Perspectives on the use of landscape genetics to detect genetic adaptive variation in the field

STÉPHANIE MANEL,*† STÉPHANE JOOST,‡ BRYAN K. EPPERSON,§ ROLF HOLDEREGGER,* ANDREW STORFER,** MICHAEL S. ROSENBERG,†† KIM T. SCRIBNER,‡‡ AURELIE BONIN §§ and MARIE-JOSÉE FORTIN ––

*Laboratoire Population Environnement Développement, UMR 151 UP/IRD, Université de Provence, 3 place Victor Hugo, 13331 Marseille Cedex 03, France, †Laboratoire d’Ecologie Alpine, UMR-CNRS 5553, Université Joseph Fourier, BP53 38041 Grenoble Cedex 9, France, ‡Laboratoire de Systèmes d’Information Géographique (LASIG), Ecole Polytechnique Fédérale de Lausanne (EPFL), Bâtiment GC, Station 18, 1015 Lausanne, Switzerland, §Department of Forestry, Michigan State University, East Lansing, MI 48824, USA, ¶ Department of Fisheries & Wildlife and Dept. of Zoology, Michigan State University, East Lansing, MI 48824, §§Department of Botany University of British Columbia 3529-6270 University Boulevard Vancouver, BC, V6T 1Z4, Canada, ––Department of & Evolutionary Biology, University of Toronto, Toronto, ON, Canada M5S 3G5

Abstract

Understanding the genetic basis of species adaptation in the context of global change poses one of the greatest challenges of this century. Although we have begun to understand the molecular basis of adaptation in those species for which whole genome sequences are available, the molecular basis of adaptation is still poorly understood for most non-model species. In this paper, we outline major challenges and future research directions for correlating environmental factors with molecular markers to identify adaptive genetic variation, and point to research gaps in the application of landscape genetics to real-world problems arising from global change, such as the ability of organisms to adapt over rapid time scales. High throughput sequencing generates vast quantities of molecular data to address the challenge of studying adaptive genetic variation in non-model species. Here, we suggest that improvements in the sampling design should consider spatial dependence among sampled individuals. Then, we describe available statistical approaches for integrating spatial dependence into landscape analyses of adaptive genetic variation.

Keywords: computational approach, genome scan, local adaptation, landscape genomics, molecular techniques, regression analysis

Introduction

Can species adapt to global change? Environmental change at all spatial scales is rapidly altering selection regimes for global flora and fauna (Reusch & Wood 2007). One of the most challenging questions of our time is whether adaptive evolution can keep pace with the rate and direction of selection that is imposed by humans (Hendry et al. 2008). It is critical to assess how genetic diversity may change and at what cost to the maintainance of population viability in the long-term (Lynch & Lande 1993). Adaptive genetic diversity dictates narrower tolerance limits to changing environmental conditions for specific populations than for a species as a whole (Etterson 2008). It is predicted that many species are able to shift their geographic ranges to track global change (Parmesan 2001), but the general potential of species to adapt to rapid change is still debated (Davis et al. 2005; Reusch & Wood 2007).
Local adaptation results from the balance between gene flow and many natural selection factors, including the climate (Savolainen et al. 2007; Hoffmann & Willi 2008). Adaptive differentiation among populations within a species has been documented through the study of clinal variation in physiological, phenological, and fitness traits in relation to gradients in climate (Davis et al. 2005; Gienapp et al. 2008). Broadly speaking, however, we are only just beginning to understand the genomic basis of phenotypic traits associated with local adaptation for species whose whole genomes have been sequenced (e.g. Begun et al. 2007; Turner et al. 2008). Additionally, for a limited number of species such as forest trees, studies have been able to build upon a long history of common garden experimentation (Bradshaw et al. 1995; Neale & Savolainen 2004; Neale 2007; Neale & Ingvarsson 2008; Gratapaglia et al. 2009), allowing the characterization of the geographic pattern of neutral and adaptive genetic variation in relation to geography and climate (Savolainen et al. 2007; Aitken et al. 2008; Eckert et al. 2009a,b). This knowledge may aid the understanding of these species’ responses to rapid climate change in the future (Sork et al. 2010).

Nonetheless, recent and upcoming advances in high throughput DNA sequencing leads to ever increasing availability of genomic sequences, facilitating an enhanced understanding of the genetic basis of current and future adaptation in a broad variety of species (Segelbacher et al. 2010). As a result, the limiting factor in future studies will no longer be the molecular laboratory workload, but rather the development of statistical, bioinformatics and modelling tools for identifying both genes or gene networks under selection (McCarthy et al. 2008), as well as the environmental factors acting as selective pressures. We refer to such a framework for understanding the spatial distribution of adaptive genetic variation using genomic tools (Box 1) as landscape genomics (Joost et al. 2007), and in Box 2 we clarify the use of this and other related terms.

