Review

CellPress

Advances and limits of using population genetics to understand local adaptation

Peter Tiffin[1] **and Jeffrey Ross-Ibarra**[2]

[1] Department of Plant Biology, University of Minnesota, 250 Biosciences, Saint Paul, MN 55108, USA
[2] Department of Plant Sciences, Center for Population Biology, and Genome Center, 262 Robbins Hall, Mail Stop 4, University of California, One Shields Ave, Davis, CA 95616, USA

Local adaptation shapes species diversity, can be a stepping stone to ecological speciation, and can facilitate species range expansion. Population genetic analyses, which complement organismal approaches in advancing our understanding of local adaptation, have become widespread in recent years. We focus here on using population genetics to address some key questions in local adaptation: what traits are involved? What environmental variables are the most important? Does local adaptation target the same genes in related species? Do loci responsible for local adaptation exhibit trade-offs across environments? After discussing these questions we highlight important limitations to population genetic analyses including challenges with obtaining high-quality data, deciding which loci are targets of selection, and limits to identifying the genetic basis of local adaptation.

Population genetics complements manipulation experiments

Local adaptation (see Glossary) results when populations of a species evolve in response to geographically variable selection. Decades of field studies and manipulative experiments have established local adaptation as being extremely common [1] and central to understanding the role of adaptation in shaping species diversity. Local adaptation also can contribute to the maintenance of genetic variation, be a stepping stone to ecological speciation, and facilitate species range expansion (reviewed in [2]).

Local adaptation has been an area of active study by evolutionary ecologists since Turresson [3] first defined the concept of ecotypes and Clausen, Keck, and Hiesey [4] established the use of reciprocal transplant and common garden experiments to investigate the role of habitat in driving population divergence. Even earlier, forest tree biologists were using provenance tests to identify phenotypic differences among trees from different geographic or climatic regions (reviewed in [5]). Field studies are powerful for identifying locally adapted traits, identifying the ecological forces that drive selection, and predicting short-term response to selection. Organismal perspectives are also necessary for interpreting results from population genetic analyses in an ecologically meaningful context. These approaches are limited, however, in that they provide no direct insight into evolutionary processes at the molecular level and because they reflect selection over fairly short periods of time that might not be representative of historical conditions.

Population genetic approaches that explore adaptation based on sampling DNA sequences from multiple individuals offer a temporal and genetic perspective that complement organism-based approaches. Moreover, because population genetic analyses are not constrained by logistical difficulties of caring for, growing, or handling live organisms, they can be used to investigate local adaptation when reciprocal transplant, common garden, or phenotypic selection analyses are not feasible due to logistical (e.g., organism size, lifespan) or ethical (e.g., humans) concerns. We focus here on recent empirical population genetic studies that have furthered our understanding of local adaptation. We first discuss some basic questions of local adaptation and then review important challenges and limitations of population genetic approaches to studying local adaptation. For discussion of other topics related to the ecological genetics of local adaptation including theory, field experiments, and statistical tests we refer readers to an excellent review by Savolainen *et al.* [2].

What traits are locally adapted?

Until recently, most population genetic analyses of local adaptation focused on candidate genes chosen because of

Glossary

Clinal adaptation: a form of local adaptation in which the adaptive phenotype changes gradually across an environmental or geographic gradient.
Isolation by distance (IBD): a negative relationship between the genetic similarity of individuals and geographic distance.
Isolation by environment (IBE): a positive relationship between the genetic similarity of individuals and the similarity of the environments in which populations are found. IBE can be caused by selection or spatial autocorrelation.
Local adaptation: adaptation in response to selection that varies geographically.
Reference genome: a genome assembly used as a reference for a species and for aligning sequencing reads for population-genomic studies. Depending on the species the reference can be based on a single individual or on a collection of individuals. Reference genomes do not capture the full extent of nucleotide or structural variation segregating within a species.

Corresponding author: Tiffin, P. (ptiffin@umn.edu).
Keywords: clinal adaptation; selection; genomics..

0169-5347/
 http://dx.doi.org/10.1016/j.tree.2014.10.004

their putative phenotypic effects (e.g., coat color in mice [6,7], flowering time in annual plants [8], immunity in plants and animals [9,10], and high-altitude adaptation [11]). These studies have investigated traits already thought to have been targets of local adaptation. Therefore the value of these studies has been in providing confirmatory evidence of adaptation, elucidating the molecular mechanisms of adaptation, and identifying which of the many genes that can affect a phenotype in the laboratory are responsible for local adaptation.

As genomic data have become more available, genome scans of local adaptation have become more commonplace than candidate gene studies. One promise of genomic scans is their potential to discover genes that have been subject to local adaptation without identifying loci of interest *a priori*. Once genes subject to selection have been identified, the phenotypes upon which selection has acted can potentially be inferred on the basis of gene function (a bottom-up [12] or reverse-ecology [13] approach). For example, a transcriptome scan for local adaptation in *Neurospora crassa* identified not only several genes affecting temperature-dependent growth but also a gene involved in circadian oscillation, suggesting a role for circadian cycles in latitude-related adaptation in this species [14]. Genome scans also can lead to a refinement of the phenotype responsible for local adaptation. Although water availability is clearly a strong selective force acting on plants, the potential adaptive responses to this selection are many and complex. Recent genome-wide scans in *Medicago truncatula* [15] and *Arabidopsis thaliana* [16] have improved our understanding of traits that might be responsible for adaptation to water availability; both find evidence of selection along precipitation gradients for genes affecting stomatal closure or photosynthetic capacity relating to the proportion of time stomata are open.

