A polygenic architecture with conditionally neutral effects underlies ecological differentiation in *Silene*

Susanne Gramlich[1], Xiaodong Liu[1,2], Adrien Favre[3], C. Alex Buerkle[4] and Sophie Karrenberg[1*]

[1] Department of Ecology and Genetics, Uppsala University, Norbyvägen 18D, 75267 Uppsala, Sweden

[2] The Bioinformatics Center, Department of Biology, University of Copenhagen, Ole Maaløes Vej 5, 2200, Copenhagen, Denmark

[3] Senckenberg Research Institute and Natural History Museum, Senckenberganlage 25, 60325 Frankfurt/Main, Germany

[4] Department of Botany, University of Wyoming, 1000 E. University Ave. Laramie, WY82071, USA

* corresponding author: Sophie Karrenberg, sophie.karrenberg@ebc.uu.se, +46 (0)76 820 7578

Word count

Summary: 196 words

Total (main text): 5569 words

Introduction: 933 words

Materials and Methods: 1820 words

Results: 1026 words

Discussion: 1679 words

Acknowledgements: 111

Number of Figures: 5 (all in color)

Number of Tables: 1

Number Supporting Information files: 2 (Figures S1-S8, Tables S1-S4)

Summary

- Ecological differentiation can drive speciation but it is unclear how the genetic architecture of habitat-dependent fitness contributes to lineage divergence. We investigated the genetic architecture of cumulative flowering, a fitness component, in second-generation hybrids between *Silene dioica* and *S. latifolia* transplanted into the natural habitat of each species.

- We used reduced-representation sequencing and Bayesian Sparse Linear Mixed Models (BSLMMs) to analyze the genetic control of cumulative flowering in each habitat.

- Our results point to a polygenic architecture of cumulative flowering. Allelic effects were mostly beneficial or deleterious in one habitat and neutral in the other. The direction of allelic effects was associated with allele frequency differences between the species: positive-effect alleles were often derived from the native species, whereas negative-effect alleles, at other loci, tended to originate from the non-native species.

- We conclude that ecological differentiation is governed and maintained by many loci with small, conditionally neutral effects. Conditional neutrality may result from differences in selection targets in the two habitats and provides hidden variation upon which selection can act. Polygenic architectures of adaptive differentiation are expected to be transient during lineage divergence and may therefore be unrelated to high genetic differentiation at the underlying loci.

Key words: adaptation, Bayesian Sparse Linear Mixed Models (BSLMM), conditional neutrality, ddRAD-Seq, reproductive isolation, speciation, *Silene*

Introduction

Adaptation to different habitats can promote divergence and speciation and can mediate population persistence under changing conditions (Nosil, 2012; Savolainen *et al.*, 2013). Evolutionary trajectories toward adaptive differentiation depend on the number and effect sizes of loci controlling fitness, on allelic effects of such loci in alternative habitats and on gene flow (Savolainen *et al.*, 2013; Tigano & Friesen, 2016; Kokko *et al.*, 2017). However, despite recent progress in the theoretical understanding of adaptation (Yeaman, 2015; Mee & Yeaman, 2019; Booker *et al.*, 2021), empirical studies on the genetic control of fitness in natural habitats is still scarce (Wadgymar *et al.*, 2017). Here we provide such a study using two ecologically differentiated but hybridizing campions (*Silene*).

Adaptive differentiation is often depicted as a classic cartoon with populations outperforming other populations in their home habitat but not in the foreign habitat (Kawecki & Ebert, 2004). A similar pattern has been proposed to operate at the genetic level: alleles with positive effects on fitness in one habitat have negative effects in the alternative habitat, termed antagonistic pleiotropy (Anderson *et al.*, 2011; Savolainen *et al.*, 2013). Empirical data, however, suggests that such antagonistic pleiotropy is rare in natural populations, although this could partially be caused by limited statistical power (Anderson *et al.*, 2011; Anderson *et al.*, 2014; Wadgymar *et al.*, 2017). An alternative pattern, conditional neutrality, where alleles have neutral effects in one environment and are under positive or negative selection in the other, may be much more prevalent (Anderson *et al.*, 2011; Savolainen *et al.*, 2013; Wadgymar *et al.*, 2017). In fact, even at the phenotypic level, populations often perform differently in one habitat but not in the other (Leimu & Fischer, 2008). Importantly, loci with antagonistic pleiotropy for fitness in alternative habitats are expected to be maintained by natural selection and contribute to population differentiation even with high gene flow, whereas conditionally neutral loci can only lead to transient allele frequency differences in high-gene flow scenarios (Mitchell-Olds *et al.*, 2007; Savolainen *et al.*, 2013; Tiffin & Ross-Ibarra, 2014; Mee & Yeaman, 2019; Booker *et al.*, 2021). If adaptive allele frequency changes are predominantly transient, even in the presence of consistent selection, prospects will be poor for finding loci underlying adaptive differentiation with commonly used methods for quantifying and comparing genetic differentiation along the genome (i.e., genome scans of differentiation).

A further major determinant of adaptive differentiation concerns the number and effect sizes of the underlying loci. In general, the distribution of effect sizes for complex traits (including fitness) is expected to be exponential (Orr, 1998, 2005) such that adaptation is mainly due to many loci with small effects, while large-effect loci, although not generally unimportant, are rare (Orr, 1998; Orr, 2005; Rockman, 2012; Savolainen *et al.*, 2013; Boyle *et al.*, 2017; Selby & Willis, 2018). A large number of segregating loci underlying adaptation further promotes transient genetic architectures of adaptation (Yeaman & Whitlock, 2011; Yeaman, 2015; Booker *et al.*, 2021). Polygenic or even omnigenic genetic architectures, involving most of the genome, render identification of all individual loci both practically impossible and undesirable (Rockman, 2012). As effect sizes decrease and the number of loci increases, it becomes progressively more difficult to detect individual genotype-phenotype associations and to distinguish them from spurious associations due to linkage. For this reason, promising methods, such as Bayesian Sparse Linear Mixed Models (BSLMMs), identify genotype-phenotype associations of all loci simultaneously rather than individually, assume polygenic genetic architectures and remove effects that can be attributed to linkage disequilibrium (Zhou *et al.*, 2013; Gompert *et al.*, 2017). The approach has been successfully applied in studies on adaptive divergence, for example in pine (Lind *et al.*, 2017) and in *Arabidopsis* (Exposito-Alonso *et al.*, 2019) .

In this study, we investigate the genetic basis of differential adaptation in two dioecious sister species of *Silene* (Caryophyllaceae), *Silene dioica* (L.) Clairv. and *S. latifolia* Poiret. Demographic models indicate that the two species diverged with gene flow within the last 120 000 years reaching a neutral sequence divergence (D_a, autosomes) of 0.0027 and genetic differentiation, F_{ST}, of 0.28 (Hu & Filatov, 2015; Liu *et al.*, 2020). The pink-flowered *S. dioica* occurs in moister and colder habitats like meadows, pastures or forests and occupies a wide range of elevations up to more than 2300 m a. s. l., while the white-flowered *S. latifolia* is found on drier, warmer and more disturbed sites like dry meadows, arable fields or road sites at elevations up to 1000 m a. s. l. (Friedrich, 1979; Karrenberg & Favre, 2008). Ecological differentiation and assortative pollination constitute strong barriers to gene flow in this species pair, but the two species are still fully cross-fertile (Goulson & Jerrim, 1997; Karrenberg *et al.*, 2019). Studies in an experimental garden point to a complex, genome wide genetic architecture of traits associated with reproductive isolation, such as flower color, flowering

120 phenology, first-year flowering and specific leaf area in this system (Liu & Karrenberg,
121 2018).

122 Here we investigated the genetic architecture of adaptation using recombinant second-
123 generation hybrids (F_2) from a multi-site field transplant experiment where the two
124 species exhibited strong evidence of habitat adaptation: each species outperformed the
125 other in its own habitat in terms of flowering and survival over four years (Fig. 1, Favre
126 *et al.*, 2017). We focused on the following questions: (1) What is the genetic
127 architecture underlying differential habitat adaptation? (2) Is antagonistic pleiotropy or
128 conditional neutrality the predominant pattern when comparing allelic effects across
129 habitats? and (3) Are fitness effects associated with allele frequency differences
130 between the two species; for example, are beneficial alleles more likely to be derived
131 from the native species?

