Title
Linked Landscapes: Metapopulation Connectivity of Secretive Wetland Birds

Author
Hall, Laurie

Publication Date
2015

Peer reviewed|Thesis/dissertation

Linked Landscapes: Metapopulation Connectivity of Secretive Wetland Birds

By

Laurie Anne Hall

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Environmental Science, Policy, and Management

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Steven R. Beissinger, Chair
Professor Justin S. Brashares
Professor Todd E. Dawson
Professor Rauri C. Bowie

Fall 2015

Linked Landscapes: Metapopulation Connectivity of Secretive Wetland Birds

By Laurie Anne Hall

Abstract

Linked Landscapes: Metapopulation Connectivity of Secretive Wetland Birds

by

Laurie Anne Hall

Doctor of Philosophy in Environmental Science, Policy, and Management

University of California, Berkeley

Professor Steven R. Beissinger, Chair

Dispersal and gene flow connect habitat patches, linking demographic and genetic processes within metapopulations. Demographic connectivity can prevent local extinctions and increase species persistence. It allows species to colonize new areas, making it possible for them to shift their ranges in response to changes in the environment. Genetic connectivity may impede local adaptation, but is often advantageous because it minimizes the negative effects of habitat loss and fragmentation, such as inbreeding.

This research explored different methodologies for estimating dispersal and gene flow, and estimated connectivity for two species of secretive wetland birds, the California black rail (*Laterallus jamaicensis coturniculus*) and the Virginia rail (*Rallus limicola*), from an inland metapopulation in the Sierra Nevada foothills and a coastal metapopulation in the San Francisco Bay Area, California. The research presented includes: (1) a synthesis of landscape genetics methods used to examine spatial patterns of genetic variation, (2) a validation of dispersal distances estimated from occupancy models using dispersal distances estimated from genetic parentage assignments within the inland metapopulation, (3) an estimate of the frequency and distance of long-distance dispersal events between the inland and coastal metapopulations, and (4) an assessment of the effects of habitat loss and fragmentation on gene flow within the coastal metapopulation.

Despite its importance in ecology and evolution, dispersal distances are poorly understood for many species, especially those that are secretive and rare, such as rails. For these species, indirect methods, including the use of occupancy models, genetics, and isotopic markers, may be optimal for estimating dispersal. In the Sierra Nevada foothills, Black and Virginia rails exhibited contrasting relationships between connectivity metrics and patch colonization. For black rails, model-averaged dispersal distances from occupancy models (Buffer Radius Metric = 3.46 km; Incidence Function Metric = 3.70 km) showed good agreement with the mean dispersal distance from genetic parentage assignments (5.58 ± 1.92 km). For Virginia rails, however, it was difficult to identify a spatial scale with the best fit in occupancy models, and the sample size for estimating dispersal distance from parentage assignments was limited. Combined inference from genetic and isotopic population assignments suggested that long-distance dispersal of black rails between the Sierra Nevada foothills and the San Francisco Bay

Area occurred infrequently, but, somewhat surprisingly, three dispersal events greater than 100 km were recorded during the study. The first was recorded from a band return for a black rail that traveled 128 km. Two other long-distance dispersal estimates were obtained using isotopic and genetic population assignments. Within the San Francisco Bay Area, estimates of gene flow for black rails among 11 wetlands (F_{ST} range: 0.014 to 0.067) indicated rails dispersed frequently between wetlands at shorter spatial scales (tens of kilometers), but that gene flow at a larger spatial scale (greater than 50 km) between the North Bay and South Bay (F_{ST} = 0.018) was limited. Although previously thought to be dispersal-limited, black rails appear to be capable of dispersing and maintaining some level of gene flow at spatial scales less than 50 km. Therefore, wetland mitigation and restoration efforts for rails should focus on protecting and creating habitat at smaller spatial scales (tens of kilometers) to maintain demographic and genetic connectivity and metapopulation viability.

For Jerry

Table of Contents

Acknowledgements

I would like to extend my sincerest thanks to Jerry Tecklin, to whom this dissertation is dedicated. Jerry is a wonderful mentor and a dear friend who taught me to think like a black rail. Because of him, I never leave home without my playback equipment, and every time I see a small patch of wetland I wonder if it might be teeming with rails.

I am grateful to my advisor, Steve Beissinger, for his mentorship. Steve always made time to talk with me when I knocked on his door, he challenged me to be a better scientist, and he taught me that persistence has its rewards. I would also like to thank my qualifying exam and dissertation committee members C. Mortiz, C. Kremen, T. Dawson, J. Brashares, and R. Bowie for their helpful comments and support. Assistance with stable isotope analysis was given by S. Mambelli, P. Brooks, and W. Yang. My colleagues in the Beissinger Lab and in other research groups at the University of California, Berkeley, including: C. Tarwater, J.P. Kelley, A. Lee, H. Streby, P. Epanchin, T.L. Morelli, M. Tingley, O. Richmond, K. Iknayan, N. Van Schmidt, K. Labarbera, R. Romero, and H. Sardiñas, provided thoughtful feedback and support throughout my dissertation research.

I am so fortunate to have an amazing community of family and friends that provided support (both financially and emotionally) during my dissertation research. I would like to thank my mother for always being proud of me; my father for sharing his love of nature and science with me; and my brother, Scott, for providing IT support, for traveling with me, and for paying for everything while traveling with me. I owe one thousand thank yous to Mike Kiselow for loving me “no less” while I struggled with some of the challenges of graduate school. To my friends T. Graham, H. Steklis, H. Robinson, J. Schuytema, K. Negrey, C. Young, C. Gibble, J. Sweeney, R. Romero, S. Ainsley, L. Thompson, and A. McAleavey, thank you for helping shape who I am both personally and professionally. I love you all dearly.

An armada of undergraduate assistants and field technicians including: C. Cleveland, C. Stegall, N. Najar, T. Nguyen, C. Chu, K. Lovett, J. Bruce, M. Goodnow, B. Gartland, E. Hunter, L. Evans, and L. Doll contributed countless hours to help me collect data. I would especially like to thank Maybellene Gamboa for her hard work and wonderful sense of humor.

A number of agencies in the San Francisco Bay area granted permission to access study sites including: The U.S. Fish and Wildlife Service, California Dept. of Fish and Wildlife, California Department of Water Resources, National Park Service, California State Parks, and East Bay Regional Parks. Thank you to C. Strong, J. Krause, J. Hitchen, D. Williams, and B. Becker for facilitating field work at these sites. In addition, many Bay Area scientists assisted with rail surveys and captures including: T. Graham, K. Spragens, V. Bui, J. Wood, J. Takekawa, J. McBroom, and J. Evens.

Many private landowners provided access to study sites in the Sierra Foothills. This research would not have been possible without their support. In addition, the California Dept. of Fish and Wildlife, City of Chico, Beale Airforce Base, and the Wildlife Heritage Foundation all permitted access to study sites. Thank you to D. Effseaf, J. Drew, T. Caldwell, D. Whitmore, C. Carroll, C. Harrison, M. Carroll, P. Shea for facilitating field work at these sites.

Financial support for this research was provided by the National Science Foundation DEB-1051342 and CNH 1115069, the A. Starker Leopold Chair funds awarded to S. Beissinger, and the Kathleen S. Anderson Award from the Manomet Center for Conservation Sciences awarded to L. Hall.

Chapter 1. A practical toolbox for design and analysis of landscape genetics studies

This article has been previously published and is reproduced here with permission from the publisher, Springer:

Hall, L.A., and S.R. Beissinger. 2014. A practical toolbox for design and analysis of landscape genetics studies. Landscape Ecology 29:1487-1504.

Abstract

Landscape genetics integrates theory and analytical methods of population genetics and landscape ecology. Research in this area has increased in recent decades, creating a plethora of options for study design and analysis. Here we present a practical toolbox for the design and analysis of landscape genetics studies following a seven-step framework: (1) define the study objectives, (2) consider the spatial and temporal scale of the study, (3) design a sampling regime, (4) select a genetic marker, (5) generate genetic input data, (6) generate spatial input data, and (7) choose an analytical method that integrates genetic and spatial data. Study design considerations discussed include choices of spatial and temporal scale, sample size and spatial distribution, and genetic marker selection. We present analytical methods suitable for achieving different study objectives. As emerging technologies generate genetic and spatial data sets of increasing size, complexity, and resolution, landscape geneticists are challenged to execute hypothesis-driven research that combines empirical data and simulation modeling. The landscape genetics framework presented here can accommodate new design considerations and analyses, and facilitate integration of genetic and spatial data by guiding new landscape geneticists through study design and implementation.

Keywords: Landscape genetics, Landscape genomics, Population genetics, Landscape ecology, Connectivity, Genetic distance, Resistance surface, Network model, Gravity model, Least-cost path

Introduction

Understanding how genetic variation of a species responds to changes in a landscape is enhanced through the integration of genetic and spatial data. This approach, typically called landscape genetics, examines the microevolutionary processes driving the distribution of genetic variation across landscapes (Manel et al. 2003; Holderegger and Wagner 2008; Balkenhol et al. 2009a; Segelbacher et al. 2010; Manel and Holderegger 2013). The effect of landscapes on genetic variation has long been recognized (Wright 1943; Dobzhansky 1947), but use of spatially-explicit analyses that integrate theory from population genetics and landscape ecology have grown rapidly over the past decade (Holderegger and Wagner 2006; Storfer et al. 2010).

The multidisciplinary nature and recent rapid growth of landscape genetics has led to an overwhelming number of options for study design and analysis (Storfer et al. 2006; Balkenhol et al. 2009b; Anderson et al. 2010; Spear et al. 2010). New researchers and managers are challenged to evaluate the design, implementation, and interpretation of landscape genetics studies. Previous reviews of landscape genetics have been limited to a single component of

design or implementation, such as analytical methods (Storfer et al. 2006; Balkenhol et al. 2009b; Wagner and Fortin 2013), spatial and temporal considerations (Anderson et al. 2010), and creation of resistance surfaces (Spear et al. 2010).

Here we provide a practical toolbox for the design and analysis of landscape genetics studies. We review the seven steps required for the execution of a landscape genetics study: (1) define the study objectives, (2) consider the spatial and temporal scale of the study, (3) design a sampling regime, (4) select a genetic marker, (5) generate genetic input data, (6) generate spatial input data, and (7) choose an analytical method that integrates genetic and spatial data (Fig. 1). For each step we summarize important design considerations for landscape genetics studies to guide new researchers and managers through study design and implementation.

Overview of the framework

Landscape genetics studies can be designed and implemented in seven steps (Fig. 1) linked by the flow of information between them. Choices made during each step are affected by decisions from previous steps, and by the life history and demographic characteristics of the study organism. First, study objectives must be defined (Step 1) to assess whether gene flow or selection will be measured. Next, the study is designed by selection of an appropriate spatial and temporal scale (Step 2), a sampling design (e.g., number and spatial distribution of genetic samples collected and selection of landscape or environmental characteristics of interest; Step 3), and a genetic marker (Step 4).

Landscape genetics studies require collection and analysis of genetic and spatial data (Steps 5 and 6). For some studies, data will first be analyzed separately to generate genetic and spatial measures that serve as input data for the final landscape genetics analysis, while other studies may utilize raw data such as genotypes or climate data. The final analysis (Step 7) integrates genetic and spatial data to examine the relationship between individual- or population-based genetic measures and spatially-explicit landscape or environmental variables. Following Wagner and Fortin (2013), we describe four types of analytical methods that are suitable for landscape genetics: (1) node-based methods that relate adaptive genes or genetic diversity to landscape or environmental characteristics at sampling locations; (2) link-based methods that relate pair-wise measures of genetic differentiation to geographic distance measures; (3) neighborhood-based methods that relate genetic diversity or differentiation to characteristics of the landscape surrounding the sampling locations; and (4) boundary-based methods that delineate populations on a landscape.

Simulations play an important role in landscape genetics theory, study design (Steps 2,3), and data analysis (Step 7; Epperson et al. 2010). Simulations allow researchers to examine statistical power of different sampling regimes and analytical methods (Cushman and Landguth 2010a; Landguth et al. 2012b; Oyler-McCance et al. 2013), to validate results from analysis of empirical data (e.g., Shirk et al. 2012; Castillo et al. 2014), and to extend inference of landscape genetics studies to larger spatial and temporal scales (e.g., Wasserman et al. 2012). We integrate results from and emphasize potential uses of simulation studies throughout the seven-step framework. Below we explain the seven steps in detail and discuss important considerations for study design and implementation in each step.

Step 1: Define the study objectives

The objectives of most landscape genetics studies have focused on understanding gene flow. Early work quantified the restriction of gene flow by landscape barriers (isolation by barriers or IBB) or geographic distance (isolation by distance or IBD; Wright 1943; Slatkin 1993). Contemporary studies test isolation by resistance (IBR; Cushman et al. 2011; Amos et al. 2012) using multivariate resistance surfaces, which model the permeability of dispersers through different habitats. Recently, isolation by environment (IBE) models have been used to understand selection by partitioning the effects of landscape and environmental factors on spatial patterns of genetic variation (Wang et al. 2013; Bradburd et al. 2013; Wang 2013). In addition, studies have assessed dispersal corridor effectiveness (Epps et al. 2007), measured genetic response to landscape change (Holzhauer et al. 2009), and examined demographic and metapopulation processes (Murphy et al. 2010a).

Some landscape genetics studies characterize adaptive genetic diversity by describing spatial patterns of selection (Table 1). The continued loss and fragmentation of habitats from development and climate change has fueled an increased interest in understanding selection and local adaptation of organisms. Studies focused on selection should increase in frequency as methods for landscape genomics develop (Schwartz et al. 2010; Manel and Holderegger 2013). For example, landscape genomics has been used recently to examine selection and local adaptation in plants and animals (Sork et al. 2010; Schoville et al. 2012; Bradburd et al. 2013).

Step 2: Consider the spatial and temporal scale of the study

Linking genetic variation to landscape and environmental characteristics that change over different spatial and temporal scales is a major challenge for landscape genetics studies (Anderson et al. 2010). The choice of scales depends on the study objectives and the life history and demography of the study organism (Fig. 1, Step 2) and should match the scale of the process being measured. Processes occurring at different spatial and temporal scales, such as seasonal changes, climate shifts, and habitat loss, have different impacts on landscapes and, in turn, gene flow (Anderson et al. 2010).

Dispersal behavior and generation time are arguably the most important characteristics affecting scale selection (Wright 1943; Slatkin 1987). Studies often use qualitative information about dispersal to inform scale selection and study design, but some have quantified dispersal distance using radio-telemetry (Epps et al. 2007; Elliot et al. 2014). Studies of organisms with greater dispersal distances require data collection over larger spatial scales, but may be able to use data collected over shorter time scales. A simulation study by Landguth et al. (2010) showed that genetic differentiation in organisms with greater dispersal distances responded more quickly to landscape changes than organisms with limited dispersal. If dispersal differs among age classes or sexes, age of individuals and mating strategy of an organism become important considerations (e.g., Elliot et al. 2014). In addition, genetic variation of organisms with short generation times responds more quickly to landscape changes (Landguth et al. 2010) so data collection can occur at shorter time scales.

Some studies collect genetic, landscape, or environmental data at multiple spatial or temporal scales (e.g., Holzhauer et al. 2009; Pavlacky Jr et al. 2009; Emaresi et al. 2009). Data

collection from multiple time periods allows researchers to account for a time lag in the response of genetic variation to landscape change (see below) and to estimate the rate of change in the response. The effects of spatial-scale variation can be assessed by calculating different transect widths (Emaresi et al. 2009; Murphy et al. 2010b) or by using different resolutions of spatial data (Cushman and Landguth 2010b). Anderson et al. (2010) recommended studies use a spatial grain size smaller than the organism's average home range size. Recently Galpern and Manseau (2013) described a method to identify an appropriate spatial grain size using moving windows and grains of connectivity.

The choice of temporal scale for collection of landscape or environmental data is important because lag time in the response of genetic variation to landscape change differs among organisms. In simulations, the amount of time to detect removal or formation of a dispersal barrier increased with decreasing dispersal distance (Landguth et al. 2010). Organisms with very short dispersal distances can maintain the signal of a historic barrier for more than 100 generations (Landguth et al. 2010). Thus, for species with limited dispersal, historic landscape characteristics may have a greater influence on genetic variation, which could potentially confound the correlation of genetic variation with contemporary landscapes (Landguth et al. 2010). Non-independence of landscape features can be accounted for by removing the effect of a correlated landscape feature (e.g., a land cover type) from historic and contemporary landscape data (Zellmer and Knowles 2009). In addition to controlling for correlated landscape characteristics, it is important to consider demographic processes, such as distribution shifts, that may influence genetic variation on a landscape over evolutionary time scales (He et al. 2013).

Step 3: Design a sampling regime

The sampling design of a landscape genetics study – the number of samples, the spatial distribution of samples, and the landscape or environmental characteristics measured – is influenced by the study objectives, spatial and temporal scale, and life history and demographic traits of the study organism (Fig. 1, Step 3). Logistical considerations, such as cost or time required for sample collection, genetic analysis, and collection of landscape or environmental data, also play a role in designing a sampling regime.

In general, the statistical power of genetic analyses increases with an increasing number of samples, loci, and alleles per locus. Simulations, implemented in programs such as CDPOP (Landguth and Cushman 2010; Landguth et al. 2012a), provide a way to evaluate how power is affected by the number of samples, loci, and alleles per locus, and allows selection of an optimal sample size for a particular study system (Cushman and Landguth 2010a; Landguth et al. 2012b). Landguth et al. (2012b) used Mantel tests to detect the effects of different landscape isolation models on genetic differentiation and found that using more loci and using loci with more alleles yielded a greater increase in power than adding samples. Thus, when there are trade-offs in cost or time between increasing the number of samples and increasing the number of loci, it may be more efficient to increase the number of loci and to use loci with more alleles. However, when using loci with many alleles to measure genetic differentiation among populations, estimates should be interpreted with caution (Hedrick 1999). The ability to increase power by increasing the number of loci evaluated is particularly helpful when sample size is limited by the rarity or secretive behavior of the study organism (e.g., Girard et al. 2010). When genetic differentiation

among populations is low, Hale et al. (2012) recommended collecting 25 to 30 samples per population for microsatellite (msat) studies; fewer samples may be required when genetic differentiation is high.

The spatial distribution of sample collection depends on the distribution of the study organism and spatial heterogeneity of the landscape or environmental characteristics of interest. Samples should be collected in clusters if organisms are distributed in discrete patches, whereas sample collection should be distributed evenly across the landscape for continuously-distributed organisms (Anderson et al. 2010). Simulations have been used to compare alternative study designs and to guide decisions about sampling. For example, Oyler-McCance et al. (2013) recommended using random, linear, or systematic sampling designs; they found that cluster and single study-site designs failed to correctly identify landscape factors effecting genetic differentiation. For landscape genomics studies of selection, Manel et al. (2012) recommended stratified sampling across environmental space rather than geographic space.

An uneven sample distribution can affect genetic distance estimates and lead to the detection of false signals of genetic differentiation. Genetic distance estimates, including F-statistics (Wright 1931) and conditional graph distance (cGD; Dyer and Nason 2004), are sensitive to under-sampled locations, and cGD is also sensitive to unsampled locations (Koen et al. 2013; Table 4). Schwartz and McKelvey (2008) compared the number of populations identified under different sampling distributions. In the presence of a genetic gradient (a scenario likely to occur in populations where nearest neighbors mate with each other), the sampling distribution influenced the number of populations identified. Thus, sample collection should occur at a scale small enough to test for spatial autocorrelation of genetic data, and the appropriate scale is best identified from a pilot study (Schwartz and McKelvey 2008). Alternatively, information about the study organism's home range size and dispersal distance can inform decisions about the spatial scale of sample collection (Anderson et al. 2010).

Landscape and environmental characteristics should be selected from hypotheses about their effects on genetic variation informed by a pilot study or expert opinion. Using large numbers of landscape or environmental characteristics should be avoided, because many characteristics are correlated (Cushman et al. 2011). Candidate landscape and environmental characteristics should be tested for correlation prior to analysis; highly correlated characteristics should be discarded, reduced to independent orthogonal components using principal components analysis (PCA) or canonical correspondence analysis (Table 3), or accounted for using Variance Inflation Factors or tolerances (O'brien 2007).

Step 4: Select a genetic marker

The genetic marker selected for a landscape genetics study should have enough variability to achieve the study objectives at the spatial and temporal scale of interest (Fig. 1, Step 4; Table 1). Each marker has unique properties that affect its suitability for landscape genetics studies, such as whether it is neutral or adaptive, its mode of inheritance, its mutation rate, and its spatial and temporal scale of inference (Table 2).

Most markers are neutral, having no (or little) effect on fitness and therefore reflect a pattern of gene flow and genetic drift across a landscape, whereas some markers, such as nuclear genes, are adaptive, reflecting a pattern of selection across an environmental gradient (Holderegger et al. 2006; Table 2). Single nucleotide polymorphism (SNP) or amplified fragment length polymorphism (AFLP) loci are commonly used as neutral markers, but may be adaptive markers when they are located in or adjacent to loci under selection (Table 2).

Markers with different modes of inheritance have different effective population sizes and inferences about gene flow. Nuclear markers, such as allozymes, msats, SNPs, AFLPs, and nuclear gene sequences, have bi-parental inheritance reflecting gene flow patterns for both sexes (Table 2). In contrast, organellar DNA (i.e., mtDNA and cpDNA) are inherited uni-parentally, usually maternally (Table 2). Organellar DNA markers have smaller effective population sizes ($1/2N_e$ in hermaphrodites and $1/4N_e$ in species with separate sexes) than nuclear markers and therefore experience greater effects of genetic drift (Latta 2006). This could lead to increased genetic variation over small spatial scales, which is useful for landscape genetics (Latta 2006). In addition, comparisons between nuclear and organellar DNA marker types can be used to examine differences in gene flow between sexes (Latta 2006).

In general, markers with higher mutation rates, such as msats and SNPs, are useful for examining genetic variation at smaller spatial and shorter temporal scales, while markers with lower rates (allozymes, mtDNA, cpDNA, AFLP, and nuclear gene sequences) are useful for studies with larger spatial and longer temporal scales (Anderson et al. 2010; Wang 2011; Table 2).

Marker selection is also affected by differences in allelic expression, mutation model, and relative cost and ease of laboratory analysis. Allozymes are multi-allelic, codominant markers, allowing for identification of heterozygous genotypes. Laboratory analysis of allozymes is fast and inexpensive, but their utility is limited because only well-documented, soluble proteins are detectable and small numbers of loci are available. Allozymes mutate more slowly than other genetic markers (Latta 2006), making them useful for studies with larger spatial and longer temporal scales. The cost efficiency of allozymes made them popular in early landscape genetics studies, particularly in plants (Epperson and Chung 2001; Saenz-Romero et al. 2001), but they have since been replaced by markers with greater variability and more loci (e.g., msats, SNPs, and AFLPs).

Microsatellites are the most commonly used markers in landscape genetics (Storfer et al. 2010), and can address a broad range of study objectives (Table 1). Their high mutation rate makes them particularly useful for studying the rapid response of genetic variation to landscape change (Wang 2011). They are multi-allelic, codominant markers that mutate following a stepwise mutation model (Kimura and Ohta 1978). Microsatellites are multi-locus markers, so statistical power of analyses can be increased by increasing the number of loci (Landguth et al. 2012b). However, msats are not suitable for all types of landscape genetics studies including studies that characterize adaptive genetic diversity (Table 1); these studies require loci that are spread across an organism's genome and have greater numbers of loci than those typically employed by msat studies.

Although mtDNA and cpDNA have been used less frequently than msats in landscape genetics (Storfer et al. 2010), they are an appropriate marker choice when comparing divergent populations over larger spatial and longer temporal scales (Anderson et al. 2010; Wang 2011; Table 1). Mitochondrial and chloroplast DNA are multi-allelic markers with uni-parental inheritance. They are commonly inherited maternally, though some organisms (e.g., mussels and pines) have paternal inheritance (Latta 2006). The mutation of mtDNA and cpDNA can be modeled using an infinite sites model (Kimura 1969), making them well suited for many population-based genetic distance measures (Kalinowski 2002; Table 4). However, the mutation rate of these markers may not be fast enough to detect a response of genetic variation to rapidly changing landscapes over small spatial or short temporal scales (Holderegger and Wagner 2008; Wang 2011; Table 2).

More recently, studies have used AFLPs and SNPs with hundreds to thousands of loci to examine neutral or adaptive genetic diversity with high statistical power (Schwartz et al. 2010; Schoville et al. 2012; Table 1). Comparison of neutral and adaptive AFLP or SNP loci may be particularly useful for disentangling the effects of selection from gene flow and genetic drift. Single nucleotide polymorphisms are bi-allelic (usually), codominant markers that mutate following an infinite sites model (Kimura 1969). Sequencing technologies have made SNP development possible for non-model organisms, paving the way for use of these markers in landscape genetics. Like msats, SNPs are useful for studies at smaller spatial and shorter temporal scales. Amplified fragment length polymorphisms are multi-allelic, dominant markers (i.e., heterozygotes cannot be differentiated from individuals that are homozygous for the dominant allele). So, unlike co-dominant SNPs, accurate calculation of heterozygosity is not possible using AFLPs. Also, the choice of genetic measure is limited when using AFLPs (Table 4). But, AFLPs have been useful for studies conducted at larger spatial and longer temporal scales (Anderson et al. 2010). As sequencing technologies increase and become more cost effective, use of AFLPs and SNPs is likely to increase (Schwartz et al. 2010).

Studies describing spatial patterns of selection can use nuclear genes. Nuclear genes are multi-allelic, codominant markers that are under selection. Often, genes of interest are chosen based on the expectation that their function varies across an environmental gradient. Nuclear genes follow an infinite sites mutation model (Kimura 1969) but, in addition to mutation and drift, selection influences allele frequencies. Therefore, genetic measures of migration or distance (Table 4) that assume marker neutrality are not appropriate for nuclear genes. Studies that employ nuclear genes often use descriptive genetic measures such as genotypes or allele frequencies for landscape genetics analyses (Table 4).

Step 5: Generate genetic input data

Genetic measures (Fig. 1, Step 5) describe basic genetic variation or estimate genetic differentiation among populations or individuals sampled across a landscape (Table 4). The spatial distribution of sampling influences the decision to use a population- or individual-based measure. Population-based measures, usually applied to organisms in discrete groups, can be used for continuously-distributed organisms if they can be assigned to groups using boundary-based analyses (Table 3) such as Bayesian clustering methods (Guillot et al. 2005; Chen et al.

2007). Individual-based measures avoid the need to group organisms *a priori*, but require careful consideration of spatial autocorrelation of genetic data (Schwartz and McKelvey 2008).

Three types of population-based genetic measures can be used: descriptive, migration, or distance (Table 4). Descriptive measures characterize basic genetic variation of each population (Table 4) and are useful for node-, link-, neighborhood-, and boundary-based analyses (Table 3). Migration measures calculate the migration rate or number of migrants among populations (Table 4), providing a measure of connectivity for link-based analyses (Table 3). Distance measures are also useful for link-based analyses (Table 3), because they measure genetic differentiation among populations (Table 4). Some distance measures, such as F-statistics (Wright 1931; Rousset 1997), cord distance (Cavalli-Sforza and Edwards 1967; Nei et al. 1983), and standard distance (Nei 1972), assume populations are in mutation-drift equilibrium (Table 4). This assumption is reasonable for populations that have not undergone recent changes in spatial distribution. However, in studies of recent habitat disturbance, these distance measures may be unsuitable; under these circumstances distance measures that do not assume mutation-drift equilibrium should be used (e.g., cGD; Table 4).

Individual-based descriptive or distance measures (Table 4) can be used in landscape genetics studies. Raw genotypes (Table 4) can be used as input data in node-, link-, neighborhood-, and boundary-based analyses (Table 3). Commonly, genetic distance measures calculated among individuals (Table 4) are used for link-based analyses (Table 3). Distance measures range in computational complexity from the simple proportion of shared alleles (Bowcock et al. 1994) to a more complex PCA-based distance that gives more weight to loci with greater genetic variation (Shirk et al. 2010; Table 4).

Most genetic measures can be calculated using any type of marker, but some measures are only suitable for certain markers (Table 4). For example, percentage of polymorphism is suitable for allozyme and AFLP markers, whereas the Jaccard (1908), Dice (1945), and simple-matching (Sokal and Michener 1958) coefficients are only suitable for AFLP markers (Table 4). In addition, population-based distance measures that assume an infinite sites mutation model, such as chord distance (Cavalli-Sforza and Edwards 1967; Nei et al. 1983) and standard distance (Nei 1972), may be less suitable than Rst (Slatkin 1995) for use with msats due to their stepwise mutation (Kalinowski 2002; Table 4).

