Chapter 3

BASICS OF POPULATION GENETICS: QUANTIFYING NEUTRAL AND ADAPTIVE GENETIC VARIATION FOR LANDSCAPE GENETIC STUDIES

Lisette P. Waits[1] *and Andrew Storfer*[2]

[1]*Fish and Wildlife Sciences, University of Idaho, USA*
[2]*School of Biological Sciences, Washington State University, USA*

3.1 INTRODUCTION

The field of landscape genetics integrates concepts and methods from landscape ecology and population genetics. Population genetics has a rich history focused around the study of the genetic composition of populations and changes in genetic variation that are driven by the processes of mutation, selection, gene flow and drift. This chapter is designed to provide an overview of molecular methods and population genetic theory for readers who have no or little background in population genetics. We introduce a conceptual framework describing how landscapes influence genetic variation and highlight the major research questions in landscape genetics. We review molecular genetic terms and molecular methods for evaluating genetic variation and present a summary of population genetic theories that are particularly relevant for landscape genetics. This chapter provides an overview of analytical methods for measuring genetic diversity of populations and individuals, and the methods used to assess genetic structure and gene flow in natural populations. We end with a discussion of future challenges in population and landscape genetics, with particular attention to interpretation and analysis of models and analyses for harnessing the power of next-generation sequencing data.

Landscape Genetics: Concepts, Methods, Applications, First Edition. Edited by Niko Balkenhol, Samuel A. Cushman, Andrew T. Storfer, and Lisette P. Waits.

3.2 OVERVIEW OF LANDSCAPE INFLUENCES ON GENETIC VARIATION

Deoxyribonucleic acid, or DNA, is the genetic material that codes for all of the diversity of life. The information in DNA is stored in four chemical bases: adenine (A), guanine (G), cytosine (C), and thymine (T). DNA bases pair up with each other, A with T and C with G, to form units called base pairs. Each base is attached to a sugar molecule and a phosphate molecule to make up the double helix backbone of DNA. The order, or sequence, of bases provides the information for building and maintaining individual organisms. DNA contains *genes* that code for proteins, which are created through the processes of transcription and translation. During transcription, DNA is translated to messenger RNA (mRNA) and during translation the sequence of mRNA is read. Each sequence of three bases codes for an amino acid and mRNA is converted to a protein one amino acid at a time. Overall, the DNA sequence is very similar among individuals within a species. For example, human DNA consists of about 3 billion base pairs and less than 1% is different among individuals. However, these small numbers of differences, known as genetic variation, are very important.

At its most basic level, genetic variation can be defined as DNA sequence differences at the same physical location in the genome. This variation can be observed at locations in the genome that are under selection (***adaptive loci*** or ***non-neutral loci***) or locations that are not under selection (***neutral loci***). Genetic variation of a population or species is influenced by four main processes – mutation, selection, gene flow, and genetic drift (Hedrick 2011; Allendorf et al. 2012). ***Mutation*** is the process that creates new genetic variants or ***alleles*** due to errors in DNA replication. Selection impacts genetic variation by increasing the frequencies of alleles that are favorable in a particular environment and decreasing the frequencies of alleles that are less favorable in that environment. If a ***locus*** or a mutation is neutral, selection does not directly influence the frequency of alleles. Gene flow occurs as a result of migration (dispersal and subsequent reproduction) and this process moves alleles between populations and tends to make them more similar genetically. Genetic drift is the change in allele frequencies due to random sampling effects as alleles are passed on from one generation to the next (Wright 1931). Genetic drift is thus influenced by the number of breeding individuals and the variance in reproductive success among breeders.

Genetic drift is directly linked to ***effective population size*** (N_e). Sewall Wright (Wright 1931, 1938) introduced the concept of N_e to mathematically and conceptually represent the effects of genetic drift, and N_e is the population genetic analog to census size used in ecology. Wright described N_e as the number of breeding individuals in an idealized population that contribute genetically to the next generation. A simplified but useful approach for understanding N_e is to think about it as the number of individuals that are able to successfully pass on their genes to the next generation. Populations with larger N_e will be able to maintain higher amounts of genetic variation because more new alleles are created by mutation due to a larger number of breeding events and fewer alleles are lost to genetic drift since there is a larger sample of breeders each generation. N_e also influences the relative impacts of genetic drift and selection in natural populations (Fisher 1930; Wright 1931). For example, as N_e decreases, the effect of selection decreases relative to the effect of genetic drift and important adaptive genetic variation can be lost (Hedrick 2011).

Population geneticists often evaluate two main components of genetic variation: genetic diversity and genetic structure. Genetic diversity is the amount of genetic variation found within an individual, a population, or a spatial area. On the other hand, genetic structure is the distribution of that variation among individuals, populations, or areas. Landscapes influence the amount and distribution of genetic variation in many ways. First, the landscape influences where an organism can live and how many individuals live in a particular location that affects the distribution and amount of genetic variation (Fig. 3.1). The landscape also influences the amount of movement and

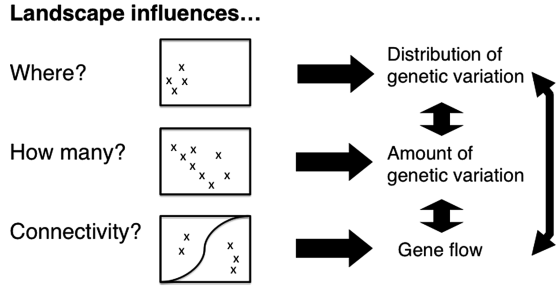

Fig. 3.1 Influence of the landscape on individual organisms and genetic variation.

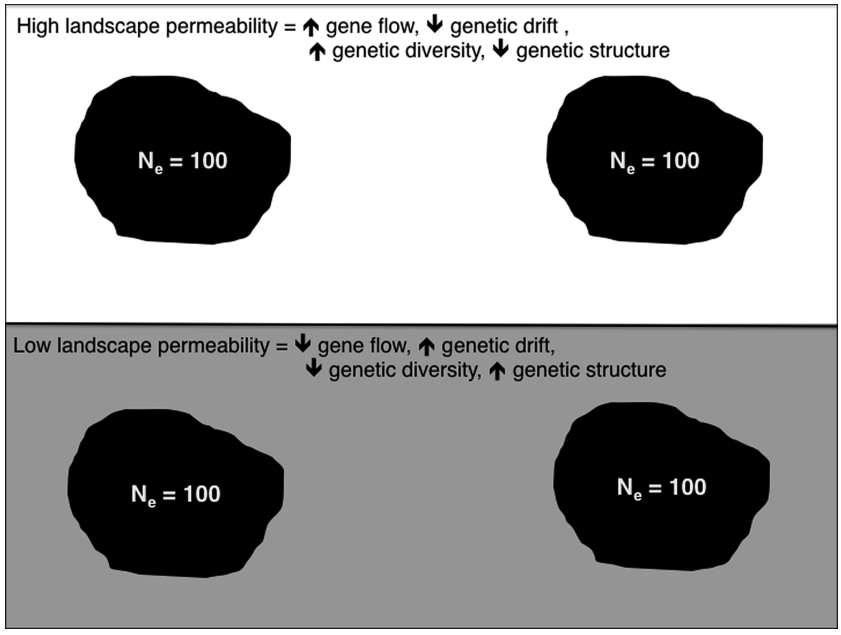

Fig. 3.2 The effect of landscape permeability on genetic diversity and structure with equal effective population sizes (N_e). The white landscape matrix in the top panel has high permeability while the gray landscape matrix in the bottom panel has low permeability. Thus, gene flow is higher and genetic drift is lower in the top panel, leading to higher levels of genetic diversity and lower levels of genetic structure in the white landscape compared to the gray landscape.

subsequent gene flow that occurs among individuals in different locations, which directly affects N_e and the amount and distribution of genetic variation (Fig. 3.1).

For example, when two populations of equal N_e are imbedded in habitat matrices of differing levels of permeability, the two populations separated by the more permeable matrix will retain more genetic variation due to increased gene flow and reduced genetic drift (Fig. 3.2). This difference in matrix permeability also creates different levels of genetic structure. With the less permeable matrix (Fig. 3.2), a greater amount of genetic structure is created due to the increased restriction of gene flow, which increases genetic drift and causes allele frequencies to diverge at a faster rate. In an alternative scenario, where matrix permeability is the same but the two populations have different N_e, the population with the smaller N_e will lose genetic variation more quickly and genetic structure will increase more rapidly due to increased genetic drift (Fig. 3.3). In this scenario, differences in N_e could be due to differences in the habitat quality of each patch, and the patch with the higher habitat quality could support a higher N_e.

For loci that are under selection, the amount of genetic variation and the frequency of particular alleles are directly affected by the environmental conditions at a site. For example, the melanocortin-1 receptor (*Mc1r*) gene has a large effect on coat color in beach mice (Hoekstra et al. 2006). One allele for this gene codes for dark color mice and is in very high frequency in the forest habitats, where dark mice match the forest floor. Another allele codes for light color and is found in high frequency in mice that match the sandy beach habitats where they reside (Linnen et al. 2009).

In the following sections, we will (1) describe molecular markers and methods used to generate genetic data for landscape genetic studies, (2) review population genetic theory and terminology, (3) overview the metrics and methods used to measure genetic diversity, genetic structure, and gene flow, and (4) discuss future directions and challenges for genetic analysis.

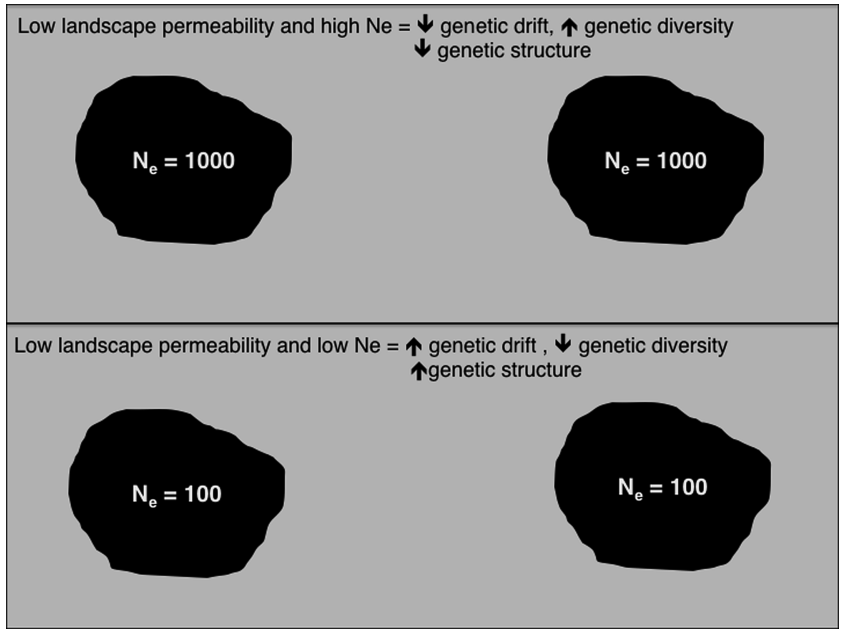

Fig. 3.3 The effect of effective population size (N_e) on genetic diversity and structure. The landscape has low permeability in the top and bottom panels but the top panel has populations with much higher N_e than the bottom panel. Thus, genetic drift is lower in the top panel, leading to higher levels of genetic diversity and less genetic structure compared to the example in the bottom panel.

