



Dioecy is associated with high genetic diversity and adaptation rates in the plant genus *Silene*

Aline Muyle, Hélène Martin, Niklaus Zemp, Maeva Mollion, Sophie Gallina, Raquel Tavares, Alexandre Da Silva, Thomas Bataillon, Alex Widmer, Sylvain Glémin, et al.

► To cite this version:

Aline Muyle, Hélène Martin, Niklaus Zemp, Maeva Mollion, Sophie Gallina, et al.. Dioecy is associated with high genetic diversity and adaptation rates in the plant genus *Silene*. *Molecular Biology and Evolution*, 2021, 38 (3), pp.805-818. 10.1093/molbev/msaa229 . hal-02959763

HAL Id: hal-02959763

<https://hal.science/hal-02959763>

Submitted on 25 Nov 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Dioecy is associated with high genetic diversity and adaptation rates in the plant genus *Silene*

Aline Muyle^{1,2,*}, Hélène Martin^{3,4,*}, Niklaus Zemp^{5,6}, Maéva Mollion⁷, Sophie Gallina³, Raquel Tavares¹, Alexandre Silva⁸, Thomas Bataillon⁷, Alex Widmer⁵, Sylvain Glémin^{9,10}, Pascal Touzet^{3,‡}, Gabriel AB Marais^{1,‡}

¹ *Laboratoire de Biométrie et Biologie Evolutive (UMR 5558), CNRS/Université Lyon 1, 69622 cedex Villeurbanne, France.*

² *Department of Ecology and Evolutionary Biology, UC Irvine, Irvine, California 92697, USA.*

³ *Univ. Lille, CNRS, UMR 8198 - Evo-Eco-Paleo, F-59000 Lille, France.*

⁴ *Département de Biologie, Institut de Biologie Integrative et des Systèmes, Université Laval, Québec, Canada.*

⁵ *Institute of Integrative Biology, ETH Zurich, 8092 Zurich, Switzerland.*

⁶ *Genetic Diversity Centre (GDC), ETH Zurich, 8092 Zurich, Switzerland.*

⁷ *Bioinformatics Research Centre, Aarhus University, C.F. Møllers Allé 8, Aarhus C, 8000, Denmark.*

⁸ *Centro de Interpretação da Serra da Estrela (CISE), 6270-423 Seia, Portugal.*

⁹ *CNRS, Univ. Rennes, ECOBIO [(Ecosystèmes, biodiversité, évolution)] - UMR 6553, F-35000 Rennes, France*

¹⁰ *Department of Ecology and Genetics, Evolutionary Biology Center and Science for Life Laboratory, Uppsala University, 75236 Uppsala, Sweden.*

* These authors contributed equally as co-first authors to this work

‡ These authors contributed equally as co-senior authors to this work

Corresponding Authors: amuyle@uci.edu, helene.martin@tzone.org, pascal.touzet@univ-lille1.fr, gabriel.marais@univ-lyon1.fr

Abstract

About 15,000 angiosperm species (~6%) have separate sexes, a phenomenon known as dioecy. Why dioecious taxa are so rare is still an open question. Early work reported lower species richness in dioecious compared to non-dioecious sister clades, raising the hypothesis that dioecy may be an evolutionary dead-end. This hypothesis has been recently challenged by macro-evolutionary analyses that detected no or even positive effect of dioecy on diversification. However, the possible genetic consequences of dioecy at the population level, which could drive the long-term fate of dioecious lineages, have not been tested so far. Here, we used a population genomics approach in the *Silene* genus to look for possible effects of dioecy, especially for potential evidence of evolutionary handicaps of dioecy underlying the dead-end hypothesis. We collected individual-based RNA-seq data from several populations in 13 closely related species with different sexual systems: seven dioecious, three hermaphroditic and three gynodioecious species. We show that dioecy is associated with increased genetic diversity, as well as higher selection efficacy both against deleterious mutations and for beneficial mutations. The results hold after controlling for phylogenetic inertia, differences in species census population sizes and geographic ranges. We conclude that dioecious *Silene* species neither show signs of increased mutational load nor genetic evidence for extinction risk. We discuss these observations in the light of the possible demographic differences between dioecious and self-compatible hermaphroditic species and how this could be related to alternatives to the dead-end hypothesis to explain the rarity of dioecy.

Introduction

Out of 261,750 angiosperm species, approximately 15,600 (~6%) are dioecious (Renner 2014). Dioecy is thus rare compared to the dominant sexual system in angiosperms, hermaphroditism, in which individuals carry bisexual flowers. Nonetheless, dioecy can be found in ~40% of angiosperm families. The current view is that many independent (871 to ~5000) and mostly recent transitions occurred from hermaphroditism to dioecy in angiosperms (Renner 2014). These events probably followed two main evolutionary paths: one through a gynodioecious intermediate (with female and hermaphrodite individuals), called the gynodioecy pathway, and another one through a monoecious intermediate (with a single individual carrying separate female and male flowers) called the monoecy or paradioecy pathway (Barrett 2002). In few cases, dioecy might have evolved through alternative pathways, such as heterostyly (where individuals have different style and stamen lengths, Barrett 2002).

Species richness was originally found to be lower in dioecious compared to non-dioecious sister clades, suggesting that dioecy might lead to an increased extinction rate in angiosperms and defining the evolutionary dead-end hypothesis (Heilbuth 2000). Theoretical work has explored this idea and shown that an increased extinction rate could result from two main evolutionary handicaps. First, in dioecious species, only females contribute to seed production and dispersal, which could result in individuals being less dispersed spatially and competing more for local resources compared to populations with hermaphrodites (Heilbuth et al. 2001). This should cause dioecious species to have smaller geographic distributions and species abundance compared to non-dioecious species. Second, in dioecious species with insect pollination, male intra-sexual competition would result in males being more attractive than females for pollinators, making reproduction and population size sensitive to variation in pollinator density (Vamosi and Otto 2002). This should lead to fluctuating population sizes in dioecious species and in turn affect levels of genetic diversity. It has been proposed that these handicaps could be the reason why dioecy is associated with a specific suite of traits in angiosperms (fleshy fruits, wind-pollination, inconspicuous flowers and woody growth) as well as ecological preferences because dioecious species are preferentially found in the tropics and on islands (Vamosi et al. 2003). Assuming that pollination mode evolves more slowly than dioecy, dioecy could be more frequent among wind-pollinated species because wind pollinated-species are not subject to the pollinator handicap. Similarly, assuming fruit type evolves more slowly than dioecy, dioecy could be more prevalent in species with fleshy fruits because seeds can be dispersed over longer distances by fruit-eating animals, thus limiting the seed-dispersal handicap.

The evolutionary dead-end hypothesis has recently been challenged (Kafer et al. 2017), in part because the sister clade analysis used in early work assumed that speciation events and trait

transitions coincide (Heilbuth 2000). However, if dioecy evolved after the split between the two diverging lineages and was not coincident with the split, then species richness is not expected to be equal between sister clades even when diversification rates are similar (Kafer and Mousset 2014). Using a corrected version of the sister clade analysis on an updated dataset, dioecious clades have been found to have actually higher species richness than non-dioecious sister clades in angiosperms (Kafer et al. 2014). Using a phylogenetic method to study trait evolution and species diversification on several genera with dioecious species (Maddison et al. 2007), dioecy was not associated with an increased extinction rate (Sabath et al. 2016; Duan et al. 2018). Another hypothesis to explain the rareness of dioecy in angiosperms posits that reversion from dioecy to other sexual systems is frequent and that dioecy is more evolutionary labile than previously thought (Kafer et al. 2017). This hypothesis was supported by a study of 2145 species from 40 genera in which the authors found evidence of frequent reversions from dioecy to other breeding systems (Goldberg et al. 2017).

However, these works relied on macro-evolutionary and comparative approaches and the population functioning of dioecious versus non-dioecious species has so far been poorly explored. Investigating consequences both at the macro-evolutionary and population scale has been fruitful in the study of the evolution of selfing from outcrossing, another frequent breeding system transition in angiosperms. Selfing was also proposed to be an evolutionary dead-end, which has been supported by several macroevolutionary analyses (Goldberg et al. 2010; de Vos et al. 2014; Zenil-Ferguson et al. 2019). In parallel, various negative genetic consequences of selfing (i.e. low genetic diversity, accumulation of deleterious mutations) have been reported in several species, even in those of recent origin (see Glémin et al. 2019 for a review). Interestingly, negative genetic consequences have been detected despite the short terms advantage of selfing provided by reproductive assurance, allowing higher colonization potential and larger species range (Randle et al. 2009; Grossenbacher et al. 2015).

If the seed-dispersal handicap of dioecy operates, it should reduce the ability of dioecious plants to spread and increase local competition, which should lead to smaller geographic distributions and lower species abundance in dioecious compared to non-dioecious species. These expectations can be tested using estimates of geographic distributions and species abundance in dioecious and non-dioecious species. If the pollinator handicap of dioecy exists, it should cause fluctuations in dioecious population sizes. The effective population size (N_e) is given by the harmonic mean of the varying population sizes over time (Wright 1938). The harmonic mean tends to be dominated by the smallest population sizes that the population goes through. Fluctuating population sizes are therefore expected to decrease the long-term N_e of a species. If the pollinator and the seed-dispersal handicap of dioecy exist, they should cause dioecious species to have smaller N_e compared to non-dioecious species. This expectation can be tested by studying genetic diversity

to estimate the long-term N_e of dioecious and non-dioecious species. A reduced effective population size should then lead to a decrease in the efficacy of selection. This expectation can be tested by comparing closely related dioecious and non-dioecious species. A study that used the ratio of nonsynonymous to synonymous substitution rate (d_N/d_S) as a proxy to selection efficacy found that one out of two dioecious *Silene* clade had a reduced efficacy of selection for some genes, but positive selection increased in other genes (Kafer et al. 2013). However, transitions to dioecy can happen after speciation, making d_N/d_S biased because it is computed on entire branches. We therefore decided to use instead a population genetics approach to confront predictions of the dead end hypothesis in *Silene*. We relied on statistics that allow quantifying the strength of natural selection, such as rates of adaptive and deleterious nonsynonymous substitutions. If dioecy is not a dead-end, as suggested by macro-evolutionary analyses, we expect no effect (or even a positive effect) of dioecy on genetic diversity and the efficacy of selection. Note that if extinction due to the handicaps of dioecy is too rapid to allow for populations to survive long enough to be sampled, our approach would not be able to detect the genetic footprints expected under the dead-end hypothesis. However, as mentioned above, similar approaches in selfing species have consistently detected the predicted effects, even in very recent species such as *Capsella rubella* that originated about 20,000 years ago from the self-incompatible *C. grandiflora* (Brandvain et al. 2013; Slotte et al. 2013).