Our focus is on studies and methods that will assess spatial correlations of particular molecular markers with environmental variables (Hamilton et al. 2002; Manel et al. 2009, 2010; Schwartz et al. 2010; Poncet et al. 2010). The pattern of genetic variation observed in such loci along environmental gradients has usually been interpreted as being caused by natural selection (Endler 1986; Schmidt et al. 2008). Further, we focus on the effects of evolutionary processes that can operate over much smaller spatial and temporal scales than those typically employed in phylogeographic studies (Manel et al. 2003). However, it is important to point out that larger time scale effects such as selective sweeps, a form of genetic hitchhiking where neutral gene flow and many natural selection factors, including the climate (Savolainen et al. 2007; Hoffmann & Willi 2008). Adaptive differentiation among populations within a species has been documented through the study of clinal variation in physiological, phenological, and fitness traits in relation to gradients in climate (Davis et al. 2005; Gienapp et al. 2008). Broadly speaking, however, we are only just beginning to understand the genomic basis of phenotypic traits associated with local adaptation for species whose whole genomes have been sequenced (e.g. Begun et al. 2007; Turner et al. 2008). Additionally, for a limited number of species such as forest trees, studies have been able to build upon a long history of common garden experimentation (Bradshaw et al. 1995; Neale & Savolainen 2004; Neale 2007; Neale & Ingvarsson 2008; Gratapaglia et al. 2009), allowing the characterization of the geographic pattern of neutral and adaptive genetic variation in relation to geography and climate (Savolainen et al. 2007; Aitken et al. 2008; Eckert et al. 2009a,b). This knowledge may aid the understanding of these species’ responses to rapid climate change in the future (Sork et al. 2010).

Nonetheless, recent and upcoming advances in high throughput DNA sequencing leads to ever increasing availability of genomic sequences, facilitating an enhanced understanding of the genetic basis of current and future adaptation in a broad variety of species (Segelbacher et al. 2010). As a result, the limiting factor in future studies will no longer be the molecular laboratory workload, but rather the development of statistical, bioinformatics and modelling tools for identifying both genes or gene networks under selection (McCarthy et al. 2008), as well as the environmental factors acting as selective pressures. We refer to such a framework for understanding the spatial distribution of adaptive genetic variation using genomic tools (Box 1) as landscape genomics (Joost et al. 2007), and in Box 2 we clarify the use of this and other related terms.

Our focus is on studies and methods that will assess spatial correlations of particular molecular markers with environmental variables (Hamilton et al. 2002; Manel et al. 2009, 2010; Schwartz et al. 2010; Poncet et al. 2010). The pattern of genetic variation observed in such loci along environmental gradients has usually been interpreted as being caused by natural selection (Endler 1986; Schmidt et al. 2008). Further, we focus on the effects of evolutionary processes that can operate over much smaller spatial and temporal scales than those typically employed in phylogeographic studies (Manel et al. 2003). However, it is important to point out that larger time scale effects such as selective sweeps, a form of genetic hitchhiking where neutral

Local adaptation results from the balance between gene flow and many natural selection factors, including the climate (Savolainen et al. 2007; Hoffmann & Willi 2008). Adaptive differentiation among populations within a species has been documented through the study of clinal variation in physiological, phenological, and fitness traits in relation to gradients in climate (Davis et al. 2005; Gienapp et al. 2008). Broadly speaking, however, we are only just beginning to understand the genomic basis of phenotypic traits associated with local adaptation for species whose whole genomes have been sequenced (e.g. Begun et al. 2007; Turner et al. 2008). Additionally, for a limited number of species such as forest trees, studies have been able to build upon a long history of common garden experimentation (Bradshaw et al. 1995; Neale & Savolainen 2004; Neale 2007; Neale & Ingvarsson 2008; Gratapaglia et al. 2009), allowing the characterization of the geographic pattern of neutral and adaptive genetic variation in relation to geography and climate (Savolainen et al. 2007; Aitken et al. 2008; Eckert et al. 2009a,b). This knowledge may aid the understanding of these species’ responses to rapid climate change in the future (Sork et al. 2010).

Nonetheless, recent and upcoming advances in high throughput DNA sequencing leads to ever increasing availability of genomic sequences, facilitating an enhanced understanding of the genetic basis of current and future adaptation in a broad variety of species (Segelbacher et al. 2010). As a result, the limiting factor in future studies will no longer be the molecular laboratory workload, but rather the development of statistical, bioinformatics and modelling tools for identifying both genes or gene networks under selection (McCarthy et al. 2008), as well as the environmental factors acting as selective pressures. We refer to such a framework for understanding the spatial distribution of adaptive genetic variation using genomic tools (Box 1) as landscape genomics (Joost et al. 2007), and in Box 2 we clarify the use of this and other related terms.