3.3 OVERVIEW OF DNA TYPES AND MOLECULAR METHODS

3.3.1 Types of DNA

There are two main types of DNA in animal cells: ***mitochondrial DNA*** (mtDNA) and ***nuclear DNA*** *(nDNA)*. Plant cells contain mtDNA, nDNA, and ***chloroplast DNA*** (cpDNA). MtDNA is a circular DNA molecule of the mitochondrion, an organelle that is the energy powerhouse of cells. MtDNA has a uniparental mode of inheritance in animals and plants as it is generally passed only from mother to offspring (Gillham 1974; Birky 1978), but paternal transmission has been documented in some animals (Zouros 2000; Zhao et al. 2004; Wolff et al. 2013) and conifer species (Neale et al. 1989). Nuclear DNA is the DNA of chromosomes found in the nucleus of a cell and has a biparental mode of inheritance since genetic material is inherited from both the mother and the father. The cpDNA is contained in the chloroplast molecule responsible for photosynthesis in plants and its mode of inheritance differs by species and can be maternal, paternal, or biparental (Gillham 1974; Neale et al. 1989). On average, the N_e of nDNA loci is four times higher than the N_e of cpDNA and mtDNA because organellar DNA is most often uniparentally inherited and ***haploid***. The type of DNA, its mutation rate, and mode of inheritance influence the research questions that can be addressed and the temporal inference that can be obtained from genetic data. In general, nDNA loci are used more frequently in landscape genetic studies (90%) than mtDNA or cpDNA loci (Storfer et al. 2010) because they provide better resolution for detecting recent changes to the landscape.

3.3.2 Adaptive versus neutral loci

At noted above, landscape genetic researchers can collect data at two main types of loci: adaptive loci, which are under selection, and neutral loci, which are not affected by selection. Adaptive loci directly or indirectly affect the phenotype and fitness of an organism, and landscape or environmental factors at a site

Table 3.1 Examples of landscape genetic research questions that can be addressed with neutral and adaptive loci.

Locus type	Research question
Neutral	What landscape features are barriers to movement and gene flow?
Neutral	What landscape features facilitate movement and gene flow?
Neutral	What spatial areas or patches are sources and sinks?
Neutral	How does the landscape influence the amount of genetic variation and has it changed over time?
Adaptive	Which loci are under selection?
Adaptive	Are there changes in allele frequencies at certain loci across an environmental gradient?
Adaptive	Which loci are associated with phenotypic differences among populations?
Adaptive	How do environmental effects on adaptive genetic variation influence individual fitness and persistence of populations and species?

will influence allele frequencies through selection (see Chapter 9). In contrast, allele frequencies at neutral loci are primarily affected by gene flow and drift. This inherent difference between neutral and adaptive loci determines the types of research questions that can be addressed with neutral and adaptive loci (Table 3.1). When using neutral loci, researchers can address research questions related to how the landscape influences gene flow, the effective population size, or neutral genetic diversity. Thus, neutral genetic data are particularly well suited for research questions on impacts of past and present landscapes on genetic connectivity (e.g., corridor and reserve design) and for predicting how future landscape change would affect this connectivity and resulting genetic variation (see Chapters 10 to 12 for empirical examples). Using adaptive loci, we can address how environmental factors, such as temperature and elevation, affect the geographic distribution of genetic variation (Chapter 9). Currently, neutral loci are used more commonly than adaptive loci in landscape genetics (Holderegger and Wagner 2008; Sork and Waits 2010), but the number of studies using adaptive loci is increasing.

3.3.3 Molecular methods

After choosing to collect data for a particular type of DNA and locus, a variety of methods are used to obtain genetic data. There are two main types of nDNA molecular methods – codominant and dominant approaches. In codominant methods like microsatellite analysis or single-nucleotide polymorphism (SNP) analysis, it is possible to visualize both alleles at a particular locus in the form of a ***genotype*** (AA versus AB, for example). While dominant loci methods, such as amplified fragment length polymorphism (AFLP), create banding patterns of tens to hundreds of bands for each individual that resemble a barcode, both alleles from a particular locus cannot be identified. Dominant loci data are generally recorded in a binary presence/absence format, where 1 indicates presence and 0 indicates absence (0, 1, 0, 0, 0, 1, for example). Mitochondrial DNA and cpDNA are generally analyzed using DNA sequencing or restriction fragment length polymorphism (RFLP), approaches that generate ***haplotype*** data rather than genotype data since the loci are haploid. Fully describing these methods is beyond the scope of this chapter, but the most commonly used approaches, including protein allozyme analysis, nDNA microsatellite analysis, AFLP analysis, SNP analysis, and cpDNA/mtDNA sequencing, are summarized in Box 3.1 and Table 3.2. For more detailed descriptions of each molecular method, see Lowe et al. (2004) or Allendorf et al. (2012).

Box 3.1 Overview of molecular methods used in landscape genetic analyses

Allozyme analysis – Allozyme analysis detects allelic variants of protein enzymes encoded by genes.
Polymerase chain reaction (PCR) – PCR is the chemical process used to make millions of copies of a particular target DNA region or locus using short single-stranded pieces of DNA known as primers (Mullis and Faloona 1987). PCR has revolutionized molecular genetics and is used in all methods described below.

Sanger DNA sequencing – Sanger DNA sequencing is the chemical process used to read the sequence of nucleotides (DNA base pairs – A, G, C, T) at a particular DNA region (Sanger et al. 1977).
Next-generation sequencing (NGS) – This group of DNA sequencing approaches parallelize the sequencing process and can produce a greater volume of sequence data in a shorter period of time (Shendure and Ji 2008) than the original Sanger sequencing method.
Microsatellite analysis – Microsatellite loci (also known as short tandem repeats (STRs) or simple sequence repeats (SSRs)) or "μsats" are repeating sequences of 2–6 base pairs (CACACA or GTCGTCGTC, for example). Nuclear DNA microsatellite analysis has been the most commonly used data collection approach in landscape genetics (Storfer et al. 2010).
Amplified fragment length polymorphism (AFLP) analysis – AFLP analysis is a PCR-based tool for assaying genetic variation that does not require knowledge of the DNA sequence of the target species. The amplified DNA fragments are separated by size on a polyacrylamide gel and visualized using fluorescence or autoradiography (Vos et al. 1995).
Single nucleotide polymorphism (SNP) analysis – SNP analysis is a newer molecular method that surveys single base pair genetic polymorphisms in many locations throughout the genome. This method is rarely used in landscape genetics, but the application of this method is predicted to expand rapidly in coming years because it surveys a greater proportion of the genome and can be automated for high throughput analyses (Morin et al. 2004; Garvin et al. 2010).

The most important methodological advance in genetics was the development of the polymerase chain reaction or PCR (Mullis & Faloona 1987). This technique makes it possible to survey genetic variation by making many copies of a particular DNA locus. A recent review of landscape genetics quantified the frequency of application of different molecular methods and found that 70% of animal and 31% of plant studies used nDNA microsatellite loci, approximately 10% of animal and plant studies used mtDNA or cpDNA sequence data, 21% of plant and 5% of animal studies used AFLP analysis, and <2% of studies used SNPs (Storfer et al. 2010).

3.3.4 Unit of analysis

The two main units of analysis for genetic data in landscape genetics are individuals and populations (i.e., groups of individuals). In general, the population is chosen as the unit of analysis for species that are patchily distributed and the individual is chosen as the unit of analysis when the species is more continuously distributed across a landscape and there are no clear population boundaries. However, recent simulation work has shown that it can also be effective to use individuals as the unit of analysis for patchily distributed organisms (Prunier et al. 2013). Choosing the unit of analysis is an important study design decision for landscape genetics research because it influences the sampling strategy (especially the sampling level, see Chapter 4) as well as the metrics and analyses methods that can be applied. In the following sections, we will describe population genetic metrics that can be applied at either the population or individual level.

Table 3.2 Characteristics of the most commonly used molecular methods in landscape genetics.

Molecular method	Inheritance	Variability	Temporal scale of Inference
Nuclear DNA microsatellite	Biparental	High	Recent
Amplified fragment length polymorphism (AFLP)	Dominant*	High	Recent
Single nucleotide polymorphism (SNP)	Biparental	Moderate	Intermediate
Mitochondrial DNA sequence	Uniparental	Low–moderate	Historic
Chloroplast DNA sequence	Uniparental	Low	Historic

* Dominant methods provide only the presence or absence of a DNA fragment.

3.4 IMPORTANT POPULATION GENETIC MODELS

3.4.1 Hardy–Weinberg equilibrium

The basic mathematical formulation for population genetics was developed by William Weinberg (1908), a German physician, and Godfrey Hardy (1908), an English mathematician, in 1908. Although Weinberg published his findings approximately six months before Hardy, their major contribution is now known as the "Hardy–Weinberg" rule. This rule states that for a diploid organism with two alleles at one locus, if p represents the frequency of one allele and q represents the frequency of the other allele, then genotype frequencies will be $p^2 + 2pq + q^2 = 1$, whereby p^2 represents the pp genotype or the p ***homozygote*** frequency, $2pq$ represents the ***heterozygote*** frequency, and q^2 represents the q ***homozygote*** frequency. These allele frequencies are expected under Hardy–Weinberg equilibrium (HWE) in one generation of random mating in a population, that: (1) is infinite in size; (2) has no selection; (3) has no mutation; (4) has no migration; and (5) has random mating. Although these conditions are unrealistic expectations for natural populations, Hardy–Weinberg equilibrium is a useful starting assumption. That is, if predictions of allele frequencies under HWE are violated with empirical data, one can assume that at least one of the five evolutionary processes above is operating. Much of population genetics theory then deals with allele frequency expectations under models such as selection, mutation, migration, or combinations of two to three of these processes. For detailed overviews of population genetics theory, texts such as Hartl and Clark (2006) and Hedrick (2011) are recommended. We discuss some of the models most relevant to landscape genetic studies below.

3.4.2 Linkage equilibrium

Under Mendel's law of independent assortment, it is assumed that the cross-generational transmission of alleles at any particular locus is independent of alleles at other loci and that the fitnesses of possible pairs of alleles are decoupled. This idea is called linkage equilibrium or gametic equilibrium. It follows, then, that ***linkage disequilibrium*** exists when there is a statistical (non-random) association between alleles at different loci. These alleles can be physically linked on chromosomes, whereby they are in near proximity to one another and those inherited together. Alternatively, other evolutionary processes such as genetic drift in small populations can create linkage disequilibrium among pairs of alleles. Moreover, selection can cause genetic "hitchhiking", whereby alleles of little-to-no effect on fitness are inherited together with alleles favored under selection.

Because landscape genetics studies often rely on a dozen or so loci, the assumption of linkage equilibrium among alleles and loci is important for estimations of population structure. That is, loci are assumed to be representative of genomic levels of variation and significant linkage disequilibrium among loci means they cannot be treated independently because they tend to be inherited together. As such, for studies with relatively few loci (e.g., fewer than 20), linkage disequilibrium among loci essentially reduces overall power to assess population structure. This is less of an issue for population genomic studies based on next-generation sequencing technologies (see Chapter 9) where hundreds to thousands of loci are generated. Loci under significant linkage disequilibrium can thus be discarded in these cases, whereby overall sample size is proportionally less affected. Most basic population genetics software packages, such FSTAT (Goudet 1995) and GENEPOP (Rousset 2008), will estimate linkage disequilibrium among loci.

