

Genotype–environment associations to reveal the molecular basis of environmental adaptation

Jesse R. Lasky ,^{1,*} Emily B. Josephs^{2,*} and Geoffrey P. Morris ^{3,*}

1 Department of Biology, Pennsylvania State University, University Park, Pennsylvania 16802, USA

2 Department of Plant Biology; Ecology, Evolution, and Behavior Program, Michigan State University, East Lansing, Michigan 48824, USA

3 Department of Soil and Crop Sciences; Cell and Molecular Biology Program, Colorado State University, Fort Collins, Colorado 80526, USA

*Author for correspondence: lasky@psu.edu (J.R.L.), josep993@msu.edu (E.B.J.), geoff.morris@colostate.edu (G.P.M.)

All authors led the writing of the manuscript and prepared the figures.

The author responsible for distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors (<https://academic.oup.com/plcell>) is: Jesse R. Lasky (lasky@psu.edu).

Review

Abstract

A fundamental goal in plant biology is to identify and understand the variation underlying plants' adaptation to their environment. Climate change has given new urgency to this goal, as society aims to accelerate adaptation of ecologically important plant species, endangered plant species, and crops to hotter, less predictable climates. In the pre-genomic era, identifying adaptive alleles was painstaking work, leveraging genetics, molecular biology, physiology, and ecology. Now, the rise of genomics and new computational approaches may facilitate this research. Genotype–environment associations (GEAs) use statistical associations between allele frequency and environment of origin to test the hypothesis that allelic variation at a given gene is adapted to local environments. Researchers may scan the genome for GEAs to generate hypotheses on adaptive genetic variants (environmental genome-wide association studies). Despite the rapid adoption of these methods, many important questions remain about the interpretation of GEA findings, which arise from fundamental unanswered questions on the genetic architecture of adaptation and limitations inherent to association-based analyses. We outline strategies to ground GEAs in the underlying hypotheses of genetic architecture and better test GEA-generated hypotheses using genetics and ecophysiology. We provide recommendations for new users who seek to learn about the molecular basis of adaptation. When combined with a rigorous hypothesis testing framework, GEAs may facilitate our understanding of the molecular basis of climate adaptation for plant improvement.

Local adaptation in the genomic era

A fundamental goal in plant biology is to characterize and understand the variation underlying plants' adaptation to their environments. Climate change has given new urgency to this goal, as society aims to adapt ecologically important plant species, endangered plant species, and crop species to new and less predictable climates. One key strategy for identifying genotypes and traits adapted to specific climates has been to study variation among different populations of the same species that inhabit different climates. This research

strategy is based on the hypothesis that these populations each carry adaptations to their local climates. Local adaptation is defined as a genotype-by-environment interaction for fitness that favors home genotypes over foreign genotypes (Kawecki and Ebert, 2004). While demonstrating the existence of local adaptation is conceptually straightforward using multiple common garden experiments (Clausen et al., 1940), dissecting its molecular basis is more challenging. Identifying existing loci that are locally adapted to climate may reveal alleles and traits that could be used to discover

ecophysiological mechanisms of adaptation or that could be deployed in breeding climate-adapted plants (Olatoye et al., 2018; Faye et al., 2019).

In the pre-genomic era, identifying adaptive alleles was painstaking work, and required leveraging genetics, molecular biology, physiology, and ecology. Now, with the rise of pangenome analyses, which we define as those that aim to characterize almost all variation present in a given species, almost all alleles (climate-adaptive and otherwise) in many important species have been, or will soon be, sequenced, and cataloged (Figure 1; Evans et al., 2014; Alonso-Blanco et al., 2016; Zhao et al., 2018; Todesco et al., 2020; Hufford et al., 2021; Tao et al., 2021). Even panels of a few hundred randomly sampled genotypes will capture nearly all variants that are perceptible in typical genome-wide association studies (GWAS; Figure 2). It is hoped, but not certain, that this abundance of genomic data and new quantitative genomic approaches will facilitate the search for adaptive alleles. Here we consider research strategies for leveraging genotype–environment associations (GEAs), which use statistical associations between allele frequency and environmental parameters at the location of origin to identify alleles adapted to local environments, i.e. locally-adapted genetic variants. When GEAs are implemented across the genome, we will refer to them as environmental GWAS hereafter.

GEAs typically use only data on the environment of origin of an organism, combined with sequence data, and for this reason can be easily implemented. But despite the ease of implementation, few studies have actually experimentally validated the inferences of GEAs. Fewer still have shown successful applications of these inferences for plant improvement for specific climates. Here, we review the biological and statistical rationale of GEAs with a specific aim of translating their correlation inferences into molecular biological knowledge and improved plant adaptation. We present strategies for using GEAs for these purposes.

The promise of GEAs

In recent years, researchers have increasingly sought to uncover the molecular basis of local adaptation. Some of the major questions confronting researchers in this field are:

- (1) How do we find locally adapted alleles?
- (2) How do we uncover the environmental and evolutionary mechanisms explaining functional natural variation?
- (3) What is the genetic architecture of local adaptation, and what factors determine this architecture?
- (4) What molecular mechanisms underpin local adaptation?

	Species	Number of accessions resequenced	Georeference data available for GEA?	References
	Arabidopsis (<i>Arabidopsis thaliana</i>)	1,135	1,057	Alonso-Blanco et al. 2016
	Rice (<i>Oryza sativa</i>)	3,188	1,443	Wang et al. 2018, Gutaker et al. 2020
	Switchgrass (<i>Panicum virgatum</i>)	732	700	Lovell et al. 2021
	Sorghum (<i>Sorghum bicolor</i>)	384	143	Bellis et al. 2020, LeBauer et al. 2020
	Black cottonwood (<i>Populus trichocarpa</i>)	544	544	Evans et al. 2014
	3 sunflower species (<i>Helianthus annuus</i> , <i>H. argophyllus</i> , <i>H. petiolaris</i>)	1,293; 309; 432	984; 309; 432	Todesco et al. 2020

Figure 1 A summary of representative pan-genome studies, which have comprehensively characterized genome-wide polymorphisms for global germplasm panels (Evans et al., 2014; Alonso-Blanco et al., 2016; Wang et al., 2018; Bellis et al., 2020; Gutaker et al., 2020; LeBauer et al., 2020; Todesco et al., 2020; Lovell et al., 2021). Photo credits Arabidopsis, rice, switchgrass, sorghum (author G.P.M.), cottonwood, sunflower.

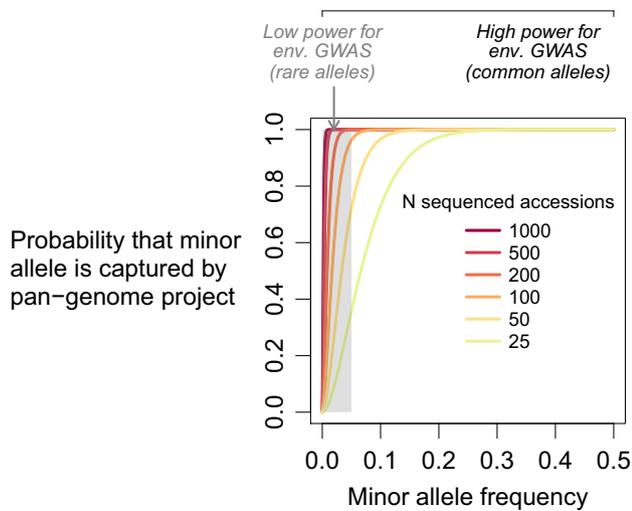


Figure 2 Most climate adaptive alleles have been sequenced in some important plant species. Assuming random sampling of genotypes, this panel shows the probability of including an allele of a given minor allele frequency (MAF) in a collection of N genotypes. GWAS approaches (including environmental GWAS) typically have low power for rare alleles, and so most alleles visible to GWAS are captured by panels of even a few hundred genotypes (assuming random sampling).

GEAs arrived in this context, enabled by the sequencing advances of the last decades (Hancock et al., 2008; Coop et al., 2010; Eckert et al., 2010; Manel et al., 2010; Sork et al., 2010; Lasky et al., 2012). GEAs are simply correlations between the environment of origin, e.g. climate and allele frequency. Other methods have been developed to detect potentially adaptive alleles by looking for excess differentiation among populations (fixation index [F_{ST}] and variants; Lewontin and Krakauer, 1973; Luu et al., 2017) or signatures of selective sweeps (Maynard Smith and Haigh, 1974; Pavlidis et al., 2013). However, GEAs are particularly useful because they cannot only identify adaptive alleles but also determine which environmental gradients drive changes in selection (Lasky et al., 2012; Capblancq and Forester, 2021). Given that the molecular mechanisms of adaptation that underlie complex traits and complex ecological variation are often unclear, implementing GEAs across the genome (i.e. environmental GWAS) may allow gene discovery without strong a priori knowledge of molecular physiological responses to specific environments. Once identified, genes with strong changes in allele frequency across environments can be further interrogated using expression responses to relevant environmental treatments (Lasky et al., 2014; Gates et al., 2019), population genetic patterns indicative of selection (Evans et al., 2014; Yoder et al., 2014; Price et al., 2018), and changes in whole plant performance across environments (Gienapp et al., 2017; Lasky et al., 2018; Gates et al., 2019). These potentially locally adapted alleles could then be used in marker-assisted breeding or genetic engineering of plant populations of interest.