Our focus is on studies and methods that will assess spatial correlations of particular molecular markers with environmental variables (Hamilton et al. 2002; Manel et al. 2009, 2010; Schwartz et al. 2010; Poncet et al. 2010). The pattern of genetic variation observed in such loci along environmental gradients has usually been interpreted as being caused by natural selection (Endler 1986; Schmidt et al. 2008). Further, we focus on the effects of evolutionary processes that can operate over much smaller spatial and temporal scales than those typically employed in phylogeographic studies (Manel et al. 2003). However, it is important to point out that larger time scale effects such as selective sweeps, a form of genetic hitchhiking where neutral

Local adaptation results from the balance between gene flow and many natural selection factors, including the climate (Savolainen et al. 2007; Hoffmann & Willi 2008). Adaptive differentiation among populations within a species has been documented through the study of clinal variation in physiological, phenological, and fitness traits in relation to gradients in climate (Davis et al. 2005; Gienapp et al. 2008). Broadly speaking, however, we are only just beginning to understand the genomic basis of phenotypic traits associated with local adaptation for species whose whole genomes have been sequenced (e.g. Begun et al. 2007; Turner et al. 2008). Additionally, for a limited number of species such as forest trees, studies have been able to build upon a long history of common garden experimentation (Bradshaw et al. 1995; Neale & Savolainen 2004; Neale 2007; Neale & Ingvarsson 2008; Gratapaglia et al. 2009), allowing the characterization of the geographic pattern of neutral and adaptive genetic variation in relation to geography and climate (Savolainen et al. 2007; Aitken et al. 2008; Eckert et al. 2009a,b). This knowledge may aid the understanding of these species’ responses to rapid climate change in the future (Sork et al. 2010).

Nonetheless, recent and upcoming advances in high throughput DNA sequencing leads to ever increasing availability of genomic sequences, facilitating an enhanced understanding of the genetic basis of current and future adaptation in a broad variety of species (Segelbacher et al. 2010). As a result, the limiting factor in future studies will no longer be the molecular laboratory workload, but rather the development of statistical, bioinformatics and modelling tools for identifying both genes or gene networks under selection (McCarthy et al. 2008), as well as the environmental factors acting as selective pressures. We refer to such a framework for understanding the spatial distribution of adaptive genetic variation using genomic tools (Box 1) as landscape genomics (Joost et al. 2007), and in Box 2 we clarify the use of this and other related terms.

Our focus is on studies and methods that will assess spatial correlations of particular molecular markers with environmental variables (Hamilton et al. 2002; Manel et al. 2009, 2010; Schwartz et al. 2010; Poncet et al. 2010). The pattern of genetic variation observed in such loci along environmental gradients has usually been interpreted as being caused by natural selection (Endler 1986; Schmidt et al. 2008). Further, we focus on the effects of evolutionary processes that can operate over much smaller spatial and temporal scales than those typically employed in phylogeographic studies (Manel et al. 2003). However, it is important to point out that larger time scale effects such as selective sweeps, a form of genetic hitchhiking where neutral
SNPs (Single Nucleotide Polymorphisms)

SNPs are the most abundant type of polymorphism in genomes (Schlötterer 2004). For example, on average there is one SNP every Kb in the 3-billion-base human genome (Zhang & Hewitt 2003). They are usually biallelic and evolve according to a simple infinite sites mutation model (Schlötterer 2004). One of the major drawbacks of SNPs is their susceptibility to ascertainment bias, i.e. the bias introduced by using a subset of the studied individuals or populations for marker discovery purposes and which can lead to a skew in the distribution of allelic frequencies (Morin et al. 2004). Detecting SNPs also requires a priori information on the studied genome sequence (Morin et al. 2004), but once this task is completed, SNPs present a high potential for an automated high-throughput analysis at a moderate cost (Schlötterer 2004). The most impressive SNP datasets have long been restricted to model species: for instance, more than 10 000 SNPs were surveyed to examine the effects of differentiation and selection in the human (Akey et al. 2004) and mouse (Harr 2006) genomes. Fortunately, next-generation sequencing technologies are expected to give a substantial boost to the use of SNPs for both model and non-model organisms. For example, Turner et al. (2010) investigated the genetic basis of adaptation to serpentine soils in Arabidopsis lyrata using about 8 millions polymorphisms (mostly SNPs) identified in Solexa sequencing data. The markers showing the highest genetic differentiation between soil types were preferentially situated in genes involved in heavy metal detoxification and calcium/magnesium transport. These genes thus constitute good candidate for serpentine adaptation.

EST (Expressed Sequence Tag)-based molecular markers and other markers derived from next-generation sequencing data.