In this study, we focused on 13 species of the large genus *Silene* (~700 species) in the Caryophyllaceae family, in which dioecy has evolved independently at least twice (Desfeux et al. 1996). The *Silene* genus is considered a model system to study the evolution of sexual systems in angiosperms, thanks to its reasonably well-described sexual systems and phylogeny (Bernasconi et al. 2009). Dioecy in *Silene* is likely to have evolved through the gynodioecy pathway (Fruchard and Marais 2017). If dioecy is associated with the handicaps described above, dioecious *Silene* species should suffer from them because they do not have any of the traits named previously that likely compensate or prevent these handicaps. Indeed, *Silene* are short-lived herbs, many of them (but not all) are insect pollinated, they do not have fleshy fruits and they live in temperate regions. The *Silene* genus is thus adequate to test for the dead-end hypothesis of dioecy.

We sampled dioecious species from two groups in which dioecy evolved independently (5 species in the *Melandrium* section and 2 species in the *Otites* section) along with several hermaphroditic and gynodioecious close relatives for each group and one outgroup (*Dianthus chinensis*). *Silene* species can have very large genome sizes (Bernasconi et al. 2009) and no complete reference genome is available in the genus (only a fourth of the *Silene latifolia* genome has been assembled to date (Papadopoulos et al. 2015)). We therefore used a population genomics approach relying on RNA-seq data, which has proven successful in organisms that lack a reference genome (Cahais et al. 2012; Gayral et al. 2013; Romiguier et al. 2014). Multiple individuals were

sampled across the species' geographic ranges and sequenced by RNA-seq. A pipeline specifically designed for such data was applied to perform SNP-calling (Tsagkogeorga et al. 2012; Gayral et al. 2013; Nabholz et al. 2014). The SNPs were then analyzed to estimate genetic diversity and the efficacy of selection in each species.

Results

Phylogeny, geographic range, species abundance and population genetics characterization of the sampled species

Our species sampling (13 species in total, see Fig. 1) includes two hermaphroditic species (*S. viscosa* and *S. paradoxa*), three gynodioecious species (*S. vulgaris*, *S. nutans* W and *S. nutans* E) and seven dioecious species (section *Melandrium*: *S. heuffelii*, *S. dioica*, *S. marizii*, *S. diclinis*, *S. latifolia* and section *Otites*: *S. pseudotites* and *S. otites*) from subgenera *Behenantha* and *Silene*, the main subgenera in the *Silene* genus (Jenkins and Keller 2010). The criteria for selecting the non-dioecious species are detailed in the Materials and Methods. Briefly, we selected (i) close relatives of both dioecious sections based on existing phylogenies, (ii) species for which the sexual system is well described, (iii) species that had similar geographic ranges and (iv) hermaphroditic species not exhibiting high levels of selfing. The latter criterion was meant to perform a fair comparison between dioecious and hermaphroditic species, not confounded by strong differences in outcrossing rates.

For each species, between 4 and 34 individuals were sampled from different populations covering the geographic range of the species (see Fig. 1 for the number of individuals sampled in every species and Fig. S5 for sample locations). Previous work has shown that reliable estimates of population genetics statistics can be obtained when individuals from populations spanning the whole geographic distribution of a species are sampled, even if the number of sampled individuals is small, i.e. 2 or 3 (Cahais et al. 2012; Gayral et al. 2013; Romiguier et al. 2014).

Seeds were collected in the wild and plants were grown in controlled greenhouse conditions. Flower buds (and in a few cases, leaves) were sampled for 130 individuals in total and sequenced by RNA-seq (see Table S1 for details on samples). *De novo* reference transcriptomes were assembled separately for each species and ORFs and orthologs were annotated (see Materials and methods and Table S1). All dioecious *Silene* species sampled here have sex chromosomes. Merging autosomal and sex-linked markers can be problematic in population genetics studies as they have different effective population sizes and selection dynamics (Gautier 2014). Therefore, only autosomal ORFs – as predicted by the SEX-DETECTOR method (Muyle et al. 2016) – were kept for downstream analyses (see Materials and methods).

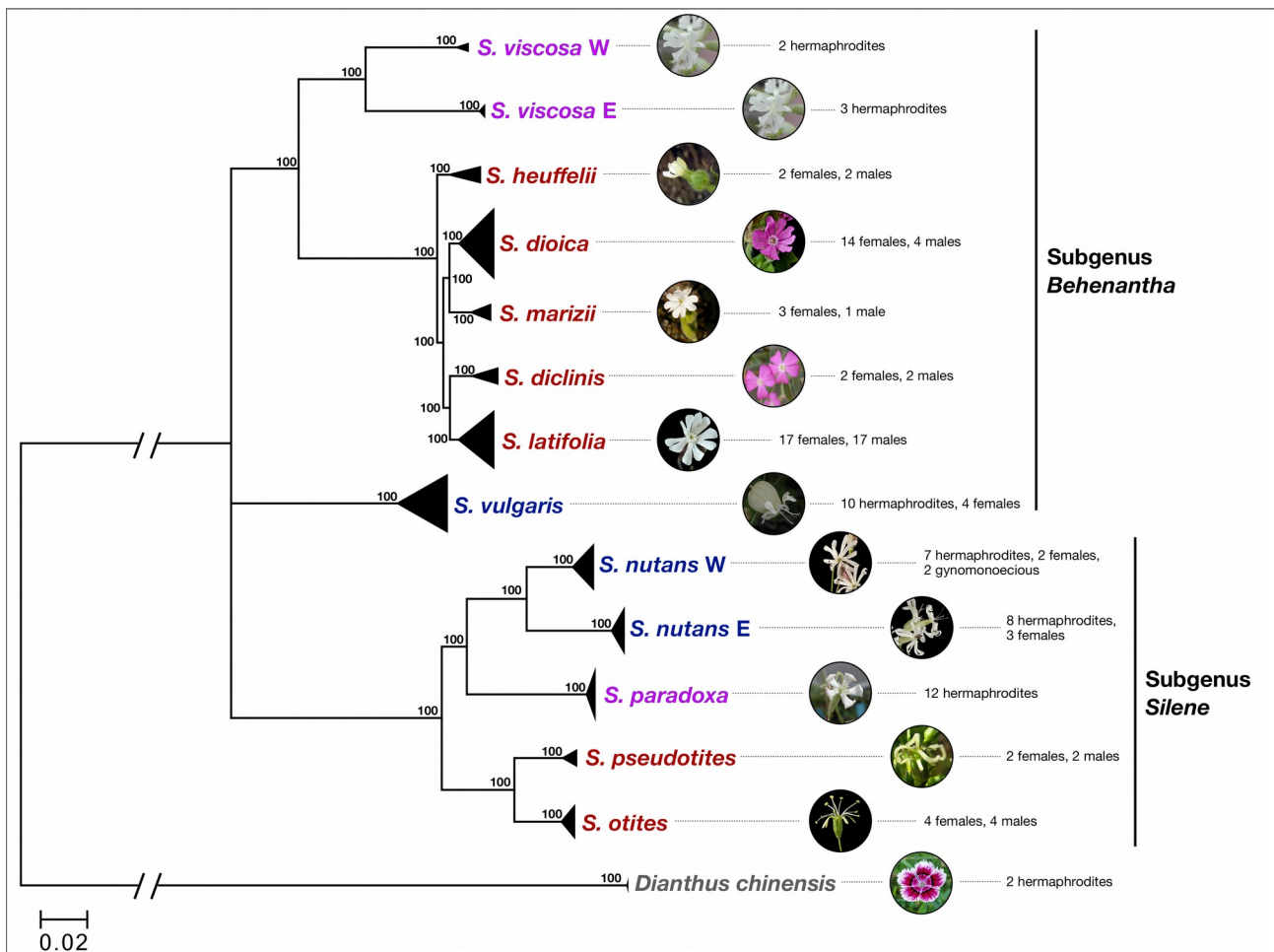


Fig. 1. Sampled species and their phylogenetic relationships. Dioecious species are shown in red, gynodioecious species in blue and hermaphroditic species in magenta. Sample sizes and sex are indicated. Bootstrap support values were generated using 100 replicates. Pictures of *S. latifolia*, *S. dioica*, *S. nutans* W, *S. nutans* E and *S. otites* are from Maarten Strack van Schijndel. Pictures of *S. marizii*, *S. vulgaris*, *S. paradoxa* and *S. viscosa* are from Oxelman *et al* (2013). Picture of *S. heuffelii* is from Dr Daniel L. Nickrent.

A phylogenetic tree was built in order to check for congruence with published phylogenies of the *Silene* genus (Jenkins and Keller 2010; Rautenberg *et al.* 2010). Orthologous ORFs were aligned among *Silene* species and *Dianthus chinensis*, which was used to root the tree. A total of 55,337 variable positions, spread over 1,294 contigs, were used to build the tree (see Materials and methods). All samples from a given species grouped together as expected (Fig. 1). The two lineages of *S. nutans* formed two clearly distinct groups that should be treated as distinct species as shown in previous studies (Martin *et al.* 2016; Martin *et al.* 2017). *S. viscosa* samples also formed two highly divergent lineages, even more divergent than the two lineages of *S. nutans*: a western lineage (*S. viscosa* W), which includes samples from Sweden and the Czech Republic, and an eastern lineage (*S. viscosa* E), which includes samples from Bulgaria, Russia and Kyrgyzstan. We treated these lineages as two differentiated taxa and analyzed them separately hereafter. Unlike in previous work

(Slancarova et al. 2013), *S. vulgaris* did not group with *S. viscosa* and the section *Melandrium*, but instead could not be reliably grouped with either subgenus (Fig. 1).

A total of 1,088,269 variable nucleotides were called (Table 1) using an RNA-seq pipeline (see Materials and methods). Mean heterozygosity levels and Weir and Cockerham F_{IT} statistics were computed for each species. F_{IT} indicates deviations from Hardy-Weinberg equilibrium caused by processes such as selfing or population structure. F_{IT} values were low to moderate (below 0.2) for most sampled species. This indicates that neither population structure nor selfing are creating strong among-species differences in genetic diversity, as already observed (Moyle 2006). However, *S. viscosa* E exhibited a high F_{IT} , a low heterozygosity level and low genetic diversity, which could be due to recent demographic shifts or a high selfing rate (Fig. 2).