The promise of reverse-ecology approaches to identify selectively important traits is limited by knowledge of gene function. Such information is generally restricted to coding regions and is dependent upon annotation information derived from mutational screens of model species in laboratory environments. However, mutational screens in controlled environments will miss genes with phenotypic effects that differ between controlled and natural environments (genotype by environment effects) or have minor effects on phenotypes. Moreover, annotation in model species might be of limited use for distantly related species, and genes underlying variation in phenotypes that have not been the subject of functional genetic analyses will be missed or misannotated. These factors can limit the utility of reverse-ecology approaches by focusing the results of genome scans on well-studied phenotypes in species closely related to genetic models. Limited information on gene function might also lead to overinterpretation because it is often easy for biologists to find biologically interesting genes that can be interpreted in the light of known selective pressures [17]. Incomplete knowledge of gene function thus serves to unjustifiably reinforce preconceived ideas of the traits and selective forces driving local adaptation [18].

The vast majority of both candidate gene and genomic scan studies have relied on analyses that treat each locus independently. However, most ecologically important traits are quantitative, with phenotypes being determined by many loci, perhaps even hundreds or thousands [19]. The molecular evidence of selection acting on quantitative traits is expected to be weak because the signal of selection is distributed across many loci [20–22]. Therefore, the signal of selection acting on quantitative traits might not be revealed via standard genome scans. Approaches that investigate the signal of selection aggregated across loci, however, show promise in identifying selection on quantitative traits using genomic data. Turchin *et al.* [23], for example, show that frequency differences at SNPs associated with variation in human height are suggestive of selection across Europe. Recent theoretical work by Berg and Coop [24] provides a general approach for detecting selection on quantitative traits using marker data that could be applied to numerous species for which common garden studies are not feasible.

What environmental variables are the most important in structuring population differences?

Population genetic approaches not only promise to help to identify locally adapted traits, but they also can be used to identify the ecological variables most important in driving adaptation. One way to achieve this goal is shown by Fumagalli *et al.* [25] who linked annotated gene functions to the strength of gene–environment correlations to identify pathogens as a major driver of local adaptation in humans. Taking advantage of gene-level sequence data and detailed functional information, Fumagalli *et al.* identified ~100 genes with unexpectedly strong correlations to pathogen environment, but none that were strongly correlated with climate or diet. In addition to requiring dense gene-level data, a potential limitation of this approach for nonmodel species is that it also requires detailed functional annotations.

Population genetic approaches are also powerful for identifying the relative importance of geographic distance and different environmental variables in structuring populations, in other words, for asking whether spatial patterns of genetic diversity are structured more by geographic distance (isolation by distance, IBD) or the environment (isolation by environment, IBE). A recent meta-analysis of population genetic studies [26] revealed that both IBD and IBE are important in structuring population genetic diversity, but that across all studies IBE was more important. This generalization, however, might be tempered by bias in study design and publication: studies are more likely to be conducted on systems in which researchers expect local adaptation or IBE to be important, and researchers might be more likely to publish studies in which IBE is detected. Finally, it can be difficult to uncouple IBE from IBD when environmental variables covary with geographic distance [26].

Recently developed statistical frameworks, including a Bayesian model [27], redundancy analyses [28], and structural equation modeling [29], provide formal means to move beyond simply asking whether IBE is statistically significant and ask more interesting questions such as the relative contributions of IBD and IBE or compare IBE among different environmental factors. Applications of these approaches should allow researchers not only to

identify the major environmental factors structuring population differentiation but also to relate the relative strengths of IBD and IBE to species characteristics and across different spatial scales [30]. Because these approaches do not require reference genomes or knowledge of gene function they can be readily applied to non-model species, as evidenced by recent work comparing the relative strength of IBD to various topological dispersal barriers in the Neotropical bird *Xenops minutes* [31].

Is local adaptation convergent at the molecular level?

As the number of population genetic studies identifying genes responsible for adaptation has grown, there is increasing interest in determining whether independent bouts of adaptation are achieved via the same molecular targets. Finding evidence for parallel or convergent adaptation provides strong evidence for the adaptive importance of a gene, informs our understanding of constraints to adaptive evolution, and can improve our ability to predict adaptive responses to selection [32,33]. Population genetic approaches have revealed several examples of convergent local adaptation at the molecular level: within the threespine stickleback (*Gaterosteus aculeatus*) the *EBT* locus has been involved in repeated adaptation to freshwater habitats [34], two species of European spruce harbor adaptive latitudinal clines at the *FTL2* and *GI* genes [35], and the *EPAS1* and *HBB* genes show patterns of high-altitude adaptation in both dogs [36] and Tibetan humans [37].

Not surprisingly, there are far more examples of adaptation occurring via different genes. Even the systems cited in the paragraph above provide many examples in which loci are putatively adaptive in only one population or species: in sticklebacks Deagle *et al.* [38] report nine instances of repeated adaptation of stream–lake differences but also 64 loci targeted in a single watershed; Chen *et al.* [35,39] identified 18 genes harboring SNPs with signatures of latitudinal adaptation in the two spruce species, but only two genes were detected in both species, and Wang *et al.* [36] identified 14 genes with strong signatures of high-altitude adaptation in dogs, but only two were targeted in both dogs and humans. Moreover, the *EPAS1* locus does not appear to have been the target of selection in Andean populations [40]. The geographic distribution of genome-wide association studies (GWAS) candidates underlying local adaptation in *A. thaliana* is also indicative that local adaptation largely occurs through different genes in different parts of the species range [41].

