on linked selection in two outcrossing Brassicaceae species. Ecol Evol 9:9532–9545. https://doi.org/10.1002/ece3.5463

15. Burns R, Mandáková T, Gunis J, et al (2021) Gradual evolution of allopolyploidy in Arabidopsis suecica. Nat Ecol Evol 5:1367–1381. https://doi.org/10.1038/s41559-021-01525-w

16. Sianta SA, Peischl S, Moeller DA, Brandvain Y (2023) The efficacy of selection may increase or decrease with selfing depending upon the recombination environment. Evolution 77:394–408. https://doi.org/10.1093/evolut/qpac013

17. Xiao Y, Lv Y-W, Wang Z-Y, et al (2024) Selfing Shapes Fixation of a Mutant Allele Under Flux Equilibrium. Genome Biol Evol 16:evae261. https://doi.org/10.1093/gbe/evae261

18. Scott MF, Mackintosh C, Immler S (2024) Gametic selection favours polyandry and selfing. PLOS Genet 20:e1010660. https://doi.org/https://doi.org/10.1371/journal.pgen.1010660

19. Gutiérrez-Valencia J, Fracassetti M, Horvath R, et al (2022) Genomic Signatures of Sexual Selection on Pollen-Expressed Genes in Arabis alpina. Mol Biol Evol 39:msab349. https://doi.org/10.1093/molbev/msab349

20. Ferrari T, Feng S, Zhang X, Mooney J (2025) Parameter Scaling in Population Genetics Simulations may Introduce Unintended Background Selection: Considerations for Scaled Simulation Design. Genome Biol Evol 17:evaf097. https://doi.org/10.1093/gbe/evaf097

21. Daigle A, Johri P (2025) Hill-Robertson interference may bias the inference of fitness effects of new mutations in highly selfing species. Evolution 79:342–363.

https://doi.org/10.1093/evolut/qpae168

22. 1001 Genomes C (2016) 1,135 genomes reveal the global pattern of polymorphism in Arabidopsis thaliana. Cell 166:481–491. https://doi.org/https://doi.org/10.1016/j.cell.2016.05.063

23. Briskine R V, Paape T, Shimizu-Inatsugi R, et al (2017) Genome assembly and annotation of Arabidopsis halleri, a model for heavy metal hyperaccumulation and evolutionary ecology. Mol Ecol Resour 17:1025–1036. https://doi.org/https://doi.org/10.1111/1755-0998.12604

24. Hu TT, Pattyn P, Bakker EG, et al (2011) The Arabidopsis lyrata genome sequence and the basis of rapid genome size change. Nat Genet 43:476–481. https://doi.org/10.1038/ng.807

25. Kolesnikova UK, Scott AD, Van de Velde JD, et al (2023) Transition to Self-compatibility Associated With Dominant S-allele in a Diploid Siberian Progenitor of Allotetraploid Arabidopsis kamchatica Revealed by Arabidopsis lyrata Genomes. Mol Biol Evol 40:. https://doi.org/10.1093/molbev/msad122

26. Barragan AC, Collenberg M, Schwab R, et al (2024) Deleterious phenotypes in wild Arabidopsis arenosa populations are common and linked to runs of homozygosity. G3 Genes|Genomes|Genetics 14:jkad290. https://doi.org/10.1093/g3journal/jkad290

27. Slotte T, Hazzouri KM, gren JA, et al (2013) The Capsella rubella genome and the genomic consequences of rapid mating system evolution. Nat Genet 45:831–835. https://doi.org/10.1038/ng.2669

28. Kasianova AM, Mityukov VD, German DA, et al (2025) Chromosome-Scale Assembly of Capsella orientalis, Maternal Progenitor of Cosmopolitan Allotetraploid C. bursa-pastoris. Genome Biol Evol 17:evaf009. https://doi.org/10.1093/gbe/evaf009

29. Penin AA, Kasianov AS, Klepikova A V, et al (2024) Origin and diversity of Capsella

665 bursa-pastoris from the genomic point of view. BMC Biol 22:52.
666 https://doi.org/10.1186/s12915-024-01832-1

667 30. Willing E-M, Rawat V, Mandáková T, et al (2015) Genome expansion of Arabis alpina
668 linked with retrotransposition and reduced symmetric DNA methylation. Nat Plants
669 1:14023. https://doi.org/10.1038/nplants.2014.23

670 31. Jiao W-B, Accinelli GG, Hartwig B, et al (2017) Improving and correcting the
671 contiguity of long-read genome assemblies of three plant species using optical
672 mapping and chromosome conformation capture data. Genome Res 27:778–786.
673 https://doi.org/10.1101/gr.213652.116

674 32. Gan X, Hay A, Kwantes M, et al (2016) The Cardamine hirsuta genome offers insight
675 into the evolution of morphological diversity. Nat Plants 2:16167.
676 https://doi.org/10.1038/nplants.2016.167

677 33. Haudry A, Platts AE, Vello E, et al (2013) An atlas of over 90,000 conserved
678 noncoding sequences provides insight into crucifer regulatory regions. Nat Genet
679 45:891–898. https://doi.org/10.1038/ng.2684

680 34. Vekemans X, Poux C, Goubet PM, Castric V (2014) The evolution of selfing from
681 outcrossing ancestors in Brassicaceae: what have we learned from variation at the
682 S-locus? J Evol Biol 27:1372–1385. https://doi.org/10.1111/jeb.12372

683 35. Dassanayake M, Oh D-H, Haas JS, et al (2011) The genome of the extremophile
684 crucifer Thellungiella parvula. Nat Genet 43:913–918.
685 https://doi.org/10.1038/ng.889