Step 6: Generate spatial input data

Landscape and environmental characteristics are used to generate spatial input data (Fig. 1, Step 6) for the final analysis. Multi- or univariate landscape or environmental data can be used in node-, link-, neighborhood-, and boundary-based analyses, whereas landscape or habitat maps serve as inputs for neighborhood-based analyses (Table 3). Spatial distance measures, including Euclidean distance and transect or resistance-surface distances, are useful for link- and neighborhood-based analyses (Table 3).

Multi- or univariate landscape or environmental data often need no prior analysis before the final analysis, but for some analytical methods it may be necessary to alter spatial input data. Methods such as PCA and canonical correspondence analysis can be used to summarize

variation in spatial data by reducing the dimensionality of multivariate data to a few, independent orthogonal components (Table 3). Other methods, like inverse distance weighting, are useful for creating maps of spatial variation from univariate landscape or environmental data (Table 3).

Landscape or habitat maps are useful input data for neighborhood-based analyses (Table 3) that relate genetic diversity or differentiation of sampling locations to landscape or habitat characteristics surrounding the locations. Many spatial pattern analyses are available in packages like FRAGSTATS (McGarigal et al. 2012) and PATCH ANALYST (Elkie et al. 1999). They can be used to quantify landscape characteristics and habitat area or fragmentation of a single patch or of a neighborhood surrounding a patch. Habitat fragmentation can also be quantified using o-ring statistics (Bruggeman et al. 2010).

Spatial input data for link and neighborhood-based analyses (Table 3) can be generated using network models that connect sampling locations on a landscape and estimate distance among locations using information about an organism's movements. Euclidean distance – the straight-line distance between two locations – is the simplest distance measure. However, gene flow often occurs via more complex routes on the landscape. Transect or resistance-surface distance measures attempt to account for this complexity by incorporating landscape or environmental characteristics that may affect gene flow.

Transects quantify proportions of different landscape or environmental characteristics along straight lines among sampled locations in a network. Transect width should be selected to match the spatial scale at which the study organism interacts with the landscape. Some studies have used multiple strip widths to test the effect of landscape or environmental characteristics on gene flow at different spatial scales (Emaresi et al. 2009; Murphy et al. 2010b).

Resistance surfaces are raster-based maps of landscape or environmental characteristics that model permeability of different habitat types to dispersal (Spear et al. 2010; Zeller et al. 2012). Each cell on a resistance surface is assigned a cost value, with high costs given to habitats that restrict dispersal and low costs to those that facilitate dispersal. Cost values can be assigned using expert opinion, habitat suitability estimated from presence/absence data, movement data from tagging and tracking studies, or genetic data (Spear et al. 2010; Zeller et al. 2012). With genetic data, cost values are assigned using optimization methods to identify parameter values for landscape or environmental characteristics that maximize the correlation of genetic distance with resistance-surface distance (Shirk et al. 2010; Graves et al. 2013).

Distance among sampling locations on a resistance surface is measured using one or more least-cost paths. A least-cost path is a predicted rectilinear path for an organism that minimizes the cost of dispersal between two locations (Spear et al. 2010). Resistance surfaces can be used to estimate least-cost distance (e.g., Epps et al. 2007), resistance distance (e.g., McRae 2006), or a least-cost transect (e.g.,Van Strien et al. 2012). Least-cost distance (total cost of a least-cost path) can easily be calculated from resistance surfaces, but it assumes organisms make movement decisions with perfect knowledge of their environment and, therefore, may not capture the true range of dispersal routes across a landscape (Spear et al. 2010). Resistance distance (the average length of several least-cost paths) accounts for different dispersal routes across a landscape (McRae 2006), but, unlike transect methods, does not account for

characteristics of the landscape surrounding the least-cost paths. The least-cost transect method is a hybrid of transect and resistance surface methods. It calculates distances among habitat patches on a resistance surface using a least-cost model, but also incorporates a proportion of the landscape or environmental characteristics surrounding the path within a specified transect width (Van Strien et al. 2012). This method may be useful for modeling movements of larger organisms that interact with the environment at larger spatial scales, or for organisms that disperse more slowly and may require wider dispersal corridors to accommodate movement among patches over longer time periods.

Step 7: Choose an analytical method that integrates genetic and spatial data

The final step in a landscape genetics study is to integrate genetic and spatial data (Fig. 1, Step 7) in a multiple-hypothesis testing framework using analytical methods to achieve the study objectives (Table 1). Four types of analytical methods can be used: node-, link-, neighborhood-, and boundary-based methods (Wagner and Fortin 2013).

Node-based analyses integrate genetic and spatial data at sampling locations, and some, including inverse distance weighting, Moran's eigen vector mapping, and PCA (Table 3), can be used to convert genetic or spatial data into inputs for the final analysis. Inverse distance weighting visualizes genetic variation as a continuous "genetic surface" (Murphy et al. 2008) that can serve as an input for link- or neighborhood-based analyses. Moran's eigen vector mapping measures positive spatial correlation between geographic locations and genetic, environmental, or landscape data. It was used to incorporate the effects of unmeasured environmental variation in a study of adaptive genetic variation of an alpine plant (*Arabis alpina*; Manel et al. 2010). Also, PCA can be used to reduce the dimensionality of multilocus genotypes or environmental data.

Node-based analyses also describe spatial patterns of genetic data to understand gene flow or selection (Table 1). Spatial principal components analysis identifies independent, orthogonal components that optimize the variance of allele frequencies while accounting for spatial correlation of individuals or populations (Table 3). This method can be used to identify global and local patterns of genetic variation (Jombart et al. 2008). Other ordination methods, including canonical correlation analysis and canonical correspondence analysis, have been used to understand patterns of selection. Canonical correlation analysis identifies linear combinations that maximize the correlation between genetic and spatial data, whereas canonical correspondence analysis measures variation in a dependent variable at geographic locations constrained by multivariate axes that describe environmental variation at those locations.

Link-based analyses, such as Mantel or partial Mantel tests, relate pair-wise measures of genetic differentiation to geographic distance measures to identify landscape or environmental characteristics that restrict or enhance gene flow. Mantel or partial Mantel tests are the most commonly used analytical methods in landscape genetics (Storfer et al. 2010; Table 3). They examine the correlation between a matrix of pair-wise genetic distances and a matrix of spatial distances (see Diniz-Filho et al. 2013 for review). Despite their frequent use, Mantel tests have been criticized for having low power and inflated type I error rates (Legendre and Fortin 2010). Cushman et al. (2013a) observed elevated type I error rates caused by high correlation among

different resistance models. In spite of these shortcomings, Mantel tests have very low type II error rates (Cushman et al. 2013a) and, when applied and interpreted correctly, can be useful for understanding spatial patterns of genetic differentiation (Diniz-Filho et al. 2013).

Other link-based analyses, including linear regression, Bayesian geographic analysis (BGA), multiple matrix regression, and generalized dissimilarity modeling (GDM), have been used to identify landscape or environmental characteristics that affect gene flow (Table 3). Linear regression models the linear relationship between genetic and spatial data while accounting for hierarchical data structure. In contrast, when the response of genetic data to the landscape is non-linear, GDM can be used (e.g., Freedman et al. 2010). Multiple matrix regression was used by Wang et al. (2013) to quantify the contributions of both landscape and environmental characteristics to gene flow. This method is especially useful for conservation and management efforts because it can be used to disentangle the effects of gene flow and selection across a landscape and identify regions of local adaptation. Finally, while most analytical methods employ frequentist statistical approaches, BGA examines the association of different loci with environmental variables to infer selection (Eckert et al. 2010).

Neighborhood-based analyses are useful for understanding demographic and metapopulation processes or measuring spatial autocorrelation, because they integrate genetic data with characteristics of the landscape surrounding sampling locations. Moran's I can be used to estimate the scale of gene flow by measuring spatial autocorrelation, and gravity models assess demographic and metapopulation processes (Table 3). Similar to node-based analyses, they measure characteristics of sampling locations, such as habitat area or population growth, but in addition, they incorporate landscape or environmental characteristics along transects among sampling locations (Table 3). For example, Murphy et al. (2010a) used a gravity model to assess connectivity among montane lakes in a metapopulation of Columbia spotted frogs (*Rana luteiventris*).

Boundary-based analyses identify dispersal barriers by measuring spatial overlap of genetic and landscape or environmental discontinuities on the landscape. In a spatial context, boundaries identify regions of change in landscape or environmental characteristics such as habitat type or precipitation. They can be measured with edge detection techniques, including Monmonier's algorithm (1973) and wombling (Womble 1951; Table 3). In a genetic context, boundaries act as barriers to gene flow, separating panmictic populations. Genetic boundaries can be identified using edge detection techniques, Bayesian clustering methods implemented in programs such as GENELAND (Guillot et al. 2005) or TESS (Chen et al. 2007; Table 3), or non-Bayesian clustering methods including PSMIX (Wu et al. 2006) and discriminant analysis of principle components (Jombart et al. 2010). Statistical power of edge detection and clustering methods to detect a genetic barrier in simulated data sets with different conditions for dispersal and genetic equilibrium has been compared by Blair et al. (2012). Their results indicated that clustering methods outperformed edge detection methods, that barriers can be detected more rapidly in species with long distance dispersal, and that isolation by distance can confound the identification of barriers using these methods. The coincidence of boundaries can be determined by overlaying maps of landscape or environmental characteristics and genetic population clusters. Statistical methods to assess boundary overlap have been developed (Jacquez 1995), but have been underutilized in landscape genetics.

Finally, it is important to consider the framework that will be used for testing multiple hypotheses or selecting among models to ensure accurate inference from statistical methods and to minimize the influence of spurious correlations on the interpretation of landscape genetics patterns. Cushman et al. (2006) introduced a causal modeling approach to test competing hypotheses using partial Mantel tests. Improvements to the approach have been made by Shirk et al. (2010), and Cushman et al. (2013a; 2013b). Others have used Akaike's information criterion (Burnham and Anderson 2002) to compare candidate models using linear modeling (Meeuwig et al. 2010) and gravity modeling (Murphy et al. 2010a). In addition to frequentist methods, Bayesian geographical analyses have used Bayes factors as measures of support for relationships between candidate landscape or environmental variables and genetic data (Eckert et al. 2010).

Concluding Remarks

A successful landscape genetics study combines genetic and landscape or environmental data to make spatially-explicit conclusions about factors affecting gene flow or selection. The projection of landscape genetics patterns to larger spatial and temporal scales through simulations is a valuable tool for management (e.g. Wasserman et al. 2012) and will likely influence future conservation decisions and mitigation efforts, especially as habitat loss and fragmentation and climate change alter landscapes. Emerging technologies, such as high throughput sequencing, advances in remote sensing, and low-cost climate data loggers, have and will continue to facilitate the collection and analysis of genetic, landscape, and environmental data at greater resolutions and finer scales. The resulting abundance of data allows researchers to explore patterns, and the processes that generate them, with increased statistical power (Schwartz et al. 2010; Schoville et al. 2012). Future progress in theory and application of landscape genetics depends upon hypothesis-driven research that utilizes empirical data from controlled, replicated experiments to test observed patterns, and uses simulation modeling to understand how relationships between patterns and processes change across spatial and temporal scales (Cushman 2014). As genetic and spatial data sets increase in size, complexity, and resolution, new study design considerations and analytical methods will surely arise. The landscape genetics framework presented here (Fig. 1) can accommodate new design considerations and analyses, and facilitate integration of genetic and spatial data by guiding new landscape geneticists through study design and implementation.

Acknowledgements

We thank the Beissinger Lab for discussions and advice during development of this manuscript. Constructive comments were also provided by N. VanSchmidt, K. Iknayan, J. Belton, two anonymous reviewers, and the associate editor. Financial support was provided by the National Science Foundation DEB-1051342 and CNH 1115069 to SRB.

Tables

Table 1. Objectives for landscape genetics studies and the applicability of genetic markers and analytical methods (Wagner and Fortin 2013). IBB = isolation by barrier; IBD = isolation by distance; IBR = isolation by resistance; IBE = isolation by environment. See text for definitions of analytical methods.

Microevolutionary Process	Study Objective	Allozyme	Msat	MtDNA/ CpDNA	AFLP	SNP	Nuclear Gene	Analytical Method
Gene Flow	Identify gene flow barriers (IBB)	x	x	x	x	x		Boundary
	Identify characteristics that restrict or enhance gene flow (IBD, IBR, or IBE)	x	x	x	x	x		Link
	Assess dispersal corridor effectiveness	x	x	x	x	x		Link
	Estimate scale of gene flow	x	x	x	x	x		Node Neighborhood
	Understand demographic and metapopulation processes		x		x	x		Neighborhood
	Measure disease emergence and spread		x		x	x		Link Neighborhood
	Measure genetic response to landscape change	x	x	x	x	x		Link Neighborhood Boundary
	Examine patterns of spatial dependency in genetic data	x	x	x	x	x		Node
Selection	Describe spatial patterns of selection				x	x	x	Node

Table 2. Advantages, disadvantages, and characteristics of genetic markers used in landscape genetics.

Marker	Effect of Selection	Mode of Inheritance	Mutation Rate[c]	Temporal Scale	Spatial Scale	Advantages	Disadvantages
Allozyme	Neutral[a]	Bi-parental	Low	Long	Large	Fast and inexpensive	Few markers available
Msat	Neutral	Bi-parental	High	Short	Small	Codominant marker with high mutation rate	Stepwise mutation model limits genetic distance measures
MtDNA & CpDNA	Neutral[a]	Uni-parental	Low	Long	Large	Smaller effective population size than bi-parentally inherited markers increases genetic differentiation at small spatial scales; detects sex-biased dispersal by comparison with bi-parentally inherited markers	Mutation rate too slow to detect effects of rapidly changing landscapes
AFLP	Neutral or Adaptive[b]	Bi-parental	Low-Moderate	Long	Large	Large number of markers spread throughout genome	Dominant marker and co-migration of non-homologous fragments can occur
SNP	Neutral or Adaptive[b]	Bi-parental	Moderate-High	Short	Small	Large number of markers spread throughout genome	Marker development requires prior knowledge of organism's genome and susceptible to ascertainment bias
Nuclear Gene	Adaptive	Bi-parental	Low-Moderate	Variable	Variable	Identify genes involved in local adaptation	Primer design requires prior knowledge of the gene sequence

[a] Assumed to be selectively neutral

[b] May be in or adjacent to regions of the genome under selection

[c] Mutation rates are marker specific, but markers can be contrasted qualitatively

Table 3. Descriptions and examples of node-, link-, neighborhood-, and boundary-based analyses (Wagner and Fortin 2013) used in landscape genetics. The type of genetic and spatial input data varies among methods.

Type of Method	Statistical Classification	Analytical Method	Genetic Measures	Spatial Measures	Description	Example
Node	Ordination	Principal components analysis	a1	a	Converts multivariate continuous data of correlated values to uncorrelated values	Shirk et al. 2010
		Spatial principal component analysis	a	d	Identifies independent, orthogonal components that optimize the variance of allele frequencies while accounting for spatial correlation among individuals or populations	Jombart et al. 2008
		Canonical correlation analysis	a1	a	Finds linear combinations that maximize correlation between genetic and spatial data	Sork et al. 2010
		Canonical correspondence analysis	a2	a	Measures genetic variation at geographic locations constrained by multivariate axes that describe environmental variation	Pease et al. 2009
	Spatial heterogeneity	Moran's eigenvector mapping	a2	a	Models positive spatial correlation of genetic, environmental, or landscape data	Manel et al. 2010
	Spatial interpolation	Inverse distance weighting	a	c	Estimates values at unmeasured locations using measurements from nearby locations	Murphy et al. 2008
Link	Correlation	Mantel and partial Mantel test	b, c	d, e	Correlation between two distance matrices assuming a linear relationship between variables	Goldberg and Waits 2010
		Linear regression	a, b, c	d, e, f	Models linear relationship between genetic and landscape variables while accounting for hierarchical data structure	Meeuwig et al. 2010
		Bayesian geographical analysis	a	f	Correlates allele frequencies with landscape variables	Eckert et al. 2010
	Spatial regression	Geographically weighted regression	a2	e	Assesses spatial heterogeneity by estimating regression parameters for each data point using values of nearby points	Spear and Storfer 2010

	Matrix regression	Multiple matrix regression	b, c	d, e	Quantifies effects of multiple dependent variables on genetic distance	Wang et al. 2013
		Generalized dissimilarity modeling	c	a, d, e	Models non-linear response of genetic data to environmental variation	Freedman et al. 2010
Neighborhood	Spatial autocorrelation	Moran's I	a, b, c	d	Tests correlation of a variable among nearby locations in space	Jones et al. 2007
	Spatial interaction	Gravity Model	c	b, c, d, e	Predicts movement based on distance between sites, site variables, and between-site variables affecting resistance	Murphy et al. 2010a
Boundary	Edge detection	Monmonier's algorithm	a2	b, c	Finds edges with highest rate of change	Manni et al. 2004
		Wombling	a2	b, c	Detects areas of abrupt change on allele frequency surfaces or maps of landscape or environmental variables	Cercueil et al. 2007
	Spatial Bayesian clustering	Boundary detection (GENELAND; Guillot et al. 2005)	a1	d	Estimates number of genetic populations by maximizing Hardy Weinberg and linkage equilibria; defines boundaries given prior information about the number of populations	Heidinger et al. 2013
		Local spatial dependence (TESS; Chen et al. 2007)	a1	d	Estimates number of genetic populations by maximizing Hardy Weinberg and linkage equilibria; defines clusters by minimizing heterozygosity reduction caused by population sub-structure and local spatial dependence	Rico et al. 2014
	Spatial overlap	Boundary overlap	d	g	Quantifies spatial overlap of boundary locations for genetic populations and landscape or environmental characteristics	Blair et al. 2012

Genetic input data:

a. Descriptive
   1. Genotypes
   2. Allele frequencies
b. Migration
c. Distance
d. Output from boundary-based analysis

Spatial input data:

a. Multivariate landscape or environmental data
b. Landscape or habitat maps
c. Univariate landscape or environmental data
d. Geographic coordinates or Euclidean distance
e. Transect or resistance-surface distance
f. Output from node-based analysis of landscape or environmental data
g. Output from boundary-based analysis

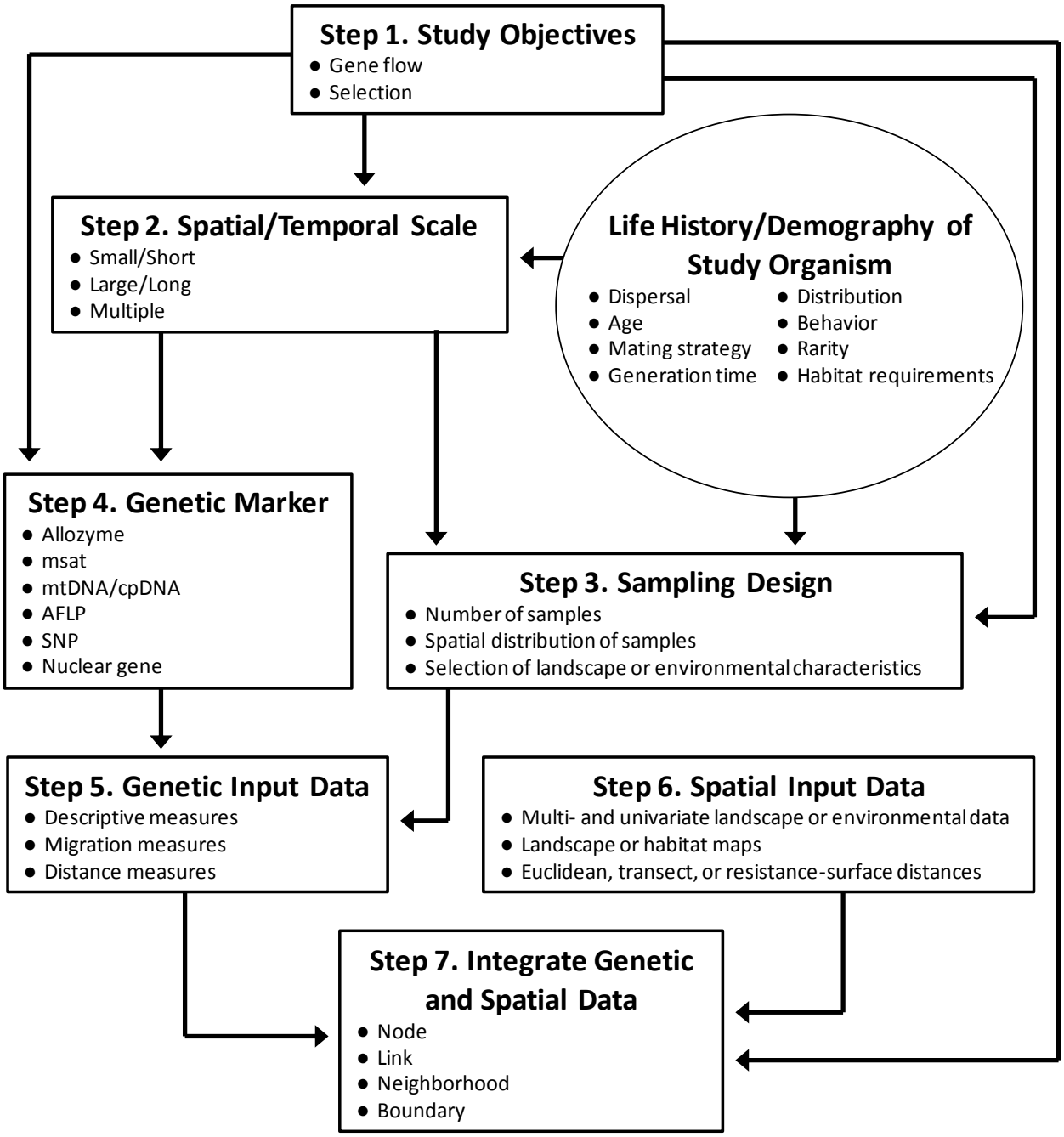
Table 4. Characteristics of population- and individual-based genetic measures including descriptive (basic statistics describing genetic variation), migration (migration rate or number of migrants), and distance (genetic differentiation) measures.

Type of Measure	Data Type	Genetic Measure	Parameter	Key Characteristics	Citation
Population	Descriptive	Heterozygosity	H_e	Commonly used measure of variation	Beebee and Rowe 2008
		Percentage of polymorphism	%P	Simple measure of variation used with allozymes or AFLPs	Beebee and Rowe 2008
		Allele frequency	p	Simple measure of variation	Beebee and Rowe 2008
		Allelic richness	A	Simple measure of variation; More sensitive to founder effects than heterozygosity	Beebee and Rowe 2008
	Migration	Migration rate	M	Migration rate over longer time scales from MIGRATE; Assumes mutation-drift equilibrium	Beerli and Felsenstein 2001
		Migration rate	m	Migration rate over shorter time scales from BAYESASS	Wilson and Rannala 2003
		No. of migrants	N_m	Number of migrants over longer time scales from MIGRATE; Assumes mutation-drift equilibrium	Beerli and Felsenstein 2001
	Distance	F-statistic	F_{ST} or $F_{ST}/(1-F_{ST})$	Common measure of differentiation among populations; Assumes mutation-drift equilibrium	Wright 1931; Rousset 1997
		F-statistic analogue	R_{ST}	Measures differentiation among populations using msats; Assumes marker follows a stepwise mutation model	Slatkin 1995
		F-statistic analogue	θ	Require no assumptions about number of populations sampled, sample size, or heterozygosity of loci; Assumes mutation-drift equilibrium	Weir and Cockerham 1984
		Jost's D	D_J	Measures fraction of allelic variation among populations	Jost 2008
		Standardized distance	G'_{ST}	Standardized measure that divides population differentiation by the maximum possible differentiation; Assumes mutation-drift equilibrium	Hedrick 2005
		Chord distance	D_C	Assumes gene frequency distribution is Gaussian; More robust when variance of gene frequencies is small; Depends on number of low-frequency alleles; Assumes mutation-drift equilibrium	Cavalli-Sforza and Edwards 1967
		Chord distance	D_A	Less dependent on number of low-frequency alleles than D_c; Assumes mutation-drift equilibrium	Nei 1983

		Standard distance	D_S	Assumes no migration and a mutation-drift equilibrium; Can be applied to organisms with different ploidys and mating schemes	Nei 1972
		Conditional graph distance	cGD	Uses differences in covariation associated with direct and indirect gene flow; Does not assume a hierarchical framework of populations; Does not use averaging statistics or coalescence; Distance matrix produced is equivalent to an AMOVA matrix	Dyer and Nason 2004
Individual	Descriptive	Genotype	pq	Set of alleles possessed by an individual	Beebee and Rowe 2008
	Distance	Jaccard coefficient	J	Used with AFLPs; Unaffected by homoplasic absent bands	Jaccard 1908
		Dice coefficient	D	Used with AFLPs; Gives weight to bands present in both individuals	Dice 1945
		Simple-matching coefficient	M	Used with AFLPs; Double-band absence and presence contribute equally; Can be used in AMOVA	Sokal and Michener 1958
		Proportion of shared alleles	P_S (D_{PS})	Easy to calculate	Bowcock et al. 1994
		Bray-Curtis percentage dissimilarity	d	Accounts for semi-quantitative nature of 3-state genetic data; Double negatives are not counted; Abundant and rare alleles contribute equally to distance	Bray and Curtis 1957
		Rousset's a	a_r	Asymptotically unbiased, except for small sample sizes; Based on isolation by distance in continuously distributed species	Rousset 2000
		PCA-based genetic distance	D_{PCA}	Gives more weight to loci with greater variation	Shirk et al. 2010

Figures

Figure 1. A framework for design and analysis of landscape genetics studies. Numbered boxes list considerations within each of the seven steps of the framework, and arrows connecting boxes depict the flow of information between steps. Life history and demographic characteristics of the study organism that inform decisions about study design, but are not controlled by the researcher, are listed in the oval.



Literature cited

Amos JN, Bennett AF, Mac Nally R, et al. (2012) Predicting landscape-genetic consequences of habitat loss, fragmentation and mobility for multiple species of woodland birds. PLoS ONE 7:e30888.

Anderson CD, Epperson BK, Fortin MJ, et al. (2010) Considering spatial and temporal scale in landscape-genetic studies of gene flow. Mol Ecol 19:3565–3575.

Balkenhol N, Gugerli F, Cushman SA, et al. (2009a) Identifying future research needs in landscape genetics: where to from here? Landsc Ecol 24:455–463.

Balkenhol N, Waits LP, Dezzani RJ (2009b) Statistical approaches in landscape genetics: an evaluation of methods for linking landscape and genetic data. Ecography 32:818–830.

Beebee T, Rowe G (2008) An introduction to molecular ecology, Second Ed. Oxford University Press, New York

Beerli P, Felsenstein J (2001) Maximum likelihood estimation of a migration matrix and effective population sizes in n subpopulations by using a coalescent approach. Proc Natl Acad Sci 98:4563–4568.

Blair C, Weigel DE, Balazik M, et al. (2012) A simulation-based evaluation of methods for inferring linear barriers to gene flow. Mol Ecol Resour 12:822–833.

Bowcock AM, Ruiz-Linares A, Tomfohrde J, et al. (1994) High resolution of human trees with polymorphic microsatellites. Nature 368:455–457.

Bradburd GS, Ralph PL, Coop GM (2013) Disentangling the effects of geographic and ecological isolation on genetic differentiation. Evolution 67:3258–3273.

Bray JR, Curtis JT (1957) An ordination of the upland forest communities of southern Wisconsin. Ecoloical Monogr 27:325–349.

Bruggeman DJ, Wiegand T, FernáNdez N (2010) The relative effects of habitat loss and fragmentation on population genetic variation in the red-cockaded woodpecker (*Picoides borealis*). Mol Ecol 19:3679–3691.

Burnham KP, Anderson DR (2002) Model selection and multimodel inference: a practical information-theoretic approach, 2nd ed. Springer-Verlag, New York

Castillo JA, Epps CW, Davis AR, Cushman SA (2014) Landscape effects on gene flow for a climate-sensitive montane species, the American pika. Mol Ecol 23:843–856.

Cavalli-Sforza LL, Edwards AW (1967) Phylogenetic analysis. Models and estimation procedures. Am J Hum Genet 19:233.

Cercueil A, François O, Manel S (2007) The genetical bandwidth mapping: a spatial and graphical representation of population genetic structure based on the Wombling method. Theor Popul Biol 71:332–341.

Chen C, Durand E, Forbes F, FrançOis O (2007) Bayesian clustering algorithms ascertaining spatial population structure: a new computer program and a comparison study. Mol Ecol Notes 7:747–756.

Cushman S, Wasserman T, Landguth E, Shirk A (2013a) Re-evaluating causal modeling with Mantel tests in landscape genetics. Diversity 5:51–72.