132 ## Materials and Methods

133 *Transplant experiment, measurements and sampling*

134 Second-generation (F_2) hybrids between *Silene dioica* (L.) Clairv. and *Silene latifolia*
135 Poir., derived from 36 F_0 individuals of three populations of each species, were
136 transplanted as juveniles into four natural sites, two in each species's habitat, as part of a
137 larger experiment with six sites (Favre *et al.*, 2017, Supporting Information Tables S1,
138 S2, and S3). Sites of *S. dioica* were situated at higher altitudes with a colder climate
139 and shorter growing season as compared to the *S. latifolia* sites (Supporting Information
140 Table S1). Leaf samples were collected prior to transplantation and silica-dried and we
141 selected four of the six sites for this study based on availability and quality of leaf
142 samples. Flowering and survival were assessed over four years and cumulative
143 flowering was calculated as the number of times an individual flowered plus 1 if it
144 survived to the end of the experiment, for detailed analyses of survival and flowering
145 see Favre et al. (2017). We sampled 4-6 F_2 individuals from each of 18 F_2 families at
146 each of the four selected sites (298 F_2 individuals in total), striving to include both high
147 and low fitness individuals from each family and site. To assess allele frequency in the
148 two species we further included samples of 32 F_0 individuals and used existing data for
149 the remaining four F_0 individuals (Liu & Karrenberg, 2018, Supporting Information
150 Table S3).

For illustration purposes (Fig. 1), we re-analyzed cumulative flowering in the two species and their first- and second-generation hybrids for the four sites used here, (18-36 families per cross type, 5-20 individuals per family, 5 blocks per site, Favre *et al.* 2017, Supporting Information Table S2). We used linear mixed models of cumulative flowering in each habitat with cross type as a fixed factor and family and block nested in site as random factors using the *lme4* package (Bates *et al.*, 2015) for R version 4.02 (R Core Team, 2020). We extracted least square means of cumulative flowering and performed multiple comparisons between cross types within sites with Holm correction of P-values in the *emmeans* package (Lenth, 2020). Cumulative flowering was log(Y+1)-transformed to yield normally distributed residuals and improve model fit. Means and standard errors are reported back-transformed to the original scale.

DNA extraction and sequencing

We extracted genomic DNA from silica-dried leaf tissue with Qiagen's DNeasy Plant Mini Kit (Qiagen, Germany) and quantified DNA using a Qubit dsDNA HS fluorometer (Life Technologies, Sweden). Double-digest RAD sequencing (ddRAD-seq) libraries were prepared with EcoRI and TaqαI restriction enzymes as described in Liu and Karrenberg (2018). After enzymatic digestion, DNA fragments were ligated with barcoded adaptors and size-selected to approximately 550 bp (Peterson *et al.*, 2012). In total, eight 48-plex libraries were sequenced on an Illumina HiSeq 2500 system at the SNP&SEQ technology platform of SciLifeLab, Uppsala, Sweden using 125-bp paired-end chemistry and two libraries per lane. F_0 individuals were included in 2 libraries to achieve higher coverage.

Bioinformatic analysis - processing of raw reads and variant filtering

The total sequencing output was 1,382,838,294 reads for 298 F_2 individuals (mean with one standard error: 4,656,021 ± 233,639 reads) and 461,127,142 reads for 32 F_0 individuals (mean: 14,410,223 ±1,093,409 reads); data on the remaining four F_0 individuals (Supplementary Information Table S3) was available from (Liu & Karrenberg, 2018). We processed the ddRAD-sequence reads following the dDocent pipeline (Puritz et al., 2014). After de-multiplexing of raw reads using STACKS 2.0b (Catchen *et al.*, 2013) and trimming with *fastp* (Chen *et al.*, 2018), we used BWA MEM 0.7.17 with default parameters (Li, 2013) to map reads to ddRAD-seq-generated

reference contigs, which were previously assembled from eight deeply sequenced individuals of both species and hybrids and corresponded to 95,040,562 bp in total, corresponding to approximately 3.4% of the *S. latifolia* genome (Liu & Karrenberg, 2018; Liu *et al.*, 2020). Only a partial genome sequence (one third of the 2.8 Gbp genome) with short scaffolds (N_{50} = 10,785 bp) is currently available for *S. latifolia* (Krasovec *et al.*, 2018).

Variants were called with FreeBayes 1.1.0 (Garrison & Marth, 2012) without population priors using the following parameters: minimum mapping quality 30, minimum base quality 20, maximum complex gap 3, minimum repeat entropy 1, binominal-obs-priors 1, and use-best-n-alleles 10. Variants were filtered following O'Leary *et al.* (2018): First, VCFtools 0.1.15 (Danecek *et al.*, 2011) was used to retain SNP sites with a minimum depth of 3, quality of 30, mean depth of 10 and allele count of 3. Secondly, we used *Vcffilter* implemented in vcflib/2017-04-04 (https://github.com/vcflib/vcflib) to retain sites with an allele balance either between 0.25 and 0.75 or lower than 0.01, a quality/depth ratio of > 0.25, and a mapping quality ratio between 0.9 and 1.05. We further used *Vcffilter* to remove loci with differences in read pairing between the alleles or with or excessive read depths (parameters as suggested in O'Leary *et al.*, 2018). We reduced the dataset to bi-allelic sites and removed 6 F_2 individuals with more than 99% missing data. SNPs in perfect linkage disequilibrium (r^2 = 1) within the F_2 individuals were removed using PLINK 1.9 (Purcell *et al.*, 2007, http://pngu.mgh.harvard.edu/purcell/plink/). The filtered dataset contained 290 F_2 individuals (*S. dioica*-habitat: 134, *S. latifolia*-habitat: 156) and 89'524 loci. Of these, 42'090 loci with both alleles in both habitats and genotypes for more than 95 F_2 individuals in each habitat were used for further analyses. These loci had an average read depth of 15.02 ± 0.03 (median 13.83) for the F_2 individuals, 36.85 ± 0.08 (median: 33.09) for the 32 F_0 individuals sequenced in this study and 41.39 ± 0.17 (median: 36.25) for the 4 F_0 individuals from Liu and Karrenberg (2018). We used genotype probabilities ranging from 0 to 2, calculated from genotype likelihoods (https:// github.com/ visoca/popgenomworkshop-gwas_gemma/tree/master/scripts/bcf2bbgeno.pl, accessed 28 January 2019), rather than called genotypes (Nielsen *et al.*, 2011). Genotype probabilities of 0 and 2 denote homozygosity for the reference and alternative allele, respectively, while 1 indicates heterozygosity.

Genetic structure

We checked for genetic structure in the F_2 individuals using Principal Components Analysis (PCA) with the *ade4* package (Dray and Dufour, 2007), based on genotype probabilities. For this analysis we reduced the dataset to 220 SNP loci with data available in all 290 F_2 individuals used in the genetic association analysis. Results did not differ from analyses on more or all loci with missing values replaced by average genotype probabilities.

Genetic association analysis

We used Bayesian Sparse Linear Mixed Models (BSLMM) in GEMMA 0.98.1 (Zhou *et al.*, 2013) to investigate the genetic architecture of cumulative flowering. BSLMMs are a combination of linear mixed models, which assume that every variant has an effect, and Bayesian Variable Selection Regression, which assumes that only a small proportion of the variants has an effect (Zhou *et al.*, 2013). BSLMM analyses were performed separately for each habitat using cumulative flowering values standardized within sites and genotype probabilities. A centered relatedness matrix was used as a covariate to take account of the family structure and the wide cross, according to the standard BSLMM method implemented in GEMMA (Zhou *et al.*, 2013).

We characterized genetic architectures using estimates of the following hyperparameters: the proportion of phenotypic variance explained by all SNPs in the model (*PVE*), the proportion of *PVE* explained by SNPs with nonzero effects (*PGE*), and the number of SNPs with a measurable effect on the phenotype (n-γ). In studies with incomplete genome coverage such as this one, *PVE* can be interpreted as broad sense heritability H^2, whereas the product of *PVE* and *PGE* can be interpreted as narrow-sense heritability h^2 (Zhou *et al.*, 2013; Gompert *et al.*, 2017; Bresadola *et al.*, 2019). For each SNP locus, we estimated the posterior inclusion probability (*PIP*, the proportion of iterations in which a SNP had a nonzero effect on phenotypic variation) and the effect size of the alternative allele on cumulative flowering. We report both raw effect estimates ($\hat{\beta}$), the effect of a locus on the phenotype when it is included in the model, and model-averaged effect estimates ($\bar{\beta}$) that take into account the posterior inclusion probability of a locus in the models (Zhou *et al.*, 2013; Gompert *et al.*, 2017). BSLMMs were run five times with 10,000,000 burn-in steps and 40,000,000 iterations with a thinning interval of 10. Convergence of the five runs was checked graphically.

Hyperparameters were estimated after combining posterior distributions across runs; *PIP*, $\hat{\beta}$ and $\bar{\beta}$ (per locus) were averaged across runs. A threshold of $PIP > 0.01$ was used to identify SNPs with sparse effects (Gompert *et al.*, 2013; Comeault *et al.*, 2014).

Allelic effects and allele frequencies

We graphically evaluated whether alleles have universal effects or display antagonistic pleiotropy or conditional neutrality by plotting per-locus effect sizes ($\hat{\beta}$ and $\bar{\beta}$) in the *S. dioica* habitat against effect sizes in the *S. latifolia* habitat. Note that β-values reflect effect estimates of the alternative allele replacing the reference allele. The *de novo* ddRADseq reference sequences used here contain alleles derived from both species (Liu & Karrenberg, 2018; Liu *et al.*, 2020) and do not inform on allelic origin.

