

# Demography and mating system shape the genome-wide impact of purifying selection in *Arabis alpina*

Benjamin Laenen<sup>1</sup>, Andrew Tedder<sup>1</sup>, Michael Nowak<sup>2</sup>, Per Toräng<sup>3</sup>, Jörg Wunder<sup>4</sup>, Stefan Wötzel<sup>4</sup>, Kim Steige<sup>5</sup>, Yiannis Kourmpetis<sup>6</sup>, Thomas Odong<sup>6</sup>, Andreas Drouzas<sup>7</sup>, Marco Bink<sup>6</sup>, Jon Ågren<sup>3</sup>, George Coupland<sup>4</sup>, Tanja Slotte<sup>1</sup>

<sup>1</sup>Stockholm University, <sup>2</sup>University of Oslo, <sup>3</sup>Uppsala University, <sup>4</sup>Max Planck Institute for Plant Breeding Research, <sup>5</sup>Universität zu Köln, <sup>6</sup>Wageningen UR, <sup>7</sup>Aristotle University of Thessaloniki

Submitted to Proceedings of the National Academy of Sciences of the United States of America

**Plant mating systems have profound effects on levels and structuring of genetic variation, and can affect the impact of natural selection. While theory predicts that intermediate outcrossing rates may allow plants to prevent accumulation of deleterious alleles, few studies have empirically tested this prediction using genomic data. Here, we study the effect of mating system on purifying selection by conducting population genomic analyses on whole-genome resequencing data from 38 European individuals of the arctic-alpine crucifer *Arabis alpina*. We find that outcrossing and mixed-mating populations maintain genetic diversity at similar levels, whereas highly self-fertilizing Scandinavian *A. alpina* show a strong reduction in genetic diversity, most likely as a result of a postglacial colonization bottleneck. We further find evidence for accumulation of genetic load in highly self-fertilizing populations, whereas the genome-wide impact of purifying selection does not differ greatly between mixed-mating and outcrossing populations. Our results demonstrate that intermediate levels of outcrossing may allow efficient selection against harmful alleles whereas demographic effects can be important for relaxed purifying selection in highly selfing populations. Thus, both mating system and demography shape the impact of purifying selection on genomic variation in *A. alpina*. These results are important for an improved understanding of the evolutionary consequences of mating system variation and the maintenance of mixed-mating strategies.**

self-fertilization | demographic history | bottleneck | fitness effects | genetic load

## Introduction

Flowering plants show a great deal of variation in their reproductive modes, and variation in outcrossing rate is particularly common. While approximately 50% of flowering plants are predominantly outcrossing, a substantial proportion (35-40%) undergo intermediate levels of outcrossing, whereas only 10-15% are predominantly self-fertilizing (selfing) (1, 2). Whether mixed-mating is evolutionarily stable or represents a transitional stage has long been debated (1, 3). Classic population genetic models predict that only high selfing and high outcrossing rates are evolutionarily stable strategies (4). However, mixed-mating can be stable in ecologically more realistic models, such as those that account for reduced outcross pollen success with increased selfing (1). Moreover, population genetic models that incorporate linkage indicate that mixed-mating populations may avoid the reduced efficacy of selection associated with high selfing rates (5-7).

While selfing can be favored due to its genetic transmission advantage, and because it can confer reproductive assurance, it also has marked population genetic consequences that might contribute to the long-term demise of highly selfing lineages (6). For instance, highly selfing populations are expected to have a reduced effective population size ( $N_e$ ), due to the direct effect of inbreeding (8, 9). Demographic processes such as frequent extinc-

tion and recolonization of local subpopulations (10) or founder events associated with the shift to selfing (6) can reduce genetic variation in selfers to an even greater extent. Self-fertilization is also expected to lead to elevated linkage disequilibrium, which means that background selection and other forms of linked selection can reduce genetic variation genome-wide, further reducing  $N_e$  in selfers (7, 11, 12).

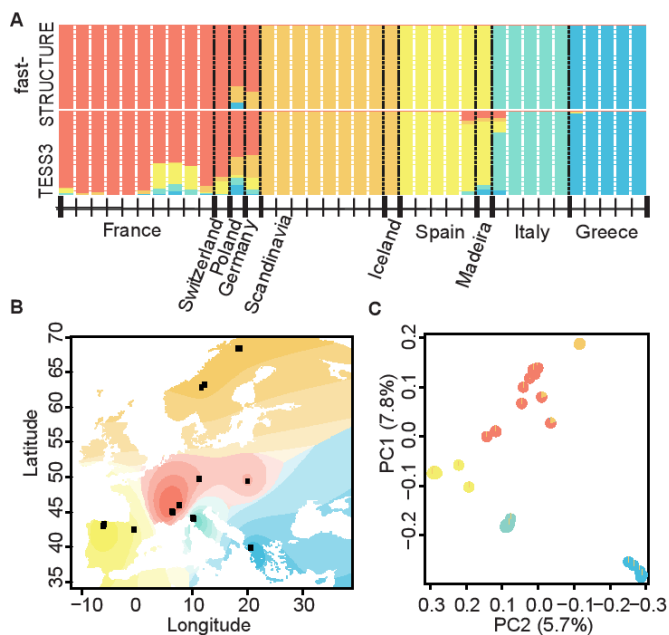
Because the strength of selection scales with  $N_e$  (13), natural selection is expected to be less efficient in selfers than in outcrossers. Selfers are therefore expected to accumulate weakly deleterious mutations at a higher rate than outcrossers (14). However, low levels of outcrossing may be sufficient to prevent accumulation of mildly deleterious alleles (15). Furthermore, selfing also increases homozygosity, exposing recessive mutations to selection. This is expected to result in purging of recessive deleterious mutations, unless selfing is associated with a strong reduction in  $N_e$  which renders such purging ineffective (16-18). Theory predicts that purging should also be efficient in mixed-mating populations (17).