Cushman SA (2014) Grand challenges in evolutionary and population genetics: the importance of integrating epigenetics, genomics, modeling, and experimentation. Front. Genet. 5:

Cushman SA, Landguth EL (2010a) Spurious correlations and inference in landscape genetics. Mol Ecol 19:3592–3602.

Cushman SA, Landguth EL (2010b) Scale dependent inference in landscape genetics. Landsc Ecol 25:967–979.
Cushman SA, Max TL, Whitham TG, Allan GJ (2013b) River network connectivity and climate gradients drive genetic differentiation in a riparian foundation tree. Ecol Appl 24:1000–1014.
Cushman SA, McKelvey KS, Hayden J, Schwartz MK (2006) Gene flow in complex landscapes: testing multiple hypotheses with causal modeling. Am Nat 168:486–499.
Cushman SA, Shirk A, Landguth EL (2011) Separating the effects of habitat area, fragmentation and matrix resistance on genetic differentiation in complex landscapes. Landsc Ecol 27:369–380.
Dice LR (1945) Measures of the amount of ecologic association between species. Ecology 26:297–302.
Diniz-Filho JAF, Soares TN, Lima JS, et al. (2013) Mantel test in population genetics. Genet Mol Biol 36:475–485.
Dobzhansky T (1947) A directional change in the genetic constitution of a natural population of *Drosophila pseudoobscura*. Heredity 1:53–64.
Dyer RJ, Nason JD (2004) Population Graphs: the graph theoretic shape of genetic structure. Mol Ecol 13:1713–1727.
Eckert AJ, Bower AD, GonzáLez-MartíNez SC, et al. (2010) Back to nature: ecological genomics of loblolly pine (*Pinus taeda*, Pinaceae). Mol Ecol 19:3789–3805.
Elkie PC, Rempel RS, Carr A, et al. (1999) Patch analyst user's manual: a tool for quantifying landscape structure. Ontario Ministry of Natural Resources, Boreal Science, Northwest Science & Technology, Thunder Bay
Elliot NB, Cushman SA, Macdonald DW, Loveridge AJ (2014) The devil is in the dispersers: predictions of landscape connectivity change with demography. J Appl Ecol. doi: 10.1111/1365-2664.12282
Emaresi G, Pellet J, Dubey S, et al. (2009) Landscape genetics of the Alpine newt (*Mesotriton alpestris*) inferred from a strip-based approach. Conserv Genet 12:41–50.
Epperson BK, Chung MG (2001) Spatial genetic structure of allozyme polymorphisms within populations of *Pinus strobus* (Pinaceae). Am J Bot 88:1006–1010.
Epperson BK, Mcrae BH, Scribner K, et al. (2010) Utility of computer simulations in landscape genetics. Mol Ecol 19:3549–3564.
Epps CW, Wehausen JD, Bleich VC, et al. (2007) Optimizing dispersal and corridor models using landscape genetics. J Appl Ecol 44:714–724.
Freedman AH, Thomassen HA, Buermann W, Smith TB (2010) Genomic signals of diversification along ecological gradients in a tropical lizard. Mol Ecol 19:3773–3788.
Galpern P, Manseau M (2013) Finding the functional grain: comparing methods for scaling resistance surfaces. Landsc Ecol 28:1269–1281.
Girard P, Takekawa JY, Beissinger SR (2010) Uncloaking a cryptic, threatened rail with molecular markers: origins, connectivity and demography of a recently-discovered population. Conserv Genet 11:2409–2418.
Goldberg CS, Waits LP (2010) Comparative landscape genetics of two pond-breeding amphibian species in a highly modified agricultural landscape. Mol Ecol 19:3650–3663.
Graves TA, Beier P, Royle JA (2013) Current approaches using genetic distances produce poor estimates of landscape resistance to interindividual dispersal. Mol Ecol 22:3888–3903.

Guillot G, Mortier F, Estoup A (2005) Geneland: a computer package for landscape genetics. Mol Ecol Notes 5:712–715.
Hale ML, Burg TM, Steeves TE (2012) Sampling for microsatellite-based population genetic studies: 25 to 30 individuals per population is enough to accurately estimate allele frequencies. PLoS ONE 7:e45170.
He Q, Edwards DL, Knowles LL (2013) Integrative testing of how environments from the past to the present shape genetic structure across landscapes. Evolution 67:3386–3402.
Hedrick PW (1999) Perspective: highly variable loci and their interpretation in evolution and conservation. Evolution 53:313.
Hedrick PW (2005) A standardized genetic differentiation measure. Evolution 59:1633–1638.
Heidinger IMM, Hein S, Feldhaar H, Poethke H-J (2013) The genetic structure of populations of Metrioptera bicolor in a spatially structured landscape: effects of dispersal barriers and geographic distance. Conserv Genet 14:299–311.
Holderegger R, Kamm U, Gugerli F (2006) Adaptive vs. neutral genetic diversity: implications for landscape genetics. Landsc Ecol 21:797–807.
Holderegger R, Wagner HH (2008) Landscape genetics. Bioscience 58:199–207.
Holderegger R, Wagner HH (2006) A brief guide to Landscape Genetics. Landsc Ecol 21:793–796.
Holzhauer SIJ, Wolff K, Wolters V (2009) Changes in land use and habitat availability affect the population genetic structure of *Metrioptera roeselii* (Orthoptera: Tettigoniidae). J Insect Conserv 13:543–552.
Jaccard P (1908) Nouvelles recherches sur la distribution florale. Bull Soc Vaudoise Sci Nat 44:223–270.
Jacquez GM (1995) The map comparison problem: Tests for the overlap of geographic boundaries. Stat Med 14:2343–2361.
Jombart T, Devillard S, Balloux F (2010) Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. BMC Genet 11:94.
Jombart T, Devillard S, Dufour AB, Pontier D (2008) Revealing cryptic spatial patterns in genetic variability by a new multivariate method. Heredity 101:92–103.
Jones TH, Vaillancourt RE, Potts BM (2007) Detection and visualization of spatial genetic structure in continuous Eucalyptus globulus forest. Mol Ecol 16:697–707.
Jost L (2008) G_{ST} and its relatives do not measure differentiation. Mol Ecol 17:4015–4026.
Kalinowski ST (2002) Evolutionary and statistical properties of three genetic distances. Mol Ecol 11:1263–1273.
Kimura M (1969) The number of heterozygous nucleotide sites maintained in a finite population due to steady flux of mutations. Genetics 61:893.
Kimura M, Ohta T (1978) Stepwise mutation model and distribution of allelic frequencies in a finite population. Proc Natl Acad Sci 75:2868–2872.
Koen EL, Bowman J, Garroway CJ, Wilson PJ (2013) The sensitivity of genetic connectivity measures to unsampled and under-sampled sites. PLoS ONE 8:e56204.
Landguth EL, Cushman SA (2010) CDPOP: A spatially explicit cost distance population genetics program. Mol Ecol Resour 10:156–161.
Landguth EL, Cushman SA, Johnson NA (2012a) Simulating natural selection in landscape genetics. Mol Ecol Resour 12:363–368.
Landguth EL, Cushman SA, Schwartz MK, et al. (2010) Quantifying the lag time to detect barriers in landscape genetics. Mol Ecol 19:4179–4191.

Landguth EL, Fedy BC, Oyler-McCance SJ, et al. (2012b) Effects of sample size, number of markers, and allelic richness on the detection of spatial genetic pattern. Mol Ecol Resour 12:276–284.
Latta RG (2006) Integrating patterns across multiple genetic markers to infer spatial processes. Landsc Ecol 21:809–820.
Legendre P, Fortin M-J (2010) Comparison of the Mantel test and alternative approaches for detecting complex multivariate relationships in the spatial analysis of genetic data. Mol Ecol Resour 10:831–844.
Manel S, Albert C, Yoccoz N (2012) Sampling in Landscape Genomics. In: Pompanon F, Bonin A (eds) Data Prod. Anal. Popul. Genomics. Humana Press, pp 3–12
Manel S, Holderegger R (2013) Ten years of landscape genetics. Trends Ecol Evol 28:614–621.
Manel S, Poncet BN, Legendre P, et al. (2010) Common factors drive adaptive genetic variation at different spatial scales in *Arabis alpina*. Mol Ecol 19:3824–3835.
Manel S, Schwartz MK, Luikart G, Taberlet P (2003) Landscape genetics: combining landscape ecology and population genetics. Trends Ecol Evol 18:189–197.
Manni F, Guerard E, Heyer E (2004) Geographic patterns of (genetic, morphologic, linguistic) variation: how barriers can be detected by using Monmonier's algorithm. Hum Biol 76:173–190.
McGarigal K, Cushman S, Ene E (2012) FRAGSTATS v4: Spatial Pattern Analysis Program for Categorical and Continuous Maps. University of Massachusetts, Amherst
McRae BH (2006) Isolation by resistance. Evolution 60:1551–1561.
Meeuwig MH, Guy CS, Kalinowski ST, Fredenberg WA (2010) Landscape influences on genetic differentiation among bull trout populations in a stream-lake network. Mol Ecol 19:3620–3633.
Monmonier MS (1973) Maximum-difference barriers: an alternative numerical regionalization method. Geogr Anal 5:245–261.
Murphy MA, Dezzani R, Pilliod DS, Storfer A (2010a) Landscape genetics of high mountain frog metapopulations. Mol Ecol 19:3634–3649.
Murphy MA, Evans JS, Cushman SA, Storfer A (2008) Representing genetic variation as continuous surfaces: an approach for identifying spatial dependency in landscape genetic studies. Ecography 31:685–697.
Murphy MA, Evans JS, Storfer A (2010b) Quantifying *Bufo boreas* connectivity in Yellowstone National Park with landscape genetics. Ecology 91:252–261.
Nei M (1972) Genetic distance between populations. Am Nat 106:283–292.
Nei M, Tajima F, Tateno Y (1983) Accuracy of estimated phylogenetic trees from molecular data. J Mol Evol 19:153–170.
O'brien RM (2007) A caution regarding rules of thumb for variance inflation factors. Qual Quant 41:673–690.
Oyler-McCance SJ, Fedy BC, Landguth EL (2013) Sample design effects in landscape genetics. Conserv Genet 14:275–285.
Pavlacky Jr DC, Goldizen AW, Prentis PJ, et al. (2009) A landscape genetics approach for quantifying the relative influence of historic and contemporary habitat heterogeneity on the genetic connectivity of a rainforest bird. Mol Ecol 18:2945–2960.
Pease KM, Freedman AH, Pollinger JP, et al. (2009) Landscape genetics of California mule deer (*Odocoileus hemionus*): the roles of ecological and historical factors in generating differentiation. Mol Ecol 18:1848–1862.

Rico Y, Holderegger R, Boehmer HJ, Wagner HH (2014) Directed dispersal by rotational shepherding supports landscape genetic connectivity in a calcareous grassland plant. Mol Ecol 23:832–842.
Rousset F (1997) Genetic differentiation and estimation of gene flow from F-statistics under isolation by distance. Genetics 145:1219–1228.
Rousset F (2000) Genetic differentiation between individuals. J Evol Biol 13:58–62.
Saenz-Romero C, Guries RP, Monk AI (2001) Landscape genetic structure of *Pinus banksiana*: allozyme variation. Can J Bot 79:871–878.
Schoville SD, Bonin A, François O, et al. (2012) Adaptive genetic variation on the landscape: methods and cases. Annu Rev Ecol Evol Syst 43:23–43.
Schwartz MK, McKelvey KS (2008) Why sampling scheme matters: the effect of sampling scheme on landscape genetic results. Conserv Genet 10:441–452.
Schwartz MK, McKelvey KS, Cushman SA, Luikart G (2010) Landscape Genomics: A Brief Perspective. In: Cushman SA, Huettmann F (eds) Spat. Complex. Inform. Wildl. Conserv. Springer Japan, Tokyo, pp 165–174
Segelbacher G, Cushman SA, Epperson BK, et al. (2010) Applications of landscape genetics in conservation biology: concepts and challenges. Conserv Genet 11:375–385.
Shirk AJ, Cushman SA, Landguth EL (2012) Simulating Pattern-Process Relationships to Validate Landscape Genetic Models. Int J Ecol 2012:1–8.
Shirk AJ, Wallin DO, Cushman SA, et al. (2010) Inferring landscape effects on gene flow: a new model selection framework. Mol Ecol 19:3603–3619.
Slatkin M (1993) Isolation by distance in equilibrium and non-equilibrium populations. Evolution 47:264–279.
Slatkin M (1987) Gene flow and the geographic structure of natural populations. Science 236:787–792.
Slatkin M (1995) A measure of population subdivision based on microsatellite allele frequencies. Genetics 139:457–462.
Sokal RR, Michener CD (1958) A statistical method for evaluating systematic relationships. Univ Kans Sci Bull 38:1409–1438.
Sork VL, Davis FW, Westfall R, et al. (2010) Gene movement and genetic association with regional climate gradients in California valley oak (*Quercus lobata* Née) in the face of climate change. Mol Ecol 19:3806–3823.
Spear SF, Balkenhol N, Fortin M-J, et al. (2010) Use of resistance surfaces for landscape genetic studies: considerations for parameterization and analysis. Mol Ecol 19:3576–3591.
Spear SF, Storfer A (2010) Anthropogenic and natural disturbance lead to differing patterns of gene flow in the Rocky Mountain tailed frog, *Ascaphus montanus*. Biol Conserv 143:778–786.
Storfer A, Murphy MA, Evans JS, et al. (2006) Putting the "landscape" in landscape genetics. Heredity 98:128–142.
Storfer A, Murphy MA, Spear SF, et al. (2010) Landscape genetics: where are we now? Mol Ecol 19:3496–3514.
Van Strien MJ, Keller D, Holderegger R (2012) A new analytical approach to landscape genetic modelling: least-cost transect analysis and linear mixed models. Mol Ecol 21:4010–4023.
Wagner HH, Fortin M-J (2013) A conceptual framework for the spatial analysis of landscape genetic data. Conserv Genet 14:253–261.

Wang IJ (2013) Examining the full effects of landscape heterogeneity on spatial genetic variation: a multiple matrix regression approach for quantifying geographic and ecological isolation. Evolution 67:3403–3411.
Wang IJ (2011) Choosing appropriate genetic markers and analytical methods for testing landscape genetic hypotheses. Mol Ecol 20:2480–2482.
Wang IJ, Glor RE, Losos JB (2013) Quantifying the roles of ecology and geography in spatial genetic divergence. Ecol Lett 16:175–182.
Wasserman TN, Cushman SA, Shirk AS, et al. (2012) Simulating the effects of climate change on population connectivity of American marten (*Martes americana*) in the northern Rocky Mountains, USA. Landsc Ecol 27:211–225.
Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population structure. Evolution 38:1358–1370.
Wilson GA, Rannala B (2003) Bayesian inference of recent migration rates using multilocus genotypes. Genetics 163:1177.
Womble WH (1951) Differential systematics. Science 114:315–322.
Wright S (1943) Isolation by distance. Genetics 28:114.
Wright S (1931) Evolution in Mendelian populations. Genetics 16:97.
Wu B, Liu N, Zhao H (2006) PSMIX: an R package for population structure inference via maximum likelihood method. BMC Bioinformatics 7:317.
Zeller KA, McGarigal K, Whiteley AR (2012) Estimating landscape resistance to movement: a review. Landsc Ecol 27:777–797.
Zellmer AJ, Knowles LL (2009) Disentangling the effects of historic vs. contemporary landscape structure on population genetic divergence. Mol Ecol 18:3593–3602.

Chapter 2. Validating dispersal distances inferred from occupancy models with genetic parentage assignments.

Abstract

Dispersal distances are commonly inferred from occupancy data but have rarely been validated. We compared dispersal distances inferred from seven years of occupancy data for 192 wetlands in metapopulations of the threatened California black rail (*Laterallus jamaicensis coturniculus*) and the more vagile Virginia rail (*Rallus limicola*) to distances between parent-offspring dyads identified from genetic parentage assignments. Buffer radius (BRM) and incidence function (IFM) connectivity measures were calculated at different spatial scales and used as covariates of colonization in multi-season occupancy models. The two rail species exhibited contrasting relationships between connectivity and colonization. For black rails, AIC weights were greatest at 3 km for both BRM and IFM, and model-averaged dispersal distances (BRM = 3.46 km; IFM = 3.70 km) showed good agreement with the mean dispersal distance from parentage assignments (5.58 ± 1.92 km). For Virginia rails, however, occupancy models without a connectivity covariate ranked first and third in the IFM and BRM model sets, respectively, making it difficult to identify a scale with the best fit. A significantly smaller proportion of Virginia rail parent-offspring dyads represented dispersal events compared to black rail dyads, suggesting more frequent dispersal outside of the 2000 km^2 study area.

Keywords: Connectivity, Parentage analysis, Dispersal, Occupancy modeling, Metapopulation, Black rail, Virginia rail

Introduction

The distance that individuals disperse between habitat patches - a key parameter in many ecological models - is a critical research gap due in part to the logistical constraints and high cost of obtaining accurate estimates (Sutherland et al. 2013, Driscoll et al. 2014). Occupancy data are easily collected and commonly used to infer dispersal distance (Driscoll et al. 2014). In metapopulations, dispersal distance can be inferred using the relationship between colonization of unoccupied patches and their connectivity at various scales (van Langevelde 2000, Pellet et al. 2007, Prugh 2009, van Langevelde and Wynhoff 2009). Connectivity, the inverse of isolation, was classically defined as the number of immigrants arriving at a patch from all other patches in the metapopulation (Hanski 1994, 1999), which is inherently difficult to measure. Instead, connectivity is often approximated using characteristics of the landscape, including patch areas to represent the number of potential dispersers and inter-patch distances to represent isolation. Two commonly used connectivity estimators include the buffer radius measure (BRM) and the incidence function measure (IFM; Moilanen and Nieminen 2002). The BRM sums the area of occupied patches within a specified radius, r, of a focal patch. In contrast, the IFM multiplies the area of each occupied patch by a dispersal parameter, α (where $1/\alpha$ represents the average dispersal distance), and then sums the products for all occupied patches in the metapopulation. Both measures have a parameter (r in BRMs and α in IFMs) that represents dispersal distance.

Few studies have empirically validated dispersal distances inferred from occupancy data against direct measures of dispersal (but see Wahlberg et al. 2002). Capture-recapture and

telemetry studies can be used to measure dispersal distances, but they require extensive field effort and expense to capture, tag, and relocate individuals, which may or may not survive, until dispersal has occurred. They may also fail to detect long-distance dispersal events, because such studies are usually limited to small sample sizes and spatial extents (Koenig et al. 1996, Nathan et al. 2003). Genetic parentage assignments offer another way to estimate dispersal distances, including both short- and long-distance dispersal, by measuring the spatial distance between parent-offspring dyads (Palsbøll 1999, Telfer et al. 2003, Nutt 2008, Woltmann et al. 2012). Parentage assignments do not require genetic differentiation of populations, allowing for the estimation of dispersal distances even when gene flow among populations is high (Marshall et al. 1998). Thus, they may offer better resolution to detect dispersal events within a metapopulation than population assignments or genetic distance measures (Palsbøll et al. 2010). Moreover, if individuals are sampled throughout the study area, then the spatial resolution of parentage assignments should match the scale of dispersal distance inferred from occupancy data.

We validated dispersal distances inferred from occupancy data using parentage assignments in two species of secretive wetland birds, the rare and threatened California black rail (*Laterallus jamaicensis coturniculus*; hereafter black rail) and the more common Virginia rail (*Rallus limicola*), that co-occur in wetlands of the Sierra Foothills (Richmond et al. 2008, 2010). Black and Virginia rails have similar life history traits, but differ in their apparent vagilities. The larger Virginia rail (mean = 91 g, n = 41; unpubl. data) is widespread across North America, whereas the smaller black rail (mean = 29 g, n = 158; unpubl. data) has a highly disjunct and restricted distribution (Eddleman et al. 1994, Conway 1995). Median dispersal distance estimates for these species from incidence function models differed greatly (8 km for black rails and 1890 km for Virginia rails; Risk et al. 2011). Thus, we predicted that dispersal distances estimated from occupancy data and parentage assignments would be shorter for black rails than for Virginia rails, and that connectivity measures calculated at smaller spatial scales would be more strongly related to colonization in black rails than Virginia rails. We also expected to detect fewer dispersal events in Virginia rails than in black rails, because the longer dispersal distances of Virginia rails should increase the probability that they emigrated from the 2000 km^2 study area.

To test these predictions for each species, we assessed the fit of BRM and IFM connectivity measures, calculated at different spatial scales that corresponded to different values of r or α, as covariates of colonization in occupancy models using seven years of occupancy data from 192 wetland patches. Dispersal distances were model-averaged across spatial scales for the BRM and IFM occupancy model sets and compared to the spatial distances between parent-offspring dyads identified with microsatellite loci. In our dynamic rail metapopulations, occupancy models and parentage assignments measure similar dispersal processes. Turnover is common (Risk et al. 2011) and patches can change suitability within months (Richmond et al. 2012). Thus, colonization could result from movements of juveniles or adults of either sex at any time of year. Further, dispersing individuals could move among multiple wetlands before detection by annual occupancy surveys or capture for parentage analysis. Because rails are secretive and rarely seen, we were unable to distinguish between age classes or sexes in our occupancy surveys. Likewise, we could not differentiate between juvenile or adult dispersal using parentage assignments, because parent-offspring dyads were often captured after offspring had matured to adults.

Methods

Study system

Black and Virginia rails in the Sierra Foothills of California occupied 136 of 192 small wetlands (mean ± SE = 1.12 ± 0.13 ha) surveyed from 2007 to 2013 (Appendix A: Fig. A1). Rails are non-migratory, year-round residents in this region, but turnover (i.e., extinction and re-colonization) occurs regularly (Richmond et al. 2008, Risk et al. 2011). Metapopulations of both species exhibited disequilibrium, and occupancy declined during the study (unpubl. data), unlike earlier years characterized by dynamic equilibrium (Risk et al. 2011). Gene flow among Sierra Foothills wetlands is unmeasured for Virginia rails but is high for black rails, meeting the assumption of panmixia for parentage assignments (Girard et al. 2010).

Occupancy data collection and modeling

Occupancy surveys for both species were conducted during the breeding season (May to August) using established call-playback methods and patches were surveyed up to three times per year (Appendix A; Richmond et al. 2008, 2010). Wetland patches were mapped in the field using a backpack Trimble GPS unit with 0.5 m accuracy, and patch area and Euclidean distances between patch edges were determined using ArcGIS v.9.0 (ESRI). Connectivity measures were calculated using values of 1, 2, 3, 4, 5, 6, 7, 10, 15, 20, 25, and 30 km for r in BRMs and $1/\alpha$ in IFMs. Annual and mean connectivity from 2007 to 2013 were calculated at each scale for each patch and log-transformed (+1). Annual connectivity measures were highly correlated at all spatial scales (Pearson correlation coefficient matrices r_p = 0.48-0.99). Therefore, we used mean connectivity measures from 2007 to 2013 in occupancy models as covariates of colonization.

To determine which scale of r or $1/\alpha$ provided the best fit for colonization patterns in our occupancy data, we implemented single-species, multi-season occupancy models, that do not assume dynamic equilibrium and account for imperfect detection, in program PRESENCE (MacKenzie et al. 2003). PRESENCE models initial occupancy (ψ), colonization (γ), extinction (ε), and detection (ρ), and assesses model fit using Akaike's Information Criterion (AIC, Burnham and Anderson 2002). For each species, we developed two occupancy model sets (one for BRMs and one for IFMs) using a three-step process. First, we identified a dynamics model with the best fit by examining the effects of patch-area and year on initial occupancy, colonization, and extinction. Second, using the best dynamics model, we identified a detection model with the best fit using year, Julian date within year, and time of day as detection covariates. Third, using the best model from step two, we assessed the fit of BRM and IFM connectivity at different scales of r and $1/\alpha$ modeled as covariates of colonization. We calculated a model-averaged dispersal distance estimate for each species by summing the products of the AIC weights and the scale of connectivity within each model set.

Genetic sample collection and parentage assignments

We collected blood samples from 213 black rails and 98 Virginia rails captured throughout the Sierra Foothills using mist nets (Appendix A: Fig. A1; Girard et al. 2010). Most captures occurred between 2007 and 2013, but two black rails captured in 2004 were included.

Blood was collected from each bird by venipuncture of a wing vein. Age (after-hatch-year or AHY; hatch-year or HY) was determined using eye color and plumage characteristics (Pyle 2008), and sex was determined genetically following Griffiths et al. (1998). Birds were banded with U.S. Geological Survey stainless steel bands, and morphometric measurements and molt were recorded.

DNA was extracted from blood using phenol/chloroform (Sambrook et al. 1989) or a DNEasy Spin Column Kit (Qiagen Inc.). Polymerase chain reaction was used to amplify 17 microsatellites in black rails, developed by Girard et al. (Molecular Ecology Resources Primer Development Consortium 2009), and 15 microsatellites in Virginia Rails, developed for this study (Appendix A: Table A1). The experimental conditions were similar for all loci following Girard et al. (2010), with annealing temperatures ranging from 52^{o}C to 59^{o}C (Appendix A: Table A1). Microsatellites were labeled with fluorescent dye (HEX or FAM), run using POP buffer and LIZ 500 size standard on an ABI3730 Genetic Analyzer (Applied Biosystems Inc.), and scored in GeneMapper v.5 (Applied Biosystems Inc.). Microsatellite genotypes were assessed for possible deviations from the expected Hardy-Weinberg equilibrium (HWE) and linkage equilibrium (LE) using GENEPOP v.4.2 (Appendix A; Raymond and Rousset 1995).

We estimated dispersal distance as the spatial distance between capture locations of individuals identified as parent and offspring with microsatellite genotypes using program ML-Relate (Kalinowski et al. 2006). We chose ML-Relate because it calculates maximum likelihood values for different types of relationships (unrelated, half sibling, full sibling, or parent-offspring) for each pair in a given data set, and it does not require estimates of the number of candidate parents or proportion of parents sampled like other parentage assignment software, such as Cervus (Kalinowski et al. 2007). Parentage assignments of rails using Cervus were highly sensitive to variability in these parameters, which we could not estimate for our secretive rail metapopulations. All rail dyads with positive LOD scores, which measure the natural log of the likelihood of a pair being related as parent and offspring divided by the likelihood that the pair is unrelated, were considered candidate parent-offspring pairs (Skaug 2001).

Parentage assignment errors are made by either incorrectly assigning two unrelated individuals as a parent-offspring dyad (type I error), or by failing to detect two related individuals as parent and offspring (type II error). The number of assignment errors in a given analysis is influenced by the number of genetic markers used, their allelic diversity, the proportion of candidate parents sampled, genotyping error rates, and the presence of null alleles (Marshall et al. 1998, Harrison et al. 2013). We assessed assignment errors by calculating *p*-values for assignments and accounting for multiple testing using a false discovery rate (FDR; Skaug et al. 2010). We calculated *p*-values by comparing the LOD score for each candidate pair to a distribution of LOD scores from 500 simulated data sets with the same sample sizes and allele frequencies as the empirical data sets for each rail species following Skaug et al. (2010) in program R (ver. 3.0.2; R Core team 2013). We used an FDR (Benjamini and Hochberg 1995) to estimate the number of correctly assigned pairs at different levels of statistical significance (FDRs of 1.00, 0.10, and 0.05; Skaug et al. 2010). We further excluded pairs that did not represent dispersal events, including juveniles that had not dispersed from their natal patches (i.e., AHY-HY pairs captured in the same patch) and sibling pairs (i.e., HY-HY pairs captured in the same patch). Following exclusions, median and mean (± SE) dispersal distances were

calculated from the remaining pairs at each FDR. Median and mean dispersal distances were compared among FDRs to explore the effects of assignment errors on dispersal distance estimates. Last, we tested for differences between species in the number of parent-offspring dyads that represented dispersal events at an FDR of 0.05 using a Fisher's exact test.

Results

Dispersal estimates from occupancy models

The best dynamics and detection models in occupancy model sets were similar for black and Virginia rails (Appendix A: Table A2). For both species, the top dynamics model included area as a covariate for initial occupancy and extinction, and year as a covariate for colonization and extinction. The top detection model for black rails included year and Julian date, and for Virginia rails included year, Julian date, and time of day. Mean site-level detection probabilities were high for black (p = 0.98; range: 0.93 – 0.99) and Virginia rails (p = 0.94; range: 0.88 – 0.98).

