Uncovering the genomic basis of local adaptation by coherent synthesis of associations with phenotypes and home environments

Jesse R. Lasky[1,2], Brenna R. Forester[3], and Matthew Reimherr[4]

[1]Department of Biology
Pennsylvania State University
408 Life Sciences Building
University Park PA 16802-5301

[2]Correspondence: jrl35@psu.edu, Phone: (814) 863-5318, Fax: (814) 865-9131

[3]Nicholas School of the Environment
Duke University
450 Research Drive
Durham, NC 27708-0001

[4]Department of Statistics
Pennsylvania State University
411 Thomas Building
University Park PA 16802-5301

Keywords: landscape genomics, reciprocal transplant, genome-wide association study, fitness tradeoff, flowering time, plant defense, plasticity

Running title: Genomic basis of G×E and local adaptation

Abstract

A substantial portion of intraspecific diversity is associated with local adaptation to environment, which is driven by genotype-by-environment interactions (G×E) for fitness. Local adaptation is often studied via 1) multiple common garden experiments comparing performance of genotypes in different environments and 2) sequencing genotypes from multiple locations and characterizing geographic patterns in allele frequency. Both approaches aim to characterize the same pattern (local adaptation), yet the complementary information from each approach has not been coherently integrated into a modeling framework. Here, we develop a genome-wide association model of genotype interactions with continuous environmental gradients (G×E). We employ an imputation approach to synthesize evidence from common garden and genome-environment associations, allowing us to identify loci exhibiting environmental clines where alleles are associated with higher fitness in home environments. We apply this model to simulations and published data on *Arabidopsis thaliana*. Simulations showed our approach increases power to detect loci causing local adaptation. In *Arabidopsis,* our approach revealed candidate genes for local adaptation based on known involvement in environmental stress response. Most identified SNPs exhibited home allele advantage and fitness tradeoffs along climate gradients, suggesting selective gradients maintain allelic clines. SNPs exhibiting G×E associations with fitness were enriched in genic regions, putative partial selective sweeps, and G×E associations with an adaptive phenotype (flowering time). We discuss extensions for situations where only adaptive phenotypes other than fitness are available. Many types of data may point toward the loci underlying G×E and local adaptation; coherent models of these diverse data provide a principled basis for synthesis.

Introduction

Populations commonly exhibit strong phenotypic differences, often due to local adaptation to environment (Leimu and Fischer 2008; Hereford 2009). Local adaptation is defined as a genotype-by-environment interaction (G×E) for fitness that favors home genotypes (Kawecki and Ebert 2004). Local adaptation has long interested empirical and theoretical biologists (Clausen *et al.* 1940, 1948; Levene 1953; Slatkin 1973). However, little is known about the genomic basis of local adaptation, *e.g.* genetic architecture, major molecular mechanisms, and how much genomic divergence among populations is driven by local adaptation. Because local adaptation involves organismal responses to environmental gradients, understanding the mechanisms of local adaptation has important applications in agriculture and biodiversity conservation under climate change (Aitken and Whitlock 2013; van Oppen *et al.* 2015; Lasky *et al.* 2015). Additionally, genotype-by-environment interactions are important in human phenotypes like disease (Anastasi 1958; Hunter 2005; Gage *et al.* 2016). Understanding the genomic basis of G×E is an emerging area of biomedical research (Thomas 2010; Keller 2014) as are the genomics of local adaptation (reviewed by (Des Marais *et al.* 2013; Manel and Holderegger 2013; Tiffin and Ross-Ibarra 2014; Adrion *et al.* 2015; Bragg *et al.* 2015; Hoban *et al.* 2016)).

A central question in local adaptation genetics is whether selective gradients can maintain allelic clines at individual loci, or whether stochastic processes, like limited dispersal, are required to explain clines at individual loci causing local adaptation (Mitchell-Olds *et al.* 2007; Anderson *et al.* 2011b). If selective gradients cause rank changes in alleles with the highest relative fitness at an individual locus, selection may maintain a cline, a pattern known as genetic tradeoff or antagonistic pleiotropy (Ågren *et*

al. 2013). Detecting loci that exhibit antagonistic pleiotropy has been challenging, partly
75 due to limited statistical power of approaches that conduct multiple tests of significance for opposing fitness effects in different environments (ANDERSON *et al.* 2013). Here, we develop a model that explicitly tests marker associations with G×E for relative fitness variation, allowing us to identify loci with patterns indicative of antagonistic pleiotropy.

Common garden experiments have been employed for over 200 years to
80 characterize genetic variation in phenotypes (Langlet 1971). In particular, reciprocal common gardens at multiple positions along environmental gradients are a powerful tool to reveal local adaptation (Clausen *et al.* 1940, 1948). One approach to identifying the loci underlying local adaptation is to combine fitness data from multiple common garden experiments with genomic data (Lowry and Willis 2010; Fournier-Level *et al.* 2011a;
85 Anderson *et al.* 2011a; Ågren *et al.* 2013). Recently, the ability to sequence large panels of diverse genotypes has allowed genome-wide association mapping of loci underlying traits in common gardens (Atwell *et al.* 2010). However, common gardens are logistically challenging, limiting biologists' ability to phenotype diverse panels in multiple environments across a species range. Additionally, it is unclear how the typically small
90 spatiotemporal scales of common gardens relate to the scales of processes that generate local adaptation in the wild (Weigel and Nordborg 2015).

An alternative approach to discovering genetic and ecological mechanisms of local adaptation is to study changes in allele frequency along environmental gradients (Hedrick *et al.* 1976; Tiffin and Ross-Ibarra 2014; Adrion *et al.* 2015; Bragg *et al.* 2015;
95 Rellstab *et al.* 2015; Hoban *et al.* 2016). In this approach, known as a genome-environment association study, individuals are sequenced from multiple locations along

environmental gradients. Genetic markers and environmental gradients showing the strongest correlations are then considered as loci and selective gradients potentially involved in local adaptation (e.g. Hancock *et al.* 2008, 2011; ECKERT *et al.* 2010; Turner *et al.* 2010; Coop *et al.* 2010; Lasky *et al.* 2012; Jones *et al.* 2012; Fitzpatrick and Keller 2015). A challenge of both genome-phenotype and genome-environment association studies is that the genomic variation is observational and is not experimentally randomized (as opposed to linkage mapping with experimental crosses) (Devlin and Roeder 1999; Hancock *et al.* 2008; Kang *et al.* 2008; Nordborg and Weigel 2008). Thus many loci may show spurious associations with phenotypes or with environment (Price *et al.* 2010; Schoville *et al.* 2012; Bragg *et al.* 2015). Spurious associations are particularly problematic for environmental gradients that are spatially autocorrelated due to confounding with population structure (Schaffer and Johnson 1974). A technique for dealing with this confounding is to control for putative population structure when testing associations (Coop *et al.* 2010), *e.g*. by controlling for genome-wide (identity-in-state) similarity among accessions (Yoder *et al.* 2014; Lasky *et al.* 2014). Thus, this approach identifies loci that show strong associations with environment that deviate from genome-wide associations with environment

Understanding the genomic basis of adaptation may benefit from synthesizing lines of evidence, *i.e.* combining multiple types of genome scans to strengthen the evidence that a locus is under selection (Lasky *et al.* 2014; Evans *et al.* 2014). For example, researchers have identified overlap between outliers for selection statistics and markers associated with putatively adaptive phenotypes (Horton *et al.* 2012) or between SNPs associated with phenotypes and those associated with climate gradients (Berg and

120 Coop 2014). (Lasky *et al.* 2015) used a Bayesian approach to combine associations with phenotype and environment, first calculating climate associations and then using each marker's association to determine the prior probability it was associated with G×E for adaptive phenotypes, yielding a posterior. Although combining multiple lines of evidence is potentially useful, the quantitative basis of synthesis in past studies has often been *ad*
125 *hoc* and lacked reasoned principles. Here we develop a modeling framework to conduct genome-wide association scans for G×E while coherently synthesizing multiple data types. Existing approaches to genome-wide association studies (GWAS) with G×E (sometimes referred to as genome-wide interaction studies, GWIS) have dealt with categorical nominal environments (Murcray *et al.* 2009; Thomas 2010; Korte *et al.* 2012;
130 Gauderman *et al.* 2013; Marigorta and Gibson 2014), but have not been applied to G×E along continuous environmental gradients. Despite the existence of studies where fitness was measured in multiple common gardens for diverse genotyped accessions (Fournier-Level *et al.* 2011a), studies where linkage mapping was conducted for fitness at multiple sites (Ågren *et al.* 2013), and studies where authors conducted association mapping for
135 G×E effects on phenotypes (Li *et al.* 2014), we found no example of association studies of G×E for fitness, *i.e.* the basis of local adaptation.

Because the underlying processes generating local adaptation are the same regardless of whether genome-environment associations or common gardens are used for inference, it is natural to synthesize these data. Furthermore, by combining datasets into a
140 single inferential framework we may increase power and accuracy for detecting causal loci. Here, we simultaneously leverage data from multiple common gardens and genome-environment associations. In the remainder, we describe our approach, present test cases

using simulations and published data on *Arabidopsis thaliana* (hereafter Arabidopsis), and discuss promising avenues for extension.

Methods

Genome-wide association study of G×E effects on fitness

In common garden experiments environment is often treated as a factor. But when more than two gardens are conducted, variation among them may be considered in a more general fashion. For a given environmental gradient, each common garden may be located along the gradient (*i.e.* single dimension) according to its conditions. Describing common gardens as such may be informative about the specific ecological mechanisms driving selective gradients, taking advantage of the ordered nature of gardens' environments. We leverage multiple common garden experiments to identify markers (single nucleotide polymorphisms, SNPs) that show the strongest associations with G×E effects, *i.e.* loci where allelic state shows the strongest interaction with environment in its association with fitness.