While there is accumulating evidence for relaxed selection on weakly deleterious mutations in highly selfing lineages (5, 19-23), several fundamental questions on the evolutionary genomic consequences of mating systems remain unanswered. Specifically, it is unclear whether partial selfing generally results in purging of recessive deleterious alleles or whether this effect is possibly

## Significance

Intermediate outcrossing rates are theoretically predicted to maintain effective selection against harmful alleles, but few studies have empirically tested this prediction using genomic data. We used whole-genome resequencing data from alpine rock-crests to study how genetic variation and purifying selection vary with mating system. We find that populations with intermediate outcrossing rates have similar levels of genetic diversity as outcrossing populations, and that purifying selection against harmful alleles is efficient in mixed-mating populations. In contrast, self-fertilizing populations from Scandinavia have strongly reduced genetic diversity, and accumulate harmful mutations, likely as a result of demographic effects of postglacial colonization. Our results suggest that mixed-mating populations can avoid some of the negative evolutionary consequences of high self-fertilization rates.

## Reserved for Publication Footnotes



**Fig. 1.** Bayesian clustering analysis supports a strong geographic component to population structure in European *A. alpina*. A. Ancestry proportions for K=5 correspond closely to geographical sampling locations. B. Geographic interpolation of genetic structure across Europe based on TESS3 results for K=5. C. PCA of genetic data, with pie charts showing ancestry proportions for K=5 and colors indicating geographic origin.

**Table 1. Population genetic summary statistics for regional populations.**

| Population <sup>1</sup> | $S_s^2$ | $\pi_s^3$ | $\pi_N/\pi_s^4$ |
|-------------------------|---------|-----------|-----------------|
| Greece (5)              | 30920   | 0.0080    | 0.2740          |
| Italy (5)               | 17971   | 0.0054    | 0.2820          |
| Spain (5)               | 15852   | 0.0046    | 0.2911          |
| France (10)             | 19661   | 0.0052    | 0.2764          |
| Scandinavia (8)         | 589     | 0.0002    | 0.3936          |

<sup>1</sup>Regional population (n)

<sup>2</sup>Segregating four-fold synonymous SNPs.

<sup>3</sup>Mean four-fold synonymous nucleotide diversity.

<sup>4</sup>Mean 0-fold degenerate nonsynonymous/4-fold degenerate synonymous polymorphism.

overridden by demographic effects, and there are few empirical genome-wide studies that have explicitly examined the selective consequences of mixed-mating in plants (but see (24)). It is therefore unclear whether mixed-mating plants avoid the negative genetic effects associated with high selfing rates.

The broadly distributed, arctic-alpine perennial herb *Arabis alpina* (Brassicaceae) is a promising plant system in which to address the impact of variation in outcrossing rates on genome-wide genetic variation and efficiency of selection. This species harbors populations that express a range of mating strategies from self-incompatible outcrossing (25) through mixed-mating to autonomous selfing (25-28). The colonization history of *A. alpina* has already begun to be characterized (26, 29-31), which facilitates interpretation of global patterns of polymorphism. Finally, the availability of a genome assembly of *A. alpina* (32) greatly facilitates population genomic studies.

In this study, we investigate the effects of mating system and demography on the efficacy of selection in *A. alpina* by population genomic analyses of whole-genome resequencing data from out-

crossing, mixed-mating and highly selfing populations. We first investigate population structure and test whether populations with higher selfing rates have lower levels of genetic diversity, and then test whether higher selfing rates are associated with relaxed selection against weakly deleterious mutations and purging of strongly deleterious mutations. To do this, we use genome-wide allele frequency distributions at nonsynonymous and synonymous sites to estimate the fraction of weakly deleterious and strongly harmful new nonsynonymous mutations. We further test whether higher rates of selfing are associated with an increase in the frequency of derived alleles with major effects on gene integrity, which would suggest relaxed purifying selection. Finally, we compare two genomic proxies of genetic load, the reduction in mean fitness of a population caused by deleterious variation (33,34), among outcrossing, mixed-mating and highly self-fertilizing populations. Our results are important for an improved understanding of the population genetic consequences of mating system variation.

## Results

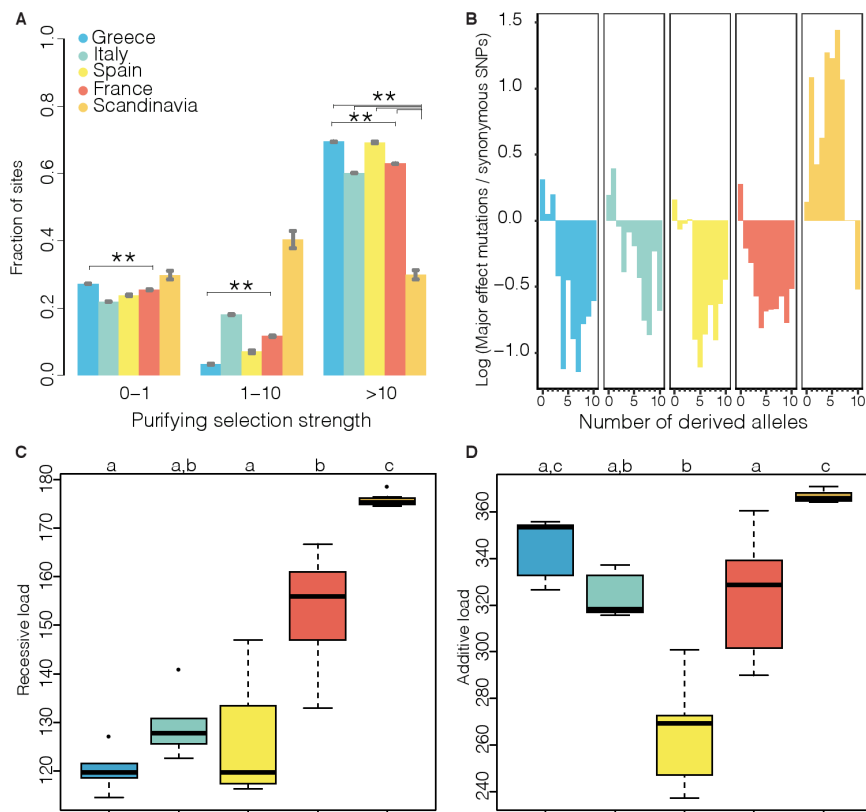
### Sequencing and single nucleotide polymorphism