686 36. Yang R, Jarvis DE, Chen H, et al (2013) The reference genome of the halophytic plant
687 Eutrema salsugineum. Front Plant Sci 4:46.
688 https://doi.org/https://doi.org/10.3389/fpls.2013.00046

689 37. Guo X, Liu J, Hao G, et al (2017) Plastome phylogeny and early diversification of
690 Brassicaceae. BMC Genomics 18:176.

https://doi.org/https://doi.org/10.1186/s12864-017-3555-3

38. Hendriks KP, Kiefer C, Al-Shehbaz IA, et al (2023) Global Brassicaceae phylogeny based on filtering of 1,000-gene dataset. Curr Biol 33:4052-4068.e6. https://doi.org/https://doi.org/10.1016/j.cub.2023.08.026

39. Novikova PY, Hohmann N, Nizhynska V, et al (2016) Sequencing of the genus Arabidopsis identifies a complex history of nonbifurcating speciation and abundant trans-specific polymorphism. Nat Genet 48:1077–1082. https://doi.org/10.1038/ng.3617

40. Hurka H, Friesen N, German DA, et al (2012) "Missing link" species Capsella orientalis and Capsella thracica elucidate evolution of model plant genus Capsella (Brassicaceae). Mol Ecol 21:1223–1238. https://doi.org/10.1111/j.1365-294X.2012.05460.x

41. Beck JB, Al-Shehbaz IA, Schaal BA (2006) Leavenworthia (Brassicaceae) revisited: testing classic systematic and mating system hypotheses. Syst Bot 31:151–159

42. Berardini TZ, Reiser L, Li D, et al (2015) The arabidopsis information resource: Making and mining the "gold standard" annotated reference plant genome. genesis 53:474–485. https://doi.org/https://doi.org/10.1002/dvg.22877

43. Ivanov R, Fobis-Loisy I, Gaude T (2010) When no means no: guide to Brassicaceae self-incompatibility. Trends Plant Sci 15:387–394. https://doi.org/10.1016/j.tplants.2010.04.010

44. Nasrallah JB (2023) Stop and go signals at the stigma–pollen interface of the Brassicaceae. Plant Physiol 193:927–948. https://doi.org/10.1093/plphys/kiad301

45. Kusaba M, Dwyer K, Hendershot J, et al (2001) Self-incompatibility in the genus Arabidopsis: characterization of the S locus in the outcrossing A. lyrata and its autogamous relative A. thaliana. Plant Cell 13:627–643. https://doi.org/https://doi.org/10.1105/tpc.13.3.627

46. Charlesworth D, Bartolome C, Schierup MH, Mable BK (2003) Haplotype Structure of the Stigmatic Self-Incompatibility Gene in Natural Populations of Arabidopsis lyrata. Mol Biol Evol 20:1741–1753. https://doi.org/10.1093/molbev/msg170

47. Hiscock SJ, McInnis SM (2003) Pollen recognition and rejection during the sporophytic self-incompatibility response: Brassica and beyond. Trends Plant Sci 8:606–613. https://doi.org/https://doi.org/10.1016/j.tplants.2003.10.007

48. Goubet PM, Bergès H, Bellec A, et al (2012) Contrasted Patterns of Molecular Evolution in Dominant and Recessive Self-Incompatibility Haplotypes in Arabidopsis. PLOS Genet 8:e1002495. https://doi.org/https://doi.org/10.1371/journal.pgen.1002495

49. Bachmann JA, Tedder A, Laenen B, et al (2019) Genetic basis and timing of a major mating system shift in Capsella. New Phytol 224:505–517. https://doi.org/https://doi.org/10.1111/nph.16035

50. Tang C, Toomajian C, Sherman-Broyles S, et al (2007) The Evolution of Selfing in Arabidopsis thaliana. Science 317:1070–1072. https://doi.org/10.1126/science.1143153

51. Beilstein MA, Nagalingum NS, Clements MD, et al (2010) Dated molecular phylogenies indicate a Miocene origin for Arabidopsis thaliana. Proc Natl Acad Sci U S A 107:18724–18728. https://doi.org/10.1073/pnas.0909766107

52. Uyenoyama MK (1995) A generalized least-squares estimate for the origin of sporophytic self-incompatibility. Genetics 139:975–992. https://doi.org/10.1093/genetics/139.2.975

53. Chantha S-C, Herman AC, Platts AE, et al (2013) Secondary Evolution of a Self-Incompatibility Locus in the Brassicaceae Genus Leavenworthia. PLOS Biol 11:1–16. https://doi.org/10.1371/journal.pbio.1001560

54. Chantha S-C, Herman AC, Castric V, et al (2017) The unusual S locus of

743 Leavenworthia is composed of two sets of paralogous loci. New Phytol 216:1247–
744 1255. https://doi.org/https://doi.org/10.1111/nph.14764

745 55. Busch JW, Sharma J, Schoen DJ (2008) Molecular Characterization of Lal2, an SRK-
746 Like Gene Linked to the S-Locus in the Wild Mustard Leavenworthia alabamica.
747 Genetics 178:2055–2067. https://doi.org/10.1534/genetics.107.083204

748 56. Llaurens V, Billiard S, Leducq J-B, et al (2008) Does frequency-dependent selection
749 with complex dominance interactions accurately predict allelic fequencies at the
750 self-incompatibility locus in Arabidopsis halleri? Evolution 62:2545–2557.
751 https://doi.org/10.1111/j.1558-5646.2008.00469.x

752 57. Uyenoyama MK, Zhang Y, Newbigin E (2001) On the Origin of Self-Incompatibility
753 Haplotypes: Transition Through Self-Compatible Intermediates. Genetics 157:1805–
754 1817. https://doi.org/10.1093/genetics/157.4.1805