To analyze whether allelic origin is associated with fitness effects, we calculated the frequency of the alternative allele (AF_{alt}) in the 18 F_0 individuals of each species (Tables S2 and S3) for loci that had genotypes for > 12 F_0 individuals per species at individual read depths of at least 6 (36,161 loci). For each SNP locus, we expressed AF_{alt} as the average genotype probability per species divided by 2. We further estimated the allele frequency difference between species (AFD_{alt}) as AF_{alt} (*S. dioica*) - AF_{alt} (*S. latifolia*). AFD_{alt} values thus range from -1 (alternative allele fixed in *S. latifolia* F_0 individuals, and reference allele fixed in *S. dioica* F_0 individuals) to 1 (alternative allele fixed in *S. dioica* F_0 individuals, and reference allele fixed in *S. latifolia* F_0 individuals). We tested whether AFD_{alt} differs between loci with positive and negative effects on cumulative flowering in each habitat using a general permutation test in the R package *coin* (Hothorn et al., 2008).

Method validation

We assessed the power of BSLMMs to detect genotype - phenotype associations in our data using a simulation approach similar to that in Gompert *et al.* (2017). Phenotypes were simulated on the basis of the observed genetic data in 290 individuals for four combinations of heritability ($h^2 = 0.2$ or $h^2 = 0.05$) and number of functional variants ($N = 10$ or $N = 50$). We simulated a normally distributed trait for which nine individuals of each family were randomly drawn. We reduced the simulated dataset to 150 individuals at random to match the sample size in our data. For each combination of h^2 and n 30 sets of simulated phenotypes were used as input for BSLMMs with

10,000,000 burn-in steps and 40,000,000 iterations. For each simulation, *PVE*, *PGE* and *n-γ* were estimated as well as the correlation between the simulated effect size and effect size estimates ($\hat{\beta}$ and $\bar{\beta}$).

Results

Cumulative flowering and genetic structure

Each species outperformed the other in its own habitat and performed best in its habitat in terms of cumulative flowering, providing strong evidence for habitat adaptation (Fig. 1, Favre *et al.*, 2017). F_1 hybrids were intermediate for cumulative flowering and F_2 hybrids flowered less often than F_1 hybrids on average (Fig. 1, Favre *et al.*, 2017).

The F_2 families exhibited extensive genetic variation in the PCA, with clustering of individuals within families but without any other emerging patterns (Supplementary Information Figs. S1 and S2). For most loci, the alternative allele occurred at low frequency in one or both species, but all combinations of allele frequencies between species were found in the data, including loci with high differentiation where the alternative allele was at high frequency in one species and at the same time near absent from the other species (Fig. 2).

Association analysis

The proportion of phenotypic variation explained (*PVE*) by BSLMMs was 0.12 and 0.10 for the *S. dioica* and the *S. latifolia* habitat, respectively (medians of posterior distributions, Table 1, posterior distributions are shown in the Supporting Information Figs. S3 and S4). Less than half of the *PVE* could be attributed to sparse effects, leading to estimates of narrow-sense heritability (h^2; *PGE* x *PVE*) of 0.05 and 0.03 for the *S. dioica* and *S. latifolia* habitat (Table 1). The number of loci with non-zero effects (*n-γ*) was estimated to be 11 in the *S. dioica* habitat and 16 in the *S. latifolia* habitat (Table 1). Median *PIP* for individual SNPs was 0.001 in both habitats and we detected 3 loci with *PIP* > 0.01, one in the *S. dioica* habitat and two in the *S. latifolia* habitat (Supporting information Table S4). *PIP* for individual SNPs could not be summed over genomic windows because no continuous reference genome is available (Krasovec *et*

al., 2018) and we used a ddRADseq-generated *de novo* reference in this study (Liu & Karrenberg, 2018; Liu *et al.*, 2020).

Allele specific effects in alternative habitats and origin of alleles

Raw effect size estimates $\hat{\beta}$ for cumulative flowering (in units of standard deviations, i.e., z-scores) ranged from -0.57 to 0.28 in the *S. dioica* habitat and from -0.52 to 0.34 in the *S. latifolia* habitat. Positive or negative allelic effects in one habitat ($\hat{\beta} >> 0$ or $\hat{\beta} << 0$) were associated with very low $\hat{\beta}$ values ($\hat{\beta} \approx 0$) in the other habitat, consistent with conditional neutrality (Fig. 3, cross pattern). A single locus showed opposing effects in the two habitats, $\hat{\beta} << 0$ in the *S. dioica* habitat and $\hat{\beta} >> 0$ in the *S. latifolia* habitat, suggestive of antagonistic pleiotropy (Fig. 3). Analyses for model-averaged effect sizes, $\bar{\beta}$, exhibited a very similar pattern (Supporting information Fig. S5), but with much smaller values due to multiplication with low posterior inclusion probabilities.

A striking pattern in our data is that alleles with effects on cumulative flowering were significantly associated with higher allele frequencies in F_0 individuals of the native species whereas alleles with negative effects on cumulative flowering were significantly more likely to be derived from the non-native species (Fig. 3, Supporting Information Figs. S5, S6 and S7). In the *S. dioica* habitat, loci with positive effects of the alternative allele on cumulative flowering had a positive median AFD_{alt} (alternative allele more common in *S. dioica*, red on Fig. 3), that were significantly different from the negative median AFD_{alt} (alternative allele more common in *S. latifolia*, blue on Fig. 3) at negative-effect alleles (Supporting Information Figs. S6 and S7). In the *S. latifolia* habitat, we observed the reverse pattern: loci with positive effects had negative median AFD_{alt} values that differed significantly from the positive median AFD_{alt} values for negative-effect loci (Supporting information Figs. S6 and S7). These patterns were observed for both $\hat{\beta}$ and $\bar{\beta}$, when using a subset of the approximately 0.5 % loci with strongest effect sizes (absolute values of $\hat{\beta}$ over 0.1 and absolute values of $\bar{\beta}$ over 0.003) and when using all loci (i.e., negative-effect loci with $\hat{\beta}$ or $\bar{\beta} > 0$ and positive-effect loci with $\hat{\beta}$ or $\bar{\beta} < 0$, Supporting Information Figs. S6 and S7).

Alleles with positive effects on cumulative flowering in each habitat ($\hat{\beta} > 0.1$) were rare in the foreign species but at appreciable frequencies in the native species; this effect was stronger in the *S. latifolia* habitat than in the *S. dioica* habitat (Fig. 4). Loci with negative effects ($\hat{\beta} < -0.1$), in contrast, were rare in the native species but at moderate or high frequencies in the foreign species (Fig. 4). These allele frequency patterns across species differ strongly from the overall allele joint allele frequency pattern for all loci (Fig. 2c).

Method validation

BSLMM estimates of narrow-sense heritability (h^2) in simulated datasets responded to changes in simulated heritabilities, as expected (Fig. 5). BSLMMs on simulated data with a low h^2 of 0.05 recovered h^2 estimates in the correct range, while models on simulated data with a high h^2 of 0.2 yielded lower than expected h^2 estimates, especially when the number of sparse-effect loci was high (N = 50, Fig. 5). The number of sparse-effect loci was correctly estimated for simulated datasets with few loci (N = 10), but strongly underestimated when the simulated number of loci was high (Fig. 5). Correlations of simulated and estimated $\hat{\beta}$ and $\bar{\beta}$-values were highest for datasets with high simulated heritabilities and declined for datasets with lower heritability and a higher number of functional loci (Fig. 5, Supporting Information Fig. S8). In simulations with $h^2 = 0.02$, $\hat{\beta}$ exhibited higher correlations with simulated effects sizes than $\bar{\beta}$ (Supporting Information Fig. S8).

These simulations thus show that our analyses can recover heritabilities and general patterns in effect size, however, estimates of the number of sparse effect loci (n-γ) must be treated with great caution. The consistency between estimated and simulated h^2 for datasets with low simulated h^2 shows that the low h^2 on our empirical data are reliable, especially as such low h^2 estimates were very rare in the high h^2 simulations. We expect that the apparent underestimation of n-γ also decreases posterior inclusion probabilities and thus $\bar{\beta}$.

Discussion

Our results suggest that ecological differentiation between the campions *Silene dioica* and *S. latifolia* has a polygenic architecture, based on a reciprocal transplant experiment

366 with recombinant second-generation hybrids between the two species. Many alleles had
367 small effects on cumulative flowering, a fitness component, but only in one of the two
368 habitats, conforming to a conditional neutrality pattern. Despite small effect sizes, the
369 direction of allelic effects was consistent with selection for native alleles and against
370 non-native alleles. This suggests that habitat adaptation is maintained by species-
371 specific variants and that habitat dependent selection on hybrids reduces the inflow of
372 non-native alleles. Below we discuss the implications of these findings for lineage
373 divergence as well as the advantages and limitations of our approach.

