

1

2 **Demography and mating system shape the genome-wide impact of**
3 **purifying selection in *Arabidopsis***

4

5 Benjamin Laenen^{a,1}, Andrew Tedder^{a,1}, Michael D. Nowak^a, Per Toräng^b, Jörg
6 Wunder^c, Stefan Wötzel^c, Kim A. Steige^{a,d}, Yiannis Kourmpetis^{e,f}, Thomas Odong^e,
7 Andreas D. Drouzas^g, Marco Bink^{e,h}, Jon Ågren^{b,2}, George Coupland^{c,2}, Tanja Slotte^{a,2}

8

9 ^aDept. of Ecology, Environment and Plant Sciences, Science for Life Laboratory,
10 Stockholm University,

11 ^bDept. of Ecology and Genetics, Uppsala University

12 ^cMax Planck Institute for Plant Breeding Research

13 ^dCurrent address: Institute of Botany, Biozentrum, University of Cologne, 50674
14 Cologne, Germany

15 ^eBiometris, Wageningen University & Research Centre, P.O. Box 16, 6700 AC
16 Wageningen, The Netherlands

17 ^fCurrent address: Nestlé Institute of Health Sciences, EPFL Campus, 1015 Lausanne,
18 Switzerland

19 ^gSchool of Biology, Aristotle University of Thessaloniki

20 ^hCurrent address: Hendrix Genetics Research, Technology & Services B.V., P.O. Box
21 114, 5830 AC Boxmeer, The Netherlands

22

23 ¹B.L. and A.T. contributed equally to this work.

24 ²Authors for correspondence, email: Jon Ågren: jon.agren@ebc.uu.se; George
25 Coupland: coupland@mpipz.mpg.de, Tanja Slotte: Tanja.Slotte@su.se

26

27

28 **Keywords**

29 self-fertilization, demographic history, bottleneck, fitness effects, genetic load, mixed-
30 mating

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

51 **Significance**

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

63

64 **Introduction**

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

77 While selfing can be favored due to its genetic transmission advantage, and
78 because it can confer reproductive assurance, it also has marked population genetic
79 consequences that might contribute to the long-term demise of highly selfing lineages
80 (6). For instance, highly selfing populations are expected to have a reduced effective
81 population size (N_e), due to the direct effect of inbreeding (8, 9). Demographic
82 processes such as frequent extinction and recolonization of local subpopulations (10)
83 or founder events associated with the shift to selfing (6) can reduce genetic variation
84 in selfers to an even greater extent. Self-fertilization is also expected to lead to

85 elevated linkage disequilibrium, which means that background selection and other
86 forms of linked selection can reduce genetic variation genome-wide, further reducing
87 N_e in selfers (7, 11, 12).

88 Because the strength of selection scales with N_e (13), natural selection is
89 expected to be less efficient in selfers than in outcrossers. Selfers are therefore
90 expected to accumulate weakly deleterious mutations at a higher rate than outcrossers
91 (14). However, selfing also increases homozygosity, exposing recessive mutations to
92 selection. This is expected to result in purging of recessive deleterious mutations,
93 unless selfing is associated with a strong reduction in N_e which renders such purging
94 ineffective (15-17).

95 While there is accumulating evidence for relaxed selection on weakly
96 deleterious mutations in highly selfing lineages (5, 18-22), several fundamental
97 questions on the evolutionary genomic consequences of mating systems remain
98 unanswered. First, it is unclear whether selfing generally results in purging of
99 recessive deleterious alleles or whether this effect is possibly overridden by
100 demographic effects. Second, it remains unclear whether intermediate levels of
101 outcrossing allow plants to avoid the negative evolutionary genetic consequences of
102 high selfing rates. Specifically, while theory predicts that intermediate levels of self-
103 fertilization could be sufficient to achieve purging of deleterious alleles (16) and
104 prevent accumulation of mildly deleterious alleles (23), few empirical genome-wide
105 studies have examined the selective consequences of mixed-mating in plants (but see
106 (24)).

107 The broadly distributed, arctic-alpine perennial herb *Arabis alpina*
108 (Brassicaceae) is a promising plant system in which to address the impact of variation
109 in outcrossing rates on genome-wide genetic variation and efficiency of selection.
110 This species harbors populations that express a range of mating strategies from self-
111 incompatible outcrossing (25) through mixed-mating to autonomous selfing (25-27).
112 The colonization history of *A. alpina* has already begun to be characterized (26, 28-
113 30), which facilitates interpretation of global patterns of polymorphism. Finally, the
114 availability of a genome assembly of *A. alpina* (31) greatly facilitates population
115 genomic studies.

116 In this study, we investigate the effects of mating system and demography on
117 the efficacy of selection in *A. alpina* by population genomic analyses of whole-
118 genome resequencing data from outcrossing, mixed-mating and highly selfing

119 populations. We first investigate population structure and test whether populations
120 with higher selfing rates have lower levels of genetic diversity, and then test whether
121 higher selfing rates are associated with relaxed selection against weakly deleterious
122 mutations and purging of strongly deleterious mutations. To do this, we use genome-
123 wide allele frequency distributions at nonsynonymous and synonymous sites to
124 estimate the fraction of weakly deleterious and strongly harmful new nonsynonymous
125 mutations. We further test whether higher rates of selfing are associated with an
126 increase in the frequency of derived alleles with major effects on gene integrity,
127 which would suggest relaxed purifying selection. Finally, we compare two genomic
128 proxies of genetic load, the reduction in mean fitness of a population caused by
129 deleterious variation (32,33), among outcrossing, mixed-mating and highly self-
130 fertilizing populations. Our results are important for an improved understanding of the
131 population genetic consequences of mating system variation.