Local adaptation requires a genotype by environment interaction for fitness at the whole genome-level. To assess this interaction at an individual locus, we assume that the relative fitness of individual i in a single location, w_i, satisfies the linear model

$$w_i = \alpha + \beta_E E_i + \beta_G G_{i,l} + \beta_{G\times E} G_{i,l} E_i + \varepsilon_i,$$

(eqn 1)

where $G_{i,l}$ is the genotype of individual i at locus l and E_i is the value of a single environmental variable at the location where w_i was measured. The β_E parameter gives

the effect of environment on fitness and β_G gives the effect of genotype on fitness. Our primary interest lies in the $\beta_{G\times E}$ parameter, which gives the strength and direction of G×E effects, *i.e.* $\beta_{G\times E}$ determines how responses to environmental gradients are mediated by genotype. α gives the fitness intercept. We assume that the vector of errors, $\boldsymbol{\varepsilon}$, can be expressed as

$$\boldsymbol{\varepsilon} = \boldsymbol{E}\boldsymbol{v} + \boldsymbol{e}$$

where $\boldsymbol{E}$ is a diagonal matrix of the environmental values, and

$$\boldsymbol{v} \sim N_n(0, \sigma^2_{G\times E}\,\boldsymbol{K}) \qquad \boldsymbol{e} \sim N_n(0, \sigma^2_e \boldsymbol{I}).$$

(eqn 2)

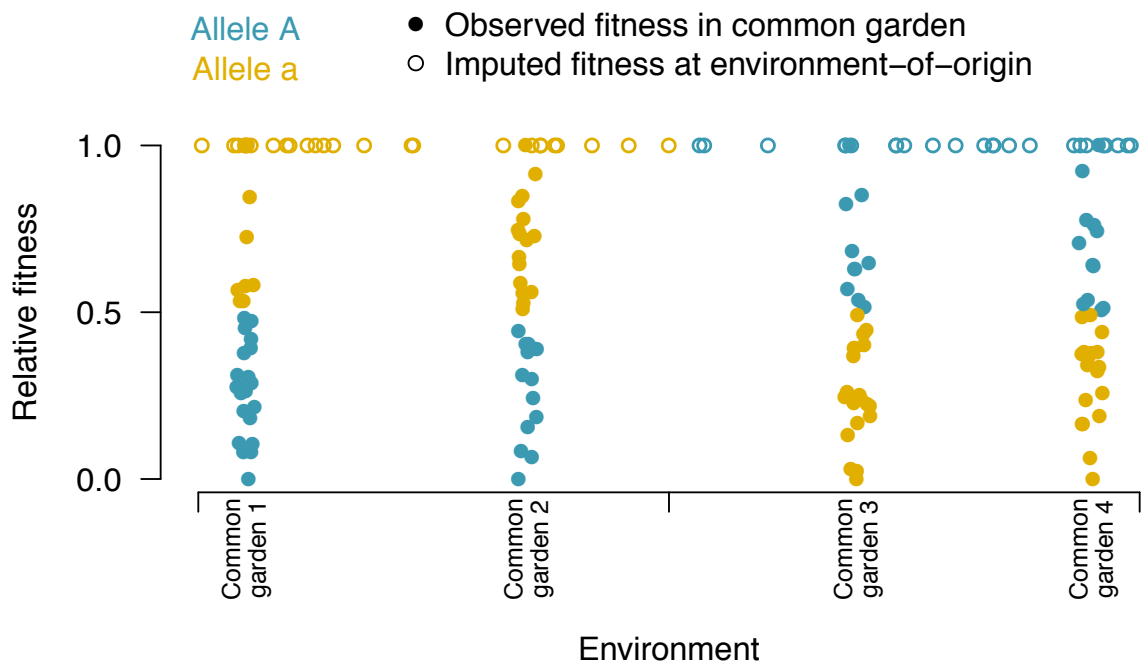
Here $\boldsymbol{v}$ and $\boldsymbol{e}$ are independent. The matrix **K** is calculated as the genome-wide identity in state for each pair of accessions (Kang *et al.* 2008). Random effects $\boldsymbol{v}$ are included because a substantial portion of G×E may be associated with population structure (Lasky *et al.* 2015); naively applying standard F-tests to assess the interaction effects can result in a dramatic increase in Type I error rates. To ameliorate this issue, the random effect $\boldsymbol{v}$ represents the genetic background interactions with environment (G×E, magnitude of their variance determined by $\sigma^2_{G\times E}$), while $\boldsymbol{e}$ represents the independent and identically distributed error in the model (variance determined by σ^2_e). However, it is important to note that incorporating random effects may also increase Type II error when causal loci covary with genomic background.

Coherent synthesis of common gardens and genome-environment associations via imputation

We now tackle the goal of synthesizing genome-environment associations and G×E observed in multiple common gardens, given that both patterns are expected to inform on the same process of local adaptation. The challenge in synthesizing these approaches is that genome-environment associations are purely observational and lack common garden phenotypic data. However, a typically unstated but implicit assumption in studies of genome-environment associations is that local adaptation occurs. That is, if a common garden were conducted at each location where genotypes are collected, the home genotype would tend to be most fit. Here, we make this assumption of genome-environment association studies explicit. A formal consequence of this assumption is an (imputed) observation of highest relative fitness for genotypes at home, which we combine with observed sequence and environment of origin data (Figure 1). Next, we scale relative fitness within each common garden so that the maximum observed fitness is given a relative fitness of unity, yielding a measure that can be directly observed or imputed in each type of study (common garden and genome-environment association). For imputation, we then assume that each genotype collected from wild populations is locally adapted at its home and thus has a relative fitness = 1 (Figure 1). After imputation, we can calculate marker associations with G×E for fitness, where each fitness observation arises from either (A) observations on a given genotype by common garden combination or (B) imputation on a given genotype collected from its natural home and subsequently sequenced.

Figure 1. Illustration of our imputation technique and stereotypical pattern captured by our approach. Here we show hypothetical data from four common gardens along an environmental gradient (solid circles in four vertical streaks, with small amount of noise added to environmental values for visualization) that have fitness scaled to a maximum of 1. We also show accessions (or ecotypes) collected in home environments and sequenced (open circles in horizontal line across top) having imputed relative fitness of 1 in their home environment (environment-of-origin). The locus in question shows strong G×E for fitness and

antagonistic pleiotropy, such that allele *A* (blue) is more fit (observed, solid circles) and more common (imputed, open circles) at the upper end (right side) of the environmental gradient while allele *a* (yellow) is more fit and more common at the lower end (left side) of the gradient.



Comparing approaches and fitting models

We compared three approaches to genome-wide G×E association study. In Approach 1, we ignored potential confounding of population structure, *i.e.* we used least squares to fit the model in eqn 1 where $\boldsymbol{\varepsilon}$ is normal, independent, and identically distributed (*i.e.* excluding random effects $\boldsymbol{\nu}$), and include imputed fitness data. In Approach 2, we fit the full mixed-effects model (including random effects $\boldsymbol{\nu}$), but only including observed fitness data from common gardens (*i.e.* excluding imputed fitness data). In Approach 3, we fit the full mixed-effects model while including imputed fitness data.

To improve computation time for the mixed-model approaches (2 & 3), we used the method of (Kang *et al.* 2010) and first fit the random effects with covariance determined by kinship, and then fixed these effects while testing the effects of each SNP

on the phenotype. We included the environmental covariate effect in this initial step, following the recommendation of (Kang *et al.* 2010) for fitting additional (i.e. non-SNP) covariates. In other words, we first fit the model:

$$w_i = \alpha + \beta_E E_i + \varepsilon_i,$$

(eqn 3)

with $\alpha + \beta_E E_i$ being removed from eqn 1 and instead fit here, and ε_i defined as in eqn 2, to obtain parameter estimates $\hat{\alpha}, \widehat{\beta_E}, \hat{\sigma}_e^2, \hat{\sigma}_{G\times E}^2$. We then take the variance parameter estimates and use them to estimate the remaining slope coefficients in eqn 1 using generalized least squares. We fit the discussed mixed-model using Minimum Norm Quadratic Unbiased Estimation, MINQUE (Rao 1971; Brown 1976; Reimherr and Nicolae 2015). This approach is equivalent to REML, but rephrased in a way that more fully exploits the linearity of the model, resulting in a flexible framework that can be quickly computed.

Simulation

We used simulation to demonstrate that our imputation technique can improve our power to identify loci causing G×E for fitness. We previously simulated local adaptation in a square two-dimensional 92160 x 92160 grid-cell landscape along a continuous environmental gradient (Forester *et al.* 2016), using the program CDPOP v1.2 (Landguth and Cushman 2010). We simulated 5,000 diploid individuals with 100 bi-allelic loci, one of which was under selection (99 neutral loci). All loci had a 0.0005 mutation rate per generation, free recombination, and no physical linkage. We ran 10 Monte Carlo

replicates of the simulation for 1,250 generations, using the first 250 generations as a burn-in, with no selection imposed, to establish a spatial genetic pattern.

In the simulation, selection changed linearly along an environmental gradient, with *AA* and *aa* genotypes favored at different ends of the gradient (North and South, Fig. S1). The selection strength of s=0.10 was mediated through density-independent mortality determined by an individual's genotype at the selected locus, where *AA* experienced 0% mortality in the North and s mortality in the South, while *aa* genotypes experienced the opposite selection gradient. *Aa* genotypes experienced uniform selection of $s/2$ across the gradient. Mating pairs of hermaphroditic individuals and dispersal locations of offspring were chosen using a random draw from the inverse-square probability function of distance, truncated at a distance equal to 10% of a landscape edge (*i.e.* truncated at 9216 pixels). Individuals near landscape edges were unable to disperse or mate with individuals beyond the edge (*i.e.* boundaries were not periodic).

The number of offspring produced from mating (fitness) was determined from a Poisson distribution ($\lambda = 4$), which produced an excess of individuals each generation, maintaining a constant population size of 5,000 individuals at every generation. Carrying capacity of the landscape surface was 5,000 individuals, and excess individuals were discarded once all 5,000 locations became occupied.

We sampled 500 individuals randomly from the 5,000 available. We then established four common gardens at equal intervals along the gradient, encompassing the extremes of the selection surface (Fig. S1). We subsampled 100 individuals from the full 500, and then averaged fitness for 10 clones of each individual (each with the identical adaptive genotype of their parent clone) in each common garden using the above

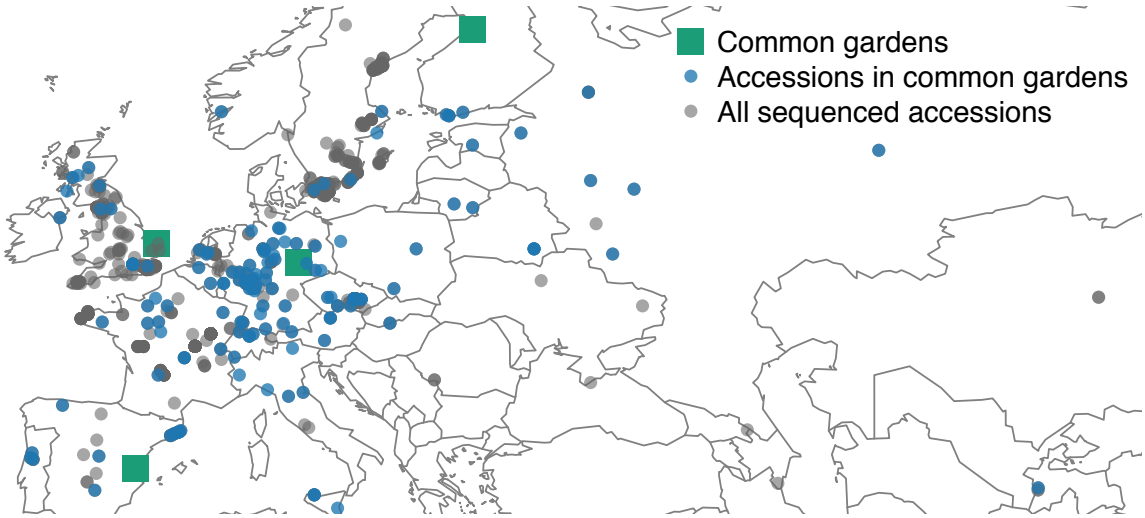
parameters for selection and fitness. After imputing fitness for the 400 individuals not included in common gardens, we had a total of 900 observations of fitness × location (i.e. 500 imputed observations + 400 real observations).