ESTs are short (~200–700 nucleotides) subsequences of transcribed and spliced DNA, generated by partially sequencing a pool of mRNAs (Bouck & Vision 2007). One of the most exciting prospects offered by next-generation sequencing technologies is the development of EST libraries for a wider range of species (Hudson 2008). These libraries can be astutely exploited to identify EST-based markers. These markers (classical microsatellites or SNPs) are usually located within a coding or a transcribed but untranslated region of a gene (Bouck & Vision 2007); but, they can also be assayed in non-transcribed sequences flanking genes by using a primer anchoring within the EST and another primer complementary to an adaptor-ligated restriction site (Bouck & Vision 2007). EST-based markers are thus tightly associated to gene-rich regions, which is particularly useful when searching for signatures of selection (Bonin 2008). Other types of promising marker systems building on next-generation sequencing data include the CRoPS (Complexity Reduction of Polymorphic Sequences; van Orsouw et al. 2007) and the RAD (Restriction-site associated DNA; Baird et al. 2008) methods. The practical and analytical shortcomings of all these new markers are nonetheless poorly understood. For example, the use of normalized EST libraries can theoretically bias the estimation of alleles frequencies at EST-based markers by favoring the sequencing of low-frequencies alleles. Similarly, the impact of sequencing errors on marker discovery remains to be explored.
Finally, we suggest statistical approaches for integrating spatial dependence in analyses of genomic data.

Molecular data in landscape genomics

The main goal of landscape genomics is to identify loci having adaptive significance in the genome by combining genomic and environmental data (Box 2) (Joost et al. 2007). Landscape genomics has the remarkable characteristic of not requiring phenotypic data on the adaptive trait(s) of interest, which can be laborious to collect especially for wild and/or endangered species. In that respect, it differs from other classical strategies aimed at unraveling the genetic basis of adaptation, such as Quantitative Trait Loci (QTL) analysis or association mapping or quantitative genetics studies (Stinchcombe & Hoekstra 2008).

A prerequisite of the landscape genomics approach is to survey many genetic loci (typically several hundred or more) scattered in the genome of many individuals in order to discover genomic regions under selection, either directly or more likely though physical linkage (Box 3) (Luikart et al. 2003; Storz 2005). Several genomic resources can advantageously be exploited to this end (Box 1). Yet until recently, the amplified fragment length polymorphism (AFLP) technique has often been the most efficient option in terms of effort and costs to screen the genome of non-model species (Luikart et al. 2003). Hundreds of AFLP markers spanning the whole genome can be obtained relatively easily for any organism, without a priori sequence knowledge (Meudt & Clarke 2007). However, it is very laborious to link markers showing a signature of selection with the actual gene or mutation under selection (Bonin 2008). Moreover, obtaining sufficient AFLP markers to adequately saturate the genome is difficult, especially in species where linkage disequilibrium decays rapidly (Bonin 2008). As a result, AFLP-based genome scans have largely failed to pinpoint potential adaptive gene(s) or mutation(s) (but see Wood et al. 2008; Manel et al. 2010; Poncet et al. 2010).

Soon such technical limitations will disappear, owing to recent advances of next-generation sequencing technologies and its increasing affordability (Box 1) (Hudson 2008). The phrase ‘next-generation sequencing’ refers to the series of recent technologies capable of producing up to millions of relatively short sequence reads

Box 2. Clarification of terms

A number of recent terms, including landscape genetics (Manel et al. 2003), landscape genomics (Luikart et al. 2003; Joost et al. 2007), molecular genecology (Hamilton et al. 2002; Skot et al. 2002), and ecological genomics (Ungerer et al. 2008), have recently been introduced to describe studies aimed at understanding the impact of the environment/landscape on genetic response. These are in fact not new research fields, but rather involve the interdisciplinary integration of multiple pre-existing research disciplines, including spatial statistics, landscape ecology, population genetics and molecular biology. These terms were initially introduced to facilitate the discussion of researchers across disciplines; however, the multiplication of similar terms has led to the need for clarification.

Landscape genetics (Manel et al. 2003) aims to provide information about the interaction between landscape features and microevolutionary processes, such as gene flow, genetic drift or selection. Most current applications of landscape genetics focus on gene flow and migration (processes that can either facilitate or constrain local adaptation), i.e. the effect of the environment on the selectively neutral component of genetic diversity (Storfer et al. 2007; Anderson et al. 2010). However, landscape genetics also aims to correlate allele frequencies with the environment in order to understand the effect of the environment on the adaptive component of genetic diversity (Hölderegger et al. 2006).