Proxies for species abundance were calculated based on the number of occurrences of each species retrieved from the Global Biodiversity Information Facility (GBIF) database. The geographic range of each species was extracted from the Atlas *Florae Europaeae* (see Materials and methods, Fig. S6, Fig. 3). Dioecious and non-dioecious species had similar geographic ranges (622.6 versus 784 Atlas dots on average, Wilcoxon one-sided rank sum test p-value=0.52) and GBIF species abundance (30,777 versus 23,968.5 on average, Wilcoxon one-sided rank sum test p-value=0.43). This is not consistent with the dead-end hypothesis which predicts smaller geographic distributions and population sizes for dioecious species. However, some *Silene* dioecious species have very limited geographic ranges, making the study of their genetic diversity informative to test whether they are genetically endangered.

Genetic diversity is higher in dioecious *Silene* species, regardless of census population size differences

In order to test whether dioecious and non-dioecious *Silene* species differ in their effective population size (N_e), synonymous genetic diversity (π_s) was used as a proxy for N_e with $\pi_s = 4N_e\mu$ (the mutation rate μ was assumed constant among species). N_e is a direct measure of the long term mean intensity of drift in populations. Under the evolutionary dead-end hypothesis, dioecious species are expected to have lower N_e and therefore lower values of π_s compared to non-dioecious species (see Introduction). Table 1 and Fig. 2 showed that dioecious species have higher π_s values than non-dioecious species (Wilcoxon one-sided rank sum test $W = 35$, p-value = 0.026), which does not support the dead-end predictions.

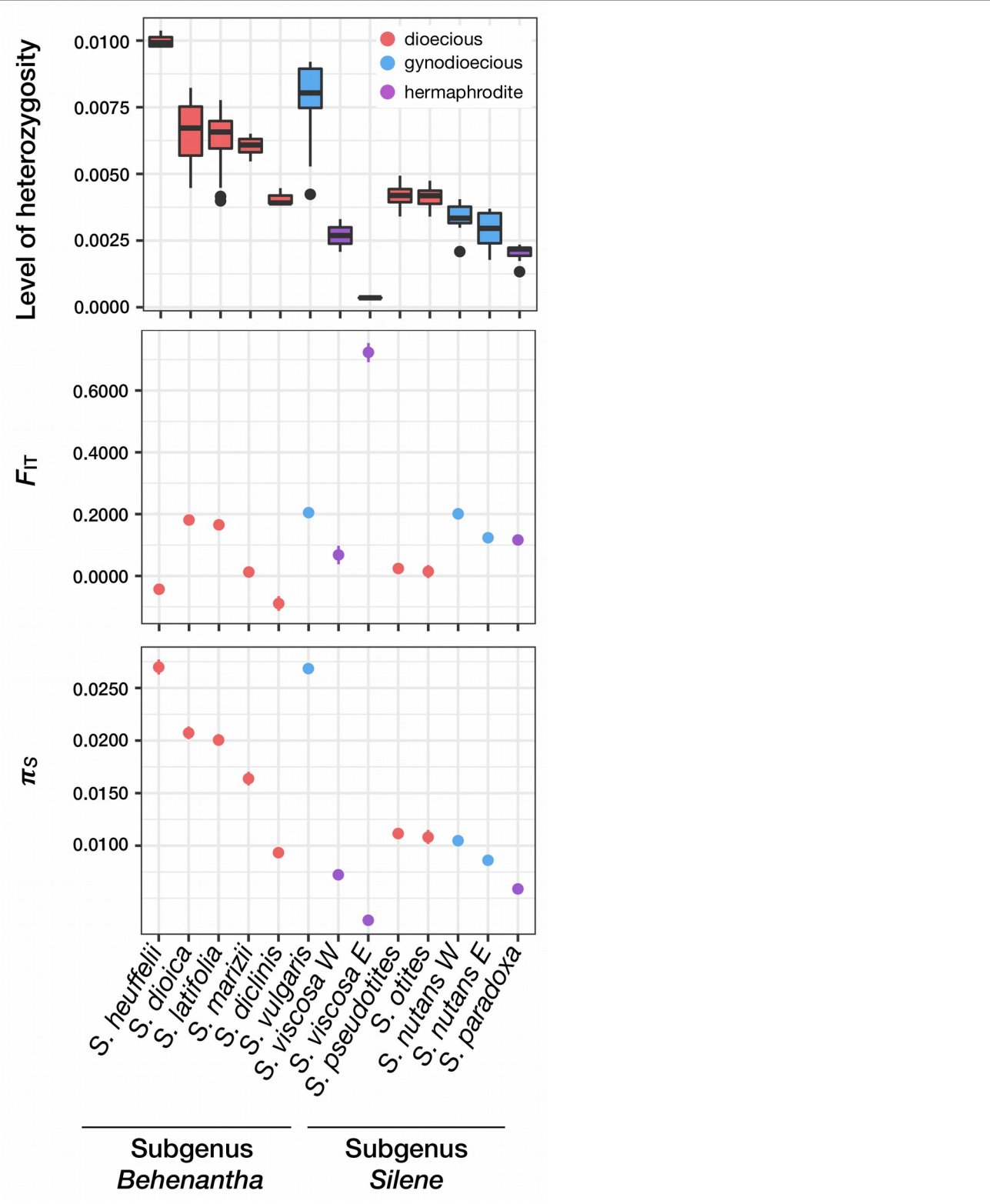


Fig. 2. Levels of individual heterozygosity (proportion of heterozygote positions), Weir and Cockerham F_{IT} statistic and π_S within each species. Dots show the mean value and vertical lines mark the 95% confidence interval (obtained by 1,000 bootstraps on contigs). Dioecious species are shown in red, gynodioecious species are shown in blue and hermaphroditic species are shown in magenta.

However, the 13 species studied here have very different census population sizes (N) with some species being cosmopolitan (e.g. *S. latifolia*, *S. vulgaris*) while others are rare or even endemic (e.g. *S. diclinis*). Genetic diversity can be affected by N , although the relationship can be complex such that N and N_e are quite different (Kalinowski and Waples 2002). In order to make a fair comparison of dioecious and non-dioecious species, we accounted for the effect of N when looking at the effect of sexual systems on genetic diversity. Two different proxies for N were used: (i) the species abundance in the GBIF dataset, and (ii) the geographic range of each species according to the *Atlas Florae Europaeae*.

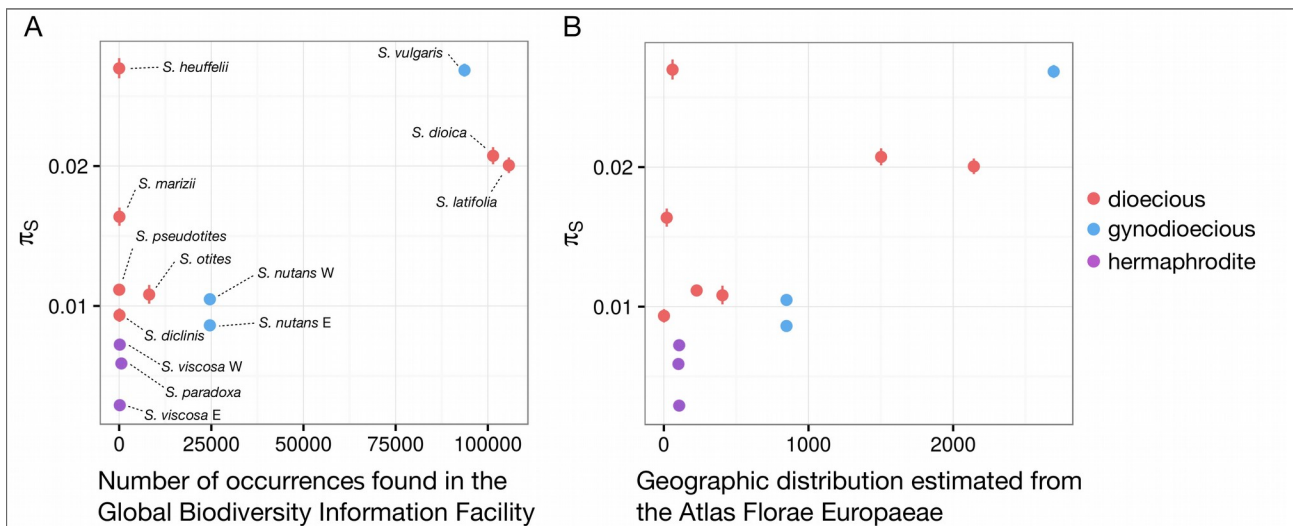


Fig. 3. Relationship between synonymous polymorphism (π_s) and **(A)** census population size estimated from the Global Biodiversity Information Facility (GBIF) and **(B)** geographic range estimated from the *Atlas Florae Europaeae*. For each species, dots represent the mean π_s and vertical lines mark the 95% confidence interval (obtained by 1,000 bootstraps on contigs). Census population sizes estimated from GBIF refer to the number of occurrences, and geographic distributions from the *Atlas Florae Europaeae* refer to the number of dots (native occurrences) counted on the Atlas maps (see Materials and methods). Dioecious species are shown in red, gynodioecious species are shown in blue and hermaphroditic species are shown in magenta.