We then compared our three approaches to genome-wide G×E association study. In Approach 1, we ignored potential confounding of population structure but included imputed fitness data. In Approach 2, we fit the full mixed-effects model, but only including observed fitness data from common gardens (*i.e*. excluding imputed fitness). In Approach 3, we fit the full mixed-effects model while including imputed fitness data.

Case study: local adaptation to climate in Arabidopsis thaliana

We next applied these approaches to published data from studies of *Arabidopsis thaliana* in its native Eurasian range. Fournier-Level *et al.* (2011) conducted replicated common gardens at four sites across Europe: Spain, England, Germany, and Finland (Figure 2). With these common garden data, (Fournier-Level *et al.* 2011a; b; Wilczek *et al.* 2014) showed evidence that genotypes are locally adapted to their home temperature and moisture regimes and that alleles associated with high fitness in a given garden tended to be found nearer to that garden than alternate alleles, suggesting these loci were involved in local adaptation. At each site the authors transplanted 157 accessions (59 in the case of Finland) on a date in the fall matched to germination of local winter-annual natural populations (Fournier-Level *et al.* 2011a). The authors calculated survival (out of individuals surviving transplant) and average fecundity (where individuals that died before reproducing had fecundity zero) giving an estimate of absolute fitness (excluding the seed to seedling transition) (Fournier-Level *et al.* 2011a).

Figure 2. Data used in case study on Arabidopsis. The location of common gardens, natural accessions in common gardens, and all other sequenced natural accessions are shown.



These accessions were part of a panel of 1,307 accessions from around the globe that were genotyped at ~250k SNPs using a custom Affymetrix SNP tiling array (AtSNPtile1), with 214,051 SNPs remaining after quality control (Figure 2) (Horton *et al.* 2012). This array was generated by resequencing 19 diverse ecotypes from across the range of Arabidopsis (Kim *et al.* 2007). Of the 1,307 genotyped accessions, we used 1,001 accessions that were georeferenced and likely not contaminant lines (Anastasio *et al.* 2011), in addition to being from the native range in Eurasia (Hoffmann 2002; Lasky *et al.* 2012a), and excluding potentially inaccurate high altitude outliers (i.e. > 2000 m). After imputing fitness for accessions in their home environments we had a total of 1,531 observations of fitness × location (*i.e.* 1001 imputed observations + 530 real observations). We removed from association tests SNPs having minor allele frequency (MAF) below 0.01, though we also considered a more conservative threshold of MAF = 0.1.

We used climate data compiled previously (Lasky *et al.* 2012a) from published global climate datasets (Hijmans *et al.* 2005; Zomer *et al.* 2008). Here we focus on four climate variables that differ among common gardens, are not strongly correlated, and may be involved in local adaptation: minimum temperature of the coldest month, average
320 monthly minimum temperature in the growing season, coefficient of variation of monthly growing season precipitation, and aridity index.

Genome-wide G×E associations

We separately tested for each SNP's interaction with each of the four environmental
325 variables, *i.e.* for each of the three approaches we fit a model for each combination of SNP and environmental variable. To characterize the types of patterns identified by our approach, we studied variation in the SNPs in the 0.01 lower tail of p-values for the hypothesis that $\beta_{G\times E} = 0$ (the coefficient for SNP×environment effects on fitness) for each climate gradient. We asked whether these SNPs showed patterns consistent with
330 home genotype advantage via changes in the allele with greatest relative fitness along the environmental gradient (*i.e.* local adaptation via antagonistic pleiotropy). For these SNPs we calculated whether the direction of allelic differentiation along environmental gradients was consistent with the sign of $\beta_{G\times E}$. For example, if one allele was more common in accessions from warmer locations, we assessed whether that same allele
335 showed an increase in relative fitness in warmer common gardens. Next, we assessed whether our model predicted that different alleles were most fit in the two common gardens at either extreme of a climate gradient, *i.e.* whether the SNP was associated with rank changes in fitness that are consistent with antagonistic pleiotropy. For example, if

one allele was estimated to be most fit in the coldest common garden, we asked whether a
340 different allele was estimated to be most fit in the warmest common garden. Furthermore, we also quantified similarity (*i.e.* rank correlation in SNP scores and proportion of SNPs common to the strongest 0.01 tail of associations) between results from our G×E approach versus those from other recent studies of association with home climate in Arabidopsis (Hancock *et al.* 2011; Lasky *et al.* 2012a, 2014).

345

Enrichment of strong SNP×environment associations across the genome

We studied whether loci we identified as likely being involved in local adaptation exhibited supportive patterns in ancillary datasets. First, to assess whether our association approach is capable of identifying the signal of local adaptation rather than spurious
350 background associations, we tested for enrichment of SNPs in genic versus intergenic regions. These tests are based on the hypothesis that loci involved in adaptation are on average more likely to be found near genes and linked to genic variation, in comparison with loci evolving neutrally (Hancock *et al.* 2011; Lasky *et al.* 2012a). For a test statistic, we calculated the portion of SNPs in the 0.01 lower p-value tail that were genic versus
355 intergenic.

Second, we hypothesized that locally-adaptive alleles may have been subject to partial (*i.e*. local) selective sweeps, especially given that much of Arabidopsis' Eurasian range was recently colonized following the last glacial maximum. We tested for an enrichment of significant (alpha = 0.05) pairwise haplotype sharing (PHS, (Toomajian *et*
360 *al.* 2006)) in the SNPs (using PHS calculated by Horton et al. 2012) showing the greatest evidence of G×E for fitness. We also tested evidence that these SNPs are enriched for

significant (alpha = 0.05) integrated extended haplotype homozygosity (standardized, iHS (Voight *et al.* 2006)), an additional metric of partial sweeps. We used ancestral SNP allele determinations from (Horton *et al.* 2012) (based on alignment with the *Arabidopsis*
365 *lyrata* genome) and the R package 'rehh' to calculate iHS (Gautier *et al.* 2012).

Third, we also studied whether loci we identified were associated with plasticity in flowering time, a trait that plays a major role in local adaptation to climate in plants (Hall and Willis 2006; Franks *et al.* 2007; Keller *et al.* 2012; Lowry *et al.* 2014). Recently (Li *et al.* 2014) tested the flowering time response of 417 natural accessions to
370 simulated warming (up to ~4ºC), and then identified SNP associations with changes in flowering time across treatments, *i.e*. G×E for flowering time. We tested whether SNPs we identified as having SNP×environment interactions for fitness (0.01 lower p-value tail) were enriched in nominally significant associations (alpha = 0.05) with G×E for flowering time.

375 To generate a null expectation for each enrichment while maintaining a signal of linkage disequilibrium in the null model, we circularly permuted SNP categories (*e.g*. as genic vs intergenic, having significant iHS or not) along the genome and recalculated the test statistics 10,000 times (Hancock *et al.* 2011; Lasky *et al.* 2012a).

All data were previously published, including fitness (Data Dryad package
380 Fournier-Level *et al.* 2011b), environment (Data Dryad package Lasky *et al.* 2012b), and SNPs (Horton *et al.* 2012, available at http://bergelson.uchicago.edu/wp-content/uploads/2015/04/call_method_75.tar.gz).

Results

385 We compared three approaches to genome-wide G×E association study. Approach 1 used observed and imputed fitness data, but no correction for population structure. Approach 2 fitted the full mixed-effects model, but only including observed fitness data from common gardens (*i.e*. excluding imputed fitness). Approach 3 fitted the full mixed-effects model while including both observed and imputed fitness data.

390

Simulations

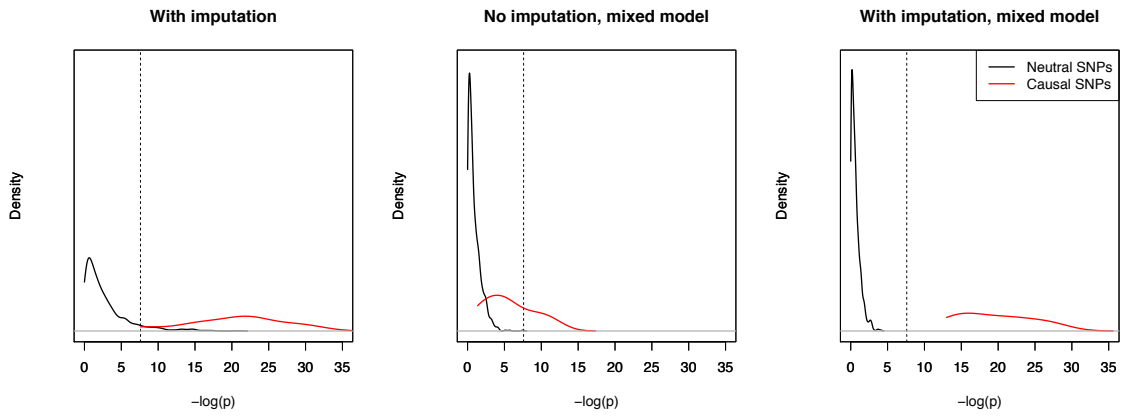
We found that imputation of relative fitness for genotypes in their home environments increased the power of mixed-model associations to identify SNPs driving G×E for fitness (Figure 3). In each of the 10 simulations, the single causal SNP was detected as

395 having the strongest (out of 100 SNPs) G×E for fitness when using imputed observations and a mixed model (Approach 3, Fig. 3 right). By contrast, when only using common gardens (Approach 2), the causal SNP did not have the strongest association in three simulations, when it was instead scattered among the top 10 SNPs (Fig. 3 middle). When including imputation but not random effects to control for population structure (Approach

400 1), the single causal SNP was detected as strongest in nine of ten simulations, though many neutral SNPs had very low p-values (false positive detections, Fig. 3 left).