374 *Low heritability of a fitness component in the field, based on many small-effect loci*

375 Heritability of cumulative flowering, a fitness component, was low to moderate in F_2
376 hybrids between *S. dioica* and *S. latifolia* in the habitats of both species. Using
377 Bayesian Sparse Linear Mixed Models (BSLMM) we estimated narrow-sense
378 heritability h^2 (*PVE* x *PGE*) to be 0.04 and 0.03 and broad-sense heritability H^2 (*PVE*)
379 to be 0.12 and 0.10 for *S. dioica* and *S. latifolia* habitats, respectively. Our simulations
380 indicate that such low h^2 values can be recovered fairly well with BSLMM analyses, as
381 also reported by Gompert *et al.* (2017). Similar low to moderate heritabilities have
382 previously been found for complex traits, for example for height in *Pinus albicaulis*
383 (Lind *et al.*, 2017) and in *Populus* (Bresadola *et al.*, 2019) as well as in A*rabidopsis*
384 *thaliana* accessions, where heritability for fitness components varied strongly between
385 experimental sites (Exposito-Alonso *et al.*, 2019). Our analyses suggest a polygenic
386 genetic architecture of cumulative flowering in *Silene* with estimates of the number of
387 sparse-effect loci estimated to 11 in the *S. dioica* habitat and to 16 in the *S. latifolia*
388 habitat; however, our simulations indicate that the number of sparse effect loci in our
389 dataset is difficult to estimate with BSLMM analyses, similar to results from Gompert
390 et al. (2017). In many other systems, oligo- or polygenic control of fitness components
391 was detected in natural settings, especially when the number of contributing loci was
392 explicitly modelled (Lind *et al.*, 2017; Bresadola *et al.*, 2019; Exposito-Alonso *et al.*,
393 2019). A polygenic genetic architecture of fitness components in our study system does
394 appear likely given that QTL studies detected many loci distributed throughout the
395 genome that were associated with ecologically relevant traits (Liu & Karrenberg, 2018;
396 Baena Díaz *et al.*, 2019).

Limitations of studies on the genetic architecture of fitness in natural habitats include the number of individuals and sites that can be studied as well as the genetic resolution in the experimental material. Where markers are used, as in our study, most associations will be the result of linkage to causal variants rather than due to causal variants themselves. This reduces estimates of effect sizes and may lead to a situation where multiple loci linked to the same causal variant are included in alternative BSLMM iterations reducing the posterior inclusion probability (*PIP*) of each linked locus (Bresadola *et al.*, 2019). We used a highly variable multi-cross F_2 population with a limited number of recombination events, where markers likely were associated with larger genomic regions derived from each species. Recombination and genetic variation included in our material differs markedly from other studies using largely homozygous selfing accessions (*Arabidopsis*, Exposito-Alonso *et al.*, 2019), within-species crosses or recombinant inbred lines derived from a small number of individuals (*Arabidopsis*, Fournier-Level *et al.*, 2013; Leinonen *et al.*, 2013; Ågren *et al.*, 2017) or naturally recombinant wild-collected hybrids (*Populus*, Bresadola *et al.*, 2019). These differences can make it difficult to compare results and model performance across studies. Despite its limitations, our setup has the advantage that the multi-cross F_2 population used here approximates a natural contact site situation, where early-generation hybrids grow in the habitats of the two species when they come into contact (Karrenberg & Favre, 2008).

Conditional neutrality is the prevailing pattern

Allelic effects conformed to a conditional neutrality pattern in our study: both positive and negative allelic effects on cumulative flowering in one habitat were associated with near-zero effects in the other habitat; only one locus deviated from this pattern. This finding further strengthens the view that conditional neutrality is more common in natural settings than antagonistic pleiotropy, as has been shown in monkeyflowers (Hall *et al.*, 2010) and *Arabidopsis* (Fournier-Level *et al.*, 2013; Ågren *et al.*, 2017; Exposito-Alonso *et al.*, 2019), switchgrass (Lowry *et al.*, 2019) and *Lycaeides* butterflies (Gompert *et al.*, 2015). However, strong evidence for conditional neutrality is generally difficult to provide, because this would require evidence for the absence of an effect in one of the habitats (Anderson *et al.*, 2014; Mee & Yeaman, 2019). Interestingly, antagonistic pleiotropy appears to be detected more readily in controlled experimental evolution studies with microorganisms than in natural settings with higher plants or

insects (Gompert & Messina, 2016; Bono *et al.*, 2017; Wadgymar *et al.*, 2017; Tusso *et al.*, 2021). This is likely due to the control of selective agents in experiments as opposed to natural sites where both selective agents and traits under selection may vary (Wadgymar *et al.*, 2017). In our system, high-elevation sites of *S. dioica* exhibit high winter mortality in *S. latifolia* and in hybrids in comparison to *S. dioica* (Favre *et al.*, 2017) suggesting that frost tolerance or the regulation of carbohydrate storage could be under selection. At lowland *S. latifolia* sites, in contrast, *S. dioica* and hybrids suffer higher summer mortality than the native *S. latifolia*, most likely due to drought exposure (Favre *et al.*, 2017). We therefore find it plausible that phenotypic and genetic targets of selection differ between habitats and this can contribute to the conditional neutrality pattern of effect sizes for fitness components as observed here.

Fitness is increased by native alleles and decreased by non-native alleles, but not at the same loci

A striking result in our study is the association of allele frequencies in the two species with allelic effects on cumulative flowering, even for loci with very small effects. Alleles with positive effects were rare in the non-native species and common or of intermediate frequency in the native species, especially in the *S. latifolia* habitat. Alleles with positive effects on cumulative flowering in one habitat were neutral in the alternative habitat and could thus easily spread through both species if not lost by drift (Savolainen *et al.*, 2013; Mee & Yeaman, 2019). Alleles with negative effects on cumulative flowering, on the other hand, had intermediate to high frequencies in the non-native species but were rare in the native species. Negative-effect alleles also included numerous variants that were rare in both species but these loci are most likely linked to causal loci with unknown allele frequencies rather than being causal loci themselves. Negative-effect alleles are most readily interpreted as deleterious load that is only exposed to selection in the alternative habitat and rare, conditionally deleterious variants are expected to arise frequently (Orr, 2005). Conditionally deleterious loci may experience reduced gene flow due to negative selection in the alternative habitat and thereby increase between-lineage differentiation (Mee & Yeaman, 2019). However, rare variants with conditionally deleterious effects likely remain rare even in the lineage they arose in and will thus not exhibit high differentiation between lineages. Variants with conditionally neutral effects, as reported here, likely play an important role for adaptation, but they are most certainly under-reported, not only because many of these

463 variants are rare, but also because studies in alternative and relevant environments are
464 needed to detect them.

465 *Implications for hybrid zones and speciation*

466 Our results suggest that selection favors hybrids that are genetically similar to the native
467 species. In early-generation hybrids with still large genomic regions of each species, as
468 in our experiment, selection will favor genomic regions that contain positive-effect
469 alleles that were more often derived from the native species. At the same time,
470 selection will disfavor regions with negative-effect alleles that were often derived from
471 the non-native species. Both types of conditionally neutral variants thus act as barriers
472 to gene flow in each habitat. In natural hybrid zones between *S. dioica* and *S. latifolia*,
473 intermediate individuals are rare and later-generation hybrids are heavily biased toward
474 individuals that are genetically similar to the native species (Minder *et al.*, 2007;
475 Karrenberg & Favre, 2008). Our study suggests that this may be due not only to back-
476 crossing with the native species but also to ecological selection in each habitat.

477 Regarding the evolution of adaptative differentiation, our results suggest adaptation
478 through selection on many loci in each species. Alleles with small positive effects had
479 intermediate to high frequencies in the native species and low frequencies in the non-
480 native species. This may point to a redundant genetic control of adaptation as also
481 suggested by a QTL study under controlled conditions in our study system (Liu &
482 Karrenberg, 2018) and by studies in other systems, for example in *Drosophila* (Barghi
483 *et al.*, 2019). Polygenic architectures of adaptation with redundant effects are expected
484 to be transient in time (Yeaman & Whitlock, 2011; Yeaman, 2015). Thus, even when
485 ecological differentiation is a likely driver of divergence, as in our study system
486 (Goulson & Jerrim, 1997; Karrenberg & Favre, 2008; Karrenberg *et al.*, 2019), this may
487 not necessarily manifest in strong, range-wide genetic differentiation at the underlying
488 loci.

489 *Conclusion*

490 Overall, our study adds to the understanding of how ecological differentiation can
491 promote reproductive isolation and lineage divergence. We detected a polygenic
492 architecture for a fitness component, cumulative flowering, in the campions *Silene*
493 *dioica* and *S. latifolia* using a reciprocal transplant experiment. Allelic effects were

494 mostly beneficial or deleterious in one habitat and neutral in the other and native-
495 derived alleles were favored. Conditionally neutral effects may result from differences
496 in selection targets in the two habitat types. Variants with conditionally neutral effects
497 provide hidden variation upon which selection can act and facilitate current and future
498 adaptation. The challenges ahead lie in understanding how within-lineage variation for
499 fitness affecting alleles can be reconciled with the evolution of genome-wide
500 differentiation between species.