This promise has led to a rapid adoption of these studies. However, the often implicit evolutionary and ecological genetic hypotheses that underpin the use of GEA remain under-appreciated. As a result, GEA may not currently be deployed optimally by researchers. Here our goal is to review these underlying hypotheses, and suggest high-level strategies for using GEA in research.

Evolutionary biology hypotheses that underpin the use of GEA

Below we discuss the evolutionary hypotheses that, when true, suggest the best-case scenario for detecting GEA for a gene that contributes to local adaptation (Figure 3). Local adaptation may occur in ways contrary to some of these hypotheses; in these situations, GEA will have less power to detect genomic regions underlying adaptation. Carefully considering these hypotheses, and whether they apply to the species of interest, can help researchers decide when and how to apply GEA.

Populations are locally adapted via consistent phenotypic clines along environmental gradients

To study the basis of environmental adaptation, one must first include individuals inhabiting a range of environments (Figure 3). Next, for locally-adapted genes to be discovered via environmental GWAS, and for a variant of interest to show a truly locally-adapted GEA, an (often implicit) hypothesis is that populations are locally adapted to their environment (Figure 3; Lasky et al., 2018). Adaptation requires genetic variation for fitness and conditions that allow effective selection. For example, population genetic theory indicates that the strength of spatially changing selection must be strong enough to counteract the homogenizing effects of gene flow among environments (Slatkin, 1973). Moreover, species with small effective population sizes or that have experienced bottlenecks might be less likely to evolve local adaptation (Siol et al., 2010; Angert et al., 2020). Meta-analyses have revealed that local adaptation is common, but importantly, local adaptation does not always occur, even when there are environmental gradients that select for different phenotypes (Leimu and Fischer, 2008; Hereford, 2009).

An additional hypothesis that is required for GEA is that populations must be locally adapted along a gradient captured by the tested environmental variables. For GEA to fully live up to its promise, environmental variables representing distinct axes of selection should be orthogonal to each other so that it is possible to determine which aspect of the environment is driving changes in selection (Figure 3). When multiple environmental variables are correlated, the best GEA can do is to identify suites of collinear variables associated with the genotype (Manel et al., 2010; Sork et al., 2010; Lasky et al., 2012; Eckert et al., 2015). Additionally, populations may not be locally adapted along every environmental gradient that researchers hypothesize, and so GEA along some

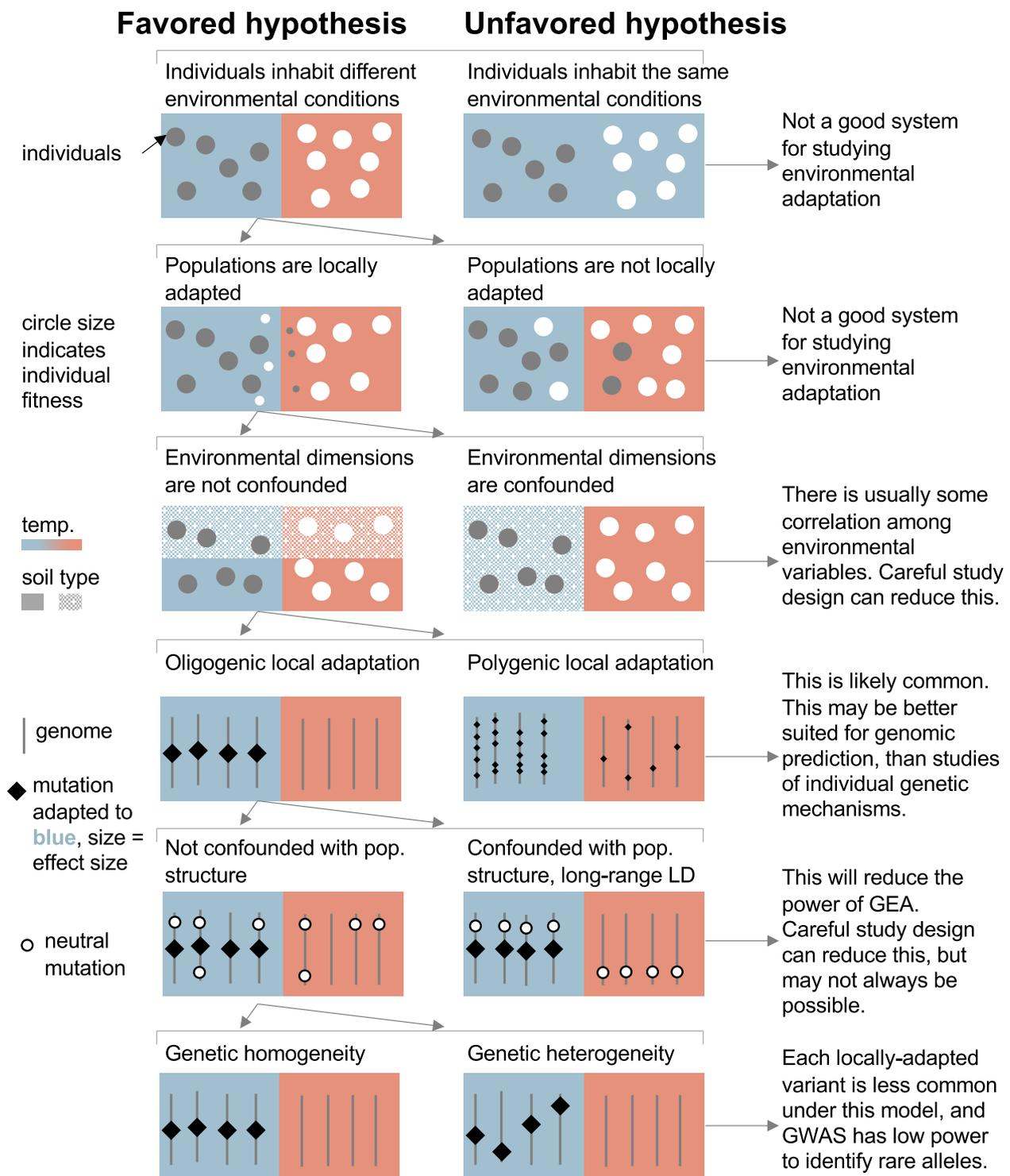


Figure 3 Evolutionary ecological and genetic hypotheses underpinning the use of GEA and environmental GWAS. The ideal scenario for using GEA, and especially environmental GWAS, is when a range of favored evolutionary hypotheses are true. When the favored hypotheses are not true, GEA and especially environmental GWAS will likely perform less well. The bottom three rows are major considerations for phenotypic GWAS as well, where the same hypotheses are favored (oligogenic; not confounded with population structure; genetic homogeneity). Hypotheses under which GEA approaches are more effective are noted as “favored,” while hypotheses under which GEA approaches are less effective are noted as “unfavored.” Acknowledging multiple hypotheses can mitigate bias toward favored hypotheses (Platt, 1964).

environmental gradients may be specious. This latter problem can be considered a pitfall of “pan-environmental adaptationism,” i.e. assuming that because a plausible

ecophysiological tradeoff can be hypothesized, populations must be adapted along a related environmental gradient.

Another important hypothesis is that different lineages will adapt to environmental gradients using the same traits. Researchers employing GEA often use diverse panels from multiple regions, where the same environmental gradients may be repeated (Hancock et al., 2011; Lasky et al., 2015; van Boheemen and Hodgins, 2020). GEA will work optimally when local adaptation in different regions occurs via consistent phenotypic changes along tested environmental gradients. This hypothesis is prior to even hypotheses about the consistency of genetic mechanisms (see genetic architecture section “Detection of allele frequency clines along environmental gradients depends on genetic architecture” below). Freezing tolerance may be an example where populations in different regions evolve similar physiological strategies to deal with the same conditions (Monroe et al., 2016). By contrast, plants may adapt along apparently similar gradients in different regions via distinct ecological strategies/phenotypic clines, because true selective gradients are not consistent between regions, or because of historical contingencies. For example in *Arabidopsis* (*Arabidopsis thaliana*), despite consistent elevation gradients in climate across regions, higher elevation populations in the Mediterranean are late flowering, while they are early flowering in Asia, which would confound global GEAs (Gamba et al., 2022). Different lineages may evolve different solutions to the same problem because locally-adaptive tradeoffs can be counterintuitive; for instance, in the case where stress-tolerating traits might be hypothesized to evolve in “stressful” environments, but the actual locally adapted populations have stress escape traits such as early flowering (Ludlow, 1989; Franks, 2011; Vigouroux et al., 2011).