Figure 3. Comparison of inferred SNP associations with G×E for fitness across 10 simulations, using imputation but no random effects (left) mixed models that used only observations from four common gardens (100 genotypes in each, middle) or mixed models combining imputed

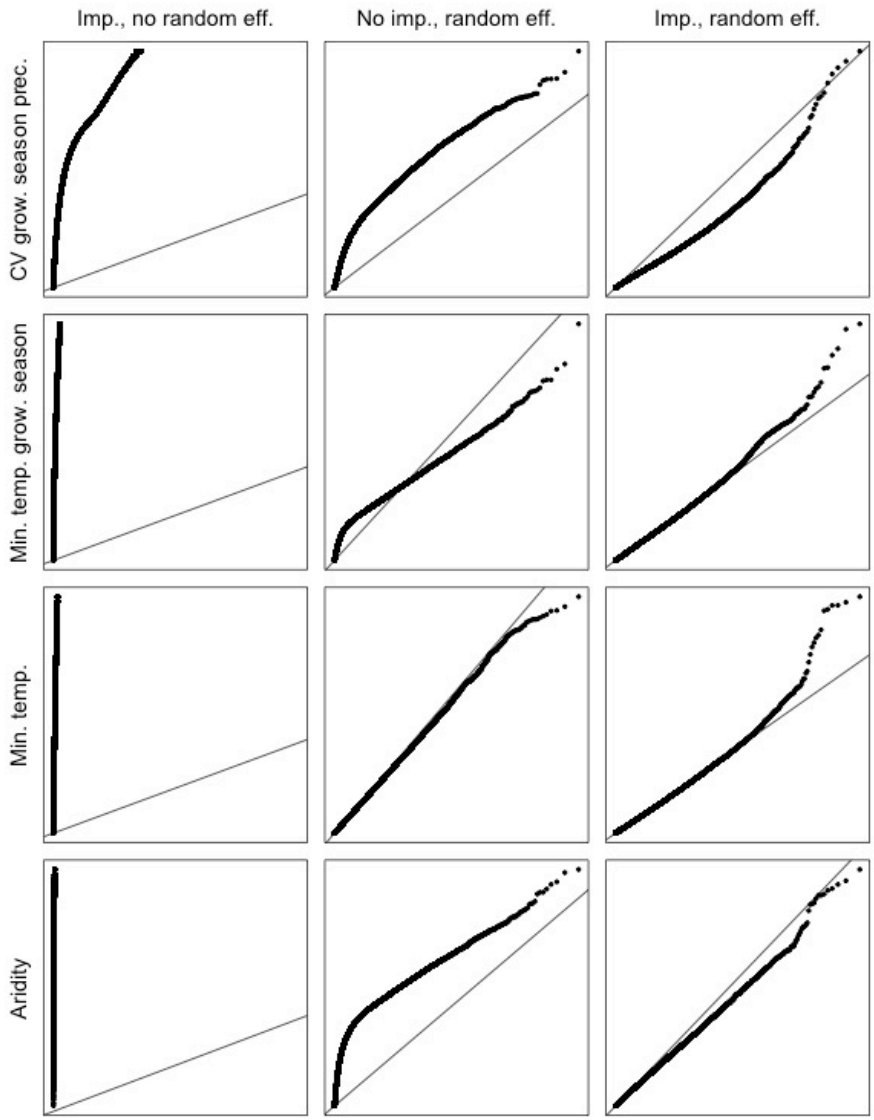
405 observations of relative fitness in home environments (500 genotypes) with common garden observations (right panel). Each simulation had 1 causal SNP and 99 neutral SNPs; density plots show aggregate distributions for all SNP by simulation combinations (*i.e.* total of 10 causal and 990 neutral SNPs). For reference, dashed line indicates a strict Bonferonni cutoff for alpha = 0.05, –log(0.05/100) = 7.6. Density curves are truncated at lower minimum values.



Case study on Arabidopsis

We found that simple linear model tests (Approach 1) of SNP×environment interactions were highly enriched in very low p-values (Figure 4) relative to the theoretical expectation. After incorporating the kinship×environment random effects (but excluding imputed fitness observations, Approach 2), we found that SNP×environment associations with fitness were closer to the theoretical expectation but still highly enriched in low p-values for three climate variables. After incorporating imputed fitness observations into the mixed model (Approach 3, right column, Figure 4), we found p-value distributions hewed closer to the theoretical expectation and were slightly conservative (*i.e.* under-enriched in low p-values) for two climate variables. These three approaches tended to identified different SNPs as having the strongest SNP×environment associations with fitness (Table S5).

Figure 4. Quantile-quantile plots of p-value distributions for three approaches to calculating genome-wide SNP×environment associations with fitness. X-axes show expected $-\log_{10}(p)$ and y-axes show observed $-\log_{10}(p)$.



Based on the results of our simulations and the p-value distributions noted above, we focus the remainder of analyses on results from mixed models with imputed fitness included (Approach 3). We found that climate variables differed in the importance of
455 kinship-climate interaction associations with fitness (*i.e*. proportion of variance in fitness explained by random effects ***v***), suggesting that population structure in Arabidopsis is more strongly correlated with some climatic axes of local adaptation (G×E for fitness) compared to other climate gradients. For growing season minimum temperatures, kinship×environment interactions explained most of the variation in fitness (R^2 =0.78,
460 Table 1, row 3). By contrast, kinship×environment interactions for fitness were weaker along a gradient in winter minimum temperature (R^2 =0.07).

Approach 3 tended to identify SNPs where SNP×environment interactions favored alleles in climates where they were relatively more common, that is, the sign of allelic differences in home climates were mirrored by the sign of fitted mixed model
465 SNP×environment associations with relative fitness (Table 1, row 1, and see outlier examples in Figure 5). In addition to characterizing SNP×environment associations, we tended to identify SNPs where we estimated a rank change in relative fitness for alternate alleles along the environmental gradient between the two extreme common gardens (*i.e*. where the fitted model expectation was that the allele with higher fitness at one extreme
470 common garden differed from the allele with higher fitness at the other extreme, Table 1, row 2). It appeared that the proportion of SNPs expected to show rank changes in relative fitness among the common gardens was related to how much of each climate variable's range was covered by gardens (Table 1, row 4). Thus the common gardens may have

been limited in their ability to capture rank changing of alleles at some loci involved in local adaptation to aridity and growing season cold.

Table 1. Characterization of patterns identified by Approach 3 (mixed model including imputations) for SNPs in 0.01 lower tail of p-values for SNP×environment interactions for fitness (first two rows of table) and for kinship×environment interactions for fitness (third row). SNPs with home allele advantage are defined as those where the sign of allelic differences in home climates were mirrored by the sign of fitted mixed model SNP×environment associations with relative fitness. Rank changing SNPs are those where we estimated a rank change in relative fitness for alternate alleles along the environmental gradient between the two extreme common gardens. The final row gives the proportion of total observed climate gradient (among ecotypes) captured by the two most extreme common gardens.

Statistic	*Aridity*	*CV grow. seas. prec.*	*Min. temp.*	*Min. temp. grow. seas.*
Proportion SNP×E with home allele advantage	0.46	0.89	0.92	>0.99
Proportion SNP×E rank changing	0.31	0.92	>0.99	0.73
Kinship×E R^2 for fitness	0.08	0.58	0.07	0.43
Proportion climate gradient covered by gardens	0.13	0.65	0.78	0.31

We found non-random, but very weak overlap between the SNPs we identified and those outliers in previous analyses (Hancock *et al.* 2011; Lasky *et al.* 2012a, 2014). When considering mixed model (Lasky *et al.* 2014) or partial Mantel (Hancock *et al.* 2011) SNP associations with the same climate variables (*i.e.* genome-environment associations with no common garden data), we found significant overlap among the previously identified SNPs in the 0.01 lower tail of p-values versus those in the 0.01 tail identified here (permutation test, all $p<0.05$, Table S6). However, rank correlations among SNP scores from previous approaches versus our current approach were very weak (all rho < 0.2, Table S6).

SNP×environment associations with fitness are enriched in regions suggestive of local adaptation

To assess whether our approach identified certain types of SNPs, we tested for enrichment of genic and intergenic regions for SNP×environment effects on fitness (again focusing on Approach 3: mixed model including imputations). We found that SNP×environment interactions for fitness were significantly enriched in genic regions (Table 2; for reference, SNPs identified via Approach 1, without random effects, were not significantly enriched in genic regions). Additionally, we found that SNP×environment interactions for fitness were enriched for high pairwise haplotype sharing (PHS) and high integrated extended haplotype homozygosity (iHS, Table 2). Finally, SNPs associated with G×E for flowering time response to growing temperature (identified in Li *et al.* 2014) tended to also have strong SNP×growing season minimum temperature interactions for fitness ($p < 0.0002$) but not for other climate variables. (Table 2) Enrichments reported above did not change qualitatively (with respect to statistical significance) when we only considered SNPs with MAF > 0.1.

Table 2. Permutation tests of enrichment p-values for various signals suggestive of local adaptation to climate. For each statistic, we tested for enrichment of signal in the SNPs in the 0.01 lower tail of p-values for SNP×environment associations with relative fitness. "Genic" tests enrichment of genic versus non-genic SNPs, "PHS" and "iHS" test for enrichment with significant (alpha = 0.05) pairwise haplotype sharing and standardized integrated extended haplotype homozygosity, respectively. The final row shows enrichment with SNPs having significant (alpha = 0.05) associations with change in flowering time (G×E) in response to warming during growth.

Statistic	*Aridity enrichment*	*CV grow. seas. prec. enrichment*	*Min. temp. enrichment*	*Min. temp. grow. seas. enrichment*
Genic	0.0036	<0.0002	0.0054	0.0006
PHS	0.0272	0.0122	0.0068	0.0272
iHS	0.2840	0.0006	0.0020	0.0008
Flowering time under warming, G×E	0.8582	0.0680	0.5844	<0.0002

520

SNP×environment associations with fitness identify genes potentially involved in local adaptation

Our approach identified a number of strong candidates for local adaptation at the top of lists of SNPs with the strongest SNP×environment associations with relative fitness

525 (Tables S1-S4). For example, the top SNP associated with aridity interaction effects on fitness (chr. 4, position 11005059) fell within LESION SIMULATING DISEASE 1 (LSD1), which affects a number of traits in Arabidopsis, including survival and fecundity under drought (Wituszyńska *et al.* 2013; Szechyńska-Hebda *et al.* 2016) (Figure 5), while the third SNP (chr. 2, position 7592008) fell within ATMLO8, MILDEW RESISTANCE

530 LOCUS O 8, homologous with barley MLO which controls resistance to the fungal pathogen powdery mildew (Büschges *et al.* 1997). The top SNP associated with winter cold interaction effects on fitness (chr. 5, position 7496047) falls within coding region of WRKY38, involved in the salicylic acid pathway and pathogen defense (Kim *et al.* 2008), and was the same locus identified as most strongly associated with multivariate