We sampled 38 *A. alpina* individuals from 17 geographical sites, with targeted sampling of 2-5 individuals from 12 geographical sites harboring self-incompatible outcrossing populations (Greece and Italy), populations with intermediate outcrossing rates (France and Spain), and highly selfing populations (Scandinavia), in addition to a sample of single individuals from 5 additional geographic locations across the European range of *A. alpina* (See SI Appendix, Table S1). Progeny-array based outcrossing estimates have shown that populations from Scandinavia are highly selfing (up to ~10% outcrossing) whereas intermediate outcrossing rates have been estimated for French and Spanish populations (~19% and ~18%, respectively) (28; See SI Appendix, SI Text). We further verified that Greek and Italian individuals produced no offspring after forced self-pollination in the greenhouse. The distribution of genomic runs of homozygosity (ROHs) and decay of linkage disequilibrium supported mating system variation among the populations studied here (See SI Appendix, Fig. S1-S2).

Each individual was resequenced to high coverage (average 26X, range 16-45X) using Illumina short-read technology. We called single nucleotide polymorphisms (SNPs) and applied stringent filtering criteria to identify a total of 1,514,615 high-quality SNPs, of which 98,564 were 0-fold degenerate nonsynonymous (i.e. sites at which any mutation will result in a nonsynonymous change) with a mean nucleotide diversity ( $\pi$ ) of 0.0027 and 65,821 were 4-fold degenerate synonymous (i.e. synonymous sites at which any mutation will result in a synonymous change) with a mean nucleotide diversity of 0.0102 (See SI Appendix, Table S2).

### Population structure has a strong geographic component

We analyzed population structure with two model-based Bayesian clustering approaches, implemented in the software fastSTRUCTURE (35) and TESS3 (36), based on 25,505 4-fold synonymous SNPs (see Methods). Both methods, as well as principal component analysis gave very similar results, and supported the presence of five clusters (Fig. 1) that were substantially differentiated (average pairwise  $F_{ST}$  0.56, range 0.39-0.82; See SI Appendix, Table S3), in good agreement with previous analyses of population structure in *A. alpina* (30). These clusters correspond to a central European population of mixed-mating individuals from France, Germany, Poland and Switzerland, a northern European population of highly self-fertilizing individuals from Sweden, Norway and Iceland, a mixed-mating population containing individuals from Spain and Madeira, and two outcrossing populations representing individuals from Italy and Greece, respectively (Fig. 1). Subsequent population genetic analyses are presented separately for regional population samples from each of these geographical regions.



**Fig. 2.** The impact of mating system on purifying selection in *A. alpina*. A. The distribution of fitness effects (DFE) in bins of  $N_e s$  (the product of the effective population size and the selection coefficient) for new nonsynonymous mutations, estimated under model with a stepwise population size change. Error bars show  $\pm 1$  standard error. Asterisks indicate significant differences ( $FDR < 0.05$ ) among populations. B. Scandinavian *A. alpina* show an increase in the frequency of derived major effect polymorphism relative to synonymous polymorphism. The figure shows the log ratio of major effect derived allele frequencies to 4-fold synonymous allele frequencies, downsampled to 10 derived alleles in all populations. C. The recessive genetic load for major effect alleles (number of derived homozygous genotypes). D. The additive genetic load for major effect alleles (number of derived alleles). In C and D letters indicate groups with statistically significant differences ( $P < 0.05$ , Kruskal-Wallis test, post-hoc Dunn test).

### Genetic diversity is maintained in mixed-mating but not in highly selfing populations

For each regional population, we quantified nucleotide diversity at three categories of sites: synonymous sites, nonsynonymous sites and intergenic sites in regions of low gene density and high recombination rate (Table 1, See SI Appendix, Table S4). At all three categories of sites, levels of nucleotide diversity ( $\pi$ ) varied by an order of magnitude or more among regional populations (Table 1, See SI Appendix, Table S4). The outcrossing Greek population was the most genetically diverse (synonymous diversity  $\pi_S = 0.008$ ), whereas the highly self-fertilizing Scandinavian populations had very low nucleotide diversity ( $\pi_S = 0.0002$ ; Table 1). Levels of nucleotide diversity were intermediate and of a similar magnitude in both outcrossing Italian and mixed-mating French and Spanish populations (Table 1), which also have similar demographic histories (See SI Appendix, Fig. S7). These results thus suggest that in *A. alpina*, mixed-mating populations maintain similar levels of genetic diversity as outcrossing populations. Similar patterns were seen for nonsynonymous sites, although the relative reduction in diversity in Scandinavia was less severe for nonsynonymous than for synonymous sites (See SI Appendix, Table S4). This resulted in a markedly elevated ratio of nonsynonymous to synonymous nucleotide diversity ( $\pi_N/\pi_S$ ) in Scandinavia (Table 1).