Detection of allele frequency clines along environmental gradients depends on genetic architecture

The home genotype advantage of local adaptation occurs at the organismal level, but the specific genetic variants that cause this home advantage can have a wide range of number, effect sizes, interactions with the environment, dominance, and epistasis patterns, i.e. genetic architectures. Genetic architecture is an important consideration for the use and deployment of locally adapted genetic variants, as GEA has more power under some genetic architectures than others.

One major axis of this architecture is the number of variants and their effect size distribution, often characterized in the extreme as polygenic (many small-effect variants) or oligogenic (few large-effect variants, Figure 3). Large-effect variants are more likely to show differences in allele frequency between locally-adapted populations than small-effect loci (Le Corre and Kremer, 2012), making GEAs harder to detect or nonexistent for small-effect loci. Adaptive loci also might show dominance and epistasis (Monnahan and Kelly, 2015) making them harder to find in association studies because most studies start with

single locus additive models, and a much larger parameter space must be tested to identify dominance (in outcrossing systems) and epistasis.

While local adaptation is due to individual-level tradeoffs in fitness between environments, the individual loci contributing to adaptation do not necessarily themselves show tradeoffs in fitness. That is, conditionally neutral loci show fitness differences in some environments but not others, while antagonistically pleiotropic loci show fitness tradeoffs between alternate alleles (Anderson et al., 2011). Only the latter example may be stably maintained as spatial clines in the face of gene flow, but many systems may be far from equilibrium in spatial allele frequency patterns, so conditionally neutral alleles could still be important contributors to local adaptation (Wadgymer et al., 2017). GEAs are more powerful when the antagonistic pleiotropy hypothesis is true, and have reduced power under conditional neutrality (Yoder and Tiffin, 2018).

GEAs also depend on a hypothesis of genetic homogeneity, i.e. that different populations inhabiting the same type of environment adapt via the same genetic mechanisms (Figure 3). When multiple loci influence a trait under spatially-varying selection, different genes may be more likely to underlie local adaptation in different isolated populations (i.e. genetic heterogeneity). Even when the same genes are involved in local adaptation in different populations, different mutations, i.e. different alleles, may underlie local adaptation (i.e. allelic heterogeneity). Both phenomena result in individual causal alleles being rare, reducing the power of environmental GWAS. Theoretical results show that limited gene flow between populations inhabiting the same environments will increase the likelihood of genetic or allelic heterogeneity by reducing the chance that a new locally-adaptive variant can spread and be used in adaptation by different populations in the same environment (Lee and Coop, 2019). By contrast, shared ancestral variation or higher gene flow between these populations increases the chance that the same genetic mechanism underlies local adaptation (Lee and Coop, 2019). In general, gene flow and demography are known in theory to play important roles in determining genetic architecture, but their influence on local adaptation is often poorly known in most empirical systems; the extent of heterogeneity may also differ between different traits in the same system (Lopez-Arboleda et al., 2021).

Recent progress with GEAs

Here we provide a brief overview of the last decade of research that has employed GEAs. We divide these examples into three major strategic categories: reverse ecological genetics, forward genetics (i.e. environmental GWAS), and genomic prediction of genotype–environment interactions ($G \times E$). An ever-growing number of studies are being published in these categories, although forward genetics environmental GWAS have been least often followed with experimental validation. We include some prominent examples outside of plant systems, even though ecologically-

relevant experiments are often logistically prohibitive in many nonplant systems.

Forward ecological genetics using environmental GWAS

There are many examples of researchers employing environmental GWAS (genome scans for GEAs) over the last decade. While there are too many to thoroughly review here, these studies often find environmental associations in genes with functions plausibly involved in local adaptation (for example (Lasky et al., 2012; Evans et al., 2014; Yoder et al., 2014; Gibson and Moyle, 2020)). However, with abundant annotations of many genes, and many traits being involved in local adaptation, functional annotations of environmental GWAS hits are not strong evidence for a role in adaptation (Pavlidis et al., 2012).

An emerging approach is to test whether environmental GWAS identifies the same variants as GWAS on putatively adaptive traits, supporting the hypothesis that the variants responsible for trait variation underlie local adaptation along the tested environmental gradient (Evans et al., 2014; Yeaman et al., 2016; Lovell et al., 2021). In *Arabidopsis*, Lasky et al. (2018) constructed a model that combined changes in performance along climate gradients with GEA patterns to identify genes locally adapted to climate, and found that putative temperature-adapted single nucleotide polymorphisms (SNPs) were also associated with flowering time variation, suggesting local adaptation via evolution of flowering time. Relatedly, Gates et al. (Gates et al., 2019) showed in maize (*Zea mays* L.) that top environmental GWAS SNPs were also associated with $G \times E$ for fitness across common gardens in Mexico, and with genetic variation in flowering time. Also in *Arabidopsis*, Exposito-Alonso and colleagues observed that SNPs associated with greater fitness in a dry common garden tended to originate from drier climates (Exposito-Alonso et al., 2019).

A separate use of GEA has been to test hypotheses about the genetic architecture of local adaptation. For example, several studies have used the number and effect size of environmental GWAS hits to test whether local adaptation is oligogenic or polygenic (Bay and Palumbi, 2014; De La Torre et al., 2019). However, a challenge with using any association studies to explore genetic architecture is the unknown role of long-range linkage disequilibrium arising from population structure, which can obscure large-effect variants and create a false signal of polygenic adaptation (Brachi et al., 2011). Additionally, there are ascertainment biases and statistical artifacts like the Beavis effect and winner's curse, where quantitative trait loci (QTL) identified as being significant have an upward bias in effect size estimation, particularly in smaller samples (Beavis, 1994; Palmer and Pe'er, 2017). These biases could cause the genetic architecture that is detectable by an environmental GWAS to not match the true genetic architecture of adaptation.

Overall, the environmental GWAS approach is being adopted in a wide range of systems to study local

adaptation. However, it is notable that these studies rarely identify cloned candidate genes with known natural variants and functions related to the tested environmental variable (see example genes in "Reverse ecological genetics using GEA" section). Furthermore, there has been no systematic attempt to validate environmental GWAS hits using near-isogenic lines (NILs) or genetically engineered variants. As a more limited (but promising) example, Ferrero-Serrano and co-authors first detected temperature-associated exonic SNPs in *Arabidopsis* and then identified those predicted to affect mRNA secondary structure, and showed in vitro that two top SNPs did cause changes in the mRNA structure in response to temperature (Ferrero-Serrano et al., 2022).

Reverse ecological genetics using GEA

In general, fewer reverse ecological genetics GEAs have been performed than there have been forward genetics (environmental GWAS) studies. However, there is a long history of reverse ecological genetics in the pre-genomic era. For example, in one of the earliest studies, Oakeshott et al. (1982) studied allozymes of alcohol dehydrogenase (ADH) in *Drosophila* (*Drosophila melanogaster*) and detected replicated latitudinal clines on several continents. However, these authors could not compare their results to the genomic background to determine whether these allozymes were distinct in their correlation with latitude, which would suggest a role in local adaptation, or whether the whole genome showed latitudinal changes in ancestry, which would suggest that the history of human introductions of this species generates parallel ancestries across continents. Recent studies have used GEAs for markers in sets of candidate genes, such as plant phenology genes (Fitzpatrick and Keller, 2015; Pluess et al., 2016; Fournier-Level et al., 2022) but some have not compared patterns at candidate genes to genomic background patterns. A lack of comparison with genome-wide patterns may be responsible for the decline in these "candidate gene" or reverse ecological genetics approaches in recent years, in favor of forward genetics approaches like genome-wide mapping studies (e.g. environmental GWAS), that are often referred to as "unbiased," but which actually have substantial biases with respect to genetic architecture (similar to those for GWAS, see "favored hypotheses" in Figure 3 and below).

Nevertheless, there are a growing number of studies where researchers have applied GEA to individual genes where natural variants have demonstrated phenotypes (i.e. with more support than simply being unverified phenotypic GWAS hits, which are often spurious) with strong ecological hypotheses for the consequences of natural variation. Below are a few examples of such genes: a seed tannin accumulation gene (*Tannin1*) associated with precipitation and a photoperiod sensitivity (*Maturity1*) gene associated with latitude in sorghum (*Sorghum bicolor*; Lasky et al., 2015); *C-REPEAT/DRE BINDING FACTOR* (CBF) transcription factor genes involved in cold acclimation associated with temperature in *Arabidopsis* (Monroe et al., 2016); a gene (*LOW GERMINATION STIMULANT 1* [LGS1]) controlling root

exudate strigolactones associated with parasitic plant prevalence in sorghum (Bellis et al., 2020), a sodium transporter (*HIGH-AFFINITY K⁺ TRANSPORTER 1 [HKT1;1]*) gene controlling salt tolerance in *Arabidopsis* associated with distance from the coast (Baxter et al., 2010); a MYB-domain transcription factor gene controlling UV-absorbing flavonol glycosides in sunflower (*Helianthus annuus*) associated with temperature and moisture (Todesco et al., 2022); and an ethylene response transcription factor gene involved in oxygen sensing in *Arabidopsis* associated with aridity (Lou et al., 2022). In most cases, the authors showed evidence that these specific genes contributed to local adaptation by comparison to the genomic background correlation with the environment. However, would these genes be found by forward ecological genetics using environmental GWAS and survive a control for multiple testing? In the case of the sorghum genes (Lasky et al., 2015; Bellis et al., 2020), these loci were not detected as environmental GWAS peaks, indicating uncertainty in the power of such genome scans. Future reverse ecological genetics studies could help our understanding of the power of environmental GWAS if these studies also tested whether such GWAS would identify their locally-adapted candidate gene.