The two rail species exhibited contrasting relationships between connectivity and colonization (Fig. 1b, d; Appendix A: Table A3). For black rails, colonization was strongly influenced by connectivity at small spatial scales. Moreover, AIC weights were greatest at 3 km for both BRM and IFM, and model-averaged dispersal distances were remarkably similar (BRM = 3.46 km and IFM = 3.70 km; Fig. 1b). Models without connectivity as a colonization covariate had AIC weights of zero in black rail model sets (Appendix A: Table A3). For Virginia rails, however, connectivity measures were less strongly related to colonization; models without connectivity as a covariate ranked first and third in the IFM and BRM model sets, respectively (Appendix A: Table A3). AIC weights for IFM models differed little among scales and were unable to resolve dispersal distance for Virginia rails (Fig. 1d; Appendix A: Table A3). The best scale for BRM models was 3 km, but it had much less support (w = 0.40) than the top model in the black rail BRM model set (w = 0.85). As a result, model-averaged dispersal estimates for Virginia rails (BRM = 6.36 km; IFM= 9.31 km; Fig. 1d) were two to three times greater than for black rails.

Dispersal estimates from parentage assignments

One microsatellite locus (C2) for black rails and four microsatellite loci (C124, C127, C188, and D126) for Virginia rails were excluded from our analysis because of deviations from the expected HWE and LE (Appendix A: Table A1). The remaining 16 black rail microsatellites and 11 Virginia rail microsatellites were used in subsequent analyses.

At FDRs of 1.00, 0.10, and 0.05, we identified 65, 38, and 32 parent-offspring dyads of black rails and 25, 18, and 14 parent-offspring dyads of Virginia rails, respectively (Appendix A: Table A4). Fourteen of 65 (22%) black rail dyads and 14 of 25 (56%) Virginia rail dyads were excluded from dispersal estimates because they were juveniles that had not yet dispersed from their natal patch (black rail n = 13, Virginia rail n = 9), or were siblings captured in the same patch (black rail n = 1, Virginia rail n = 5; Appendix A: Table A4). At the strictest FDR (0.05) a significantly smaller proportion of Virginia rail parent-offspring dyads represented dispersal

events compared to black rail dyads (4 of 14 = 29% and 23 of 32 = 72%, respectively; Fisher's exact test, $p < 0.01$).

Spatial distances between all 65 black rail dyads ranged from 0 to 81.38 km and for all 25 Virginia rail dyads ranged from 0 to 63.08 km. The choice of FDR employed in our parentage assignments strongly affected the median and mean estimates of dispersal distance (Fig. 1a, c). For black rails, median and mean (± SE) dispersal distances decreased from 5.81 km and 15.96 ± 3.30 km, respectively, when all 65 pairs were considered (FDR = 1.00) to 1.05 km and 8.54 ± 2.67 km at an FDR of 0.10, and to 0.18 km and 5.58 ± 1.92 km at an FDR of 0.05. There were too few Virginia rail parent-offspring dyads detected to make meaningful comparisons of dispersal distance estimates at different FDRs, so we report only the median (1.79 km) and mean ± SE (4.66 ± 3.04 km) dispersal distances for an FDR of 0.05 (Fig. 1c). We compared the median and mean dispersal distances of 23 black rails and 4 Virginia rails selected using an FDR of 0.05 to our model-averaged dispersal distances inferred from occupancy models for validation (Fig. 1b, d; Appendix A: Table A4).

Discussion

The performance of dispersal distance estimators from occupancy models differed between black and Virginia rails. For black rails, AIC weights were greatest for BRM and IFM models using connectivity calculated at 3 km, and model-averaged dispersal distances (3.46 km and 3.70 km from the BRM and IFM models, respectively) showed good agreement with median (0.18 km) and mean (5.58 ± 1.92 km) dispersal distances from parentage assignments (Fig. 1b). For Virginia rails, however, the AIC weights for BRM and IFM models differed less among scales, especially for IFM models, making it difficult to identify a single scale with the best fit for our occupancy data (Fig. 1d). In contrast to black rails, that had model-averaged dispersal distances that fell between the median and mean distances from parentage assignments, the model-averaged dispersal distances for Virginia rails (6.36 km and 9.31 km from the BRM and IFM models, respectively) exceeded the median (1.79 km) and mean dispersal distances from parentage assignments (4.66 ± 3.04 km; Fig. 1b, d). However, our inference is limited by the small number of Virginia rail parent-offspring dyads that represented dispersal events.

A significantly smaller proportion of Virginia rail parent-offspring dyads represented dispersal events compared to black rail dyads ($p < 0.01$) suggesting that Virginia rail dyads had lower capture probabilities than black rail dyads and supporting the prediction that Virginia rails disperse greater distances to patches outside our study area. Further support was provided by model-averaged dispersal distances from BRM and IFM for Virginia rails that were two to three times greater than for black rails (Fig. 1b, d). If Virginia rails dispersed outside our study area, our estimates of Virginia rail dispersal distances from occupancy models and parentage assignments would be biased downward. Nonetheless, our results obtained during non-equilibrium conditions agree with those of Risk et al. (2011), who suggested that Virginia rails disperse greater distances than black rails, based on incidence function analysis of occupancy data collected primarily under conditions of stable occupancy (dynamic equilibrium).

We estimated dispersal distances by fitting occupancy models with colonization covariates that measured connectivity (the inverse of isolation) at different scales based on the

premise that isolation affects colonization and that the effect of isolation on colonization should be stronger in species with limited dispersal (Hanski and Thomas 1994, Hanski 1999, Prugh et al. 2008). As predicted, we observed a stronger relationship between the scale of connectivity and colonization in black rails than in Virginia rails. In general, the effect of isolation on colonization appears to be highly variable among taxa, with some studies detecting significant effects (van Langevelde 2000, Wahlberg et al. 2002, van Langevelde and Wynhoff 2009) and others detecting weak or no effects (Pellet et al. 2007, Robles and Ciudad 2012). Moreover, a meta-analysis by Prugh et al. (2008) concluded that the ability of isolation to predict occupancy was only weakly related to dispersal limitation. A more meaningful comparison among species of the effect of isolation on colonization and occupancy would include the relationship between the scale of dispersal and the size of the study area rather than simply categorizing species as more or less dispersive.

Dispersal distance estimators from occupancy data, such as r from BRMs and $1/\alpha$ from IFMs, are influenced by characteristics of the study organism, the landscape configuration, and the timing of occupancy surveys (Hanski 1999, Wahlberg et al. 2002). Our estimate of α for black rails (0.27) was similar to α estimated for juvenile water voles from a metapopulation in which isolation affects colonization ($\alpha = 0.33$; Sutherland et al. 2014), and our estimate of α for Virginia rails (0.11) was similar to α estimated for woodpeckers from a metapopulation in which isolation did not have a strong effect on colonization ($\alpha = 0.19$; Robles and Ciudad 2012).

In our dynamic rail metapopulations, colonization occurs from dispersal of juveniles and adults of both sexes. We were unable to distinguish between age classes or sexes in our occupancy surveys, or differentiate between juvenile and adult dispersal using parentage assignments (Appendix A: Table A4). In addition, the timing of dispersal events and the number of wetlands visited by rails during dispersal were unknown because our occupancy and genetic data were collected annually during the breeding season over multiple years. These limitations prevented us from making inferences about juvenile and adult dispersal distances, sex-biased dispersal, the timing of dispersal events, and whether or not dispersal occurred as a single movement or multiple movements between sampling periods. Nevertheless, limitations were similar for both the occupancy models and parentage assignments, suggesting the use of dispersal distance estimates from parent-offspring dyads to validate dispersal distances inferred from occupancy models is an appropriate comparison.

In conclusion, our study suggests that a reasonably accurate mean dispersal distance can be inferred from occupancy data for dispersal-limited species. For vagile species, however, connectivity may not strongly affect colonization, making it difficult to infer dispersal distances from occupancy data, even when the data are collected at a relatively large spatial scale.

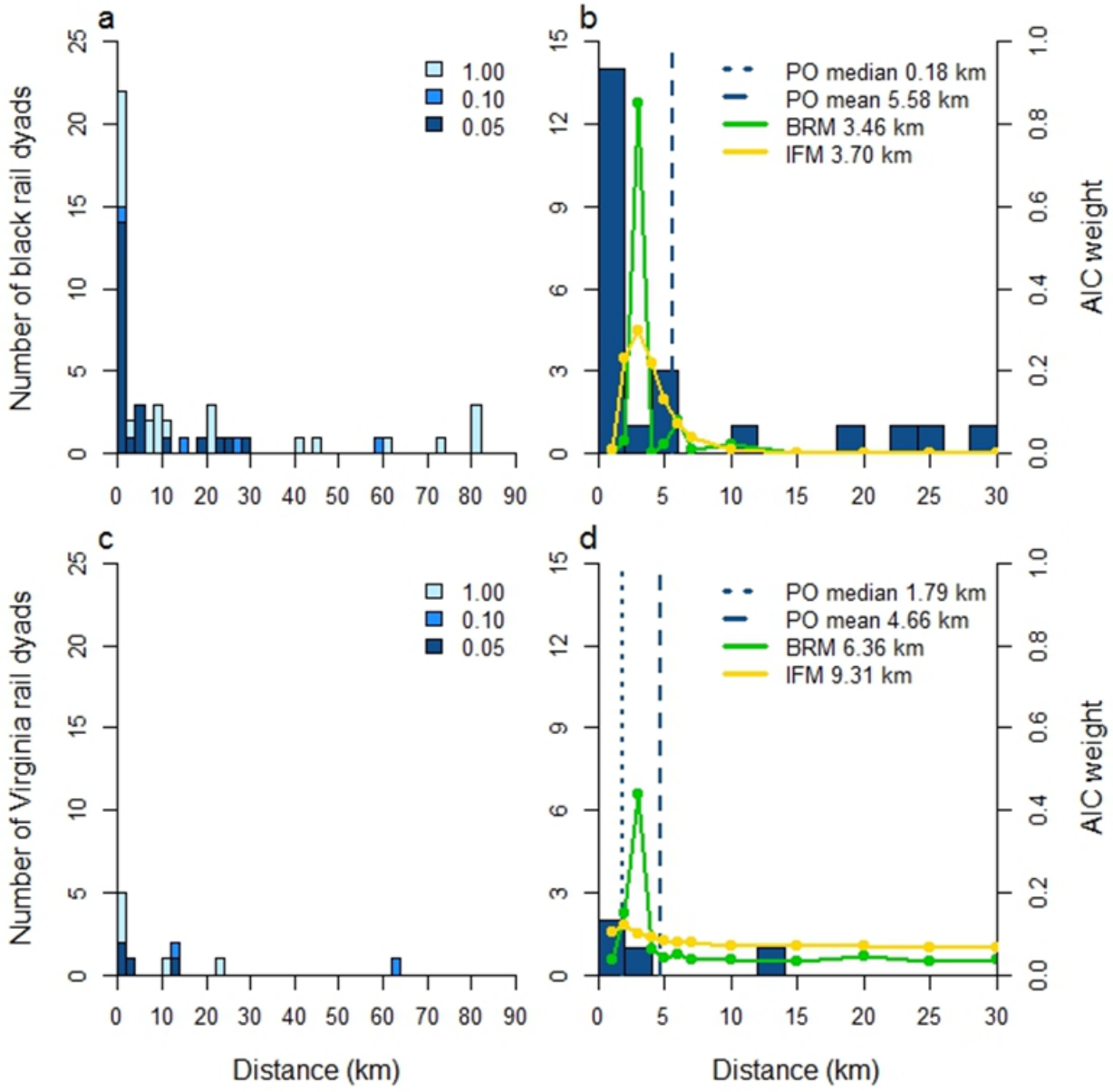
Acknowledgements

We thank J. Tecklin, O. Richmond, M. Gamboa, C. Cleveland, C. Stegall, N. Najar, T. Nguyen, C. Chu, K. Lovett, J. Bruce, M. Goodnow, B. Gartland, E. Hunter, L. Evans, L. Doll, T. Graham, and K. Spragens for assistance. Reviews by P. Palsbøll, P. Sunnucks, A. Lee, and the

Beissinger Lab greatly improved this manuscript. Financial support was provided by the National Science Foundation DEB-1051342 and CNH 1115069.

Figures

Figure 1. Frequency of dispersal distances between parent-offspring dyads of (a) California black rails (*Laterallus jamaicensis coturniculus*) and (c) Virginia rails (*Rallus limicola*) at false discovery rates (FDRs) of 1.00, 0.10, and 0.05. The increasingly blue color gradient indicates increasing statistical confidence (smaller FDRs) in parentage assignments. The median and mean dispersal distances from parentage assignments at an FDR of 0.05 compared to model-averaged dispersal distance estimates from occupancy models in (b) California black rails and (d) Virginia rails. AIC weights for occupancy models that incorporated buffer radius (BRM, solid yellow line) and incidence function (IFM, solid brown line) connectivity measures calculated at scales from one to 30 km are overlaid on histograms of dispersal distances from parentage assignments.



Literature Cited

Benjamini, Y., and Y. Hochberg. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society Series B. 57:289–300.

Burnham, K. P., and D. R. Anderson. 2002. Model selection and multimodel inference: a practical information-theoretic approach. 2nd edition. Springer-Verlag, New York.

Conway, C. J. 1995. Virginia Rail (*Rallus limicola*). The Birds of North America Online (A. Poole, Ed.). Ithaca: Cornell Lab of Ornithology; http://bna.birds.cornell.edu/bna/species/173.

Driscoll, D. A., S. C. Banks, P. S. Barton, K. Ikin et al. 2014. The trajectory of dispersal research in conservation biology. Systematic review. PLoS ONE 9:e95053.

Eddleman, W. R., R. E. Flores, and M. Legare. 1994. Black Rail (*Laterallus jamaicensis*). The Birds of North America Online (A. Poole, Ed.). Ithaca: Cornell Lab of Ornithology; http://bna.birds.cornell.edu/bna/species/123.

Girard, P., J. Y. Takekawa, and S. R. Beissinger. 2010. Uncloaking a cryptic, threatened rail with molecular markers: origins, connectivity and demography of a recently-discovered population. Conservation Genetics 11:2409–2418.

Griffiths, R., M. C. Double, K. Orr, and R. J. Dawson. 1998. A DNA test to sex most birds. Molecular Ecology 7:1071–1075.

Hanski, I. 1994. A practical model of metapopulation dynamics. Journal of Animal Ecology 63:151–162.

Hanski, I. 1999. Metapopulation Ecology. Oxford University Press, New York, New York.

Hanski, I., and C. Thomas. 1994. Metapopulation dynamics and conservation: A spatially explicit model applied to butterflies. Biological Conservation 68:167–180.

Harrison, H. B., P. Saenz-Agudelo, S. Planes, G. P. Jones, and M. L. Berumen. 2013. Relative accuracy of three common methods of parentage analysis in natural populations. Molecular Ecology 22:1158–1170.

Kalinowski, S. T., M. L. Taper, and T. C. Marshall. 2007. Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. Molecular Ecology 16:1099–1106.

Kalinowski, S. T., A. P. Wagner, and M. L. Taper. 2006. ML-Relate: a computer program for maximum likelihood estimation of relatedness and relationship. Molecular Ecology Notes 6:576–579.

Koenig, W. D., D. Van Vuren, and P. N. Hooge. 1996. Detectability, philopatry, and the distribution of dispersal distances in vertebrates. Trends in Ecology & Evolution 11:514–517.

van Langevelde, F. 2000. Scale of habitat connectivity and colonization in fragmented nuthatch populations. Ecography 23:614–622.

van Langevelde, F., and I. Wynhoff. 2009. What limits the spread of two congeneric butterfly species after their reintroduction: quality or spatial arrangement of habitat?: Spread of butterfly species after reintroduction. Animal Conservation 12:540–548.

MacKenzie, D. I., J. D. Nichols, J. E. Hines, M. G. Knutson, and A. B. Franklin. 2003. Estimating site occupancy, colonization, and local extinction when a species is detected imperfectly. Ecology 84:2200–2207.

Marshall, T. C., J. Slate, L. E. B. Kruuk, and J. M. Pemberton. 1998. Statistical confidence for likelihood-based paternity inference in natural populations. Molecular Ecology 7:639–655.

Moilanen, A., and M. Nieminen. 2002. Simple connectivity measures in spatial ecology. Ecology 83:1131–1145.
Molecular Ecology Resources Primer Development Consortium. 2009. Permanent genetic resources added to Molecular Ecology Resources Database 1 May 2009-31 July 2009. Molecular Ecology Resources 9:1460–1559.
Nathan, R., G. Perry, J. T. Cronin, A. E. Strand, and M. L. Cain. 2003. Methods for estimating long-distance dispersal. Oikos 103:261–273.
Nutt, K. J. 2008. A comparison of techniques for assessing dispersal behaviour in gundis: revealing dispersal patterns in the absence of observed dispersal behaviour. Molecular Ecology:3541–3556.
Palsbøll, P. J. 1999. Genetic tagging: contemporary molecular ecology. Biological Journal of the Linnean Society 68:3–22.
Palsbøll, P. J., M. Z. Peery, and M. Berube. 2010. Detecting populations in the "ambiguous" zone: kinship-based estimation of population structure at low genetic divergence. Molecular Ecology Resources 10:797–805.
Pellet, J., E. Fleishman, D. S. Dobkin, A. Gander, and D. D. Murphy. 2007. An empirical evaluation of the area and isolation paradigm of metapopulation dynamics. Biological Conservation 136:483–495.
Prugh, L. R. 2009. An evaluation of patch connectivity measures. Ecological Applications 19:1300–1310.
Prugh, L. R., K. E. Hodges, A. R. Sinclair, and J. S. Brashares. 2008. Effect of habitat area and isolation on fragmented animal populations. Proceedings of the National Academy of Sciences 105:20770–20775.
Pyle, P. 2008. Identification guide to North American birds: part II Anatidae to Alcidae. Slate Creek Press, Point Reyes Station, California.
Raymond, M., and F. Rousset. 1995. GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. Journal of Heredity 86:248.
R Core team. 2013. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
Richmond, O. M., J. Tecklin, and S. R. Beissinger. 2008. Distribution of California black rails in the Sierra Nevada foothills. Journal of Field Ornithology 79:381–390.
Richmond, O. M. W., J. E. Hines, and S. R. Beissinger. 2010. Two-species occupancy models: a new parameterization applied to co-occurrence of secretive rails. Ecological Applications 20:2036–2046.
Risk, B. B., P. De Valpine, and S. R. Beissinger. 2011. A robust-design formulation of the incidence function model of metapopulation dynamics applied to two species of rails. Ecology 92:462–474.
Robles, H., and C. Ciudad. 2012. Influence of habitat quality, population size, patch size, and connectivity on patch-occupancy dynamics of the middle spotted woodpecker. Conservation Biology 26:284–293.
Sambrook, J., E. F. Fritsch, and T. Maniatis. 1989. Molecular Cloning: A Laboratory Manual. Second edition. Cold Spring Harbor Laboratory Press, New York.
Skaug, H. J. 2001. Methods for estimation of population size. Biometrics 57:750–756.
Skaug, H. J., M. BéRubé, and P. J. Palsbøll. 2010. Detecting dyads of related individuals in large collections of DNA-profiles by controlling the false discovery rate. Molecular Ecology Resources 10:693–700.

Sutherland, W. J., R. P. Freckleton, H. C. J. Godfray, S. R. Beissinger et al. 2013. Identification of 100 fundamental ecological questions. Journal of Ecology 101:58–67.
Telfer, S., S. B. Piertney, J. F. Dallas, W. A. Stewart, F. Marshall, J. L. Gow, and X. Lambin. 2003. Parentage assignment detects frequent and large-scale dispersal in water voles. Molecular Ecology 12:1939–1949.
Wahlberg, N., T. Klemetti, V. Selonen, and I. Hanski. 2002. Metapopulation structure and movements in five species of checkerspot butterflies. Oecologia 130:33–43.
Woltmann, S., T. W. Sherry, and B. R. Kreiser. 2012. A genetic approach to estimating natal dispersal distances and self-recruitment in resident rainforest birds. Journal of Avian Biology 43:33–42.

Chapter 3. Dissecting metapopulation connectivity with combined inference from genetic and isotopic population assignments allows for differentiation of recent and older migrants.

Abstract

The stochastic and infrequent nature of long-distance dispersal makes it difficult to quantify dispersal between metapopulations at large spatial scales. We quantified the frequency and distance of long-distance dispersal of a non-migratory, secretive wetland bird, the California black rail (*Laterallus jamaicensis coturniculus*), between an inland and a coastal metapopulation separated by greater than 100 km. We combined inference from genetic and isotopic population assignments to classify rails as residents, recent migrants, or older migrants. Most rails (334 of 336 or 99%) were classified as residents of their capture populations, but we observed two migrants that had moved greater than 100 km to their capture populations more than one year before capture. In addition, isotopic population assignments alone classified seven rails as migrants. However, mapping of isotope values from wetland soils revealed high variation in some of the isotope values across small spatial scales (tens of meters). This spatial variation in the environment mirrored the variation observed in rail feathers, indicating that isotope values in feathers likely resulted from natural environmental variation (e.g., source element affects) rather than long-distance dispersal of individuals. Thus, these seven rails were most likely misassigned by isotopic population assignments due to small-scale, spatial variation in the isoscape. Combining inferences from genetic and isotopic population assignments allowed us to differentiate between residents, recent migrants, and older migrants, and to successfully track long-distance movements of non-migratory rails over regional spatial scales.

Keywords: Connectivity, Genetic population assignments, Stable isotopes, Linear discriminant function analysis, Dispersal, Metapopulation, Black rail

Introduction

Dispersal of individuals among habitat patches plays an important role in metapopulation dynamics at smaller spatial scales and can connect metapopulations at larger, regional spatial scales (Hanski 1999). Many studies have estimated dispersal distances within metapopulations (e.g., Wahlberg et al. 2002, Robles and Ciudad 2012, Sutherland et al. 2014), but the stochastic and infrequent nature of long-distance dispersal makes it difficult to quantify the frequency of dispersal between metapopulations (Koenig et al. 1996, Nathan et al. 2003). Genetic and isotopic markers offer a practical method for tracking long-distance movements. The frequency of dispersal and subsequent gene flow between metapopulations can be assessed with measures of genetic differentiation (F_{ST}; Wright 1931), or by estimating recent migration rates (*m*; Wilson and Rannala 2003). However, these measures cannot be used to estimate dispersal distances, and they are less accurate when gene flow is high and genetic differentiation between populations is low (Faubet et al. 2007). On the other hand, genetic population assignments, that assign individuals to a population of origin, can measure both frequency and distance of long-distance dispersal events (Cornuet et al. 1999, Pritchard et al. 2000, Berry et al. 2004). Similarly, the analysis of the stable isotope composition of a range of animal tissues can assist in tracking long-distance dispersal of an individual between its capture population and its most probable

population of origin if populations occur at isotopically distinct locations (Hobson 1999, Rubenstein and Hobson 2004, Hobson and Wassenaar 2008). In combination, genetic and isotopic markers often contain sufficient variability to track annual movements of migratory species across large, continental spatial scales (Caccamise et al. 2000, Wennerberg et al. 2002, Kelly et al. 2005, Rundel et al. 2013), but they can lead to conflicting population assignments for non-migratory species at regional scales due to the timing of dispersal events and spatial variation of isotopes across the landscape (i.e., "isoscape"; West et al. 2010).

The timing of dispersal events does not affect population assignments made with genetic data because the genetic signature of an individual does not change over time, but can affect assignments made with isotopic data because the isotopic signature of an individual can change over time depending upon the individual's diet and the turnover rate of the tissue being sampled (Hobson and Wassenaar 2008). Even isotopically inert tissues, such as feathers or hair, only retain an isotopic signature until the tissue is molted or shed (Hobson and Wassenaar 2008). Therefore, if an individual was captured soon after dispersing and before isotopic turnover had occurred, both genetic and isotopic data should assign the individual to the same population of origin. However, if an individual was captured long enough after dispersing for isotopic turnover to have occurred, genetic data would assign the individual to its population of origin, but isotopic data would assign it to the population where it was recently foraging. This mismatch in inference between genetic and isotopic data in non-migratory species could be used to parse captured individuals into three groups: (1) resident individuals that have genetic and isotopic signatures assigning them to the population in which they were captured, (2) recent migrants that have genetic and isotopic signatures assigning them to a population that differs from their capture population, and (3) older migrants that have isotopic signatures assigning them to the population in which they were captured, but genetic signatures assigning them to a population that differs from their capture population.

Spatial variation in the isoscape may also lead to mismatches between genetic and isotopic population assignments. Differences in isotope values among populations are used to assign individuals to a population of origin. Thus, assignment accuracy depends on the isotopic similarity of individuals captured in close proximity to one another. However, small-scale, spatial variation in the biogeochemical processes that effect environmental concentrations of isotopes could lead to assignment errors. In this case, an individual would be correctly assigned to its capture population by its genetic signature, but it would be incorrectly assigned to a population that differs from its capture population by its isotopic signature.

We quantified the frequency and distance of dispersal events in a non-migratory, threatened wetland bird, the California black rail (*Laterallus jamaicensis coturniculus*), between coastal and inland metapopulations separated by 100 km of unsuitable habitat (Richmond et al. 2008). Generally, organisms foraging in marine ecosystems have more enriched stable isotope values than organisms foraging in freshwater ecosystems (Hobson and Wassenaar 2008). Thus, we predicted feathers of rails from more saline, coastal wetlands in the San Francisco Bay Area (Bay Area) would be more enriched in the heavier isotopes of carbon (^{13}C), nitrogen (^{15}N), and sulfur (^{34}S) than feathers of rails from inland, freshwater wetlands of the Sierra Foothills (Foothills). Further, black rails in the Bay Area metapopulation are genetically differentiated from the metapopulation in the Foothills, and previous genetic population assignments and

migration rate estimates suggest these metapopulations are linked by long-distance dispersal (Girard et al. 2010). We quantified the frequency and distance of long-distance dispersal for black rails between the Bay Area and Foothills by estimating F_{ST} and recent migration rates with genetic data, and by performing population assignments using isotopic and genetic data. We combined inferences from genetic and isotopic population assignments to classify individuals as: (1) residents, (2) recent (less than one year) migrants, and (3) older (greater than or equal to one year) migrants.

Methods

Sample collection

Blood (n = 336) and feather (n = 251) samples were collected from black rails captured throughout the Bay Area and Foothills (Fig. 1) from 2004 to 2013. Rails were captured during the day with mist nets and at night with spotlights and hand-nets (Tsao et al. 2009, Girard et al. 2010). Blood was collected by venipuncture of a wing vein and stored in Longmire's solution (Longmire et al. 1988). One secondary feather from each wing and four to six breast feathers were collected and stored in paper envelopes. Only breast feathers were collected for birds molting their secondary feathers. Age (after-hatch-year or AHY; hatch-year or HY) was determined using eye color and plumage characteristics (Pyle 2008), and sex was determined genetically with primers from Griffiths et al. (1998). Birds were banded with U.S. Geological Survey stainless steel bands, and morphometric measurements and molt were recorded.

Genetic data collection

DNA was extracted from blood using either phenol/chloroform (Sambrook et al. 1989) or a DNEasy Spin Column Kit (Qiagen Inc.). Seventeen microsatellite loci developed by Girard et al. (Molecular Ecology Resources Primer Development Consortium 2009) were amplified in black rails by polymerase chain reaction (PCR; Appendix 1). The experimental conditions were similar for all loci following Girard et al. (2010), with PCR annealing temperatures ranging from 52^{o}C to 59^{o}C (Appendix 1). Microsatellites were labeled with fluorescent dye (HEX or FAM), run using POP buffer and LIZ 500 size standard on an ABI3730 Genetic Analyzer (Applied Biosystems Inc.), and scored in GeneMapper ver. 5 (Applied Biosystems Inc.).

Isotopic data collection

Prior to isotopic analysis, feathers were soaked in a 2:1 chloroform/methanol solution for 24 hours to remove dirt and oil, dried for 24 hours, and then the vane was separated from the rachis and placed in tin capsules to obtain a mass ranging from 1.25 mg to 1.60 mg (Hobson and Wassenaar 2008). Feather samples were analyzed simultaneously for carbon, nitrogen, and sulfur stable isotope content (% dry weight) and ratios at the Center for Stable Isotope Geochemistry at the University of California, Berkeley using a continuous flow CHNOS Elemental Analyzer (vario ISOTOPE cube, Elementar, Hanau, Germany) coupled with an IsoPrime100 IRMS (Isoprime, Cheadle, UK) following the methods of (Mambelli et al. In review). We included feather standards from wild turkey (*Meleagris gallopavo*), common murre (*Uria aalge*), and Pacific loon (*Gavia pacifica*) with each batch of rail samples to assess deviation among runs.