### Selfing, but not mixed-mating, is associated with a genomic signature of relaxed purifying selection

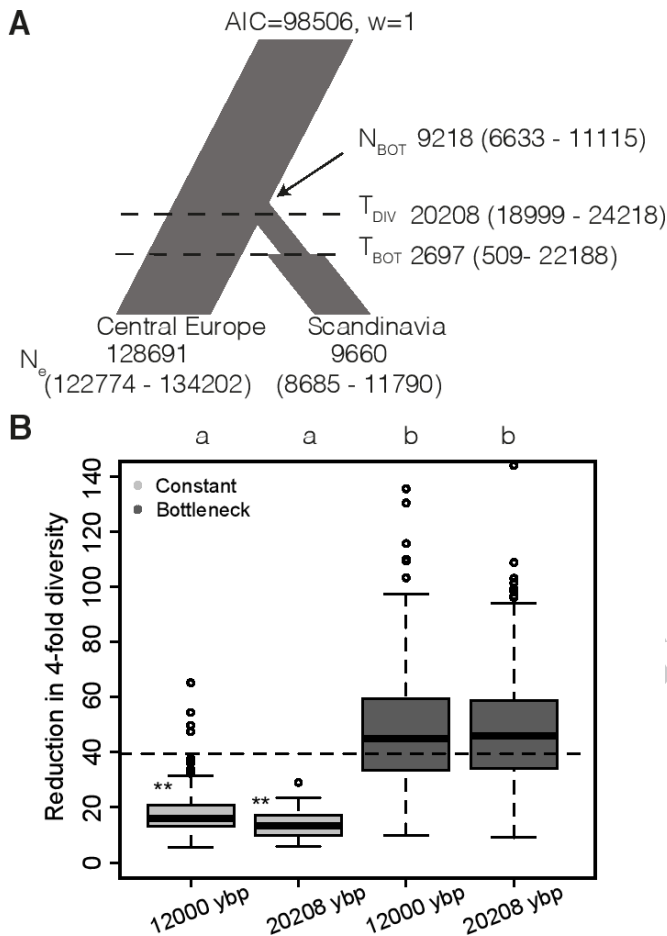
To test whether the impact of purifying selection varied with mating system, we estimated the distribution of negative fitness effects (DFE) of new nonsynonymous mutations using DFE-alpha (37) (See Methods for details). This method can detect changes in selection in association with plant mating system shifts, including purging of strongly deleterious mutations, which should be visible as an increase in the proportion of new mutations under strong purifying selection (38), and corrects for effects of

demographic changes on allele frequency spectra using a simple population size change model. We summarized the strength of selection, defined as the product of the effective population size  $N_e$  and the selection coefficient  $s$ , in three bins, ranging from nearly neutral to strongly deleterious ( $0 < N_e s < 1$ ;  $1 < N_e s < 10$ ;  $N_e s > 10$ ). While the DFE of the mixed-mating Spanish and outcrossing Greek populations differed, there were no significant differences in the DFE of the outcrossing Italian population and the mixed-mating Spanish and French populations (Fig. 2A). We detected no signature of purging of strongly deleterious mutations in mixed-mating populations based on DFE analyses (Fig. 2A), as the inferred proportion of strongly deleterious new nonsynonymous mutations was not generally higher in the mixed-mating French and Spanish than in the outcrossing Greek and Italian populations. Thus, *A. alpina* populations undergoing as much as 80% selfing do not show population genomic evidence of relaxed purifying selection or increased purging of recessive deleterious mutations.

In contrast to results for mixed-mating populations, several lines of evidence suggest that selection against deleterious alleles is compromised in highly self-fertilizing Scandinavian *A. alpina*. First, there was a strong difference in the DFE of nonsynonymous mutations, with a strong, significant reduction in selection against strongly deleterious nonsynonymous mutations ( $N_e s > 10$ ) in Scandinavia compared to all other regional populations (Fig. 2A). Thus, DFE analyses suggest that the Scandinavian population could be accumulating strongly deleterious nonsynonymous mutations.

In line with our inference of accumulation of deleterious alleles in Scandinavian *A. alpina*, we found that derived alleles with a major effect on gene integrity (see Methods) were at a markedly higher frequency relative to derived synonymous alleles in Scandinavia compared to all other regional populations (Fig. 2B). Because selected and neutral allele frequency spectra can be





**Fig. 3.** A recent bottleneck and selfing explain the reduction of polymorphism in Scandinavia. **A.** The best-fit demographic model of the colonization of Scandinavia from a Central European population. Estimated times are given in years before present (ybp). **B.** Background selection alone does not explain the reduction in diversity in Scandinavian *A. alpina*. Boxplots show the ratio of synonymous polymorphism between an outcrossing population and a 90% selfing population experiencing either a constant population size or a 10-fold bottleneck, with the two populations diverging either 12,000 ybp or 20,208 ybp. The dashed line indicates the observed ratio of synonymous polymorphism in Central Europe to that in Scandinavia. Letters indicate significant difference between models (Mann-Whitney test  $P < 0.001$ ). Asterisks indicate an observed neutral diversity reduction significantly greater than that expected based on 300 simulations.

differently affected by demographic changes even in the absence of an actual change in selection (39-41), we considered additional proxies for genetic load. First, we estimated the average number of homozygous derived major-effect genotypes (Fig. 2C), which is expected to be proportional to genetic load if deleterious mutations act recessively, and then we estimated the average number of derived major-effect alleles (Fig. 2D), which is proportional to genetic load if deleterious mutations have additive effects (42). According to both statistics, the highly selfing Scandinavian regional population had an elevated genetic load compared to mixed-mating populations, and outcrossing and mixed-mating populations were not different from each other (Fig. 2C-D). Estimates of additive load were lower in mixed-mating than in outcrossing populations, but the difference was not statistically significant (Fig. 2D). In addition, the number of fixed derived major effect variants in Scandinavia was highly inflated (~70% increase) compared to both the outcrossing, and mixed mating populations (See SI Appendix, Table S5). A similar pattern was seen for derived nonsynonymous variants as well as derived

nonsynonymous variants at evolutionarily constrained sites, and when considering each geographical population separately (See SI Appendix, Fig. S3-S4; Table S5). Overall, this suggests that the highly selfing *A. alpina* from Scandinavia are accumulating deleterious mutations, whereas we observe no statistically significant evidence for purging of genetic load in mixed-mating populations.