In the end, it is difficult to know whether convergent evolution takes place more or less than expected by chance, both because we seldom know the number of possible paths by which these adaptations could evolve and because reports of parallel adaptation are given greater attention than cases in which adaptation has involved separate genes. The lack of rigorous null expectations for the extent of convergent adaptation suggests the need for further theoretical work (e.g., [42]) on the topic.

Do local adaptation loci exhibit trade-offs or conditional neutrality?

Local adaptation can result from alleles with environment-dependent fitness trade-offs (antagonistic pleiotropy) or alleles that confer a selective advantage in one environment but are neutral in others (conditional neutrality [43]). Of these two possibilities, only antagonistic pleiotropy is expected to lead to strong balancing selection and the maintenance of genetic variation. While antagonistic pleiotropy is found in some studies [44,45], field experiments indicate that conditional neutrality might be more common than antagonistic pleiotropy [46], although this conclusion needs to be viewed with some caution given that statistical issues make it easier to identify loci subject to conditional neutrality [45]. Population genetic analyses have also provided evidence of conditional neutrality; Fournier-Level *et al.* [47] found that alleles associated with lower fitness in common gardens had greater climate specificity than either control or beneficial alleles, leading to the conclusion that these alleles can be effectively neutral where they occur and deleterious elsewhere in the range.

Understanding the extent to which locally adapted loci exhibit conditional neutrality or antagonistic pleiotropy not only advances our understanding of genetic mechanisms but also is important for evaluating the power of statistical tests used to identify genes responsible for local adaptation. Despite empirical support for conditional neutrality, simulation studies evaluating the power of statistical approaches for identifying targets of local adaptation [48–50] have often assumed that adaptive loci exhibit antagonistic pleiotropy. These tests can have much less power to identify loci that are conditionally neutral than antagonistically pleiotropic (Box 1).

Challenge of obtaining high-quality data

The quality of the inferences made from population genetic analyses depend on the quality of the data. In the era of PCR amplification and Sanger sequencing obtaining high-quality data required considerable time and effort. With high-throughput sequencing, obtaining data is fast, but ensuring data quality is far more difficult. While genomes can be relatively easily sequenced, going from a collection of short sequence reads to accurately representing genomic variation segregating within and among populations is a challenge that has yet to be completely resolved.

The standard practice for making sense of sequence reads is to align them to a single reference genome or transcriptome. Because a single reference genome will not capture all the variation segregating within a species, alignment of reads to a reference genome will result in poor coverage and thus missing sequence information in highly diverged genomic regions including intergenic sequences, presence–absence variants (PAV, including transposable elements), and copy-number variants (CNV) [51]. Variants in these regions can play important roles in phenotypic variation and adaptation (e.g., [52–54]). Developing pan-genomes from *de novo* assemblies as references for aligning sequences will alleviate some limitations of aligning reads to a single reference (e.g., [55]), but the data and bioinformatic demands are such that widespread *de novo* assemblies will be limited to relatively few systems at least for the near future.

The challenges with genome-scale sequence data do not end with mapping reads. Once reads are mapped it is necessary to identify the segregating variants (i.e., SNPs,

Box 1. Gene action matters

Simulation studies evaluating the power of statistical methods to identify targets of local adaptation have assumed that alleles under selection exhibit antagonistic pleiotropy (AP) in which an allele that is beneficial in one part of the range is disadvantageous elsewhere in the range. However, empirical work shows that at least some alleles that contribute to local adaptation exhibit conditional neutrality (CN) in which an allele is acted upon by selection in one part of the range but is neutral elsewhere. Figure I shows results from simulations showing that the signal of adaptation from commonly used tests of local adaptation is considerably weaker for CN than AP alleles.

In the simulations, selection was imposed on a new mutation that entered one of 100 equal-sized populations that were evenly distributed across a range and connected through a stepping-stone model of migration. The left panel shows the geographic pattern of selection at a single biallelic locus; in the western part of the range an allele is favored by selection whereas in the eastern part of the range the allele is either neutral (CN) or selectively disadvantageous (AP). In both cases the total strength of selection acting across the range is equivalent. The right panels show results from selection analyses [mean and 95% confidence interval (CI) of the LOD (logarithm of the odds) score, or F_{ST} values for neutral (N), AP, and CN loci from 100 simulations). Association analyses were conducted using a pseudo-variable representing a selectively important environmental factor that ranged linearly from high to low values from west to east. Results are from sampling 100 individuals, 2000 generations after the start of selection. Simulations and analyses by J.B. Yoder.