755 58. Castric V, Bechsgaard J, Schierup MH, Vekemans X (2008) Repeated Adaptive
756 Introgression at a Gene under Multiallelic Balancing Selection. PLoS Genet
757 4:e1000168

758 59. Edh K, Widén B, Ceplitis A (2009) The Evolution and Diversification of S-Locus
759 Haplotypes in the Brassicaceae Family. Genetics 181:977–984.
760 https://doi.org/10.1534/genetics.108.090837

761 60. Thompson KF (1957) Self-incompatibility in marrow-stem kale, Brassica oleracea
762 var. acephala I. Demonstration of a sporophytic system. J Genet 55:45–60.
763 https://doi.org/https://doi.org/10.1007/BF02981615

764 61. Yasuda S, Kobayashi R, Ito T, et al (2021) Homology-Based Interactions between
765 Small RNAs and Their Targets Control Dominance Hierarchy of Male Determinant
766 Alleles of Self-Incompatibility in Arabidopsis lyrata. Int J Mol Sci 22:.
767 https://doi.org/10.3390/ijms22136990

768 62. Durand E, Méheust R, Soucaze M, et al (2014) Dominance hierarchy arising from the

evolution of a complex small RNA regulatory network. Science 346:1200–1205. https://doi.org/10.1126/science.1259442

63. Tarutani Y, Shiba H, Iwano M, et al (2010) Trans-acting small RNA determines dominance relationships in Brassica self-incompatibility. Nature 466:983–986. https://doi.org/10.1038/nature09308

64. Prigoda NL, Nassuth A, Mable BK (2005) Phenotypic and Genotypic Expression of Self-incompatibility Haplotypes in Arabidopsis lyrata Suggests Unique Origin of Alleles in Different Dominance Classes. Mol Biol Evol. https://doi.org/10.1093/molbev/msi153

65. Schierup MH, Vekemans X, Christiansen FB (1997) Evolutionary Dynamics of Sporophytic Self-Incompatibility Alleles in Plants. Genetics 147:835–846. https://doi.org/10.1093/genetics/147.2.835

66. Vekemans X, Castric V, Hipperson H, et al (2021) Whole-genome sequencing and genome regions of special interest: Lessons from major histocompatibility complex, sex determination, and plant self-incompatibility. Mol Ecol 30:6072–6086. https://doi.org/10.1111/mec.16020

67. Guo Y-L, Zhao X, Lanz C, Weigel D (2011) Evolution of the S-Locus Region in Arabidopsis Relatives. Plant Physiol 157:937–946. https://doi.org/10.1104/pp.111.174912

68. Lucek K, Flury JM, Willi Y (2025) Genomic implications of the repeated shift to self-fertilization across a species' geographic distribution. J Hered 116:43–53. https://doi.org/10.1093/jhered/esae046

69. Genete M, Castric V, Vekemans X (2020) Genotyping and De Novo Discovery of Allelic Variants at the Brassicaceae Self-Incompatibility Locus from Short-Read Sequencing Data. Mol Biol Evol 37:1193–1201. https://doi.org/10.1093/molbev/msz258

70. Le Veve A, Genete M, Lepers-Blassiau C, et al (2024) The genetic architecture of the load linked to dominant and recessive self-incompatibility alleles in Arabidopsis halleri and Arabidopsis lyrata. Elife 13:RP94972. https://doi.org/10.7554/eLife.94972

71. Le Veve A, Burghgraeve N, Genete M, et al (2023) Long-term balancing selection and the genetic load linked to the self-incompatibility locus in Arabidopsis halleri and A. lyrata. Mol Biol Evol 40:msad120. https://doi.org/https://doi.org/10.1093/molbev/msad120

72. Llaurens V, Gonthier L, Billiard S (2009) The Sheltered Genetic Load Linked to the S Locus in Plants: New Insights From Theoretical and Empirical Approaches in Sporophytic Self-Incompatibility. Genetics 183:1105–1118. https://doi.org/10.1534/genetics.109.102707

73. David P, Pujol B, Viard F, et al (2007) Reliable selfing rate estimates from imperfect population genetic data. Mol Ecol 16:2474–2487. https://doi.org/https://doi.org/10.1111/j.1365-294X.2007.03330.x

74. Wright S (1951) The genetical structure of populations. Ann Eugen 15:323–354. https://doi.org/https://doi.org/10.1111/j.1469-1809.1949.tb02451.x

75. Hardy GH (1908) Mendelian proportions in a mixed population. Science 28:49–50. https://doi.org/10.1126/science.28.706.49

76. Ansell SW, Grundmann M, Russell SJ, et al (2008) Genetic discontinuity, breeding-system change and population history of Arabis alpina in the Italian Peninsula and adjacent Alps. Mol Ecol 17:2245–2257. https://doi.org/10.1111/j.1365-294X.2008.03739.x

77. Tedder A, Ansell SW, Lao X, et al (2011) Sporophytic self-incompatibility genes and mating system variation in Arabis alpina. Ann Bot 108:699–713. https://doi.org/10.1093/aob/mcr157

78. Toräng P, Vikström L, Wunder J, et al (2017) Evolution of the selfing syndrome: Anther

orientation and herkogamy together determine reproductive assurance in a self-compatible plant. Evolution 71:2206–2218. https://doi.org/10.1111/evo.13308

79. Fenster CB, Vekemans X, Hardy OJ (2003) Quantifying gene flow from spatial genetic structure data in a metapopulation of Chamaecrista fasciculata (Leguminosae). Evolution 57:995–1007. https://doi.org/10.1111/j.0014-3820.2003.tb00311.x