3.4.3 Effective population size and genetic drift

Populations are not infinite in size and N_e is almost always smaller than the census population size or the number of individuals estimated from field or other surveys. Unequal sex ratio, assortative mating, and overlapping generations are all examples of demographic factors that make effective population sizes smaller than census population sizes. A review by Frankham (1995) suggests that effective population sizes are generally 10% (or lower) of the census population size across different wildlife species. There are two commonly estimated measures of N_e. Variance effective size (N_{eV}) is the size of an ideal population experiencing drift at the same rate as the actual population. ***Inbreeding*** effective size (N_{eI}) is the size of an ideal population losing heterozygosity, due to increased relatedness, at the same rate as the actual population. For stable, large populations N_{eI} and N_{eV} are similar, but when populations are growing or

shrinking they can differ greatly (Crandall et al. 1999). Detailed descriptions of methods for estimating N_e are beyond the scope of this chapter but good reviews of methods can be found in Leberg (2005), Wang (2005), and Hare et al. (2011).

Genetic drift results in random changes in allele frequencies across generations resulting from sampling of gametes, and thus has larger effects on smaller populations than larger populations. Consider, for example, a population of 10 haploid individuals, with an allele frequency of 0.1 for allele A at a locus versus a population of 100 individuals with the same allele frequency. In the smaller population, allele A is present in 1 individual, while in the larger population, it is present in 10 individuals. Therefore, in the smaller population allele A can be lost easily if the single individual with that allele does not breed for some reason. It is harder to imagine that a stochastic event will cause all 10 individuals in the larger population to fail to breed, however. In fact, the probability that an allele is lost in a population is 1 minus its starting allele frequency, so rare alleles in small populations tend to be lost.

In diploid organisms, heterozygosity (the proportion of individuals with 2 different alleles at a locus) is lost at a rate of $1/(2N_e)$ at each generation due to genetic drift. As such, the reduction of genetic diversity due to drift has become a management concern for small populations or populations of endangered species because maintenance of genetic diversity is recognized as important for the evolutionary potential of populations to respond to future environmental change. Extreme and rapid declines of (effective) population size result in population ***bottlenecks*** and thereby often rapid reduction in genetic diversity resulting from drift. Conservation biologists have proposed maintaining N_e above 50 to preserve short-term viability and avoid inbreeding depression (e.g., Franklin 1980) and at least 500 to avoid loss of evolutionary potential (Franklin 1980; Lande 1995). However, this general guideline has been controversial (Jamieson & Allendorf 2012; Frankham et al. 2013) and recent studies have suggested higher thresholds are needed to avoid inbreeding depression and maintain minimum viable populations (Lynch & Lande 1998; Frankham et al. 2013).

3.4.4 Mutation

Mutation is the ultimate source of genetic variation in populations and species. Mutations occur when mistakes are made during DNA replication and can take two main forms: point mutations or insertions or deletions of segments of DNA. Point mutations are changes of a single nucleotide, such as A to G or G to C. As much of the genome is non-coding, point mutations are often considered "silent" as they do not result in changes in protein structure. Even when point mutations do occur in segments of DNA that code for amino acids (i.e., exons), they often do not result in changes of amino acid sequences (and are called ***synonymous substitutions***) due to the conservation of the genetic code (i.e., 64 possible codons code for only 20 amino acids). Point mutations that do change amino acid sequences are called ***non-synonymous substitutions***. Insertions or deletions change the length of a strand of DNA and are hence called frameshift mutations when they occur in exons because they shift the reading frame that is transcribed, often resulting in major changes in amino acid sequences or even premature termination of protein formation.

Before we even knew what DNA was, mutations were already considered in population genetic models. In the 1930s, Fisher (1930) and Wright (1931) modeled the fate of neutral mutations, or those that had no fitness effect on the individual with a particular allele. The frequency of an allele in a generation ranges from 0 to $2N_e$ for diploid organisms. Mutation introduces variation at a rate of $2N_e\mu$, where μ = the mutation rate. As above, genetic drift reduces genetic variation at a rate of $1/(2N_e)$ per generation. Since the average number of new mutations entering the population each generation is $2N_e\mu$ and the chance of allelic extinction (loss of an allele) is $1/(2N_e)$, then the average substitution rate is $2N_e\mu \times 1/(2N_e) = \mu$.

Later, the idea of the "neutral theory" of molecular evolution (Kimura 1968) embraced the neutrality of mutations and therefore considered genetic variation in populations as a balance between the rate at which new alleles were introduced into new populations via mutation and the rate at which they were lost due to genetic drift. This view regards selection as a minor evolutionary force and heated debates about the relative influence of mutation, selection, and genetic drift ensued. A revision to the neutral theory called "nearly neutral theory" (Ohta 1973) was later developed. Nearly neutral theory suggests that, while most mutations can be considered selectively neutral, mutations that are slightly deleterious or advantageous are also allowed in the resulting models. In these models, alleles are chosen at random from the previous generation

and extinction or fixation of alleles happens much more quickly in small populations than larger populations due to genetic drift. This also means that even slightly advantageous alleles can be lost in small populations due to chance (i.e., genetic drift).

In population and landscape genetics studies, the model of mutation can be important because it introduces new alleles and thus results in allele frequency changes in populations, thereby affecting how genetically diverged or similar the study populations are. Three main models of mutation are commonly used in landscape genetics. The first is the infinite alleles model (IAM). The infinite alleles model assumes essentially that an infinite number of alleles are possible at a particular locus and, as such, any new mutation results in a new allele. At equilibrium, the number of alleles in a population remains constant because mutation introduces new alleles at the same rate at which they are lost by genetic drift. A more recently developed model that applies to microsatellites is the stepwise mutation model (SMM) (Ohta & Kimura 1973). Under the SMM, it is assumed that mutations occur stepwise to the next possible allelic state. Thus, for microsatellites, for example, an allele with 8 repeats can only mutate to 7 repeats or 9 repeats. As microsatellite studies revealed that mutations can occur at greater than one mutational step, the two-phase model (TPM) was developed (Di Rienzo et al. 1994). The two-phase model addresses this by considering mutations of 1 repeat (one-phase) occurring with probability p and mutations of ≥ 1 repeat (two-phase) with probability $1 - p$, with a geometric distribution of lengths greater than 1 repeat.

3.4.5 Migration (gene flow)

Although the term "migration" has a few different definitions, in population genetics it is synonymous with gene flow. Gene flow is defined as the movement of genes (via successful mating) among populations. High rates of gene flow tend to homogenize populations, thereby resulting in similar allele frequencies. Conversely, low rates of gene flow tend to result in population isolation and genetic differentiation. In turn, if populations have a small N_e, genetic drift will act to differentiate these populations genetically via random changes in allele frequencies. However, it only takes small amounts of migration (about 1 effective migrant per generation in theory) to overcome drift, which is a relatively weak evolutionary force (Wright 1969).

In reality, maintenance of population connectivity from a management standpoint may require artificial movements of greater than one actual individual per generation, depending on the effective size of the population(s) being managed (Hedrick 1995). For example, when the Florida panther reached a very small population size and showed evidence of inbreeding depression, 20 Texas cougars were introduced as a one-time measure to increase genetic diversity in the panther population (Hedrick 1995). The introduction was successful, but required detailed analyses and, in this case, involved concern about introgression of a separate subspecies. These concerns were addressed with a single, rather than repeated introductions. Thus, management strategies for maintaining genetic diversity in small or declining populations, while based on Wright's (1969) original theory, need to be modified with current data that are often likely to be species-specific.

Several models of migration have been developed for use in population genetics and application in landscape genetics studies. The original "mainland-island" model of migration was developed by Wright in 1931 (Fig. 3.4a). In essence, the island model assumes that there is a mainland of infinite size exchanging migrants with an island of finite size at a rate of m. Given the larger size of the mainland, the allele frequencies on the island will come to resemble those of the mainland because the mainland has a disproportionate effect on the island. Alternatively, the island model can be thought of exchange of an infinite number of genes (as under the IAM) among an infinite number of islands, with the chance of any particular allele moving at rate m and that allele not moving as $1 - m$. Essentially the mainland island model can be thought of as an ideal situation or null model because real populations generally do not conform to its expectations. Later, a more restricted migration model called the "stepping stone" model (Fig. 3.4b) was developed (Kimura & Weiss 1964). The stepping stone model only allows migration between nearest neighbor populations. Consider a two-dimensional lattice of populations. For migration to occur from a population to another population two steps away in the lattice, individuals must move to an intermediate population one step away first before making the second step. In contrast to the mainland island model, this is at the other end of the continuum, with highly restricted gene flow among populations.

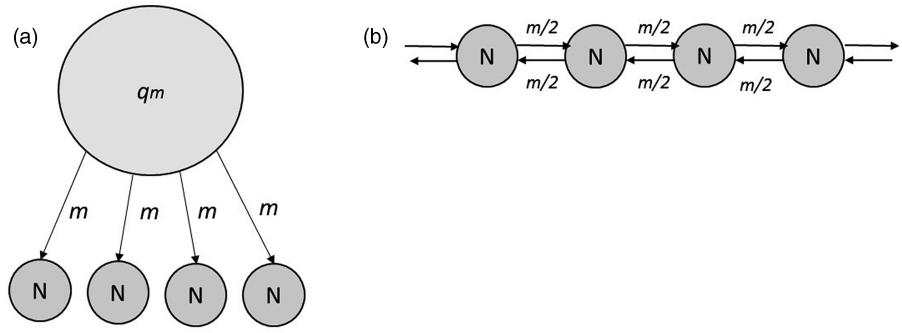

Fig. 3.4 Representations of different models of population structure. (a) Continent-island (or island) models. Shows an island of essentially infinite population size (q) that sends out migrants unidirectional at an equal rate (m) to subpopulations with finite population size (N). (b) Stepping-stone model. Populations of finite size (N) share migrants only with their nearest neighbor population at a rate of $m/2$ since gene flow is bidirectional and symmetrical.

3.4.6 Isolation-by-distance and landscape

The above "classical" methods of conceiving gene flow formed the basis for later tests of isolation-by-distance (IBD; Wright 1969). The concept of isolation-by-distance was first envisioned by Sewall Wright (1943, 1951) and assumes that genetic distance among populations is positively correlated with geographic distance among those populations. Implicit in this theoretical framework is that populations are finite in size and subdivided. The main approach for testing for this correlation has utilized Mantel tests (Mantel 1967) of matrix correlations between straight-line geographic distance and genetic distance. In general, most population genetics studies find significant positive correlations of genetic distance and geographic distance (Slatkin 1995; Whitlock & MacCauley 1999); this spatial scale generally correlates with the dispersal capabilities of the study organism. However, strict isolation-by-distance assumes movement in straight lines ("as the crow flies") and fails to account for which environmental characteristics might influence movement of individuals among populations. Landscape genetics as a discipline directly addressed this gap, recognizing that different components of the landscape, such as mountains, rivers and forests, can affect dispersal and consequent of different species in different ways (Manel et al. 2003; Storfer et al. 2007). For example, a mountain might act as a barrier to a species of vole, but not to a hawk. As such, the resulting general approach of "isolation-by-landscape" (IBL) considers the influence of intervening landscape characteristics on dispersal patterns and consequent gene flow of individuals or ***demes*** (see Chapter 8). Nearly all landscape genetics studies have shown that inclusion of landscape variables explains significantly more genetic variation among demes than (what is now most often considered) the null model of isolation-by-distance (Storfer et al. 2010).