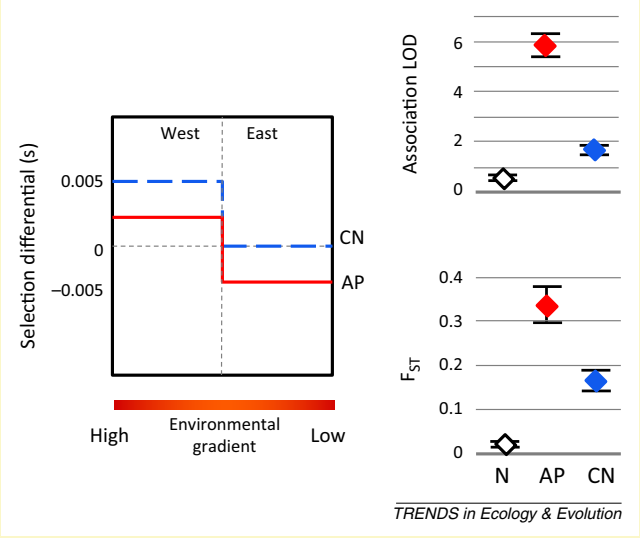

Figure I. Selection differentials defining antagonistically pleiotropic (AP) and conditionally neutral (CN) loci (left) and the signals of local adaptation at neutral (N), AP, and CN loci.

short indels) that form the basis for statistical analyses of local adaptation (reviewed in [2]). The standard approach for calling variants has been to align reads from each individual to a reference genome and then use computational tools (e.g., GATK [56] or SAMTools [57]) to identify the allelic state in each individual. However, owing to sequencing errors and misalignments, this call-based approach can lead to considerable error, especially in calling rare variants or when sequence coverage is $<10\times$ (i.e., an average of 10 reads aligned to each base of the reference) [58]. Coverage of $10\times$ is considerably higher than what has been used in many population genomic studies. Lower error rates are associated with direct estimation approaches, such as that implemented in ANGSD [59], which make sample-wide inferences of variant frequencies without calling genotypes for each sampled individual. An additional advantage of these approaches is that they can be easily applied to the sequencing of pools of individuals, and this can greatly reduce the costs of data collection.

The expense and complexity of handling full-genome sequence data has led to considerable interest in the use of reduced-representation data (GBS, RAD-tag, gene capture, SNP chips, RNA-seq), especially in non-model systems for which reference genomes are not available. Although reduced representation approaches are attractive for characterizing population structure, and useful for IBD versus IBE comparisons (e.g., [27]; see above) it is important to remember that they sample only a small portion of the genome and cannot be expected to identify a meaningful portion of the genes underlying local adaptation or phenotypic variation (Box 2). While reduced representation data might be powerful for identifying structural variants, such as large inversion polymorphisms associated with local adaptation [60,61], they are poorly suited for evaluating the relative importance of chromosomal rearrangements and single-nucleotide variants.

Challenge of identifying adaptive loci

Genome scans, as well as candidate-gene studies, can also be vexed by difficulty in identifying which genes have evolved in response to geographically variable selection. One approach for identifying selected loci is to look for those with extreme values of a statistic relative to expectations under a demographic model. However, the robustness of these inferences can be highly dependent on the assumed demographic model (Box 3) [50,62–64].

An alternative to modeling demographic history is to compare statistics calculated at loci of interest (candidate loci) to statistics found at non-candidate or reference loci. Under the assumption that reference loci provide an estimate of the expected distribution of the statistic in the absence of adaptation, then the distribution of the statistic calculated for reference loci can be used to estimate the probability of obtaining a statistical value at candidate genes (e.g., [24,50,65]). For example, candidate loci with fixation index F_{ST} values that exceed 99% of the F_{ST} values of randomly selected reference loci can be considered likely targets of selection (e.g., [66]). Of course, if only 1% of candidate loci exceed that threshold, then there is no reason to conclude that candidate loci bear different signatures of selection than the reference loci. Potential drawbacks of this approach, as well as of model-based approaches that make use of defined reference loci, are that they require *a priori* separation of reference and genes of interest and lose statistical power if reference loci harbor many genes that are actually involved in adaptation [50]. In addition, care is needed in the selection of reference loci – references should be drawn from portions of the genome with similar levels of recombination and linked

Box 2. Incomplete data produce an incomplete picture

A variety of approaches can be used to sample a large number of genetic markers distributed across a genome without requiring whole-genome sequencing (e.g., genotype by sequencing [73], RADseq [74], multiplexed genotype sequencing [75], and sequence capture). These technologies are powerful for collecting data for mapping genes in genetic crosses and characterizing population structure. However, because they sample only a small portion of the genome they have very limited power to identify loci responsible for phenotypic variation in association analyses or in identifying targets of local adaptation. To illustrate this limited power we calculated the expected probability of detecting recent hard-sweeps (in which selection drives a new mutation to fixation) as a function of SNP density, recombination rates, and the strength of selection (Figure I). The results show that, even for hard-sweeps (the easiest form of selection to detect), the great majority of targets of selection will be missed: only with fairly high SNP density (1 SNP per 5 kb), a large number of sweeps (100), and low to normal levels of recombination, is there a high probability of detecting even a single sweep. For studies to successfully identify targets of adaptation using reduced-representation approaches, a very high density of SNPs relative to the rate of decay in linkage disequilibrium will be necessary. Reduced-representation approaches were generally not designed to obtain SNPs at such high densities, although sequence capture can do so for some regions of the genome (e.g., all exons).