501

502 ## Acknowledgements

503 We thank to Rasmus Janson and Karin Steffen for providing help in the molecular lab.
504 We are grateful for support from the Science for Life Laboratory (SciLifeLab) and the
505 National Genomics Infrastructure, NGI, for massive parallel sequencing. Computations
506 were performed on resources provided by SNIC through the Uppsala Multidisciplinary
507 Center for Advanced Computational Science (UPPMAX) under SNIC projects 2017/7-
508 406 and 2019/8-21. This work was funded by a project grants of the Swiss National
509 Science Foundation (SNF, no. 3100A-118221), the Swedish Research Council
510 (Vetenskapsrådet, no. 2012-03622) and the Carl Tryggers foundation to SK, as well as
511 by a grant of the German Science Foundation (Deutsche Forschungsgemeinschaft,
512 project no. FA1117/1-2) to AF.

513 ## Author contribution

514 The study was designed by SK with input by AF and XL. AF conducted the transplant
515 experiment with help from SK, XL performed molecular lab work, XL and SG did
516 bioinformatic analyses, SG performed the association analysis and simulations with
517 help from AB and SK. SK analyzed phenotypes and produced figures with help from
518 SG. SK and SG wrote the paper with input from all other authors.

519 ## Data availability

520 Double-digest RAD (ddRAD) sequencing data is available on NCBI's Short Read
521 Archive (SRA, SUB8324195 https://www.ncbi.nlm.nih.gov/sra/PRJNA669447). The

variant call format (VCF) file as well as GEMMA input files will be submitted to dryad (project number to be added).

References

Ågren J, Oakley CG, Lundemo S, Schemske DW. 2017. Adaptive divergence in flowering time among natural populations of *Arabidopsis thaliana*: Estimates of selection and QTL mapping. *Evolution* **71**(3): 550-564.

Anderson JT, Lee CR, Mitchell-Olds T. 2014. Strong selection genome-wide enhances fitness trade-offs across environments and episodes of selection. *Evolution* **68**(1): 16-31.

Anderson JT, Willis JH, Mitchell-Olds T. 2011. Evolutionary genetics of plant adaptation. *Trends in Genetics* **27**(7): 258-266.

Baena□Díaz F, Zemp N, Widmer A. 2019. Insights into the genetic architecture of sexual dimorphism from an interspecific cross between two diverging *Silene* (Caryophyllaceae) species. *Molecular Ecology* **28**(23): 5052-5067.

Barghi N, Tobler R, Nolte V, Jaksic AM, Mallard F, Otte KA, Dolezal M, Taus T, Kofler R, Schlotterer C. 2019. Genetic redundancy fuels polygenic adaptation in *Drosophila*. *PLoS Biology* **17**(2): e3000128.

Bates D, Maechler M, Bolker B, Walker S. 2015. Fitting linear mixed-effects models using *lme4*. *Journal of Statistical Software* **67**: 1-48.

Bono LM, Smith LB, Jr., Pfennig DW, Burch CL. 2017. The emergence of performance trade-offs during local adaptation: insights from experimental evolution. *Molecular Ecology* **26**(7): 1720-1733.

Booker TR, Yeaman S, Whitlock MC. 2021. Global adaptation complicates the interpretation of genome scans for local adaptation. *Evolution Letters* **5**(1): 4-15.

Boyle EA, Li YI, Pritchard JK. 2017. An Expanded View of Complex Traits: From Polygenic to Omnigenic. *Cell* **169**(7): 1177-1186.

Bresadola L, Caseys C, Castiglione S, Buerkle CA, Wegmann D, Lexer C. 2019. Admixture mapping in interspecific *Populus* hybrids identifies classes of genomic architectures for phytochemical, morphological and growth traits. *New Phytologist* **223**(4): 2076-2089.

Catchen J, Hohenlohe PA, Bassham S, Amores A, Cresko WA. 2013. Stacks: an analysis tool set for population genomics. *Molecular Ecology* **22**: 3124-3140.

554 **Chen S, Zhou Y, Chen Y, Gu J. 2018.** Fastp: An ultra-fast all-in-one FASTQ
555 preprocessor. *Bioinformatics* **34**(17): i884-i890.
556 **Comeault AA, Soria-Carrasco V, Gompert Z, Farkas TE, Buerkle CA, Parchman**
557 **TL, Nosil P. 2014.** Genome-wide association mapping of phenotypic traits
558 subject to a range of intensities of natural selection in *Timema cristinae*. *The*
559 *American Naturalist* **183**(5): 711-727.
560 **Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker**
561 **RE, Lunter G, Marth GT, Sherry ST, et al. 2011.** The variant call format and
562 VCFtools. *Bioinformatics* **27**(15): 2156-2158.
563 **Exposito-Alonso M, Burbano HA, Bossdorf O, Nielsen R, Weigel D. 2019.** Natural
564 selection on the *Arabidopsis thaliana* genome in present and future climates.
565 *Nature* **573**(7772): 126-129.
566 **Favre A, Widmer A, Karrenberg S. 2017.** Differential adaptation drives ecological
567 speciation in campions (*Silene*): evidence from a multi-site transplant
568 experiment. *New Phytologist* **213**(3): 1487-1499.
569 **Fournier-Level A, Wilczek AM, Cooper MD, Roe JL, Anderson J, Eaton D,**
570 **Moyers BT, Petipas RH, Schaeffer RN, Pieper B, et al. 2013.** Paths to
571 selection on life history loci in different natural environments across the native
572 range of *Arabidopsis thaliana*. *Molecular Ecology* **22**(13): 3552-3566.
573 **Friedrich HC 1979.** Caryophyllaceae. In: Rechinger KH ed. *Illustrierte Flora von*
574 *Mitteleuropa*. Hamburg: Parey.
575 **Garrison E, Marth G. 2012.** Haplotype-based variant detection from short-read
576 sequencing. 1-9.
577 **Gompert Z, Egan SP, Barrett RD, Feder JL, Nosil P. 2017.** Multilocus approaches
578 for the measurement of selection on correlated genetic loci. *Molecular Ecology*
579 **26**(1): 365-382.
580 **Gompert Z, Jahner JP, Scholl CF, Wilson JS, Lucas LK, Soria-Carrasco V,**
581 **Fordyce JA, Nice CC, Buerkle CA, Forister ML. 2015.** The evolution of
582 novel host use is unlikely to be constrained by trade-offs or a lack of genetic
583 variation. *Molecular Ecology* **24**(11): 2777-2793.
584 **Gompert Z, Lucas LK, Nice CC, Buerkle CA. 2013.** Genome divergence and the
585 genetic architecture of barriers to gene flow between *Lycaeides idas* and
586 *L.melissa*. *Evolution* **67**(9): 2498-2514.

587 **Gompert Z, Messina FJ. 2016.** Genomic evidence that resource-based trade-offs limit
588 host-range expansion in a seed beetle. *Evolution* **70**(6): 1249-1264.
589 **Goulson D, Jerrim K. 1997.** Maintenance of the species boundary between *Silene*
590 *dioica* and *S. latifolia* (red and white campion). *Oikos* **79**(1): 115-126.
591 **Hall MC, Lowry DB, Willis JH. 2010.** Is local adaptation in *Mimulus guttatus* caused
592 by trade-offs at individual loci? *Molecular Ecology* **19**(13): 2739-2753.
593 **Hothorn T, Hornik K, van de Wiel MA, Zeileis A. 2008.** Implementing a class of
594 permutation tests: the *coin* package. *Journal of Statistical Software* **28**(8): 1-23.
595 **Hu XS, Filatov DA. 2015.** The large-X effect in plants: Increased species divergence
596 and reduced gene flow on the *Silene* X-chromosome. *Molecular Ecology* **25**:
597 2609-2619.
598 **Karrenberg S, Favre A. 2008.** Genetic and ecological differentiation in the hybridizing
599 campions *Silene dioica* and *S. latifolia*. *Evolution* **62**(4): 763-773.
600 **Karrenberg S, Liu X, Hallander E, Favre A, Herforth□Rahmé J, Widmer A. 2019.**
601 Ecological divergence plays an important role in strong but complex
602 reproductive isolation in campions (*Silene*). *Evolution* **73**(2): 245-261.
603 **Kawecki TJ, Ebert D. 2004.** Conceptual issues in local adaptation. *Ecology Letters*
604 **7**(12): 1225-1241.
605 **Kokko H, Chaturvedi A, Croll D, Fischer MC, Guillaume F, Karrenberg S, Kerr**
606 **B, Rolshausen G, Stapley J. 2017.** Can evolution supply what ecology
607 demands? *Trends in Ecology and Evolution* **32**(3): 187-197.
608 **Krasovec M, Chester M, Ridout K, Filatov DA. 2018.** The mutation rate and the age
609 of the sex chromosomes in *Silene latifolia*. *Current Biology* **28**(11): 1832-1838.
610 **Leimu R, Fischer M. 2008.** A meta-analysis of local adaptation in plants. *PloS One*
611 **3**(12): e4010. doi:4010.1371/journal.pone.0004010.
612 **Leinonen PH, Remington DL, Leppälä J, Savolainen O. 2013.** Genetic basis of local
613 adaptation and flowering time variation in *Arabidopsis lyrata*. *Molecular*
614 *Ecology* **22**(3): 709-723.
615 **Lenth RV 2020**. *emmeans*: Estimated marginal means, aka least-squares means. R
616 package version 1.5.3. https://CRAN.R-project.org/package=emmeans.
617 **Li H. 2013.** Aligning sequence reads, clone sequences and assembly contigs with
618 BWA-MEM. *arXiv* **1303.3997**.
619 **Lind BM, Friedline CJ, Wegrzyn JL, Maloney PE, Vogler DR, Neale DB, Eckert**
620 **AJ. 2017.** Water availability drives signatures of local adaptation in whitebark