132

133 **Results**

134

135 **Sequencing and single nucleotide polymorphism**

136 We sampled 38 *A. alpina* individuals, with targeted sampling of 2-5 individuals from
137 12 geographical sites harboring self-incompatible outcrossing populations (Greece
138 and Italy), populations with intermediate outcrossing rates (France and Spain), and
139 highly selfing populations (Scandinavia), in addition to a sample of single individuals
140 from 5 additional geographic locations across the European range of *A. alpina* (SI
141 Appendix, Table S1). Progeny-array based outcrossing estimates have shown that
142 populations from Scandinavia are highly selfing (up to ~10% outcrossing) whereas
143 intermediate outcrossing rates have been estimated for two French and Spanish
144 populations (~20% and ~18%, respectively) (SI Appendix, SI Text). We further
145 verified that Greek and Italian individuals produced no offspring after forced selfing
146 in the greenhouse. The distribution of genomic runs of homozygosity (ROHs) further
147 supported mating system variation among the populations studied here (SI Appendix,
148 SI Text).

149 Each individual was resequenced to high coverage (average 26X, range 16-
150 45X) using Illumina short-read technology. We called single nucleotide
151 polymorphisms (SNPs) and applied stringent filtering criteria to identify a total of
152 1,514,615 high-quality SNPs, of which 98,564 were 0-fold degenerate

153 nonsynonymous (i.e. sites at which any mutation will result in a nonsynonymous
154 change) with a mean nucleotide diversity (π) of 0.0027 and 65,821 were 4-fold
155 degenerate synonymous (i.e. synonymous sites at which any mutation will result in a
156 synonymous change) with a mean nucleotide diversity of 0.0102 (SI Appendix, Table
157 S2).

158

159 **Population structure has a strong geographic component**

160 We analyzed population structure with two model-based Bayesian clustering
161 approaches, implemented in the software fastSTRUCTURE (34) and TESS3 (35),
162 based on 25,505 4-fold synonymous SNPs (see Methods). Both methods, as well as
163 principal component analysis gave very similar results, and supported the presence of
164 five clusters (Fig. 1; SI Appendix, Fig. S1), in good agreement with previous analyses
165 of population structure in *A. alpina* (29). These clusters correspond to a central
166 European population of mixed-mating individuals from France, Germany, Poland and
167 Switzerland, a northern European population of highly self-fertilizing individuals
168 from Sweden, Norway and Iceland, a mixed-mating population containing individuals
169 from Spain and Madeira, and two outcrossing populations representing individuals
170 from Italy and Greece, respectively (Fig. 1). Subsequent population genetic analyses
171 are presented separately for regional population samples from each of these
172 geographical regions.

173

174 **Genetic diversity is maintained in mixed-mating but not in highly selfing** 175 **populations**

176 For each regional population, we quantified nucleotide diversity at three categories of
177 sites: 4-fold degenerate sites, 0-fold degenerate sites and intergenic sites in regions of
178 low gene density and high recombination rate (Table 1, SI Appendix, Table S3). At
179 all three categories of sites, levels of nucleotide diversity (π) varied by an order of
180 magnitude among regional populations (Table 1, SI Appendix, Table S3). The
181 outcrossing Greek population was the most genetically diverse (4-fold synonymous
182 diversity $\pi_S = 0.008$), whereas the highly self-fertilizing Scandinavian populations had
183 very low nucleotide diversity ($\pi_S = 0.0002$; Table 1). Levels of nucleotide diversity
184 were intermediate and of a similar magnitude in both outcrossing Italian and mixed-
185 mating French and Spanish populations (Table 1), suggesting that in *A. alpina*, mixed-
186 mating populations maintain similar levels of genetic diversity as outcrossing

187 populations. Similar patterns were seen for 0-fold nonsynonymous sites, although the
188 relative reduction in diversity in Scandinavia was less severe for 0-fold sites than for
189 4-fold sites (SI Appendix, Table S3). This resulted in a markedly elevated ratio of
190 nonsynonymous to synonymous nucleotide diversity (π_N/π_S) in Scandinavia (Table 1).

191

192 **Selfing, but not mixed-mating, is associated with relaxed purifying selection**

193 To test whether the impact of purifying selection varied with mating system, we
194 estimated the distribution of negative fitness effects (DFE) of new nonsynonymous
195 mutations using DFE-alpha (37) (See Methods for details). This method can detect
196 changes in selection in association with plant mating system shifts, including purging
197 of strongly deleterious mutations (38), and corrects for effects of demographic
198 changes on allele frequency spectra using a simple population size change model (37).
199 We summarized the strength of selection, defined as the product of the effective
200 population size N_e and the selection coefficient s , in three bins, ranging from nearly
201 neutral to strongly deleterious ($0 < N_e s < 1$; $1 < N_e s < 10$; $N_e s > 10$). While the DFE of the
202 mixed-mating Spanish and outcrossing Greek populations differed, there were no
203 significant differences in the DFE of the outcrossing Italian population and the mixed-
204 mating Spanish and French populations (Fig. 2A). We detected no signature of
205 purging of strongly deleterious mutations in mixed-mating populations based on DFE
206 analyses (Fig. 2A), as the inferred proportion of strongly deleterious new
207 nonsynonymous mutations was not generally higher in the mixed-mating French and
208 Spanish than in the outcrossing Greek and Italian populations. This suggests that *A.*
209 *alpina* populations undergoing as much as 80% selfing do not show evidence of
210 relaxed purifying selection or increased purging of recessive deleterious mutations.