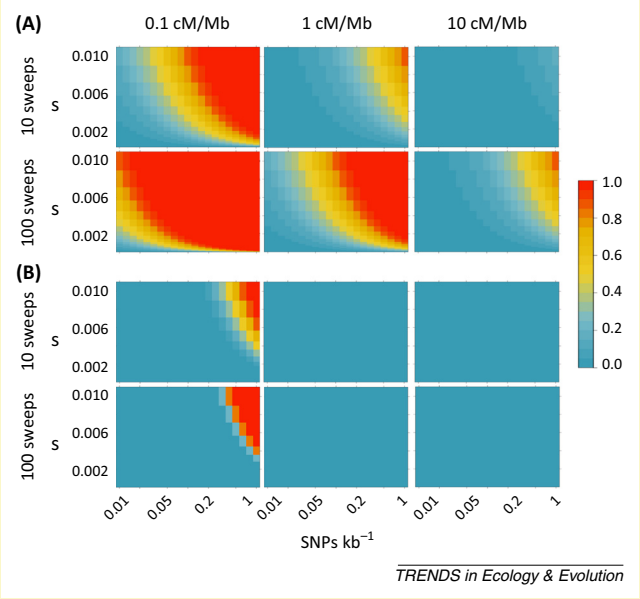

Figure I. Probability of detecting (A) $\geq$ 1 or (B) $>$ 50% of hard sweeps as a function of the number of sweeps, the strength of selection (s), and the number of SNPs kb-1 for three rates of recombination.

selection (e.g., intergenic regions might not provide a good set of reference loci for testing for evidence of selection on coding regions).

An alternative to using a model-derived or reference loci-based probability threshold to identify loci responsible for adaptation is to simply rank loci on the basis of *P* values or posterior probabilities; loci with the most extreme values are those most likely to be targets of selection (e.g., [15,49]). Given the sample sizes of many local adaptation studies, and the fact that *P* values are dependent on sample sizes, the use of ranks might capture functionally important loci that would not be captured if formal probability-based rejection of a null model is required. The exclusion of these loci has the potential to lead to a problem of 'missing local adaptation' analogous to the association genetics problem of 'missing heritability' (although of course nothing is actually missing, it has only not been detected with the statistical tools applied). Ranking loci to identify candidates is clearly a heuristic approach: it not only requires researchers to arbitrarily decide the number of loci to consider as candidates, but also will identify candidates even when none of the sampled loci are

Box 3. How many sweeps in Sweden?

Two recent papers searching for evidence of local adaptation in Swedish populations of *Arabidopsis thaliana* highlight how our ability to identify local adaptation depends on our understanding of demographic history. In the first of these papers, Long *et al.* [76] used full-genome sequence data from 180 individuals to search for evidence of recent selective sweeps in northern (50 individuals) and southern (130 individuals) Sweden. Based on previous demographic analyses of European populations of *A. thaliana*, Long *et al.* treated the northern and southern populations separately and applied SweepFinder, a program designed to detect recent sweeps based on the site frequency spectrum [77], to each population. SweepFinder identified 22 strong signals of adaptation in the northern sample, but only a single signal in the southern sample. This difference is not expected to be due to statistical power because more than twice as many individuals were sampled from the south than from the north.

Huber *et al.* [78] used data from Long *et al.* and extensive simulations implemented in 'diffusion approximations for demographic inference' (δaδi), which infers demography based on a diffusion approximation to the site frequency spectrum [79], to ask whether the observed asymmetry in the number of sweeps might result from a complex demographic history. They find that the data are most consistent with a secondary contact model in which northern and southern populations separated more than 100 kyr ago and then within the past 40 kyr started exchanging migrants again. Once migration started, gene flow was asymmetric with more gene flow from the north to the south. Simulations showed that the critical value of the composite likelihood ratio (CLR) test statistic used in SweepFinder is strongly skewed under a model of secondary contact. Using simulations to determine critical values, Huber *et al.* identified six loci in the north and 10 loci in the south that have experienced local sweeps – one third less and 10-fold more than that found using a naïve demographic model.

Huber *et al.* also estimated that the strength of selection was much greater (>fourfold) in the north than in the south. Potential explanations include the northern population being farther from its selective optimum owing to its relatively recent arrival, or the existence of more heterogeneous environment in the south. Based on a higher ratio of non-synonymous to synonymous site diversity in the southern versus northern populations (despite a larger effective population size in the south) the authors conclude that the southern population was drawn from a more heterogeneous environment. These results highlight that care must be taken when assigning individuals to populations when conducting tests of local adaptation: geography-based assignment might not validly capture the selective environments, and ignorance of demography can greatly skew estimates of the loci under selection and the strength of that selection.

responsible for adaptation. The potential for a sample to contain no targets of selection can be particularly problematic when reduced-representation approaches are used to survey genomic diversity (Box 2) or if populations have not diverged in response to differential local selection. Nevertheless, if genome scans are viewed as exploratory analyses for identifying candidates for subsequent analyses, then simple ranking has merits.

Regardless of the statistical approach used to identify candidates, independent validation of functional importance is an important step in elevating the status of genes from candidates to causative. For candidate gene studies, functional information is generally central to the selection of candidates. For genome scans, however, this is not the case. Support for the importance of a candidate can be obtained from functional analyses (e.g., [67,68]) or biparental QTL mapping (e.g., [69]). Identifying the same gene in independent occurrences of local adaptation is also strong evidence of a functionally important role [32] but, because of the potential for independent paths of adaptation, the repeatability of a candidate should not be viewed as a necessary step to establishing importance. An alternative to focusing on individual loci is to test whether genes identified by tests of local adaptation, as a group, predict performance in an independent sample. This approach was used by Fournier-Level *et al.* [47] who tested for correlations between the frequency of putatively adaptive alleles and the geographic distance from the experimental location at which those alleles were identified. This approach has also been inverted to predict the performance of genotypes on the basis of candidate genes identified through climate associations: Hancock *et al.* [16] applied this to predict performance in a common garden and Yoder *et al.* [15] found a positive correlation between predicted performance and the growth rate of accessions not used to identify adaptation candidates.