Genomic prediction of genotype–environment interactions

If GEAs really do identify genes involved in local adaptation, then by the definition of local adaptation, alleles at these genes should predict changes in fitness (or components of fitness) in response to environmental changes. The genetic prediction of whole organism phenotypes is a major field of research, either based on sets of GWAS loci or whole-genome similarity (Gienapp et al., 2017; McGaugh et al., 2021).

Genomic predictions of population fitness based on GEAs have become popular in conservation genetics. For example, Bay et al. (2018) showed how climate-associated SNPs that were mismatched with future climates were correlated with current population declines. However, Láruson et al. (2022) recently showed that these patterns of mismatch might arise when stronger neutral drift in small populations generates spurious GEAs. Even if GEAs are not effective at identifying individual causal loci, multi-locus GEAs may still reflect local adaptation when many small-effect loci contribute. Gienapp et al. (2017) showed in *Arabidopsis* how whole-genome associations with winter cold predicted change in the most fit accessions across multiple common gardens, Fitzpatrick et al. (2021) and Láruson et al. (2022) showed that GEAs with randomly chosen SNPs or whole-genome variation, respectively, performed just as well as GEAs with causal SNPs for predicting population response to environmental change. Overall these methods show the promise of combining environmental GWAS results with prediction.

Most genomic prediction studies have focused on predicting traits in a single environment, and fewer have focused on predicting G \times E (but see e.g. Technow et al., 2015; Tiezji

et al., 2017; Messina et al., 2018). GEA-based genomic prediction might be a useful complement to traditional phenotype-based predictions of plant performance across environments. Lasky et al. (2015) showed how dozens or hundreds of SNPs with the strongest GEAs could predict the response to experimental drought and aluminum toxicity in sorghum, and Gates et al. (2019) showed in maize that such SNPs can predict response to drought.

A future avenue of research for prediction is to integrate functional genomic information into predictions, so as to leverage existing functional genetic knowledge. In a promising finding highlighting this potential (though not a GEA), the ecologically important trait of flowering time in maize was better predicted by genome-wide allele frequencies in open chromatin (potentially *cis*-regulatory) regions than when predicted by all genomic markers (Rodgers-Melnick et al., 2016).

Tactics for GEAs: A brief overview of technical considerations

Our goal in this section is to provide some brief suggestions for sound use of GEAs, to point the interested reader to more in-depth technical publications, and to suggest some areas where GEA techniques might be improved. For recent comparisons of GEA approaches using simulations, see Lotterhos and Whitlock (2014, 2015); Forester et al. (2016, 2018); Yoder and Tiffin (2018) and for reviews of these methods see Rellstab et al. (2015); Hoban et al. (2016). Despite their relative ease of implementation, there are several points where care should be taken with the use of GEAs. These points are misspecified statistical models, spurious associations due to population structure confounding, and a failure to capture environmental selective gradients.

Statistical models

In principle, if one knew certain details about their study system, including demographic rates and history, and the genetic architecture of adaptive traits, one could rationally select the best GEA model. However, these details are rarely known a priori, and the best GEA approaches might be those that work well across a wide parameter space. Further simulation studies, as well as empirical data on demography and life history, are needed to determine the robustness of GEA inferences to variation in genetic architecture, demographic patterns, and life history.

The main differences among GEA models are in their linearity, environmental dimensionality, and consideration of population structure. Linear model GEAs are often based on the same association models used for phenotypic GWAS, but instead these environmental GWAS test whether different alleles occupy different environments (with normally distributed errors; Lasky et al., 2014; Yoder et al., 2014), though environmental outliers may have unwanted influence (Hancock et al., 2011). Others have implemented nonlinear approaches such as using logit transforms of allele frequencies (Coop et al., 2010), rank correlations between climate

and allele frequency (Hancock et al., 2011), and regression trees (Fournier-Level et al., 2022).

An additional consideration is whether to include a single or multiple environmental variables in the model. Univariate environmental correlations may be simplest to interpret, at the price of ecological realism (see “Modeling environmental variables” section). An alternate approach is to implement models with multiple environmental variables, such as (linear model-based) redundancy analysis (Lasky et al., 2012; Forester et al., 2018) or (regression tree-based) gradient forests (Martínez-Berdeja et al., 2020).

Confounding with population structure

In some (perhaps many) cases, GEAs can reflect demographic processes rather than changing selection and local adaptation. Individuals are often more closely related to nearby individuals compared to individuals further away (Vekemans and Hardy, 2004). This pattern is called isolation by distance (IBD). IBD can arise for a number of reasons, such as limited migration (Wright, 1943; Slatkin, 1993) and recent population expansion (Excoffier et al., 2009). Crucially, if environmental gradients are spatially autocorrelated (as climate gradients often are), then IBD could generate spurious associations between genotype and the environment (Hoban et al., 2016). For example, Arabidopsis exhibits IBD and also has recently spread into Europe from southern glacial refugia (Alonso-Blanco et al., 2016). Since many of the climate factors that matter for Arabidopsis local adaptation also vary north-to-south (Ågren and Schemske, 2012; Monroe et al., 2016; Gienapp et al., 2017), IBD might create spurious associations between genotype and climate (Hancock et al., 2011). These patterns could also be important for many crop species, which have often recently spread from one or a few domestication locations (Romay et al., 2013; Gutaker et al., 2020) and show substantial patterns of IBD (Gutaker et al., 2020; Lasky et al., 2022).

It is often considered important to attempt to control for IBD or relatedness (population structure) to reduce spurious GEAs that inflate genome-wide false positive rates (Rellstab et al., 2015). However, researchers are often most interested in the top environmental GWAS hits and will rarely follow up on all significant genome-wide loci, so it is not always clear if a genome-wide false positive control is necessary (Forester et al., 2018; Price et al., 2020). Nevertheless, commonly used GEA methods control for population structure using model parameters that account for genome-wide (putatively indicating relatedness) associations with environmental gradients (Coop et al., 2010; Hancock et al., 2011) or fitted latent factors that model relatedness (Frichot et al., 2013). An alternative approach is to control for spatial structure in GEAs that may arise from IBD (Lasky et al., 2012). In general, when conducting many statistical tests across the genome (as in GWAS), many markers will be “significantly” associated even if the null model of no association is true for all of them. This genome-wide false discovery rate can be estimated or reduced with additional tests or statistical adjustments (Han et al., 2009; Benjamini, 2010).

Separate from the problem of discovery in environmental GWAS, researchers may be interested in testing whether a functional natural variant shows evidence of local adaptation to different environments. In this case, controlling for false positives is important, so it is important to use one of the GEAs that account for population structure. However, if a known ecologically important variant shows a GEA, but the GEA is not significantly different from the null expectation when controlling for population structure, this does not mean the variant does not play a role in local adaptation, but merely that the environmental association is not strong support for this hypothesis. For example, Lasky et al. (2015) noted that loss of photoperiod sensitivity alleles at *Maturity1* in sorghum landraces increased nearly to fixation with decreasing latitude in southern Africa, a GEA that was consistent with the ecological hypothesis of changing selection for photoperiod sensitivity. However, this association was not significant after accounting for population structure, indicating that many loci all across the genome show similar latitudinal clines in southern Africa.

Modeling environmental variables

There are several considerations for modeling environments for GEAs. Ideally, GEA would use an environmental variable that perfectly captures changing selection. This variable would likely be a combination of commonly measured environmental variables. In reality, researchers have to make their best educated guess to select or create environmental variables to test in GEA. Ecological variation that drives changing selection between populations is more complex than, and may not be well-captured by, global gridded monthly climate datasets. For this reason, a treatment of climate that specifically aims to characterize the potentially stressful aspects may be better (e.g. bioclimatic variables; Hijmans et al., 2005).

Relatedly, models of growing seasons and plant growth and development, built from heuristics or from first ecophysiological principles, may be one way to build more ecologically realistic proxies for true selective gradients (Lasky et al., 2012, 2015; Messina et al., 2018). Alternatively, naive, data-driven approaches to statistical models of GEAs (e.g. gradient forests) could potentially capture some interactions among environmental gradients (Fitzpatrick and Keller, 2015).

Are GEAs living up to their promise? And how would we know?