We used a linear model to jointly assess the effects of N and sexual systems on π_s . GBIF data and Atlas geographic range (estimates of N) were significantly positively associated with π_s (p -values = 0.030 and 0.012, t -values = 2.529 and 3.054, R^2 = 0.390 and 0.483 respectively, and see Fig. 3). This shows that genetic diversity is strongly influenced by N in our *Silene* sample set. This also suggests that, in the case of the *Silene* genus, GBIF abundance and Atlas geographic ranges are sensible proxies for N . After controlling for the effect of N on genetic diversity, dioecious species

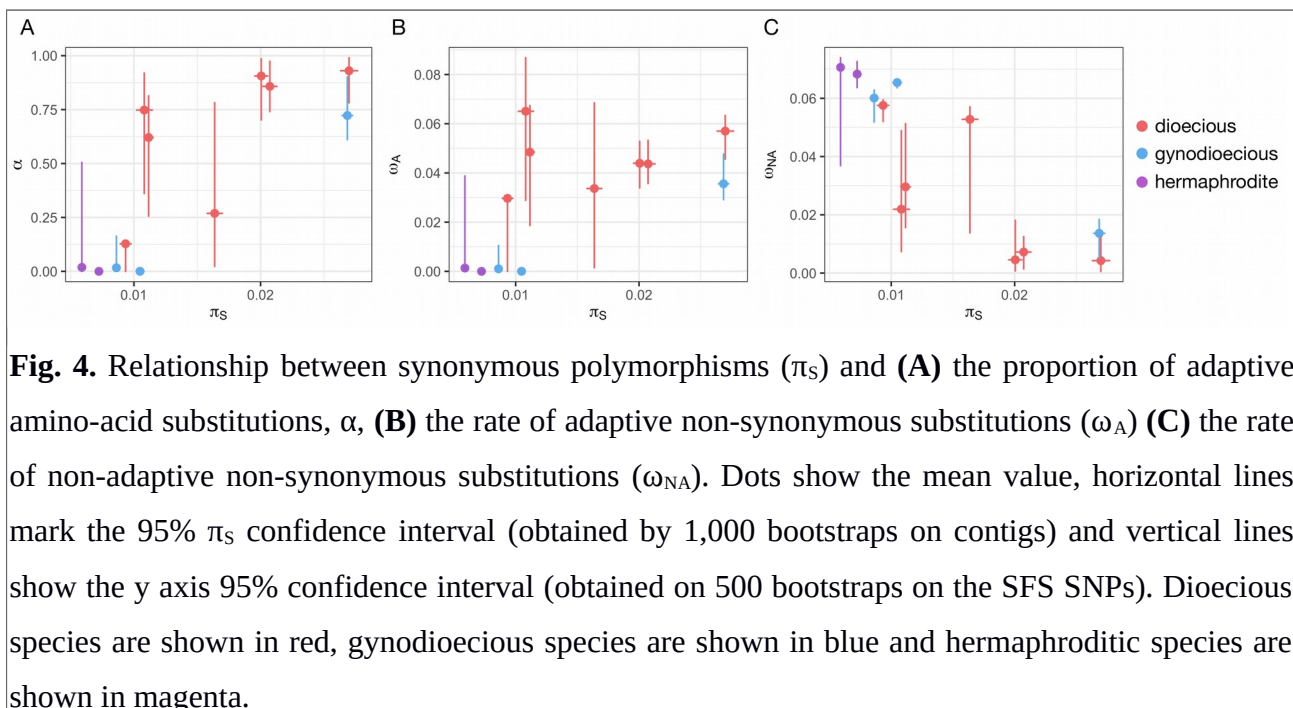
had a significantly higher π_s than non-dioecious species (linear model pairwise t-test p -values = 0.142 and 0.048, t -values = 1.595 and 2.247 using GBIF data and Atlas geographic range estimates respectively). Adding F_{IT} to the linear model correlating genetic diversity and N did not significantly improve our linear model (anova $p=0.102$ and $p=0.097$ for GBIF and Atlas geographic range estimates respectively, see Materials and Methods for details). This suggests that the increase in genetic diversity observed under dioecy is not attributable to inbreeding evasion.

These effects remained after correcting for phylogenetic relationships among species using generalized least square models in the R package *ape* (Paradis and Schliep 2019): N was significantly positively correlated to π_s (p -values = 0.012 and 0.03, t -values = 3.05 and 2.53 for Atlas and GBIF respectively) and dioecious species had a significantly higher π_s compared to non-dioecious species when using Atlas geographic range estimates (p -value = 0.048, t -value = 2.25), but not when using GBIF estimates (p -value = 0.14, t -value = 1.6). In this analysis, different sets of genes were used in different species. As this might affect the results, the analysis was repeated on two smaller sets of genes, one shared by every species in the subgenus *Behenantha* and the second shared by every species in the subgenus *Silene*. Similar results were found using these shared gene subsets, i.e. GBIF data and Atlas geographic range have a significant positive effect on π_s (p -values = 0.021 and 0.012, t -values = 2.725 and 3.074, R^2 = 0.429 and 0.488, respectively). And dioecious species had higher π_s compared to other species after correcting for N (linear model pairwise t-test p -values = 0.127 and 0.045, t -values = 1.662 and 2.283 using GBIF data and Atlas geographic range estimates respectively, see Fig. S1).

Selection is more efficient in dioecious *Silene* species, regardless of differences in N_e

We then tested whether selection efficacy differs in dioecious and non-dioecious *Silene* species. Selection should be more efficient in species with larger N_e . The ratio of non-synonymous to synonymous substitution rate, d_N/d_S , could not be used here because the studied species are closely related and some shifts in sexual systems might have occurred too recently to allow for fixed differences to evolve (Desfeux et al. 1996; Bernasconi et al. 2009). Rather, the ratio of non-synonymous over synonymous polymorphism, π_N/π_S , was used, because it is a more appropriate statistic to study the efficacy of selection on a short evolutionary timescale (Mugal et al. 2014). One expects increased π_N/π_S values when purifying selection is reduced and hence the dead-end hypothesis would predict elevated π_N/π_S in dioecious species. However, we found no statistical association between sexual systems and π_N/π_S (Wilcoxon rank sum test $p=0.295$, Table 1), perhaps reflecting the opposing effects that purifying and positive selection have on π_N/π_S , or perhaps reflecting the fact that breeding systems do not have an effect on π_N/π_S .

We thus turned to a more sophisticated approach to quantify selection efficacy and used a modified version of the McDonald-Kreitman framework, Implemented in the method polyDFE (Tataru et al. 2017; Tataru and Bataillon 2020). This method estimates the distribution of fitness effects (DFE) of both deleterious and beneficial mutations using polymorphism data only, thus avoiding the problem of shifts in population sizes. *Silene* species underwent a post-glacial age population expansion in Europe and for this reason it would be biased to use methods that also rely on divergence (Keightley and Eyre-Walker 2007; Galtier 2016). Indeed, other methods to estimate the DFE use both polymorphism and divergence data and usually return estimates of proportion of adaptive substitution that are assuming that population size remained constant since divergence with the outgroup. *S. viscosa* E was not included because its diversity was too low due to selfing and the polyDFE model did not apply. PolyDFE was used to estimate the DFE (Fig. S2) for each species with a gamma distribution for deleterious mutations and exponential distribution for beneficial mutations (see estimated parameter ranges in Fig. S3). From this, the predicted proportion of adaptive substitutions (α), the predicted rate of adaptive substitutions (ω_A), and the predicted rate of deleterious substitutions (ω_{NA}) were estimated. Under the dead-end hypothesis, dioecious species are expected to have a higher ω_{NA} , a lower α and a lower ω_A because of a faster accumulation of deleterious mutations and a lower rate of adaptation due to genetic drift, when compared to non-dioecious species. Because N_e is known to influence the efficacy of selection, we correct for differences in N_e (using π_S as a proxy) between species in order to make a fair comparison of dioecious and non-dioecious species.



A linear model was used to jointly assess the effects of N_e (using π_s as a proxy) and of the sexual system on α . π_s was significantly positively correlated to α (p -value = 0.003, t -value = 4.07, Fig. 4). The same result has previously been observed in animal species (Galtier 2016) where the correlation between α and π_s was solely due to the link between π_s and ω_{NA} . Because $\alpha = \omega_A / (\omega_A + \omega_{NA})$, when selection is more efficient (higher N_e and higher π_s), the rate of deleterious substitutions ω_{NA} is reduced, which in turn increases α . However, in the present dataset, the correlation between α and π_s was both due to higher adaptive evolutionary rates and lower deleterious evolutionary rates when effective population size is high. Indeed, π_s was significantly positively linked to ω_A (linear model p -value = 0.011, t -value = 3.19) and significantly negatively correlated to ω_{NA} (linear model p -value = 0.00017, t -value = -6.15, see Fig. 4). These effects remained after correcting for phylogenetic relationships among species using *gls* models in the R package *ape* (Paradis and Schliep 2019): π_s was significantly positively correlated to α and ω_A (p -values = 0.026 and 0.0054, t -values = 4.11 and 3.64 for α and ω_A respectively) and π_s was negatively correlated to ω_{NA} (p -value = 0.0015 and t -value = -4.49).

Estimates of α , ω_A and ω_{NA} were then compared between dioecious and non-dioecious species, after controlling for the effect of N_e (using π_s as a proxy). Non-dioecious species had a significantly lower α and ω_A and a significantly higher ω_{NA} compared to dioecious species (linear model pairwise t -test p -values = 0.015, 0.00031 and 0.019, t -values = -2.99, -5.65 and 2.87 for α , ω_A and ω_{NA} respectively). These effects remained significant after correcting for phylogenetic relationships among species using *gls* models in the R package *ape* (p -values = 0.032, 0.0007 and 0.058, t -values = -2.53, -5.03 and 2.17 for α , ω_A and ω_{NA} respectively). Dioecious species thus experience fewer fixations of slightly deleterious mutations due to genetic drift and more adaptive evolution, which is opposite to what we expect under the dead-end hypothesis. The analyses were repeated using orthologous genes shared within the subgenus *Behenantha* or within the subgenus *Silene* as different sets of genes in different species could affect the results. The conclusions remained unchanged; differences were still in the same direction although statistical power was reduced in some cases (see Fig. S4 legend for details).

Discussion

In the set of *Silene* species studied here, dioecy is associated with higher genetic diversity, more efficient purifying selection and increased adaptive evolution. These results contradict the dead-end hypothesis and are consistent with macro-evolutionary analysis that rejected the dead-end from different lines of evidence. High levels of genetic diversity in dioecious species had previously been observed in *Silene dioica* (Giles and Goudet 1997), *Silene chlorantha* (Lauterbach et al. 2011),

Silene otites (Lauterbach et al. 2012), *Ficus* (Nazareno et al. 2013), *Antennaria dioica* (Rosche et al. 2018) and a comparison of 12 monoecious and 11 dioecious plants (Nazareno et al. 2013). Some of these studies have shown that the sex ratio strongly influences genetic diversity through its influence on the outcrossing rate. We do not have access to sex ratios of the species sampled here but their study would be insightful to better understand patterns of diversity in *Silene* species in future. Our results clearly show that dioecious *Silene* species are not at risk of extinction given their high genetic diversity. Observed rates of adaptive non-synonymous substitution patterns suggest that dioecious species undergo more adaptation than non-dioecious species, however, the underlying causes of this pattern remain to be determined.

The *Silene* genus is a good model system to test for the dead-end hypothesis because dioecious *Silene* species do not have any of the ecological traits thought to possibly compensate for the handicaps of dioecy (see Introduction). In other genera in which dioecy is ancient, it is likely that enough time has passed for species to evolve compensating mechanisms to counteract the handicaps of dioecy, if these truly exist. This is less likely to be an issue in the *Silene* genus where dioecy has evolved fairly recently: ~11 Mya old in the *Melandrium* section (Krasovec et al. 2018) and even more recently in section *Otites* (Slancarova et al. 2013).