Limits in identifying the genetic basis of local adaptation

Many of the important questions in local adaptation being pursued with population genetics approaches begin – rather than end – with identifying loci responsible for variation. It is therefore important to realize that a full accounting of local adaptation at the molecular level goes beyond having high-quality data to analyze and statistical methods to identify causative genes. The crux of the challenge is that most ecologically important traits responsible for local adaptation are quantitative, and identifying all of the genes responsible for variation in quantitative traits is likely not possible. Even the cumulative explanatory power of individual loci identified in human genotype–phenotype association studies, which often involve tens of thousands of individuals, is generally only a small percentage of the phenotypic variation [70].

Although some association analyses of traits implicated in local adaptation have been reported to explain a high proportion of phenotypic variance, these studies have generally used the same data to identify causative loci and predict the proportion of variance explained by those loci. The potential inflationary bias of using the same data for both gene identification and phenotypic prediction is illustrated by Stanton-Geddes *et al.* [71]. In their study, when the same data were used to identify causative variants with association approaches and estimate the amount of phenotypic variation explained by those SNPs, on average more than 60% of the phenotypic variance could be explained even when phenotypes were randomly assigned to genotypes and no causation was present.

Concluding remarks

Population genetic analyses of local adaptation have come a long way since Lewontin and Krakauer [72] first used F_{ST} to investigate local adaptation at the molecular level. With high-throughput sequencing now commonplace, and greater interest in using population genetics to understand geographically variable selection, empirical population genetic analyses now promise to greatly advance our understanding of local adaptation. To do this it will be important for researchers to move beyond simply searching for genes – after all, we already know that genetic variation contributes to variation in many ecologically relevant phenotypes. Rather, the challenge is to use population genetic data to advance our understanding of local adaptation as a process.

In this review we have focused on some of the local adaptation questions that have been addressed with population genetic analyses. Population genetic approaches are certainly well suited for informing other aspects of the patterns and process of local adaptation, including the genetic architecture of local adaptation and the strength and stability of selection. Improved understanding of local adaptation as a process is also likely to require advances in theory that provide empiricists with testable predictions, and analytical tools that incorporate quantitative traits, non-equilibrium conditions, and simultaneous estimation of gene-flow and selection (topics not covered in this review). Finally, a stronger link between population genetic analyses and organism-based studies is likely a fruitful direction. Not only can organismal approaches be used to validate the putative importance of genes identified through genome scans, but organism-based follow-up analyses of population genetic studies serve to inform questions related to evolutionary process.

Acknowledgments

We thank Amanda Gorton, members of the laboratory of J.R-I., Gideon Bradburd, and Jesse Lasky for comments and discussion, Jeremy Yoder for the analyses presented in Box 1, two anonymous reviewers for insightful and constructive comments that improved the manuscript, and National Science Foundation awards (IOS-1238014 and 1237993) for financial support.

References

1 Hereford, J. (2009) A quantitative survey of local adaptation and fitness trade-offs. *Am. Nat.* 173, 579–588
2 Savolainen, O. *et al.* (2013) Ecological genomics of local adaptation. *Nat. Rev. Genet.* 14, 807–820
3 Turresson, G. (1922) The genotypical response of the plant species to the habitat. *Hereditas* 3, 211–350
4 Clausen, J.K. *et al.* (1940) *Experimental Studies on the Nature of Species. I. Effects of Varied Environments on Western North American Plants*, Carnegie Institute of Washington
5 Langlet, O. (1971) Two hundred years of genecology. *Taxon* 20, 653–722
6 Nachman, M. *et al.* (2003) The genetic basis of adaptive melanism in pocket mice. *Proc. Natl. Acad. Sci. U.S.A.* 100, 5268–5273