621 pine (*Pinus albicaulis* Engelm.) across fine spatial scales of the Lake Tahoe
622 Basin, USA. *Molecular Ecology* **26**(12): 3168-3185.
623 **Liu X, Glémin S, Karrenberg S. 2020.** Evolution of putative barrier loci at an
624 intermediate stage of speciation with gene flow in campions (*Silene*). *Molecular*
625 *Ecology* **29**(18): 3511-3525.
626 **Liu X, Karrenberg S. 2018.** Genetic architecture of traits associated with reproductive
627 barriers in *Silene*: Coupling, sex chromosomes and variation. *Molecular Ecology*
628 **27**(19): 3889-3904.
629 **Lowry DB, Lovell JT, Zhang L, Bonnette J, Fay PA, Mitchell RB, Lloyd-Reilley J,**
630 **Boe AR, Wu Y, Rouquette FM, Jr., et al. 2019.** QTL x environment
631 interactions underlie adaptive divergence in switchgrass across a large latitudinal
632 gradient. *Proceeding of the National Academy of Sciences, U S A* **116**(26):
633 12933-12941.
634 **Mee JA, Yeaman S. 2019.** Unpacking conditional neutrality: genomic signatures of
635 selection on conditionally beneficial and conditionally deleterious mutations.
636 *American Naturalist* **194**(4): 529-540.
637 **Minder AM, Rothenbuehler C, Widmer A. 2007.** Genetic structure of hybrid zones
638 between *Silene latifolia* and *Silene dioica* (Caryophyllaceae): evidence for
639 introgressive hybridization. *Molecular Ecology* **16**(12): 2504-2516.
640 **Mitchell-Olds T, Willis JH, Goldstein DB. 2007.** Which evolutionary processes
641 influence natural genetic variation for phenotypic traits? *Nature Reviews:*
642 *Genetics* **8**(11): 845-856.
643 **Nosil P. 2012.** *Ecological Speciation*. Oxford: Oxford University Press.
644 **O'Leary SJ, Puritz JB, Willis SC, Hollenbeck CM, Portnoy DS. 2018.** These aren't
645 the loci you'e looking for: Principles of effective SNP filtering for molecular
646 ecologists. *Molecular Ecology* **27**(16): 3193-3206.
647 **Orr HA. 1998.** The population genetics of adaptation: the distribution of factors fixed
648 during adaptive evolution. *Evolution* **52**(4): 935-949.
649 **Orr HA. 2005.** The genetic theory of adaptation: a brief history. *Nature Reviews*
650 *Genetics* **6**(2): 119-127.
651 **Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE. 2012.** Double digest
652 RADseq: an inexpensive method for de novo SNP discovery and genotyping in
653 model and non-model species. *PloS One* **7**(5): e37135.

654 **Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J,**
655 **Sklar P, de Bakker PIW, Daly MJ, et al. 2007.** PLINK: a tool set for whole-
656 genome association and population-based linkage analyses. *American Journal of*
657 *Human Genetics* **81**(3): 559-575.
658 **Puritz JB, Hollenbeck CM, Gold JR. 2014.** dDocent: a RADseq, variant-calling
659 pipeline designed for population genomics of non-model organisms. *PeerJ* **2**:
660 e431.
661 **R Core Team. 2020.** R: A language and environment for statistical computing. R
662 Foundation for Statistical Computing, Vienna, Austria. https://www.R-
663 project.org/.
664 **Rockman MV. 2012.** The QTN program and the alleles that matter for evolution: All
665 that's gold does not glitter. *Evolution* **66**(1): 1-17.
666 **Savolainen O, Lascoux M, Merila J. 2013.** Ecological genomics of local adaptation.
667 *Nature Reviews: Genetics* **14**(11): 807-820.
668 **Selby JP, Willis JH. 2018.** Major QTL controls adaptation to serpentine soils in
669 *Mimulus guttatus*. *Molecular Ecology* **27**(24): 5073-5087.
670 **Tiffin P, Ross-Ibarra J. 2014.** Advances and limits of using population genetics to
671 understand local adaptation. *Trends in Ecology and Evolution* **29**(12): 673-680.
672 **Tigano A, Friesen VL. 2016.** Genomics of local adaptation with gene flow. *Molecular*
673 *Ecology* **25**(10): 2144-2164.
674 **Tusso S, Nieuwenhuis BPS, Weissensteiner B, Immler S, Wolf JBW. 2021.**
675 Experimental evolution of adaptive divergence under varying degrees of gene
676 flow. *Nature Ecology and Evolution* **5**(3): 338-349.
677 **Wadgymar SM, Lowry DB, Gould BA, Byron CN, Mactavish RM, Anderson JT.**
678 **2017.** Identifying targets and agents of selection: innovative methods to evaluate
679 the processes that contribute to local adaptation. *Methods in Ecology and*
680 *Evolution* **8**(6): 738-749.
681 **Yeaman S. 2015.** Local adaptation by alleles of small effect. *The American Naturalist*
682 **186**(S1): S74-S89.
683 **Yeaman S, Whitlock MC. 2011.** The genetic architecture of adaptation under
684 migration-selection balance. *Evolution* **65**(7): 1897-1911.
685 **Zhou X, Carbonetto P, Stephens M. 2013.** Polygenic modeling with bayesian sparse
686 linear mixed models. *Plos Genetics* **9**(2): e1003264.
687

Supporting Information

Figure S1. Principal components analysis including F_0 and F_2 generations.

Figure S2. Principal components analysis of F_2 families.

Figure S3. Posterior distributions of hyperparameters, *S. dioica* habitat.

Figure S4. Posterior distributions of hyperparameters, *S. latifolia* habitat.

Figure S5. Model-averaged effect sizes ($\bar{\beta}$) in the *S. dioica* habitat, plotted against $\bar{\beta}$ in the *S. latifolia* habitat.

Figure S6. Permutation test comparing between-species allele frequency differences among positive- and negative-effect alleles, based on raw effect size estimates ($\hat{\beta}$).

Figure S7. Permutation test comparing between-species allele frequency differences among positive- and negative-effect alleles, based on model-averaged effect size estimates ($\bar{\beta}$).

Figure S8. Correlations among simulated effects and raw and model-averaged estimates of effect sizes ($\hat{\beta}$ and $\bar{\beta}$).

Table S1. Description of source populations and transplant sites.

Table S2. Crossing scheme.

Table S3. Individuals used for crosses.

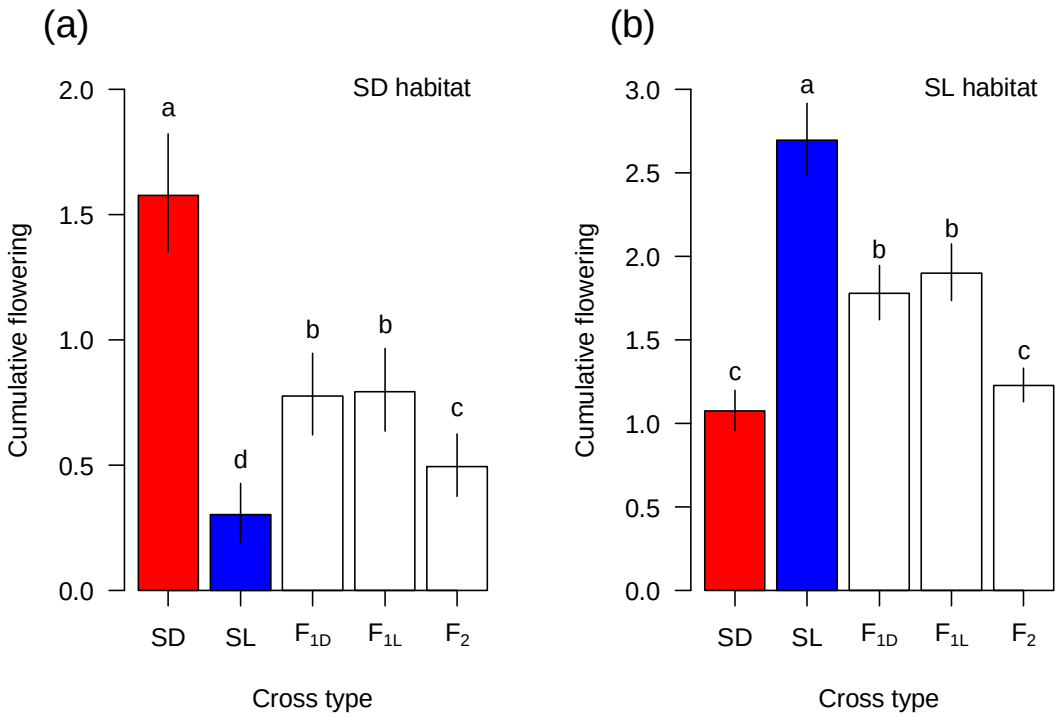
Table S4. Sparse-effect loci with posterior inclusion probabilities over 0.01.