Feather standards were prepared using a Wiley mill to create a fine, homogenous powder from a combination of breast, body, and flight feathers from each species. Deviation across runs was similar for all three species used as standards with the smallest deviation observed for delta nitrogen, a moderate deviation observed for delta carbon, and the greatest deviation observed for delta sulfur (Appendix 2). Isotope ratios are reported in parts per thousand (‰) using delta notation with $\delta^{h}N = (R_{sample}/R_{standard}-1) \times 1000$, where R is the ratio of enriched to depleted isotopes for the sample and standard, N is the element of interest, and h is the mass of the enriched isotope. For $\delta^{13}C$, $\delta^{15}N$, and $\delta^{34}S$ values, Vienna Pee Dee Belemnite (VPDB), air (AIR), and Vienna Cañon Diablo Triolite (VCDT) were used as standards respectively. Bovine liver, reference material NIST (National Institute of Standards and Technology, Gaithersburg, MD, USA) SMR 1547, was used as a calibration standard for $\delta^{15}N$ and $\delta^{13}C$. Two organic in-house standards with different S isotope composition, fish material ($\delta^{34}S$ value of 18.40 ‰) and algae material ($\delta^{34}S$ value of 1.98 ‰), were calibrated against IAEA sulfur sulfides and were used as a calibration standard for $\delta^{34}S$. The long-term analytical precision of $\delta^{13}C$, $\delta^{15}N$, and $\delta^{34}S$ are ±0.10, ±0.15 and ±0.40 ‰, respectively. A full discussion of the accuracy and precision of the simultaneous analysis of all three isotopes in a simple sample can be found in Mambelli et al. (In review).

Data analysis

Microsatellite genotypes were assessed for possible deviations from the expected Hardy-Weinberg equilibrium (HWE) and linkage equilibrium (LE) using exact tests with 20 batches of 10,000 iterations and a dememorization of 10,000 in GENEPOP ver. 4.2 (Raymond and Rousset 1995). We adjusted *p*-values to account for multiple testing using a false discovery rate (FDR; Benjamini and Hochberg 1995) of 0.05. Loci that deviated significantly from expected HWE in either metapopulation were removed from the data set. We identified significant deviations from LE by splitting the data set for each metapopulation into two equally-sized data sets containing randomly assigned individuals. We then compared deviations from LE in the entire data set with the two smaller data sets. Pairs of loci that deviated from LE across all three data sets were considered linked, and one locus from each pair was removed from the data set (Waples 2015). We detected a marginally significant deviation from HWE at one microsatellite locus (C2) in the Foothills and one microsatellite locus in the Bay Area (D105) due to heterozygote deficiencies (Appendix 1). In addition, locus C2 was part of a pair of loci that consistently yielded significant deviations from LE (Appendix 1). Therefore, C2 and D105 were removed from subsequent analyses.

We measured genetic differentiation between metapopulations in the Bay Area and Foothills by calculating F_{ST} with the R package PopGenReport (Adamack and Gruber 2014), and migration rates (m) were estimated using Bayesass (Wilson and Rannala 2003) with 10,000,000 iterations, sampled every 100 iterations, following a burnin of 1,000,000 iterations. The number of genetic populations (K) was estimated in STRUCTURE (Pritchard et al. 2000) using the admixture model with the degree of admixture (α) inferred from the data, and the distribution of allele frequencies (λ) set to 1. For each value of K, ranging from one to four, we performed 10 runs of 100,000 iterations that were preceded by a burn-in of 1,000,000 iterations. The metapopulation where each rail was captured was used to inform cluster assignments for individuals using the LOCPRIOR setting (Hubisz et al. 2009). First generation migrants were

identified in GENECLASS2 using Λ, the ratio of the likelihood of drawing an individual's genotype from the population in which it was captured to the maximum likelihood observed for the individual's genotype in any population (Paetkau et al. 2004, Piry et al. 2004). Likelihoods were calculated from observed allele frequencies in each population using the Bayesian method of Rannala and Mountain (1997), and *p*-values were calculated for each individual by comparing the observed Λ to a distribution of Λ -values generated with 10,000 simulations (Paetkau et al. 2004). Individuals were identified as first generation migrants if their *p*-value was less than 0.01.

Differences in delta values of carbon, nitrogen, and sulfur isotopes between metapopulations (Bay Area and Foothills), age classes (AHY and HY), and sexes (male and female) were assessed using generalized linear models implemented in program R (R Core team 2013). Assignment probabilities used to identify recent migrants were generated using linear discriminant function analysis with cross validation of isotopic data in program R (R Core team 2013). Individuals were classified as migrants if the assignment probability for the population in which they were captured was less than the assignment probability for the other population. We determined which isotopes contributed significantly to the discriminant function and calculated the percent contribution of each isotope using a Wilks' Lambda test in R (R Core team 2013). In addition, assignment accuracy was assessed by calculating the percentage of correct classifications from simulated isotopic signatures. The discriminant function, trained with our empirical data, was used to classify 2000 (1000 from each metapopulation) simulated isotopic signatures matching those observed in the Bay Area and Foothills.

Linear discriminant function analysis assumes covariances are equal among groups and that data are sampled from a multivariate normal distribution. We tested for equal covariances among groups using a Box's M test in the R package biotools (Anderson 2015) and found significant differences among isotopes (Box's M $\chi^2_{0.05, 6} = 89.95$, $p < 0.001$). However, Box's M tests are highly sensitive to deviations from normality so the significant differences may have been caused by slight deviations from normality in the distributions of our isotope data (Appendix 3). It is not possible to test for multivariate normality, but if data are normally distributed within groups, it is reasonable to assume the data exhibit multivariate normality. Although the distributions of our isotope data deviated slightly from normality, linear discriminant function analysis is probably robust to minor violations of the assumptions of equal covariances and multivariate normality (Williams 1983).

Environmental sampling and analysis

After preliminary data analysis, we observed highly enriched sulfur isotope values from four rails captured in the Foothills at Bidwell Park in Chico, CA. Although the sulfur isotope values of these individuals were similar to values observed from rails in the Bay Area, the carbon and nitrogen isotope values from these individuals were similar to five other rails captured at Bidwell Park, making it unlikely that these four birds were migrants from the Bay Area. To examine small-scale spatial variation of stable isotope values within wetlands and to determine the cause of the highly enriched sulfur values we observed from rails in Bidwell Park, we collected topsoil (~ 6 cm) from 29 randomly selected locations within home ranges (Hall, unpublished data) of black rails in the Foothills. Wetland soils were collected with a trowel and transported in plastic bags to the lab, where they were dried at 60ºC in a drying oven for a

minimum of six weeks. After drying, soils were homogenized with a mortar and pestle, and approximately 30 to 50 g of each sample was weighed and packaged in a tin capsule for isotopic analysis. Simultaneous analysis of carbon, nitrogen, and sulfur isotopes was performed using the same methodology as described for feathers. Soils that were highly enriched in ^{34}S were reanalyzed using approximately five grams of sample. Differences in delta carbon, nitrogen, and sulfur isotope values between soils collected at Bidwell Park and soils collected at other wetlands in the Foothills were assessed using generalized linear models in R (R Core team 2013). In addition, delta sulfur isotope values from soils and rail feathers collected at Bidwell Park were mapped to examine spatial variability in the sulfur isoscape.

Results

Genotypes clustered into two groups: rails captured in the Bay Area were assigned to one group, and rails captured in the Foothills were assigned to a second group. Although we detected distinct genetic populations in STRUCTURE, gene flow between the Bay Area and Foothills was moderate (F_{ST} = 0.018), and recent migration rate estimates indicated a low level of gene flow between the metapopulations with greater migration from the Foothills to the Bay Area (m = 0.075) than from the Bay Area to the Foothills (m = 0.016). Genetic population assignments identified two individuals as first generation migrants: one rail from the Foothills was assigned to the Bay Area ($p < 0.004$), and one rail from the Bay Area was assigned to the Foothills ($p <$ 0.001; Fig. 2). The migrant rail captured in the Foothills dispersed a minimum of 115 km from the nearest known breeding location in the Bay Area, and the migrant rail captured in the Bay Area dispersed a minimum of 160 km from the nearest known breeding location in the Foothills.

Feathers of rails from the Bay Area had significantly ($p < 0.001$) enriched $\delta^{13}C$, $\delta^{15}N$, and $\delta^{34}S$ values compared with feathers of rails from the Foothills (Tables 1 & 2; Appendix 4). Delta values did not differ significantly between male and female birds for any of the isotopes analyzed, but $\delta^{13}C$ values were significantly different (p = 0.011) between the AHY and HY age classes (Table 2; Appendix 4). Therefore, only AHY (n = 204) rails were classified in our discriminant function analysis.

Stable carbon, nitrogen, and sulfur isotopes all contributed significantly to the linear discriminant function ($p < 0.001$), with sulfur having the greatest percent contribution (66%), followed by nitrogen (50%), and carbon (31%; Table 1). Linear discriminant function analysis assigned 197 of 204 (96.6%) AHY rails to their capture population (Fig. 3a, b). Of the seven rails assigned to a population that differed from their capture population, four were captured in the Bay Area but were assigned to the Foothills and three were captured in the Foothills but assigned to the Bay Area. Three rails captured in the eastern fringe of the Bay Area metapopulation at wetlands in Lodi (n = 1) and Oakley (n = 2) had isotope values similar to rails from freshwater wetlands in the Foothills. Another rail captured in Martinez in the Bay Area had $\delta^{13}C$ and $\delta^{15}N$ values similar to the Bay Area but was assigned to the Foothills because of its depleted $\delta^{34}S$ value. In addition, three rails from Bidwell Park in the Foothills had enriched $\delta^{34}S$ values similar to rails from marine wetlands in the Bay Area. Assignments of simulated isotopic signatures using the discriminant function were highly accurate for the Bay Area (93.0%) and Foothills (97.8%).

When inferences from genetic and isotopic population assignments were combined, 195 of 204 (95.6%) rails were classified as residents of the populations in which they were captured. We did not detect any recent migrants for which both the genetic and isotopic data assigned an individual to a population that differed from its capture population. We did, however, detect two older migrants that were assigned to their capture population using isotopic data, but were assigned to a population that differed from their capture population using genetic data. In addition, seven rails were classified as residents by genetic population assignments, but were classified as migrants by isotopic population assignments.

Delta values of stable carbon, nitrogen, and sulfur isotopes of wetland soils from the Foothills ranged from -29.18 to -25.00 ‰, 1.97 to 7.17 ‰, and -3.69 to 22.35 ‰, respectively (Fig. 4a). Delta values of stable carbon and nitrogen isotopes were not significantly different (p = 0.103 and p = 0.997, respectively) between soils collected at Bidwell Park and soils collected at other wetland sites in the Foothills. However, delta values of stable sulfur isotopes from soils collected at Bidwell Park were significantly greater ($p < 0.001$) than delta sulfur values from soils collected at other wetland sites in the Foothills (Fig 3a). Within the wetland at Bidwell Park, delta sulfur values were highly variable across small spatial scales (tens of meters; Fig. 4b).

Surprisingly, one band return from our study provided a direct estimate of dispersal distance and confirmed the ability of black rails to disperse between the Foothills and the Bay Area. A black rail banded on June 12, 2012 in southern Yuba County in the Foothills was found dead 61 days later at a wind energy site in Solano County near the San Francisco Bay Delta, more than 128 km from its banding location.

Discussion

Combined inference from genetic and isotopic population assignments

By combining inferences from genetic and population assignments, we successfully tracked long-distance movements of non-migratory rails over smaller spatial scales than many previous studies that combined genetic and isotopic data (Wennerberg et al. 2002, Clegg et al. 2003, Kelly et al. 2005, Rundel et al. 2013). Cook et al. (2007) tracked southern pygmy perch among streams over small spatial scales (tens of meters), but used genetic and isotopic data as independent methods to assess dispersal of perch. In contrast, we combined inferences from genetic and isotopic data to classify sampled individuals as: (1) residents, (2) recent migrants, or (3) older migrants. Although we did not detect any recent migrants, our combined method allowed us to detect two older migrants. Further, we observed seven individuals that were assigned by genetic signatures to the population in which they were captured, but were assigned by isotopic signatures to a population that differed from their capture population. These individuals could be residents that dispersed away from their natal population, foraged in a different population, and then recently returned to their natal population. However, it is more likely these individuals were misassigned by isotopic population assignments due to small-scale, spatial variation in the isoscape.

While genetic population assignments are unaffected by spatial processes in the environment, small-scale, spatial variation in biogeochemical processes in the environment may

cause variation in isotopic signatures measured in organisms. For example, differences in stable carbon isotope ratios in wetland environments could be driven by different photosynthetic pathways or differences between terrestrial and marine plant food sources, whereas differences in stable nitrogen and sulfur isotope ratios may be caused by variation in biogeochemical cycles as well as "source effects" (Mambelli et al. In review, Hobson and Wassenaar 2008). If these processes occur at small spatial scales, then variability among isotopic signatures of individuals captured in close proximity to one another would be high, resulting in an increased number of assignment errors.

Seven of 204 black rails were misclassified by isotopic population assignments using linear discriminant function analysis. Four of the misclassified rails were captured in wetlands located at the eastern edge of the San Francisco Bay Delta that have lower salinity than wetlands of San Pablo Bay and San Francisco Bay. Cloern et al. (2002) observed less enriched delta values of stable carbon and nitrogen isotopes in freshwater marsh plants and sediments collected in the San Francisco Bay Delta compared to marine marsh plants and sediments collected in San Pablo Bay and San Francisco Bay. The remaining three misclassified rails were captured in the Foothills at Bidwell Park in Chico and had highly enriched delta sulfur values that matched the highly enriched delta sulfur values we observed in soils collected from Bidwell Park (Fig. 3a, b). These enriched delta sulfur values may have been caused by the presence of volcanic sediments in the region from the adjacent Lassen Volcanic National Park (Thode 1991). If only isotopic data had been used for population assignments, the seven misclassified rails would have been identified as migrants. But, by combining inferences from genetic and isotopic data we ascertained valuable information about small-scale, spatial variation in the Bay Area and Foothills isoscapes.

Accuracy of genetic and isotopic population assignments

Accuracy of genetic population assignments is affected by the level of genetic differentiation among populations, the quantity and variability of markers used, and the completeness of sampling across all possible source populations (Cornuet et al. 1999, Berry et al. 2004). Errors can be made by either the incorrect identification of a resident as a migrant (type I error) or the failure to identify a migrant individual (type II error). Generally, assignment accuracy is greater for highly differentiated populations ($F_{ST} \geq 0.1$), and increases as the quantity and variability of markers increases, with highly accurate assignments achieved using 10 or more moderately variable ($H \approx 0.6$) loci (Cornuet et al. 1999). Rails from the Bay Area and Foothills metapopulations are only moderately differentiated ($F_{ST} = 0.018$). Given this, we maximized our assignment accuracy by: (1) using 15 polymorphic microsatellite loci (mean $H_o \approx 68$; Appendix 1), and (2) controlling the type I error rate by comparing a *p*-value, generated for each assignment using the simulation procedure of Paetkau et al. (2004), to a strict α-value of 0.01. At this error rate, up to three type I assignment errors could have occurred in our sample of 336 rails. Therefore, there is a small probability that the two migrants observed in our sample are assignment errors rather than true migrants.

To further minimize assignment errors, we collected samples from all known breeding areas of black rails in the region (Evens et al. 1991, Evens and Nur 2002, Richmond et al. 2008). However, due to the secretive nature of this species, a small number of individuals in our sample

may have originated from an unknown, and thus un-sampled, breeding area. Breeding areas of black rails are poorly documented throughout the San Francisco Bay Delta and the coastal region northwest of San Francisco near Bolinas Lagoon and Tomales Bay (Evens et al. 1991, Evens and Nur 2002, Spautz et al. 2002). In fact, one of the two migrants detected by our population assignments was captured at Tomales Bay, the only known breeding area for black rails in the coastal region northwest of San Francisco. If black rails breed in other areas of this region, it is possible that this migrant originated from a closer breeding area rather than a more distant area in the Foothills.

Isotopic population assignments of black rails performed with linear discriminant function analysis were highly accurate for both the Bay Area and Foothills (93.0% and 97.8%, respectively) because the delta values of stable carbon, nitrogen, and sulfur isotopes from feathers of rails captured in marine wetlands of the Bay Area were significantly more enriched than feathers of rails from freshwater wetlands of the Foothills (Tables 1 & 2). Previous studies have reported a high contribution of sulfur isotopes for distinguishing between birds foraging in marine and freshwater environments (Caccamise et al. 2000, Lott et al. 2003), and we obtained similar results with delta sulfur having the greatest contribution (66%) to differentiation in the discriminant function (Table 1).

Connectivity of rail metapopulations via long-distance dispersal

Black rail metapopulations of the Bay Area and Foothills are genetically distinct, but appear to be connected by a low to moderate level of gene flow (F_{ST} = 0.018, Foothills to Bay Area m = 0.075, Bay Area to Foothills m = 0.016) suggesting that dispersal between metapopulations may occur infrequently. When inferences from genetic and isotopic population assignments were combined, no rails were identified as recent migrants and only two rails were identified as older migrants. Of these two older migrants, one dispersed a minimum distance of 160 km from the Foothills to the Bay Area, and the other dispersed a minimum distance of 115 km from the Bay Area to the Foothills. These indirect dispersal distance estimates were similar in magnitude to the direct dispersal distance measurement from a band return (128 km).

Our estimates of gene flow were very similar to the migration rate estimates (Foothills to Bay Area m = 0.11, Bay Area to Foothills m = 0.02) and F_{ST} value (0.020) of a previous study of black rails in the Foothills and Bay Area by Girard et al. (2010). Although Girard et al. (2010) detected limited gene flow, they detected a high proportion of migrant individuals relative to their sample size (7 of 62 or 11%). In contrast, we detected a much lower proportion of migrants (2 of 336 or 0.6%). This difference likely resulted from differences in sample size and differences in analytical methods. First, the larger sample size of our study compared to that of Girard et al. (2010) is expected to improve the accuracy of allele frequency estimates, which in turn, increases the accuracy of population assignments. Second, unlike Girard et al. (2010), we used the capture location of individuals to inform their assignment probabilities in STRUCTURE. We used an informative prior because it has been shown to improve cluster assignments among populations with low genetic differentiation (Hubisz et al. 2009). Further, because assignment accuracy declines with decreasing genetic differentiation, the low to moderate level of gene flow between the Foothills and Bay Area could affect the accuracy of population assignments. Therefore, we performed our population assignments in GENECLASS2

so that we could calculate *p*-values for each assignment and control our type I error rate using the method of Paetkau et al. (2004). Given the increased accuracy of our assignments black rails probably disperse less frequently between the Foothills and Bay Area than suggested by Girard et al. (2010).

We report the first direct observation of dispersal distance for black rails of 128 km. Although black rails are secretive and rarely observed flying, the eastern subspecies, *L. j. jamaicensis*, is migratory across at least part of its range (Eddleman et al. 1994). Therefore, it is not surprising that black rails are capable of long-distance movements, even in California, where the subspecies is non-migratory. Dispersal distance estimates for other species of rails are similar in magnitude to those observed for black rails in our study. Takekawa et al. (2014) reported a Ridgway's rail (*Rallus obsoletus*) in the Bay Area that dispersed 45 km in 43 days. In addition, a study by Adams and Quay (1958) of migratory clapper rails (*Rallus crepitans*) indicated rails may travel greater than 480 km in three to four days during migration and suggested that the total distance was probably covered by one or a few longer movements. Clearly, like other rails, black rails are capable of dispersing long distances (> 100 km), but whether long-distance dispersal occurs as the result of a single movement or multiple smaller movements requires further investigation.

Acknowledgements

We thank J. Tecklin, S. Mambelli, P. Brooks, W. Yang, M. Gamboa, C. Cleveland, C. Stegall, N. Najar, T. Nguyen, C. Chu, K. Lovett, J. Bruce, M. Goodnow, B. Gartland, E. Hunter, L. Evans, L. Doll, T. Graham, K. Spragens, V. Bui, J. Wood, J. Takekawa, C. Strong, J. McBroom, J. Krause, J. Hitchen, D. Williams, and B. Becker for assistance. The U.S. Fish and Wildlife Service, California Dept. of Fish and Wildlife, National Park Service, California State Parks, and East Bay Regional Parks granted permission to access study sites. Reviews by T. Dawson and the Beissinger Lab greatly improved this manuscript. Financial support was provided by the National Science Foundation DEB-1051342 and CNH 1115069.

Tables

Table 1. Mean (± standard error) delta values of stable carbon ($\delta^{13}C$), nitrogen ($\delta^{15}N$), and sulfur ($\delta^{34}S$) isotopes measured in California black rail (*Laterallus jamaicensis coturniculus*) feathers from the San Francisco Bay Area (Bay Area) and Sierra Foothills (Foothills) used in linear discriminant function analysis. The percent contribution of each isotope to the discrinimant function was calculated using a Wilks' Lambda test. All isotopes contributed significantly to the discriminant function ($p < 0.001$).

	Mean ± SE (‰)		
Stable Isotope	**Bay Area**	**Foothills**	**% Contribution from Wilks' Lambda**
$\delta^{13}C$	-20.93 ± 0.12	-23.15 ± 0.09	31
$\delta^{15}N$	15.14 ± 0.24	9.57 ± 0.12	50
$\delta^{34}S$	16.79 ± 0.20	5.61 ± 0.25	66

Table 2. Generalized linear model results for individual and interaction effects of area (San Francisco Bay Area and Sierra Foothills), age (after-hatch-year and hatch-year), and sex (male and female) on delta values of carbon ($\delta^{13}C$), nitrogen ($\delta^{15}N$), and sulfur ($\delta^{34}S$) isotopes from feathers of 251 California black rails (*Laterallus jamaicensis coturniculus*). Estimates of three-way interaction effects for area, age, and sex were not calculated because only one hatch year, male bird was captured in the San Francisco Bay Area. *P*-values for statistically significant ($\alpha = 0.05$) effects are shown in bold.

	$\delta^{13}C$			$\delta^{15}N$			$\delta^{34}S$		
	Estimate	**SE**	***p***	**Estimate**	**SE**	***p***	**Estimate**	**SE**	***p***
(Intercept)	-20.771	0.347	< **0.001**	15.494	0.594	< **0.001**	17.662	0.847	< **0.001**
Area	-1.991	0.447	< **0.001**	-6.030	0.765	< **0.001**	-11.568	1.090	< **0.001**
Age	-2.227	0.868	**0.011**	-2.161	1.486	0.147	0.468	2.117	0.825
Sex	-0.048	0.406	0.906	-0.211	0.695	0.762	-1.510	0.991	0.129
Area*Age	1.770	1.014	0.082	2.019	1.735	0.246	-0.522	2.472	0.833
Area*Sex	-0.329	0.527	0.533	0.040	0.901	0.965	1.136	1.284	0.377
Age*Sex	-0.270	0.699	0.700	0.037	1.195	0.975	-0.465	1.703	0.785
Area*Age*Sex	NA	NA	NA	NA	NA	NA	NA	NA	NA

Figures

Figure 1. Capture locations of 336 California black rails (*Laterallus jamaicensis coturniculus*) in the San Francisco Bay Area (Bay Area; blue) and Sierra Foothills (Foothills; green) of California.

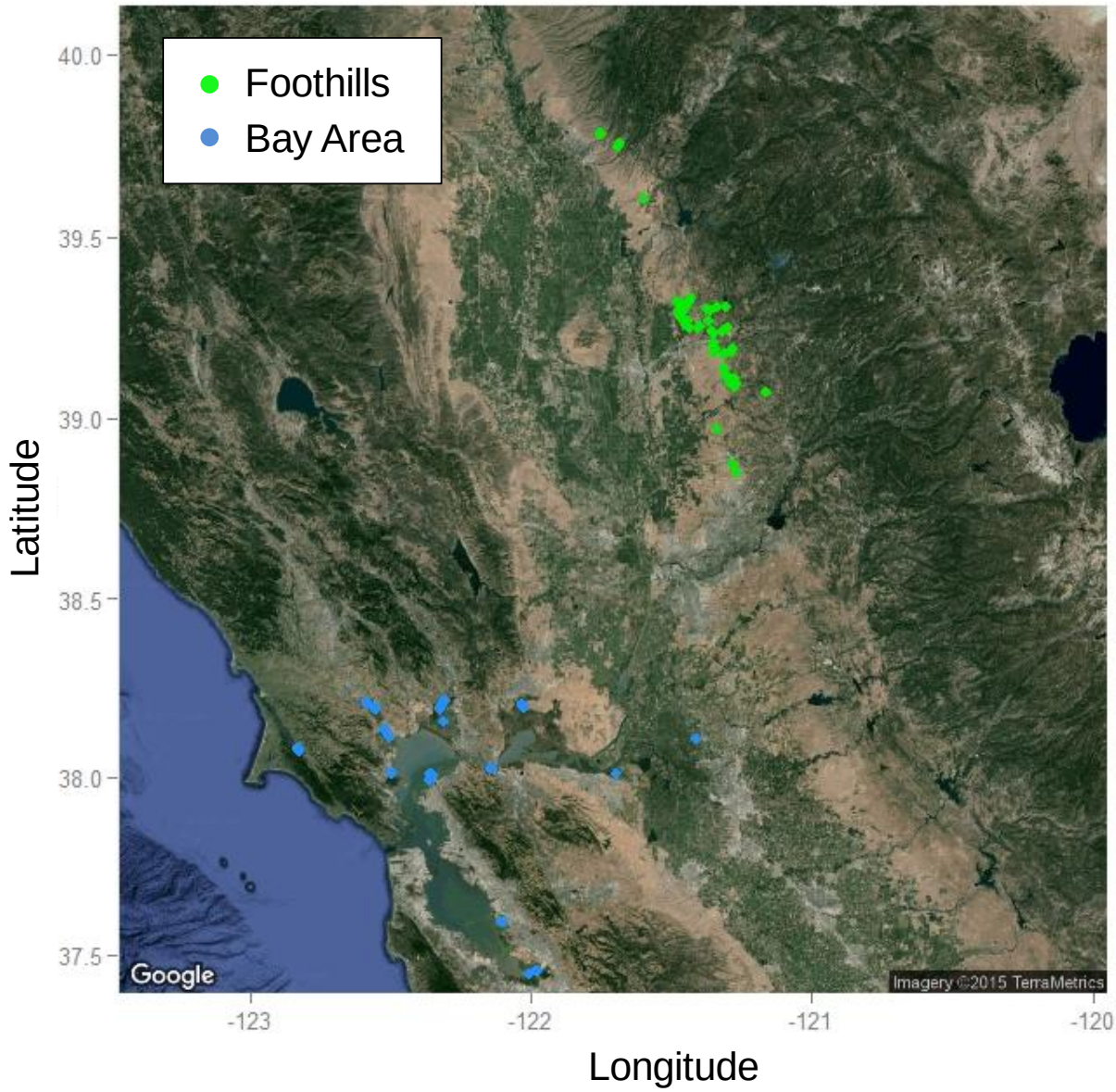


Figure 2. Genetic population assignments of 336 California black rails (*Laterallus jamaicensis coturniculus*) to the San Francisco Bay Area (Bay Area) or Sierra Foothills (Foothills). For each rail, the negative log likelihood value of assignment to the Bay Area is plotted on the x-axis, and the negative log likelihood value of assignment to the Foothills is plotted on the y-axis. Rails captured in the Bay Area (n = 123) are shown in blue and rails captured in the Foothills (n = 213) are shown in green. Most rails (334 of 336) were assigned to the population in which they were captured and were assigned as residents (open circles), but two rails were assigned to a population that differed from their capture location and were classified as migrants (filled squares).