Landscape genomics (Luikart et al. 2003; Joost et al. 2007) uses correlation studies between the genomic data and the environment to identify genes either potentially linked to candidate genes or the genes themselves under selection. Landscape genomics is included in landscape genetics, but refers more specifically to the use of the future large amount of genetic data due to high-throughput sequencing. Landscape genomics is thus at the interface of bioinformatics, genomics, spatial statistics and landscape ecology.

Molecular genecology (Hamilton et al. 2002) is the study of geographical clines in the frequencies of alleles and their relationship to ecological clines in environmental conditions. Its objectives are largely the same as for the other research fields listed above.

Ecological genomics (Ungerer et al. 2008) integrates over several disciplines and seeks to understand the genetic mechanisms underlying responses of organisms to their natural environment. It is broader than landscape genetics and genomics, since it further includes experimental and laboratory approaches.
(35–1000 bases) at once thanks to the high parallelization of the sequencing process. Data throughput will continue to scale up in the near future with the ongoing development of real-time single-molecule sequencing technologies targeting longer reads (Hudson 2008). As a result, companies like VisiGen are aiming for a $1000 (human) genome, and we expect that studies of non-model organisms too will necessarily benefit from this ‘genomic revolution’. Currently, and for a wide range of species, it is financially feasible to sequence Expressed Sequence Tag (EST; see Box 1) libraries and develop EST-associated molecular markers (e.g. Vera et al. 2008). Interestingly, unlike AFLPs or ‘classical’ microsatellites or Single Nucleotide Polymorphisms (SNPs), these markers occur in gene-rich regions of the genome, i.e. those most likely to be under selection. EST-based genome scans have already been used to identify promising candidate genes for adaptation in various species such as white spruce (Namroud et al. 2008), salmon (Vasemagi et al. 2005) and seagrass (Oetjen & Reusch 2007). However, the use of high-throughput sequencing techniques in a landscape genomics context is still in its infancy (but see Eckert et al. 2009a; 2010; for recent applications). Additionally, increasing the number of analysed loci will inevitably raise concerns about linkage disequilibrium, as is discussed in Box 3 (Segelbacher et al. 2010). Furthermore, current landscape genomic studies, currently for the most part at the exploratory stage, need to move forward to the confirmatory stage of proving the adaptive significance of identified loci linked to genes under selection (Reusch & Wood 2007).

Environmental data

Landscape genomic studies either use environmental data collected in the field or take advantage of existing GIS databases. Recent increases in the availability of digital environmental data from remote sensors and weather stations have now made many global environmental data sets freely available (Box 4). In the absence of local environmental data, global environmental data can serve as valuable surrogates in landscape genomics studies. Yet, depending on the spatial (and temporal) scale of study question, detailed local measurements with high precision (e.g. spatial resolution ≤1 m²) may often be needed to understand local microevolutionary processes (Anderson et al. 2010). Micro-environmental data may be gathered using special sensor networks installed in the field, providing high-resolution eco-climatic data. The US National Ecological Observatory Network (NEON) will likely encompass sensor networks throughout the USA to gather long-term data on ecological responses to changes in land use and climate at a cost of $400–450 million USD (Keller et al. 2008). Such high-resolution data form the very foundation for future research to investigate the current local adaptation of organisms, which has been shaped by past selection.
Future environmental data acquisition for use in landscape genomics should: (1) use measures of environmental conditions within the home range of mobile organisms (Moorcroft & Lewis 2006); (2) complement coarser environmental data sets acquired over several decades (e.g. LANDSAT data) with local high resolution environmental data (e.g. fine scale IKONOS data); (3) make use of performance increases in data from new satellites or sensors-networks (4) make use of underexploited Digital Elevation Models (DEM); and (5) use spatio-temporal three-dimensional data (Gugerli et al. 2008) instead of point environmental data, as is especially important in studies of vagile animals. We could then precisely match genetic data to environment at adequate spatial and temporal scales. For example Sork et al. (2010) used fine scale climate data at a scale appropriate for the genetic data to understand how climate change shapes the evolutionary response of Californian valley oak (Quercus lobata).

**Spatial aspects specific to Quercus lobata**

To predict the future geographical range of a species, it is crucial to understand how species biologically respond to spatial heterogeneity of the environment or landscape at multiple spatial and temporal scales (Fortin & Dale 2005). Current species distributions are the result of many confounding processes, including population demography and history, phylogeographic history, behavior, physiological tolerances, competition, response to human land use change and adaptation to the environment (Gaston 2003). The interplay between selection and gene flow strongly influences biotic processes linked to adaptation (Savolainen et al. 2007; Holderegger & Wagner 2008).