In this section, we lay out how GEAs may be integrated into a research program, based on existing knowledge of the power of GEAs. To summarize, we suggest that GEAs may be most useful for generating and testing hypotheses of local adaptation for natural variants with known or highly likely phenotypic effects. This approach is similar to “reverse genetics” and might be termed “reverse ecological genetics.” Our overarching rationale for this suggestion is that many variants identified in GWAS (and likely environmental

GWAS) will be false positives with no phenotypic influence due to issues caused by population structure, synthetic associations, and model misspecification, and also because we lack experimental validation of candidate genes identified by environmental GWAS. By contrast, there are many cases of genetic variation in known QTL or genes with some functional information that can reasonably be hypothesized to be ecologically important, and reverse ecological genetics with GEAs can help test these hypotheses.

Testing hypotheses generated by forward ecological genetics

Despite the proliferation of environmental GWAS for discovery of locally-adapted loci, most validations of environmental GWAS inferences have not validated specific genes. In part, this gap may be due to the fact that population genomics in nonmodel plants has advanced much more rapidly than functional genomics. Also, experimental validation generally requires development of genetic stocks via methods (e.g. NILs or gene editing) that are, at best, time-consuming or, at worst, not possible in nonmodel plants. Because of this lack of gene-level validation, we are left without a solid understanding of the reliability of environmental GWAS inferences. This is a problem in general with GWAS, where many more association studies are published than there are follow-up experiments showing that the strongest associated genes underlie a given QTL for the trait (Alsheikh et al., 2022; Figure 4).

While computer simulations have indicated that environmental GWAS should help discover loci (Lotterhos and Whitlock, 2015; Forester et al., 2016, 2018; Lasky et al., 2018; Yoder and Tiffin, 2018), it is unclear how well these simulations reflect the true underlying genetic architecture. Furthermore, it is hard to assess environmental GWAS based on the performance of the few known locally-adaptive positive control genes. Indeed, as mentioned above, environmental GWAS often fails to find these loci as signals that stand out across the genome. This does not mean that the environmental GWAS results are incorrect, because these “positive controls” were identified via alternate approaches and are not necessarily expected to be genome-wide outliers. For instance, in a global sorghum diversity panel, cloned climate-adaptive natural variants at the WD40 transcriptional regulator gene *Tannin1* and *Maturity1*, encoding PSEUDO-RESPONSE REGULATOR 37 (PRR37), were associated with expected climate variables, but were not genome-wide outliers (Lasky et al., 2015). However, these known locally-adapted genes cannot support the hypothesis that environmental GWAS are working as intended in empirical systems. Better negative controls, such as testing randomly selected genes (Chong and Stinchcombe, 2019), could also help assess the utility of environmental GWAS. Environmental GWAS may turn out to be a good route for the discovery of locally-adapted loci, but conclusions about the efficacy of environmental GWAS are premature.

To remedy this lack of certainty surrounding the question of whether environmental GWAS is a useful and efficient route to discovery, the field requires more studies that systematically test individual locus hypotheses generated from environmental GWAS. We suggest several potential routes to the validation of loci identified by environmental GWAS. To conclusively demonstrate that a locus is involved in local adaptation requires the generation of NILs or genetically engineered genotypes (e.g. transgenic or gene-edited lines) that recapitulate natural variation isolated to the single gene of interest, and the demonstration that the gene-specific natural variation causes fitness tradeoffs across the candidate environmental gradient. This latter condition has not yet been met by any environmental GWAS study. Ideally, these experiments would be conducted in contrasting field environments that represent the environments to which each allele is putatively adapted; if the single locus variation resulted in performance tradeoffs under these conditions, these results would indicate likely local adaptation. When the relevant field environments are logistically prohibitive, it would still be a major advance to use naturalistic conditions replicated elsewhere (e.g. field, greenhouse, or growth chamber). A separate approach to validating GEAs or environmental GWAS is to test whether alleles at multiple identified loci allow prediction of plant fitness responses to a given environment (Lasky et al., 2015; Gienapp et al., 2017; Gates et al., 2019), although this approach does not allow for strong conclusions about any individual genes.

Using reverse ecological genetics to test hypotheses

Researchers may often develop a hypothesis that a particular gene is ecologically important, and in particular that variation in the function of this gene (e.g. expression level, context, or protein function) might contribute to tradeoffs in plant fitness across environments. If natural variation exists in this gene, the researchers may also hypothesize this variation is maintained due to changes in selection among populations, contributing to local adaptation. By contrast, despite the initial hypothesis of ecological importance, the candidate gene might show very little natural variation, suggesting that positive or purifying selection dominates over spatially varying selection.

If such a candidate gene, when genotyped in multiple populations, does exhibit variation, then GEAs are one way to test the hypothesis that this variation is locally adapted. One benefit of testing the hypothesis of local adaptation using GEAs and the geographic distributions of alleles at this gene is that these population genetic patterns can suggest the potential value of each allele under a much wider range of environmental and genetic contexts than can be quickly evaluated by researchers experimentally. GEAs can suggest potential axes for ecophysiological tradeoffs not previously considered by researchers.

An example of our suggested implementation can be found in a GEA study (Bellis et al., 2020) that built on the identification of the *LGS1* gene (Gobena et al., 2017). First, Gobena et al. (Gobena et al., 2017) fine-mapped a QTL in

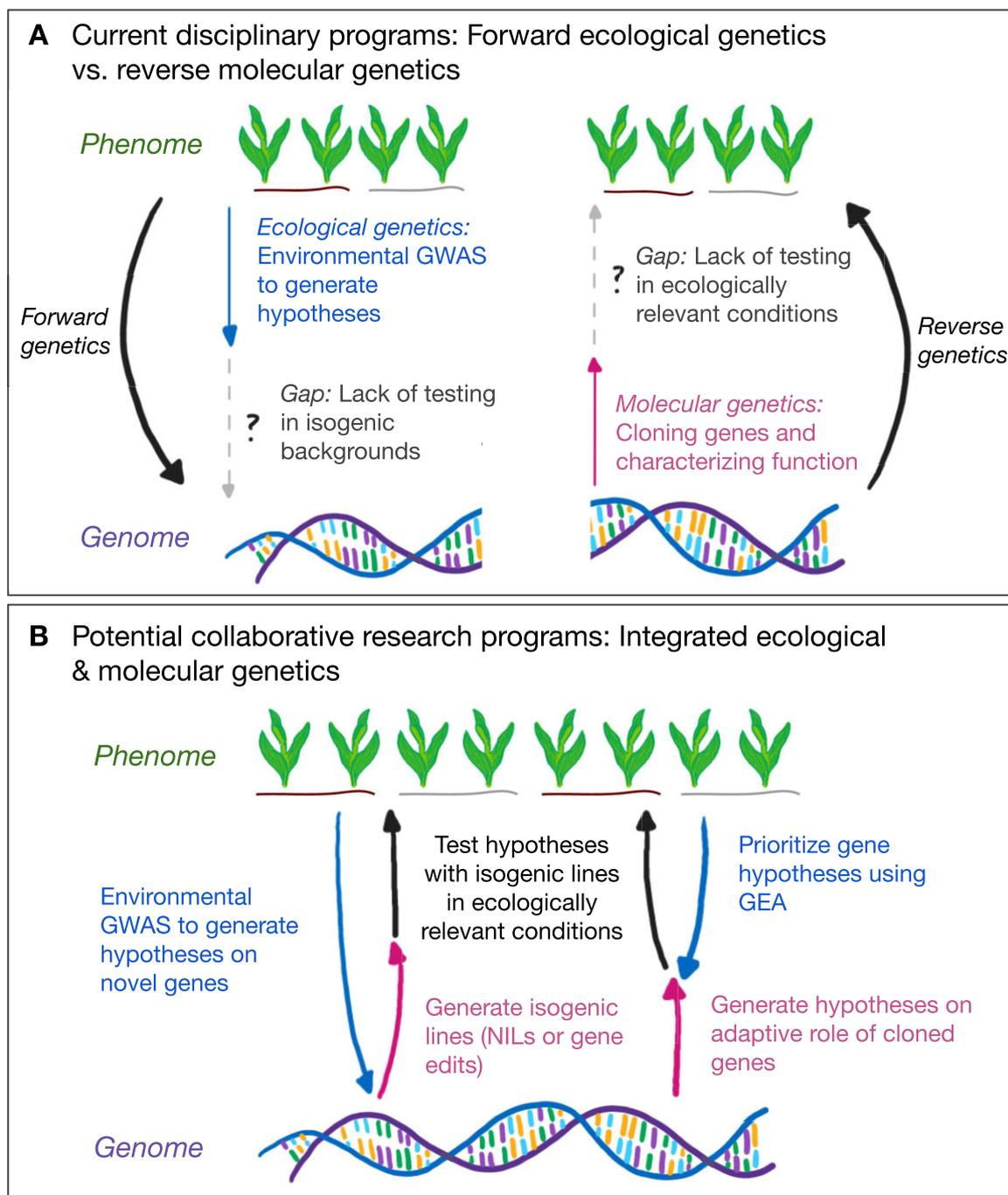


Figure 4 The role of GEA in understanding the genome–phenome landscape for adaptation. Understanding the molecular basis of adaptation is a major goal of plant biology, but various subdisciplines focus on different parts of the genome–phenome landscape (A). For instance, ecological geneticists generally use a forward genetics approach, starting with proxies for fitness or traits thought to be involved in local adaptation, then using genome-wide mapping approaches such as environmental GWAS to generate hypotheses on the molecular basis of adaptation. Note that the phenome is the complete set of traits of an organism, and environmental GWAS may capture multiple phenotypes locally adapted along a gradient. Unfortunately, due to the large number of hypotheses generated by environmental GWAS and the technical challenge of testing each one, most environmental GWAS-based hypotheses have not yet been tested (like phenotypic GWAS). By contrast, molecular geneticists more often use reverse genetics approaches and may not have direct access to ecologically relevant field environments to test hypotheses on the adaptive role of genes and molecular variants. B, A collaborative interdisciplinary approach that integrates ecological and molecular genetics may be needed to more rigorously and efficiently test hypotheses. Illustration by Gina Cerimele.

sorghum for resistance to the parasitic plant purple witchweed (*Striga hermonthica*), using a biparental mapping population and other subsequent crosses (Gobena et al., 2017).