#### A recent bottleneck and selfing explain polymorphism in the Scandinavian population

Theory predicts that high levels of self-fertilization should result in effective purging of recessive deleterious variation, unless selfing is associated with strong reductions in the effective population size (17). Given that we observed no evidence for purging and a strong reduction in genetic variation in Scandinavian *A. alpina*, we asked whether this could be due to a bottleneck, or if increased background selection due to selfing is sufficient to explain the reduction of diversity.

Previous phylogeographic analyses have shown that *A. alpina* likely originated in Asia Minor and subsequently spread westward about 500 kya (29, 31). Population genetic analyses have identified Central European populations as the most likely source for Scandinavian *A. alpina* populations (30), but so far no explicit demographic model has been fit to estimate the timing or demographics of colonization of Scandinavia by *A. alpina*. For this purpose, we used a maximum likelihood-based approach to estimate the parameters of a demographic model of colonization of Scandinavia using a two-dimensional site frequency spectrum (2D-SFS, See SI Appendix, Fig S5) based on a scattered sample of individuals from Central Europe and the highly selfing Scandinavian *A. alpina* (see Methods). For analyses of demographic history, we used a set of 12,967 SNPs in intergenic regions with low gene density and high recombination rate, which are expected to be less affected by linked selection and thus useful for demographic inference (12, 43). According to our best-fit model, the split between Central Europe and Scandinavia occurred approximately 20 kya and was associated with a prolonged bottleneck (Fig. 3A; See SI Appendix, Fig. S6-7, Table S6).

We conducted population genetic simulations to assess whether the reduction in diversity in Scandinavia could be explained by a stronger impact of background selection due to selfing, without additional demographic changes. These simulations, which used realistic settings for *A. alpina* genome structure, including variation in recombination rate, gene density and mutation rates, and a realistic distribution of negative fitness effects (see Methods), show that models that incorporate a transition to selfing but no demographic change cannot explain the reduction in diversity we observe in the Scandinavian population ( $P = 0.013$  based on 300 simulations), whereas a model which includes a 10-fold bottleneck and a transition to selfing is consistent with the observed reduction in genetic diversity ( $P = 0.58$  based on 300 simulations, Fig. 3B). These conclusions are robust if we assume a more recent split between the Central European and the Scandinavian population, as late as 12,000 years ago (Fig. 3B). The results also hold under a different DFE (see Methods) (See SI Appendix, Fig S9). These demographic modelling results suggest that a postglacial colonization bottleneck reduced diversity and affected the impact of natural selection in Scandinavian *A. alpina*.

#### Discussion

We have used population genomic analyses to investigate effects of mating system and demographic history on genetic variation and purifying selection in the arctic-alpine crucifer species *A. alpina*. Our results show that populations with intermediate levels of self-fertilization maintain genetic variation and experience similar levels of purifying selection as outcrossing populations. We further find that highly selfing Scandinavian *A. alpina* experience relaxed purifying selection, most likely as a result of having undergone a severe bottleneck, the timing of which is consistent

with postglacial recolonization of Northern Europe. Our results suggest a strong effect of demography on the impact of purifying selection in selfing populations of this species, and demonstrate that intermediate levels of outcrossing can allow populations to avoid the negative population genetic consequences of self-fertilization.

The fact that mixed-mating populations maintained similar levels of genetic diversity as outcrossing populations suggests that the loss of self-incompatibility in *A. alpina* was not associated with a recent and strong bottleneck. Empirical studies in other plant species have frequently found higher diversity in outcrossing and mixed-mating species relative to highly self-fertilizing species (7, 44), in good agreement with our results. Taken together these findings agree well with the expectation that levels of diversity should be higher in outcrossing than in self-fertilizing populations (6).

We detected no strong evidence for purging of deleterious alleles in mixed-mating *A. alpina* populations based on analyses of the DFE. This suggests that, in contrast to expectations from single-locus theory (17), strongly deleterious alleles may not be rapidly purged as a result of partial selfing in *A. alpina*, perhaps due to selective interference (5). While our power to detect purging may be limited if strongly deleterious recessive mutations are rare, a recent simulation-based study showed that the analysis method used here can detect the signal of purging following shifts to higher selfing rates (38). Indeed, no strong differences between outcrossing and mixed-mating populations were found for the DFE of new nonsynonymous mutations, derived allele frequencies of major-effect alleles. Estimates of additive load were lower for mixed-mating than outcrossing populations, which suggests that non-random mating could result in some purging of deleterious alleles (17), however these differences were not statistically significant. Our results thus suggest that the genome-wide impact of purifying selection is similar in outcrossing and mixed-mating populations of *A. alpina*.

While there is a dearth of genome-wide studies contrasting purifying selection in outcrossing and mixed-mating plants, one previous study also found modest differences in purifying selection between outcrossing and mixed-mating species (24). These observations agree with simulation-based results that suggest a low degree of outcrossing is sufficient to maintain efficient purifying selection (15). Thus, mixed-mating populations may avoid some of the negative evolutionary genetic effects associated with high selfing rates, which could imply that population genetic effects contribute to the stable maintenance of intermediate outcrossing rates (5). However, ultimately, direct estimates of inbreeding depression are needed to fully elucidate the maintenance of mating system variation in *A. alpina*.

