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

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

[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]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 identify 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 climatic clines where alleles are associated with higher fitness in home environments. We apply this model to published data on natural *Arabidopsis thaliana* accessions. Our approach reveals candidate genes for local adaptation based on known involvement in environmental stress response. Most outlier SNPs exhibit home allele advantage and fitness tradeoffs along climate gradients, suggesting selective gradients may maintain allelic clines. SNPs exhibiting G×E associations with fitness are enriched in genic regions, putative partial selective sweeps, and G×E associations with an important 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 linked to 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 disease (Hunter 2005; Gage *et al.* 2016) and approaches for detecting the genomic basis of G×E are an emerging area of biomedical research (Thomas 2010; Keller 2014). The combination of these applications with advances in sequencing technology has generated increasing interest in the genomic basis of local adaptation and G×E (reviewed by (Des Marais *et al.* 2013; Manel and Holderegger 2013; Tiffin and Ross-Ibarra 2014; Adrion *et al.* 2015; Bragg *et al.* 2015)).

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). However, G×E at individual loci contributing to local adaptation might not involve changes (among sites) in which of the alleles has a greater local fitness, only changes in the magnitude of fitness difference between alleles (when alleles have equal fitness in some environments this is referred to as conditional neutrality (Verhoeven *et al.* 2004; Mitchell-Olds *et al.* 2007; HALL *et al.* 2010)). Detecting loci that exhibit antagonistic pleiotropy has been challenging, partly 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 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.* 2011; 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 (and prohibitive for some species), limiting biologists' ability to phenotype diverse panels at many locations across a species range. Additionally, it is unclear how

the typically small spatiotemporal scales of common gardens relate to the scales of processes that generate local adaptation in wild populations (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; Rellstab *et al.* 2015). 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

An emerging approach to understanding the genomic basis of adaptation involves synthesizing lines of evidence, *i.e.* multiple types of genome scans are combined to strengthen the evidence that a locus is under selection. For example, (Horton *et al.* 2012) found that loci that were outliers for selection statistics were also strongly associated with putatively adaptive phenotypes. (Evans *et al.* 2014) mapped overlaps between loci associated with environment and loci associated with putatively adaptive phenotypes, while (Lasky *et al.* 2014) characterized overlap between genes showing genetic variation in expression responses to abiotic stress and loci associated with environment. (Berg and Coop 2014) found that SNPs having strong associations with some human phenotypes also had significant associations with climate gradients. Additionally, (Lasky *et al.* 2015) used a Bayesian approach to combine associations with phenotype and environment. Based on the hypothesis that loci associated with climatic gradients were candidates for loci underlying G×E in response to abiotic stress, the authors first calculated climate associations and used each marker's association to determine the prior probability it was associated with G×E, yielding a posterior. Although combining multiple lines of evidence is potentially useful, the quantitative basis of synthesis in past studies has often been *ad hoc* and lacked strongly-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 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; Gauderman *et al.* 2013; Marigorta and Gibson 2014), but have not been applied to G×E along continuous environmental gradients. 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 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 a test case using published data on *Arabidopsis thaliana* (hereafter Arabidopsis), and discuss promising avenues for extension.

Methods

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

Local adaptation requires a genotype by environment interaction for fitness. To assess this interaction, we assume that the relative fitness, F_i, satisfies the linear model

$$F_i = \alpha + G_i\beta_G + G_iE_i\beta_{G\times E} + \varepsilon_i,$$

(eqn 1)

where E_i and G_i are the values of the environmental and genotype variables, respectively, for the ith individual. The $\beta_{G\times E}$ parameter 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 (a direct environmental effect is included in a separate step, see eqn. 3 below). In common garden experiments environment is often treated as a factor. But when multiple 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
160 the gradient (*i.e.* single dimension) according to its conditions. Describing common gardens as such may be particularly informative about the specific ecological mechanisms driving selective gradients, taking advantage of the ordered nature of multiple gardens' environments.