80. Ceballos FC, Joshi PK, Clark DW, et al (2018) Runs of homozygosity: Windows into population history and trait architecture. Nat Rev Genet 19:220–234. https://doi.org/10.1038/nrg.2017.109

81. Zeitler L, Gilbert KJ (2024) Using Runs of Homozygosity and Machine Learning to Disentangle Sources of Inbreeding and Infer Self-Fertilization Rates. Genome Biol Evol 16:evae139. https://doi.org/10.1093/gbe/evae139

82. Abbott RJ, Gomes MF (1989) Population genetic structure and outcrossing rate of Arabidopsis thaliana (L.) Heynh. Heredity 62:411–418. https://doi.org/https://doi.org/10.1038/hdy.1989.56

83. Sellinger TPP, Abu Awad D, Moest M, Tellier A (2020) Inference of past demography, dormancy and self-fertilization rates from whole genome sequence data. PLOS Genet 16:e1008698. https://doi.org/https://doi.org/10.1371/journal.pgen.1008698

84. Li H, Durbin R (2011) Inference of human population history from individual whole-genome sequences. Nature 475:493–496. https://doi.org/10.1038/nature10231

85. Schiffels S, Durbin R (2014) Inferring human population size and separation history from multiple genome sequences. Nat Publ Gr 46:. https://doi.org/10.1038/ng.3015

86. Malaspinas A-S, Westaway MC, Muller C, et al (2016) A genomic history of Aboriginal Australia. Nature 538:207–214. https://doi.org/http://dx.doi.org/10.1038/nature18299

87. Charlesworth D, Vekemans X (2005) How and when did Arabidopsis thaliana become highly self-fertilising. Bioessays 27:472–476.

https://doi.org/https://doi.org/10.1002/bies.20231

88. Durvasula A, Fulgione A, Gutaker RM, et al (2017) African genomes illuminate the early history and transition to selfing in Arabidopsis thaliana. Proc Natl Acad Sci 114:5213–5218. https://doi.org/https://doi.org/10.1073/pnas.1616736114

89. Nasrallah ME, Liu P, Nasrallah JB (2002) Generation of self-incompatible Arabidopsis thaliana by transfer of two S locus genes from A. lyrata. Science 297:247–249. https://doi.org/10.1126/science.1072205

90. Tsuchimatsu T, Suwabe K, Shimizu-Inatsugi R, et al (2010) Evolution of self-compatibility in Arabidopsis by a mutation in the male specificity gene. Nature 464:1342. https://doi.org/https://doi.org/10.1038/nature08927

91. Hoebe PN, Stift M, Tedder A, Mable BK (2009) Multiple losses of self-incompatibility in North-American Arabidopsis lyrata?: Phylogeographic context and population genetic consequences. Mol Ecol 18:4924–4939. https://doi.org/https://doi.org/10.1111/j.1365-294X.2009.04400.x

92. Li Y, Mamonova E, Köhler N, et al (2023) Breakdown of self-incompatibility due to genetic interaction between a specific S-allele and an unlinked modifier. Nat Commun 14:3420. https://doi.org/10.1038/s41467-023-38802-0

93. Mable BK, Hagmann J, Kim ST, et al (2017) What causes mating system shifts in plants? Arabidopsis lyrata as a case study. Heredity 118:52–63. https://doi.org/https://doi.org/10.1038/hdy.2016.99

94. Porcher E, Lande R (2005) Loss of gametophytic self-incompatibility with evolution of inbreeding depression. Evolution 59:46–60. https://doi.org/10.1111/j.0014-3820.2005.tb00893.x

95. Bechsgaard JS, Castric V, Charlesworth D, et al (2006) The transition to self-compatibility in Arabidopsis thaliana and evolution within S-haplotypes over 10 Myr. Mol Biol Evol 23:1741–1750. https://doi.org/https://doi.org/10.1093/molbev/msl042

96. Guo Y-L, Bechsgaard JS, Slotte T, et al (2009) Recent speciation of Capsella rubella from Capsella grandiflora, associated with loss of self-incompatibility and an extreme bottleneck. Proc Natl Acad Sci U S A 106:5246–5251. https://doi.org/10.1073/pnas.0808012106

97. Koch MA, Haubold B, Mitchell-Olds T (2000) Comparative evolutionary analysis of chalcone synthase and alcohol dehydrogenase loci in Arabidopsis, Arabis, and related genera (Brassicaceae). Mol Biol Evol 17:1483–1498. https://doi.org/https://doi.org/10.1093/oxfordjournals.molbev.a026248

98. Ossowski S, Schneeberger K, Lucas-Lledó JI, et al (2010) The rate and molecular spectrum of spontaneous mutations in arabidopsis thaliana. Science 327:92–94. https://doi.org/10.1126/science.1180677

99. Brandvain Y, Slotte T, Hazzouri KM, et al (2013) Genomic identification of founding haplotypes reveals the history of the selfing species Capsella rubella. PLoS Genet 9:e1003754. https://doi.org/https://doi.org/10.1371/journal.pgen.1003754

100. Lucek K, Willi Y (2021) Drivers of linkage disequilibrium across a species' geographic range. PLOS Genet 17:e1009477. https://doi.org/https://doi.org/10.1371/journal.pgen.1009477

101. Gutenkunst RN, Hernandez RD, Williamson SH, Bustamante CD (2009) Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. PLoS Genet 5:. https://doi.org/10.1371/journal.pgen.1000695