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

219 To follow up on this observation, we used the closely related species *Arabis*
220 *montbretiana* as an outgroup to polarize alleles as ancestral or derived. We then

221 examined the allele frequency distribution of derived alleles, and found that derived
222 alleles with a major effect on gene integrity (see Methods) were found at a markedly
223 higher frequency relative to derived 4-fold synonymous alleles in Scandinavia
224 compared to all other regional populations (Fig. 2B). Because selected and neutral
225 allele frequency spectra can be differently affected by demographic changes even in
226 the absence of an actual change in selection (39-41), we considered two additional
227 proxies for genetic load. First, we estimated the average number of homozygous
228 derived major-effect genotypes (Fig. 2C), which is expected to be proportional to
229 genetic load if deleterious mutations act recessively, and then we estimated the
230 average number of derived major-effect alleles (Fig. 2D), which is proportional to
231 genetic load if deleterious mutations have additive effects (42). According to both
232 statistics, the highly selfing Scandinavian regional population had an elevated genetic
233 load compared to mixed-mating populations, and outcrossing and mixed-mating
234 populations were not different from each other (Fig. 2C-D). In addition, the number of
235 fixed derived major effect variants in Scandinavia was highly inflated (~70%
236 increase) compared to both the outcrossing, and mixed mating populations (SI
237 Appendix, Table S4). A similar pattern was seen for derived nonsynonymous variants
238 and when considering each geographical population separately (SI Appendix, Fig. S3;
239 Fig. S4). Overall, this suggests that the highly selfing *A. alpina* from Scandinavia are
240 accumulating deleterious mutations, whereas mixed-mating and outcrossing
241 populations have similar levels of genetic load.

242

243 **A recent bottleneck and selfing explain polymorphism in the Scandinavian** 244 **population**

245 Theory predicts that high levels of self-fertilization should result in effective purging
246 of recessive deleterious variation, unless selfing is associated with strong reductions
247 in the effective population size (16). Given that we observed no evidence for purging
248 and a strong reduction in genetic variation in Scandinavian *A. alpina*, we asked
249 whether this could be due to a bottleneck, or if increased background selection due to
250 selfing is sufficient to explain the reduction of diversity.

251 Previous phylogeographic analyses have shown that *A. alpina* likely originated
252 in Asia Minor and subsequently spread westward about 500 kya (28, 30). Population
253 genetic analyses have identified Central European populations as the most likely
254 source for Scandinavian *A. alpina* populations (29), but so far no explicit

255 demographic model has been fit to estimate the timing or demographics of
256 colonization of Scandinavia by *A. alpina*. For this purpose, we used a maximum
257 likelihood-based approach to estimate the parameters of a demographic model of
258 colonization of Scandinavia using a two-dimensional site frequency spectrum (2D-
259 SFS) based on a scattered sample of individuals from Central Europe and the highly
260 selfing Scandinavian *A. alpina* (see Methods). For analyses of demographic history,
261 we used a set of 12,967 SNPs in intergenic regions with low gene density and high
262 recombination rate, which are expected to be less affected by linked selection and thus
263 useful for demographic inference (12, 43). According to our best-fit model, the split
264 between Central Europe and Scandinavia occurred approximately 20 kya and was
265 associated with a prolonged bottleneck (Fig. 3A; SI Appendix, Fig. S5).

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

281

282 **Discussion**

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

289 of having undergone a severe bottleneck, the timing of which is consistent with
290 postglacial recolonization of Northern Europe. Our results suggest a strong effect of
291 demography on the impact of purifying selection in selfing populations of this species,
292 and demonstrate that intermediate levels of outcrossing can allow populations to
293 avoid the negative population genetic consequences of self-fertilization.

294 The fact that mixed-mating populations maintained similar levels of genetic
295 diversity as outcrossing populations suggests that the loss of self-incompatibility in *A.*
296 *alpina* was not associated with a recent and strong bottleneck. Empirical studies in
297 other plant species have frequently found higher diversity in outcrossing and mixed-
298 mating species relative to highly self-fertilizing species (7, 44), in good agreement
299 with our results. Taken together these findings agree well with the expectation that
300 levels of diversity should be higher in outcrossing than in self-fertilizing populations
301 (6).

302 We detected no evidence for purging of deleterious alleles in mixed-mating *A.*
303 *alpina* populations. This suggests that, in contrast to expectations from single-locus
304 theory (16), strongly deleterious alleles may not be rapidly purged as a result of
305 partial selfing in *A. alpina*, perhaps due to selective interference (5). While our power
306 to detect purging may be limited if strongly deleterious recessive mutations are rare, a
307 recent simulation-based study showed that the analysis method used here can detect
308 the signal of purging following shifts to higher selfing rates (38). Indeed, no strong
309 differences between outcrossing and mixed-mating populations were found for the
310 distribution of fitness effects of new nonsynonymous mutations, derived allele
311 frequencies of major-effect alleles or for two different proxies of genetic load that
312 assume different dominance levels of deleterious alleles. Our results thus suggest that
313 the genome-wide impact of purifying selection is similar in outcrossing and mixed-
314 mating populations.

315 While there is a dearth of genome-wide studies contrasting purifying selection
316 in outcrossing and mixed-mating plants, one previous study also found modest
317 differences in purifying selection between outcrossing and mixed-mating species (24).
318 These observations agree with simulation-based results that suggest a low degree of
319 outcrossing is sufficient to maintain efficient purifying selection (23). Thus, mixed-
320 mating populations may avoid some of the negative evolutionary genetic effects
321 associated with high selfing rates, which could imply that population genetic effects
322 contribute to the stable maintenance of intermediate outcrossing rates (5). However,

323 ultimately, direct estimates of inbreeding depression are needed to fully elucidate the
324 maintenance of mating system variation in *A. alpina*.