We leverage multiple common garden experiments to identify markers, *e.g.* single
165 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. 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 1 error rates. To ameliorate
170 this issue, we control for genomic background associations by including an interaction term between genomic background associations and the effect of environment on fitness (i.e. G×E), into the variance-covariance structure of the errors, ε_i. In particular, we assume that the vector of errors, $\boldsymbol{\varepsilon}$, can be expressed as

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

175 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). The random effect $\boldsymbol{v}$ represents the
180 effects on fitness conditioned on environment (G×E), while $\boldsymbol{e}$ represents the iid error in the model. Despite the existence of studies where fitness was measured in multiple

common gardens for diverse genotyped accessions (Fournier-Level *et al.* 2011), studies where linkage mapping was conducted for fitness at multiple sites (Ågren *et al.* 2013), and studies where authors conducted association mapping for 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.

We fit the discussed model, eqns 1 and 2, 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.

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

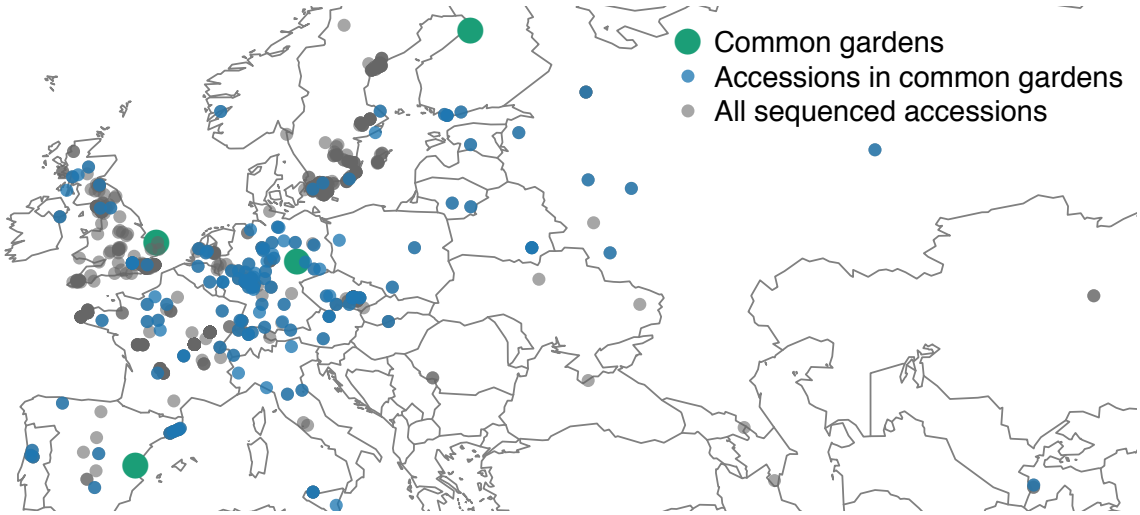
We now tackle the goal of synthesizing genome-environment associations and G×E observed in common gardens, given that both patterns are expected to inform on the same process of local adaptation. However, each approach does not use exactly the same types of data. A major limitation of genome-environment associations is that they are purely observational; fitness is not observed. 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. If we make this assumption of genome-environment associations explicit, one formal result would be an (imputed) observation of highest relative fitness for genotypes in their home environment for which we only collected sequence and environmental data. If we standardize fitness within each

common garden so that the maximum observed fitness is given a relative fitness of unity, we have a metric of relative fitness that can be directly measured or imputed in each type of study (common garden and genome-environment association). We then assume that each genotype collected from wild populations is locally adapted at its home and thus has a relative fitness = 1. We can then conduct a genome scan for markers associated with G×E as in eqn. 2, where each observation is either (A) a given genotype by common garden combination or (B) a given genotyped collected from its natural home and subsequently sequenced (i.e. data typically used in genome-environment associations).

Case study: local adaptation to climate in Arabidopsis thaliana

Here we apply these approaches to published data from multiple studies of *Arabidopsis thaliana* in its native Eurasian range. First, Fournier-Level *et al.* (2011) conducted replicated common gardens at four sites in Europe of differing climate, in Spain, England, Germany, and Finland (Figure 1). With these common garden data, (Fournier-Level *et al.* 2011; 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 planted 157 accessions (59 in the case of Finland) and calculated survival and fecundity (Fournier-Level *et al.* 2011). We converted these phenotypes to absolute fitness by multiplying survival rate with average fecundity of surviving plants.