7 Linnen, C.R. *et al.* (2009) On the origin and spread of an adaptive allele in deer mice. *Science* 325, 1095–1098
8 Le Corre, V. *et al.* (2002) DNA polymorphism at the FRIGIDA gene in *Arabidopsis thaliana*: extensive nonsynonymous variation is consistent with local selection for flowering time. *Mol. Biol. Evol.* 19, 1261–1271
9 Moeller, D.A. and Tiffin, P. (2008) Geographic variation in adaptation at the molecular level: a case study of plant immunity genes. *Evolution* 62, 3069–3081
10 Prugnolle, F. *et al.* (2005) Pathogen-driven selection and worldwide HLA class I diversity. *Curr. Biol.* 15, 1022–1027
11 Storz, J.F. *et al.* (2007) The molecular basis of high-altitude adaptation in deer mice. *PLoS Genet.* 3, 448–459
12 Ross-Ibarra, J. *et al.* (2007) Plant domestication, a unique opportunity to identify the genetic basis of adaptation. *Proc. Natl. Acad. Sci. U.S.A.* 104, 8641–8648
13 Li, Y.F. *et al.* (2008) 'Reverse ecology' and the power of population genomics. *Evolution* 62, 2984–2994
14 Ellison, C.E. *et al.* (2011) Population genomics and local adaptation in wild isolates of a model microbial eukaryote. *Proc. Natl. Acad. Sci. U.S.A.* 108, 2831–2836
15 Yoder, J.B. *et al.* (2014) Genomic signature of adaptation to climate in *Medicago truncatula*. *Genetics* 196, 1263–1275
16 Hancock, A.M. *et al.* (2011) Adaptation to climate across the *Arabidopsis thaliana* genome. *Science* 334, 83–86
17 Pavlidis, P. *et al.* (2012) A critical assessment of storytelling: gene ontology categories and the importance of validating genomic scans. *Mol. Biol. Evol.* 29, 3237–3248
18 Barrett, R.D.H. and Hoekstra, H.E. (2011) Molecular spandrels: tests of adaptation at the genetic level. *Nat. Rev. Genet.* 12, 767–780
19 Rockman, M.V. (2012) The QTN program and the alleles that matter for evolution: all that's gold does not glitter. *Evolution* 66, 1–17
20 Kelly, J. (2006) Geographical variation in selection, from phenotypes to molecules. *Am. Nat.* 167, 481–495
21 Kemper, K.E. *et al.* (2014) Selection for complex traits leaves little or no classic signatures of selection. *BMC Genomics* 15, 246
22 Le Corre, V. and Kremer, A. (2012) The genetic differentiation at quantitative trait loci under local adaptation. *Mol. Ecol.* 21, 1548–1566
23 Turchin, M.C. *et al.* (2012) Evidence of widespread selection on standing variation in Europe at height-associated SNPs. *Nat. Genet.* 44, 1015–1019
24 Berg, J. and Coop, G. (2014) A population genetic signal of polygenic adaptation. *PLoS Genet.* 10, e1004412
25 Fumagalli, M. *et al.* (2011) Signatures of environmental genetic adaptation pinpoint pathogens as the main selective pressure through human evolution. *Plos Genet.* 7, e1002355
26 Sexton, J.P. *et al.* (2014) Genetic isolation by environment or distance: which pattern of gene flow is most common? *Evolution* 68, 1–15
27 Bradburd, G.S. *et al.* (2013) Disentangling the effects of geographic and ecological isolation on genetic differentiation. *Evolution* 67, 3258–3273
28 Lasky, J.R. *et al.* (2012) Characterizing genomic variation of *Arabidopsis thaliana*: the roles of geography and climate. *Mol. Ecol.* 21, 5512–5529
29 Wang, I.J. *et al.* (2013) Quantifying the roles of ecology and geography in spatial genetic divergence. *Ecol. Lett.* 16, 175–182
30 Richardson, J.L. *et al.* (2014) Microgeographic adaptation and the spatial scale of evolution. *Trends Ecol. Evol.* 29, 165–176
31 Harvey, M.G. and Brumfield, R.T. (2014) Genomic variation in a widespread Neotropical bird (*Xenops minutus*) reveals divergence, population expansion, and gene flow. *Ar-Xiv.Org* ar-Xiv:1405.6571v2.
32 Conte, G.L. *et al.* (2012) The probability of genetic parallelism and convergence in natural populations. *Proc. R. Soc. B: Biol. Sci.* 279, 5039–5047
33 Streisfeld, M.A. and Rausher, M.D. (2011) Population genetics, pleiotropy, and the preferential fixation of mutations during adaptive evolution. *Evolution* 65, 629–642
34 Jones, F.C. *et al.* (2012) A genome-wide SNP genotyping array reveals patterns of global and repeated species-pair divergence in sticklebacks. *Curr. Biol.* 22, 83–90
35 Chen, J. *et al.* (2014) Clinal variation at phenology-related genes in spruce: parallel evolution in *FTL2* and *Gigantea*? *Genetics* 197, 1025–1038
36 Wang, G. *et al.* (2014) Genetic convergence in the adaptation of dogs and humans to the high altitude environment of the Tibetan plateau. *Genome Biol. Evol.* 56, 2122–2128
37 Yi, X. *et al.* (2010) Sequencing of 50 human exomes reveals adaptation to high altitude. *Science* 329, 75–78
38 Deagle, B.E. *et al.* (2012) Population genomics of parallel phenotypic evolution in stickleback across stream–lake ecological transitions. *Proc. R. Soc. B: Biol. Sci.* 279, 1277–1286
39 Chen, J. *et al.* (2012) Disentangling the roles of history and local selection in shaping clinal variation of allele frequencies and gene expression in Norway spruce (*Picea abies*). *Genetics* 191, 865–881
40 Bigham, A. *et al.* (2010) Identifying signatures of natural selection in Tibetan and Andean populations using dense genome scan data. *Plos Genet.* 6, e1001116
41 Fournier-Level, A. *et al.* (2013) Paths to selection on life history loci in different natural environments across the native range of *Arabidopsis thaliana*. *Mol. Ecol.* 22, 3552–3566
42 Ralph, P.L. and Coop, G. (2014) Convergent evolution during local adaptation to patchy landscapes. *BioRxiv* http://dx.doi.org/10.1101/006940
43 Schnee, F. and Thompson, J. (1984) Conditional neutrality of polygene effects. *Evolution* 38, 42–46
44 Agren, J. *et al.* (2013) Genetic mapping of adaptation reveals fitness tradeoffs in *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. U.S.A.* 110, 21077–21082
45 Anderson, J.T. *et al.* (2013) Genetic trade-offs and conditional neutrality contribute to local adaptation. *Mol. Ecol.* 22, 699–708
46 Anderson, J.T. *et al.* (2011) Evolutionary genetics of plant adaptation. *Trends Genet.* 27, 258–266
47 Fournier-Level, A. *et al.* (2011) A map of local adaptation in *Arabidopsis thaliana*. *Science* 334, 86–89
48 De Mita, S. *et al.* (2013) Detecting selection along environmental gradients: analysis of eight methods and their effectiveness for outbreeding and selfing populations. *Mol. Ecol.* 22, 1383–1399
49 Jones, M.R. *et al.* (2013) Integrating landscape genomics and spatially explicit approaches to detect loci under selection in clinal populations. *Evolution* 67, 3455–3468
50 Lotterhos, K.E. and Whitlock, M.C. (2014) Evaluation of demographic history and neutral parameterization on the performance of F-ST outlier tests. *Mol. Ecol.* 23, 2178–2192
51 Sims, D. *et al.* (2014) Sequencing depth and coverage: key considerations in genomic analyses. *Nat. Rev. Genet.* 15, 121–132
52 Fischer, I. *et al.* (2011) Adaptation to drought in two wild tomato species: the evolution of the *Asr* gene family. *New Phytol.* 190, 1032–1044
53 Hanikenne, M. *et al.* (2013) Hard selective sweep and ectopic gene conversion in a gene cluster affording environmental adaptation. *PLoS Genet.* 9, e1003707
54 Demuth, J.P. and Hahn, M.W. (2009) The life and death of gene families. *Bioessays* 31, 29–39
55 Gan, X. *et al.* (2011) Multiple reference genomes and transcriptomes for *Arabidopsis thaliana*. *Nature* 477, 419–423
56 McKenna, A. *et al.* (2010) The genome analysis toolkit: a mapreduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 20, 1297–1303
57 Li, H. *et al.* (2009) The sequence alignment/map format and SAMtools. *Bioinformatics* 25, 2078–2079
58 Han, E. *et al.* (2014) Characterizing bias in population genetic inferences from low-coverage sequencing data. *Mol. Biol. Evol.* 31, 723–735
59 Nielsen, R. *et al.* (2012) SNP calling, genotype calling, and sample allele frequency estimation from new-generation sequencing data. *PLoS ONE* 7, e37558
60 Pyhajarvi, T. *et al.* (2013) Complex patterns of local adaptation in teosinte. *Genome Biol. Evol.* 5, 1594–1609
61 Hoffmann, A. *et al.* (2004) Chromosomal inversion polymorphisms and adaptation. *Trends Ecol. Evol.* 19, 482–488
62 Excoffier, L. *et al.* (2009) Detecting loci under selection in a hierarchically structured population. *Heredity* 103, 285–298
63 Poh, Y. *et al.* (2014) On the prospect of identifying adaptive loci in recently bottlenecked populations. *Biorxiv* http://dx.doi.org/10.1101/009456