After the identification of a loss-of-function allele at the *LGS1* locus as being putatively resistant, several questions remained unanswered. For example, might this resistance

allele underlie local adaptation to *S. hermonthica* parasitism by sorghum landraces? If so, might this allele be beneficial in a wide range of field settings across tropical Africa where *S. hermonthica* is prevalent? Is this allele beneficial in a wide range of genetic backgrounds? Are there potential tradeoffs associated with this allele? Are additional alleles at *LGS1* involved in local adaptation in sorghum landraces? GEAs and population genomics were well-suited to make quick progress on these questions. Using GEA, Bellis et al. (Bellis et al., 2020) found that the alleles identified by Gobena and colleagues were present in greater frequencies in regions where *S. hermonthica* was most prevalent, which also held true for diverse genetic backgrounds and across West and East Africa. However, Bellis et al. (Bellis et al., 2020) also showed that these putative resistance alleles were never carried by a majority of local sorghum landraces, even in the worst *S. hermonthica*-prone regions, and that where *S. hermonthica* is rare, the resistance alleles are essentially absent. These GEA findings supported a hypothesis of local adaptation at *LGS1* and motivated Bellis et al. (2020) to further search for tradeoffs at *LGS1* with additional phenotyping.

Overall, published GEA analyses of verified natural variants in plants suggest that the simplest models for the genetic basis of environmental adaptation (e.g. a limited number of relatively common major variants in master regulators or structural genes) may not be prevalent in nature. Instead, more complex genetic architectures that encompass many genes and their direct and indirect interactions, may be prevalent for ecologically important traits (Rockman, 2012; Mathieson, 2021). In such case, it will be important to use GEAs that explicitly consider these architectures (e.g. multivariate and nonlinear approaches; Fitzpatrick and Keller, 2015; Forester et al., 2018), although the challenges remain as to how to explore the large number of potential genetic interactions and how to apply knowledge of polygenic local adaptation.

Conclusion

Despite the rapid adoption of GEAs and environmental GWAS, many important questions remain on the validity and interpretation of their findings, largely arising from fundamental unanswered questions on the genetic architecture of adaptation. We outlined strategies that could better ground GEAs and method development in the underlying hypotheses on genetic architecture and better test environmental GWAS-generated hypotheses using genetics and eco-physiology. When combined with a rigorous hypothesis testing framework, GEAs will have the potential to facilitate our understanding of the molecular basis of climate adaptation and plant improvement.

Acknowledgments

We thank two anonymous reviewers for comments on an earlier version of this manuscript. The contents are the sole responsibility of the authors and do not necessarily reflect

the views of USAID or the United States Government. Photos were obtained from https://commons.wikimedia.org/wiki/Category:Arabidopsis_thaliana#/media/File:Arabidopsis_thaliana_2019-04-27_2246.jpg [https://commons.wikimedia.org/wiki/File:Rice_Plants_\(IRRI\).jpg](https://commons.wikimedia.org/wiki/File:Rice_Plants_(IRRI).jpg) https://commons.wikimedia.org/wiki/File:Panicum_virgatum_Heavy_Metal_1zz.jpg https://commons.wikimedia.org/wiki/Populus_trichocarpa#/media/File:Populus_trichocarpa_Umatilla.jpg https://commons.wikimedia.org/wiki/File:Helianthus_annuus_00001.jpg

Funding

J.R.L. was supported by National Institutes of Health (NIH) R35 GM138300. E.B.J. was supported by NIH R35 GM142829. G.P.M. was supported by United States Agency for International Development (USAID): this study is made possible by the support of the American People provided to the Feed the Future Innovation Lab for Collaborative Research on Sorghum & Millet and the Innovation Lab for Crop Improvement through the USAID under Cooperative Agreement No. AID-OAA-A-13-00047 and 7200AA-19LE-00005.

Conflict of interest statement. None declared.

References

- Ågren J, Schemske DW (2012) Reciprocal transplants demonstrate strong adaptive differentiation of the model organism *Arabidopsis thaliana* in its native range. *New Phytol* **194**: 1112–1122
- Alonso-Blanco C, Andrade J, Becker C, Bemm F, Bergelson J, Borgwardt KM, Cao J, Chae E, Dezaan TM, Ding W et al. (2016) 1,135 genomes reveal the global pattern of polymorphism in *Arabidopsis thaliana*. *Cell* **166**: 481–491
- Alsheikh AJ, Wollenhaupt S, King EA, Reeb J, Ghosh S, Stolzenburg LR, Tamim S, Lazar J, Davis JW, Jacob HJ (2022) The landscape of GWAS validation; systematic review identifying 309 validated non-coding variants across 130 human diseases. *BMC Med Genom* **15**: 74
- Anderson JT, Willis JH, Mitchell-Olds T (2011) Evolutionary genetics of plant adaptation. *Trends Genet* **27**: 258–266
- Angert AL, Bontrager MG, Ågren J (2020) What do we really know about adaptation at range edges? *Annu Rev Ecol Evol Syst* **51**: 341–361
- Baxter I, Brazelton JN, Yu D, Huang YS, Lahner B, Yakubova E, Li Y, Bergelson J, Borevitz JO, Nordborg M, et al. (2010) A coastal cline in sodium accumulation in *Arabidopsis thaliana* is driven by natural variation of the sodium transporter *AtHKT1;1*. *PLoS Genet* **6**: e1001193
- Bay RA, Harrigan RJ, Underwood VL, Gibbs HL, Smith TB, Ruegg K (2018) Genomic signals of selection predict climate-driven population declines in a migratory bird. *Science* **359**: 83–86
- Bay RA, Palumbi SR (2014) Multilocus adaptation associated with heat resistance in reef-building corals. *Curr Biol* **24**: 2952–2956
- Beavis WD (1994) The power and deceit of QTL experiments: lessons from comparative QTL studies. *In* Proceedings of the forty-ninth annual corn and sorghum industry research conference (Chicago, IL), p 266
- Bellis ES, Kelly EA, Lorts CM, Gao H, DeLeo VL, Rouhan G, Budden A, Bhaskara GB, Hu Z, Muscarella R et al. (2020) Genomics of sorghum local adaptation to a parasitic plant. *Proc Natl Acad Sci USA* **117**: 4243–4251