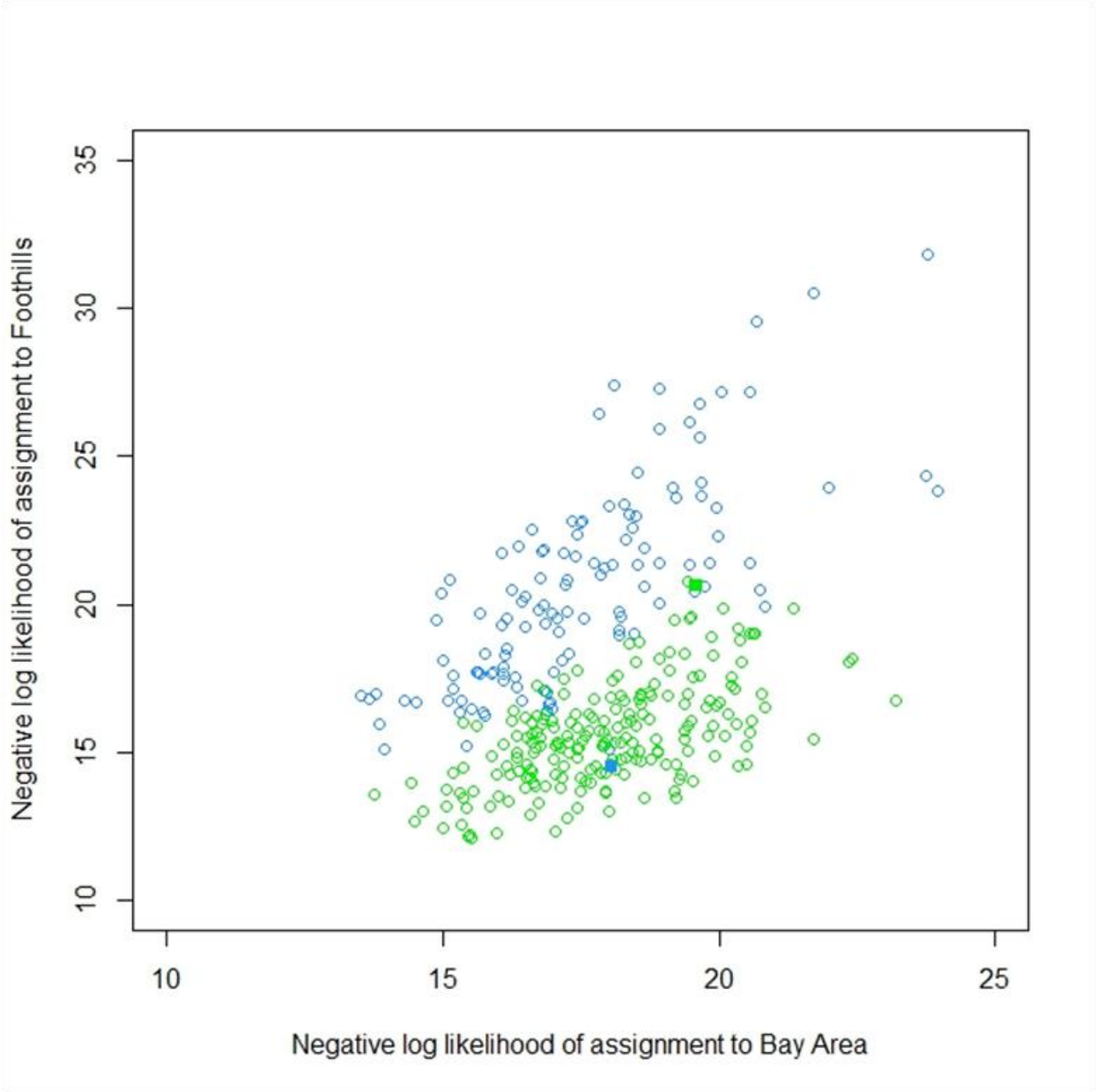


Figure 3. (a) Delta values and (b) discriminant scores (median ± 1 interquartile range) from linear discriminant function analysis of stable carbon ($\delta^{13}C$), nitrogen ($\delta^{15}N$), and sulfur ($\delta^{34}S$) isotopes from feathers of 204 after-hatch-year California black rails (*Laterallus jamaicensis coturniculus*) captured in marine wetlands in the San Francisco Bay Area (BA; blue) or freshwater wetlands in the Sierra Foothills (FH; green) of California. Three rails from the Foothills and four rails from the Bay Area were misclassified by the discriminant function analysis (circled in black). These same individuals were classified as residents by genetic population assignments. Therefore, they were not considered migrants, but yield information about spatial variation of stable isotope values in the environment.

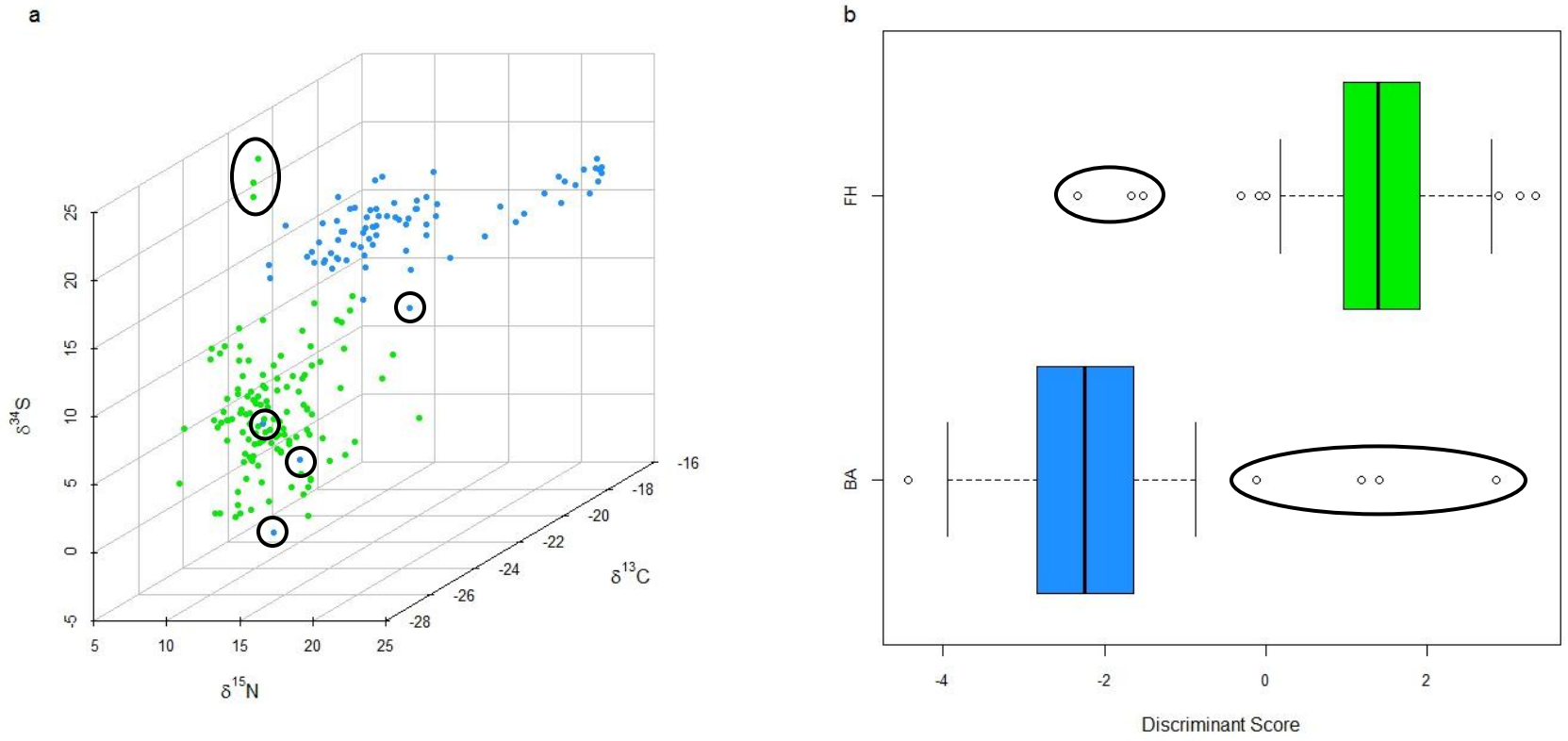
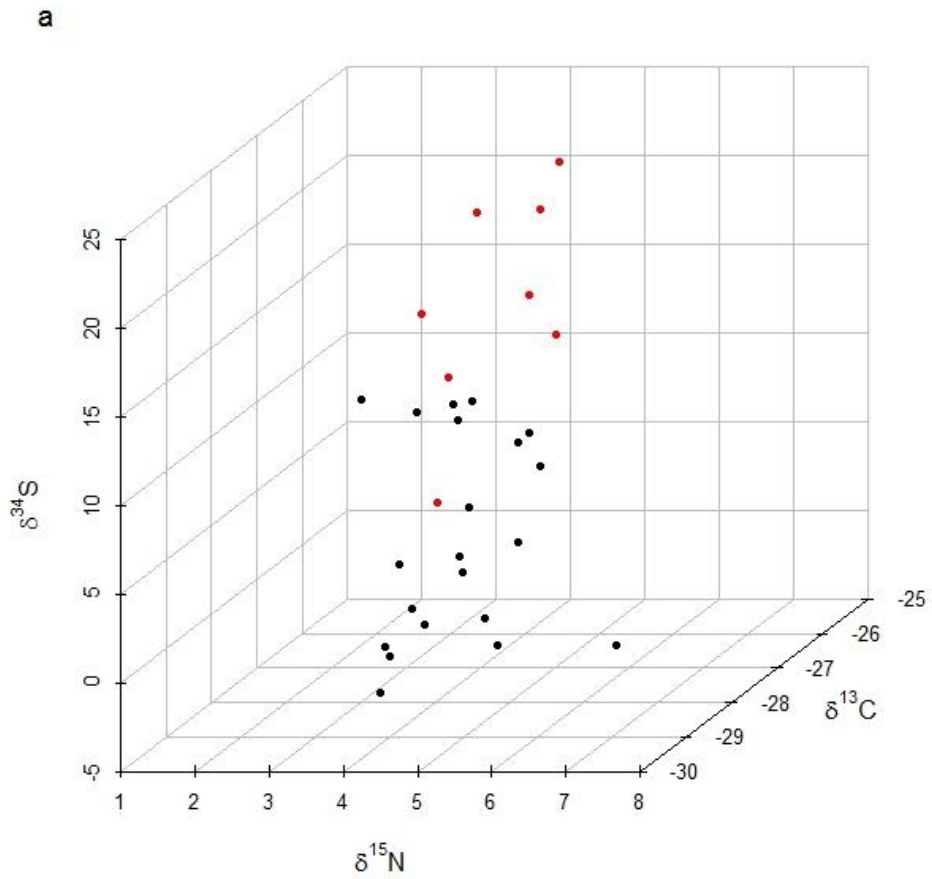
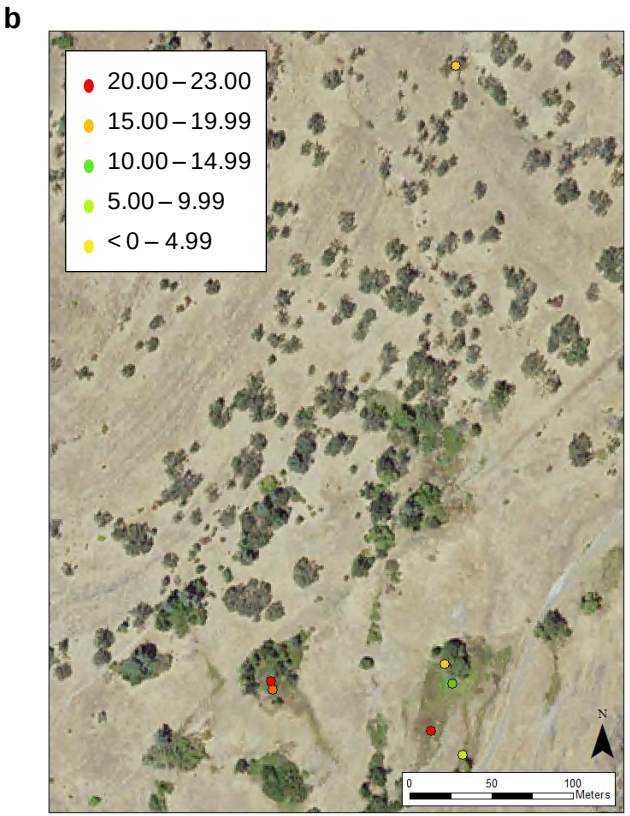


Figure 4. (a) Delta values of stable carbon ($\delta^{13}C$), nitrogen ($\delta^{15}N$), and sulfur ($\delta^{34}S$) isotopes from soils collected in the Sierra Foothills at Bidwell Park in Chico (red) and 28 other wetlands (black) occupied by California black rails (*Laterallus jamaicensis coturniculus*). (b) Delta sulfur isotope values (reported as parts per thousand) measured in Bidwell Park were highly variable at small spatial scales (tens of meters) with some soils having highly enriched delta sulfur isotope values (red and orange) and other soils have lesser enriched delta values (yellow, yellow-green, and green).





Literature Cited

Adamack, A. T., and B. Gruber. 2014. PopGenReport: simplifying basic population genetic analyses in R. Methods in Ecology and Evolution 5:384–387.

Adams, D. A., and T. L. Quay. 1958. Ecology of the clapper rail in southeastern North Carolina. The Journal of Wildlife Management 22:149.

Anderson, R. da S. 2015. Biotools: tools for biometry and applied statistics in agricultural science.

Benjamini, Y., and Y. Hochberg. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society Series B. 57:289–300.

Berry, O., M. D. Tocher, and S. D. Sarre. 2004. Can assignment tests measure dispersal? Molecular Ecology 13:551–561.

Caccamise, D. F., L. M. Reed, P. M. Castelli, S. Wainright, and T. C. Nichols. 2000. Distinguishing migratory and resident Canada geese using stable isotope analysis. The Journal of Wildlife Management 64:1084.

Clegg, S. M., J. F. Kelly, M. Kimura, and T. B. Smith. 2003. Combining genetic markers and stable isotopes to reveal population connectivity and migration patterns in a Neotropical migrant, Wilson's Warbler (Wilsonia pusilla). Molecular Ecology 12:819–830.

Cloern, J. E., E. A. Canuel, and D. Harris. 2002. Stable carbon and nitrogen isotope composition of aquatic and terrestrial plants of the San Francisco Bay estuarine system. Limnology and oceanography 47:713–729.

Cook, B. D., S. E. Bunn, and J. M. Hughes. 2007. Molecular genetic and stable isotope signatures reveal complementary patterns of population connectivity in the regionally vulnerable southern pygmy perch (*Nannoperca australis*). Biological Conservation 138:60–72.

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

Eddleman, W. R., R. E. Flores, and M. Legare. 1994. Black Rail (*Laterallus jamaicensis*). The Birds of North America Online (A. Poole, Ed.). Ithaca: Cornell Lab of Ornithology; http://bna.birds.cornell.edu/bna/species/123.

Evens, J. G., and N. Nur. 2002. California Black Rails in the San Francisco Bay region: spatial and temporal variation in distribution and abundance. Bird Populations 6:1–12.

Evens, J. G., G. W. Page, S. A. Laymon, and R. W. Stallcup. 1991. Distribution, relative abundance and status of the California Black Rail in western North America. The Condor 93:952–966.

Faubet, P., R. S. Waples, and O. E. Gaggiotti. 2007. Evaluating the performance of a multilocus Bayesian method for the estimation of migration rates. Molecular Ecology 16:1149–1166.

Girard, P., J. Y. Takekawa, and S. R. Beissinger. 2010. Uncloaking a cryptic, threatened rail with molecular markers: origins, connectivity and demography of a recently-discovered population. Conservation Genetics 11:2409–2418.

Griffiths, R., M. C. Double, K. Orr, and R. J. Dawson. 1998. A DNA test to sex most birds. Molecular Ecology 7:1071–1075.

Hanski, I. 1999. Metapopulation Ecology. Oxford University Press, New York, New York.

Hobson, K. A. 1999. Tracing origins and migration of wildlife using stable isotopes: a review. Oecologia 120:314–326.
Hobson, K. A., and Wassenaar, editors. 2008. Tracking animal migration with stable isotopes. Academic Press, San Diego.
Hubisz, M. J., D. Falush, M. Stephens, and J. K. Pritchard. 2009. Inferring weak population structure with the assistance of sample group information. Molecular Ecology Resources 9:1322–1332.
Kelly, J. F., K. C. Ruegg, and T. B. Smith. 2005. Combining isotopic and genetic markers to identify breeding origins of migrant birds. Ecological Applications 15:1487–1494.
Koenig, W. D., D. Van Vuren, and P. N. Hooge. 1996. Detectability, philopatry, and the distribution of dispersal distances in vertebrates. Trends in Ecology & Evolution 11:514–517.
Longmire, J. L., A. K. Lewis, N. C. Brown, J. M. Buckingham, L. M. Clark, M. D. Jones, L. J. Meincke, J. Meyne, R. L. Ratliff, F. A. Ray, R. P. Wagner, and R. K. Moyzis. 1988. Isolation and molecular characterization of a highly polymorphic centromeric tandem repeat in the family falconidae. Genomics 2:14–24.
Lott, C. A., T. D. Meehan, and J. A. Heath. 2003. Estimating the latitudinal origins of migratory birds using hydrogen and sulfur stable isotopes in feathers: influence of marine prey base. Oecologia 134:505–510.
Mambelli, S., P. D. Brooks, R. Sutka, S. Hughes, K. Finstad, J. P. Nelson, and T. E. Dawson. In review. High throughput method for simultaneous analysis of N, C and S stable isotopes and concentrations in organics and soils. Rapid Communication in Mass Spectrometry.
Molecular Ecology Resources Primer Development Consortium. 2009. Permanent genetic resources added to Molecular Ecology Resources Database 1 May 2009-31 July 2009. Molecular Ecology Resources 9:1460–1559.
Nathan, R., G. Perry, J. T. Cronin, A. E. Strand, and M. L. Cain. 2003. Methods for estimating long-distance dispersal. Oikos 103:261–273.
Paetkau, D., R. Slade, M. Burden, and A. Estoup. 2004. Genetic assignment methods for the direct, real-time estimation of migration rate: a simulation-based exploration of accuracy and power. Molecular Ecology 13:55–65.
Piry, S., A. Alapetite, J. M. Cornuet, D. Paetkau, L. Baudouin, and A. Estoup. 2004. GENECLASS2: a software for genetic assignment and first-generation migrant detection. Journal of Heredity 95:536.
Pritchard, J. K., M. Stephens, and P. Donnelly. 2000. Inference of population structure using multilocus genotype data. Genetics 155:945.
Pyle, P. 2008. Identification guide to North American birds: part II Anatidae to Alcidae. Slate Creek Press, Point Reyes Station, California.
Rannala, B., and J. L. Mountain. 1997. Detecting immigration by using multilocus genotypes. Proceedings of the National Academy of Sciences 94:9197.
Raymond, M., and F. Rousset. 1995. GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. Journal of Heredity 86:248.
R Core team. 2013. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
Richmond, O. M., J. Tecklin, and S. R. Beissinger. 2008. Distribution of California black rails in the Sierra Nevada foothills. Journal of Field Ornithology 79:381–390.

Robles, H., and C. Ciudad. 2012. Influence of habitat quality, population size, patch size, and connectivity on patch-occupancy dynamics of the middle spotted woodpecker. Conservation Biology 26:284–293.
Rubenstein, D. R., and K. A. Hobson. 2004. From birds to butterflies: animal movement patterns and stable isotopes. Trends in Ecology & Evolution 19:256–263.
Rundel, C. W., M. B. Wunder, A. H. Alvarado, K. C. Ruegg, R. Harrigan, A. Schuh, J. F. Kelly, R. B. Siegel, D. F. DeSante, T. B. Smith, and J. Novembre. 2013. Novel statistical methods for integrating genetic and stable isotope data to infer individual-level migratory connectivity. Molecular Ecology 22:4163–4176.
Sambrook, J., E. F. Fritsch, and T. Maniatis. 1989. Molecular Cloning: A Laboratory Manual. Second edition. Cold Spring Harbor Laboratory Press, New York.
Spautz, H., N. Nur, D. Stralberg, C. J. Ralph, and T. D. Rich. 2002. California Black Rail (*Laterallus jamaicensis coturniculus*) distribution and abundance in relation to habitat and landscape features in the San Francisco Bay Estuary. Pages 20–24 Bird conservation implementation and integration in the Americas: Proceedings of the Third International Partners in Flight Conference.
Sutherland, C. S., D. A. Elston, and X. Lambin. 2014. A demographic, spatially explicit patch occupancy model of metapopulation dynamics and persistence. Ecology 95:3149–3160.
Takekawa, J. Y., M. Casazza, C. Overton, T. Bui, A. G. Vandergast, D. A. Wood, and S. Estrella. 2014. Applied studies on California clapper rail dynamics in support of comprehensive multispecies tidal marsh recovery. Unpublished Data Summary Report, U.S. Geological Survey, Western Ecological Research Center, Vallejo, Dixon, and San Diego.
Thode, H. G. 1991. Sulfur isotopes in nature and the environment: an overview. Pages 1–26 Stable isotopes: natural and anthropogenic sulfur in the environment. John Wiley & Sons.
Tsao, D. C., J. Y. Takekawa, I. Woo, J. L. Yee, and J. G. Evens. 2009. Home range, habitat selection, and movements of California black rails at tidal marshes at San Francisco Bay, California. The Condor 111:599–610.
Wahlberg, N., T. Klemetti, V. Selonen, and I. Hanski. 2002. Metapopulation structure and movements in five species of checkerspot butterflies. Oecologia 130:33–43.
Waples, R. S. 2015. Testing for Hardy-Weinberg proportions: have we lost the plot? Journal of Heredity 106:1–19.
Wennerberg, L., M. Klaassen, and A. Lindstrom. 2002. Geographical variation and population structure in the White-rumped Sandpiper *Calidris fuscicollis* as shown by morphology, mitochondrial DNA and carbon isotope ratios. Oecologia 131:380–390.
West, J. B., G. J. Bowen, T. E. Dawson, and K. P. Tu, editors. 2010. Isoscapes: understanding movement, pattern, and process on Earth through isotope mapping. Springer, New York.
Williams, B. K. 1983. Some observations of the use of discriminant analysis in ecology. Ecology 64:1283.
Wilson, G. A., and B. Rannala. 2003. Bayesian inference of recent migration rates using multilocus genotypes. Genetics 163:1177.
Wright, S. 1931. Evolution in Mendelian populations. Genetics 16:97.

Chapter 4. Gene flow remains high in a secretive wetland bird following a century of habitat loss in the San Francisco Bay Area.

Abstract

Habitat loss and fragmentation can alter the demographic processes of metapopulations and may lead to decreased genetic diversity and increased genetic differentiation. Wetlands of the San Francisco Bay Area, California have been reduced by greater than 80 percent from intense urbanization and agriculture over the past century. We assessed genetic differentiation and gene flow of a secretive and threatened wetland bird, the California black rail (*Laterallus jamaicensis coturniculus*), from 11 wetlands throughout the San Francisco Bay Area. Despite the extreme loss and fragmentation of tidal wetland habitat, population- and individual-level measures of genetic differentiation indicated that gene flow was moderate to high for black rails among remaining wetlands across the Bay Area (F_{ST} range: 0.014 to 0.067). Gene flow was especially high across wetland sites in the North Bay, but was lower between sites in the North Bay and South Bay (F_{ST} = 0.018). Bayesian clustering analysis indicated that black rails from the South Bay represent a unique genetic population that is distinct from rails in the North Bay and Delta. In the future, habitat loss and fragmentation will be exacerbated by rises in sea level due to climate change. Our results indicate that black rails should be capable of adapting to rising sea levels and colonizing protected or newly created wetlands if these habitats are separated by only tens of kilometers.

Keywords: Connectivity, Dispersal, Gene flow, Metapopulation, Habitat loss, Habitat fragmentation, Black rail

Introduction

Habitat loss and fragmentation can alter the demographic processes of metapopulations by reducing population sizes and dispersal among habitat patches (With and Crist 1995, Hanski et al. 1996, Fahrig 2003, Swift and Hannon 2010). Reductions in population size and dispersal may lead to decreased genetic diversity and increased genetic differentiation of populations, and, if severe, could increase local extinction rates through inbreeding depression, accumulation of deleterious mutations, and loss of adaptive potential (Frankham 2005, Keyghobadi 2007).

Despite the paradigm that habitat fragmentation negatively effects genetic diversity and differentiation, 30 to 40 percent of studies fail to document significant genetic effects of habitat fragmentation (Keyghobadi 2007). This may be because the genetic response of populations to habitat fragmentation differs among species and landscapes. Differences in the habitat quality and configuration, the matrix between habitat patches, the edge/area sensitivity of species, the mating system, and the dispersal propensity of species can all driving differences in gene flow among fragmented populations (Gibbs 1998, Callens et al. 2011). It is generally accepted that taxa with greater vagilities, such as birds, are less impacted by habitat fragmentation, but several studies have documented reduced gene flow in birds caused by fragmentation in both forested (Van Houtan et al. 2007, Callens et al. 2011) and urban landscapes (Delaney et al. 2010, Barr et al. 2015).

We measured gene flow and characterized genetic differentiation in the secretive and threatened California black rail (*Laterallus jamaicensis coturniculus*; hereafter black rail) from wetlands in the highly-urbanized San Francisco Bay Area (Bay Area), California. During the last century, intense urban and agricultural pressure in the Bay Area has caused a greater than 80 percent reduction in tidal wetland habitat (Foin et al. 1997, Goals Project 1999). Because this region provides habitat for more than 80 percent of the total black rail population (Evens et al. 1991), negative effects of habitat loss and fragmentation in the Bay Area could increase the extinction risk of the species. Black rails are patchily distributed throughout the Bay Area, with greater numbers of rails inhabiting wetlands of the North Bay (Evens & Nur 2002; Evens et al. 1991). Black rails inhabit densely vegetated wetlands and have small home ranges (mean fixed-kernel home range = 0.59 ha; Tsao et al. 2009). However, black rail occupancy in the Bay Area was significantly positively correlated with marsh size and negatively correlated with distance to the nearest large (> 100 ha) marsh, so they may be sensitive to habitat loss and fragmentation (Spautz et al. 2002). Further, black rails may be poor dispersers with a median dispersal distance of only eight kilometers (Risk et al. 2011). A genetics study by Girard et al. (2010) detected local genetic differentiation and inbreeding depression in a small sample of black rails from five sites in the North Bay, indicating that gene flow among wetlands may be limited.

Here, we assessed genetic differentiation and gene flow of black rails from 11 wetlands throughout the Bay Area by: (1) estimating the number of genetic populations in the Bay Area, (2) calculating population-level, pair-wise measures of F_{ST} among wetlands, (3) comparing individual-level, pair-wise measures of genetic distance to measures of Euclidian distance, and (4) examining spatial genetic patterns using Moran's Eigenvector maps. Because of their short median dispersal distance (8 km; Risk et al. 2011), we predicted that gene flow would be low among the highly-fragmented wetlands of the Bay Area. We also predicted that genetic distances among individual black rails would increase with increasing Euclidean distances following a pattern of isolation by distance.

Methods

Sample collection

Blood samples (n = 123) were collected from black rails captured at 11 wetlands in the Bay Area from 2005 to 2014 (Fig. 1a). Rails were captured during the day with mist nets and at night with spotlights and hand-nets (Tsao et al. 2009, Girard et al. 2010). Blood was collected by venipuncture of a wing vein and stored in Longmire's solution (Longmire et al. 1988). Age (after-hatch-year or AHY; hatch-year or HY) was determined using eye color and plumage characteristics (Pyle 2008), and sex was determined genetically with primers from Griffiths et al. (1998). Birds were banded with U.S. Geological Survey stainless steel bands, and morphometric measurements and molt were recorded.

Genetic data collection

DNA was extracted from blood using either phenol/chloroform (Sambrook et al. 1989) or a DNEasy Spin Column Kit (Qiagen Inc.). Seventeen microsatellite loci developed by Girard et al. (Molecular Ecology Resources Primer Development Consortium 2009) were amplified by

polymerase chain reaction (PCR; Appendix B: Table B1). The experimental conditions were similar for all loci following Girard et al. (2010), with PCR annealing temperatures ranging from 52°C to 59°C (Appendix B: Table B1). Microsatellites were labeled with fluorescent dye (HEX or FAM), run using POP buffer and LIZ 500 size standard on an ABI3730 Genetic Analyzer (Applied Biosystems Inc.), and scored in GeneMapper ver. 5 (Applied Biosystems Inc.).

Data analysis

Microsatellite genotypes were assessed for possible deviations from the expected Hardy-Weinberg equilibrium (HWE) and linkage equilibrium (LE) using exact tests with 20 batches of 10,000 iterations and a dememorization of 10,000 in GENEPOP ver. 4.2 (Raymond and Rousset 1995). We adjusted *p*-values to account for multiple testing using a false discovery rate (FDR; Benjamini and Hochberg 1995) of 0.05. Loci that deviated significantly from expected HWE were removed. We identified significant deviations from LE by splitting the data set into two equally-sized data sets containing randomly assigned individuals. We then compared deviations from LE in the entire data set with the two smaller data sets. Pairs of loci that deviated from LE across all three data sets were considered linked, and one locus from each pair was removed (Waples 2015). No loci deviated from LE. One locus (D105) deviated significantly from HWE and was excluded from genetic analyses (Appendix B: Table B1). Deviations from HWE resulted from a heterozygote deficiency. The possible causes for a heterozygote deficiency in a single locus include: sex-linkage, assortative mating, non-random sampling, or null alleles (Waples 2015). The locus that deviated from HWE was not sex-linked, and it was unlikely that the locus affected rail phenotypes in a manner that would cause non-random mating or sampling. Therefore, the presence of null alleles was the most reasonable explanation for deviation from HWE.