In contrast to the lack of differences in purifying selection between outcrossing and mixed-mating populations, we found strong evidence for accumulation of deleterious mutations in highly self-fertilizing Scandinavian *A. alpina*. Indeed, our DFE analyses indicate less efficient selection against strongly deleterious alleles, and we observe an increase in the relative allele frequency and fixed derived major-effect mutations, as well as elevated additive and expressed genetic load in Scandinavian *A. alpina*. These results suggest that deleterious alleles are not efficiently selected against and have been able to increase in frequency in these populations. In the context of human population history, theory and simulations have shown that range expansions (45) or strong and extended bottlenecks (39, 41, 42) can lead to increased genetic load. It has previously been shown that genetic variation is strongly reduced over a vast geographical area representing the northern part of the distribution of *A. alpina* (30). Our demographic inference and population genetic simulations suggest that this reduction of genetic variation is most likely a result of postglacial colonization bottlenecks. Thus, bottlenecks

associated with range expansion appear to have strongly reduced the effective population size and increased the impact of drift in the Northern part of the species range, causing accumulation of deleterious variants in Scandinavian *A. alpina*. The impact of demographic history on genetic load is currently a strongly debated topic in human population genetics, where studies on the effect of the out-of-Africa bottleneck on genetic load have come to different conclusions depending on the statistics they applied (39, 46-48). Here, we extend these studies to a plant species and document increased accumulation of deleterious mutations in bottlenecked Scandinavian *A. alpina*.

Increased background selection can lead to sharply reduced genetic diversity in highly selfing populations (7, 12), but our forward population genetic simulations show that this effect alone cannot explain the reduction in diversity in Scandinavia, whereas a model with both selfing and a population size reduction is consistent with the observed reduction in diversity. At present, however, we cannot rule out a contribution of positive selection, or other changes in selection during range expansion, to reduced diversity in the Scandinavian *A. alpina* population. Indeed, reciprocal transplant experiments have documented adaptive differentiation between Spanish and Scandinavian populations of *A. alpina* (49). While elevated load can be associated with enhanced local adaptation following a range expansion (50), this raises the possibility that some of the increase in frequencies of major-effect variants in Scandinavia may have been directly or indirectly driven by positive selection. However, we believe that it is unlikely to result in the genome-wide signature of relaxed purifying selection that we observe in Scandinavian *A. alpina*. Empirical identification of genomic regions responsible for local adaptation and population genomic studies using larger sample sizes will be needed to explore the genomic impact of positive selection.

Here, we have investigated the impact of demographic history and mating system on genomic patterns of variation in *A. alpina*. Our results show that mixed-mating populations maintain genetic variation and purifying selection at similar levels as outcrossing populations. In contrast, we find an increase in genetic load in highly self-fertilizing populations, most likely as a result of demographic effects associated with postglacial range expansion. Our results are important for a more general understanding of the impact of mating system and demographic history on genomic variation and selection in plants.

## Methods

### Data and sequencing

We performed paired-end (100 bp) whole-genome resequencing of 38 *Arabis alpina* individuals from Europe (See SI Appendix, Table S1), utilizing libraries with an insert size of 300-400 bp (See SI Appendix, SI Text) on an Illumina HiSeq 2000 instrument (Illumina, San Diego, CA, USA). We obtained a total of 10,079 Gbp (QC > 30) with a mean coverage of 26X ranging from 16X to 45X.

### Quality assessment, trimming, genotype calling and filtering

Adapter trimmed reads were mapped to the *A. alpina* V4 reference genome assembly (31) using BWA-MEM v0.7.8 (51), and duplicate free BAM alignment files were further processed using the Genome Analysis Toolkit (GATK v.3.4.0; 52) (See SI Appendix, SI Text). The *A. alpina* genome assembly is enriched for repetitive elements relative to Brassicaceae relatives (32) so we employed a variety of hard and custom filtering techniques (See SI Appendix, SI Text) to avoid calling SNPs in regions of the genome that putatively represent copy number variants. After applying all filters the dataset contained 1,514,615 SNPs and 43,209,020 invariant sites.

### Inference of population structure and population genetic analyses of selection

Population structure was inferred using a combination of principal component analysis (PCA), and Bayesian clustering analysis using fastSTRUCTURE v1.0 (35) and TESS v3 (36) (See SI Appendix, SI Text).

For each of the five regional populations identified, we estimated summary statistics ( $S$ ,  $\pi$ , Tajima's  $D$ ) at 4-fold degenerate synonymous sites, 0-fold degenerate nonsynonymous sites, and intergenic sites in regions with high recombination and low gene density (Table 1, See SI Appendix, SI Text). We estimated the distribution of fitness effects (DFE) using DFE-alpha v. 2.15 (37) on folded 4-fold and 0-fold site frequency spectra (SFS) under

681  
682  
683  
684  
685  
686  
687  
688  
689  
690  
691  
692  
693  
694  
695  
696  
697  
698  
699  
700  
701  
702  
703  
704  
705  
706  
707  
708  
709  
710  
711  
712  
713  
714  
715  
716  
717  
718  
719  
720  
721  
722  
723  
724  
725  
726  
727  
728  
729  
730  
731  
732  
733  
734  
735  
736  
737  
738  
739  
740  
741  
742  
743  
744  
745  
746  
747  
748

a stepwise population size change model (See SI Appendix, SI Text, Table S7). We compared the DFE of regional populations based on 200 bootstrap replicates.