Species distributional response to environmental conditions is a phenomenon that is often referred to as spatial dependence (Legendre 1993; Fortin & Dale 2005; Wagner & Fortin 2005). Species spatial aggregation occurs as well due to biotic processes such as dispersal and species interactions. These spatial structures create spatial autocorrelated genetic data. The degree of spatial autocorrelation in genetic data can be measured though various spatial autocorrelation coefficients (Fortin & Dale 2005). For animal species, a hypothetical example of the effects of habitat (Fig. 1) on spatial dependence of genetic associations among individuals for a neutral genetic locus versus a locus under selection is represented by Fig. 1. Spatial distributions of genotypes at the two loci are characterized by different spatial autocorrelation patterns. Measures of spatial structure for genotypes at the neutral locus often exhibit an isolation by distance pattern (Wright 1943) reflecting localized breeding and gene flow (Fig. 1b). In the example, environmental variables intrinsic to forest habitats

---

**Box 4. Initiatives to map the environment**

**Large scale database measures.** The Global Map project (http://www.globalmap.org/) exemplifies the trend toward constructing freely available, large-scale environmental data sets. It will include elevation, land cover (including vegetation) and land use data, as well as transportation infrastructure and political boundaries. The project is supervised by the International Steering Committee for Global Mapping (Secretariat of ISCGM 1998) with over 90 participating countries (Verdin & Jenson 1996). The main international global environmental geodata sources are included into the Global Map project and are available over the Internet from the Secretariat of ISCGM housed within the Geographical Survey Institute of Japan.

Several important international or national agencies have made efforts to freely distribute geo-environmental data describing the earth at different resolutions and for different periods. Primary among these are the European Environment Agency (EEA; http://www.eea.europa.eu/), American agencies such as USGS and NASA, and LANDSAT satellite images (http://www.landsat.org), which have offered global orthorectified data free of charge. Moreover, the Global Biodiversity Information Facility (GBIF; http://www.gbif.org/) is an international organization which aims to make the world’s biodiversity geodata digitally available (including data on livestock species). Finally, UNEP documents the Global Environment Outlook (http://www.unep.org/geo). This UN report presents the challenges facing the Earth in safeguarding the environment and moving towards a more sustainable future, and it proposes a data compendium with a list of all key data providers (http://geocompendium.grid.unep.ch/).

**Local scale sensor measures.** With regard to local scale, research in landscape genomics will benefit from an ongoing major technological revolution in the acquisition of high spatial and temporal resolution environmental data. Sensor networks can be used for survey of the environment at many different scales, from continental systems designed to measure global change to recent advances allowing high resolution monitoring of specific habitats. They can be combined with computational tools including high-performance communication networks, data storage systems, GIS and visualization environments (Rundel et al. 2009). Moreover, resulting data can be easily integrated with remote sensing or other types of standard sets of ecoclimatic parameters. The main quality of sensor networks lies in their capacity to extend spatial and tem-
or gradients (de Gruijter & ter Braak 2004). Also, it is important to assess the effective distance and pathway that an organism would use to move in a heterogeneous landscape (Spear et al. 2010), in order to relate the genetic diversity to the appropriate landscape features. For example, Vignieri (2005) tested whether individuals of the Pacific jumping mouse were using riparian zones or mountains to move between areas.

In quantifying the spatial structure of genetic data, it is difficult to tease apart the relative proportion of spatial dependence versus spatial autocorrelation described in the previous section, which are always confounded in both plants and animals. A potential solution to this problem is spatially-nested sampling designs, whereby for animals the distance between sampling locations varies from less than that of the species daily movement (i.e. to capture the degree of spatial autocorrelation) to beyond natal dispersal (i.e. to determine the environmental-species relationship) (Fortin & Dale 2005). Another way would be to perform model-based sampling to account for known environmental structure or gradients (de Gruijter & ter Braak 2004).

Challenges in spatial analysis of adaptive loci

The null hypothesis is that there is no correlation between a particular allele and environmental factors such as temperature or moisture apart from that which may be caused by limited dispersal and genetic drift. Relating specific alleles to an environmental variable is similar in some regards to association studies (Gupta et al. 2005; Balding 2006) that link alleles to phenotypes or to studies correlating species occurrence to environmental variables, as in ecological niche models. Methods used to tackle this problem have ranged from simple approaches such as linear regression to more sophisticated approaches such as generalized additive multifunctional landscapes. This underexploited tool can provide multiscale data (Lassueur et al. 2006) to be used in landscape genomics studies.

Multiscale measures and Digital Elevation Models (DEM). DEM, using elevation measures (from database or direct measures) and numeric models, can provide a diversity of morphometric (slope, aspect, curvature), hydro-morphometric (e.g. wetness), and also climatic indicators (e.g. solar radiation). The USGS Earth Resources Observation and Science Center (http://eros.usgs.gov) distributes global digital raster data sets with spatial resolutions ranging from 1 km (GTOPO30) to 90 m resolution (SRTM), and even 30 m for the United States and territorial islands. These data sets can be completed with increasingly available Very High Resolution DEMs (1 m for XY coordinates, and ~0.5 m for Z) acquired with LIDAR (Light Detection And Ranging) technology, and are able to generate high-resolution habitat predictors (Andrew & Ustin 2009). This underexploited tool can provide multiscale data (Lassueur et al. 2006) to be used in landscape genomics studies.