The number of genetic populations (K) was estimated in STRUCTURE (Pritchard et al. 2000) using the admixture model with the degree of admixture (α) inferred from the data, and the distribution of allele frequencies (λ) set to 1. For each value of K, ranging from one to four, we performed 10 runs of 100,000 iterations that were preceded by a burn-in of 1,000,000 iterations. The wetland where each rail was captured was used to inform cluster assignments for individuals (Hubisz et al. 2009). Population-level genetic differentiation was estimated among 11 Bay Area wetlands with F_{ST} (Wright 1931) using GenAlEx v. 6.5 (Peakall and Smouse 2006, 2012). Individual-level genetic distances and Euclidean distances were calculated, and isolation by distance was tested using a Mantel test with 9999 permutations in GenAlEx v. 6.5 (Peakall and Smouse 2006, 2012). Spatial patterns of genetic variation were examined using the R package Memgene to extract the spatial component of genetic distance data with a redundancy analysis of Moran's Eigenvectors and an individual-level, pair-wise genetic distance matrix (R Core team 2013, Galpern et al. 2014).

Results

Two distinct genetic populations were identified using Structure: rails captured in the North Bay and Delta were assigned to one genetic population and rails captured in the South Bay were assigned to another (Fig. 1b). All six rails captured at site 10, located between the North and South Bay regions, were assigned with approximately equal probabilities to each of the two

genetic populations. Pair-wise F_{ST} values between the North Bay, Delta, and South Bay support the results from Structure, with F_{ST} = 0.006 between the North Bay and Delta, F_{ST} = 0.018 between the North Bay and South Bay, and F_{ST} = 0.024 between the South Bay and Delta. Pair-wise F_{ST} values were generally low across the 11 wetlands sampled, with the exception of F_{ST} values from site 9, which had only one sample (Table 1). When F_{ST} values for site 9 were removed, the mean F_{ST} across sites was 0.033 ± 0.005 SE (range: 0.014 to 0.067).

Euclidian distance explained only a small amount of the variation observed among individual-level, pair-wise genetic distances (R^2 = 0.003; Fig. 2). Further, the Mantel test revealed no significant pattern of isolation by distance (p = 0.102). However, when Moran's Eigenvector mapping was used to examine spatial patterns of genetic variation, sites 8 and 9 in the Delta and sites 10 and 11 in the South Bay were genetically distinct from sites 1 to 7 in the North Bay and Delta (Fig. 3). The values extracted using Memgene indicated that relative to rails captured in the North Bay, genetic differentiation of rails captured in the South Bay was of greater magnitude than the differentiation of rails captured in the Delta.

Discussion

Despite the extreme loss and fragmentation of tidal wetland habitat from intense urbanization and agriculture over the past century, population- and individual-level measures of genetic differentiation indicated that gene flow was moderate to high for black rails among remaining wetlands across the Bay Area. Gene flow was especially high across wetland sites in the North Bay, where the greatest numbers of black rails breed (Evens et al. 1991, Evens and Nur 2002). Wetland sites in the North Bay and Delta were separated by an average of 46 ± 8 km (range: 11 to 124 km); this distance was much greater than the median dispersal distance of 8 km reported by Risk et al. (2011). However, the average distance was similar in magnitude to dispersal distances for black rails from a neighboring metapopulation in the Sierra foothills of California (range: 0 to 30 km; Hall, unpublished data). Our results are in contrast to those of Girard et al. (2010) who observed deviations from HWE indicative of low gene flow and inbreeding depression. Differences in the level of gene flow detected between our study and that of Girard et al. (2010) may have resulted from differences in sample sizes and distributions. The results of Girard et al. (2010) were likely an artifact of sampling bias caused by small sample sizes and the partitioning of samples at too-small of spatial scales. To minimize sampling bias, we collected at least eight, but as many as 22, samples at each site (with the exception of sites 8 and 9 in the Delta where too few rails were encountered to collect at least eight samples), and we grouped sample collection locations from contiguous wetland areas to avoid detection of deviations from HWE and false genetic differentiation (Tucker et al. 2014).

Both population- and individual-level measures of genetic differentiation indicated black rails from the South Bay represent a unique genetic population that is distinct from rails in the North Bay and Delta. Previous studies detected few to no black rails in the South Bay and suggested that the species no longer bred in this region (Manolis 1978, Evens et al. 1991). However, we captured black rails from the South Bay in a small patch of historic tidal marsh where the birds appeared to be year-round residents and held territories in pairs during the breeding season (Hall, unpublished data). The genetic differentiation we observed between rails in the North Bay and South Bay confirms that black rails breed in the South Bay, but whether the

population was historically present yet undetected or whether it represents a recent colonization is unknown.

Somewhat surprisingly, the F_{ST} value between the North Bay and South Bay (F_{ST} = 0.018) was equal to the F_{ST} value between the Bay Area and a neighboring metapopulation of black rails from the Sierra foothills (F_{ST} = 0.018; Hall, unpublished data), indicating that gene flow between the North Bay and South Bay occurs infrequently. Wetland sites in the North Bay and South Bay were separated by an average of 72 ± 17 km (range: 50 to 99 km) compared to the greater than 100 km distance between wetlands in the Bay Area and the Sierra foothills. Our results indicate black rail dispersal and subsequent gene flow occurs frequently at smaller spatial scales (tens of kilometers) in the North Bay, but may be limited at larger spatial scales exceeding 50 km. Reed (2004) suggested dispersal among subpopulations partially alleviates the effect of fragmentation on population extinction rates. In black rails, dispersal and gene flow among wetlands in the North Bay may help to increase viability of the Bay Area metapopulation, but the limited gene flow between the North Bay and South Bay detected in our study indicates that these two regions may support demographically independent populations that should be managed separately. Whether populations in the North Bay and South Bay have divergent mitochondrial DNA haptlotypes and represent evolutionarily significant units remains a question for future research.

Black rails from the eastern most sites in the Delta (sites 8 and 9) may also be genetically differentiated from the North Bay and South Bay. Wetlands of the Delta and South Bay had the greatest F_{ST} value (F_{ST} = 0.024) observed in our study. Gene flow may be limited between these wetlands because they are separated by long distances, forcing rails to disperse either 50 to 70 km across a highly-urbanized landscape or 140 to 160 km along the coastline of San Francisco Bay. However, these results should be interpreted with caution because samples sizes from sites 8 and 9 were small (n = 4 at site 8 and n = 1 at site 9).

The patterns of genetic differentiation observed for black rails were similar to those described for Ridgway's rails (*Rallus obsoletus*) in the Bay Area. Takekawa et al. (2014) reported genetic differentiation of Ridgway's rails among wetland sites in the North Bay, East Bay, and South Bay, but, in contrast to black rails, Ridgway's rails exhibited a pattern of isolation by distance. Although black and Ridgway's rails occur sympatrically in some Bay Area wetlands, they likely utilize different regions of wetlands because of differences in their size and their affinities for different salinities. Ridgway's rails tend to occur in wetland areas with higher salinities, whereas black rails can be found in both freshwater drainages and saline wetlands near the Bay's edge. In addition, larger Ridgway's rails are capable of inhabiting lower elevation marshes with greater tidal inundation but, they require taller vegetation to provide cover from predation. Thus, differences in habitat selection between these species may drive differences in their distribution and resulting patterns of gene flow in the Bay Area.

In the future, habitat loss and fragmentation from urbanization and agriculture will be exacerbated by rises in sea level due to climate change. A sea level rise of 13 to 76 cm is projected to impact San Francisco Bay wetlands by 2099 (Cayan et al. 2008). Increased tidal inundation resulting from sea level rise will likely further reduce and fragment wetlands. While wetlands in the North Bay have higher elevations and may be less impacted by sea level rise,

lower elevation wetlands in the South Bay, and the unique genetic population of black rails that inhabits these wetlands, may be particularly vulnerable to sea level rise. Our results indicate that black rails should be capable of adapting to rising sea levels and colonizing protected or newly created wetlands if these habitats are separated by only tens of kilometers. However, despite their ability to disperse and adapt to changes in wetland distribution, sea level rise may indirectly increase mortality of black rails from predation. As wetland habitats move up in elevation with rising sea levels, black rails may be pushed closer to the urban-wetland interface where there are greater numbers of predators. High predation rates have been reported for Ridgway's rails in the Bay Area (Foin et al. 1997, Overton et al. 2014), and black rails likely suffer similarly high rates of predation.

Many local agencies are currently designing and implementing multi-million dollar management and restoration plans to mitigate the impacts of sea level rise on Bay Area wetlands, but the ability to make robust predictions about species specific responses to changes in wetland size and distribution on the landscape is limited by the paucity of information on dispersal and gene flow for many wetland species. The results of this study, in combination with data on dispersal and gene flow for other Bay Area wetland species, will inform agencies about the scale at which wetland management and restoration activities should take place to reduce the negative impacts of sea level rise and maintain viable populations.

Acknowledgements

We thank J. Tecklin, M. Gamboa, C. Stegall, N. Najar, T. Nguyen, C. Chu, K. Lovett, L. Evans, L. Doll, T. Graham, K. Spragens, V. Bui, J. Wood, J. Takekawa, C. Strong, J. McBroom, J. Krause, J. Hitchen, D. Williams, and B. Becker for assistance. The U.S. Fish and Wildlife Service, California Dept. of Fish and Wildlife, National Park Service, California State Parks, and East Bay Regional Parks granted permission to access study sites. Financial support was provided by the National Science Foundation DEB-1051342 and CNH 1115069.

Tables

Table 1. Sample sizes (n) and pair-wise F_{ST} values for California black rails (*Laterallus jamaicensis coturniculus*) from 11 wetland sites in the San Francisco Bay Area, California.

n											
10	**1**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
22	**2**	0.029									
12	**3**	0.035	0.020								
16	**4**	0.025	0.016	0.023							
15	**5**	0.021	0.014	0.019	0.016						
13	**6**	0.019	0.025	0.025	0.022	0.020					
8	**7**	0.036	0.021	0.029	0.023	0.021	0.034				
4	**8**	0.047	0.035	0.046	0.044	0.035	0.044	0.040			
1	**9**	0.208	0.184	0.185	0.189	0.181	0.205	0.202	0.218		
6	**10**	0.050	0.043	0.049	0.047	0.046	0.044	0.056	0.067	0.278	
16	**11**	0.031	0.029	0.030	0.026	0.023	0.029	0.043	0.055	0.220	0.035

Figures

Figure 1. (a) Collection locations (red and yellow points) of 123 genetic samples from California black rails (*Laterallus jamaicensis coturniculus*) in the San Francisco Bay Area, California. Collection locations were grouped into 11 wetland sites (numbered circles), and the sample size for each site is given. The color (red or yellow) of each collection location represents the genetic population with the greatest assignment probabilities for individuals captured at that location according to (b). (b) Population assignment probabilities generated by Structure for California black rails sampled from 11 sites shown in (a) distributed throughout the North Bay, Delta, and South Bay regions of the San Francisco Bay Area. Rails were assigned to two distinct genetic populations, shown in red and yellow.

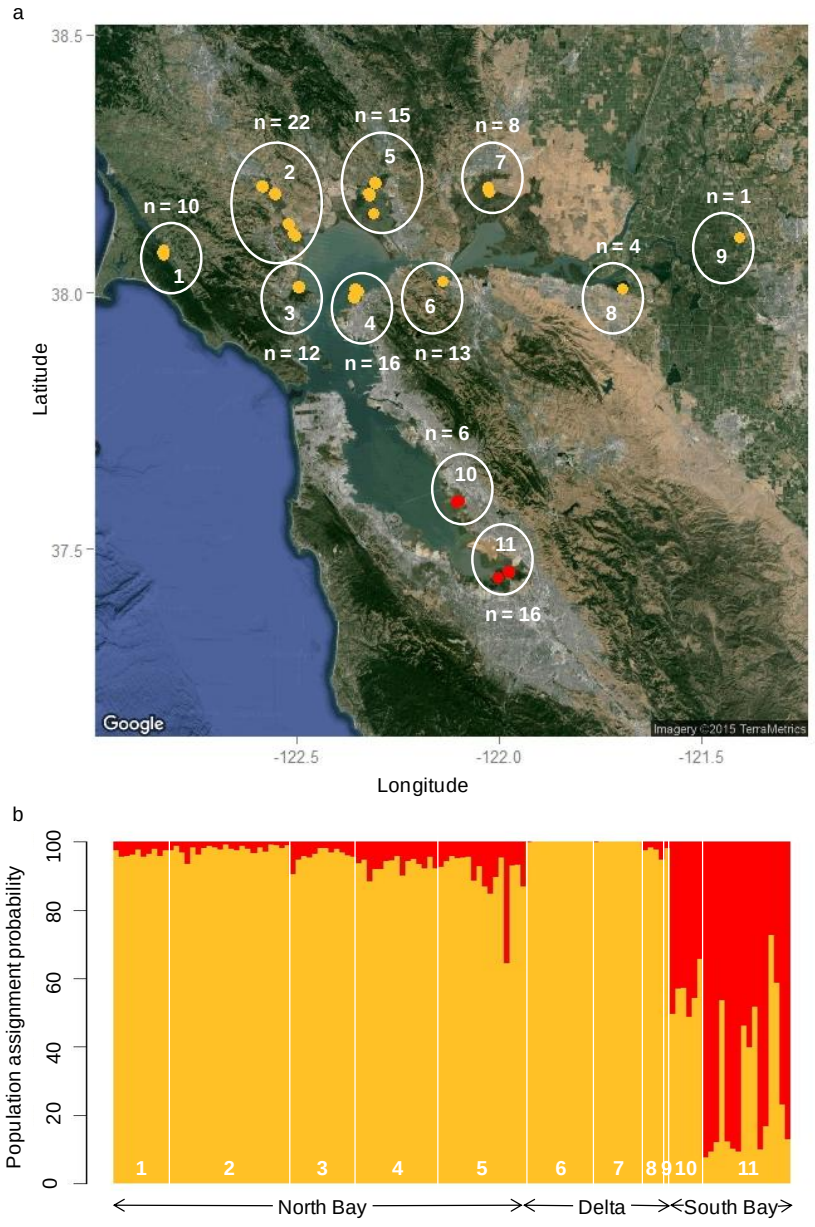


Figure 2. The relationship between individual-level, pair-wise genetic distance and pair-wise Euclidian (geographic) distance (km) between capture locations in the San Francisco Bay Area for 123 California black rails (*Laterallus jamaicensis coturniculus*). The equation and coefficient of determination (R^2) for the linear regression are given.

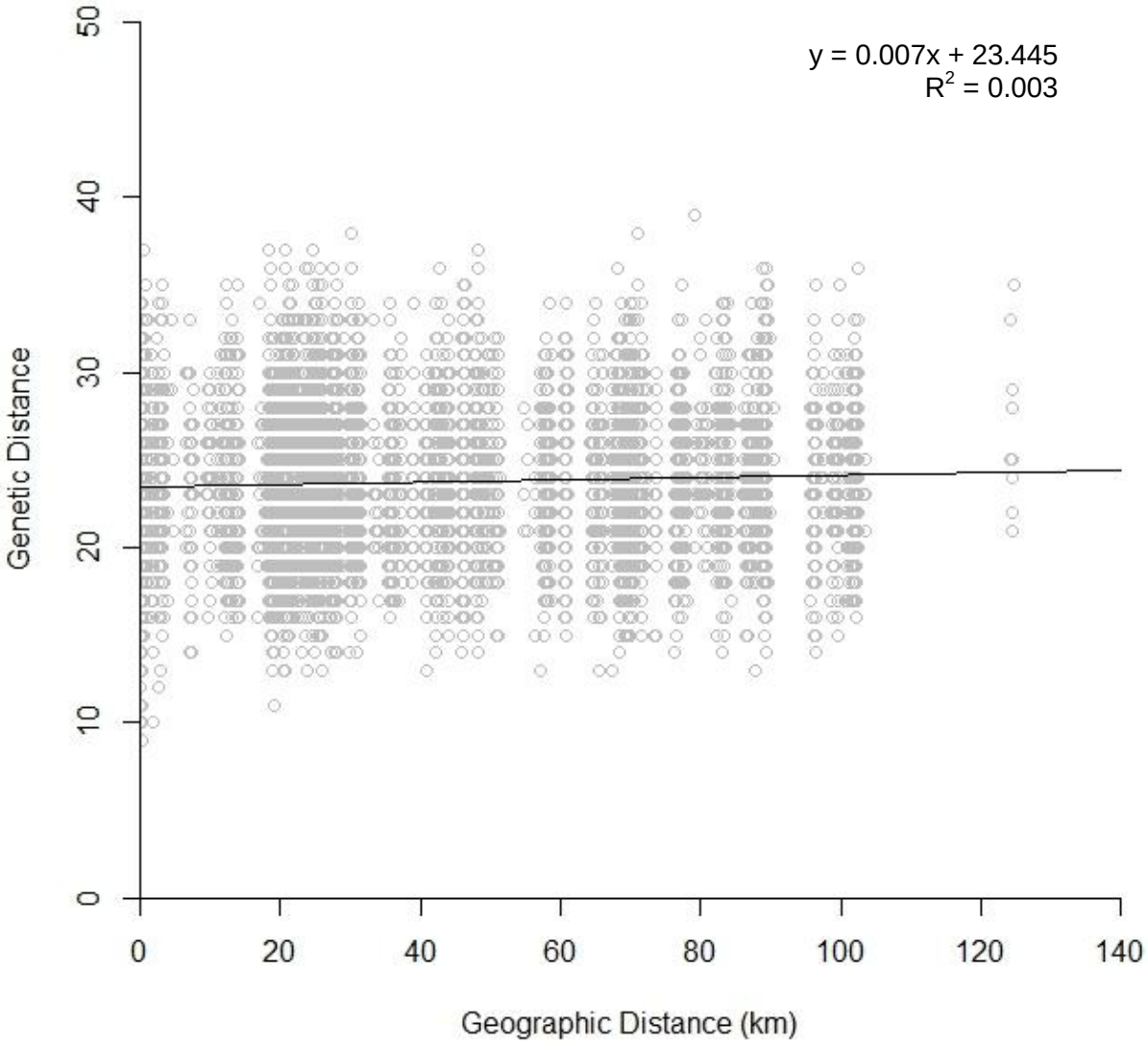
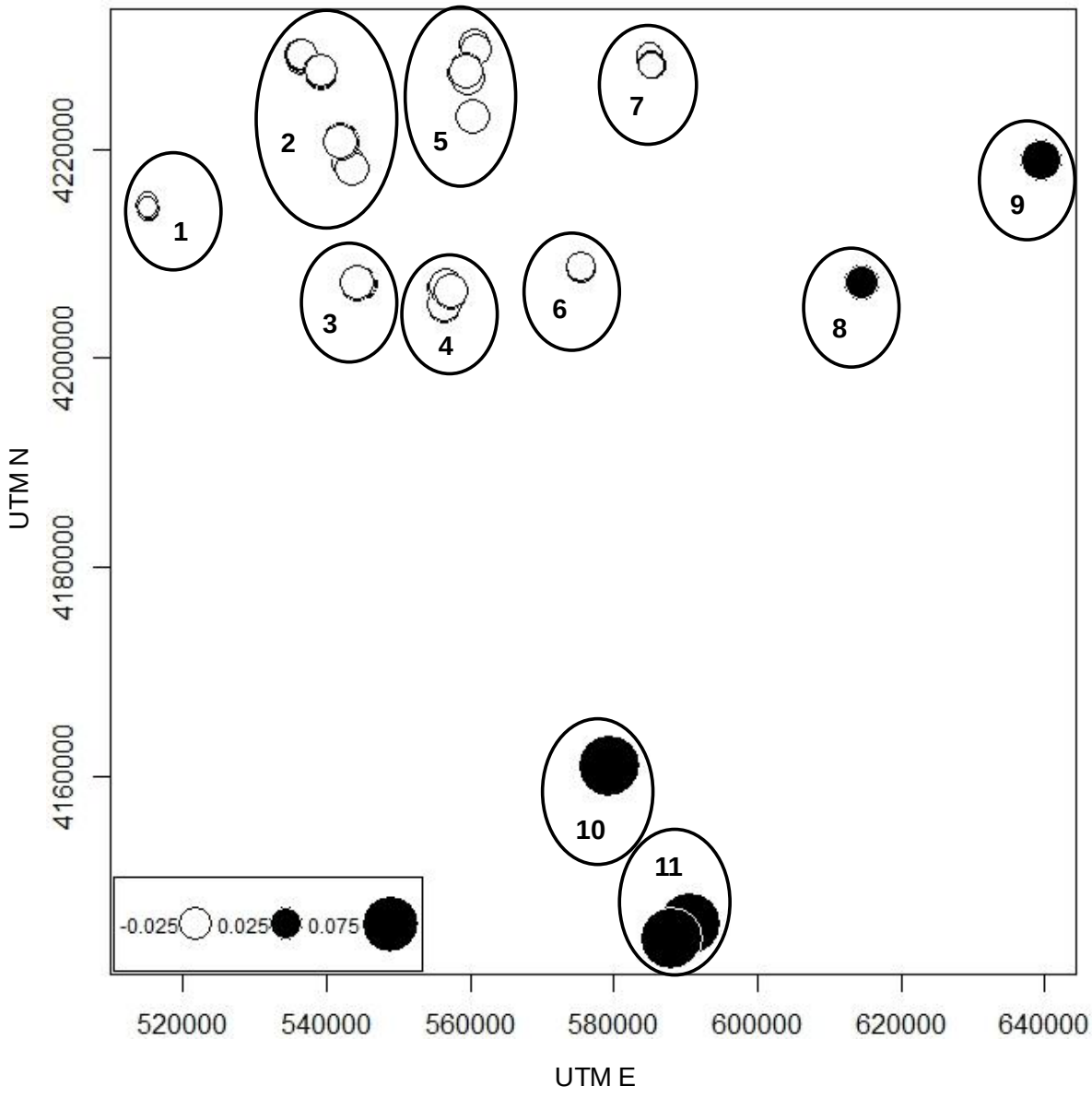


Figure 3. Spatial genetic variation extracted using a redundancy analysis of Moran's Eigenvectors and the individual-level, pair-wise genetic distance matrix for 123 California black rails (*Laterallus jamaicensis coturniculus*) from the San Francisco Bay Area. The size and color of each point represents the magnitude of differences in spatial genetic variation among sample collection locations, such that black points are distinct from white points, and larger points are more distinct than smaller points. Sample collection locations were grouped into 11 wetland sites (numbered circles).



Literature Cited

Barr, K. R., B. E. Kus, K. L. Preston, S. Howell, E. Perkins, and A. G. Vandergast. 2015. Habitat fragmentation in coastal southern California disrupts genetic connectivity in the cactus wren (*Campylorhynchus brunneicapillus*). Molecular Ecology 24:2349–2363.

Benjamini, Y., and Y. Hochberg. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society Series B. 57:289–300.

Callens, T., P. Galbusera, E. Matthysen, E. Y. Durand, M. Githiru, J. R. Huyghe, and L. Lens. 2011. Genetic signature of population fragmentation varies with mobility in seven bird species of a fragmented Kenyan cloud forest. Molecular Ecology 20:1829–1844.

Cayan, D. R., P. D. Bromirski, K. Hayhoe, M. Tyree, M. D. Dettinger, and R. E. Flick. 2008. Climate change projections of sea level extremes along the California coast. Climatic Change 87:57–73.

Delaney, K. S., S. P. D. Riley, and R. N. Fisher. 2010. A rapid, strong, and convergent genetic response to urban habitat fragmentation in four divergent and widespread vertebrates. PLoS ONE 5:e12767.

Evens, J. G., and N. Nur. 2002. California Black Rails in the San Francisco Bay region: spatial and temporal variation in distribution and abundance. Bird Populations 6:1–12.

Evens, J. G., G. W. Page, S. A. Laymon, and R. W. Stallcup. 1991. Distribution, relative abundance and status of the California Black Rail in western North America. The Condor 93:952–966.

Fahrig, L. 2003. Effects of habitat fragmentation on biodiversity. Annual Review of Ecology, Evolution, and Systematics 34:487–515.

Foin, T., E. J. Garcia, R. E. Gill, S. D. Culberson, and J. N. Collins. 1997. Recovery strategies for the California clapper rail (*Rallus longirostris obsoletus*) in the heavily-urbanized San Francisco estuarine ecosystem. Landscape and Urban Planning 38:229–243.

Frankham, R. 2005. Genetics and extinction. Biological Conservation 126:131–140.

Galpern, P., P. Peres-Neto, J. Pulfos, and M. Manseau. 2014. MEMGENE: spatial pattern detection in genetic distance data.

Gibbs, J. P. 1998. Distribution of woodland amphibians along a forest fragmentation gradient. Landscape Ecology 13:263–268.

Girard, P., J. Y. Takekawa, and S. R. Beissinger. 2010. Uncloaking a cryptic, threatened rail with molecular markers: origins, connectivity and demography of a recently-discovered population. Conservation Genetics 11:2409–2418.

Goals Project. 1999. Baylands ecosystem habitat goals. A report of habitat recommendations prepared by the San Francisco Bay Area Wetlands Ecosystem Goals Project. S. F. U.S. Environmental Protection Agency, CA and San Francisco Regional Water Quality Control Board, Oakland, CA.

Griffiths, R., M. C. Double, K. Orr, and R. J. Dawson. 1998. A DNA test to sex most birds. Molecular Ecology 7:1071–1075.

Hanski, I., A. Moilanen, and M. Gyllenberg. 1996. Minimum viable metapopulation size. The American Naturalist 147:527–541.

Hubisz, M. J., D. Falush, M. Stephens, and J. K. Pritchard. 2009. Inferring weak population structure with the assistance of sample group information. Molecular Ecology Resources 9:1322–1332.

Keyghobadi, N. 2007. The genetic implications of habitat fragmentation for animals. Canadian Journal of Zoology 85:1049–1064.

Longmire, J. L., A. K. Lewis, N. C. Brown, J. M. Buckingham, L. M. Clark, M. D. Jones, L. J. Meincke, J. Meyne, R. L. Ratliff, F. A. Ray, R. P. Wagner, and R. K. Moyzis. 1988. Isolation and molecular characterization of a highly polymorphic centromeric tandem repeat in the family falconidae. Genomics 2:14–24.

Manolis, T. 1978. Status of the black rail in central California. Western Birds 9:151–158.

Molecular Ecology Resources Primer Development Consortium. 2009. Permanent genetic resources added to Molecular Ecology Resources Database 1 May 2009-31 July 2009. Molecular Ecology Resources 9:1460–1559.

Overton, C. T., M. L. Casazza, J. Y. Takekawa, D. R. Strong, and M. Holyoak. 2014. Tidal and seasonal effects on survival rates of the endangered California clapper rail: does invasive Spartina facilitate greater survival in a dynamic environment? Biological Invasions 16:1897–1914.

Peakall, R., and P. E. Smouse. 2006. genalex 6: genetic analysis in Excel. Population genetic software for teaching and research. Molecular Ecology Notes 6:288–295.

Peakall, R., and P. E. Smouse. 2012. GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research--an update. Bioinformatics 28:2537–2539.

Pritchard, J. K., M. Stephens, and P. Donnelly. 2000. Inference of population structure using multilocus genotype data. Genetics 155:945.

Pyle, P. 2008. Identification guide to North American birds: part II Anatidae to Alcidae. Slate Creek Press, Point Reyes Station, California.

Raymond, M., and F. Rousset. 1995. GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. Journal of Heredity 86:248.

R Core team. 2013. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Reed, D. H. 2004. Extinction risk in fragmented habitats. Animal Conservation 7:181–191.

Risk, B. B., P. De Valpine, and S. R. Beissinger. 2011. A robust-design formulation of the incidence function model of metapopulation dynamics applied to two species of rails. Ecology 92:462–474.

Sambrook, J., E. F. Fritsch, and T. Maniatis. 1989. Molecular Cloning: A Laboratory Manual. Second edition. Cold Spring Harbor Laboratory Press, New York.

Spautz, H., N. Nur, D. Stralberg, C. J. Ralph, and T. D. Rich. 2002. California Black Rail (*Laterallus jamaicensis coturniculus*) distribution and abundance in relation to habitat and landscape features in the San Francisco Bay Estuary. Pages 20–24 Bird conservation implementation and integration in the Americas: Proceedings of the Third International Partners in Flight Conference.

Swift, T. L., and S. J. Hannon. 2010. Critical thresholds associated with habitat loss: a review of the concepts, evidence, and applications. Biological Reviews 85:35–53.