Our approach, however, also has its limitations. First, our set of species represents only two independent events of dioecy evolution and a small sample of the whole *Silene* genus (~700 species). However, we did not find a significant phylogenetic signal on genetic diversity in our species set (Pagel's $\lambda=0.848$, $p=0.133$), suggesting that each species can be considered as independent even though some species share the same transition event to dioecy. Nonetheless, other genera should be studied in the future to assess whether our findings can be generalized to other angiosperm species, especially to groups in which dioecy evolved through different routes (see Introduction).

A second limitation is that sampling was limited by the number of *Silene* species for which the sexual system is well described and for which the geographic distribution is similar among species, especially for hermaphrodites (see Materials and Methods for detailed criteria used to select species). None of the *Silene* hermaphrodite species have geographic distributions as wide as the widest dioecious species range. However, dioecious species that have geographic ranges comparable to those of the hermaphroditic species we sampled still have higher diversity and selection efficacy compared to hermaphrodites (Fig. 3 and 4), which is conservatively in disagreement with the dead-end hypothesis. For two species out of thirteen, namely *S. otites* (dioecious) and *S. paradoxa* (non-dioecious), we were not able to sample individuals across the whole geographic range of the species and their genetic diversity might be underestimated compared to others.

A third limitation comes from the comparison of dioecious and non dioecious species which implies comparing species with sex chromosomes to species with only autosomes. Sex-linked genes, which are located on the non-recombining region of sex chromosomes, are known to suffer from lower N_e , which could bias comparisons of dioecious to non dioecious species. Because sex-linked genes in *Silene* represent <10% of genes (discussed in Muyle et al. 2012), we chose to subset our study to autosomal genes only. Our results therefore overlook the potential negative effects of sex chromosomes on selection efficacy in dioecious species.

A fourth limitation of our study is that, if the extinction dynamics of dioecious species is extremely fast, there might not be enough time for deleterious genetic footprints to accumulate in dioecious species before they go extinct. The micro-evolutionary approach taken here can only detect handicaps of dioecy that do not lead to immediate extinction. Such a study, as others testing evolutionary dead-end hypotheses, suffers from a “survivor bias” as only non-extinct species can be analyzed. However, some of the dioecious species studied here have small geographic distributions and one is even endemic. These species should be strongly affected by the handicaps of dioecy, if there are any. But even these species exhibited high levels of diversity and selection efficacy.

In the *Silene* genus, all hermaphrodites are self-compatible and therefore able to self-fertilize. Outcrossing promotes genetic diversity by increasing effective population size and minimizing the effects of linkage (Nordborg 2000; Glémin et al. 2019). Dioecy could be beneficial in *Silene* by preventing the negative effects of selfing that occur in hermaphroditic and gynodioecious species (Charlesworth and Charlesworth 1978). One selfing hermaphrodite subspecies (*S. viscosa* E) could not be included in the analysis of selection efficacy because its genetic diversity was too low. However, two hermaphrodite species that are not heavy selfers (Figure 2) were included in our comparison of dioecious and non-dioecious species. However, the fact that significant differences in terms of selection efficacy between dioecious and non-dioecious species remain, even without this selfing species, highlights that the advantages of dioecy that we observe do not result from comparing selfing and outcrossing species. What is more, the increased genetic diversity observed in dioecious species could not be explained by differences in inbreeding levels in our dataset.

However, the main difference between dioecious species and hermaphrodites is the ability to *occasionally* self-fertilize, which can strongly affect the demography of a species. Indeed, in hermaphroditic self-compatible species, a single individual can start a new population (Schoen and Brown 1991). This reproductive assurance provides an ecological advantage to hermaphroditic self-compatible species over dioecious species. However, the other side of the coin is that hermaphroditic self-compatible species can experience recurrent bottlenecks, even if the mean selfing rate is close to zero. This reduces genetic diversity and N_e and makes selection less efficient.

On the other hand, dioecious populations require both males and females to be present in sufficient density in a location in order to reproduce. Therefore, dioecious population sizes cannot decline below a given threshold at risk of declining to extinction, a phenomenon called the Allee effect (Courchamp et al. 2008). A counter-intuitive prediction is thus that, *conditioning on survival*, the Allee effect promotes genetic diversity in dioecious populations, especially during colonization (Roques et al. 2012; Romiguier et al. 2014). *Silene* species are known to have colonized Europe northward after the last ice age and have experienced population expansion. The Allee effect could have played a strong role during this expansion with self-compatible hermaphrodites colonizing new habitats fast with as few as a single individual and undergoing strong bottlenecks and diversity loss. Dioecious species that survived, on the other hand, would have expanded slowly, maintaining sufficiently dense populations throughout the expansion, resulting in the maintenance of genetic diversity. In established dioecious species, the Allee effect could provide a long-term safeguard against genetic depletion.

Our results suggest that established dioecious *Silene* species are genetically healthy. Our work revisits the fascinating question of why dioecy is rare in angiosperms, since the dead-end hypothesis does not seem to apply, at least to our sample of species. However, our dataset might suffer from a “survivor bias” and we cannot exclude that potential handicaps of dioecy could lead to very early extinction during the emergence of new dioecious populations from hermaphrodite lineages. At a macro-evolutionary scale, this rapid dead-end would lead to an observed low rate of transition towards dioecy. However, current estimates do not show low transition rates to dioecy in angiosperms (Goldberg et al. 2017), which goes against the possibility of an early dead-end of dioecy.

An alternative hypothesis to explain the rarity of dioecy is that rates of transition to dioecy could be low in angiosperms and/or dioecy could be more evolutionary labile than previously thought and easily lost and reversed to non-dioecious states depending on environmental conditions (Goldberg et al. 2017; Kafer et al. 2017). The combination of a low transition rate to dioecy and a high reversibility to non-dioecy together would lead to a low probability to observe dioecious angiosperms in nature, as they would only be transiently existing. This would make dioecy rare without the need for it to be an evolutionary dead-end. Larger and fully resolved angiosperm phylogenies with annotated breeding systems will be required to precisely estimate more precisely transition and reversion rates to and from dioecy.

Materials and methods

Sampling and sequencing

We investigated patterns of polymorphism and divergence of three sexual systems (hermaphroditism, gynodioecy and dioecy) in two subgenera of the *Silene* genus, subgenus *Behenantha* and subgenus *Silene* (sample sizes and sexes are indicated in Fig. 1). For subgenus *Behenantha*, we sampled five dioecious species (*S. latifolia*, *S. diclinis*, *S. dioica*, *S. marizii* and *S. heuffelli*), one gynodioecious species (*S. vulgaris*), and one hermaphroditic species (*S. viscosa*). For the subgenus *Silene*, we sampled two dioecious species (*S. otites* and *S. pseudotites*), two gynodioecious sub-species (*S. nutans* W - Western lineage, and *S. nutans* E - Eastern lineage, as defined in Martin et al. (2016)) and one hermaphroditic species (*S. paradoxa*). Two samples of *Dianthus chinensis* were added to root the phylogeny. Some of the individuals were already used in previous studies (S1).

The two known dioecious sections in the genus *Silene* were sampled: section *Melandrium* (all five species of the section were selected) and section *Otites* (2 closely related dioecious species were selected). The selection of the non-dioecious species was done using different criteria. First, close relatives of the two dioecious sections were selected based on available phylogenies (Desfeux et al. 1996; Jenkins and Keller 2010; Slancarova et al. 2013). We focused on those for which the sexual system is well described. The *Silene* genus is a large genus comprising ~700 species but only ~60 have a well-described sexual system (dioecious species included, Jürgens et al. 2002). Highly selfing species were avoided by focusing on hermaphroditic species that are perennial, insect pollinated (or at least not autogamous/cleistogamous) and with a high pollen/ovule ratio (Jürgens et al. 2002). These criteria left us with only a handful of hermaphroditic species, of which *S. viscosa* and *S. paradoxa* were selected for seed availability.

Seeds were sampled across the geographic distribution of each species (Fig. S5) and were sown and grown in controlled greenhouse conditions. In case plants did not flower, young leaves were sampled instead (details on individual names, sex, origin and sampled tissue can be found in Table S1). Total RNA was extracted through the Spectrum Plant Total RNA kit (Sigma, Inc., USA) following the manufacturer's protocol and treated with DNase. Libraries were prepared with the TruSeq RNA sample Preparation v2 kit (Illumina Inc., USA). Each 2 nM cDNA library was sequenced using a paired-end protocol on an Illumina HiSeq2000 sequencer, producing 100 bp reads. Demultiplexing was performed using CASAVA 1.8.1 (Illumina) to produce paired-sequence files containing reads for each sample in Illumina FASTQ format. RNA extraction and sequencing were done by the sequencing platform in the AGAP laboratory, Montpellier, France (<http://umr-agap.cirad.fr/>). The *Melandrium* and *Dianthus* samples were sequenced as described in Zemp et al. (2016: 2016) at the Quantitative Genomics Facility (QGF, ETH Zürich, Switzerland) and by GATC Biotech AG (Konstanz, Germany).

Reference transcriptome assemblies

Reference transcriptomes were assembled *de novo* for all species (except for *S. marizii*, *S. heuffelli*, *S. diclinis* and *S. dioica*). As assemblies were carried in different labs, they differ slightly for the two *Silene* taxa. For the subgenus *Behenantha*, reads from all individuals were pooled for each species. Steps in the assembly are represented in Fig. S3. 100% identical reads, assumed to be PCR duplicates, were filtered using the ConDeTri trimming software (Smeds and Künstner 2011). Illumina reads were then filtered for sequencing adapters and low read quality using the ea-utils FASTQ processing utilities (Aronesty 2011). Filtered reads were assembled using trinity with default settings (Haas et al. 2013). Poly-A tails were removed using PRINSEQ (Schmieder and Edwards 2011) with parameters -trim_tail_left 5 -trim_tail_right 5. rRNA-like sequences were removed using riboPicker version 0.4.3 (Schmieder et al. 2012) with parameters -i 90 -c 50 -l 50 and the following databases: SILVA Large subunit reference database, SILVA Small subunit reference database, the GreenGenes database and the Rfam database. Obtained transcripts were further assembled, inside of trinity components, using the program Cap3 (Huang and Madan 1999) with parameter -p 90 and custom perl scripts.

Raw data of the five taxa of subgenus *Silene* were filtered for sequencing adapters using Cutadapt (Martin 2011), low read quality and poly-A tails with PRINSEQ (Schmieder and Edwards 2011). For each focal species of the subgenus, reference transcriptomes were assembled *de novo* by combining filtered reads from two individuals from the same population (the two individuals with the highest number of reads) using default settings of Trinity (Haas et al. 2013) and an additional Cap3 (Huang and Madan 1999: 3) run. Finally, in both subgenera, coding sequences were predicted using TransDecoder (Haas et al. 2013) with PFAM domain searches as ORF retention criteria (see website: <http://transdecoder.github.io>). Assembly statistics are presented in Table S2.