Spatial patterns in genotypic variation in the locus under selection is therefore the end result of convergence, isolation by distance and environmental effects intrinsic to forest as well as to grassland habitats.

In addition, current species distributions may have resulted from adaptations to environmental conditions that no longer exist (i.e. ancestral vs. current niche) (Wiens & Graham 2005; James et al. 2007; Roe et al. 2009). Indeed, a species optimal habitat may have already been lost or changed due to either natural or human influence. In such circumstances, current environmental–genotypic relationships would not be reliable as indicators of a species’ genetic responses to environmental changes (Cushman et al. 2009).

Sampling design

Sampling of populations adapted to different habitats, climates, land uses or management systems (e.g. for livestock) must be carefully designed and statistically analysed over appropriate geographic scales (Loehr 1999; Fortin & Dale 2005; Muirhead et al. 2008; Schwartz & McKelvey 2009; Anderson et al. 2010). Sampling effort that is too low (Muirhead et al. 2008; Schwartz & McKelvey 2009), in light of the multiplicity of landscape and environmental factors acting at multiple spatial and temporal scales, limits the signal to noise ratio of spatial genetic structure and could lead to misinterpretation of spatial statistics. To interpret detection of significant relationships between genetic and spatial environmental data, one should distinguish among the effects of sampling effort (Fortin & Dale 2005; Muirhead et al. 2008), sampling design (Fortin et al. 1989; Legendre et al. 2002), the power of statistical methods employed (Fortin & Dale 2005) and the confounding effects of multiple spatio-temporal scales (Dungan et al. 2002; Geffen et al. 2004; Boulet et al. 2007). Accordingly, sampling designs should be stratified across environmental variables of interest using current landscapes features and environmental conditions as a quasi-experimental design to test specific hypotheses. In fact, landscape heterogeneity itself can be used as a quasi-experimental design to test specific hypotheses. For example, samples taken along an altitudinal gradient could be used to determine local adaptation to climatic conditions (e.g. Bonin et al. 2006).

© 2010 Blackwell Publishing Ltd
models (Guisan et al. 2002; Pearman et al. 2008). Potential solutions have been proposed to consider explicitly the spatially dependent nature of the data using spatial regression methods (Dormann et al. 2007; Diniz et al. 2009; Dormann 2009). However, applications to detect loci potentially under selection are still lacking (but see Manel et al. 2010). Ideally, statistical methods in landscape genomics should consider both (1) spatial autocorrelation in allele frequencies generated by biotic processes (i.e. gene flow) which are distance related; and (2) unaccounted spatially structured environmental variables resulting in a spatial structuring of allele frequency distribution (Manel et al. 2010). Spatial regression methods have been put forth (e.g. conditional autoregressive models and simultaneous autoregressive models) to consider spatial dependence between individuals/loci and biotic processes by incorporating geographic space in the model structure. A promising spatial regression approach is the method of Moran’s eigenvector maps (MEM) (Borcard & Legendre 2002; Dray et al. 2006; Diniz-Filho et al. 2009). MEM variables are the eigenvectors of a spatial weighting matrix calculated from the sampling locations’ geographic coordinates. MEM analysis produces uncorrelated spatial eigenfunctions used to dissect the spatial patterns of the studied variation (allele frequencies in the present context) into separate scales to be used as predictors in regression. To detect loci potentially linked to genes under selection, Manel et al. (2010) used multiple linear regressions to correlate single AFLP allele frequencies from a large genome scan of Arabis alpina with environmental variables. To consider unmeasured variables in the analysis which potentially create spatial structure in allele distribution, they used only broad-scale principal coordinates of neighbour matrices (MEMs) as explanatory variables.

When sample size is small, spatial regression methods may not be appropriate given that the signal to noise ratio is generally low; geographically weighted regression has been proposed as one promising alternative (Fotheringham et al. 2002). In non-stationary circumstances, i.e. when spatial autocorrelation and effects of environmental correlates are not constant across the region, regression tree methods (e.g. CART, random forest, boosted regression; Elith & Graham 2009) offer alternatives to spatial regression (Dormann et al. 2007; Fortin & Melles 2009). Regression tree methods are based on an iterative procedure that splits the observations (samples) into a series of two groups in a hierarchical ‘tree’ (dendrogram-like) structure where the values of dependent variable are similar within each group based on a specific value of one of the independent variables (quantitative or qualitative independent values). Usually the first deeper splits reflect mostly large spatial scales processes while the last shallower splits in the tree structure correspond to localize spatial effects.