Table

Table 1. Genetic architecture of cumulative flowering, a fitness component, in recombinant F_2 hybrids between the campions *Silene dioica* and *S. latifolia* in a field transplant experiment at sites of each species; given are medians, means and 90 % equal tail credible intervals (90% cred. int.) of hyperparameters estimated in Bayesian Sparse Linear Mixed Models (BSLMMs): *PVE* (proportion variation explained, broad-sense heritability, H^2), *PGE* (proportion variation explained by measurable effects), h^2 (narrow-sense heritability, *PVE* * *PGE*) and $n\text{-}\gamma$ (number of loci with measurable effects). Posterior distributions of hyperparameters are shown in Supplementary Figures S3 and S4.

	Hyperparameter			
	PVE	*PGE*	h^2	$n\text{-}\gamma$
Silene dioica habitat				
median	0.117	0.411	0.037	11
mean	0.136	0.434	0.060	42.2
90% cred. int.	0.017 - 0.321	0 - 0.945	0 - 0.198	0 - 205
Silene latifolia habitat				
median	0.096	0.411	0.029	16
mean	0.124	0.431	0.054	51.4
90% cred. int.	0.008 - 0.337	0 - 0.938	0 - 0.193	0 - 225

724 FIGURE 1
725

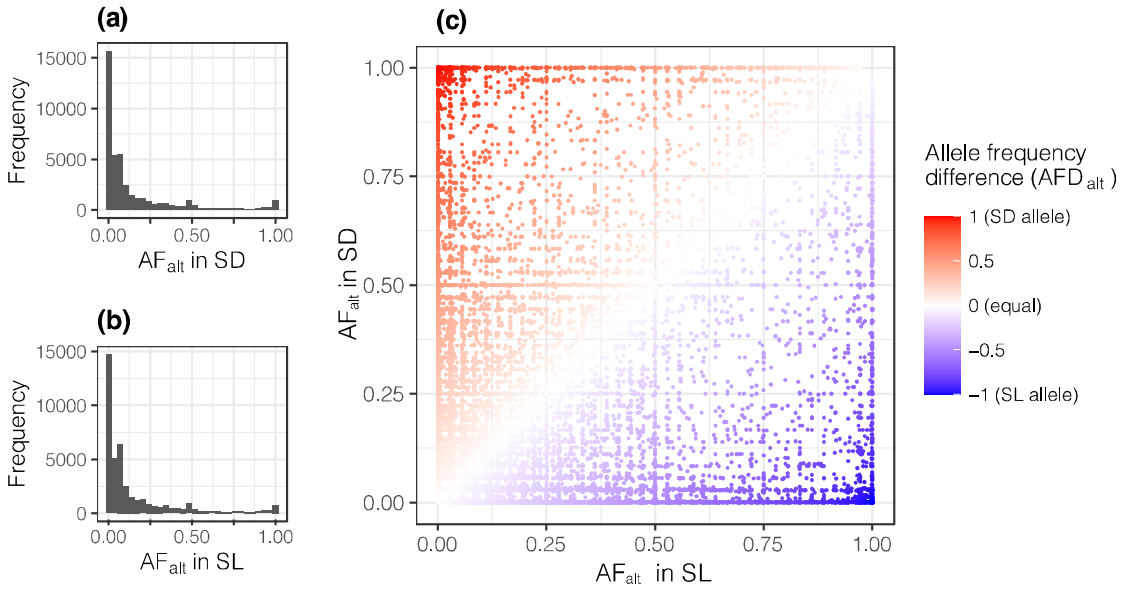


726
727

728 **Figure 1.** Habitat adaptation in campions (*Silene*): cumulative flowering (number of times
729 flowered over four years, plus 1 if the plant survived to the end of the experiment), a fitness
730 component, in crosses within each of two campion species, *Silene dioica* (SD) and *S. latifolia*
731 (SL), as well as in their first-generation hybrids (F_{1D} with SD mother, F_{1L}, SL mother) and
732 second-generation hybrids (F_2) in the habitat of each species, (a) SD habitat and (b) SL habitat;
733 least square means and 95% confidence intervals are given and different letters indicate
734 significant differences between cross types within each habitat. This data is published in Favre
735 et al. (2017 and re-analyzed here for illustration (see text).

736
737

738 FIGURE 2
739
740

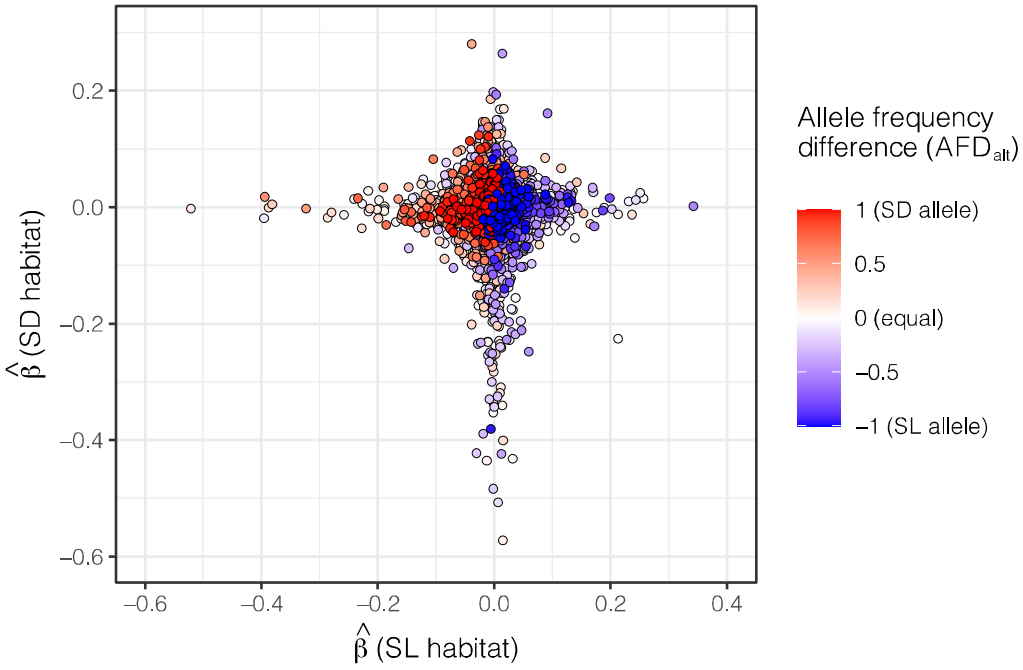


741
742
743

744 **Figure 2.** Frequencies of the alternative allele (AF_{alt}) in two campion species, *Silene dioica*
745 (SD) and *S. latifolia* (SL) for 42'090 loci used in an association study: **(a)** distribution of alleles
746 frequencies in SD, **(b)** distribution of allele frequencies in SL and **(c)** SD allele frequency
747 (y-axis) against SL allele frequency (x-axis) with allele frequency differences (AFD_{alt}) between
748 the two species (SD - SL) indicated as a color gradient from red (allele fixed in SD and absent
749 in SL) to white (equal frequency in both species) to blue (allele fixed in SL and absent in SD).

750
751

752 FIGURE 3



753
754

755 **Figure 3.** Locus-specific effect sizes of the alternative allele on cumulative flowering, a fitness
756 component, measured in recombinant hybrids between two campion species, *Silene dioica* (SD)
757 and *S. latifolia* (SL) that were transplanted into the habitat of each species (SD habitat, y-axis;
758 SL habitat, x-axis). Given are raw effect estimates, $\hat{\beta}$, from Bayesian Sparse Linear Mixed
759 Models (BSLMMs). Allele frequency differences for the alternative allele (AFD_{alt}) between the
760 two species (SD-SL) are indicated as a color gradient from blue (allele fixed in SL and absent in
761 SD) to white (equal frequency in both species) to red (allele fixed in SD and absent in SL).
762 Points are plotted in the order of increasing absolute AFD_{alt}, such that highly differentiated loci
763 are most visible. Loci with evidence for antagonistic pleiotropy would be have $\hat{\beta}$-values of
764 opposite signs in two habitats (top left, and bottom right), whereas loci consistent with
765 conditional neutrality lie at zero on one axis, but deviate substantially from zero on the other.