325 In contrast to the lack of differences in purifying selection between
326 outcrossing and mixed-mating populations, we found strong evidence for
327 accumulation of deleterious mutations in highly self-fertilizing Scandinavian *A.*
328 *alpina*. Indeed, our DFE analyses indicate less efficient selection against strongly
329 deleterious alleles, and we observe an increase in the relative allele frequency and
330 fixed derived major-effect mutations, as well as elevated additive and expressed
331 genetic load in Scandinavian *A. alpina*. These results suggest that deleterious alleles
332 are not efficiently selected against and have been able to increase in frequency in this
333 population. In the context of human population history, theory and simulations have
334 shown that range expansions (45) or strong and extended bottlenecks (39, 41, 49) can
335 lead to increased genetic load. It has previously been shown that genetic variation is
336 strongly reduced over a vast geographical area representing the northern part of the
337 distribution of *A. alpina* (29). Our demographic inference and population genetic
338 simulations suggest that this reduction of genetic variation is most likely a result of
339 postglacial colonization bottlenecks. Thus, bottlenecks associated with range
340 expansion appear to have strongly reduced the effective population size and increased
341 the impact of drift in the Northern part of the species range, causing accumulation of
342 deleterious variants in Scandinavian *A. alpina*. The impact of demographic history on
343 genetic load is currently a strongly debated topic in human population genetics, where
344 studies on the effect of the out-of-Africa bottleneck on genetic load have come to
345 different conclusions depending on the statistics they applied (39, 46-48). Here, we
346 extend these studies to a plant species and document increased accumulation of
347 deleterious mutations in bottlenecked Scandinavian *A. alpina*.

348 Increased background selection can lead to sharply reduced genetic diversity
349 in highly selfing populations (7, 12), but our forward population genetic simulations
350 show that this effect alone cannot explain the reduction in diversity in Scandinavia,
351 whereas a model with both selfing and a population size reduction is consistent with
352 the observed reduction in diversity. At present, however, we cannot rule out a
353 contribution of positive selection during range expansion to reduced diversity in the
354 Scandinavian *A. alpina* population. Indeed, reciprocal transplant experiments have
355 documented adaptive differentiation between Spanish and Scandinavian populations
356 of *A. alpina* (49). While elevated load can be associated with enhanced local

357 adaptation following a range expansion (50), this raises the possibility that some of
358 the increase in frequencies of major-effect variants in Scandinavia may have been
359 directly or indirectly driven by positive selection. However, we believe that it is
360 unlikely to result in the genome-wide signature of relaxed purifying selection that we
361 observe in Scandinavian *A. alpina*. Empirical identification of genomic regions
362 responsible for local adaptation and population genomic studies using larger sample
363 sizes will be needed to explore the genomic impact of positive selection.

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

372

373 **Methods**

374 *Data and sequencing*

375 We performed paired-end (100 bp) whole-genome resequencing of 38 *Arabidopsis alpina*
376 individuals from Europe (see SI Appendix, Table S1), utilizing libraries with an insert
377 size of 300-400 bp (see SI Appendix, SI Text) on an Illumina HiSeq 2000 instrument
378 (Illumina, San Diego, CA, USA). We obtained a total of 10,079 Gbp and ~245 Gbp
379 (QC > 30) per sample on average with a mean coverage of 26X ranging from 16X to
380 45X.

381

382 *Quality assessment, trimming, genotype calling and filtering*

383 Adapter trimmed reads were mapped to the *A. alpina* V4 reference genome assembly
384 (30) using BWA-MEM v0.7.8 (51), and duplicate free BAM alignment files were
385 further processed using the Genome Analysis Toolkit (GATK v.3.4.0; 52) (see SI
386 Appendix, SI Text). The *A. alpina* genome assembly is exceptionally enriched for
387 repetitive elements relative to previously sequenced Brassicaceae relatives (31) and so
388 we employed a variety of hard and custom filtering techniques (see SI Appendix, SI
389 Text) to avoid calling SNPs in regions of the genome that putatively represent copy

390 number variants. After applying all filters the dataset was composed of 1,514,615
391 SNPs and 43,209,020 invariant sites amenable to further analysis.

392

393 *Inference of population structure and population genetic analyses of selection*

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

397 For each of the five regional populations identified by population structure
398 analysis, we obtained estimates of summary statistics (S , π , Tajima's D) (Table 1, see
399 SI Appendix, SI Text). We estimated the distribution of fitness effects (DFE) using
400 DFE-alpha v. 2.15 (37) on folded 4-fold and 0-fold site frequency spectra (SFS) under
401 a stepwise population size change model (SI Appendix, SI Text). DFE estimates were
402 compared among regional populations based on 200 bootstrap replicates.

403

404 *Major effect mutations and genetic load*

405 Presence and frequency of major effect mutations, i.e. loss of start and stop codons;
406 gain of stop codons and changes in splice sites were calculated per population using
407 snpEFF v4.2 (53). To avoid reference-biased inference of alternate alleles we
408 polarized the SNPs using the *A. montbretiana* genome assembly (ASM148412v1, 54)
409 as an outgroup (SI Appendix, SI text). As a proxy for genetic load, we estimated the
410 average number of derived nonsynonymous or major-effect homozygous genotypes
411 (41, 46), and the average number of derived alleles for nonsynonymous and major-
412 effect alleles (42) for each regional population. In addition, we counted the total
413 number of fixed derived nonsynonymous or major-effect alleles for each regional
414 population.