Mapping, genotyping and alignment

Reads from each species were mapped onto their respective reference transcriptome using BWA (Li and Durbin 2009) version 0.7.12 with parameter -n 5 for the subgenus *Behenantha* and Bowtie2 (Langmead and Salzberg 2012) for the subgenus *Silene*. Mapping statistics can be found in Table S3. For the *Melandrium* section (*S. latifolia*, *S. marizii*, *S. heuffelli*, *S. diclinis* and *S. dioica*), all reads were mapped to the *S. latifolia* reference transcriptome. SAM files were compressed and sorted using SAMtools (Li et al. 2009). Individuals were genotyped using reads2snp_2.0 (Tsagkogeorga et al. 2012; Gayral et al. 2013) with parameters -par 1 -min 3 (10 for the subgenus *Silene*) -aeb -bqt 20 -rqt 10 that is to say: stringent filtering of paralogous positions, minimum coverage of 3 (or 10) reads for calling a genotype, accounting for allelic expression bias, minimum

read mapping quality threshold of 10 and minimum base quality threshold of 20. The output obtained was two unphased sequences for each individual and each ORF.

Orthologous ORFs were identified using OrthoMCL (Li et al. 2003) Basic Protocol 2 with default parameters. Best reciprocal blast hits were considered orthologs among species and orthologous ORFs were aligned for all individuals using MACSE (Ranwez et al. 2011). Species from the subgenus *Behenantha* were aligned separately with *S. paradoxa*, considered as an outgroup for this section and species from the subgenus *Silene* were aligned with *S. viscosa* considered as an outgroup for this section.

Phylogenetic reconstruction and phylogenetic signal

Orthologous ORFs among species were obtained using best reciprocal blast hits among the 14 species (all species from the *Silene* genus and *Dianthus chinensis*). For each orthologous ORF, the nine species were added one by one into the alignment using MACSE (Ranwez et al. 2011). The data were filtered so that a maximum of 20% missing data per site and per species was allowed, for a total of 55,337 variable nucleotides across 1,294 ORFs. We used the maximum number of ORFs available and did not restrict our analysis to autosomal ORFs identified in dioecious species. The phylogenetic tree was reconstructed using the GTRGAMMA model in RaxML (Stamatakis 2006). Bootstrap support values were generated using 100 replications.

Phylogenetic signal and its significance were computed using the R function *phylosig* in the *phytools* R package (Revell 2012).

Data filtering

Analyses were carried out on autosomal genes only, in order to avoid potential confounding effects of sex chromosomes and sexual systems on estimates of selection efficiency. Autosomal genes were predicted for each dioecious species separately by SEX-DETECTOR (Muyle et al. 2016) thanks to the study of a cross (parents and progeny) sequenced by RNA-seq (Martin et al. (2019) and Muyle et al. *in prep*). 11,951 contigs were inferred as autosomal in *S. latifolia*, 7,695 for *S. diclinis*, 11,675 for *S. heuffelli*, 8,997 for *S. marizii*, 10,698 for *S. dioica*, 2,035 in *S. otites* and 6,234 in *S. pseudotites*.

The data were further filtered in order to remove ORFs with frameshifts. Only biallelic sites were kept for analyses (sites with two or less alleles). Only ORFs longer than 100 codons and with a maximum of 20% of missing individuals of the focal species in the alignment were retained, after removing incomplete codons and gaps. We obtained final datasets ranging from 893 retained ORFs for *S. otites* up to 6,413 retained ORFs for *S. viscosa* E (see Table 1). Finally, to verify that observed trends are not due to differences in studied ORFs among species, we performed the same

analyses on a smaller set of orthologous ORFs present in all species within a subgenus, so that parameters were estimated on the same set of ORFs. For these tests, 188 orthologous ORFs were analysed for the subgenus *Silene*, and 298 for the subgenus *Behenantha*.

Polymorphism, heterozygosity and F_{IT}

Population statistics were computed on the filtered dataset for each ORF and averaged using the PopPhyl tool dNdSpiNpiS_1.0 (Gayral et al. 2013) with parameter kappa=2. The following statistics were computed for each contig: the mean fixation index F_{IT} (Weir and Cockerham 1984), per-individual heterozygosity (proportion of heterozygous positions), per-site synonymous polymorphism (π_s), per-site non-synonymous polymorphism (π_N) and the ratio of non-synonymous to synonymous polymorphism (π_N/π_s). Note that we could not compute F_{IS} or F_{ST} as we do not have the required 2 individuals / population. Confidence intervals were obtained by 1,000 bootstraps on contigs.

Estimation of geographic distribution and population size

Geographic distributions of the *Silene* species were estimated by counting the black dots (native occurrence) on the Atlas maps by Jalas and Suominen (1986). This was partly achieved using the Dot Count program (<http://reuter.mit.edu/software/dotcount>). Since no map was available for *S. pseudotites*, dots were counted for Italy and adjacent regions of France and the Balkans using the *S. vulgaris* map (the Atlas indicates these are the regions where this plant is abundant).

Census population sizes of the *Silene* species were estimated using raw species occurrence numbers from GBIF (<http://www.gbif.org> 17th January 2017 Download of *S. dioica*: <http://doi.org/10.15468/dl.fim8f2>, *S. marizii*: <http://doi.org/10.15468/dl.yywbrk>, *S. diclinis*: <http://doi.org/10.15468/dl.vizuiv> and 1st December 2016 Download of *S. pseudotites*: <http://doi.org/10.15468/dl.yrvenq>, *S. otites*: <http://doi.org/10.15468/dl.fxooxf>, *S. paradoxa*: <http://doi.org/10.15468/dl.wizsb6>, *S. nutans*: <http://doi.org/10.15468/dl.poqxc1>, *S. viscosa*: <http://doi.org/10.15468/dl.bwdu9z>, *S. vulgaris*: <http://doi.org/10.15468/dl.vmvebr>, *S. latifolia*: <http://doi.org/10.15468/dl.mfvvds>). Some countries contribute poorly to this database, however, we assumed that this affected all *Silene* species in the same way. The GBIF database was filtered for showing occurrence based on human observation and literature. Only data from European countries and Russia, which represent the native range of those species, were kept. For each species, all existing names of the species were included.

Both estimates (geographic range and population size) were divided by two when two subspecies were identified with our phylogeny (i.e. for *S. viscosa* and *S. nutans*). Atlas geographic

ranges and GBIF species abundance were compared among dioecious and non dioecious species using a one-sided Wilcoxon rank sum test in R (R Core Team 2016). Geographic range and GBIF data were correlated to π_S and sexual systems or F_{IT} using R linear model function `lm` as follows:

`lm(piS ~ Atlas + sexual system)`

`lm(piS ~ GBIF + sexual system)`

`lm(piS ~ Atlas + F_{IT})`

`lm(piS ~ GBIF + F_{IT})`

The significance of the effect of sexual system and F_{IT} were tested thanks to a comparison to a nested model using the `anova` function in R. The nested models are:

`lm(piS ~ Atlas)`

`lm(piS ~ GBIF)`

The same models were used and corrected for phylogenetic relationships among species using *gls* models with Martins and Hansen's correlation structure (`corMartins`) in the R package *ape* (Paradis and Schliep 2019).

Selection efficacy

The unfolded synonymous and non-synonymous site frequency spectra (SFS, the distribution of derived allele counts across SNPs) were computed for each species with the `sfs` program provided by Nicolas Galtier (Galtier 2016). These SFS were used to estimate α (the proportion of adaptive amino-acid substitutions) in each species. We estimated α according to a modified version of the Eyre-Walker and Keightley method (Keightley and Eyre-Walker 2007), which relies on polymorphism data only to estimate α (Tataru et al. 2017). `polyDFE` (<https://github.com/paula-tataru/polyDFE>) was used to estimate the distribution of fitness effects of mutations (DFEM), α , the (per synonymous substitution) rate of adaptive non-synonymous substitution $\omega_A = \alpha d_N/d_S$ and the rate of non-adaptive non-synonymous substitutions $\omega_{NA} = (1-\alpha) d_N/d_S$. For each species separately, four different models were fitted to the SFS:

- Model 1: full DFEM with adaptive mutations ($p_b > 0$), estimation of an orientation error rate for ancestral alleles in the SFS (`eps_an`), estimation of demography ($r_i \neq 1$) for a total

of 6 parameters plus the number of demography parameters r_i which depends on the sample size of each species.

- Model 2: same as Model 1 but without ancestral allele orientation error rate ($\text{eps_an}=0$), for a total of 5 parameters plus the number of demography parameters r_i .
- Model 3: same as Model 1 but without adaptive mutations ($p_b=0$), for a total of 5 parameters plus the number of demography parameters r_i .
- Model 4: same as Model 1 but without adaptive mutations ($p_b=0$) and without ancestral allele orientation error rate ($\text{eps_an}=0$), for a total of 4 parameters plus the number of demography parameters r_i .

Each model provided estimates for α , ω_A and ω_{NA} . These values were averaged using Akaike weights as follows:

$$\alpha_{\text{average}} = \frac{\sum_{m=1}^4 \alpha_m e^{-1/2\Delta\text{AIC}_m}}{\sum_{k=1}^4 e^{-1/2\Delta\text{AIC}_k}}$$

where m stands for model number m , α_m stands for the estimated α using model m , and

$$\Delta\text{AIC}_m = \text{AIC}_m - \text{AIC}_{\min}$$

with AIC the Akaike criterion:

$$\text{AIC}_m = 2n_m - 2\log(L_m)$$

Where n_m stands for the number of parameters and L_m for the likelihood of model m . Model averaged values for ω_A and ω_{NA} were obtained using a similar procedure.

For each species, 500 bootstrapped datasets were randomly generated by random sampling SNPs with replacement from the original SFS. Average values among models were computed as explained previously for each bootstrap. The distribution of values for the 500 bootstraps was used to compute approximate 95% confidence intervals for α , ω_A and ω_{NA} . For this, 2.5% of values were removed from each side of the bootstrap distribution (Efron and Tibshirani 1993). α , ω_A and ω_{NA} values were analysed in a linear model using R `lm` function (R Core Team 2016). The model tested the effect of π_s and sexual system (dioecious, gynodioecious, hermaphroditic) as fixed effect, while taking into account the variance of the estimations with weights computed as follow:

`lm($\alpha \sim \pi_s$ + sexual system, weights)`

`lm($\omega_A \sim \pi_s$ + sexual system, weights)`

`lm($\omega_{NA} \sim \pi_s$ + sexual system, weights)`

The same models (without weights) were used and corrected for phylogenetic relationships among species using *gls* models with Martins and Hansen's correlation structure (`corMartins`) in the R package *ape* (Paradis and Schliep 2019).