Figure 1. 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 1307 accessions from around the globe
230 that were genotyped at ~250k SNPs using a custom Affymetrix SNP tiling array (AtSNPtile1), with 214,051 SNPs remaining after quality control (Horton *et al.* 2012). Of the 1307 genotyped accessions, we used 1001 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.* 2012), and excluding potentially
235 inaccurate high altitude outliers (i.e. > 2000 m). After imputing fitness for accessions in their home environments we had a total of 1531 observations of fitness × location (i.e. 1001 imputed observations + 530 real observations).

We used climate data compiled previously (Lasky *et al.* 2012) from published global climate datasets (Hijmans *et al.* 2005; Zomer *et al.* 2008). Here we focus on four
240 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 monthly minimum temperature in the growing season, coefficient of variation of monthly growing season precipitation, and aridity index.

245 *Genome-wide G×E associations*

We tested the models presented in eqns. 1 and 2, where the main parameter of interest was $\beta_{G\times E}$, the coefficient for SNP×environment effects on fitness. In the first approach (eqn 1), we did not incorporate a random effect of kinship. In the second approach (eqn 2) we incorporated kinship effects of genotype x environment on fitness. In order to

250 improve computation time for each individual association at all ~200 k SNPs, we used the approach 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)

255 covariates. In other words, we first fit the model:

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

(eqn 3)

with ε_i defined as in eqn 2, to obtain parameter estimates $\widehat{\beta_E}, \hat{\sigma}_e^2, \hat{\sigma}_{G\times E}^2$. We then take the variance parameter estimates and use them to estimate the slope coefficients in eqn 2

260 using generalized least squares.

To characterize the types of patterns identified by our approach, we studied variation in the SNPs arising from the 0.01 lower tail of p-values for each climate gradient. We asked whether these SNPs showed patterns consistent with home genotype advantage via changes in the allele with greatest relative fitness along the environmental

265 gradient. 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 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
270 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 different allele was estimated to be most fit in the warmest common garden.

275

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. To assess whether our association approach is capable of identifying the signal of local adaptation rather than spurious
280 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.* 2012). For a test statistic, we calculated the portion of SNPs in the 0.01 lower p-value tail that were found in each
285 of the SNP categories.

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 colonized following the last glacial maximum. We tested for an enrichment of pairwise haplotype sharing (PHS, (Toomajian *et al.* 2006)) in the SNPs (using PHS
290 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 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 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 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 significant associations ($p < 0.05$) with G×E for flowering time.

To generate a null expectation for each enrichment, 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.

Results

G×E variance components

We found that climate variables differed in the importance of kinship-climate interaction associations with fitness (*i.e.* values of $\hat{\sigma}^2_{G\times E}$), 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 intra-growing season precipitation

variability, kinship×environment interactions explained most of the fitness error (i.e. $\hat{\sigma}^2_{G\times E} > \hat{\sigma}^2_e$, Table 1). By contrast, kinship×environment interactions for fitness were weaker along temperature and aridity gradients (i.e. $\hat{\sigma}^2_e > \hat{\sigma}^2_{G\times E}$).