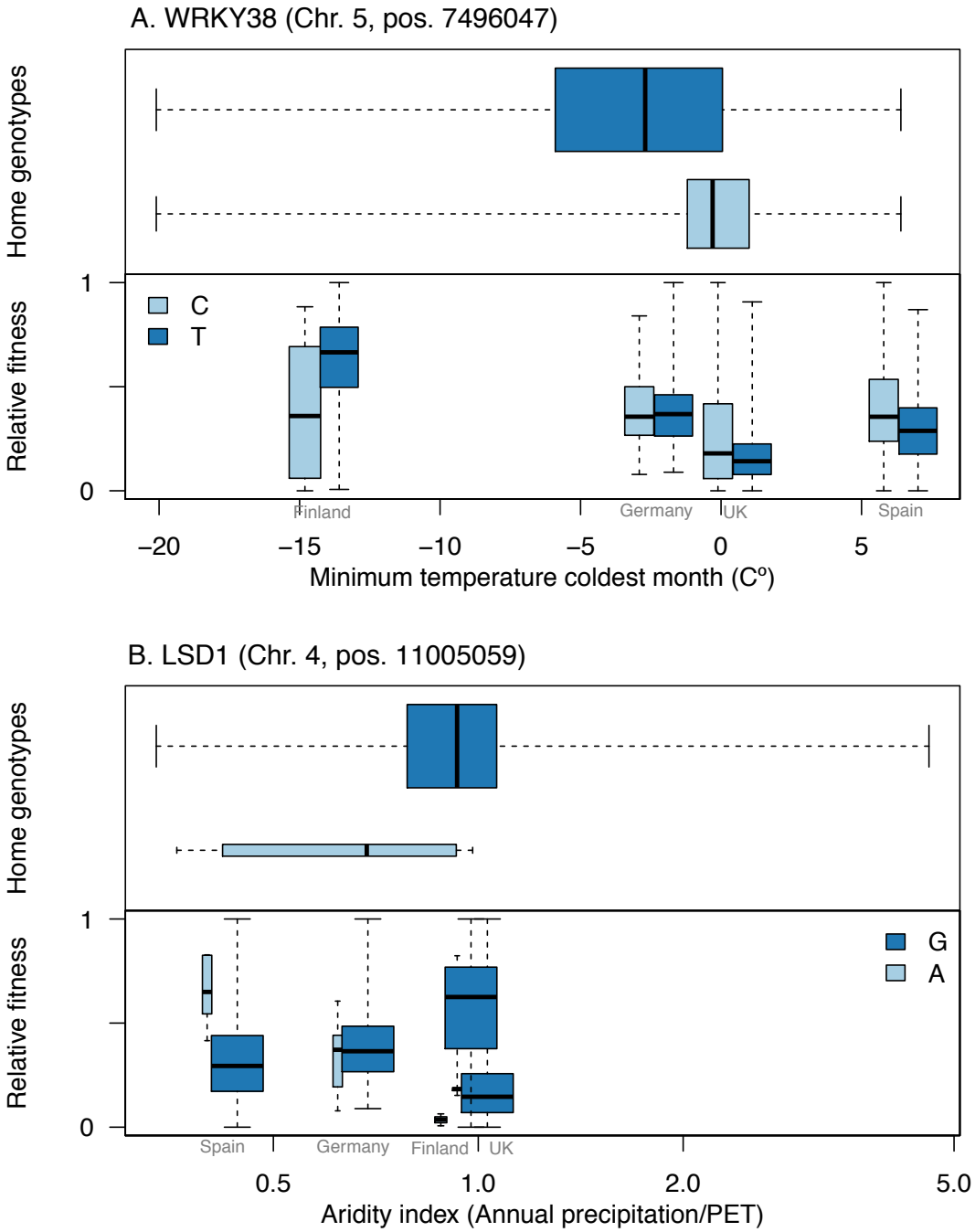
535 climate in Lasky et al. (2012) (Figure 5). The top SNP associated with variability in growing season precipitation interaction effects on fitness (chr. 2, position 18504858) falls 380 bp from ABA HYPERSENSITIVE GERMINATION 11, AHG11, which mediates the effect of abscisic acid (ABA), a major hormone of abiotic stress response, on germination (Murayama *et al.* 2012). The fifth highest SNP (and second highest locus)

540 associated with growing season cold interaction effects on fitness (chr. 3, position 8454439) fell within ABERRANT LATERAL ROOT FORMATION 5, ALF5, a gene

that confers resistance to toxins (Diener *et al.* 2001) belonging to the MATE gene family, which play a variety of roles responding to environment (Shoji 2014).

545

Figure 5. Example SNPs with the strongest associations (i.e. lowest p-value) with cold winter temperatures (A) and aridity (B). Top subpanels show the climate distribution of alleles in home genotypes (*i.e*. natural geographic patterns), known as genotype-environment associations. Bottom subpanels show relative fitness of alleles in four common gardens, where common gardens' climates determine position on x-axes. Each SNP falls within the coding region of indicated genes (WRKY38 and LSD1). Boxes are scaled to number of accessions with each allele. In both (A) and (B), the allele with the greatest relative fitness in common gardens changes along the environmental gradient consistent with change in allele frequency in native accessions (i.e. ecotypes).



Discussion

Genetic variation in environmental responses (G×E) is ubiquitous but poorly understood at large spatial scales, *e.g*. across a species range. Replicated common garden experiments and genome scans for loci exhibiting evidence for local adaptation have been important in understanding the genetic basis of G×E and local adaptation (Hancock *et al.* 2008; ECKERT *et al.* 2010; Turner *et al.* 2010; Fournier-Level *et al.* 2011a; Lasky *et al.* 2012a, 2015; Ågren *et al.* 2013; Evans *et al.* 2014). However, the complementary information in common gardens and geographic variation in allele frequency have not been coherently synthesized. Previous association studies of G×E have modeled discrete, categorical environmental effects (Murcray *et al.* 2009; Thomas 2010; Korte *et al.* 2012; Marigorta and Gibson 2014). Here, we demonstrated an approach to association study of G×E for fitness and an imputation technique that allowed us to coherently synthesize evidence from common gardens and genome-environment associations. Our imputation method relied on making explicit the often implicit assumption of local adaptation that underlies genome-environment association studies (Coop *et al.* 2010; Hancock *et al.* 2011; Lasky *et al.* 2012a). Using simulation, we demonstrated that this imputation can increase power to identify SNPs causing G×E for fitness and local adaptation. Our approach also identified strong candidate genes in Arabidopsis associated with SNPs that exhibit fitness tradeoffs along climate gradients such that locally common alleles had greater relative fitness. An advantage of studying Arabidopsis was that we had published measures of fitness, whereas below we discuss how our approach could be applied when only data on components of fitness or adaptive traits are available.

580 *Model extensions*

Above we described a method of imputation based on the assumption of local adaptation, *i.e*. home genotypes had greater fitness than away genotypes. However, local adaptation in nature is typically imperfect, such that the optimal genotype for a given location might not be the home genotype (Leimu and Fischer 2008; Hereford 2009). Local adaptation
585 may fail due to immigration of maladaptive alleles (Slatkin 1973), limited genetic variation (Barton 2001), and other processes (Bridle and Vines 2007). Thus our imputation can be considered a heuristic to be improved by further development.

We consider two approaches that would extend the generality of our approach by treating relative fitness as a parameter rather than imputed data. First, instead of assuming
590 that each sequenced genotype is most fit in its home environment, an alternative approach could treat the unobserved fitness of home genotypes as a free parameter. To constrain estimates of unobserved fitness one could use informative priors, such that the prior probability of relative fitness at home for each genotype would be monotonically increasing, *i.e.* local adaptation is the most likely state, but minor maladaptation is
595 expected to be common. Inferences about unobserved fitness could be further constrained using hierarchical models, such that home fitness parameters for multiple genotypes arise from a distribution (GELMAN and HILL 2007). Second, for situations in which fitness is not measured, components of fitness (*e.g.* survival) or traits thought to be locally adaptive (e.g. physiological or behavioral) can be measured and used to infer the genomic basis of
600 local adaptation. For example, instead of modeling SNP×environment associations with fitness, one could model SNP×environment associations with components of fitness or adaptive traits measured in common gardens, and estimate unobserved traits for

sequenced genotypes using informative priors. To be clear, in our case study of Arabidopsis, we had near but not complete lifetime fitness data (missing germination stage). Here we do not attempt to parameterize these model extensions, given the current computational challenge of fitting many more parameters in a Bayesian framework.

Genotype-by-environment interactions in genome-wide association studies

Recent advances in association models have included explicit modeling of G×E (Murcray *et al.* 2009; Thomas 2010; Korte *et al.* 2012; Marigorta and Gibson 2014; Li *et al.* 2014; Kooperberg *et al.* 2016; Windle 2016), but to our knowledge there are no published genome-wide association studies accounting for SNP interactions with continuous environmental gradients. Some of the aforementioned categorical treatments of SNP×environment interactions were used in association studies for human disease. However, many of the environmental variables that may mediate genetic risk of disease are continuous in nature, *e.g*. exposure to ultraviolet radiation and tobacco smoke. Future research on local adaptation and human disease may benefit from exchange of approaches given the shared importance across disciplines of understanding the genomic basis of G×E.

Case study on Arabidopsis thaliana

Our approach leveraged information from both common gardens and geographic patterns in allele frequency, which simulations indicated can increase power and accuracy to detect variants driving local adaptation. Furthermore, although there was overlap with signal identified by previous approaches using the same data (Hancock *et al.* 2011;

Lasky *et al.* 2012a, 2014), overlap was generally weak, indicating our approach identified distinct loci as causing local adaptation. In our case study on Arabidopsis, the SNPs that exhibited the strongest evidence for SNP×climate interaction effects on fitness often fell within the coding regions of strong candidate genes based on known roles in environmental responses, suggesting our approach is a useful for identifying loci underlying local adaptation.

Our approach identified many SNPs where allelic variation was associated with rank-changing relative fitness tradeoffs along climate gradients (*e.g.* all 214 of the SNPs with strongest interaction, *i.e.* 0.001 quantile, with winter minimum temperature association for fitness), loci where selective gradients may maintain population differentiation (Anderson *et al.* 2011b; Ågren *et al.* 2013). Studies of local adaptation genomics often find limited evidence for loci with antagonistic pleiotropy. A previous study of the common garden data used here (Fournier-Level *et al.* 2011a) found that the SNPs with the strongest association with fitness in one common garden were rarely among those with the strongest associations in another garden, which may be evidence for conditional neutrality. By contrast with previous approaches that model phenotypes at a single site, our model was explicitly focused on detecting alleles with the strongest evidence for SNP×climate interactions favoring home alleles. Thus using our model, loci with patterns indicative of antagonistic pleiotropy were most likely to be detected. Additionally, local adaptation to many climate gradients may involve evolution of complex traits governed by variants at many loci. Thus loci exhibiting antagonistic pleiotropy and loci exhibiting G×E but no tradeoffs may both underlie genome-level local adaptation. Note that our study, like that of (Fournier-Level *et al.* 2011a) is based on

association mapping, which may suffer from identification of more false positives
650 compared with linkage mapping approaches (HALL *et al.* 2010; ANDERSON *et al.* 2013; Ågren *et al.* 2013). Follow-up experimental study of phenotypic effects of variation at individual loci is required to confirm the results of association mapping (*e.g.* Verslues *et al.* 2014; Broekgaarden *et al.* 2015).