A commonly used application of the isolation-by-landscape approach considers isolation-by-resistance (IBR) as a way of considering the effects of landscape features on the distribution of genetic variation in an explicitly spatial manner (Spear et al. 2010). The landscape under study is converted to a resistance surface, and resistance values are calculated between pairs of individuals or populations. Generally, resistance values are assigned to each landscape variable or environmental feature contained in a spatial layer of a raster GIS environment (Spear et al. 2010). Higher assigned values translate to greater hypothesized resistance to gene flow by that particular variable, whereas lower values suggest lower resistance. Thus, each raster cell in a GIS layer is assigned a cumulative value based on hypothesized habitat use and movement through each of the landscape variables contained in that cell. A resistance surface is comprised of all cells that comprise the intervening habitat matrix between samples collected across a study area.

Two major modeling frameworks have been used to convert resistance values into measures of population connectivity – least-cost paths and circuit-based analyses. Least-cost paths are based on the underlying assumption that costs, in terms of cumulative resistance values between populations, are minimized such that

individuals tend to move along an optimal path (Adriaensen et al. 2003). In contrast, circuit theory-based models simultaneously integrate all possible pathways that connect populations, while still assuming that genetic distance is positively correlated with resistance (McRae 2006). Additional details and pros and cons of these two modeling frameworks will be discussed in more detail in Chapter 8.

A more recent approach is "isolation-by-environment" (IBE; Wang et al. 2013), which predicts a correlation between genetic distance and environmental dissimilarity. It is expected that the more dissimilar habitats are among populations, where the stronger divergent selection will be consequently reducing fitness of dispersers (Lee & Mitchell-Olds 2011). This approach shows promise, as results of IBE analysis explained more spatial variation in gene flow than IBL analyses for 17 species of Caribbean *Anolis* lizards (Wang 2013).

3.5 MEASURING GENETIC DIVERSITY

3.5.1 Population level

There are a variety of ways that genetic diversity can be quantified (Table 3.3). When the population is the unit of analysis, there are four main summary metrics for codominant nDNA loci: (1) number of alleles, (2) allelic richness, (3) heterozyosity, and (4) proportion of polymorphic loci. The average number of alleles per locus

Table 3.3 Commonly used genetic diversity metrics for (a) populations and (b) individuals.

(a) Population level

Metric	Symbol	Equation	Method type
Polymorphism	P	Number of polymorphic loci/number of loci	nDNA SNP or μsat, AFLP*
Observed heterozygosity	H_o	Number of heterozygotes/number of individuals	nDNA SNP or μsat
Expected heterozygosity	H_e	$2n(1-\Sigma p_i^2)/(2n-1)$, n = sample size, p_i freq ith allele	nDNA SNP or μsat, AFLP*
Allelic richness	A_r	Rarefaction at smallest sample size	nDNA SNP or μsat
Haplotypic diversity	h	$n(1-\Sigma p_i^2)/(n-1)$, n = sample size, p_i freq ith allele	mtDNA or cpDNA sequence
Nucleotide diversity	π	Average number of nucleotide differences per site	mtDNA,cpDNA or nDNA sequence, AFLP*

(b) Individual level

Heterozygosity	H	Number of heterozygosity loci/number of loci	nDNA SNP or μsat
Standardized heterozygosity	SH	Proportion of heterozygous loci/mean H of loci	nDNA SNP or μsat
Internal relatedness	IR	$(2H-\Sigma f_i)/(2N-\Sigma f_i)$, H = homozygous loci, N = number of loci, f_i = freq ith allele	nDNA SNP or μsat
Homozygosity per locus	HL	$\Sigma E_h/(\Sigma E_h+\Sigma E_j)$, E_h is expected H of homozygous loci and E_j is expected H of heterozygous loci	nDNA SNP or μsat

AFLP = amplified fragment length polymorphism, cpDNA = chloroplast DNA, mtDNA = mitochondrial DNA, nDNA = nuclear DNA, SNP = single nucleotide polymorphism, μsat = microsatellite.

* An analogue to this metric can be estimated for dominant loci like AFLP when making specific assumptions.

(A) can be used to evaluate genetic diversity; however, this metric is very sensitive to differences in sample size. This problem can be avoided by using allelic richness (A_r), which can correct for differences in sample sizes using subsampling or rarefaction techniques (El Mousadik & Petit 1996; Leberg 2002; Kalinowski 2004). There are two main metrics for heterozygosity: observed and expected heterozygosity. Observed heterozygosity (H_o) is the proportion of individuals that are heterozygous at a particular locus. Expected heterozygosity (H_e) is calculated based on the observed frequency of alleles, assuming Hardy–Weinberg equilibrium, and can be corrected for sample size differences among populations (Nei 1978). Both H_o and H_e are estimated for each locus for a population sample and then averaged across loci. The polymorphism metric (P) estimates the proportion of loci out of the total analyzed that contain more than one allele. The polymorphism metric is used less frequently because loci are generally selected for landscape genetic studies because they are known to be polymorphic and thus informative for measuring diversity. The A_r metric is the most sensitive for detecting recent losses of genetic diversity (Allendorf et al. 2012). When sampling is more continuous and the location of populations is less clear, an alternative approach is to calculate these metrics in a spatially-explicit manner based on grouping individuals into genetic neighborhoods that match the population structure (Shirk & Cushman 2011).

For cpDNA and mtDNA sequence data, the main diversity metrics are the number of haplotypes, haplotypic diversity, and nucleotide diversity. A haplotype is the haploid version of a genotype and for cpDNA and mtDNA sequence data refers to a unique DNA sequence. Haplotypic diversity (h), also known as gene diversity, is the haploid version of expected heterozygosity. Nucleotide diversity (π) is the average number of differences in the nucleotides at each position along a DNA sequence. For dominant loci, like AFLPs, the standard diversity metric is the average number of bands per population, but analogues of H_e and π can be estimated under specific assumptions (Bonin et al. 2007).

3.5.2 Individual level

There are many metrics used to estimate individual heterozygosity using codominant loci: heterozygosity (H), standardized heterozygosity (SH), internal relatedness (IR), and homozygosity by loci (HL). Heterozygosity is just the number of heterozygous loci divided by the total number of loci analyzed for an individual sample. However, this simple metric can be biased when loci differ in number or frequency of alleles or when all individuals are not genotyped for the same set of loci. Thus, other metrics have been derived to address these potential biases. For standardized heterozygosity, the score for each locus is weighted by the average heterozygosity at that locus (Coltman et al. 1999). Amos et al. (2001) proposed a different measure, IR, based on allele sharing, where the frequency of each allele counts towards the final score, thereby allowing the sharing of rare alleles to be weighted more than the sharing of common alleles. The IR and SH metrics are highly correlated and both have been shown to perform well in studies of individual genetic diversity and fitness (Amos et al. 2001). However, IR underestimates heterozygosity of individuals carrying rare alleles. Homozyosity per locus (HL) has been developed as an alternative metric that avoids this weakness and reduces the sample sizes required to achieve a given statistical power (Aparacio et al. 2006). All of these metrics can be calculated in the R function GENHET (Coulon 2005).

3.6 EVALUATING GENETIC STRUCTURE AND DETECTING BARRIERS

One of the key components of a landscape genetic analysis is evaluating genetic structure (i.e., differentiation) among populations and individuals. When populations are *panmictic*, or completely randomly mating, there are no discernable differences in allele frequencies and estimates of gene flow become infinite. However, as noted above, geographic distance plus landscape and environmental variables generally limit gene flow and over multiple generations this will produce specific patterns of population subdivision and genetic structure. These patterns can be evaluated using population-based or individual-based measures of genetic structure or distance, Bayesian clustering analysis, and barrier detection methods.

3.6.1 Population-based measures

The oldest and most widely used population-level metric of genetic structure is Wright's F_{ST}, one of three F-

statistics derived by Sewall Wright. F-statistics were developed to describe the partitioning of genetic variation within a species (Wright 1931, 1951). Three different F-statistics were derived to describe the partitioning of genetic variability among the total population (*T*) or all populations being considered, the subpopulation (*S*) or individuals within a subpopulation (I). F_{ST} is a measure of subpopulation-level genetic differentiation relative to the total population that ranges from 0 (***panmixia***) to 1 (complete genetic isolation) and measures allele frequency divergence among subpopulations:

$$F_{ST} = \frac{H_T - H_S}{H_T}$$

where H_T is the estimated heterozygosity if all subpopulations were randomly mating and H_S is the average heterozygosity among subpopulations. Based on Wright's island model, F_{ST} is estimated to be inversely correlated with the number of migrants per generation using the formula

$$F_{ST} = \frac{1}{1 + 4N_e m}$$

where m is the number of migrants per generation. Due to the unrealistic assumptions of the continent-island model, such as the infinite alleles model and infinite continent population size, however, the utility of this direct calculation of m from F_{ST} has been widely cautioned (Whitlock & MacCauley 1999). The other two F-statistics, F_{IS} and F_{IT}, are thought of as inbreeding estimates within and among populations, respectively, and are also standard population genetic estimates. For more detailed discussion on Wright's F-statistics and their derivations, see Hedrick (2011).

Since Wright, several F_{ST} analogs have been developed to estimate genetic differentiation among populations (Table 3.4). Nei (1972) developed G_{ST}, an extension of F_{ST} that accounts for multiple alleles at a given locus instead of two, as in Wright's original model. Weir and Cockerham's (1984) developed an analysis of variance (ANOVA) approach to estimate an F_{ST} analog theta (θ). Assuming a step-wise mutational model, Slatkin (1995) developed R_{ST} specifically for loci like microsatellites with high numbers of alleles and heterozygosity. However, R_{ST} does not perform well unless the loci strictly conform to the step-wise mutational model (Balloux et al. 2000). Since very few loci conform to the assumptions of this model, this metric is no longer recommended (Meirmans & Hedrick 2011). More recently, additional F_{ST} analogs like G'_{ST} and G''_{ST} and alternative metrics (Jost's *D*) have been developed to improve the performance for data sets with high levels of heterozygosity and/or a small number of populations, or those likely to be out of equilibrium (Hedrick 2005; Meirmans 2006; Jost 2008; Meirmans & Hedrick 2011).