#### Major effect mutations and genetic load

Presence and frequency of major effect mutations, i.e. loss of start and stop codons; gain of stop codons and changes in splice sites were calculated per population using snpEFF v4.2 (53). To avoid reference biases we polarized the SNPs using the *A. montbretiana* genome assembly (ASM148412v1, 54) as an outgroup (See SI Appendix, SI text). As a proxy for genetic load, we estimated the average number of derived nonsynonymous and major-effect homozygous genotypes (41, 46), and the average number of derived alleles for nonsynonymous and major-effect alleles (42) for each regional population. We repeated these analyses for nonsynonymous variants at highly constrained sites (See SI Appendix, SI text). In addition, we counted the total number of fixed derived nonsynonymous and major-effect alleles for each regional population.

#### Demographic modelling and simulations

We conducted demographic inference in the software fastsimcoal2 v. 2.5.2.21 (55). To estimate parameters associated with the origin of Scandi-

navian *A. alpina*, we compared three demographic models (See SI Appendix, Fig. S5), using two-dimensional joint SFS based on a scattered sample from central Europe and the Scandinavian population (See SI Appendix, SI Text; Fig. S6). We used 12,967 intergenic sites, a mutation rate of  $7 \times 10^{-9}$  and a generation time of 1.5 years. We used forward simulation in SLiM2 v2.1 (56) to assess the impact of demography and selection associated with a shift to selfing on genetic diversity in the Scandinavian population under four demographic models with varying bottleneck severity and population split time (See SI Appendix, SI Text).

#### Acknowledgements

We thank Tiina Mattila for advice on population genetic analyses and Cindy Canton for help with plant care and extractions. Computations were performed on resources provided by SNIC through UPPMAX under Projects snic2014-1-194, b2013022 and b2013237. This work was funded by grants from the Swedish Research Council to JÄ and TS, by a grant from the DFG through SPP1529 to GC, and from SciLifeLab to TS.

the selfing syndrome: Anther orientation and herkogamy together determine reproductive assurance in a self-compatible plant. *Evolution*, 71:2206–2218.