Acknowledgements

We would like to thank Jos Kafer for his help with fieldwork and a number of colleagues for sending us seeds: J. J. Wieringa, Deborah Charlesworth, Bohuslav Janousek, Tatiana Giraud, Honor C. Prentice, Salvatore Cozzolino, Delphine Griveau, Henk Schat, Cristina Gonnelli, Alessia Guggisberger, Sophie Karrenberg, Adrien Favre, Maria Domenica Moccia, Aria Minder, Carmen Rothenbuehler, Massimiliano Gnesotto and Kerstin Kustas. AM, RT and GABM thank Elise Lacroix technician of the Univ. Lyon 1 greenhouse and HM, SGa and PT thank Eric Schmitt for taking care of the *Silene* plants. We are grateful to Sylvain Santoni and the AGAP platform staff for the RNA extraction and RNA-seq. The numerical results presented in this paper were carried out using the computing cluster of Laboratoire Biométrie et Biologie-Pôle Rhone-Alpin de Bioinformatique (LBBE-PRABI) at Univ. Lyon 1 and the HPC service of the Centre de Ressources Informatiques (CRI) of Univ. Lille 1, for which we thank Stéphane Delmotte and Bruno Spataro. We thank Nicolas Galtier for help with his program sfs. HM, SGa and PT are grateful to Jonathan Aceituno for technical support in R and to Céline Poux for discussion on phylogenetic inference. We are thankful to Brandon Gaut for commenting on the manuscript. This work was supported by the Agence Nationale de la Recherche (ANR-11-BSV7-013-03, TRANS) to SGI, GM, and PT, the Swiss National Science Foundation (31003A-182675), and a PhD fellowship from the French Research Ministry to HM.

Data availability

All the RNA-seq reads generated for this study can be downloaded from the European Nucleotide Archive Accession PRJEB39526.

References

- Aronesty E. 2011. ea-utils: “Command-line tools for processing biological sequencing data.” Available from: <http://code.google.com/p/ea-utils>
- Barrett SCH. 2002. The evolution of plant sexual diversity. *Nat. Rev. Genet.* 3:274–284.
- Bernasconi G, Antonovics J, Biere A, Charlesworth D, Delph LF, Filatov D, Giraud T, Hood ME, Marais G a. B, McCauley D, et al. 2009. *Silene* as a model system in ecology and evolution. *Heredity (Edinb)* 103:5–14.
- Brandvain Y, Slotte T, Hazzouri KM, Wright SI, Coop G. 2013. Genomic Identification of Founding Haplotypes Reveals the History of the Selfing Species *Capsella rubella*. Glemin S, editor. *PLoS Genetics* 9:e1003754.
- Cahais V, Gayral P, Tsagkogeorga G, Melo-Ferreira J, Ballenghien M, Weinert L, Chiari Y, Belkhir K, Ranwez V, Galtier N. 2012. Reference-free transcriptome assembly in non-model

- animals from next-generation sequencing data. *Mol Ecol Resour* 12:834–845.
- Charlesworth B, Charlesworth D. 1978. A Model for the Evolution of Dioecy and Gynodioecy. *The American Naturalist* 112:975–997.
- Courchamp F, Berec L, Gascoigne J. 2008. Allee Effects in Ecology and Conservation. New York: Oxford University Press Available from: <https://academic.oup.com/jmammal/article/91/6/1530/892517>
- Desfeux C, Maurice S, Henry JP, Lejeune B, Gouyon PH. 1996. Evolution of reproductive systems in the genus *Silene*. *Proc. Biol. Sci.* 263:409–414.
- Duan T, Deng X, Chen S, Luo Z, Zhao Z, Tu T, Khang NS, Razafimandimbison SG, Zhang D. 2018. Evolution of sexual systems and growth habit in *Mussaenda* (Rubiaceae): Insights into the evolutionary pathways of dioecy. *Mol. Phylogenet. Evol.* 123:113–122.
- Efron B, Tibshirani RJ. 1993. An Introduction to the Bootstrap. 1st ed. New York: Chapman and Hall/CRC
- Fruchard C, Marais G. 2017. The evolution of sex determination in plants. In: Evolutionary Developmental Biology: A Reference Guide. Springer. p. 1–14.
- Galtier N. 2016. Adaptive Protein Evolution in Animals and the Effective Population Size Hypothesis. *PLoS Genet.* 12:e1005774.
- Gautier M. 2014. Using genotyping data to assign markers to their chromosome type and to infer the sex of individuals: a Bayesian model-based classifier. *Mol Ecol Resour* 14:1141–1159.
- Gayral P, Melo-Ferreira J, Glémin S, Bierne N, Carneiro M, Nabholz B, Lourenco JM, Alves PC, Ballenghien M, Faivre N, et al. 2013. Reference-free population genomics from next-generation transcriptome data and the vertebrate-invertebrate gap. *PLoS Genet.* 9:e1003457.
- Giles B, Goudet G. 1997. Genetic differentiation in *Silene dioica* metapopulations: estimation of spatiotemporal effects in a successional plant species. *The American Naturalist* 149:507–526.
- Glémin S, François CM, Galtier N. 2019. Genome Evolution in Outcrossing vs. Selfing vs. Asexual Species. *Methods Mol. Biol.* 1910:331–369.
- Goldberg EE, Kohn JR, Lande R, Robertson KA, Smith SA, Igic B. 2010. Species Selection Maintains Self-Incompatibility. *Science* 330:493–495.
- Goldberg EE, Otto SP, Vamasi JC, Mayrose I, Sabath N, Ming R, Ashman T-L. 2017. Macroevolutionary synthesis of flowering plant sexual systems. *Evolution* 71:898–912.
- Grossenbacher D, Briscoe Runquist R, Goldberg EE, Brandvain Y. 2015. Geographic range size is predicted by plant mating system. *Ecol. Lett.* 18:706–713.
- Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, Couger MB, Eccles D, Li B, Lieber M, et al. 2013. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat Protoc* 8:1494–1512.
- Heilbut JC. 2000. Lower species richness in dioecious clades. *The American Naturalist* 156:221–241.

- Heilbuth JC, Ilves KL, Otto SP. 2001. The consequences of dioecy for seed dispersal: modeling the seed-shadow handicap. *Evolution* 55:880–888.
- Huang X, Madan A. 1999. CAP3: A DNA Sequence Assembly Program. *Genome Res.* 9:868–877.
- Jalas J, Suominen J, The Committee for Mapping the Flora of Europe. 1986. Distribution of Vascular Plants in Europe. 7. Caryophyllaceae (Silenoideae). In: *Atlas Florae Europaeae*.
- Jenkins C, Keller SR. 2010. A phylogenetic comparative study of preadaptation for invasiveness in the genus *Silene* (Caryophyllaceae). *Biol Invasions* 13:1471–1486.
- Jürgens A, Witt T, Gottsberger G. 2002. Pollen grain numbers, ovule numbers and pollen-ovule ratios in Caryophylloideae: correlation with breeding system, pollination, life form, style number, and sexual system | SpringerLink. *Sexual Plant Reproduction* 14:279–289.
- Kafer J, de Boer HJ, Mousset S, Kool A, Dufay M, Marais GAB. 2014. Dioecy is associated with higher diversification rates in flowering plants. *Journal of Evolutionary Biology* 27:1478–1490.
- Kafer J, Marais GAB, Pannell JR. 2017. On the rarity of dioecy in flowering plants. *Mol. Ecol.* 26:1225–1241.
- Kafer J, Mousset S. 2014. Standard sister clade comparison fails when testing derived character States. *Syst. Biol.* 63:601–609.
- Kafer J, Talianová M, Bigot T, Michu E, Guéguen L, Widmer A, žlůvová J, Glémin S, Marais GAB. 2013. Patterns of molecular evolution in dioecious and non-dioecious *Silene*. *Journal of Evolutionary Biology* 26:335–346.
- Kalinowski ST, Waples RS. 2002. Relationship of Effective to Census Size in Fluctuating Populations. *Conservation Biology* 16:129–136.
- Keightley PD, Eyre-Walker A. 2007. Joint inference of the distribution of fitness effects of deleterious mutations and population demography based on nucleotide polymorphism frequencies. *Genetics* 177:2251–2261.
- Krasovec M, Chester M, Ridout K, Filatov DA. 2018. The Mutation Rate and the Age of the Sex Chromosomes in *Silene latifolia*. *Curr. Biol.* 28:1832–1838.e4.
- Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9:357–359.
- Lauterbach D, Ristow M, Gemeinholzer B. 2011. Genetic population structure, fitness variation and the importance of population history in remnant populations of the endangered plant *Silene chlorantha* (Willd.) Ehrh. (Caryophyllaceae). *Plant Biology* 13:667–777.
- Lauterbach D, Ristow M, Gemeinholzer B. 2012. Population genetics and fitness in fragmented populations of the dioecious and endangered *Silene otites* (Caryophyllaceae). *Plant Systematics and Evolution* 298:155–164.
- Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25:1754–1760.
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R,