102. Excoffier L, Dupanloup I, Huerta-Sánchez E, et al (2013) Robust Demographic Inference from Genomic and SNP Data. PLoS Genet 9:e1003905. https://doi.org/https://doi.org/10.1371/journal.pgen.1003905

103. Kamm J, Terhorst J, Durbin R, Song YS (2020) Efficiently Inferring the Demographic History of Many Populations With Allele Count Data. J Am Stat Assoc 115:1472–1487. https://doi.org/10.1080/01621459.2019.1635482

104. Terhorst J, Kamm JA, Song YS (2017) Robust and scalable inference of population history from hundreds of unphased whole genomes. Nat Genet 49:. https://doi.org/10.1038/ng.3748

105. Foxe JP, Stift M, Tedder A, et al (2010) Reconstructing origins of loss of self-incompatibility and selfing in North American Arabidopsis lyrata: a population genetic context. Evolution 64:3495–3510. https://doi.org/https://doi.org/10.1111/j.1558-5646.2010.01094.x

106. Willi Y, Lucek K, Bachmann O, Walden N (2022) Recent speciation associated with range expansion and a shift to self-fertilization in North American Arabidopsis. Nat Commun 13:7564. https://doi.org/10.1038/s41467-022-35368-1

107. Pyhäjärvi T, Aalto E, Savolainen O (2012) Time Scales of divergence and speciation among natural populations and subspecies of Arabidopsis lyrata (Brassicaceae). Am J Bot. https://doi.org/10.3732/ajb.1100580

108. Sellinger T, Johannes F, Tellier A (2024) Improved inference of population histories by integrating genomic and epigenomic data. Elife 12:RP89470. https://doi.org/10.7554/eLife.89470

109. Korfmann K, Zauchner A, Huo B, et al (2025) Methylomes reveal recent evolutionary changes in populations of two plant species. bioRxiv. https://doi.org/10.1101/2024.09.30.615871

110. Niederhuth CE, Bewick AJ, Ji L, et al (2016) Widespread natural variation of DNA methylation within angiosperms. Genome Biol 17:194. https://doi.org/10.1186/s13059-016-1059-0

111. Strütt S, Sellinger T, Glémin S, et al (2023) Joint inference of evolutionary transitions to self-fertilization and demographic history using whole-genome sequences. Elife 12:e82384. https://doi.org/10.7554/eLife.82384

112. Shimizu KK, Tsuchimatsu T (2015) Evolution of selfing: recurrent patterns in

molecular adaptation. Annu Rev Ecol Evol Syst 46:593–622. https://doi.org/https://doi.org/10.1146/annurev-ecolsys-112414-054249

113. Weng M-L, Becker C, Hildebrandt J, et al (2019) Fine-Grained Analysis of Spontaneous Mutation Spectrum and Frequency in Arabidopsis thaliana. Genetics 211:703–714. https://doi.org/10.1534/genetics.118.301721

114. Salomé PA, Bomblies K, Fitz J, et al (2012) The recombination landscape in Arabidopsis thaliana F2 populations. Heredity 108:447–455. https://doi.org/10.1038/hdy.2011.95

115. Dukić M, Bomblies K (2022) Male and female recombination landscapes of diploid Arabidopsis arenosa. Genetics 220:iyab236. https://doi.org/10.1093/genetics/iyab236

116. Soni V, Pfeifer SP, Jensen JD (2024) The Effects of Mutation and Recombination Rate Heterogeneity on the Inference of Demography and the Distribution of Fitness Effects. Genome Biol Evol 16:evae004. https://doi.org/10.1093/gbe/evae004

117. Monroe JG, Srikant T, Carbonell-Bejerano P, et al (2022) Mutation bias reflects natural selection in Arabidopsis thaliana. Nature 602:101–105. https://doi.org/10.1038/s41586-021-04269-6

118. Belfield EJ, Brown C, Ding ZJ, et al (2021) Thermal stress accelerates Arabidopsis thaliana mutation rate. Genome Res 31:40–50. https://doi.org/10.1101/gr.259853.119

119. Lu Z, Cui J, Wang L, et al (2021) Genome-wide DNA mutations in Arabidopsis plants after multigenerational exposure to high temperatures. Genome Biol 22:160. https://doi.org/10.1186/s13059-021-02381-4

120. Barringer BC (2007) Polyploidy and self-fertilization in flowering plants. Am J Bot 94:1527–1533. https://doi.org/https://doi.org/10.3732/ajb.94.9.1527

121. Van de Peer Y, Mizrachi E, Marchal K (2017) The evolutionary significance of

polyploidy. Nat Rev Genet 18:411–424. https://doi.org/10.1038/nrg.2017.26

122. . Novikova PY, Hohmann N, Van de Peer Y (2018) Polyploid Arabidopsis species originated around recent glaciation maxima. Curr Opin Plant Biol 42:8–15. https://doi.org/https://doi.org/10.1016/j.pbi.2018.01.005

123. Jakobsson M, Säll T, Lind-halldén C, Halldén C (2007) The evolutionary history of the common chloroplast genome of Arabidopsis thaliana and A. suecica. J Evol Biol 20:104–121. https://doi.org/10.1111/j.1420-9101.2006.01217.x

124. Mummenhoff K, Hurka H (1995) Allopolyploid Origin of Arabidopsis suecica (Fries) Norrlin: Evidence from Chloroplast and Nuclear Genome Markers. Bot Acta 108:449–456. https://doi.org/https://doi.org/10.1111/j.1438-8677.1995.tb00520.x

125. Säll T, Lind-halldén C, Jakobsson M, Halldén C (2004) Mode of reproduction in Arabidopsis suecica. Hereditas 141:313–317. https://doi.org/10.1111/j.1601-5223.2004.01833.x