We found evidence that SNP×climate interaction effects on fitness were enriched
655 in genic regions, suggesting that our model captured a signal of local adaptation rather than population structure. We found that enrichments in genic SNPs only emerged after using a mixed model to control for the putative effects of population structure (*i.e.* genome-wide similarity), suggesting that the genic-enriched patterns of divergence we modeled were not simply associated with overall patterns of among-population
660 divergence. This enrichment is consistent with other findings in Arabidopsis (Hancock *et al.* 2011; Lasky *et al.* 2012a) and other species ((Coop *et al.* 2009; Fumagalli *et al.* 2011; Lasky *et al.* 2015), but see (Pyhäjärvi *et al.* 2013)). We do not interpret this enrichment as indicating that changes in amino acid sequences are more important than regulatory evolution in local adaptation, but rather as supporting the hypothesis that local adaptation
665 is more likely to involve sequence evolution near genes as opposed to at locations farther from genes, where many intergenic SNPs are found.

We found evidence that loci we identified as candidates for local adaptation were enriched in evidence for partial selective sweeps (PHS and iHS statistics), suggesting that recent local sweeps in particular environments are an important mode of local adaptation
670 (Voight *et al.* 2006; Toomajian *et al.* 2006). That many locally adaptive variants were swept recently may be expected based on the range dynamics of Arabidopsis, which has

colonized much of its Eurasian range following the retreat of glaciers (Sharbel *et al.* 2000), a process that may have involved recent local adaptation. It is important to note that extended haplotype patterns suggestive of partial sweeps may occur at the shoulders (*i.e*. away from causal loci) of complete sweeps (Schrider *et al.* 2015), thus caution is warranted in attributing our observed PHS and iHS enrichment to localized sweeps versus global sweeps at nearby loci.

Finally, we found significant overlap between SNPs associated with G×E for fitness along growing season cold gradients and SNPs associated with G×E for flowering time across growing season temperature treatments (Li *et al.* 2014). Our findings suggest that evolution of plasticity in flowering time is a mechanism of local adaptation along growing season cold gradients and that our model has captured the signal of this adaptation. For organisms inhabiting seasonal environments, timing of the life cycle may have large impacts on fitness. Previous common garden experiments have provided strong evidence that flowering time is a central trait involved in local adaptation (Hall and Willis 2006; Franks *et al.* 2007; Keller *et al.* 2012; Lowry *et al.* 2014) with molecular study further supporting the role of flowering time (Stinchcombe *et al.* 2004; Caicedo *et al.* 2004; Shindo *et al.* 2005; Lovell *et al.* 2013) and the role of plasticity (Fraser 2013; Lasky *et al.* 2014) in local adaptation.

Conclusions

Local adaptation to environment involves genotype-by-environment interactions for fitness. Genome-wide association studies are a promising approach for identifying the

genomic basis of local adaptation and G×E. Additional approaches, *e.g*. genome-wide expression profiling, may also be useful for uncovering the genomic basis of local adaptation (Des Marais *et al.* 2013). Future approaches that use a principled basis for quantitative synthesis of these data types may enhance our ability to characterize adaptation in an integrative fashion.

Acknowledgements

We thank David Lowry, Thomas Juenger, and two anonymous reviewers for comments on an earlier version of this manuscript.

Literature Cited

Adrion J. R., Hahn M. W., Cooper B. S., 2015 Revisiting classic clines in Drosophila melanogaster in the age of genomics. Trends Genet.

Ågren J., Oakley C. G., McKay J. K., Lovell J. T., Schemske D. W., 2013 Genetic mapping of adaptation reveals fitness tradeoffs in Arabidopsis thaliana. Proc. Natl. Acad. Sci. **110**: 21077–21082.

Aitken S. N., Whitlock M. C., 2013 Assisted Gene Flow to Facilitate Local Adaptation to Climate Change. Annu. Rev. Ecol. Evol. Syst. **44**: 367–388.

Anastasi A., 1958 Heredity, environment, and the question“ How?” Psychol. Rev. **65**: 197.

Anastasio A. E., Platt A., Horton M., Grotewold E., Scholl R., Borevitz J. O., Nordborg M., Bergelson J., 2011 Source verification of mis-identified Arabidopsis thaliana accessions. Plant J. **67**: 554–566.

Anderson J. T., Lee C.-R., Mitchell-Olds T., 2011a Life history QTLs and natural selection on flowering time in Boechera stricta, a perennial relative of Arabidopsis. Evol. Int. J. Org. Evol. **65**: 771–787.

Anderson J. T., Willis J. H., Mitchell-Olds T., 2011b Evolutionary genetics of plant adaptation. Trends Genet. **27**: 258–266.

ANDERSON J. T., LEE C.-R., RUSHWORTH C. A., COLAUTTI R. I., MITCHELL-OLDS T., 2013 Genetic trade-offs and conditional neutrality contribute to local adaptation. Mol. Ecol. **22**: 699–708.

Atwell S., Huang Y. S., Vilhjalmsson B. J., Willems G., Horton M., Li Y., Meng D., Platt A., Tarone A. M., Hu T. T., Jiang R., Muliyati N. W., Zhang X., Amer M. A., Baxter I., Brachi B., Chory J., Dean C., Debieu M., Meaux J. de, Ecker J. R., Faure N., Kniskern J. M., Jones J. D. G., Michael T., Nemri A., Roux F., Salt D. E., Tang C., Todesco M., Traw M. B., Weigel D., Marjoram P., Borevitz J. O., Bergelson J., Nordborg M., 2010 Genome-wide association study of 107 phenotypes in Arabidopsis thaliana inbred lines. Nature **465**: 627–631.

Barton N. H., 2001 Adaptation at the edge of a species' range. Spec. Publ.-Br. Ecol. Soc. **14**: 365–392.

Berg J. J., Coop G., 2014 A Population Genetic Signal of Polygenic Adaptation. PLoS Genet **10**: e1004412.

Bragg J. G., Supple M. A., Andrew R. L., Borevitz J. O., 2015 Genomic variation across landscapes: insights and applications. New Phytol.: n/a-n/a.

740 Bridle J. R., Vines T. H., 2007 Limits to evolution at range margins: when and why does adaptation fail? Trends Ecol. Evol. **22**: 140–147.

Broekgaarden C., Bucher J., Bac-Molenaar J., Keurentjes J. J. B., Kruijer W., Voorrips R. E., Vosman B., 2015 Novel Genes Affecting the Interaction between the Cabbage Whitefly and Arabidopsis Uncovered by Genome-Wide Association
745 Mapping. PLoS ONE **10**: e0145124.

Brown K. G., 1976 Asymptotic Behavior of Minque-Type Estimators of Variance Components. Ann Stat.: 746–754.

Büschges R., Hollricher K., Panstruga R., Simons G., Wolter M., Frijters A., Daelen R. van, Lee T. van der, Diergaarde P., Groenendijk J., Töpsch S., Vos P., Salamini
750 F., Schulze-Lefert P., 1997 The Barley Mlo Gene: A Novel Control Element of Plant Pathogen Resistance. Cell **88**: 695–705.

Caicedo A. L., Stinchcombe J. R., Olsen K. M., Schmitt J., Purugganan M. D., 2004 Epistatic interaction between Arabidopsis FRI and FLC flowering time genes generates a latitudinal cline in a life history trait. Proc. Natl. Acad. Sci. U. S. A.
755 **101**: 15670–15675.

Clausen J., Keck D. D., Hiesey W. M., 1940 Experimental Studies on the Nature of Species. I. Effect of Varied Environments on Western North American Plants. Carnegie Institution of Washington Publication No. 520. Wash. DC.

Clausen J., Keck D. D., HIESEY W. M., 1948 Experimental studies on the nature of species. III. Environresponses of climatic races of Achillea. Exp. Stud. Nat. Species III Environresponses Clim. Races Achillea.

Coop G., Pickrell J. K., Novembre J., Kudaravalli S., Li J., Absher D., Myers R. M., Cavalli-Sforza L. L., Feldman M. W., Pritchard J. K., 2009 The Role of Geography in Human Adaptation. PLoS Genet **5**: e1000500.

Coop G., Witonsky D., Di Rienzo A., Pritchard J. K., 2010 Using Environmental Correlations to Identify Loci Underlying Local Adaptation. Genetics **185**: 1411–1423.

Des Marais D. L., Hernandez K. M., Juenger T. E., 2013 Genotype-by-Environment Interaction and Plasticity: Exploring Genomic Responses of Plants to the Abiotic Environment. Annu. Rev. Ecol. Evol. Syst. **44**: 5–29.

Devlin B., Roeder K., 1999 Genomic Control for Association Studies. Biometrics **55**: 997–1004.

Diener A. C., Gaxiola R. A., Fink G. R., 2001 Arabidopsis ALF5, a Multidrug Efflux Transporter Gene Family Member, Confers Resistance to Toxins. Plant Cell **13**: 1625–1637.

ECKERT A. J., BOWER A. D., GONZÁLEZ-MARTÍNEZ S. C., WEGRZYN J. L., COOP G., NEALE D. B., 2010 Back to nature: ecological genomics of loblolly pine (Pinus taeda, Pinaceae). Mol. Ecol. **19**: 3789–3805.

Evans L. M., Slavov G. T., Rodgers-Melnick E., Martin J., Ranjan P., Muchero W., Brunner A. M., Schackwitz W., Gunter L., Chen J.-G., Tuskan G. A., DiFazio S. P., 2014 Population genomics of Populus trichocarpa identifies signatures of selection and adaptive trait associations. Nat Genet **46**: 1089–1096.

Fitzpatrick M. C., Keller S. R., 2015 Ecological genomics meets community-level modelling of biodiversity: mapping the genomic landscape of current and future environmental adaptation. Ecol. Lett. **18**: 1–16.

Forester B. R., Jones M. R., Joost S., Landguth E. L., Lasky J. R., 2016 Detecting spatial genetic signatures of local adaptation in heterogeneous landscapes. Mol. Ecol. **25**: 104–120.

Fournier-Level A., Korte A., Cooper M. D., Nordborg M., Schmitt J., Wilczek A. M., 2011a A Map of Local Adaptation in Arabidopsis thaliana. Science **334**: 86–89.

Fournier-Level A., Korte A., Cooper MD, Nordborg M, Schmitt J, Wilczek AM, 2011b Data from: A map of local adaptation in Arabidopsis thaliana.

Franks S. J., Sim S., Weis A. E., 2007 Rapid evolution of flowering time by an annual plant in response to a climate fluctuation. Proc. Natl. Acad. Sci. **104**: 1278–1282.

Fraser H. B., 2013 Gene expression drives local adaptation in humans. Genome Res. **23**: 1089–1096.

Fumagalli M., Sironi M., Pozzoli U., Ferrer-Admettla A., Pattini L., Nielsen R., 2011 Signatures of Environmental Genetic Adaptation Pinpoint Pathogens as the Main Selective Pressure through Human Evolution. PLoS Genet **7**: e1002355.

Gage S. H., Davey Smith G., Ware J. J., Flint J., Munaf M. R., 2016 G = E: What GWAS Can Tell Us about the Environment. PLoS Genet **12**: e1005765.

Gauderman W. J., Zhang P., Morrison J. L., Lewinger J. P., 2013 Finding Novel Genes by Testing G×E Interactions in a Genomewide Association Study. Genet. Epidemiol. **37**: 603–613.