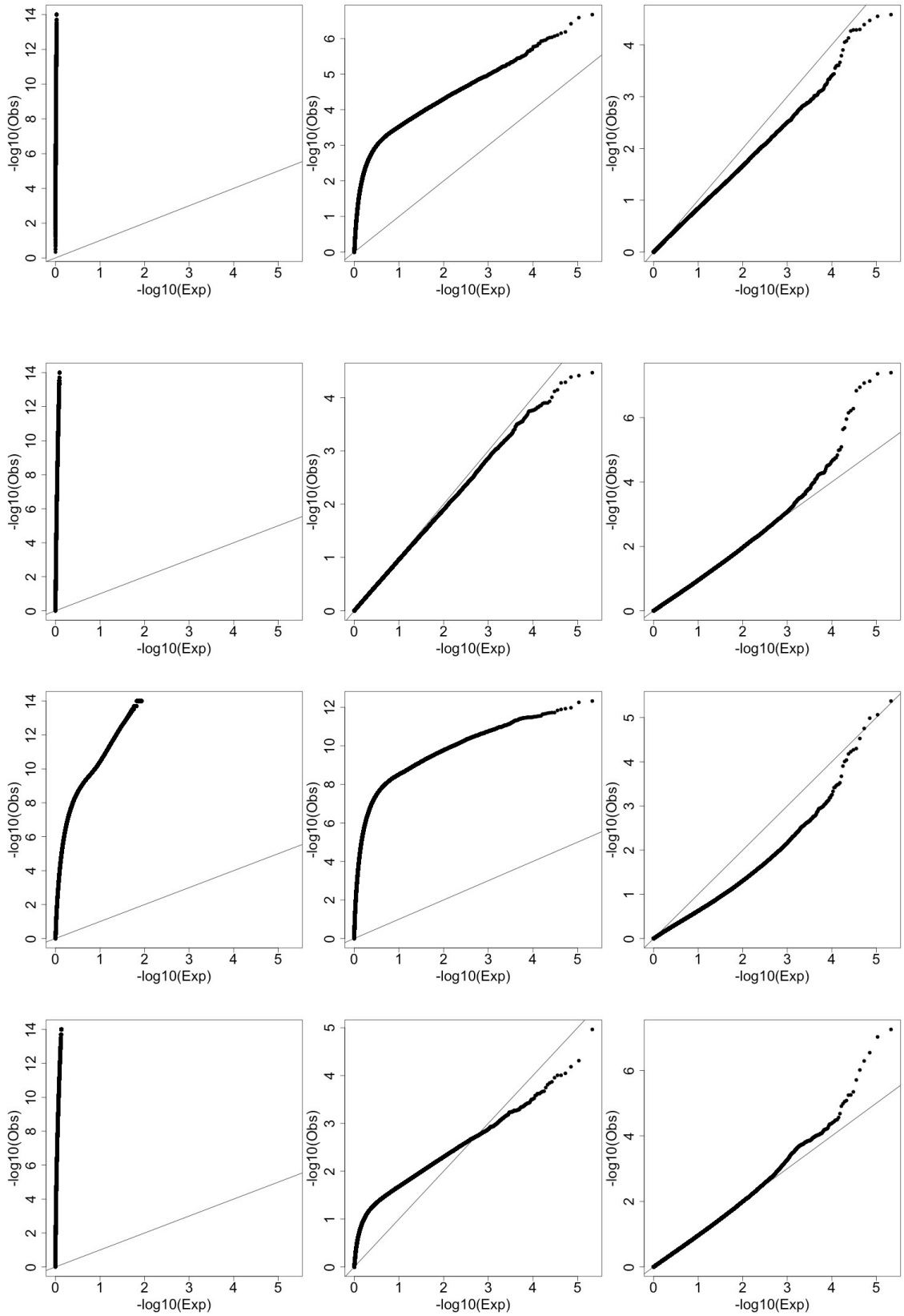
Table 1. Variance components of relative fitness variation including both observed and imputed observations, as in eqn. 3.

Effect	CV growing season precip. (intra-seasonal)	Minimum temp. growing season	Min. temp. coldest month.	Aridity
$\hat{\sigma}^2_{G\times E}$	0.137410	0.028428	0.045418	0.028738
$\hat{\sigma}^2_e$	0.080658	0.107870	0.094755	0.095757

P-value distributions

We found that simple linear model tests of SNP×environment interactions were highly enriched in very low p-values (Figure 2) relative to the theoretical expectation. After incorporating the kinship×environment random effects (but excluding imputed fitness observations), 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 (right column, Figure 2), 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.

Figure 2. Quantile-quantile plots of p-value distributions for three approaches to calculating genome-wide SNP×environment associations with fitness. In the left panels, random effects with covariance structure determined by **K** *were not* included, but imputed fitness observations were included. In the center panels, random effects with covariance structure determined by **K** *were* included, but imputed fitness observations were excluded (i.e. only direct fitness observations included). In the right panels, random effects and imputed observations were included. The rows represent associations with (from top to bottom) aridity, winter minimum temperature, CV of growing season precipitation, and minimum temperatures in the growing season.