Many additional population-level measures of genetic differentiation, or genetic distance, have been developed and used in contemporary population genetic studies including: Cavalli-Sforza chord distance (D_c) (Cavalli-Sforza & Edwards 1967), Nei's genetic distance (Nei's *D*) (Nei 1972), and Bowcock et al.'s (1994) proportion of shared alleles (D_{ps}). These statistics rely on estimates of heterozygosity or differences in allelic similarity among populations. A more recently derived statistic, conditional genetic distance (or cGD) is based on graph theory and creating networks of populations with edges (connections) and nodes (populations) (Dyer et al. 2010). For a detailed discussion of graph theory and network modeling, see Chapter 10. The choice of a genetic distance statistic is an important decision in landscape genetic studies and results can differ depending on the metric chosen (Lindsay et al. 2008; Pavlacky et al. 2009; Spear & Storfer 2010; Goldberg & Waits 2010), making it important for researchers to understand the assumptions of each metric and evaluate whether results are consistent across multiple metrics. See Table 3.4 for a summary of these statistics and their assumptions. For landscape genetics studies, the most commonly used statistic has been F_{ST} and its analogs followed by *D* and D_c (Storfer et al. 2010). Commonly available free software can be used to estimate different statistics associated with gene flow among populations, including: FSTAT (Goudet 1995), GENEPOP (Rousset 2008), and GENALEX (Peakall & Smouse 2006). A large number of *R* software packages also exist for population genetics calculations, including: gstudio (Dyer 2014), pegas (Paradis et al. 2015), and genetics (Warnes 2015).

3.6.2 Individual-based genetic distance metrics

The population-based genetic distance metrics discussed above provide pairwise estimates of genetic connectivity between populations that consist of groups of individuals. However, for species that are more continuously distributed on the landscape, such as many plants and highly mobile animals (e.g., grizzly

Table 3.4 Estimators of gene flow and genetic differentiation among (a) populations and (b) individuals.

(a) Population-level metrics		
Statistic	**Description and/or assumptions**	**Citations**
F_{ST}	Island model of migration	Wright (1951)
D_c	Chord distance – geometric estimate of genetic distances	Cavalli-Sforza and Edwards (1967)
G_{ST}	Island model, accounts for multiple alleles	Nei (1972)
D (Nei's)	Related to the number of changes per locus, assumes mutation rate is constant among loci. Genetic differences arise among populations due to mutation and genetic drift	Nei and Chesser (1983)
D (Jost's)	Measures the fraction of allelic variation among populations as an absolute rather than a relative measure (such as F_{ST}).	Jost (2008)
Θ (theta)	Accounts for variation in sample size among populations; accounts for error in incomplete population sampling	Weir and Cockerham (1984)
D_{ps}	Proportion of shared alleles – measures similarity in allelic identity among populations	Bowcock et al. (1994)
R_{ST}	Developed especially for microsatellites; accounts for mutation rates at microsatellite loci	Slatkin (1995)
G'_{ST}	Corrects for the bias associated with hypervariable loci, such as microsatellites, and normalizes the standard F_{ST} estimate by dividing by maximum F_{ST}	Hedrick (2005)
G''_{ST}	Corrects for the bias in G'_{ST} associated with sampling a small number of populations; should be used for estimating Hedrick's (2005) standardized measure whenever the number of sampled populations is small	Meirmans and Hedrick (2011)
(b) Individual-level metrics		
a_r	Rousset's *a* assumes that individuals are drawn from a two-dimensional population at migration-drift equilibrium, analogous to the $F_{ST}/(1 - F_{ST})$ ratio using pairs of individuals	Rousset (2000)
D_{ps}	Proportion of shared alleles – measures the degree of similarity in allelic identity among individuals	Bowcock et al. (1994)
r	Relatedness is a measure of identity by descent or shared ancestry between individuals; r is the genetic similarity between two individuals relative to that between random individuals from some reference population. Multiple estimators have been derived	Queller and Goodnight (1989), Smouse and Peakall (1999), Lynch and Ritland (1999)
B_C	Bray Curtis dissimilarity metric was originally derived to measure dissimilarity between species composition and abundance at two sites, but can also be used to measure dissimilarity of alleles between two individuals	Legendre and Legendre (1998)

bears or cougars), it is often more appropriate to use the individual as the unit of analysis when estimating genetic connectivity. A review of landscape genetic studies (Storfer et al. 2010) showed that the most commonly used individual-based genetic distance metrics are pairwise estimates of relatedness or kinship (Queller & Goodnight 1989; Loiselle et al. 1995; Lynch & Ritland 1999; Smouse & Peakall 1999; Wang 2002). Other commonly used metrics include: the proportion of shared alleles (Bowcock et al. 1994), Rousset's a (2000), and Bray Curtis dissimilarity metric (Legendre & Legendre 1998) (Table 3.4). These individual-based genetic distances can be calculated between all (or a subset of) pairs of individuals and then used as a response variable in analyses that evaluate the relationships between genetic distance and geographic or effective landscape distance. Methods to generate individual-based genetic distances between any two individual genotypes are available in software packages such as: SPAGeDI (Hardy & Vekemans 2002), GENALEX (Peakall & Smouse 2006), and ALLELES IN SPACE (Miller 2005).

3.6.3 Bayesian clustering methods

Another approach to detecting genetic structure in population and landscape genetics is Bayesian clustering analysis. In these methods, the individual is the unit of analysis and the number of distinct genetic clusters is inferred from a set of individual multilocus genotypes by estimating genetic ancestry from unobserved source populations. The inferred clusters can then be used for assigning group membership to genotypes and individuals. This methodological approach was introduced by Pritchard et al. (2000) and used to evaluate genetic structure among Taita thrush (*Turdus helleri*) in Africa. This new individual-based approach to evaluating genetic structure and defining populations is implemented in the software, STRUCTURE, and has become one of the most commonly used methods in population and landscape genetic studies. Additional Bayesian clustering methods have been developed that explicitly include spatial locations of individuals in the model such as GENELAND (Guillot et al. 2005), BAPS5 (Corander et al. 2008), and TESS (Chen et al. 2007). These approaches can be particularly effective for defining populations and detecting landscape barriers when genetic differentiation among subpopulations is moderate to high (Latch et al. 2006; Chen et al. 2007; Safner et al. 2011; Blair et al. 2012), but can provide misleading results if samples are not collected evenly across space (Frantz et al. 2009; Schwartz & McKelvey 2009). For a more detailed description of Bayesian clustering and other clustering methods used to evaluate genetic structure see Chapter 7.

3.6.4 Barrier detection methods

In a 2010 review, approximately 35% of all landscape genetics papers had a focus on detecting barriers or genetic breaks among groups of individuals. There is particular interest in identifying barriers from a basic research standpoint to better understand evolution of different populations via genetic structuring (Manel et al. 2003; Storfer et al. 2007, 2010). In addition, identification of barriers is important from a conservation and management perspective because barriers cause genetic isolation among populations and potential loss of evolutionary potential via loss of diversity due to genetic drift. Two main methods used to detect barriers in population and landscape genetics studies are Bayesian clustering methods and edge detection methods (Safner et al. 2011). As discussed above, Bayesian clustering methods determine the most likely groups of genetically related individuals based on their multilocus genotypes. When explicit spatial (i.e., georeferenced) locations of individual genotypes are included in clustering analyses, barriers between resulting clusters of individuals are included. For example, GENELAND and TESS both use tessellation, which creates polygons around genotype locations that partition the study area without leaving gaps. Barriers are then drawn along edges of these polygons where genotypes on one side of the polygon belong to one cluster and genotypes on the other side belong to the other (Guillot et al. 2005).

Edge detection methods are established methods in spatial statistics that can be used to detect barriers among groups of genotypes. For example, Wombling (Womble 1951) and Monmonier's maximum difference algorithm (Monmonier 1973) essentially assign likely barriers to geographic areas where there are maximum differences in allele frequencies or genetic distances over short geographic distances. In addition to population or group level analyses, applications of individual-based estimators of gene flow (Table 3.4) have been developed to help detect barriers among continuously distributed genotypes. For example,

Manel et al. (2007) developed an individual-based method to detecting genetic breaks or discontinuities within populations. This method uses ***assignment tests*** that are applied in a moving window over individuals across a study area to generate a probability surface of finding that genotype across the landscape. Genetic discontinuities are then found by finding the areas of highest slope of the probability surfaces across all individuals (Manel et al. 2007). A related approach uses individual genotypes and their relationship to nearby genotypes to create a heat map of possible membership to one population or another (Murphy et al. 2008). These heat maps can show areas of genetic boundaries where individuals located on either side have a high probability of belonging to one population or another.

A recent comparison of edge detection methods and Bayesian clustering algorithms on both simulated data and actual empirical data showed better performance of Bayesian clustering methods (Safner et al. 2011). Specifically, with small dispersal distances in the simulated data, the software TESS performed best and was able to most accurately detect the number of barriers. At higher dispersal distances, GENELAND had almost 100% accuracy in detecting barriers (Safner et al. 2011). In addition, Bayesian methods were superior to edge detection methods with particular regard to detecting permeable barriers.

3.7 ESTIMATING GENE FLOW USING INDIRECT AND DIRECT METHODS

Before the development of molecular techniques, researchers estimated gene flow indirectly with mark–recapture studies of animals and pollen and seed dispersal studies in plants. These estimates, while measuring direct dispersal rates and exchange of migrants among populations, fail to account for whether these individuals have successfully mated in those new populations. Estimates of gene flow require successful mating in addition to successful dispersal because only then can the dispersing individual's genes influence allele frequencies in the new population. Today, most studies use molecular estimates of gene flow due to the time it takes to conduct a successful mark–recapture study or the difficulty associated with tracking pollen from plant to plant, combined with the rapidly decreasing costs associated with molecular studies. Nonetheless, the combination of molecular data with mark–recapture dispersal estimates make inferences about dispersal most meaningful.

The two main approaches to estimating gene flow among populations using molecular methods can be classified as indirect or direct methods. Indirect methods estimate the number of migrants per generation between populations using different model-based approaches. Many of the indirect methods have been discussed above. When estimating genetic distance, it is assumed that gene flow is inversely related to genetic distance. An explicit calculation for estimating gene flow from Wright's F_{ST} statistic is discussed above. In contrast, direct methods can directly detect migrants or offspring of migrants, and identify the parents of offspring to document direct movement and gene flow. The following sections provide an overview of additional indirect estimates using ***coalescent*** methods, as well as direct methods using assignment tests and parentage analysis to estimate dispersal and gene flow.

3.7.1 Indirect measures of gene flow – coalescent approaches

Gene flow and migration rates among populations can be estimated with coalescent methods. Modern formulations of coalescent theory extend from Wright–Fisher models of identity by descent (Wright 1931; Fisher 1930). These models assume the conditions for Hardy–Weinberg equilibrium, except that population size is finite (N). The models assume that, for each neutral locus, alleles are chosen at random among breeding adults to form the gene pool of the next generation. Thus, in a population of size N, there are $2N$ gene copies and the probability an allele chosen at random in the next generation coming from a particular parent is $1/(2N_e)$.Coalescent theory considers allele frequencies in contemporary populations and back-calculates the time since alleles coalesce to a common ancestor from which they descended. Alleles coming from the same ancestor are thus *identical by descent*, based on the Wright–Fisher model (Kingman 1982; Hudson 1990). In generation 1, the probability of two alleles bring identical by descent is $1/(2N_e)$ and the probability of two alleles *not* being identical by descent is $1 - 1/(2N_e)$. Therefore, after t generations, the probability that any two alleles coalesce is

$$P_c(t) = \left(1 - \frac{1}{2N_e}\right)^{t-1}\left(\frac{1}{2N_e}\right)$$

This model assumes a diploid population with a constant effective population size and no selection. However, more recent mathematical formulations have allowed for selection, recombination, fluctuations in population size, etc. (Kingman 1982; Hudson 1990; Nielsen & Wakeley 2001; Rannala & Yang 2003).