415

416 *Demographic modelling and simulations*

417 We conducted demographic inference in the software fastsimcoal2 v. 2.5.2.21 (55).
418 To estimate parameters associated with the origin of Scandinavian *A. alpina*, we
419 tested two demographic models (SI Appendix, Fig. S5), using two-dimensional joint
420 SFS (2D-SFS) based on a scattered sample from central Europe and the Scandinavian
421 population (SI Appendix, SI Text). We used 12,967 intergenic sites, a mutation rate of
422 7×10^{-9} and a generation time of 1.5 years. We used forward simulation in SLiM2
423 v2.1 (56) to assess the impact of demography and selection associated with a shift to

424 selfing on genetic diversity in the Scandinavian population under four demographic
425 models with varying bottleneck severity and population split time (SI Appendix, SI
426 Text).

427

428 **Acknowledgements**

429 TS thanks Cindy Canton for help with plant care and extractions. Computations were
430 performed on resources provided by SNIC through UPPMAX under Projects
431 snic2014-1-194, b2013022 and b2013237. This work was funded by grants from the
432 Swedish Research Council to JÅ and TS, by a grant from the DFG through SPP1529
433 to GC, and from SciLifeLab to TS.

434 **References**

- 435 1. Goodwillie C, Kalisz S, Eckert C (2005) The evolutionary enigma of mixed
436 mating systems in plants: Occurrence, theoretical explanations, and empirical
437 evidence. *Annu Rev Ecol Evol S* 36:47–79.
- 438 2. Igic B, Kohn JR (2006) The distribution of plant mating systems: study bias
439 against obligately outcrossing species. *Evolution* 60(5):1098–1103.
- 440 3. Igic B, Busch JW (2013) Is self-fertilization an evolutionary dead end? *New*
441 *Phytol* 198(2):386–397.
- 442 4. Lande R, Schemske D (1985) The evolution of self-fertilization and inbreeding
443 depression in plants. I. Genetic models. *Evolution* 39:24-40.
- 444 5. Hartfield M (2016) Evolutionary genetic consequences of facultative sex and
445 outcrossing. *J Evol Biol* 29(1):5–22.
- 446 6. Wright SI, Kalisz S, Slotte T (2013) Evolutionary consequences of self-
447 fertilization in plants. *Proc Biol Sci* 280(1760):20130133.
- 448 7. Barrett SCH, Arunkumar R, Wright SI (2014) The demography and population
449 genomics of evolutionary transitions to self-fertilization in plants. *Philos T Roy*
450 *Soc B* 369(1648). doi:10.1098/rstb.2013.0344.
- 451 8. Pollak E (1987) On the theory of partially inbreeding finite populations. I.
452 Partial selfing. *Genetics* 117(2):353–360.
- 453 9. Nordborg M (2000) Linkage disequilibrium, gene trees and selfing: an
454 ancestral recombination graph with partial self-fertilization. *Genetics*
455 154(2):923–929.
- 456 10. Ingvarsson PK (2002) A metapopulation perspective on genetic diversity and
457 differentiation in partially self-fertilizing plants. *Evolution* 56(12):2368–2373.
- 458 11. Charlesworth D, Wright SI (2001) Breeding systems and genome evolution.
459 *Curr Opin Genet Dev* 11(6):685–690.
- 460 12. Slotte T (2014) The impact of linked selection on plant genomic variation.
461 *Brief Funct Genomics* 13(4):268–275.
- 462 13. Kimura M (1983) *The neutral theory of molecular evolution* (Cambridge
463 University Press, Cambridge).
- 464 14. Ohta T (1973) Slightly deleterious mutant substitutions in evolution. *Nature*
465 246(5428):96–98.
- 466 15. Bataillon T, Kirkpatrick M (2000) Inbreeding depression due to mildly
467 deleterious mutations in finite populations: size does matter. *Genet Res*

- 468 75(1):75–81.
- 469 16. Glémin S (2003) How are deleterious mutations purged? Drift versus
470 nonrandom mating. *Evolution* 57(12):2678–2687.
- 471 17. Glémin S (2007) Mating systems and the efficacy of selection at the molecular
472 level. *Genetics* 177(2):905–916.
- 473 18. Qiu S, Zeng K, Slotte T, Wright S, Charlesworth D (2011) Reduced efficacy of
474 natural selection on codon usage bias in selfing *Arabidopsis* and *Capsella*
475 species. *Genome Biol Evol* 3:868–880.
- 476 19. Ness RW, Siol M, Barrett SCH (2012) Genomic consequences of transitions
477 from cross- to self-fertilization on the efficacy of selection in three
478 independently derived selfing plants. *BMC Genomics* 13:611.
- 479 20. Brandvain Y, Slotte T, Hazzouri KM, Wright SI, Coop G (2013) Genomic
480 identification of founding haplotypes reveals the history of the selfing species
481 *Capsella rubella*. *PLoS Genet* 9(9):e1003754.
- 482 21. Slotte T, et al. (2013) The *Capsella rubella* genome and the genomic
483 consequences of rapid mating system evolution. *Nat Genet* 45(7):831–835.
- 484 22. Douglas GM, et al. (2015) Hybrid origins and the earliest stages of
485 diploidization in the highly successful recent polyploid *Capsella bursa-pastoris*.
486 *Proc Natl Acad Sci USA* 112(9):2806–2811.
- 487 23. Kamran-Disfani A, Agrawal AF (2014) Selfing, adaptation and background
488 selection in finite populations. *J Evol Biol* 27(7):1360–1371.
- 489 24. Salcedo A, Kalisz S, Wright SI (2014) Limited genomic consequences of
490 mixed mating in the recently derived sister species pair, *Collinsia concolor* and
491 *Collinsia parryi*. *J Evol Biol* 27(7):1400–1412.
- 492 25. Tedder A, Ansell SW, Lao X, Vogel JC, Mable BK (2011) Sporophytic self-
493 incompatibility genes and mating system variation in *Arabis alpina*. *Ann Bot-*
494 *London* 108(4):699–713.
- 495 26. Ansell SW, Grundmann M, Russell SJ, Schneider H, Vogel JC (2008) Genetic
496 discontinuity, breeding-system change and population history of *Arabis alpina*
497 in the Italian Peninsula and adjacent Alps. *Mol Ecol* 17(9):2245–2257.
- 498 27. Buehler D, Graf R, Holderegger R, Gugerli F (2012) Contemporary gene flow
499 and mating system of *Arabis alpina* in a Central European alpine landscape.
500 *Ann Bot-London* 109(7):1359–1367.
- 501 28. Koch MA, et al. (2006) Three times out of Asia Minor: the phylogeography of