766

FIGURE 4

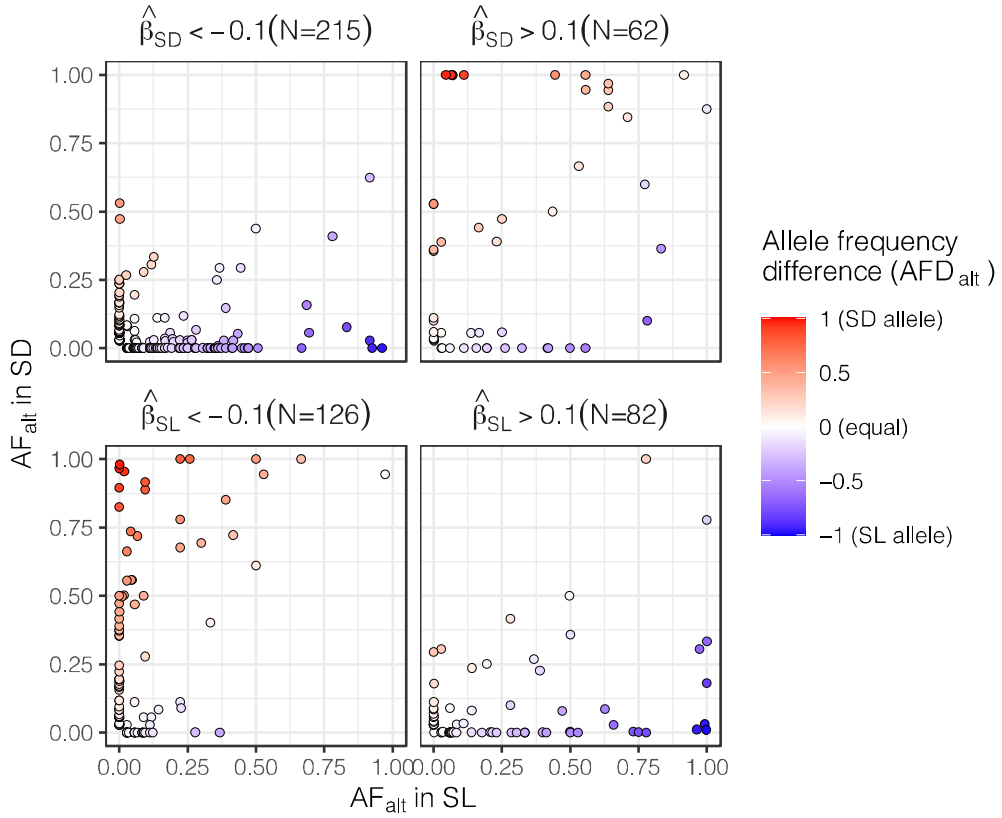
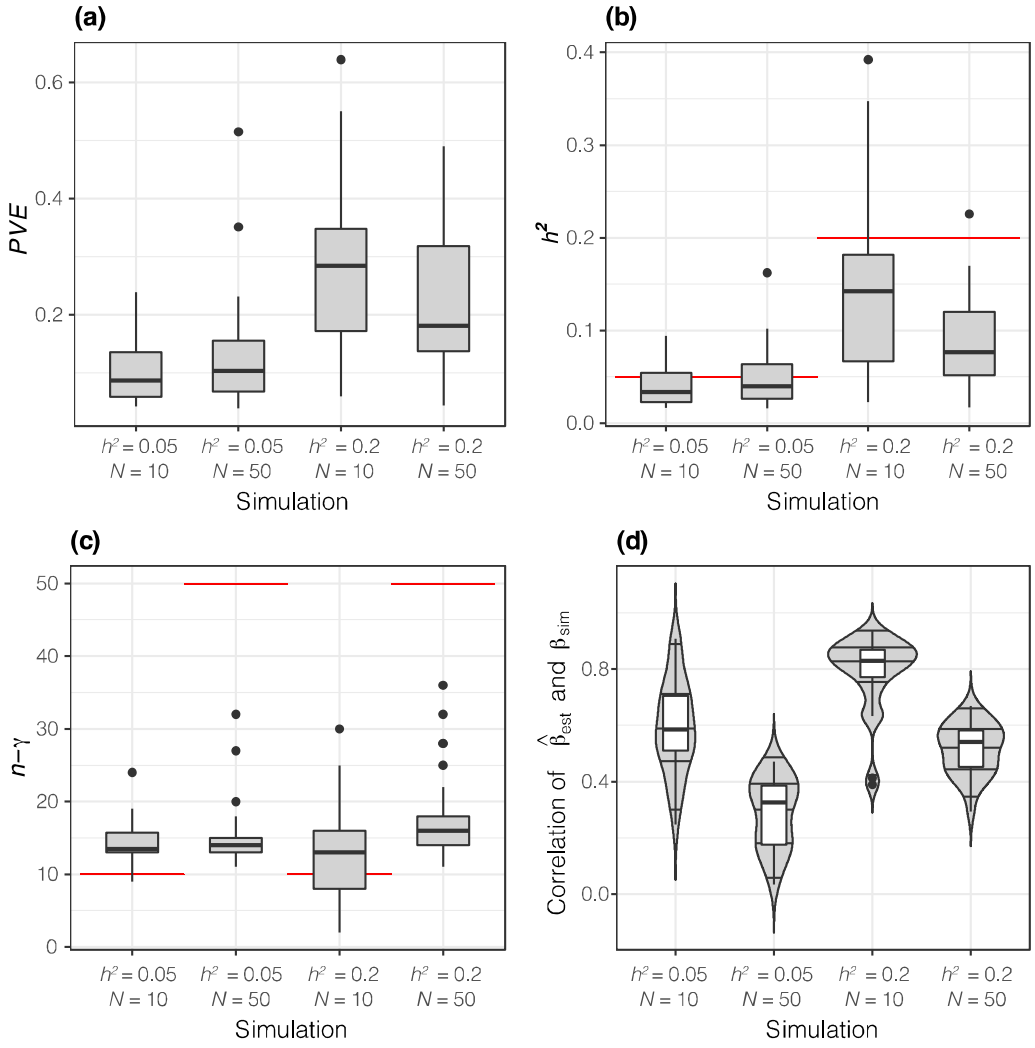


Figure 4. Frequencies of the alternative allele (AF_{alt}) in two campion species, *Silene dioica* (SD, y-axis) and *S. latifolia* (SL, x-axis) for loci associated with cumulative flowering, a fitness component, in the habitat of each species. Shown are loci with absolute raw effect sizes, $|\hat{\beta}|$, over 0.1 (in standard deviation units) in Bayesian Sparse Linear Mixed Models (BSLMMs) with the number of loci (N): $\hat{\beta}$ < -0.1 in the SD habitat (negative selection), $\hat{\beta}$ > 0.1 in the SD habitat (positive selection), $\hat{\beta}$ < -0.1 in the SL habitat (negative selection), and $\hat{\beta}$ > 0.1 in the SL habitat (positive selection). Allele frequency differences for the alternative allele (AFD_{alt}) between the two species (SD-SL) are indicated as a color gradient from blue (allele fixed in SL and absent in SD) to white (equal frequency in both species) to red (allele fixed in SD and absent in SL). Positive selection is associated with native alleles, whereas alleles under negative selection are more common in the non-native species (see also text and Supporting Information Figs. S6 and S7).

786 FIGURE 5



787
788

789 **Figure 5.** Performance analysis of Bayesian Sparse Linear Mixed Models (BSLMM) on
790 simulated data based on a dataset from campions (*Silene*, 150 individuals, 42'090 loci).
791 Simulated phenotypes had narrow-sense heritabilities (h^2) of 0.05 or 0.20 with 10 or 50
792 functional loci (*N*); 30 replicate simulations per scenario were used and boxplots or
793 violin plots summarize variation in medians of posterior distributions of
794 hyperparameters among simulation replicates as well as of correlations of simulated and
795 estimated per-locus effect sizes. **(a)** percent variation explained (*PVE*, interpreted as
796 broad-sense heritability, H^2), **(b)** narrow-sense heritability h^2 (as *PVE* * *PGE* [*PGE*,
797 proportion of sparse effects in *PVE*]), **(c)** number of sparse-effect loci (n-γ) and **(d)**,
798 correlation of raw estimated effect sizes for individual loci $\hat{\beta}_{est}$ with simulated effect
799 sizes β_{sim}. On **(b)** and **(c)** simulated values are indicated with red lines.