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.

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