- 502 *Arabis alpina* L. (Brassicaceae). *Mol Ecol* 15(3):825–839.
- 503 29. Ehrlich D, et al. (2007) Genetic consequences of Pleistocene range shifts:
504 contrast between the Arctic, the Alps and the East African mountains. *Mol Ecol*
505 16(12):2542–2559.
- 506 30. Ansell SW, et al. (2011) The importance of Anatolian mountains as the cradle
507 of global diversity in *Arabis alpina*, a key arctic-alpine species. *Ann Bot-*
508 *London* 108(2):241–252.
- 509 31. Willing E-M, et al. (2015) Genome expansion of *Arabis alpina* linked with
510 retrotransposition and reduced symmetric DNA methylation. *Nature Plants* 1
511 SP - EP -(2):14023.
- 512 32. Morton NE, Crow JF, Muller HJ, (1956) An estimate of the mutational damage
513 in man from data on consanguineous marriages. *Proc Natl Acad Sci USA*
514 42:855–863.
- 515 33. Kimura M, Maruyama T, Crow JF (1963) Mutation load in small populations.
516 *Genetics* 48:1303–1312.
- 517 34. Raj A, Stephens M, Pritchard JK (2014) fastSTRUCTURE: variational
518 inference of population structure in large SNP data sets. *Genetics* 197(2):573–
519 589.
- 520 35. Caye K, Deist TM, Martins H, Michel O, François O (2016) TESS3: fast
521 inference of spatial population structure and genome scans for selection. *Mol*
522 *Ecol Resour* 16(2):540–548.
- 523 36. Gibson J, Morton NE, Collins A (2006) Extended tracts of homozygosity in
524 outbred human populations. *Hum Mol Genet* 15:789-795.
- 525 37. Keightley PD, Eyre-Walker A (2007) Joint inference of the distribution of
526 fitness effects of deleterious mutations and population demography based on
527 nucleotide polymorphism frequencies. *Genetics* 177(4):2251–2261.
- 528 38. Arunkumar R, Ness RW, Wright SI, Barrett SCH (2014) The evolution of
529 selfing is accompanied by reduced efficacy of selection and purging of
530 deleterious mutations. *Genetics* 199(3):817–829.
- 531 39. Gravel S (2016) When is selection effective? *Genetics* 203(1):451–462.
- 532 40. Brandvain Y, Wright SI (2016) The limits of natural selection in a
533 nonequilibrium world. *Trends Genet* 32(4):201–210.
- 534 41. Pedersen C-ET, et al. (2017) The effect of an extreme and prolonged
535 population bottleneck on patterns of deleterious variation: insights from the

- 536 greenlandic inuit. *Genetics* 205(2):787-801.
- 537 42. Henn BM, Botigué LR, Bustamante CD, Clark AG, Gravel S (2015)
538 Estimating the mutation load in human genomes. *Nat Rev Genet* 16(6):333–343.
- 539 43. Messer PW, Petrov DA (2013) Frequent adaptation and the McDonald-
540 Kreitman test. *Proc Natl Acad Sci USA*. 110(21):8615-8620.
- 541 44. Hamrick JL, Godt MJW (1996) Effects of life history traits on genetic diversity
542 in plant species. *Phil Trans Roy Soc B* 351(1345):1291–1298.
- 543 45. Peischl S, Dupanloup I, Kirkpatrick M, Excoffier L (2013) On the
544 accumulation of deleterious mutations during range expansions. *Mol Ecol*
545 22(24):5972–5982.
- 546 46. Henn BM, et al. (2016) Distance from sub-Saharan Africa predicts mutational
547 load in diverse human genomes. *Proc Natl Acad Sci USA* 113(4):E440–9.
- 548 47. Lohmueller KE, et al. (2008) Proportionally more deleterious genetic variation
549 in European than in African populations. *Nature* 451(7181):994–997.
- 550 48. Simons YB, Turchin MC, Pritchard JK, Sella G (2014) The deleterious
551 mutation load is insensitive to recent population history. *Nat Genet* 46(3):220–
552 224.
- 553 49. Toräng P, et al. (2015) Large-scale adaptive differentiation in the alpine
554 perennial herb *Arabis alpina*. *New Phytol* 206(1):459–470.
- 555 50. Gilbert KJ, et al. (2017) Local adaptation interacts with expansion load during
556 range expansion: maladaptation reduces expansion load. *Am Nat* 189:368-380.
- 557 51. Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-
558 Wheeler transform. *Bioinformatics* 25(14):1754–1760.
- 559 52. McKenna A, et al. (2010) The Genome Analysis Toolkit: a MapReduce
560 framework for analyzing next-generation DNA sequencing data. *Genome Res*
561 20(9):1297–1303.
- 562 53. Cingolani P, et al. (2012) A program for annotating and predicting the effects
563 of single nucleotide polymorphisms, SnpEff: SNPs in the genome of
564 *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly* 6(2):80–92.
- 565 54. Kiefer C, et al. (2017) Divergence of annual and perennial species in the
566 Brassicaceae and the contribution of *cis*-acting variation at *FLC* orthologues.
567 *Mol Ecol* In press.
- 568 55. Excoffier L, et al. (2013) Robust demographic inference from genomic and
569 SNP data. *PLoS Genet* 9(10):e1003905.