Gautier, M., Vitalis, R., 2012 rehh: An R package to detect footprints of selection in genome-wide SNP data from haplotype structure. Bioinformatics **0**: accepted.

GELMAN A., HILL J., 2007 *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge University Press, Cambridge.

Hall M. C., Willis J. H., 2006 DIVERGENT SELECTION ON FLOWERING TIME CONTRIBUTES TO LOCAL ADAPTATION IN MIMULUS GUTTATUS POPULATIONS. Evolution **60**: 2466–2477.

HALL M. C., LOWRY D. B., WILLIS J. H., 2010 Is local adaptation in Mimulus guttatus caused by trade-offs at individual loci? Mol. Ecol. **19**: 2739–2753.

Hancock A. M., Witonsky D. B., Gordon A. S., Eshel G., Pritchard J. K., Coop G., Di Rienzo A., 2008 Adaptations to Climate in Candidate Genes for Common Metabolic Disorders. PLoS Genet **4**: e32.

Hancock A. M., Brachi B., Faure N., Horton M. W., Jarymowycz L. B., Sperone F. G., Toomajian C., Roux F., Bergelson J., 2011 Adaptation to Climate Across the Arabidopsis thaliana Genome. Science **334**: 83–86.

Hedrick P. W., Ginevan M. E., Ewing E. P., 1976 Genetic Polymorphism in Heterogeneous Environments. Annu. Rev. Ecol. Syst. **7**: 1–32.

Hereford J., 2009 A Quantitative Survey of Local Adaptation and Fitness Trade-Offs. Am. Nat. **173**: 579–588.

Hijmans R. J., Cameron S. E., Parra J. L., Jones P. G., Jarvis A., 2005 Very high resolution interpolated climate surfaces for global land areas. Int. J. Climatol. **25**: 1965–1978.

Hoban S., Kelley J. L., Lotterhos K. E., Antolin M. F., Bradburd G., Lowry D. B., Poss M. L., Reed L. K., Storfer A., Whitlock M. C., 2016 Finding the Genomic Basis of Local Adaptation: Pitfalls, Practical Solutions, and Future Directions. Am. Nat. **188**: 379–397.

Hoffmann M. H., 2002 Biogeography of Arabidopsis thaliana (L.) Heynh. (Brassicaceae). J. Biogeogr. **29**: 125–134.

Horton M. W., Hancock A. M., Huang Y. S., Toomajian C., Atwell S., Auton A., Muliyati N. W., Platt A., Sperone F. G., Vilhjalmsson B. J., Nordborg M., Borevitz J. O., Bergelson J., 2012 Genome-wide patterns of genetic variation in worldwide Arabidopsis thaliana accessions from the RegMap panel. Nat Genet **44**: 212–216.

Hunter D. J., 2005 Gene-environment interactions in human diseases. Nat Rev Genet **6**: 287–298.

Jones F. C., Grabherr M. G., Chan Y. F., Russell P., Mauceli E., Johnson J., Swofford R., Pirun M., Zody M. C., White S., Birney E., Searle S., Schmutz J., Grimwood J., Dickson M. C., Myers R. M., Miller C. T., Summers B. R., Knecht A. K., Brady S. D., Zhang H., Pollen A. A., Howes T., Amemiya C., Lander E. S., Di Palma F., Lindblad-Toh K., Kingsley D. M., 2012 The genomic basis of adaptive evolution in threespine sticklebacks. Nature **484**: 55–61.

Kang H. M., Zaitlen N. A., Wade C. M., Kirby A., Heckerman D., Daly M. J., Eskin E., 2008 Efficient Control of Population Structure in Model Organism Association Mapping. Genetics **178**: 1709–1723.

Kang H. M., Sul J. H., Service S. K., Zaitlen N. A., Kong S., Freimer N. B., Sabatti C., Eskin E., 2010 Variance component model to account for sample structure in genome-wide association studies. Nat Genet **42**: 348–354.

Kawecki T. J., Ebert D., 2004 Conceptual issues in local adaptation. Ecol. Lett. **7**: 1225–1241.

Keller S. R., Levsen N., Olson M. S., Tiffin P., 2012 Local Adaptation in the Flowering-Time Gene Network of Balsam Poplar, Populus balsamifera L. Mol. Biol. Evol. **29**: 3143–3152.

Keller M. C., 2014 Gene × Environment Interaction Studies Have Not Properly Controlled for Potential Confounders: The Problem and the (Simple) Solution. Temperam. Genet. Environ. Factors **75**: 18–24.

Kim S., Plagnol V., Hu T. T., Toomajian C., Clark R. M., Ossowski S., Ecker J. R., Weigel D., Nordborg M., 2007 Recombination and linkage disequilibrium in Arabidopsis thaliana. Nat Genet **39**: 1151–1155.

Kim K.-C., Lai Z., Fan B., Chen Z., 2008 Arabidopsis WRKY38 and WRKY62 Transcription Factors Interact with Histone Deacetylase 19 in Basal Defense. Plant Cell Online **20**: 2357–2371.

Kooperberg C., Dai J. Y., Hsu L., 2016 Two-stage procedures for the identification of gene x environment and gene x gene interactions in genome-wide association studies. In: *Statistical approaches to gene x environment interactions for complex phenotypes*, MIT Press, Cambridge, Massachusetts, pp. 11–34.

Korte A., Vilhjalmsson B. J., Segura V., Platt A., Long Q., Nordborg M., 2012 A mixed-model approach for genome-wide association studies of correlated traits in structured populations. Nat Genet **44**: 1066–1071.

875 Landguth E. L., Cushman S. A., 2010 cdpop: A spatially explicit cost distance population genetics program. Mol. Ecol. Resour. **10**: 156–161.

Langlet O., 1971 Two hundred years genecology. Taxon: 653–721.

Lasky J. R., Des Marais D. L., McKay J. K., Richards J. H., Juenger T. E., Keitt T. H., 2012a Characterizing genomic variation of Arabidopsis thaliana: the roles of
880 geography and climate. Mol. Ecol. **21**: 5512–5529.

Lasky J. R., Des Marais DL, McKay JK, Richards JH, Juenger TE, Keitt TH, 2012b Data from: Characterizing genomic variation of Arabidopsis thaliana: the roles of geography and climate.

Lasky J. R., Des Marais D. L., Lowry D. B., Povolotskaya I., McKay J. K., Richards J. H.,
885 Keitt T. H., Juenger T. E., 2014 Natural variation in abiotic stress responsive gene expression and local adaptation to climate in Arabidopsis thaliana. Mol. Biol. Evol. **31**: 2283–2296.

Lasky J. R., Upadhyaya H. D., Ramu P., Deshpande S., Hash C. T., Bonnette J., Juenger T. E., Hyma K., Acharya C., Mitchell S. E., Buckler E. S., Brenton Z., Kresovich S.,
890 Morris G. P., 2015 Genome-environment associations in sorghum landraces predict adaptive traits. Sci. Adv. **1**: e1400218.

Leimu R., Fischer M., 2008 A Meta-Analysis of Local Adaptation in Plants. PLoS ONE **3**: e4010.

Levene H., 1953 Genetic Equilibrium When More Than One Ecological Niche is Available. Am. Nat. **87**: 331–333.

Li Y., Cheng R., Spokas K. A., Palmer A. A., Borevitz J. O., 2014 Genetic Variation for Life History Sensitivity to Seasonal Warming in Arabidopsis thaliana. Genetics **196**: 569–577.

Lovell J. T., Juenger T. E., Michaels S. D., Lasky J. R., Platt A., Richards J. H., Yu X., Easlon H. M., Sen S., McKay J. K., 2013 Pleiotropy of FRIGIDA enhances the potential for multivariate adaptation. Proc. R. Soc. B Biol. Sci. **280**: 20131043.

Lowry D. B., Willis J. H., 2010 A Widespread Chromosomal Inversion Polymorphism Contributes to a Major Life-History Transition, Local Adaptation, and Reproductive Isolation. PLoS Biol **8**: e1000500.

Lowry D., Behrman K., Grabowski P., Morris G., Kiniry J., Juenger T., 2014 Adaptations between ecotypes and along environmental gradients in Panicum virgatum. Am. Nat.

Manel S., Holderegger R., 2013 Ten years of landscape genetics. Trends Ecol. Evol. **28**: 614–621.

Marigorta U. M., Gibson G., 2014 A simulation study of gene-by-environment interactions in GWAS implies ample hidden effects. Front. Genet. **5**.

Mitchell-Olds T., Willis J. H., Goldstein D. B., 2007 Which evolutionary processes influence natural genetic variation for phenotypic traits? Nat Rev Genet **8**: 845–856.

Murayama M., Hayashi S., Nishimura N., Ishide M., Kobayashi K., Yagi Y., Asami T., Nakamura T., Shinozaki K., Hirayama T., 2012 Isolation of Arabidopsis ahg11, a weak ABA hypersensitive mutant defective in nad4 RNA editing. J. Exp. Bot. **63**: 5301–5310.

Murcray C. E., Lewinger J. P., Gauderman W. J., 2009 Gene-Environment Interaction in Genome-Wide Association Studies. Am. J. Epidemiol. **169**: 219–226.

Nordborg M., Weigel D., 2008 Next-generation genetics in plants. Nature **456**: 720–723.

Oppen M. J. H. van, Oliver J. K., Putnam H. M., Gates R. D., 2015 Building coral reef resilience through assisted evolution. Proc. Natl. Acad. Sci. **112**: 2307–2313.

Price A. L., Zaitlen N. A., Reich D., Patterson N., 2010 New approaches to population stratification in genome-wide association studies. Nat Rev Genet **11**: 459–463.

Pyhäjärvi T., Hufford M. B., Mezmouk S., Ross-Ibarra J., 2013 Complex Patterns of Local Adaptation in Teosinte. Genome Biol. Evol. **5**: 1594–1609.

Rao C. R., 1971 Minimum variance quadratic unbiased estimation of variance components. J. Multivar. Anal. **1**: 445–456.

Reimherr M., Nicolae D., 2015 Estimating Variance Components in Functional Linear Models with Applications to Genetic Heritability. J. Am. Stat. Assoc.: 00–00.

Rellstab C., Gugerli F., Eckert A. J., Hancock A. M., Holderegger R., 2015 A practical guide to environmental association analysis in landscape genomics. Mol. Ecol. **24**: 4348–4370.

Schaffer H. E., Johnson F. M., 1974 ISOZYME ALLELIC FREQUENCIES RELATED TO SELECTION AND GENE-FLOW HYPOTHESES. Genetics **77**: 163–168.

Schoville S. D., Bonin A., François O., Lobreaux S., Melodelima C., Manel S., 2012 Adaptive Genetic Variation on the Landscape: Methods and Cases. Annu. Rev. Ecol. Evol. Syst. **43**: 23–43.