When two populations become separated, allelic differences between them may arise through the process of "lineage sorting", whereby different alleles are chosen at random (assuming selective neutrality) to form the gene pool in each successive generation in different populations (Hudson 1990). In these cases, coalescent theory can thus be used to estimate time since divergence of these populations and gene flow between them.

In modern molecular studies, coalescent theory can be applied in a number of different ways, including estimation of effective population sizes within, as well as gene flow among, populations. Commonly used software packages to accomplish these tasks include MIGRATE-N (Beerli & Palczewski 2010) and IMa2 (Hey & Neilsen 2007; Hey 2010), both of which use MCMC randomization methods to sample gene geneaologies and estimate model parameters with Bayesian inference to estimate posterior probabilities. Note that due to computational intensity, there are limits to using the coalescent approach (Wakeley 2005), which is evident in the fact that IMa2 can only handle up to 10 total populations.

3.7.2 Direct measures – assignment tests

Assignment tests are analytical methods that use genotypic information to ascertain population membership of individuals or groups of individuals. There are two main groups of assignment tests. The first group of methods assumes that population units are known and defined across a landscape. In the original ***likelihood-based*** assignment test (Paetkau et al. 1995), an individual is assigned to its putative population of origin, based on its multilocus genotype and the expected probabilities of that genotype in each of the potential source populations under the assumptions of Hardy–Weinberg and linkage equilibrium. Later derivations of assignment methods also used Bayesian models (Rannala & Mountain 1997) and included a modification that can statistically exclude potential source populations for a given individual rather than attempting to assign the individual to the most likely source population (Cornuet et al. 1999). These methods can be used for direct estimates of dispersal between populations and geographic areas since they can identify individuals that have dispersed away from their natal site. For example, Dixon et al. (2006) used assignment tests to evaluate the effectiveness of a regional corridor in connecting two Florida black bear (*Ursus americanus floridanus*) populations and found that the corridor did provide functional connectivity but that the movement was primarily unidirectional. All of the above methods can be implemented in the software GENECLASS2 (Piry et al. 2004).

The second group of assignment test approaches is the aspatial and spatial Bayesian clustering methods that were discussed above. These methods do not require *a priori* definitions of populations, but population memberships are estimated for each individual in the form of ancestry coefficients for the inferred population units. Thus, these methods can also be used to detect dispersers and offspring with mixed ancestry from two or more genetic groups documenting evidence of gene flow (Manel et al. 2005). Multiple studies have shown that assignment test approaches can be very successful in detecting individuals that disperse between populations (Paetkau et al. 1995; Rannala & Mountain 1997; Manel et al. 2002, 2005; Hauser et al. 2006) and that accuracy improves with increases in (i) the number of loci, (ii) the number of alleles per locus, (iii) the levels of genetic differentiation, and (iv) sample sizes from contributing populations (Cornuet et al. 1999; Paetkau et al. 2004; Berry et al. 2004; Latch & Rhodes 2005; Hauser et al. 2006; Chapter 7).

3.7.3 Parentage analysis

Another direct approach to studying dispersal and gene flow across a landscape is parentage analysis. In this approach, multilocus genotypes are used to determine the parents of progeny from a pool of candidate parents. Two main approaches are used in parentage analysis: parentage exclusion and statistically based parentage assignment (Marshall et al. 1998; Jones & Arden 2003; Jones et al. 2010). For example, if the offspring does not share alleles with the putative parent at one or more loci then this parent can be excluded (Table 3.5). However, exclusion methods are rarely used in natural populations since it often not possible to sample all potential parents, loci may not be powerful enough to exclude all but one parent, and mutation and

Table 3.5 Example of paternity analysis using exclusion. Genotypes are provided for a known mother, offspring, and two possible fathers. Alleles contributed by the mother are highlighted in bold and inferred alleles of the father are in italic. Father 1 can be excluded because of mismatches at loci 1 and 3. Father 2 cannot be excluded as a possible father.

	Locus 1	Locus 2	Locus 3	Locus 4
Mother	122/128	130/134	222/240	188/190
Offspring	**128**/*130*	**134**/*140*	*218*/**222**	**190**/*194*
Father 1	124/126	134/140	220/232	188/194
Father 2	126/130	132/140	216/218	186/194

genotyping errors can cause genotype mismatches between parent and offspring that can exclude true parents. More commonly, a statistically based parentage assignment using maximum likelihood or Bayesian methods is used to determine the most likely mother or father (for a review see Jones & Ardren 2010). In general, 10–15 microsatellite loci and 40–60 SNPs are needed to determine the father when the mother is known and twice this number of loci would be needed if neither parent is known (Allendorf et al. 2012).

While this technique has great potential in landscape genetics, it has been used primarily for plants (see Chapter 10) because it is much easier to sample potential mothers and fathers compared to animals, which can be more difficult to locate and capture. In plants, there are two dispersal processes – seed dispersal and pollen dispersal. Pollen is the paternal contribution to gene flow and dispersal distances with pollen are generally longer than seed dispersal distances, but seed dispersal is critical for demographic connectivity in plants (Sork & Smouse 2006). To quantify pollen sources and pollen movement across the landscape, paternity analysis and two-generation pollen pool structure analysis, known as Two-Gener, are used (Smouse et al. 2001; Smouse & Sork 2004). Paternity analysis can directly reveal the number of fathers and the spatial location of fathers in the landscape. For example, Kamm et al. (2010) genotyped 1183 offspring from 49 mother trees in the rare, insect-pollinated forest tree, *Sorbus domestica*, to determine paternity using maximum likelihood-based parentage assignment (Marshall et al. 1998). In this system, all 167 possible parental trees were known and georeferenced and paternity analysis revealed that 108 trees contributed to the offspring. They also demonstrated that none of the landscape features (settlements, open land, deep valleys, and closed forest) were an impermeable barrier to gene flow from pollen, which moved up to 16 km.

3.8 CONCLUSION AND FUTURE DIRECTIONS

The field of population genetics has a long, rich history that is over a century old. In this chapter, we have only been able to give a brief overview of the field and focus on a small subset of the theory and genetic diversity/population structure methods that are most widely used in landscape genetics. Several full textbooks have been written, including *Genetics of Populations* (Hedrick 2011) and *Principles of Population Genetics* (Hartl & Clark 2006), that provide much more comprehensive overviews of the subject.

As we look to the future, the major challenges in population and landscape genetics involve developing predictive models for changes in landscape genetic structure in response to global change, as well as the challenges in harnessing the full potential of data gathered in next-generation, high-throughput DNA sequencing platforms. With our ability to generate thousands of markers (e.g., SNPs) using high-throughput sequencing, we are now able to assess the distribution of neutral and adaptive genetic variation (Joost et al. 2007; Schoville et al. 2012). Models have shown that we can virtually saturate the statistical power for our ability to detect population genetic structure with approximately 1000 alleles from neutral loci (e.g., Schoville et al. 2012). However, complete genomic sequencing has provided new avenues for understanding the genetic basis of functionally adaptive trait variation. For model organisms such as humans and mice, we now know that structural genomic features, such as copy number

variation, chromosomal inversions, and transposable elements, are also directly involved in functional adaptive trait variation (Fedoroff 2012; Ellegren 2014). A challenge for population genetics may involve development of new theory to incorporate how these features evolve because they behave quite differently from models of point mutations or slippage mutations that are commonly used in models to estimate genetic distance and population structure. Rapid advances in our ability to obtain genomic data have also caused a paradigm shift in the way we view "genes". Once thought to be directly related to phenotype, genes operate in complex genomic landscapes, rather than in isolation. Genes are also expressed differently in different ecological landscapes. The major challenge for landscape genetics in the future is to integrate data from the complex genomic landscape, as well as the ecological landscapes in which individuals and populations exist (see Chapter 9 for more details).

REFERENCES

Adriaensen, F., Chardon, J.P., De Blust, G., Swinnen, E., Villalba, S., Gulinck, H., & Matthysen, E. (2003) The application of "least-cost" modeling as a functional landscape model. *Landscape and Urban Planning* **64**, 233–47.

Allendorf, F.W., Luikart, G.H., & Aitken, S.N. (2012) *Conservation and the Genetics of Populations*. Wiley-Blackwell, Oxford.

Amos, W., Worthington Wilmer, J., Fullard, K., Burg, T.M., Croxall, J.P., Bloch, D., & Coulson, T. (2001) The influence of parental relatedness on reproductive success. *Proceedings of the Royal Society London B* **268**, 2021–7.

Aparacio, J.M., Ortego, J., & Cordero, P.J. (2006) What should we weigh to estimate heterozygosity, alleles or loci? *Molecular Ecology* **15**, 4459–6.

Balloux, F., Brünner, H., Lugon-Moulin, N., Hausser, J., & Goudet, J. (2000) Microsatellites can be misleading: an empirical and simulation study. *Evolution* **54**, 1414–22.

Beerli, P. & Palczewski, M. (2010) Unified framework to evaluate panmixia and migration direction among multiple sampling locations. *Genetics* **185**, 313–26.

Berry, O., Tocher, M.D., & Sarre, S.D. (2004) Can assignment tests measure dispersal? *Molecular Ecology* **13**, 551–61.

Birky, C.W. (1978) Transmission genetics of mitochondria and chloroplasts. *Annual Review Genetics* **12**, 471–512.

Blair, C., Weigel, D.E., Balazik, M., Keeley, A.T.H., Walker, F.M., Landguth, E.L., Cushman, S.A., Murphy, M., Waits, L.P., & Balkenhol, N. (2012) A simulation-based evaluation of methods for inferring linear barriers to gene flow. *Molecular Ecology Resources* **12**, 822–33.

Bonin, A., Ehrich, D., & Manel, S. (2007) Statistical analysis of amplified fragment length polymorphism data: a toolbox for molecular ecologists and evolutionists. *Molecular Ecology* **16**, 3737–58.

Bowcock, A.M., Ruiz-Linares, A., Tomfohrde, J., Minch, E., Kidd, J.R., & Cavalli-Sforza, L.L. (1994) High resolution of human evolutionary trees with polymorphic microsatellites. *Nature* **368**, 455–7.

Cavalli-Sforza, L.L. & Edwards, A.W.F. (1967) Phylogenetic analysis – models and estimation procedures. *The American Journal of Human Genetics* **19**, 233–257.

Chen, C., Durand, E. Forbes, F., & François, O. (2007) Bayesian clustering algorithms ascertaining spatial population structure: a new computer program and a comparison study. *Molecular Ecology Notes* **7**, 747–56.

Coltman, D.W., Pilkington, J.G., Smith, J.A., & Pemberton, J.M. (1999) Parasite-mediated selection against inbred Soay sheep in a free-living, island population. *Evolution* **53**, 1259–67.

Corander, J., Marttinen, P., Sirén, J., & Tang, J. (2008) Enhanced Bayesian modeling in BAPS software for learning genetic structures of populations. *BMC Bioinformatics* **9**, 539.