340

Identification of SNPs with home allele advantage

Our mixed model (eqn 2, including imputed data) tended to identify SNPs where SNP×environment interactions favored alleles in climates where they were relatively more common (ignoring population structure), that is the sign of allelic differences in home climates were mirrored by the sign of fitted mixed model SNP×environment associations with relative fitness (aridity: 52% of SNPs in the lower 0.01 tail of p-values have consistent signs indicating simple SNP×environment interactions favored local alleles, CV of growing season precipitation: 51%, growing season minimum temperature: 93%, minimum temp coldest month: 74%, see outlier examples in Figure 3). In addition to characterizing SNP×environment associations, our mixed model (including imputed data) 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 (SNPs in the lower 0.01 tail of p-values for the mixed model had estimated rank changes for aridity: 31%, CV of growing season precipitation: 93%, growing season minimum temperature: 73%, minimum temp coldest month: >99%). 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. Specifically, the gardens covered only 13% of the total observed range of aridity (i.e. range of gardens / range of home climates of sequenced accessions) compared to 31% for growing season minimum temperature, 65% for CV of growing season precipitation, and 78% for minimum temperature of coldest month. 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.

SNP×environment associations with fitness are enriched in genic regions and for evidence of local selective sweeps

To assess whether our models identified certain types of SNPs, we tested for enrichment of genic and intergenic SNPs for SNP×environment effects on fitness. We found that simple linear model tests of SNP×environment interactions were not significantly enriched in any SNP type, for any of the climate variables tested (permutation two-tailed test, all $p > 0.05$). By contrast, after incorporating the kinship×environment random effects (focusing on inference including imputed fitness observations), we found that SNP×environment associations with fitness were significantly enriched in genic regions for aridity ($p = 0.0012$), CV of growing season precipitation ($p < 0.0002$), growing season minimum temperature ($p=0.0004$) and minimum temperature of coldest month ($p = 0.0036$).

We found the strongest SNP×environment interactions for fitness (including imputed observations) were significantly enriched for high PHS for all four climate variables studied (aridity $p=0.0064$, CV of growing season precipitation $p = 0.0172$, growing season minimum temperature $p=0.0302$, minimum temperature of coldest month $p = 0.0090$). Similarly, we found that the strongest SNP×environment interactions for fitness were enriched for high iHS (aridity $p = 0.0738$, CV of growing season precipitation $p = 0.0008$, growing season minimum temperature $p=0.0048$, minimum temperature of coldest month $p = 0.0036$).

Finally, we found that the SNPs with the strongest SNP×environment associations with relative fitness exhibited evidence of enrichment with G×E for