- 570 56. Haller BC, Messer PW (2017) SLiM 2: Flexible, interactive forward genetic
571 simulations. *Mol Biol Evol* 34(1):230–240.

572 **Tables**

573

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

Population ¹	S_S ²	π_S ³	π_N/π_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

575 ¹Regional population (*n*)

576 ²Segregating four-fold synonymous SNPs.

577 ³Mean four-fold synonymous nucleotide diversity.

578 ⁴Mean 0-fold degenerate nonsynonymous/4-fold degenerate synonymous
579 polymorphism.

580 **Figure Legends**

581

582 Figure 1. Bayesian clustering analysis supports a strong geographic component to
583 population structure in European *A. alpina*. A. Ancestry proportions for K=5
584 correspond closely to geographical sampling locations. B. Geographic interpolation of
585 genetic structure across Europe based on TESS3 results for K=5.

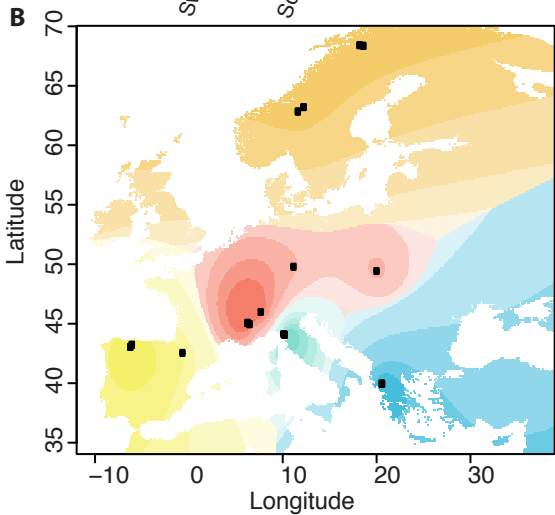
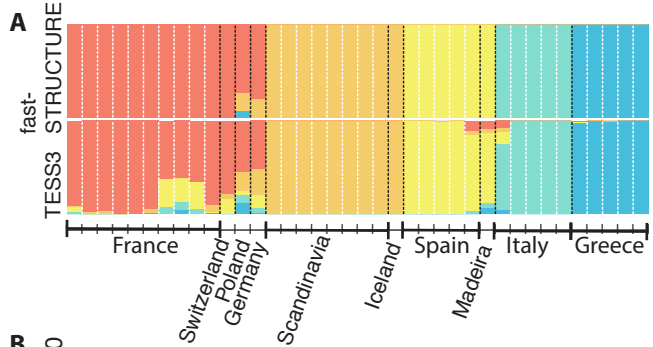
586