Cornuet, J.M., Piry, S., Luikart, G., Estoup, A., & Solignac, M. (1999) New methods employing multilocus genotypes to select or exclude populations as origins of individuals. *Genetics* **153**, 1989–2000.

Coulon, A. (2010) GENHET: an easy-to-use *R* function to estimate individual heterozygosity. *Molecular Ecology Resources* **10**, 167–9.

Crandall, K.A., Posada, D., & Vasco, D. (1999) Effective population sizes: missing measures and missing concepts. *Animal Conservation* **2**, 317–19.

Di Rienzo, A., Peterson, A.C., Garza, J.C., Valdes, A.M., Slatkin, M., & Freimen, N.B. (1994) Mutational processes of simple-sequence repeat loci in human populations. *Proceedings of the National Academies of Sciences USA* **91**, 3166–70.

Dixon, J.D., Oli, M.K., Wooten, M.C., Eason, T.H., McCown, J.W., & Paetkau, D. (2006) Effectiveness of a regional corridor in connecting two florida black bear populations. *Conservation Biology* **20**, 155–62.

Dyer, R.J. (2014) gstudio: analyses and functions related to the spatial analysis of genetic marker data. R package, version 1.3.

Dyer, R.J., Nason, J.D., & Garrick, R.C. (2010) Landscape modeling of gene flow: improved power using conditional genetic distance derived from the topology of population networks. *Molecular Ecology* **19**, 3746–59.

El Mousadik, A. & Petit, R.J. (1996) High level of genetic differentiation for allelic richness among populations of the argantree (*Argania spinosa* (L.) Skeels) endemic to Morocco. *Theoretical and Applied Genetics* **92**, 832–9.

Ellegren, H. (2014) Genome sequencing and population genomics in non-model organisms *Trends in Ecology and Evolution* **29**, 51–63.

Fedoroff, N. (2012) Transposable elements, epigenetics, and genome evolution. *Science* **338**, 758–67.

Fisher, R.A. (1930) *The Genetical Theory of Natural Selection*. Oxford University Press, New York.

Frankham, R. (1995) Effective population size/adult population size ratios in wildlife: a review. *Genetics Research* **66**, pp. 95–107.

Frankham, R., Brook, B.W., Bradshaw, C.J.A., Traill, L.W., & Spielman, D. (2013) 50/500 rule and minimum viable populations: response to Jamieson and Allendorf. *Trends in Ecology and Evolution* **28**, 187–8.

Franklin, I.R. (1980), Evolutionary change in small populations. In: Soulé, M.E. & Wilcox, B.A. (eds.), *Conservation Biology: An Evolutionary Ecological Perspective*. Sinauer Associates, Sunderland, MA.

Frantz, A.C., Cellina, S., Krier, A., Schley, L., & Burke, T. (2009) Using spatial Bayesian methods to determine the genetic structure of a continuously distributed population: clusters or isolation-by-distance? *Journal of Applied Ecology* **46**, 493–505.

Garvin, M.R., Saitoh, K., & Gharrett, A.J. (2010) Application of single nucleotide polymorphisms to non-model species: a technical review. *Molecular Ecology Resources* **10**, 915–34.

Gillham, N.W. (1974) Genetic analysis of the chloroplast and mitochondrial genomes. *Annual Review Genetics* **8**, 347–91.

Goldberg, C.S. & Waits, L.P. (2010) Comparative landscape genetics of two pond-breeding amphibian species in a highly modified agricultural landscape. *Molecular Ecology* **19**, 3650–63.

Goudet, J. (1995) FSTAT (Version 1.2): A computer program to calculate F-statistics. *Journal of Heredity* **86**, 485–6.

Guillot, G., Mortier, F., & Estoup, A. (2005) Geneland: a program for landscape genetics. *Molecular Ecology Notes* **5**, 712–15.

Hardy, G.H. (1908) Mendelian proportions in a mixed population. *Science* **28**, 49–50.

Hardy, O.J. & Vekemans, X. (2002) Spagedi: a versatile computer program to analyze spatial genetic structure at individual or population levels. *Molecular Ecology Notes* **2**, 618–20.

Hare, M.P., Nunney, L., Schwartz, M.K., Ruzzante, D.E., Burford, M., Waples, R.S., Ruegg, K., & Palstra, F. (2011) Understanding and estimating effective population size for practical application in marine species management. *Conservation Biology* **25**, 438–49.

Hauser, L., Seamons, T.R., Dauer, M., Naish, K.A., & Quinn, T.P. (2006) An empirical verification of population assignment methods by marking and parentage data: hatchery and wild steelhead (*Oncorhynchus mykiss*) in Forks Creek, Washington, USA. *Molecular Ecology* **15**, 3157–73.

Hartl, D.L. & Clark, A. G. (2006) *Principles of Population Genetics*. Sinauer Associates, Sunderland, MA.

Hedrick, P.W. (1995) Gene flow and genetic restoration: the Florida panther as a case study. *Conservation Biology* **9**, 996–1007.

Hedrick, P.W. (2005) A standardized genetic differentiation measure. *Evolution* **59**, 1633–8.

Hedrick, P.W. (2011) *Genetics of Populations*, 4th edition. Jones & Bartlett Publishers, Boston, MA.

Hey J. (2010) Isolation with migration models for more than two populations. *Molecular Biology and Evolution* **27**, 905–20.

Hey, J. & Nielsen, R. (2007) Integration within the Felsenstein equation for improved Markov chain Monte Carlo methods in population genetics. *Proceedings of the National Academy of Sciences USA* **104**, 2785–90.

Hoekstra, H.E., Hirschmann, R.J., Bundney, R.A., Insel, P.A., & Crossland, J.P. (2006) A single amino acid mutation contributes to adaptive beach mouse color pattern. *Science* **313**, 101–4.

Holderegger, R. & Wagner, H.H. (2008) Landscape genetics. *BioScience* **58**, 199–207.

Hudson, R.R. (1990) Gene genealogies and the coalescent process. In: Futuyma, D. & Antonovics, J. (eds.), *Oxford Surveys in Evolutionary Biology*, Vol. **7**. Oxford University Press, New York.

Jamieson, I.G. & Allendorf, F.W. (2012) How does the 50/500 rule apply to MVPs? *Trends in Ecology and Evolution* **27**, 578–84.

Jones, A.G. & Ardren, W.R. (2003) Methods of parentage analysis in natural populations. *Molecular Ecology* **12**, 2511–23.

Jones, A.G., Small, C.M., Paczolt, K.A., & Ratterman, N.L. (2010) A practical guide to methods of parentage analysis. *Molecular Ecology Resources* **10**, 6–30.

Joost, S., Bonin, A., Bruford, M.W., Després, L., Conord, C., Erhardt, G., & Taberlet, P. (2007) A spatial analysis method (SAM) to detect candidate loci for selection: towards a landscape genomics approach to adaptation. *Molecular Ecology* **16**, 3955–69.

Jost, L. (2008) GST and its relatives do not measure differentiation. *Molecular Ecology* **17**, 4015–26.

Kalinowski, S.T. (2004) Counting alleles with rarefaction: private alleles and hierarchical sampling designs. *Conservation Genetics* **5**, 539–43.

Kamm, U., Gugerli, F., Rotach, P., Edwards, P.J. & Holderegger, R. (2010) Open areas in a landscape enhance pollen-mediated gene flow of a tree species: evidence from northern Switzerland. *Landscape Ecology* **25**, 903–11.

Kimura, M. (1968) Evolutionary rate at the molecular level. *Nature* **217**, 624–6.

Kimura, M. & Weiss, G.H. (1964) The stepping stone model of population structure and the decrease of genetic correlation with distance. *Genetics* **49**, 561–76.

Kingman, J.F.C. (1982) The coalescent. *Stochastic Processes and Their Applications* **13**, 235–48.

Lande, R. (1995) Mutation and conservation. *Conservation Biology* **9**, 782–91.

Latch, E.K. & Rhodes, O.E. Jr. (2005) The effects of gene flow and population isolation on the genetic structure of

reintroduced wild turkey populations: Are genetic signatures of source populations retained? *Conservation Genetics* **6**, 981–97.

Latch, E.K., Dharmarajan, G., Glaubitz, J.C., & Rhodes, O.E. Jr. (2006) Relative performance of Bayesian clustering software for inferring population substructure and individual assignment at low levels of population differentiation. *Conservation Genetics* **7**, 295–302.

Leberg, P.L. (2002) Estimating allelic richness: effects of sample size and bottlenecks. *Molecular Ecology* **1**, 2445–9.

Leberg, P.L. (2005) Genetic approaches for estimating the effective size of populations. *Journal of Wildlife Management* **69**, 1385–99.

Lee, C.-R. & Mitchell-Olds, T. (2011) Complex trait divergence contributes to environmental niche differentiation in ecological speciation of *Boechera stricta*. *Molecular Ecology* **22**, 2204–17.

Legendre, P. & Legendre, L. (1998) *Numerical Ecology*. Elsevier, Amsterdam.

Lindsay, D.L., Barr, K.R., Lance, R.F., Tweddale, S.A., Hayden, T.J., & Leberg, P.L. (2008) Habitat fragmentation and genetic diversity of an endangered, migratory songbird, the golden-cheeked warbler (*Dendroica chrysoparia*). *Molecular Ecology* **17**, 2122–33.

Linnen, C.R., Kingsley, E.P., Jensen, J.D., & Hoekstra, H.E. (2009) On the origin and spread of an adaptive allele in deer mice. *Science* **325**, 1095–8.

Loiselle, B.A., Sork, V.L., Nason, J.D., & Graham, C.H. (1995) Spatial genetic structure of a tropical understory shrub, *Psychotria officinalis* (Rubiaceae). *American Journal of Botany* **82**, 1420–5.

Lowe, A.J., Harris, S.A., & Ashton, P.A. (2004) *Ecological Genetics; Design, Analysis and Application*. Wiley-Blackwell, Chichester.

Lynch, M. & Lande, R. (1998) The critical effective size for a genetically secure population. *Animal Conservation* **1**, 701.

Lynch, M. & Ritland, R. (1999) Estimation of pairwise relatedness with molecular markers. *Genetics* vol. **152**, 1753–66.

Manel S., Berthier P., & Luikart G. (2002) Detecting wildlife poaching: identifying the origin of individuals using Bayesian assignment tests and multi-locus genotypes. *Conservation Biology* **16**, 650–9.

Manel, S., Schwartz, M.K., Luikart, G., & Taberlet, P. (2003) Landscape genetics: combining landscape ecology and population genetics. *Tree* **18**, 189–97.

Manel S., Gaggiotti O. & Waples R. (2005) 'Assignment methods: matching biological questions with appropriate techniques', *Trends in Ecology and Evolution* **20**, 136–42.

Manel, S., Berthoud, F., Bellemain, E., Gaudeul, M., Luikart, G., Swenson, J.E., Waits, L.P., & Taberlet, P. (2007) A new individual-based spatial approach for identifying genetic discontinuities in natural populations. *Molecular Ecology* **16**, 2031–43.

Mantel, N. (1967) The detection of disease clustering and a generalized regression approach. *Cancer Research* **27**, 209–20.

Marshall, T.C., Slate, J., Kruuk, L.E.B., & Pemberton, J.M. (1998) Statistical confidence for likelihood-based paternity inference in natural populations. *Molecular Ecology* **7**, 639–55.