1. Goodwillie C, Kalisz S, Eckert C (2005) The evolutionary enigma of mixed mating systems in plants: Occurrence, theoretical explanations, and empirical evidence. *Annu Rev Ecol Evol S* 36:47–79.
2. Igic B, Kohn JR (2006) The distribution of plant mating systems: study bias against obligately outcrossing species. *Evolution* 60(5):1098–1103.
3. Igic B, Busch JW (2013) Is self-fertilization an evolutionary dead end? *New Phytol* 198(2):386–397.
4. Lande R, Schemske D (1985) The evolution of self-fertilization and inbreeding depression in plants. I. Genetic models. *Evolution* 39:24–40.
5. Hartfield M (2016) Evolutionary genetic consequences of facultative sex and outcrossing. *J Evol Biol* 29(1):5–22.
6. Wright SI, Kalisz S, Slotte T (2013) Evolutionary consequences of self-fertilization in plants. *Proc Biol Sci* 280(1760):20130133.
7. Barrett SCH, Arunkumar R, Wright SI (2014) The demography and population genomics of evolutionary transitions to self-fertilization in plants. *Philos T Roy Soc B* 369(1648). doi:10.1098/rstb.2013.0344.
8. Pollak E (1987) On the theory of partially inbreeding finite populations. I. Partial selfing. *Genetics* 117(2):353–360.
9. Nordborg M (2000) Linkage disequilibrium, gene trees and selfing: an ancestral recombination graph with partial self-fertilization. *Genetics* 154(2):923–929.
10. Ingvarsson PK (2002) A metapopulation perspective on genetic diversity and differentiation in partially self-fertilizing plants. *Evolution* 56(12):2368–2373.
11. Charlesworth D, Wright SI (2001) Breeding systems and genome evolution. *Curr Opin Genet Dev* 11(6):685–690.
12. Slotte T (2014) The impact of linked selection on plant genomic variation. *Brief Funct Genomics* 13(4):268–275.
13. Kimura M (1983) *The neutral theory of molecular evolution* (Cambridge University Press, Cambridge).
14. Ohta T (1973) Slightly deleterious mutant substitutions in evolution. *Nature* 246(5428):96–98.
15. Kamran-Disfani A, Agrawal AF (2014) Selfing, adaptation and background selection in finite populations. *J Evol Biol* 27(7):1360–1371.
16. Bataillon T, Kirkpatrick M (2000) Inbreeding depression due to mildly deleterious mutations in finite populations: size does matter. *Genet Res* 75(1):75–81.
17. Glémin S (2003) How are deleterious mutations purged? Drift versus nonrandom mating. *Evolution* 57(12):2678–2687.
18. Glémin S (2007) Mating systems and the efficacy of selection at the molecular level. *Genetics* 177(2):905–916.
19. Qiu S, Zeng K, Slotte T, Wright S, Charlesworth D (2011) Reduced efficacy of natural selection on codon usage bias in selfing *Arabidopsis* and *Capsella* species. *Genome Biol Evol* 3:868–880.
20. Ness RW, Sjol M, Barrett SCH (2012) Genomic consequences of transitions from cross- to self-fertilization on the efficacy of selection in three independently derived selfing plants. *BMC Genomics* 13:611.
21. 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*. *PLoS Genet* 9(9):e1003754.
22. Slotte T, et al. (2013) The *Capsella rubella* genome and the genomic consequences of rapid mating system evolution. *Nat Genet* 45(7):831–835.
23. Douglas GM, et al. (2015) Hybrid origins and the earliest stages of diploidization in the highly successful recent polyploid *Capsella bursa-pastoris*. *Proc Natl Acad Sci USA* 112(9):2806–2811.
24. Salcedo A, Kalisz S, Wright SI (2014) Limited genomic consequences of mixed mating in the recently derived sister species pair, *Collinsia concolor* and *Collinsia parryi*. *J Evol Biol* 27(7):1400–1412.
25. Tedder A, Ansell SW, Lao X, Vogel JC, Mable BK (2011) Sporophytic self-incompatibility genes and mating system variation in *Arabis alpina*. *Ann Bot-London* 108(4):699–713.
26. Ansell SW, Grundmann M, Russell SJ, Schneider H, Vogel JC (2008) Genetic discontinuity, breeding-system change and population history of *Arabis alpina* in the Italian Peninsula and adjacent Alps. *Mol Ecol* 17(9):2245–2257.
27. Buehler D, Graf R, Holderegger R, Gugerli F (2012) Contemporary gene flow and mating system of *Arabis alpina* in a Central European alpine landscape. *Ann Bot-London* 109(7):1359–1367.
28. Toräng, P., L. Vikström, J. Wunder, S. Wötzel, G. Coupland, and J. Ågren. 2017. Evolution of
29. Koch MA, et al. (2006) Three times out of Asia Minor: the phylogeography of *Arabis alpina* L. (Brassicaceae). *Mol Ecol* 15(3):825–839.
30. Ehrlich D, et al. (2007) Genetic consequences of Pleistocene range shifts: contrast between the Arctic, the Alps and the East African mountains. *Mol Ecol* 16(12):2542–2559.
31. Ansell SW, et al. (2011) The importance of Anatolian mountains as the cradle of global diversity in *Arabis alpina*, a key arctic-alpine species. *Ann Bot-London* 108(2):241–252.
32. Willing E-M, et al. (2015) Genome expansion of *Arabis alpina* linked with retrotransposition and reduced symmetric DNA methylation. *Nature Plants* 1 SP - EP -(2):14023.
33. Morton NE, Crow JF, Muller HJ, (1956) An estimate of the mutational damage in man from data on consanguineous marriages. *Proc Natl Acad Sci USA* 42:855–863.
34. Kimura M, Maruyama T, Crow JF (1963) Mutation load in small populations. *Genetics* 48:1303–1312.
35. Raj A, Stephens M, Pritchard JK (2014) fastSTRUCTURE: variational inference of population structure in large SNP data sets. *Genetics* 197(2):573–589.
36. Caye K, Deist TM, Martins H, Michel O, François O (2016) TESS3: fast inference of spatial population structure and genome scans for selection. *Mol Ecol Resour* 16(2):540–548.
37. 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(4):2251–2261.
38. Arunkumar R, Ness RW, Wright SI, Barrett SCH (2014) The evolution of selfing is accompanied by reduced efficacy of selection and purging of deleterious mutations. *Genetics* 199(3):817–829.
39. Gravel S (2016) When is selection effective? *Genetics* 203(1):451–462.
40. Brandvain Y, Wright SI (2016) The limits of natural selection in a nonequilibrium world. *Trends Genet* 32(4):201–210.
41. Pedersen C-ET, et al. (2017) The effect of an extreme and prolonged population bottleneck on patterns of deleterious variation: insights from the greenlandic inuit. *Genetics* 205(2):787–801.
42. Henn BM, Botigué LR, Bustamante CD, Clark AG, Gravel S (2015) Estimating the mutation load in human genomes. *Nat Rev Genet* 16(6):333–343.
43. Messer PW, Petrov DA (2013) Frequent adaptation and the McDonald-Kreitman test. *Proc Natl Acad Sci USA*. 110(21):8615–8620.
44. Hamrick JL, Godt MJW (1996) Effects of life history traits on genetic diversity in plant species. *Phil Trans Roy Soc B* 351(1345):1291–1298.
45. Peischl S, Dupanloup I, Kirkpatrick M, Excoffier L (2013) On the accumulation of deleterious mutations during range expansions. *Mol Ecol* 22(24):5972–5982.
46. Henn BM, et al. (2016) Distance from sub-Saharan Africa predicts mutational load in diverse human genomes. *Proc Natl Acad Sci USA* 113(4):E440–9.
47. Lohmueller KE, et al. (2008) Proportionally more deleterious genetic variation in European than in African populations. *Nature* 451(7181):994–997.
48. Simons YB, Turchin MC, Pritchard JK, Sella G (2014) The deleterious mutation load is insensitive to recent population history. *Nat Genet* 46(3):220–224.
49. Toräng P, et al. (2015) Large-scale adaptive differentiation in the alpine perennial herb *Arabis alpina*. *New Phytol* 206(1):459–470.
50. Gilbert KJ, et al. (2017) Local adaptation interacts with expansion load during range expansion: maladaptation reduces expansion load. *Am Nat* 189:368–380.
51. Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25(14):1754–1760.
52. McKenna A, et al. (2010) The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 20(9):1297–1303.
53. Cingolani P, et al. (2012) A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly* 6(2):80–92.
54. Kiefer C, et al. (2017) Divergence of annual and perennial species in the Brassicaceae and the contribution of cis-acting variation at *FLC* orthologues. *Mol Ecol* 26(13):3437–3457.
55. Excoffier L, et al. (2013) Robust demographic inference from genomic and SNP data. *PLoS Genet* 9(10):e1003905.
56. Haller BC, Messer PW (2017) SLiM 2: Flexible, interactive forward genetic simulations. *Mol Biol Evol* 34(1):230–240.

749  
750  
751  
752  
753  
754  
755  
756  
757  
758  
759  
760  
761  
762  
763  
764  
765  
766  
767  
768  
769  
770  
771  
772  
773  
774  
775  
776  
777  
778  
779  
780  
781  
782  
783  
784  
785  
786  
787  
788  
789  
790  
791  
792  
793  
794  
795  
796  
797  
798  
799  
800  
801  
802  
803  
804  
805  
806  
807  
808  
809  
810  
811  
812  
813  
814  
815  
816