- Benjamini Y** (2010) Discovering the false discovery rate. *J R Stat Soc B Stat Methodol* **72**: 405–416
- Brachi B, Morris GP, Borevitz JO** (2011) Genome-wide association studies in plants: the missing heritability is in the field. *Genome Biol* **12**: 232
- Capblancq T, Forester BR** (2021) Redundancy analysis: A Swiss Army Knife for landscape genomics. *Methods Ecol Evol* **12**: 2298–2309
- Chong VK, Stinchcombe JR** (2019) Evaluating population genomic candidate genes underlying flowering time in *Arabidopsis thaliana* using T-DNA insertion lines. *J Hered* **110**: 445–454
- Clausen J, Keck DD, Hiesey WH** (1940) Experimental Studies on the Nature of Species I. Effects of Varied Environments on Western North American Plants. Carnegie Institution of Washington. Washington, DC, USA
- Coop G, Witonsky D, Di Rienzo A, Pritchard JK** (2010) Using environmental correlations to identify loci underlying local adaptation. *Genetics* **185**: 1411–1423
- De La Torre AR, Wilhite B, Neale DB** (2019) Environmental genome-wide association reveals climate adaptation is shaped by subtle to moderate allele frequency shifts in loblolly pine. *Genome Biol Evol* **11**: 2976–2989
- Eckert AJ, Bower AD, González-Martínez SC, Wegrzyn JL, Coop G, Neale DB** (2010) Back to nature: ecological genomics of loblolly pine (*Pinus taeda*, Pinaceae). *Mol Ecol* **19**: 3789–3805
- Eckert AJ, Maloney PE, Vogler DR, Jensen CE, Mix AD, Neale DB** (2015) Local adaptation at fine spatial scales: an example from sugar pine (*Pinus lambertiana*, Pinaceae). *Tree Genet Genomes* **11**: 1–17
- Evans LM, Slavov GT, Rodgers-Melnick E, Martin J, Ranjan P, Muchero W, Brunner AM, Schackwitz W, Gunter L, Chen J-G, et al.** (2014) Population genomics of *Populus trichocarpa* identifies signatures of selection and adaptive trait associations. *Nat Genet* **46**: 1089–1096
- Excoffier L, Foll M, Petit RJ** (2009) Genetic consequences of range expansions. *Annu Rev Ecol Evol Syst* **40**: 481–501
- Exposito-Alonso M, Burbano HA, Bossdorf O, Nielsen R, Weigel D** (2019) Natural selection on the *Arabidopsis thaliana* genome in present and future climates. *Nature* **573**: 126–129
- Faye JM, Maina F, Hu Z, Fonceka D, Cisse N, Morris GP** (2019) Genomic signatures of adaptation to Sahelian and Soudanian climates in sorghum landraces of Senegal. *Ecol Evol* **9**: 1–14
- Ferrero-Serrano Á, Sylvia MM, Forstmeier PC, Olson AJ, Ware D, Bevilacqua PC, Assmann SM** (2022) Experimental demonstration and pan-structurome prediction of climate-associated riboSNitches in *Arabidopsis*. *Genome Biol* **23**: 101
- Fitzpatrick MC, Chhatre VE, Soolanayakanahally RY, Keller SR** (2021) Experimental support for genomic prediction of climate maladaptation using the machine learning approach Gradient Forests. *Mol Ecol Res* **21**: 2749–2765
- Fitzpatrick MC, Keller SR** (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 BR, Jones MR, Joost S, Landguth EL, Lasky JR** (2016) Detecting spatial genetic signatures of local adaptation in heterogeneous landscapes. *Mol Ecol* **25**: 104–120
- Forester BR, Lasky JR, Wagner HH, Urban DL** (2018) Comparing methods for detecting multilocus adaptation with multivariate genotype–environment associations. *Mol Ecol* **27**: 2215–2233
- Fournier-Level A, Taylor MA, Paril JF, Martínez-Berdeja A, Stitzer MC, Cooper MD, Roe JL, Wilczek AM, Schmitt J** (2022) Adaptive significance of flowering time variation across natural seasonal environments in *Arabidopsis thaliana*. *New Phytol* **234**: 719–734
- Franks SJ** (2011) Plasticity and evolution in drought avoidance and escape in the annual plant *Brassica rapa*. *New Phytol* **190**: 249–257
- Frichot E, Schoville SD, Bouchard G, François O** (2013) Testing for associations between loci and environmental gradients using latent factor mixed models. *Mol Biol Evol* **30**: 1687–1699
- Gamba D, Lorts C, Haile A, Sahay S, Lopez L, Xia T, Kulesza E, Elango D, Kerby J, Yifru M et al.** (2022) The genomics and physiology of abiotic stressors associated with global elevation gradients in *Arabidopsis thaliana*. *bioRxiv* 2022.03.22.485410
- Gates DJ, Runcie D, Janzen GM, Navarro AR, Willcox M, Sonder K, Snodgrass S, Rodríguez-Zapata F, Sawers RJ, Rellán-Alvarez R** (2019) Single-gene resolution of locally adaptive genetic variation in Mexican maize. *bioRxiv*: 706739
- Gibson MJS, Moyle LC** (2020) Regional differences in the abiotic environment contribute to genomic divergence within a wild tomato species. *Mol Ecol* **29**: 2204–2217
- Gienapp P, Fior S, Guillaume F, Lasky JR, Sork VL, Csilléry K** (2017) Genomic quantitative genetics to study evolution in the wild. *Trends Ecol Evol* **32**: 897–908
- Gobena D, Shimels M, Rich PJ, Ruyter-Spira C, Bouwmeester H, Kanuganti S, Mengiste T, Ejeta G** (2017) Mutation in sorghum LOW GERMINATION STIMULANT 1 alters strigolactones and causes Striga resistance. *Proc Natl Acad Sci USA* **114**: 4471–4476
- Gutaker RM, Groen SM, Bellis ES, Choi JJ, Pires JS, Bocinsky RK, Slayton ER, Wilkins O, Castillo CC, Negrao S et al.** (2020) Genomic history and ecology of the geographic spread of rice. *Nat Plants* **6**: 492–502
- Han B, Kang HM, Eskin E** (2009) Rapid and accurate multiple testing correction and power estimation for millions of correlated markers. *PLoS Genet* **5**: e1000456
- Hancock AM, Brachi B, Faure N, Horton MW, Jarymowycz LB, Sperone FG, Toomajian C, Roux F, Bergelson J** (2011) Adaptation to climate across the *Arabidopsis thaliana* genome. *Science* **334**: 83–86
- Hancock AM, Witonsky DB, Gordon AS, Eshel G, Pritchard JK, Coop G, Di Rienzo A** (2008) Adaptations to climate in candidate genes for common metabolic disorders. *PLoS Genet* **4**: e32
- Hereford J** (2009) A quantitative survey of local adaptation and fitness trade-offs. *Am Nat* **173**: 579–588
- Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A** (2005) Very high resolution interpolated climate surfaces for global land areas. *Int J Climatol* **25**: 1965–1978
- Hoban S, Kelley JL, Lotterhos KE, Antolin MF, Bradburd G, Lowry DB, Poss ML, Reed LK, Storfer A, Whitlock MC** (2016) Finding the genomic basis of local adaptation: pitfalls, practical solutions, and future directions. *Am Nat* **188**: 379–397
- Hufford MB, Seetharam AS, Woodhouse MR, Chougule KM, Ou S, Liu J, Ricci WA, Guo T, Olson A, Qiu Y et al.** (2021) De novo assembly, annotation, and comparative analysis of 26 diverse maize genomes. *Science* **373**: 655–662
- Kawecki TJ, Ebert D** (2004) Conceptual issues in local adaptation. *Ecol Lett* **7**: 1225–1241
- Láruson AJ, Fitzpatrick MC, Keller SR, Haller BC, Lotterhos KE** (2022) Seeing the forest for the trees: assessing genetic offset predictions from gradient forest. *Evol Appl* **15**: 403–416
- Lasky JR, Upadhyaya HD, Ramu P, Deshpande S, Hash CT, Bonnette J, Juenger TE, Hyma K, Acharya C, Mitchell SE et al.** (2015) Genome-environment associations in sorghum landraces predict adaptive traits. *Sci Adv* **1**: e1400218
- Lasky JR, Des Marais DL, Lowry DB, Povolotskaya I, McKay JK, Richards JH, Keitt TH, Juenger TE** (2014) Natural variation in abiotic stress responsive gene expression and local adaptation to climate in *Arabidopsis thaliana*. *Mol Biol Evol* **31**: 2283–2296
- Lasky JR, Des Marais DL, McKay JK, Richards JH, Juenger TE, Keitt TH** (2012) Characterizing genomic variation of *Arabidopsis thaliana*: the roles of geography and climate. *Mol Ecol* **21**: 5512–5529
- Lasky JR, Forester BR, Reimherr M** (2018) Coherent synthesis of genomic associations with phenotypes and home environments. *Mol Ecol Res* **18**: 91–106