- 1000 Genome Project Data Processing Subgroup. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25:2078–2079.
- Li L, Stoeckert CJ, Roos DS. 2003. OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res.* 13:2178–2189.
- Maddison WP, Midford PE, Otto SP. 2007. Estimating a binary character's effect on speciation and extinction. *Syst. Biol.* 56:701–710.
- Martin H, Carpentier F, Gallina S, Godé C, Schmitt E, Muyle A, Marais GA, Touzet P. 2019. Evolution of young sex chromosomes in two dioecious sister plant species with distinct sex determination systems. *Genome Biol Evol.*
- Martin H, Touzet P, Dufay M, Godé C, Schmitt E, Lahiani E, Delph L, Van Rossum F. 2017. Lineages of *Silene nutans* developed rapid, strong, asymmetric postzygotic reproductive isolation in allopatry. *Evolution* 71:1519–1531.
- Martin H, Touzet P, Van Rossum F, Delalande D, Arnaud J-F. 2016. Phylogeographic pattern of range expansion provides evidence for cryptic species lineages in *Silene nutans* in Western Europe. *Heredity (Edinb)* 116:286–294.
- Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* 17:10–12.
- Moyle LC. 2006. Correlates of genetic differentiation and isolation by distance in 17 congeneric *Silene* species. *Mol. Ecol.* 15:1067–1081.
- Mugal CF, Wolf JBW, Kaj I. 2014. Why time matters: codon evolution and the temporal dynamics of dN/dS. *Mol. Biol. Evol.* 31:212–231.
- Muyle A, Käfer J, Zemp N, Mousset S, Picard F, Marais GA. 2016. SEX-DETECTOR: A Probabilistic Approach to Study Sex Chromosomes in Non-Model Organisms. *Genome Biol Evol* 8:2530–2543.
- Muyle A, Marais G. 2016. Genome Evolution and Mating Systems in Plants. In: Encyclopedia of Evolutionary Biology. Vol. 2. Kliman, R.M. (ed.). Oxford: Academic Press. p. 480–492.
- Muyle A, Zemp N, Deschamps C, Mousset S, Widmer A, Marais GAB. 2012. Rapid de novo evolution of X chromosome dosage compensation in *Silene latifolia*, a plant with young sex chromosomes. *PLoS Biol.* 10:e1001308.
- Nabholz B, Sarah G, Sabot F, Ruiz M, Adam H, Nidelet S, Ghesquière A, Santoni S, David J, Glémin S. 2014. Transcriptome population genomics reveals severe bottleneck and domestication cost in the African rice (*Oryza glaberrima*). *Mol. Ecol.* 23:2210–2227.
- Nazareno AG, Alzate-Marin AL, Pereira RA. 2013. Dioecy, more than monoecy, affects plant spatial genetic structure: the case study of *Ficus*. *Ecology and Evolution* 3:3495–3508.
- Nordborg M. 2000. Linkage disequilibrium, gene trees and selfing: an ancestral recombination graph with partial self-fertilization. *Genetics* 154:923–929.
- Oxelman B, Rautenberg A, Thollesson M, Larsson A, Frajman B, Eggens F, Petri A, Aydin Z, Töpel M, Brandtberg-Falkman A. 2013. *Sileneae* taxonomy and systematics. Available from: <http://www.sileneae.info/>

- Papadopoulos AST, Chester M, Ridout K, Filatov DA. 2015. Rapid Y degeneration and dosage compensation in plant sex chromosomes. *Proc. Natl. Acad. Sci. U.S.A.* 112:13021–13026.
- Paradis E, Schliep K. 2019. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* 35:526–528.
- R Core Team. 2016. R: A language and environment for statistical computing. *R Foundation for Statistical Computing, Vienna, Austria* [Internet]. Available from: <https://www.R-project.org>
- Randle AM, Slyder JB, Kalisz S. 2009. Can differences in autonomous selfing ability explain differences in range size among sister-taxa pairs of *Collinsia* (Plantaginaceae)? An extension of Baker's Law. *New Phytol.* 183:618–629.
- Ranwez V, Harispe S, Delsuc F, Douzery EJP. 2011. MACSE: Multiple Alignment of Coding SEquences accounting for frameshifts and stop codons. *PLoS ONE* 6:e22594.
- Rautenberg A, Hathaway L, Oxelman B, Prentice HC. 2010. Geographic and phylogenetic patterns in *Silene* section *Melandrium* (Caryophyllaceae) as inferred from chloroplast and nuclear DNA sequences. *Mol. Phylogenet. Evol.* 57:978–991.
- Renner SS. 2014. The relative and absolute frequencies of angiosperm sexual systems: dioecy, monoecy, gynodioecy, and an updated online database. *Am. J. Bot.* 101:1588–1596.
- Revell LJ. 2012. phytools: an R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution* 3:217–223.
- Romiguier J, Gayral P, Ballenghien M, Bernard A, Cahais V, Chenuil A, Chiari Y, Dernat R, Duret L, Faivre N, et al. 2014. Comparative population genomics in animals uncovers the determinants of genetic diversity. *Nature* 515:261–263.
- Roques L, Garnier J, Hamel F, Klein EK. 2012. Allee effect promotes diversity in traveling waves of colonization. *Proc. Natl. Acad. Sci. U.S.A.* 109:8828–8833.
- Rosche C, Schrieber K, Lachmuth S, Durka W, Hirsch H, Wagner V, Schleuning M, Hensen I. 2018. Sex ratio rather than population size affects genetic diversity in *Antennaria dioica*. *Plant Biology* 20:789–796.
- Sabath N, Goldberg EE, Glick L, Einhorn M, Ashman T-L, Ming R, Otto SP, Vamosi JC, Mayrose I. 2016. Dioecy does not consistently accelerate or slow lineage diversification across multiple genera of angiosperms. *New Phytol.* 209:1290–1300.
- Schmieder R, Edwards R. 2011. Quality control and preprocessing of metagenomic datasets. *Bioinformatics* 27:863–864.
- Schmieder R, Lim YW, Edwards R. 2012. Identification and removal of ribosomal RNA sequences from metatranscriptomes. *Bioinformatics* 28:433–435.
- Schoen DJ, Brown AH. 1991. Intraspecific variation in population gene diversity and effective population size correlates with the mating system in plants. *Proc. Natl. Acad. Sci. U.S.A.* 88:4494–4497.
- Slancarova V, Zdanska J, Janousek B, Talianova M, Zschach C, Zluvova J, Siroky J, Kovacova V, Blavet H, Danihelka J, et al. 2013. Evolution of sex determination systems with

- heterogametic males and females in silene. *Evolution* 67:3669–3677.
- Slotte T, Hazzouri KM, Ågren JA, Koenig D, Maumus F, Guo Y-L, Steige K, Platts AE, Escobar JS, Newman LK, et al. 2013. The *Capsella rubella* genome and the genomic consequences of rapid mating system evolution. *Nature Genetics* 45:831–835.
- Smeds L, Künstner A. 2011. ConDeTri - A Content Dependent Read Trimmer for Illumina Data. *PLoS ONE* 6:e26314.
- Stamatakis A. 2006. RAxML-VI-HPG: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22:2688–2690.
- Tataru P, Bataillon T. 2020. polyDFE: Inferring the Distribution of Fitness Effects and Properties of Beneficial Mutations from Polymorphism Data. *Methods Mol. Biol.* 2090:125–146.
- Tataru P, Mollion M, Glémin S, Bataillon T. 2017. Inference of Distribution of Fitness Effects and Proportion of Adaptive Substitutions from Polymorphism Data. *Genetics* 207:1103–1119.
- Tsagkogeorga G, Cahais V, Galtier N. 2012. The population genomics of a fast evolver: high levels of diversity, functional constraint, and molecular adaptation in the tunicate *Ciona intestinalis*. *Genome Biol Evol* 4:740–749.
- Vamosi JC, Otto SP. 2002. When looks can kill: the evolution of sexually dimorphic floral display and the extinction of dioecious plants. *Proceedings of the Royal Society B: Biological Sciences* 269:1187–1194.
- Vamosi JC, Otto SP, Barrett SCH. 2003. Phylogenetic analysis of the ecological correlates of dioecy in angiosperms. *J. Evol. Biol.* 16:1006–1018.
- de Vos JM, Hughes CE, Schneeweiss GM, Moore BR, Conti E. 2014. Heterostyly accelerates diversification via reduced extinction in primroses. *Proc. Biol. Sci.* 281:20140075.
- Weir BS, Cockerham CC. 1984. Estimating F-statistics for the analysis of population structure. *Evolution* 38:1358–1370.
- Wright S. 1938. Size of a population and breeding structure in relation to evolution. *Science* 87:430–431.
- Zemp N, Tavares R, Muyle A, Charlesworth D, Marais GAB, Widmer A. 2016. Evolution of sex-biased gene expression in a dioecious plant. *Nat Plants* 2:16168.
- Zenil-Ferguson R, Burleigh JG, Freyman WA, Igić B, Mayrose I, Goldberg EE. 2019. Interaction among ploidy, breeding system and lineage diversification. *New Phytol.* 224:1252–1265.

Table 1

Sexual system	Focal species	Number of samples	Number of clean contigs	Number of snps	Weir & Cockerham F_{IT}	π_S in focal species (%)	π_N in focal species (%)	π_N/π_S
Section <i>Benantha</i>								
dioecious	<i>S. diclinis</i>	3	2729	23774	-0.089 [-0.113;-0.065]	0.934 [0.886;0.982]	0.179 [0.170;0.187]	0.191 [0.182;0.200]
	<i>S. dioica</i>	17	3730	143297	0.181 [0.176;0.186]	2.073 [2.013;2.135]	0.242 [0.234;0.251]	0.117 [0.113;0.121]
	<i>S. heuffelli</i>	4	3999	98957	-0.043 [-0.051;-0.035]	2.698 [2.628;2.771]	0.342 [0.331;0.353]	0.127 [0.123;0.131]
	<i>S. latifolia</i>	34	4132	214646	0.165 [0.161;0.169]	2.005 [1.950;2.062]	0.244 [0.236;0.252]	0.122 [0.118;0.126]
	<i>S. marizii</i>	4	2878	44868	0.013 [-0.005;0.030]	1.637 [1.573;1.703]	0.246 [0.235;0.256]	0.150 [0.144;0.156]
gynodioecious	<i>S. vulgaris</i>	14	6248	281300	0.205 [0.200;0.209]	2.685 [2.639;2.732]	0.351 [0.343;0.360]	0.131 [0.128;0.134]
hermaphrodite	<i>S. viscosa</i> E	3	6413	13712	0.723 [0.691;0.754]	0.291 [0.277;0.305]	0.056 [0.053;0.059]	0.194 [0.184;0.204]
	<i>S. viscosa</i> W	2	5959	24305	0.068 [0.038;0.098]	0.723 [0.693;0.757]	0.134 [0.127;0.142]	0.186 [0.178;0.193]
Section <i>Otites</i>								
dioecious	<i>S. pseudotites</i>	4	2339	24698	0.024 [0.006;0.041]	1.116 [1.074;1.157]	0.195 [0.185;0.204]	0.174 [0.166;0.183]
	<i>S. otites</i>	8	893	11364	0.014 [-0.007;0.035]	1.081 [1.016;1.150]	0.180 [0.167;0.194]	0.167 [0.154;0.180]
gynodioecious	<i>S. nutans</i> E	11	5100	65820	0.124 [0.106;0.140]	0.861 [0.829;0.895]	0.130 [0.125;0.136]	0.151 [0.146;0.157]
	<i>S. nutans</i> W	11	5059	88853	0.201 [0.182;0.219]	1.048 [1.011;1.084]	0.170 [0.165;0.176]	0.163 [0.157;0.169]
hermaphrodite	<i>S. paradoxa</i>	12	5955	52675	0.116 [0.108;0.124]	0.589 [0.568;0.611]	0.102 [0.098;0.105]	0.172 [0.166;0.179]