McRae, B.H. (2006) Isolation by resistance. *Evolution* **60**, 1551–61.

Meirmans, P.G. (2006) Using the AMOVA framework to estimate a standardized genetic differentiation measure. *Evolution* **60**, 2399–402.

Meirmans, P.G. & Hedrick, P.W. (2011) Assessing population structure: FST and related measures. *Molecular Ecology Resources* **11**, 5–18.

Miller, M.P. (2005) Alleles in space (AIS): computer software for the joint analysis of interindividual spatial and genetic information. *Journal of Heredity* **96**, 722–4.

Monmonier, M.S. (1973) Maximum-difference barriers: an alternative numerical regionalization method. *Geographical Analysis* **5**, 245–61.

Morin, P.A., Luikart, G., & Wayne, R.K. (2004) SNPs in ecology, evolution and conservation. *Trends in Ecology and Evolution* **19**, 208–16.

Mullis, K.B. & Faloona, F.A. (1987) Specific synthesis of DNA *in vitro* via a polymerase-catalyzed chain reaction. *Methods in Enzymology* **155**, 335–50.

Murphy, M.A., Evans, J.S., Cushman, S.A., & Storfer, A. (2008) Representing genetic variation as continuous surfaces: an approach for identifying spatial dependency in landscape genetic studies. *Ecography* **31**, 685–97.

Neale, D.B., Marshall, K.A., & Sederoff, R.R. (1989) Chloroplast and mitochondrial DNA are paternally inherited in *Sequia sempervirens* D. Don Endl. *Proceedings of the National Academy of Sciences USA* **86**, 9347–9.

Nei, M. (1972) Genetic distance between populations. *American Naturalist* **106**, 283–92.

Nei, M. (1978) Estimation of heterozygosity and genetic distance from small number of individuals. *Genetics* **89**, 583–90.

Nei, M. & Chesser, R.K. (1983) Estimation of fixation indices and gene diversities. *Annals of Human Genetics* **47**, 253–9.

Nielsen, R. & Wakeley, J. (2001) Distinguishing migration from isolation: a Markov chain Monte Carlo approach. *Genetics* **158** (June 1, 2001), 885–96.

Ohta, T. (1973) Slightly deleterious mutant substitutions in evolution. *Nature*, **246**, 96–8.

Ohta T. & Kimura, M. (1973) A model of mutation appropriate to estimate the number of electrophoretically detectable alleles in a finite population. *Genetical Research* **33**, 201–4.

Paetkau, D., Calvert, W., Stirling, I., & Strobeck C. (1995) Microsatellite analysis of population structure in Canadian polar bears. *Molecular Ecology* **4**, 347–54.

Paetkau, D., Slade, R., Burden, M., & Estoup, A. (2004) Genetic assignment methods in direct, real-time estimation of migration rate: a simulation-based exploration of accuracy and power. *Molecular Ecology* **13**, 55–65.

Paradis, E., Schliep, K., Potts, A., & Winter, D. (2015) Pegas – population and evolutionary genetics analysis system. R package, version 0.6.

Pavlacky, D.C., Goldizen, A.W., Prentis, P.J., Nicholls, J.A., & Lowe, A.J. (2009) A landscape genetics approach for quantifying the relative influence of historic and contemporary habitat heterogeneity on the genetic connectivity of a rainforest bird. *Molecular Ecology* **18**, 2945–60.

Peakall, R. & Smouse, P.E. (2006) Genalex 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes* **6**, 288–95.

Piry, S., Alapetite, A., Cornuet, J.-M., Paetkau, D., Baudouin, L., & Estoup, A. (2004) Geneclass2: a software for genetic assignment and first-generation migrant detection. *Journal of Heredity* **95**, 536–9.

Pritchard, J.K., Stephens, M., & Donnelly, P. (2000) Inference of population structure using multilocus genotype data. *Genetics* **155**, 945–59.

Prunier, J.G., Kaufmann, B., Fenet, S., Picard, D., Pompanon, F., Joly, P., & Lena, J.P. (2013) Optimizing the trade-off between spatial and genetic sampling efforts in patchy populations: towards a better assessment of functional connectivity using an individual-based sampling scheme. *Molecular Ecology* **22**, 5516–30.

Queller, D.C. & Goodnight, K.F. (1989) Estimating relatedness using molecular markers. *Evolution* **43**, 258–75.

Rannala, B. & Mountain, J.L. (1997) Detecting migration using multilocus genotypes. *Proceedings of the National Academy of Sciences USA* **94**, 9197–201.

Rannala, B. & Yang, J. (2003) Bayes estimation of species divergence times and ancestral population sizes using DNA sequences from multiple loci. *Genetics* **164**, 1645–56.

Rousset, F. (2000) Genetic differentiation between individuals. *Journal of Evolutionary Biology* **13**, 58–62.

Rousset, F. (2008) Genepop'007: a complete reimplementation of the Genepop software for Windows and Linux. *Molecular Ecology Resources* **8**, 103–6.

Safner, T., Miller, M.P., McRae, B.H., Fortin, M.J., & Manel, S. (2011) Comparison of Bayesian clustering and edge detection methods for inferring boundaries in landscape genetics. *International Journal of Molecular Sciences* **12**, 865–89.

Sanger, F., Nicklen, S., & Coulson, A.R. (1977) DNA sequencing with chain-terminating inhibitors. *Proceedings of the Royal Academy of Sciences USA* **74**, 5463–7.

Schoville, S.D., Bonin, A., François, O., Lobreaux, S., Melodelima, C., & Manel, S. (2012) Adaptive genetic variation on the landscape: methods and cases. *Annual Review of Ecology, Evolution and Systematics* **43**, 23–43.

Schwartz, M.K. & McKelvey, K.S. (2009) Why sampling scheme matters: the effect of sampling scheme on landscape genetic results. *Conservation Genetics* **10**, 441–52.

Shendure, J. & Ji, N. (2008) Next-generation DNA sequencing. *Nature Biotechnology* **26**, 1135–45.

Shirk, A.J. & Cushman, S.A. (2011) sGD software for estimating spatially explicit indices of genetic diversity. *Molecular Ecology Resources* **11**, 923–34.

Slatkin, M. (1995) A measure of population subdivision based on microsatellite allele frequencies. *Genetics* **139**, 457–62.

Smouse, P.E. & Peakall, R. (1999) Spatial autocorrelation analysis of individual multiallele and multilocus genetic structure. *Heredity* **82**, 561–73.

Smouse, P.E. & Sork, V.L. (2004) Measuring pollen flow in forest trees: an exposition comparison of alternative approaches. *Forest Ecology and Management* **197**, 21–38.

Smouse, P.E., Dyer, R.J., Westfall, R.D., & Sork, V.L. (2001) Two-generation analysis of pollen flow across a landscape I: Male gamete heterogeneity among females. *Evolution* **55**, 260–71.

Sork, V.L. & Smouse, P.E. (2006) Genetic analysis of landscape connectivity in tree populations. *Landscape Ecology* **21**, 821–36.

Sork, V.L. & Waits, L.P. (2010) Contributions of landscape genetics – approaches, insights and future potential. *Molecular Ecology* **19**, 3489–95.

Spear, S.F. & Storfer, A. (2010) Anthropogenic and natural disturbance lead to differing patterns of gene flow in the Rocky Mountain tailed frog *Ascaphus montanus*. *Biological Conservation* **143**, 778–86.

Spear, S. F., Balkenhol, N., Fortin, M.-J., McRae, B. H., & Scribner, K. T. (2010) Use of resistance surfaces for landscape genetic studies: considerations for parameterization and analysis. *Molecular Ecology* **19**, 3576–91.

Storfer, A., Murphy, M.A., Evans, J.S., Goldberg, C.S., Robinson, S., Spear, S.F., Dezzani, R., Delmelle, E., Vierling, L., & Waits, L.P. (2007) Putting the "landscape" in landscape genetics. *Heredity* **98**, 128–42.

Storfer, A., Murphy, M.A., Spear, S.F., Holderegger, R., & Waits, L.P. (2010) Landscape genetics: Where are we now? *Molecular Ecology* **19**, 3496–514.

Vos, P., Hogers, R., Bleeker, M., Reijans, M., van de Lee, T., Hornes, M., Frijters, A., Pot, J. Peleman, J. Kuiper, M., & Zabeau, M. (1995) AFLP: a new technique for DNA fingerprinting. *Nucleic Acid Research* **23**, 4407–14.

Wakeley, J. (2005) The limits of theoretical population genetics. *Genetics* **169**, 1–7.

Wang, J. (2002) An estimator for pairwise relatedness using molecular markers. *Genetics* **160**, 1203–15.

Wang, J. (2005) Estimation of effective population sizes from data on genetic markers. *Philosophical Transactions of the Royal Society B* **360**, 1395–409.

Wang, I.J. (2013) Examining the full effects of landscape heterogeneity on spatial genetic variation: a multiple matrix regression approach for quantifying geographic and ecological isolation. *Evolution* **67**, 3403–11.

Wang, I.J., Glor, R.E., & Losos, J.B. (2013) Quantifying the roles of ecology and geography in spatial genetic divergence. *Ecology Letters* **16**, 175–82.

Warnes, G. (2015) Genetics – population genetics. R package, version 1.3.8.1.

Weinberg, W. (1908) Über den Nachweis der Vererbung beim Menschen. *Jahreshauptversammlung des Vereins für Vaterländische Naturkunde in Württemberg* **64**, 369–82.

Weir, B.S. & Cockerham, C.C. (1984) Estimating F-statistics for the analysis of population structure. *Evolution* **38**, 1358–70.

Whitlock, M.C. & McCauley, D.E. (1999) Indirect measures of gene flow and migration: FST doesn't equal $1/(4Nm + 1)$. *Heredity* **82**, 117–25.

Wolff, J.N., Nafisinia, M., Sutivsky, P., & Ballard, J.W.O. (2013) Paternal transmission of mitochondrial DNA as an integral part of mitochondrial inheritance in metapopulations of *Drosophila simulana*. *Heredity* **110**, 57–62.

Womble, W.H. (1951) Differential systematics. *Science* **114**, 315–22.

Wright, S. (1931) Evolution in Mendelian populations. *Genetics* **16**, 97–159.

Wright, S. (1938) Size of population and breeding structure in relation to evolution. *Science* **87**, 430–1.

Wright, S. (1943) Isolation-by-distance. *Genetics* **28**, 114–38.

Wright, S. (1951) The genetical structure of populations. *Annals of Eugenics*. **15**, 323–54.

Wright, S. (1969) *Evolution and the Genetics of Populations*, Vol. **II**. University of Chicago Press, Chicago.

Zhao, H., Li, R., Wang, Q., Yan, Q., Deng, J.H., Han, D., Bai, Y., Young, W.Y., & Guan, M.X. (2004) Maternally inherited aminoglycoside-induced and nonsyndromic deafness is associated with the novel C1494T mutation in the mitochondrial 12S rRNA gene in a large Chinese family. *American Journal of Genetics* **74**, 139–52.

Zouros, E. (2000) The exceptional mitochondrial DNA system of the mussel family Mytilidae. *Genes and Genetic Systems* **75**, 313–18.