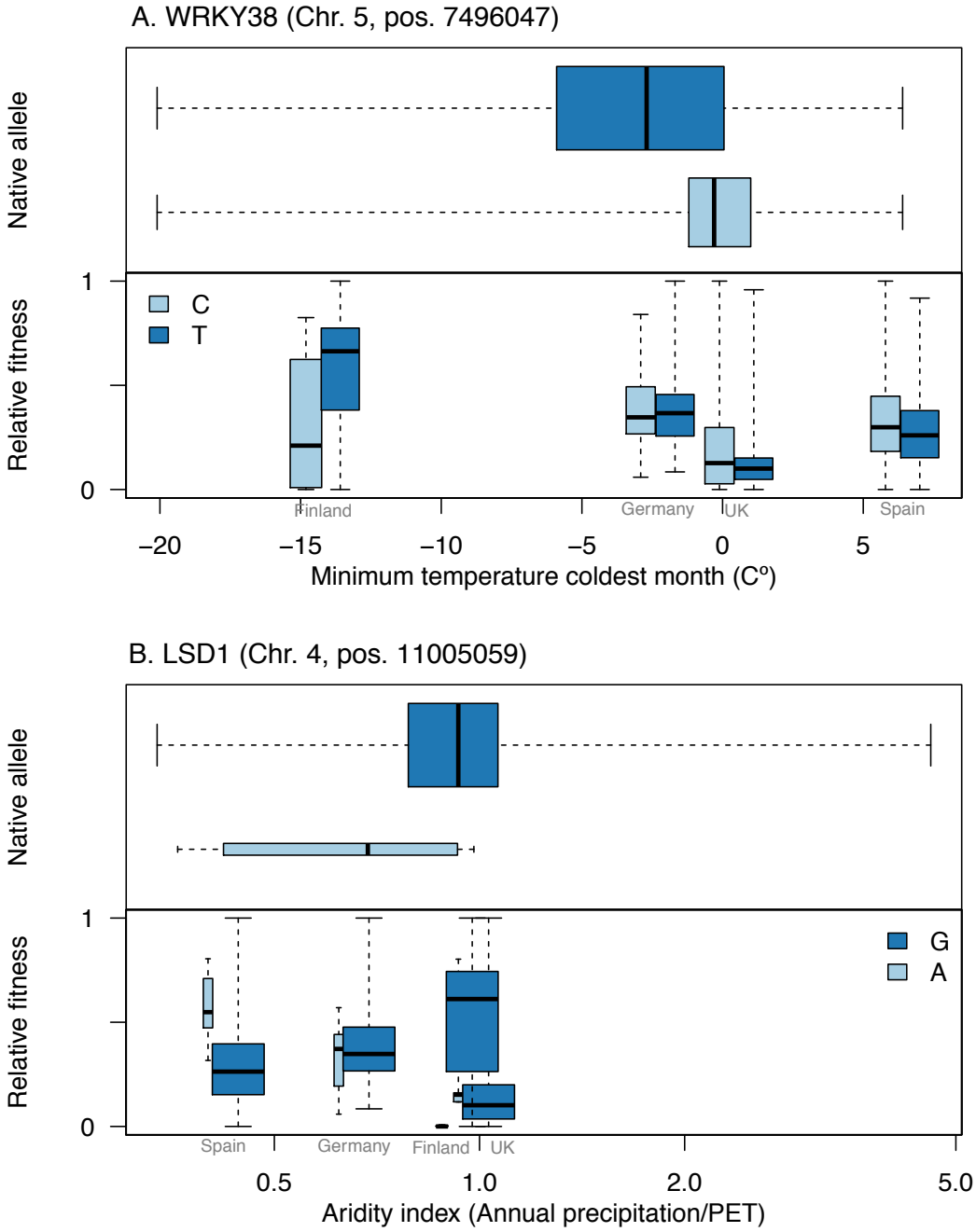
flowering time. We found that SNPs in the 0.01 lower tail of SNP interactions with growing season minimum temperature for fitness were most strongly enriched for SNPs associated with G×E for flowering time response to growing temperature ($p < 0.0002$). Additionally, SNP interactions with precipitation variability during the growing season for fitness were significantly enriched in flowering time G×E ($p = 0.0272$) although SNP interactions with aridity ($p = 0.368$) and winter cold ($p = 0.3398$) were not significantly enriched.

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 (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 3), while the second SNP (chr. 2, position 7592008) fell within ATMLO8, MILDEW RESISTANCE 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 climate in Lasky et al. (2012) (Figure 3). The top SNP associated with

variability in growing season precipitation interaction effects on fitness (chr. 2, position
415 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 sixth highest SNP (and second highest locus) associated with growing season cold interaction effects on fitness (chr. 3, position 8454439) fell within ABERRANT LATERAL ROOT FORMATION 5, ALF5, a
420 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).

Figure 3. 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 native (i.e. home) genotypes, 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 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.* 2011; Lasky *et al.* 2012, 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.* 2012). Our approach 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.

Model extensions

Above we described a method of imputation based on the assumption of local adaptation, *i.e*. home genotypes had the greatest fitness compared to other 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 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 that each sequenced genotype is most fit in its home environment, an alternative approach could treat the 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 expected to be common. Inferences about unobserved fitness could be further constrained using hierarchical models, such that 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 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. Here we do not attempt to parameterize these model extensions, given the current computational challenge of fitting many more parameters in a Bayesian framework.

485 *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), but to our knowledge there are no published genome-wide association studies accounting for SNP interactions with continuous environmental gradients. Some of the
490 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 cigarette smoke. Future research on local adaptation and human disease may benefit from exchange of approaches given the shared importance across
495 disciplines of understanding the genomic basis of G×E.