- Lasky JR, Gamba D, Keitt TH (2022) Wavelet characterization of spatial pattern in allele frequency. 2022.03.21.485229
- Le Corre V, Kremer A (2012) The genetic differentiation at quantitative trait loci under local adaptation. *Mol Ecol* **21**: 1548–1566
- LeBauer D, Maxwell B, Demieville J, Fahlgren N, French A (2020) Data from: TERRA-REF, an open reference data set from high resolution genomics, phenomics, and imaging sensors. Dryad Dataset **10**
- Lee KM, Coop G (2019) Population genomics perspectives on convergent adaptation. *Philos Trans R Soc B Biol Sci* **374**: 20180236
- Leimu R, Fischer M (2008) A meta-analysis of local adaptation in plants. *PLoS One* **3**: e4010
- Lewontin RC, Krakauer J (1973) Distribution of gene frequency as a test of the theory of the selective neutrality of polymorphisms. *Genetics* **74**: 175–195
- Lopez-Arboleda WA, Reinert S, Nordborg M, Korte A (2021) Global Genetic Heterogeneity in Adaptive Traits. *Mol Biol Evol* **38**: 4822–4831
- Lotterhos KE, Whitlock MC (2014) Evaluation of demographic history and neutral parameterization on the performance of FST outlier tests. *Mol Ecol* **23**: 2178–2192
- Lotterhos KE, Whitlock MC (2015) The relative power of genome scans to detect local adaptation depends on sampling design and statistical method. *Mol Ecol* **24**: 1031–1046
- Lou S, Guo X, Liu L, Song Y, Zhang L, Jiang Y, Zhang L, Sun P, Liu B, Tong S et al. (2022) Allelic shift in cis-elements of the transcription factor RAP2.12 underlies adaptation associated with humidity in *Arabidopsis thaliana*. *Sci Adv* **8**: eabn8281
- Lovell JT, MacQueen AH, Mamidi S, Bonnette J, Jenkins J, Napier JD, Sreedasyam A, Healey A, Session A, Shu S et al. (2021) Genomic mechanisms of climate adaptation in polyploid bioenergy switchgrass. *Nature* **590**: 438–444
- Ludlow MM (1989) Strategies of response to water stress. In KH Kreeb, H Richter, TM Hinckley, eds, *Structural and Functional Responses to Environmental Stresses*, SPB Academic, The Hague, pp 269–281
- Luu K, Bazin E, Blum MGB (2017) pcadapt: an R package to perform genome scans for selection based on principal component analysis. *Mol Ecol Res* **17**: 67–77
- Manel S, Poncet BN, Legendre P, Gugerli F, Holderegger R (2010) Common factors drive adaptive genetic variation at different spatial scales in *Arabis alpina*. *Mol Ecol* **19**: 3824–3835
- Martínez-Berdeja A, Stitzer MC, Taylor MA, Okada M, Ezcurra E, Runcie DE, Schmitt J (2020) Functional variants of DOG1 control seed chilling responses and variation in seasonal life-history strategies in *Arabidopsis thaliana*. *Proc Natl Acad Sci USA* **117**: 2526–2534
- Mathieson I (2021) The omnigenic model and polygenic prediction of complex traits. *Am J Hum Genet* **108**: S0002929721002718
- Maynard Smith J, Haigh J (1974) The hitch-hiking effect of a favourable gene. *Genet Res* **23**: 23–35
- McGaugh SE, Lorenz AJ, Flagel LE (2021) The utility of genomic prediction models in evolutionary genetics. *Proc R Soc B Biol Sci* **288**: 20210693
- Messina CD, Technow F, Tang T, Totir R, Gho C, Cooper M (2018) Leveraging biological insight and environmental variation to improve phenotypic prediction: Integrating crop growth models (CGM) with whole genome prediction (WGP). *Eur J Agron* **100**: 151–162
- Monnahan PJ, Kelly JK (2015) Epistasis is a major determinant of the additive genetic variance in *Mimulus guttatus*. *PLoS Genet* **11**: e1005201
- Monroe JG, McGovern C, Lasky JR, Grogan K, Beck J, McKay JK (2016) Adaptation to warmer climates by parallel functional evolution of CBF genes in *Arabidopsis thaliana*. *Mol Ecol* **25**: 3632–3644
- Oakeshott JG, Gibson JB, Anderson PR, Knibb WR, Anderson DG, Chambers GK (1982) Alcohol dehydrogenase and glycerol-3-phosphate dehydrogenase clines in *Drosophila melanogaster* on different continents. *Evolution* **36**: 86–96
- Olatoye MO, Hu Z, Maina F, Morris GP (2018) Genomic signatures of adaptation to a precipitation gradient in Nigerian sorghum. *G3* **8**: 3269–3281
- Palmer C, Pe'er I (2017) Statistical correction of the Winner's Curse explains replication variability in quantitative trait genome-wide association studies. *PLoS Genet* **13**: e1006916
- Pavlidis P, Jensen JD, Stephan W, Stamatakis A (2012) A critical assessment of storytelling: gene ontology categories and the importance of validating genomic scans. *Mol Biol Evol* **29**: 3237–3248
- Pavlidis P, Živković D, Stamatakis A, Alachiotis N (2013) SweeD: likelihood-based detection of selective sweeps in thousands of genomes. *Mol Biol Evol* **30**: 2224–2234
- Platt JR (1964) Strong inference. *Science* **146**: 347–353
- Pluess AR, Frank A, Heiri C, Laägüe H, Vendramin GG, Oddou-Muratorio S (2016) Genome–environment association study suggests local adaptation to climate at the regional scale in *Fagus sylvatica*. *New Phytol* **210**: 589–601
- Price N, Lopez L, Platts AE, Lasky JR (2020) In the presence of population structure: from genomics to candidate genes underlying local adaptation. *Ecol Evol* **10**: 1889–1904
- Price N, Moyers BT, Lopez L, Lasky JR, Monroe JG, Mullen JL, Oakley CG, Lin J, Ågren J, Schridder DR, et al. (2018) Combining population genomics and fitness QTLs to identify the genetics of local adaptation in *Arabidopsis thaliana*. *Proc Natl Acad Sci USA* **115**: 5028–5033
- Reilistab C, Gugerli F, Eckert AJ, Hancock AM, Holderegger R (2015) A practical guide to environmental association analysis in landscape genomics. *Mol Ecol* **24**: 4348–4370
- Rockman MV (2012) The Qtn program and the alleles that matter for evolution: all that's gold does not glitter. *Evolution* **66**: 1–17
- Rodgers-Melnick E, Vera DL, Bass HW, Buckler ES (2016) Open chromatin reveals the functional maize genome. *Proc Natl Acad Sci USA* **113**: E3177–E3184
- Romay MC, Millard MJ, Glaubitz JC, Peiffer JA, Swarts KL, Casstevens TM (2013) Comprehensive genotyping of the USA national maize inbred seed bank. *Genome Biol* **14**: R55
- Siol M, Wright SI, Barrett SCH (2010) The population genomics of plant adaptation. *New Phytol* **188**: 313–332
- Slatkin M (1973) Gene flow and selection in a cline. *Genetics* **75**: 733–756
- Slatkin M (1993) Isolation by distance in equilibrium and non-equilibrium populations. *Evolution* **47**: 264–279
- Sork VL, Davis FW, Westfall R, Flint A, Ikegami M, Wang H, Grivet D (2010) Gene movement and genetic association with regional climate gradients in California valley oak (*Quercus lobata* Née) in the face of climate change. *Mol Ecol* **19**: 3806–3823
- Tao Y, Luo H, Xu J, Cruickshank A, Zhao X, Teng F, Hathorn A, Wu X, Liu Y, Shatte T et al. (2021) Extensive variation within the pan-genome of cultivated and wild sorghum. *Nat Plants* **7**: 766–773
- Technow F, Messina CD, Totir LR, Cooper M (2015) Integrating crop growth models with whole genome prediction through approximate Bayesian computation. *PLoS One* **10**: e0130855
- Tiezzi F, de los Campos G, Parker Gaddis KL, Maltecca C (2017) Genotype by environment (climate) interaction improves genomic prediction for production traits in US Holstein cattle. *J Dairy Sci* **100**: 2042–2056
- Todesco M, Owens GL, Bercovich N, Soudi J-S, Burge S, Huang DO, Ostevik K, Drummond KL, EBM et al. (2020) Massive haplotypes underlie ecotypic differentiation in sunflowers. *Nature* **584**: 602–607

- Todesco M, Bercovich N, Kim A, Imerovski I, Owens GL, Dorado Ruiz Ó, Holalu SV, Madilao LL, Jahani M, Légaré J-S, et al.** (2022) Genetic basis and dual adaptive role of floral pigmentation in sunflowers. *eLife* **11**: e72072
- van Boheemen LA, Hodgins KA** (2020) Rapid repeatable phenotypic and genomic adaptation following multiple introductions. *Mol Ecol* **29**: 4102–4117
- Vekemans X, Hardy OJ** (2004) New insights from fine-scale spatial genetic structure analyses in plant populations. *Mol Ecol* **13**: 921–935
- Vigouroux Y, Mariac C, De Mita S, Pham JL, Gérard B, Kapran I, Sagnard F, Deu M, Chantreau J, Ali A et al.** (2011) Selection for earlier flowering crop associated with climatic variations in the sahel. *PLoS One* **6**: e19563
- Wadgyar SM, Lowry DB, Gould BA, Byron CN, Mactavish RM, Anderson JT** (2017) Identifying targets and agents of selection: innovative methods to evaluate the processes that contribute to local adaptation. *Methods Ecol Evol* **8**: 738–749
- Wang W, Mauleon R, Hu Z, Chebotarov D, Tai S, Wu Z, Li M, Zheng T, Fuentes RR, Zhang F** (2018) Genomic variation in 3,010 diverse accessions of Asian cultivated rice. *Nature* **557**: 43
- Wright S** (1943) Isolation by distance. *Genetics* **28**: 114
- Yeaman S, Hodgins KA, Lotterhos KE, Suren H, Nadeau S, Degner JC, Nurkowski KA, Smets P, Wang T, Gray LK** (2016) Convergent local adaptation to climate in distantly related conifers. *Science* **353**: 1431–1433
- Yoder JB, Stanton-Geddes J, Zhou P, Briskine R, Young ND, Tiffin P** (2014) Genomic signature of adaptation to climate in *Medicago truncatula*. *Genetics* **196**: 1263–1275
- Yoder JB, Tiffin P** (2018) Effects of gene action, marker density, and timing of selection on the performance of landscape genomic scans of local adaptation. *J Hered* **109**: 16–28
- Zhao Q, Feng Q, Lu H, Li Y, Wang A, Tian Q, Zhan Q, Lu Y, Zhang L, Huang T et al.** (2018) Pan-genome analysis highlights the extent of genomic variation in cultivated and wild rice. *Nat Genet* **1**: 278–284