126. Paape T, Briskine R V, Halstead-Nussloch G, et al (2018) Patterns of polymorphism and selection in the subgenomes of the allopolyploid Arabidopsis kamchatica. Nat Commun 9:3909. https://doi.org/10.1038/s41467-018-06108-1

127. Shimizu-Inatsugi R, LihovÁ J, Iwanaga H, et al (2009) The allopolyploid Arabidopsis kamchatica originated from multiple individuals of Arabidopsis lyrata and Arabidopsis halleri. Mol Ecol 18:4024–4048. https://doi.org/10.1111/j.1365-294X.2009.04329.x

128. Tsuchimatsu T, Kaiser P, Yew C-L, et al (2012) Recent loss of self-incompatibility by degradation of the male component in allotetraploid Arabidopsis kamchatica. PLoS Genet 8:e1002838. https://doi.org/https://doi.org/10.1371/journal.pgen.1002838

129. Douglas GM, Gos G, Steige KA, et al (2015) Hybrid origins and the earliest stages of diploidization in the highly successful recent polyploid Capsella bursa-pastoris. Proc Natl Acad Sci U S A 112:2806–2811. https://doi.org/10.1073/pnas.1412277112

130. Arnold B, Kim S-T, Bomblies K (2015) Single geographic origin of a widespread autotetraploid arabidopsis arenosa lineage followed by interploidy admixture. Mol Biol Evol 32:1382–1395. https://doi.org/10.1093/molbev/msv089

131. Novikova PY, Kolesnikova UK, Scott AD (2023) Ancestral self-compatibility facilitates the establishment of allopolyploids in Brassicaceae. Plant Reprod 36:125–138. https://doi.org/10.1007/s00497-022-00451-6

132. Novikova PY, Tsuchimatsu T, Simon S, et al (2017) Genome sequencing reveals the origin of the allotetraploid Arabidopsis suecica. Mol Biol Evol 34:957–968. https://doi.org/https://doi.org/10.1093/molbev/msw299

133. Burns R, Kulkarni A, Glushkevich A, et al (2024) Diploid origins, adaptation to polyploidy, and the beginning of rediploidization in allotetraploid Arabidopsis suecica. bioRxiv. https://doi.org/10.1101/2024.12.06.627142

134. Bachmann JA, Tedder A, Fracassetti M, et al (2021) On the origin of the widespread self-compatible allotetraploid Capsella bursa-pastoris (Brassicaceae). Heredity 127:124–134. https://doi.org/10.1038/s41437-021-00434-9

135. Duan T, Zhang Z, Genete M, et al (2024) Dominance between self-incompatibility alleles determines the mating system of Capsella allopolyploids. Evol Lett 8:550–560. https://doi.org/10.1093/evlett/qrae011

136. Munoz F, Violle C, Cheptou PO (2016) CSR ecological strategies and plant mating systems: outcrossing increases with competitiveness but stress-tolerance is related to mixed mating. Oikos 125:. https://doi.org/10.1111/oik.02328

137. Orsucci M, Milesi P, Hansen J, et al (2020) Shift in ecological strategy helps marginal populations of shepherd's purse (Capsella bursa-pastoris) to overcome a high genetic load. Proc R Soc B Biol Sci 287:. https://doi.org/10.1098/rspb.2020.0463

138. Grossenbacher D, Briscoe Runquist R, Goldberg EE, Brandvain Y (2015) Geographic range size is predicted by plant mating system. Ecol Lett 18:706–713.

https://doi.org/https://doi.org/10.1111/ele.12449

139. Grant A-G, Kalisz S (2020) Do selfing species have greater niche breadth? Support from ecological niche modeling. Evolution 74:73–88. https://doi.org/10.1111/evo.13870

140. Cheptou P-O (2012) Clarifying Baker's Law. Ann Bot 109:633–641. https://doi.org/10.1093/aob/mcr127

141. McCauley MA, Milligan WR, Lin J, et al (2024) An empirical test of Baker's law: dispersal favors increased rates of self-fertilization. Evolution qpae177. https://doi.org/10.1093/evolut/qpae177

142. Orsucci M, Yang X, Vanikiotis T, et al (2022) Competitive ability depends on mating system and ploidy level across Capsella species. Ann Bot 129:. https://doi.org/10.1093/aob/mcac044

143. Bastias CC, Estarague A, Vile D, et al (2024) Ecological trade-offs drive phenotypic and genetic differentiation of Arabidopsis thaliana in Europe. Nat Commun 15:5185. https://doi.org/10.1038/s41467-024-49267-0

144. Schmickl R, Jorgensen M, Brysting A, Koch M (2010) The evolutionary history of the Arabidopsis lyrata complex: a hybrid in the amphi-Beringian area closes a large distribution gap and builds up a genetic barrier. BMC Evol Biol 10:98. https://doi.org/https://doi.org/10.1186/1471-2148-10-98

145. Li Y, van Kleunen M, Stift M (2019) Sibling competition does not magnify inbreeding depression in North American Arabidopsis lyrata. Heredity 123:723–732. https://doi.org/10.1038/s41437-019-0268-1

146. Griffin PC, Willi Y (2014) Evolutionary shifts to self-fertilisation restricted to geographic range margins in North American Arabidopsis lyrata. Ecol Lett 17:484–490. https://doi.org/https://doi.org/10.1111/ele.12248

147. Laenen B, Tedder A, Nowak MD, et al (2018) Demography and mating system shape

the genome-wide impact of purifying selection in Arabis alpina. Proc Natl Acad Sci 201707492