64 de Villemereuil, P. *et al.* (2014) Genome scan methods against more complex models: when and how much should we trust them? *Mol. Ecol.* 23, 2006–2019
65 Wright, S. and Charlesworth, B. (2004) The HKA test revisited: a maximum-likelihood-ratio test of the standard neutral model. *Genetics* 168, 1071–1076
66 Keller, S.R. *et al.* (2012) Local adaptation in the flowering-time gene network of balsam poplar, *Populus balsamifera* L. *Mol. Biol. Evol.* 29, 3143–3152
67 Carneiro, M. *et al.* (2014) Rabbit genome analysis reveals a polygenic basis for phenotypic change during domestication. *Science* 345, 1074–1079
68 Prasad, K.V.S.K. *et al.* (2012) A gain-of-function polymorphism controlling complex traits and fitness in nature. *Science* 337, 1081–1084
69 Huang, X. *et al.* (2012) A map of rice genome variation reveals the origin of cultivated rice. *Nature* 490, 497–501
70 Visscher, P.M. *et al.* (2012) Five years of GWAS discovery. *Am. J. Hum. Genet.* 90, 7–24
71 Stanton-Geddes, J. *et al.* (2013) Candidate genes and genetic architecture of symbiotic and agronomic traits revealed by whole-genome, sequence-based association genetics in *Medicago truncatula*. *PLoS ONE* 8, e65688
72 Lewontin, R. and Krakauer, J. (1973) Distribution of gene frequency as a test of the theory of the selective neutrality of polymorphisms. *Genetics* 74, 175–195
73 Elshire, R.J. *et al.* (2011) A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS ONE* 6, e19379
74 Baird, N.A. *et al.* (2008) Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS ONE* 3, e3376
75 Andolfatto, P. *et al.* (2011) Multiplexed shotgun genotyping for rapid and efficient genetic mapping. *Genome Res.* 21, 610–617
76 Long, Q. *et al.* (2013) Massive genomic variation and strong selection in *Arabidopsis thaliana* lines from Sweden. *Nat. Genet.* 45, 884–890
77 Nielsen, R. *et al.* (2005) Genomic scans for selective sweeps using SNP data. *Genome Res.* 15, 1566–1575
78 Huber, C.D. *et al.* (2014) Keeping it local: evidence for positive selection in Swedish *Arabidopsis thaliana*. *Mol. Biol. Evol.* http://dx.doi.org/10.1093/molbev/msu247
79 Gutenkunst, R.N. *et al.* (2009) Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *PLoS Genet.* 5, e1000695