Consideration of spatial autocorrelation (i.e. biotic and abiotic processes) in the models allows the determination of processes governing allele frequency variation, but results may be strongly affected by sampling and stochastic variation (Slatkin & Arter 1991). Population and family structures have also been highlighted as a confounding issue in the inference of natural selection (Balding 2006; Excoffier et al. 2009). Bayesian geographical analysis approaches have been recently introduced to address this problem by testing for correlations between allele frequencies and environmental variables after correcting for background levels of population structure and differences in sample size (Felsenstein 2002; Yu et al. 2006; Hancock et al. 2008). Using this approach, Hancock et al. (2008) found evidence of a selective effect of the climate on metabolism genes in humans from the analysis of the association between 973 SNPs and climatic variables. This approach requires...
that populations are known or defined in advanced (i.e. from genetic structure) to be able to estimate allele frequency, which is not always possible depending on the species and the sampling (Manel et al. 2005). Studies of adaptive genetic variation can benefit by genotyping many populations, in a broad range of conditions (Turner et al. 2010). It appears that in some cases, it is more effective to sample a large number of locations with fewer individuals than to sample many individuals in only a few locations (Poncet et al. 2010).

Once a model has been chosen, it is necessary to choose among multiple, ideally uncorrelated explanatory variables. Model selection procedures are commonly used for this purpose by giving a weight (score of importance) to each explanatory variable (Burnham & Anderson 2002). Such analyses result in choosing the factors that explain the highest proportion of variation in the dependent variable (usually allele frequency variation in landscape genomics studies).

Models will likely become increasingly complex in addressing landscape genomics issues, able to account for longer term effects, various modes of adaptive selection, linkage disequilibrium (Servin & Stephens 2007), pleiotropy and epistasis, structural versus regulatory genetic effects, as well as being able to compare multiple null and alternate hypotheses, and tailored to the characteristics of the study species. In light of the great complexity of landscape genetic processes, the goal of predicting the population genetic effects following projected global changes for a given species will require that appropriate models be constructed carefully, taking into account as many details of organismal biology as possible. The modes of selection responsible for current adaptation must be determined and implemented, and projection models should be spatially explicit and include both stochastic and uncertainty components. Due to such complexity, most models will be based on computer simulations (see Epperson et al. 2010). We are currently working on programs that ultimately will be able to model multiple distributions of a very wide range of patterns of environmental variables (and how these impose selection), include complex patterns of dispersal and genetic transmission, are multilocus, and allow environmental patterns to change over time. Again, such approaches to projection will not necessarily be simple, and careful attention should be paid to model assumptions and sources of error.

Conclusions

Forthcoming whole genome data sets will propel molecular ecology into a new dimension of genetic and evolutionary analysis. Landscape genomics, via studying the spatial distribution of loci of adaptive or ecological significance in natural populations, will contribute to the better understanding of plant and animal adaptation to their environment and inform management of genetic resources in response to adaptation to global change. Recent studies investigated the geographic and environmental pattern in SNP’s associated with candidate genes, opening new insights in the understanding of the potential of populations to adapt to climate change (Eckert et al. 2009a,b; Eckert et al. 2010). Such studies provide an opportunity to resolve unanswered questions such as: does adaptation to local environments involve new mutations or standing genetic variation? How many genes influence ecologically important traits (Orr 2005; Stinchcombe & Hoekstra 2008)? The next step is to model spatially explicit forecasts of population genetic responses to climate shifts. What is needed are spatially explicit metapopulation and continuous space models (Wade & McCauley 1988; Harding et al. 1998) with directional spatial shifting of the environment. Models for genes that are differentially selected in direct response to climate change will necessarily add other layers, as well as the potentially complex interactions between dispersal and selection.

Acknowledgements

This work was conducted as a part of the Landscape Genetics Working Group supported by the National Center for Ecological Analysis and Synthesis, a Center funded by NSF (Grant #DEB-0553768), the University of California, Santa Barbara, and the State of California. SM was also supported by the Institut Universitaire de France. RH was also supported by the CCES-BIOCHANGE project of the ETH-domain. MSR was also supported by the National Evolutionary Synthesis Center (NESCent: NSF #EF-0905606) and NSF # DBI-0542599. We also thank Victoria Sork and anonymous reviewers for helpful comments on an earlier version of the manuscript.

References


© 2010 Blackwell Publishing Ltd


Parmesan C (2001) Detection of range shifts: general methodological issues and case studies of butterflies. In:
ADAPTIVE GENETIC VARIATION AND LANDSCAPE GENETICS