148. García-Dorado A (2012) Understanding and Predicting the Fitness Decline of Shrunk Populations: Inbreeding, Purging, Mutation, and Standard Selection. Genetics 190:1461–1476. https://doi.org/10.1534/genetics.111.135541

149. Zeitler L, Parisod C, Gilbert KJ (2023) Purging due to self-fertilization does not prevent accumulation of expansion load. PLOS Genet 19:1–27. https://doi.org/10.1371/journal.pgen.1010883

150. Toräng P, Wunder J, Obeso JR, et al (2015) Large-scale adaptive differentiation in the alpine perennial herb rabis alpina. New Phytol 206:459–470. https://doi.org/https://doi.org/10.1111/nph.13176

151. Excoffier L, Foll M, Petit RJ (2009) Genetic consequences of range expansions. Annu Rev Ecol Evol Syst 40:481–501. https://doi.org/10.1146/annurev.ecolsys.39.110707.173414

152. Bonchev G, Willi Y (2018) Accumulation of transposable elements in selfing populations of Arabidopsis lyrata supports the ectopic recombination model of transposon evolution. New Phytol 219:767–778. https://doi.org/https://doi.org/10.1111/nph.15201

153. Takou M, Hämälä T, Koch EM, et al (2021) Maintenance of Adaptive Dynamics and No Detectable Load in a Range-Edge Outcrossing Plant Population. Mol Biol Evol 38:1820–1836. https://doi.org/10.1093/molbev/msaa322

154. Charlesworth D, Charlesworth B (1995) Transposable elements in inbreeding and outbreeding populations. Genetics 140:415–417. https://doi.org/10.1093/genetics/140.1.415

155. Lockton S, Gaut B (2010) The evolution of transposable elements in natural populations of self-fertilizing Arabidopsis thaliana and its outcrossing relative

Arabidopsis lyrata. BMC Evol Biol 10:10. https://doi.org/https://doi.org/10.1186/1471-2148-10-10

156. Lockton S, Ross-Ibarra J, Gaut BS (2008) Demography and weak selection drive patterns of transposable element diversity in natural populations of Arabidopsis lyrata. Proc Natl Acad Sci 105:13965–13970. https://doi.org/10.1073/pnas.0804671105

157. Ågren JA, Wang W, Koenig D, et al (2014) Mating system shifts and transposable element evolution in the plant genus Capsella. BMC Genomics 15:602. https://doi.org/https://doi.org/10.1186/1471-2164-15-602

158. Hämälä T, Ning W, Kuittinen H, et al (2022) Environmental response in gene expression and DNA methylation reveals factors influencing the adaptive potential of Arabidopsis lyrata. Elife 11:e83115. https://doi.org/10.7554/eLife.83115

159. De Kort H, Legrand S, Honnay O, Buckley J (2022) Transposable elements maintain genome-wide heterozygosity in inbred populations. Nat Commun 13:7022. https://doi.org/10.1038/s41467-022-34795-4

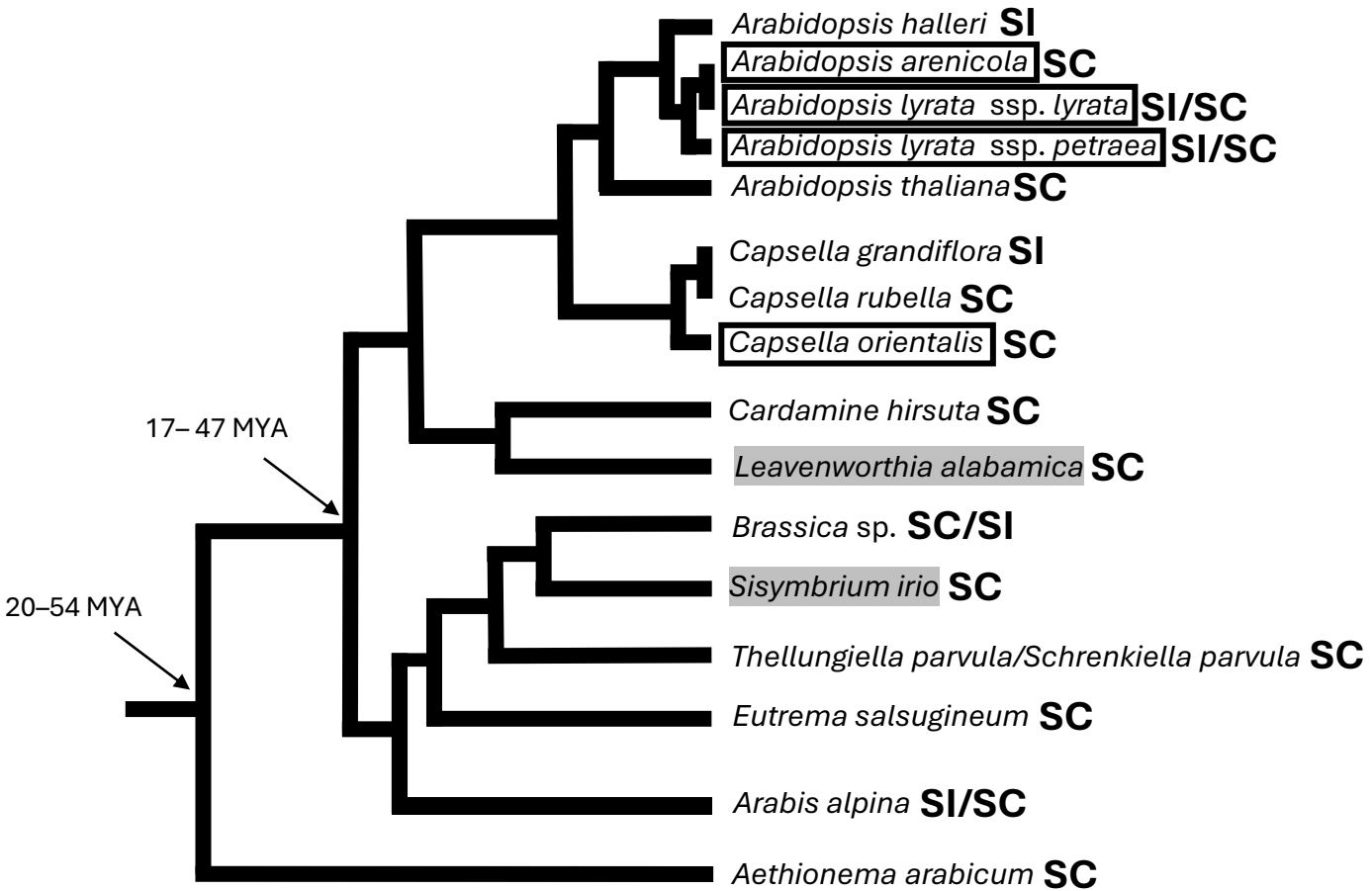
1072 Figure legends:

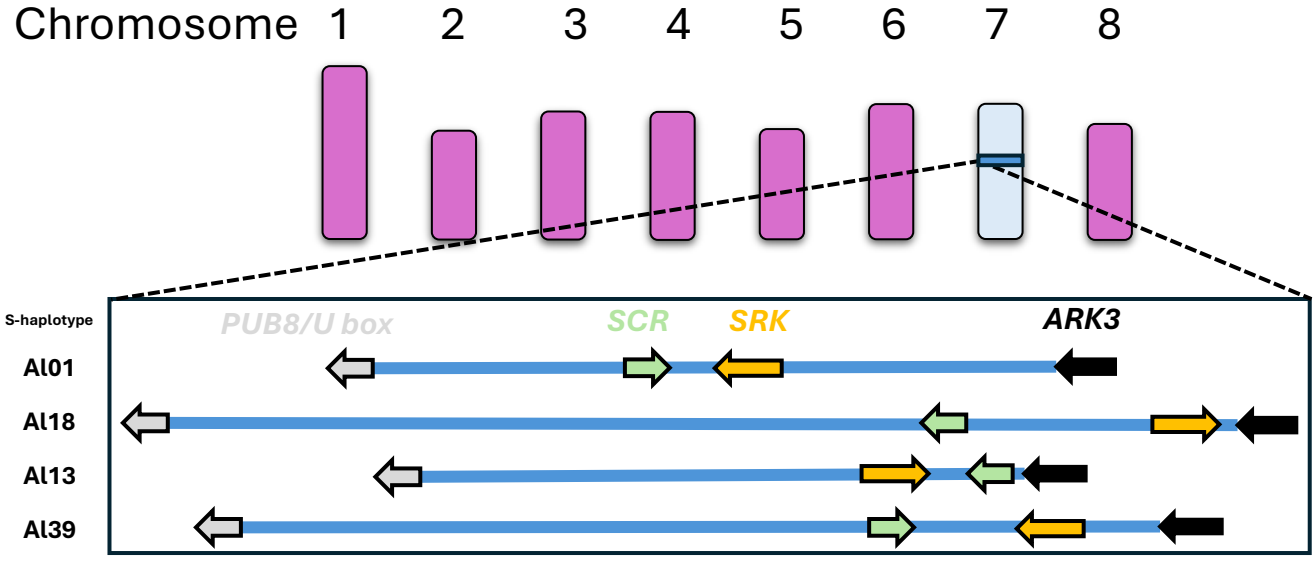
1073 Figure 1. Schematic illustration (not drawn to exact scale) of the branching order of selected diploid
1074 Brassicaceae lineages investigated in mating system genetics and genomics research. The self-recognition
1075 phenotypes of lineages or populations discussed in this review are indicated after the species names (SI = self-
1076 incompatible, SC = self-compatible). Topology and the split times are summarized from plastome gene-based
1077 phylogenies [37, 38]. Approximate positions of *Arabidopsis lyrata*, *A. arenicola,* and *Capsella orientalis* (black
1078 boxes) on the tree were inferred from [39] and [40]. *Leavenworthia alabamica* and *Sisybrium irio* were absent in
1079 the original trees (grey background) but were instead based on species from the same genus, whose
1080 monophyly was verified from [38, 41]. Time estimates for the two deepest nodes are taken from multiple time
1081 estimates compiled by [37]

1082

1083 Figure 2. Schematic illustration of the S-locus structure in the *Arabidopsis lyrata* genome. The model
1084 is based on *A. lyrata* ssp. *lyrata* genome structure [24] and selected S-alleles (Al01, Al18, Al13 and
1085 Al39) in *A. lyrata* [48]

1086


Arabidopsis halleri SI
Arabidopsis arenicola SC
Arabidopsis lyrata ssp. lyrata SI/SC
Arabidopsis lyrata ssp. petraea SI/SC
Arabidopsis thaliana SC
Capsella grandiflora SI
Capsella rubella SC
Capsella orientalis SC
Cardamine hirsuta SC
Leavenworthia alabamica SC
Brassica sp. SC/SI
Sisymbrium irio SC
Thellungiella parvula/Schrenkiella parvula SC
Eutrema salsugineum SC
Arabis alpina SI/SC
Aethionema arabicum SC
17– 47 MYA
20–54 MYA


Chromosome 1 2 3 4 5 6 7 8
S-haplotype
PUB8/U box
SCR
SRK
ARK3
Al01
Al18
Al13
Al39