Schrider D. R., Mendes F. K., Hahn M. W., Kern A. D., 2015 Soft Shoulders Ahead: Spurious Signatures of Soft and Partial Selective Sweeps Result from Linked Hard Sweeps. Genetics **200**: 267–284.

Sharbel T. F., Haubold B., Mitchell-Olds T., 2000 Genetic isolation by distance in Arabidopsis thaliana: biogeography and postglacial colonization of Europe. Mol. Ecol. **9**: 2109–2118.

Shindo C., Aranzana M. J., Lister C., Baxter C., Nicholls C., Nordborg M., Dean C., 2005 Role of FRIGIDA and FLOWERING LOCUS C in Determining Variation in Flowering Time of Arabidopsis. Plant Physiol **138**: 1163–1173.

Shoji T., 2014 ATP-binding cassette and multidrug and toxic compound extrusion transporters in plants: A common theme among diverse detoxification mechanisms. Int Rev Cell Mol Biol **309**: 303–346.

Slatkin M., 1973 GENE FLOW AND SELECTION IN A CLINE. Genetics **75**: 733–756.

955 Stinchcombe J. R., Weinig C., Ungerer M., Olsen K. M., Mays C., Halldorsdottir S. S., Purugganan M. D., Schmitt J., 2004 A latitudinal cline in flowering time in Arabidopsis thaliana modulated by the flowering time gene FRIGIDA. Proc. Natl. Acad. Sci. U. S. A. **101**: 4712–4717.

Szechyńska-Hebda M., Czarnocka W., Hebda M., Karpiński S., 2016 PAD4, LSD1 and
960 EDS1 regulate drought tolerance, plant biomass production, and cell wall properties. Plant Cell Rep.: 1–13.

Thomas D., 2010 Methods for Investigating Gene-Environment Interactions in Candidate Pathway and Genome-Wide Association Studies. Annu. Rev. Public Health **31**: 21–36.

965 Tiffin P., Ross-Ibarra J., 2014 Advances and limits of using population genetics to understand local adaptation. Trends Ecol. Evol. **29**: 673–680.

Toomajian C., Hu T. T., Aranzana M. J., Lister C., Tang C., Zheng H., Zhao K., Calabrese P., Dean C., Nordborg M., 2006 A Nonparametric Test Reveals Selection for Rapid Flowering in the Arabidopsis Genome. PLoS Biol **4**: e137.

970 Turner T. L., Bourne E. C., Von Wettberg E. J., Hu T. T., Nuzhdin S. V., 2010 Population resequencing reveals local adaptation of Arabidopsis lyrata to serpentine soils. Nat Genet **42**: 260–263.

Verslues P. E., Lasky J. R., Juenger T. E., Liu T.-W., Kumar M. N., 2014 Genome-Wide Association Mapping Combined with Reverse Genetics Identifies New
975 Effectors of Low Water Potential-Induced Proline Accumulation in Arabidopsis. Plant Physiol. **164**: 144–159.

Voight B. F., Kudaravalli S., Wen X., Pritchard J. K., 2006 A Map of Recent Positive Selection in the Human Genome (L Hurst, Ed.). PLoS Biol. **4**: e72.

Weigel D., Nordborg M., 2015 Population Genomics for Understanding Adaptation in
980 Wild Plant Species. Annu. Rev. Genet. **49**: 315–338.

Wilczek A. M., Cooper M. D., Korves T. M., Schmitt J., 2014 Lagging adaptation to warming climate in Arabidopsis thaliana. Proc. Natl. Acad. Sci. **111**: 7906–7913.

Windle M., 2016 *Statistical Approaches to Gene X Environment Interactions for*
985 *Complex Phenotypes*. MIT Press.

Wituszyńska W., Ślesak I., Vanderauwera S., Szechyńska-Hebda M., Kornaś A., Van Der Kelen K., Mühlenbock P., Karpińska B., Maćkowski S., Van Breusegem F., Karpiński S., 2013 LESION SIMULATING DISEASE1, ENHANCED DISEASE SUSCEPTIBILITY1, and PHYTOALEXIN DEFICIENT4 Conditionally Regulate

990 Cellular Signaling Homeostasis, Photosynthesis, Water Use Efficiency, and Seed Yield in Arabidopsis. Plant Physiol. **161**: 1795–1805.

Yoder J. B., Stanton-Geddes J., Zhou P., Briskine R., Young N. D., Tiffin P., 2014 Genomic Signature of Adaptation to Climate in Medicago truncatula. Genetics.

Zomer R. J., Trabucco A., Bossio D. A., Verchot L. V., 2008 Climate change mitigation:
995 A spatial analysis of global land suitability for clean development mechanism afforestation and reforestation. Agric. Ecosyst. Environ. **126**: 67–80.

Supplemental material

1000 **Tables S1-S4 (attached csv files)**. List of genes within 1 kb of SNPs in the lower 0.001 quantile for p-values for SNP×environment interactions for each climate variable, including imputed observations and accounting for kinship.

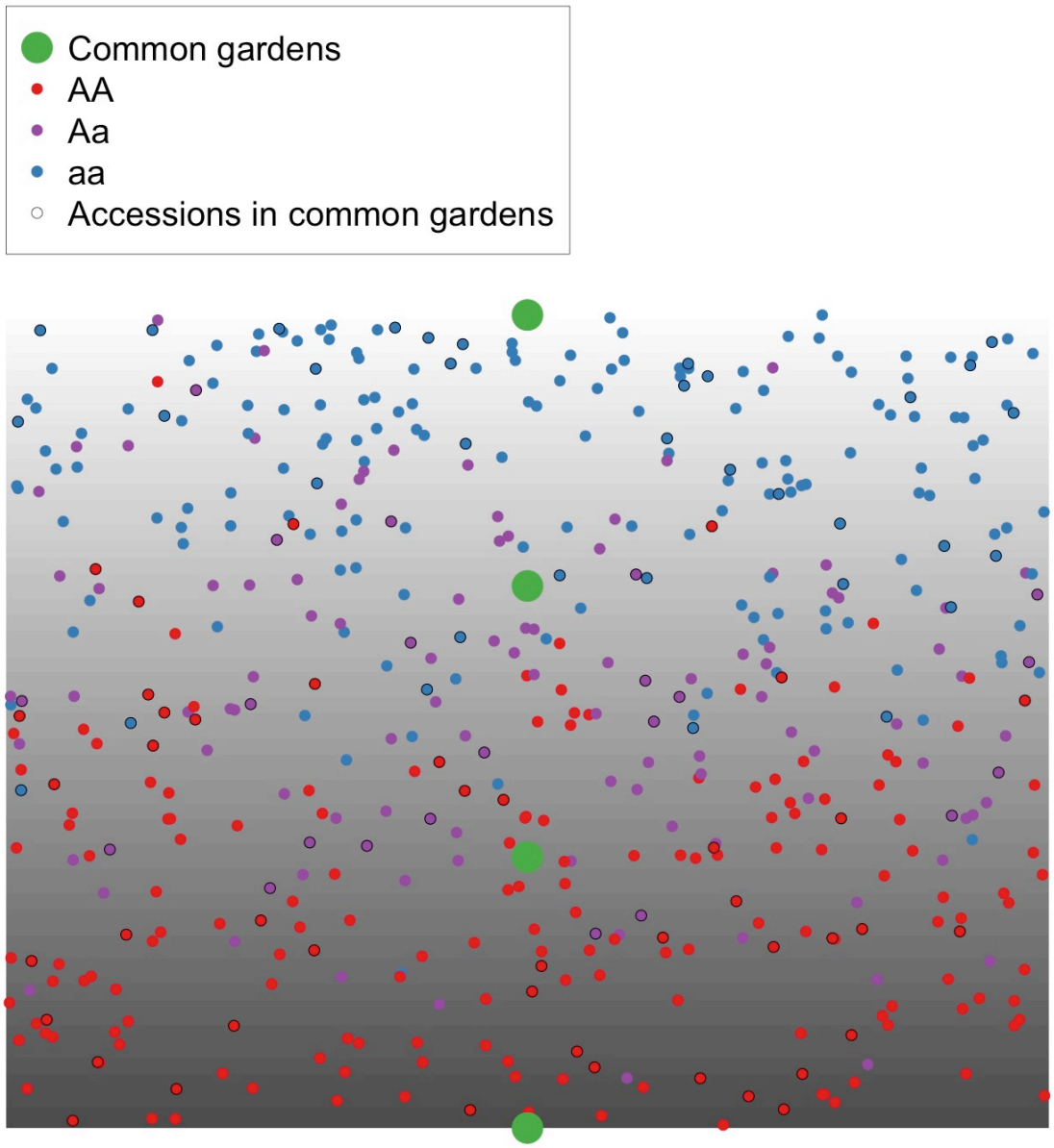
1005 **Table S5**. Rank correlation among SNPs for SNP×environment effect p-values comparing the three tested approaches with Arabidopsis.

	Random effect, no imputation	*Random effect, imputation*
Min. temp. growing season		
No random effect, including imputation	0.396	-0.001
Random effect, no imputation		0.041
Aridity		
No random effect, including imputation	0.318	0.008
Random effect, no imputation		-0.032
Min. temp. coldest month		
No random effect, including imputation	-0.021	-0.012
Random effect, no imputation		0.104
CV growing season prec.		
No random effect, including imputation	0.178	0.090
Random effect, no imputation		-0.006

Table S6. Overlap (one-tailed permutation tests) and rank correlations (Spearman's rho) between different approaches to identifying SNPs associated with local adaptation in Arabidopsis. Citations are given for the original publication of previous approaches. For partial Mantel and RDA, we used absolute value of SNP scores to rank SNPs.

Climate var.	Mixed model, overlap in 0.01 tail strongest associations	Mixed model, rank correlation between associations	Partial Mantel, overlap in 0.01 tail strongest associations	Partial Mantel, rank correlation between associations	First axis of RDA, overlap in 0.01 tail strongest associations	First axis of RDA, rank correlation between associations	First axis of RDA after removing spatial structure,, overlap in 0.01 tail strongest associations	First axis of RDA after removing spatial structure, rank correlation between associations
	Lasky et al. 2014	Lasky et al. 2014	Hancock et al. 2011	Hancock et al. 2011	Lasky et al. 2012	Lasky et al. 2012	Lasky et al. 2012	Lasky et al. 2012
Min. temp. growing season	<0.0001	0.079			0.0489	-0.093	<0.0001	-0.132
Aridity	<0.0001	0.028	<0.0001	0.065	<0.0001	-0.015	0.1107	0.013
Min. temp. coldest month	<0.0001	0.042	0.0128	0.044	<0.0001	-0.213	<0.0001	-0.057
CV growing season prec.	<0.0001	0.187			0.9591	-0.028	0.9953	0.030

1015 **Figure S1**. Example outcome of selection on a simulated landscape, with selective gradient (grayscale background), 500 randomly sampled accessions (small circles), and four common gardens (green circles). 100 accessions were simulated in the four common gardens (i.e. reciprocal transplant).



1020