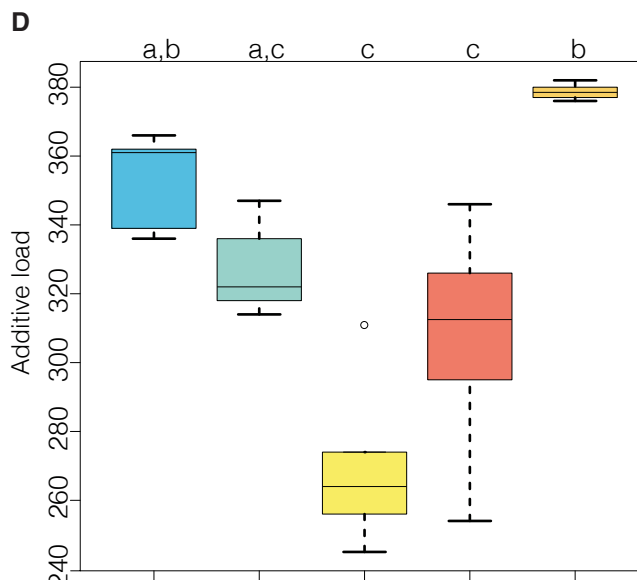
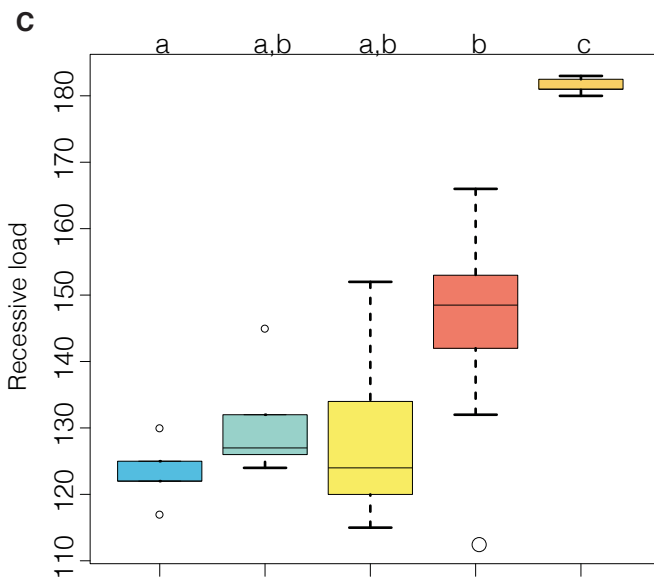
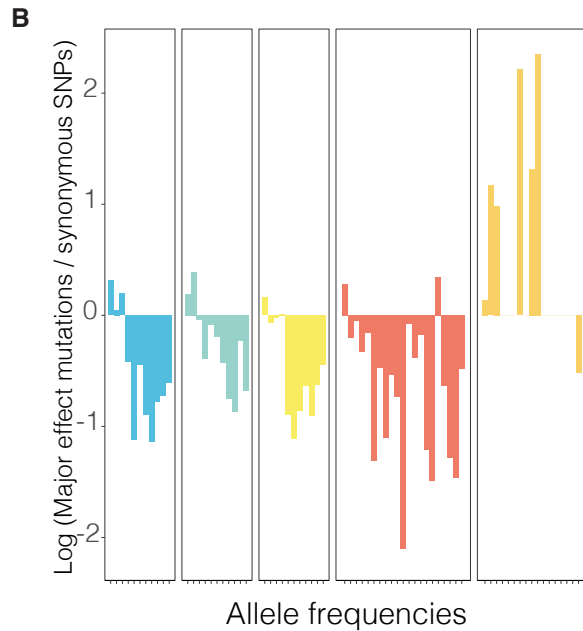
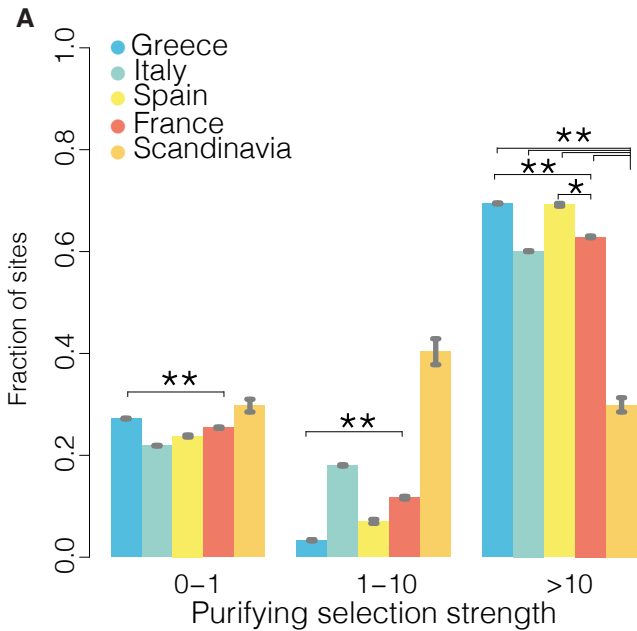
587 Figure 2. The impact of mating system on purifying selection in *A. alpina*. A. The
588 distribution of fitness effects (DFE) in bins of $N_e s$ (the product of the effective
589 population size and the selection coefficient) for new nonsynonymous mutations,
590 estimated under model with a stepwise population size change. Error bars show ± 1
591 standard error. Asterisks indicate significant differences (FDR<0.05) among
592 populations. B. Scandinavian *A. alpina* show an increase in the frequency of derived
593 major effect polymorphism relative to synonymous polymorphism. The figure shows
594 the log ratio of major effect derived allele frequencies to 4-fold synonymous allele
595 frequencies. C. Boxplots of the recessive genetic load for major effect alleles. Letters
596 indicate groups with statistically significant differences ($P<0.05$, Kruskal-Wallis test,
597 post-hoc Dunn test). D. Boxplots of the additive genetic load for major effect alleles.
598 Significance is indicated as in C.

599

600 Figure. 3. A recent bottleneck and selfing explain the reduction of polymorphism in
601 Scandinavia. A. Schematic showing the best-fit demographic model of the
602 colonization of Scandinavia from an ancestral Central European population.
603 Estimated times are given in years before present (ybp), assuming a generation time
604 of 1.5 years. B. Background selection alone does not explain the reduction in diversity
605 in Scandinavian *A. alpina*. Boxplots show the ratio of synonymous polymorphism
606 between an outcrossing population and a 90% selfing population experiencing either a
607 constant population size or a 10-fold bottleneck, with the two populations diverging
608 either 12,000 ybp or 20,208 ybp. The dashed line indicates the observed ratio of
609 synonymous polymorphism in a scattered sample from Central Europe to that in the
610 Scandinavian regional population. Letters at the top indicate significant difference
611 between models (Mann-Whitney test $P<0.001$). Asterisks represent the significance

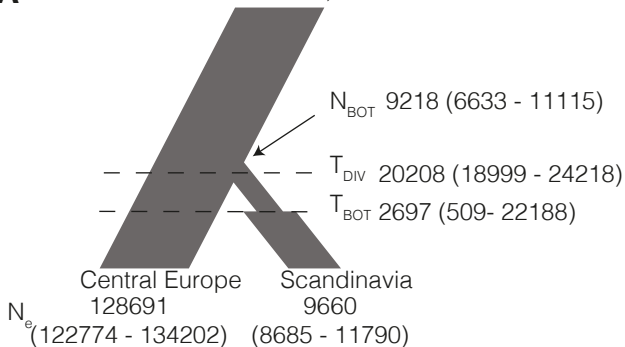
612 level of a test of whether the observed neutral diversity reduction is greater than
613 expected based on 300 simulations.





A

AIC=98506, w=1

**B**