Case study on Arabidopsis thaliana

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
500 candidate genes based on known roles in environmental responses, suggesting our approach is a useful for identifying loci underlying local adaptation. Several top genes exhibiting SNP×climate interactions have known functions in pathogen defense. Although climate might not directly select for spatial variation defense phenotypes,

climate often drives pathogen community turnover and defense phenotypes in
505 Arabidopsis are known to covary with climate (Brachi *et al.* 2015).

Our model 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 with winter minimum temperature association for fitness), loci where selective gradients may maintain population differentiation (Anderson *et al.*
510 2011b; Ågren *et al.* 2013). A previous study of the common garden data used here (Fournier-Level *et al.* 2011) 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, our model was explicitly focused on detecting alleles with the strongest evidence for SNP×climate
515 interactions favoring home alleles, which means that 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
520 of (Fournier-Level *et al.* 2011) is based on association mapping, which may suffer from identification of more false positives compared with linkage mapping approaches (HALL *et al.* 2010; ANDERSON *et al.* 2013; Ågren *et al.* 2013). Experimental study of phenotypic effects of variation at individual loci is required to confirm results of association mapping (e.g. Verslues *et al.* 2014; Broekgaarden *et al.* 2015).

525 We found evidence that SNP×climate interaction effects on fitness were enriched 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
530 modeled were not simply associated with overall patterns of among-population divergence. This enrichment is consistent with other findings in Arabidopsis (Hancock *et al.* 2011; Lasky *et al.* 2012) 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
535 evolution in local adaptation, but rather as supporting the hypothesis that local adaptation 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
540 recent local sweeps in particular environments are an important mode of local adaptation (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 13 kya (Sharbel *et al.* 2000), a process that may have involved recent local adaptation. Furthermore,
545 (Hancock *et al.* 2011) identified an enrichment of PHS in their climate-associated loci while (Fournier-Level *et al.* 2011) found limited evidence for sweeps involving loci associated with local survival and fecundity. 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

550 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 gradients, especially growing season cold, and SNPs associated with G×E for flowering time across growing season temperature
555 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
560 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 in local adaptation (Fraser 2013; Lasky *et al.* 2014).

565

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
570 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.

575 **Acknowledgements**

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

Literature Cited

580

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.
585 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.

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
590 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.

610 Brachi B., Meyer C. G., Villoutreix R., Platt A., Morton T. C., Roux F., Bergelson J., 2015 Coselected genes determine adaptive variation in herbivore resistance throughout the native range of Arabidopsis thaliana. Proc. Natl. Acad. Sci. **112**: 4032–4037.

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

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
620 Cabbage Whitefly and Arabidopsis Uncovered by Genome-Wide Association 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.
625 van, Lee T. van der, Diergaarde P., Groenendijk J., Töpsch S., Vos P., Salamini 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

630 generates a latitudinal cline in a life history trait. Proc. Natl. Acad. Sci. U. S. A. **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.

635 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
640 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
645 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.

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

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.

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 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.

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.

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., 2012 Characterizing genomic variation of Arabidopsis thaliana: the roles of geography and climate. Mol. Ecol. **21**: 5512–5529.

Lasky J. R., Des Marais D. L., Lowry D. B., Povolotskaya I., McKay J. K., Richards J. H., 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., 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.

750 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
755 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
760 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
765 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.

805 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.

Stinchcombe J. R., Weinig C., Ungerer M., Olsen K. M., Mays C., Halldorsdottir S. S.,
810 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 EDS1 regulate drought tolerance, plant biomass production, and cell wall
815 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.

Tiffin P., Ross-Ibarra J., 2014 Advances and limits of using population genetics to
820 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.

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.

Verhoeven K. J. F., Vanhala T. K., Biere A., Nevo E., Damme J. M. M. van, 2004 THE GENETIC BASIS OF ADAPTIVE POPULATION DIFFERENTIATION: A QUANTITATIVE TRAIT LOCUS ANALYSIS OF FITNESS TRAITS IN TWO WILD BARLEY POPULATIONS FROM CONTRASTING HABITATS. Evolution **58**: 270–283.

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 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 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.

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 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: A spatial analysis of global land suitability for clean development mechanism afforestation and reforestation. Agric. Ecosyst. Environ. **126**: 67–80.

Tables S1-S4. 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.