Takekawa, J. Y., M. Casazza, C. Overton, T. Bui, A. G. Vandergast, D. A. Wood, and S. Estrella. 2014. Applied studies on California clapper rail dynamics in support of comprehensive multispecies tidal marsh recovery. Unpublished Data Summary Report, U.S. Geological Survey, Western Ecological Research Center, Vallejo, Dixon, and San Diego.

Tsao, D. C., J. Y. Takekawa, I. Woo, J. L. Yee, and J. G. Evens. 2009. Home range, habitat selection, and movements of California black rails at tidal marshes at San Francisco Bay, California. The Condor 111:599–610.

Tucker, J. M., M. K. Schwartz, R. L. Truex, S. M. Wisely, and F. W. Allendorf. 2014. Sampling affects the detection of genetic subdivision and conservation implications for fisher in the Sierra Nevada. Conservation Genetics 15:123–136.

Van Houtan, K. S., S. L. Pimm, J. M. Halley, R. O. Bierregaard, and T. E. Lovejoy. 2007. Dispersal of Amazonian birds in continuous and fragmented forest. Ecology Letters 10:219–229.

Waples, R. S. 2015. Testing for Hardy-Weinberg proportions: have we lost the plot? Journal of Heredity 106:1–19.

With, K. A., and T. O. Crist. 1995. Critical Thresholds in Species' Responses to Landscape Structure. Ecology 76:2446.

Wright, S. 1931. Evolution in Mendelian populations. Genetics 16:97.

Appendices

Appendix A: Additional methods and results from Chapter 2.

Table of Contents

Occupancy survey methods

Occupancy surveys for California black rails (*Laterallus jamaicensis coturniculus*; hereafter black rail) and Virginia rails (*Rallus limicola*) were conducted annually during the breeding season (May to August) at 192 wetlands in the Sierra Foothills region of California. Wetlands were surveyed in the morning from thirty mintutes before civil sunrise until 210 minutes after civil sunrise, or in the evening from 180 minutes before civil sunset until 30 minutes after civil sunset. Each wetland was surveyed for both species up to three times per year; we did not revisit a wetland once both species had been detected in a given year. The playback protocol used for each survey followed Richmond et al. (2008, 2010). Surveys were initiated by selecting a playback station within suitable rail. At each station, the playback sequence consisted of 1 minute of silent listening, 30 seconds of black rail "*kic-kic-kerr*" calls, 30 seconds of silent listening, 30 seconds of black rail "*kic-kic-kerr*" calls, 30 seconds of silent listening, 30 seconds of black rail "*grr*" calls, 30 seconds of silent listening, 30 seconds of black rail "*grr*" calls, 2 minutes of silent listening, 30 seconds of Virginia rail "*tic-it*" and "*wep wep wepwepwepppprrr*" (grunt) calls, 30 seconds of silent listening, 30 seconds of Virginia rail "*tic-it*" and "*wep wep wepwepwepppprrr*" (grunt) calls, and 1 minute of silent listening. Calls were played on portable mp3 players (Sony model TCS-580V) and broadcast with small speakers (Radio Shack mini-amplifier-speaker Cat. No. 277-1008C) at 81 to 85 dB measured at 1 m from the speaker. When a rail was detected, calls for that species were discontinued at that wetland, but playback surveys were still conducted for the remaining species of rail. If either or both species were not detected at a playback station, the surveyor moved 50 m within the wetland and repeated the playback sequence. Playback continued at a wetland until both species were detected or the entire wetland was surveyed with playback stations spaced 50 m apart. Surveys were not conducted in rain or wind exceeding 25 km/hr.

Methods and results for tests of Hardy-Weinberg equilibrium and linkage equilibrium

Microsatellite genotypes were assessed for possible deviations from the expected Hardy-Weinberg equilibrium (HWE) and linkage equilibrium (LE) using exact tests with 20 batches of 10,000 iterations and a dememorization of 10,000 in GENEPOP v.4.2 (Raymond and Rousset 1995). We adjusted *p*-values to account for multiple testing using a false discovery rate (FDR; Benjamini and Hochberg 1995) of 0.05. Loci that deviated significantly from expected HWE were removed from the data set. We identified significant deviations from LE by splitting the data set for each species into two equally-sized data sets containing randomly assigned individuals. We then compared deviations from LE in the entire data set with the two smaller data sets. Pairs of loci that deviated from LE across all three data sets were considered linked, and one locus from each pair was removed from the data set (Waples 2015).

In black rails, we detected a marginally significant deviation from HWE, after adjusting *p*-values using an FDR of 0.05, at one microsatellite locus (C2) due to a heterozygote deficit (Appendix A: Table A1). The same locus was also part of a pair of loci that consistently yielded significant deviations from LE so it was removed from our analysis. Three Virginia rail microsatellite loci (C127, C188, and D126) deviated significantly from HWE due to a heterozygote deficit and were excluded from our analysis (Appendix A: Table A1). Possible causes for a heterozygote deficit in a small number of loci are sex-linkage, assortative mating,

non-random sampling, or null alleles (Waples 2015). The loci that deviated from HWE were not sex-linked, and it was unlikely that they affected rail phenotypes in a manner that would cause non-random mating or sampling. Therefore, the presence of null alleles was the most reasonable explanation for deviations from HWE. In addition, one pair of loci deviated from LE consistently across all Virginia rail data sets so one locus (C124) of the pair was excluded from our analysis (Appendix A: Table A1). The remaining 16 black rail microsatellites and 11 Virginia rail microsatellites were used to perform parentage assignments.

Literature cited

Benjamini, Y., and Y. Hochberg. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society Series B. 57:289–300.

Raymond, M., and F. Rousset. 1995. GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. Journal of Heredity 86:248.

Richmond, O. M., J. Tecklin, and S. R. Beissinger. 2008. Distribution of California black rails in the Sierra Nevada foothills. Journal of Field Ornithology 79:381–390.

Richmond, O. M. W., J. E. Hines, and S. R. Beissinger. 2010. Two-species occupancy models: a new parameterization applied to co-occurrence of secretive rails. Ecological Applications 20:2036–2046.

Waples, R. S. 2015. Testing for Hardy-Weinberg proportions: have we lost the plot? Journal of Heredity 106:1–19.

Table A1. Polymerase chain reaction annealing temperatures (T_m) and expected (H_e) and observed (H_o) heterozygosities for microsatellite loci amplified from California black rails (BLRA, *Laterallus jamaicensis coturniculus*) and Virginia rails (VIRA, *Rallus limicola*). Loci were tested for deviations from Hardy-Weinberg Equilibrium (HWE) and linkage equilibrium (LE) using exact tests in GENEPOP v. 4.2. Significance was determined by comparing *p*-values from exact tests to adjusted *p*-values using a false discovery rate of 0.05 to account for multiple testing. One locus from each pair of loci that deviated significantly from LE and all loci that deviated significantly from HWE are shown in bold and were excluded from genetic parentage analysis.

Species	Locus	T_m	H_o	H_e	*p*	Linked	Excluded
BLRA	**C2**	**52**	**0.6901**	**0.7967**	**0.0050**	**A122**	**Y**
	C4	52	0.8028	0.8188	0.0381	-	N
	B10	52	0.7793	0.7467	0.0404	-	N
	D8	54	0.7840	0.7716	0.0791	-	N
	A10	56	0.7606	0.7759	0.1405	-	N
	D105	54	0.8169	0.8243	0.1424	-	N
	A122	52	0.6667	0.6989	0.2677	C2	N
	A121	56	0.4836	0.4501	0.2894	-	N
	B116	56	0.3803	0.3857	0.3723	-	N
	B106	52	0.7700	0.7718	0.4522	-	N
	D112	54	0.6526	0.7205	0.4787	-	N
	C102	56	0.8545	0.8578	0.5071	-	N
	A117	56	0.5540	0.5793	0.5786	-	N
	A116	52	0.1972	0.2011	0.5840	-	N
	D11	52	0.7746	0.7850	0.9283	-	N
	C11	56	0.8310	0.8464	0.9736	-	N
	B121	52	0.6948	0.6808	0.9780	-	N
VIRA	**C188**	**59**	**0.4592**	**0.8705**	**< 0.0001**	**-**	**Y**
	D126	**57**	**0.3980**	**0.8731**	**< 0.0001**	**-**	**Y**
	C127	**55**	**0.6122**	**0.8713**	**< 0.0001**	**-**	**Y**
	B140	55	0.7755	0.8518	0.0156	-	N
	C153	57	0.9082	0.7821	0.0619	-	N
	B161	55	0.6531	0.7714	0.1158	-	N
	A147	54	0.8469	0.8773	0.3101	-	N
	C149	54	0.6837	0.6795	0.3421	-	N
	A104	57	0.7653	0.7806	0.3736	-	N
	C124	**56**	**0.8367**	**0.8314**	**0.4402**	**C186**	**Y**
	C113	56	0.7857	0.8392	0.4406	-	N
	C186	56	0.8367	0.8314	0.4536	C124	N
	A105	57	0.7347	0.7360	0.4560	-	N
	B159	54	0.4286	0.4645	0.5216	-	N
	C118	57	0.8469	0.9013	0.7886	-	N

Table A2. Model selection results for dynamics and detection model sets of single-species, multi-season occupancy models for California black rails (BLRA, *Laterallus jamaicensis coturniculus*) and Virginia rails (VIRA, *Rallus limicola*) from a metapopulation in the Sierra Foothills, CA. Model parameters included occupancy (ψ), colonization (γ), extinction (ε), and detection (ρ). Covariates included the area (ha) of wetland patches in the metapopulation (area), annual effects (year), the Julian date within year of each occupancy survey (Julian day), and whether each survey was performed in the morning or evening (time of day). The number of parameters (K), Akaike's Information Criterion (AIC) scores, the differences in AIC relative to the best model (ΔAIC), the AIC weights indicating relative support for each model (w), and the model likelihoods are given.

Species	Model Set	Model	K	AIC	ΔAIC	*w*	Model Likelihood
BLRA	Dynamics	**ψ(area), γ(year), ε(area, year), ρ(.)**	**16**	**1794.15**	**0.00**	**0.74**	**1.00**
		ψ(area), γ(year), ε(area), ρ(.)	11	1797.61	3.46	0.13	0.18
		ψ(area), γ(.), ε(area, year), ρ(.)	11	1797.99	3.84	0.11	0.15
		ψ(area), γ(.), ε(area), ρ(.)	6	1800.99	6.84	0.02	0.03
		ψ(area), γ(.), ε(.), ρ(.)	5	1825.25	31.10	0.00	0.00
		ψ(.), γ(.), ε(.), ρ(.)	4	1871.34	77.19	0.00	0.00
	Detection	**ψ(area), γ(year), ε(area, year), ρ(year, Julian day)**	**23**	**1784.59**	**0.00**	**0.51**	**1.00**
		ψ(area), γ(year), ε(area, year), ρ(year, Julian day, time of day)	24	1786.18	1.59	0.23	0.45
		ψ(area), γ(year), ε(area, year), ρ(year)	22	1786.76	2.17	0.17	0.34
		ψ(area), γ(year), ε(area, year), ρ(year, time of day)	23	1788.18	3.59	0.08	0.17
		ψ(area), γ(year), ε(area, year), ρ(.)	16	1794.15	9.56	0.00	0.01
		ψ(area), γ(year), ε(area, year), ρ(Julian day)	17	1795.46	10.87	0.00	0.00
		ψ(area), γ(year), ε(area, year), ρ(time of day)	17	1796.04	11.45	0.00	0.00
		ψ(area), γ(year), ε(area, year), ρ(Julian day, time of day)	18	1797.42	12.83	0.00	0.00
		ψ(.), γ(.), ε(.), ρ(.)	4	1871.34	86.75	0.00	0.00
VIRA	Dynamics	**ψ(area), γ(year), ε(area, year), ρ(.)**	**16**	**2263.94**	**0.00**	**0.94**	**1.00**
		ψ(area), γ(.), ε(area, year), ρ(.)	11	2269.71	5.77	0.05	0.06
		ψ(area), γ(year), ε(area), ρ(.)	11	2274.70	10.76	0.00	0.00
		ψ(area), γ(.), ε(area), ρ(.)	6	2283.76	19.82	0.00	0.00

	ψ(area), γ(.), ε(.), ρ(.)	5	2327.93	63.99	0.00	0.00
	ψ(.), γ(.), ε(.), ρ(.)	4	2368.12	104.18	0.00	0.00
Detection	**ψ(area), γ(year), ε(area, year), ρ(year, Julian day, time of day)**	**24**	**2251.98**	**0.00**	**0.41**	**1.00**
	ψ(area), γ(year), ε(area, year), ρ(year, time of day)	23	2252.04	0.06	0.40	0.97
	ψ(area), γ(year), ε(area, year), ρ(Julian day, time of day)	18	2254.20	2.22	0.14	0.33
	ψ(area), γ(year), ε(area, year), ρ(time of day)	17	2257.41	5.43	0.03	0.07
	ψ(area), γ(year), ε(area, year), ρ(year)	22	2258.83	6.85	0.01	0.03
	ψ(area), γ(year), ε(area, year), ρ(year, Julian day)	23	2258.95	6.97	0.01	0.03
	ψ(area), γ(year), ε(area, year), ρ(Julian day)	17	2261.49	9.51	0.00	0.01
	ψ(.), γ(.), ε(.), ρ(.)	4	2368.12	116.14	0.00	0.00

Table A3. Model selection results for buffer radius (BRM) and incidence function (IFM) connectivity measure model sets of single-species, multi-season occupancy models for California black rails (BLRA, *Laterallus jamaicensis coturniculus*) and Virginia rails (VIRA, *Rallus limicola*) from a metapopulation in the Sierra Foothills, CA. Model parameters included occupancy (ψ), colonization (γ), extinction (ε), and detection (ρ). BRM and IFM connectivity measures were calculated at scales ranging from one to 30 km and used as covariates of colonization. Additional covariates included the area (ha) of wetland patches in the metapopulation (area), annual effects (year), the Julian date of each occupancy survey (Julian day), and whether each survey was performed in the morning or evening (time of day). The number of parameters (K), Akaike's Information Criterion (AIC) scores, the differences in AIC relative to the best model (ΔAIC), the AIC weights indicating relative support for each model (w), and the model likelihoods are given.

Species	Model Set	Model	K	AIC	ΔAIC	*w*	Model Likelihood
BLRA	BRM	**ψ(area), γ(year, BRM3km), ε(area, year), ρ(year, Julian day)**	**24**	**1756.77**	**0.00**	**0.85**	**1.00**
		ψ(area), γ(year, BRM6km), ε(area, year), ρ(year, Julian day)	24	1761.59	4.82	0.08	0.09
		ψ(area), γ(year, BRM2km), ε(area, year), ρ(year, Julian day)	24	1763.45	6.68	0.03	0.04
		ψ(area), γ(year, BRM10km), ε(area, year), ρ(year, Julian day)	24	1764.52	7.75	0.02	0.02
		ψ(area), γ(year, BRM5km), ε(area, year), ρ(year, Julian day)	24	1764.75	7.98	0.02	0.02
		ψ(area), γ(year, BRM7km), ε(area, year), ρ(year, Julian day)	24	1766.30	9.53	0.01	0.01
		ψ(area), γ(year, BRM4km), ε(area, year), ρ(year, Julian day)	24	1769.19	12.42	0.00	0.00
		ψ(area), γ(year, BRM15km), ε(area, year), ρ(year, Julian day)	24	1773.25	16.48	0.00	0.00
		ψ(area), γ(year, BRM1km), ε(area, year), ρ(year, Julian day)	24	1774.20	17.43	0.00	0.00
		ψ(area), γ(year, BRM20km), ε(area, year), ρ(year, Julian day)	24	1776.48	19.71	0.00	0.00
		ψ(area), γ(year, BRM25km), ε(area, year), ρ(year, Julian day)	24	1779.00	22.23	0.00	0.00
		ψ(area), γ(year, BRM30km), ε(area, year), ρ(year, Julian day)	24	1781.06	24.29	0.00	0.00
		ψ(area), γ(year), ε(area, year), ρ(year, Julian day)	23	1784.59	27.82	0.00	0.00
		ψ(.), γ(.), ε(.), ρ(.)	4	1871.34	114.57	0.00	0.00
	IFM	**ψ(area), γ(year, IFM3km), ε(year, area), ρ(year, Julian day)**	**24**	**1756.36**	**0.00**	**0.30**	**1.00**
		ψ(area), γ(year, IFM2km), ε(year, area), ρ(year, Julian day)	24	1756.90	0.54	0.23	0.76
		ψ(area), γ(year, IFM4km), ε(year, area), ρ(year, Julian day)	24	1756.95	0.59	0.22	0.74
		ψ(area), γ(year, IFM5km), ε(year, area), ρ(year, Julian day)	24	1758.05	1.69	0.13	0.43
		ψ(area), γ(year, IFM6km), ε(year, area), ρ(year, Julian day)	24	1759.29	2.93	0.07	0.23
		ψ(area), γ(year, IFM7km), ε(year, area), ρ(year, Julian day)	24	1760.49	4.13	0.04	0.13

		ψ(area), γ(year, IFM1km), ε(year, area), ρ(year, Julian day)	24	1762.73	6.37	0.01	0.04
		ψ(area), γ(year, IFM10km), ε(year, area), ρ(year, Julian day)	24	1763.35	6.99	0.01	0.03
		ψ(area), γ(year, IFM15km), ε(year, area), ρ(year, Julian day)	24	1766.19	9.83	0.00	0.01
		ψ(area), γ(year, IFM20km), ε(year, area), ρ(year, Julian day)	24	1767.84	11.48	0.00	0.00
		ψ(area), γ(year, IFM25km), ε(year, area), ρ(year, Julian day)	24	1768.96	12.60	0.00	0.00
		ψ(area), γ(year, IFM30km), ε(year, area), ρ(year, Julian day)	24	1769.79	13.43	0.00	0.00
		ψ(area), γ(year), ε(year, area), ρ(year, Julian day)	23	1784.59	28.23	0.00	0.00
		ψ(.), γ(.), ε(.), ρ(.)	4	1871.34	114.98	0.00	0.00
VIRA	BRM	**ψ(area), γ(year, BRM3km), ε(area, year), ρ(year, Julian day, time of day)**	**25**	**2248.76**	**0.00**	**0.40**	**1.00**
		ψ(area), γ(year, BRM2km), ε(area, year), ρ(year, Julian day, time of day)	25	2250.90	2.14	0.14	0.34
		ψ(area), γ(year), ε(area, year), ρ(year, Julian day, time of day)	24	2251.98	3.22	0.08	0.20
		ψ(area), γ(year, BRM4km), ε(area, year), ρ(year, Julian day, time of day)	25	2252.66	3.90	0.06	0.14
		ψ(area), γ(year, BRM6km), ε(area, year), ρ(year, Julian day, time of day)	25	2253.17	4.41	0.04	0.11
		ψ(area), γ(year, BRM20km), ε(area, year), ρ(year, Julian day, time of day)	25	2253.25	4.49	0.04	0.11
		ψ(area), γ(year, BRM5km), ε(area, year), ρ(year, Julian day, time of day)	25	2253.48	4.72	0.04	0.09
		ψ(area), γ(year, BRM10km), ε(area, year), ρ(year, Julian day, time of day)	25	2253.69	4.93	0.03	0.09
		ψ(area), γ(year, BRM7km), ε(area, year), ρ(year, Julian day, time of day)	25	2253.74	4.98	0.03	0.08
		ψ(area), γ(year, BRM1km), ε(area, year), ρ(year, Julian day, time of day)	25	2253.76	5.00	0.03	0.08
		ψ(area), γ(year, BRM30km), ε(area, year), ρ(year, Julian day, time of day)	25	2253.80	5.04	0.03	0.08
		ψ(area), γ(year, BRM25km), ε(area, year), ρ(year, Julian day, time of day)	25	2253.86	5.10	0.03	0.08
		ψ(area), γ(year, BRM15km), ε(area, year), ρ(year, Julian day, time of day)	25	2253.96	5.20	0.03	0.07
		ψ(.), γ(.), ε(.), ρ(.)	4	2368.12	119.36	0.00	0.00
	IFM	**ψ(area), γ(year), ε(area, year), ρ(year, Julian day, time of day)**	**24**	**2251.98**	**0.00**	**0.16**	**1.00**
		ψ(area), γ(year, IFM2km), ε(area, year), ρ(year, Julian day, time of day)	25	2252.84	0.86	0.10	0.65
		ψ(area), γ(year, IFM1km), ε(area, year), ρ(year, Julian day, time of day)	25	2253.13	1.15	0.09	0.56
		ψ(area), γ(year, IFM3km), ε(area, year), ρ(year, Julian day, time of day)	25	2253.17	1.19	0.09	0.55
		ψ(area), γ(year, IFM4km), ε(area, year), ρ(year, Julian day, time of day)	25	2253.40	1.42	0.08	0.49
		ψ(area), γ(year, IFM5km), ε(area, year), ρ(year, Julian day, time of day)	25	2253.56	1.58	0.07	0.45

ψ(area), γ(year, IFM6km), ε(area, year), ρ(year, Julian day, time of day)	25	2253.66	1.68	0.07	0.43
ψ(area), γ(year, IFM7km), ε(area, year), ρ(year, Julian day, time of day)	25	2253.73	1.75	0.06	0.42
ψ(area), γ(year, IFM10km), ε(area, year), ρ(year, Julian day, time of day)	25	2253.85	1.87	0.06	0.39
ψ(area), γ(year, IFM15km), ε(area, year), ρ(year, Julian day, time of day)	25	2253.92	1.94	0.06	0.38
ψ(area), γ(year, IFM20km), ε(area, year), ρ(year, Julian day, time of day)	25	2253.95	1.97	0.06	0.37
ψ(area), γ(year, IFM25km), ε(area, year), ρ(year, Julian day, time of day)	25	2253.97	1.99	0.06	0.37
ψ(area), γ(year, IFM30km), ε(area, year), ρ(year, Julian day, time of day)	25	2253.97	1.99	0.06	0.37
ψ(.), γ(.), ε(.), ρ(.)	4	2368.12	116.14	0.00	0.00

Table A4. Candidate parent-offspring pairs from genetic parentage assignments of California black rails (BLRA, *Laterallus jamaicensis coturniculus*) and Virginia rails (VIRA, *Rallus limicola*) from a metapopulation in the Sierra Foothills, CA. Parentage assignments were performed in program MLRelate using microsatellite genotypes to calculate pair-wise LOD scores (the natural log of the likelihood of a pair being related as parent and offspring divided by the likelihood that the pair is unrelated). All pairs with positive LOD scores were considered candidate parent-offspring (PO) pairs. A false discovery rate (FDR) was used to identify the number of correctly assigned pairs at different levels of statistical significance (FDR = 1.00, 0.10, and 0.05). Candidate pairs in the table that fall below each FDR were treated as assignment errors for the given FDR. We estimated dispersal using pairs at an FDR of 0.05 and further excluded any pairs that did not represent dispersal events, including juveniles captured in the same patch and same year as their parent and full siblings captured in the same patch. For each pair, spatial distance (km) between capture locations, information about capture location and year (year 1 and 2), age of each bird (age 1 and 2; AHY = after hatch year, HY = hatch year, U = unknown), dispersal ability of each bird (capable of dispersal bird 1 and 2) at the time of capture (determined by molt of flight feathers), and sex of each bird (sex 1 and 2; M = male, F = female, U = unknown) are given. This information was collectively used to make possible inferences about dispersal for each pair.

										Capable of dispersal					
Species	**FDR**	**Pair**	**LOD**	**Distance (km)**	**Same patch**	**Year 1**	**Year 2**	**Age 1**	**Age 2**	**Bird 1**	**Bird 2**	**Sex 1**	**Sex 2**	**Excluded**	**Inference**
BLRA		1	17.47	0.01	Y	2009	2010	HY	HY	N	N	M	M	Y	a
		2	13.63	1.28	N	2009	2010	AHY	HY	Y	N	U	M	N	b
		3	13.11	0.13	Y	2012	2013	AHY	AHY	Y	Y	F	M	N	c
		4	12.38	0.18	Y	2011	2011	AHY	AHY	Y	Y	M	F	N	c
		5	12.20	0.43	Y	2010	2010	AHY	HY	Y	Y	M	F	Y	d
		6	12.10	5.91	N	2012	2012	HY	AHY	Y	Y	M	M	N	e
		7	11.96	0.38	Y	2012	2012	AHY	HY	Y	N	M	F	Y	d
		8	11.68	19.39	N	2012	2013	AHY	U	Y	Y	M	U	N	e
		9	11.38	0.13	Y	2012	2012	AHY	AHY	Y	Y	M	M	N	c
		10	11.30	10.18	N	2009	2009	AHY	AHY	Y	Y	F	M	N	e
		11	11.03	5.81	N	2012	2012	HY	AHY	Y	Y	M	M	N	e
		12	10.91	0.08	Y	2012	2012	HY	AHY	Y	Y	U	M	Y	d
		13	10.83	23.85	N	2007	2008	AHY	AHY	Y	Y	M	M	N	e
		14	10.36	0.08	Y	2012	2013	AHY	AHY	Y	N	M	F	N	c
		15	10.27	0.00	Y	2009	2009	HY	AHY	Y	Y	M	M	Y	d

	16	10.18	0.01	Y	2012	2013	AHY	AHY	Y	Y	F	M	N	c
	17	10.05	0.13	Y	2012	2013	AHY	AHY	Y	Y	M	M	N	c
	18	10.01	25.18	N	2011	2012	AHY	AHY	Y	Y	M	M	N	e
	19	9.87	5.19	N	2012	2012	HY	AHY	Y	Y	M	M	N	e
	20	9.51	2.07	N	2013	2013	AHY	HY	Y	Y	F	M	N	e
	21	9.37	0.24	N	2012	2013	AHY	HY	Y	Y	F	M	N	e
	22	9.28	0.02	Y	2010	2010	AHY	HY	Y	N	M	M	Y	d
	23	8.91	0.01	Y	2012	2012	AHY	AHY	Y	Y	F	M	N	c
	24	8.76	0.08	Y	2012	2012	HY	AHY	Y	Y	F	M	Y	d
	25	8.69	0.18	Y	2010	2010	AHY	AHY	Y	Y	M	F	N	c
	26	8.66	0.12	Y	2009	2012	AHY	AHY	Y	Y	F	M	N	c
	27	8.54	0.08	Y	2011	2011	AHY	AHY	Y	Y	M	U	N	c
	28	8.51	28.07	N	2011	2013	AHY	AHY	Y	Y	M	U	N	e
	29	8.42	0.08	Y	2012	2012	AHY	AHY	Y	Y	U	M	N	c
	30	8.40	0.00	Y	2012	2013	AHY	AHY	Y	Y	M	M	N	c
	31	8.34	0.03	Y	2009	2010	HY	AHY	N	Y	M	M	Y	d
0.05	32	7.64	0.03	Y	2012	2012	AHY	HY	Y	N	M	M	Y	d
	33	7.35	14.70	N	2004	2013	U	AHY	Y	Y	M	U	N	e
	34	6.90	0.00	Y	2010	2010	HY	AHY	N	Y	F	M	Y	d
	35	6.64	58.49	N	2010	2013	AHY	AHY	Y	Y	M	U	N	e
	36	6.62	27.90	N	2009	2013	AHY	AHY	Y	Y	M	F	N	e
	37	6.54	0.11	Y	2012	2013	AHY	HY	Y	N	M	F	Y	d
0.10	38	6.50	1.05	N	2010	2011	AHY	AHY	Y	Y	M	M	N	e
	39	6.20	0.23	N	2010	2010	AHY	HY	Y	N	F	F	N	f
	40	6.11	1.27	N	2009	2010	AHY	AHY	Y	Y	U	M	N	e
	41	6.09	0.41	Y	2012	2012	AHY	HY	Y	N	F	F	Y	d
	42	6.01	11.90	N	2009	2013	HY	HY	Y	N	U	M	N	g
	43	5.90	6.31	N	2012	2013	AHY	AHY	Y	Y	M	M	N	e
	44	5.86	60.83	N	2009	2012	AHY	AHY	Y	Y	M	M	N	e
	45	5.58	40.20	N	2011	2012	AHY	AHY	Y	Y	U	F	N	e

		46	5.56	0.10	Y	2009	2012	AHY	HY	Y	N	F	M	Y	d
		47	5.45	20.62	N	2010	2013	HY	AHY	Y	N	M	F	N	e
		48	5.32	20.79	N	2010	2012	AHY	AHY	Y	Y	M	M	N	e
		49	5.11	80.07	N	2009	2012	HY	AHY	Y	Y	F	M	N	e
		50	4.83	0.02	Y	2012	2013	AHY	HY	Y	Y	M	U	Y	d
		51	4.81	72.33	N	2009	2012	AHY	AHY	Y	Y	M	M	N	e
		52	4.59	0.06	Y	2010	2012	AHY	AHY	Y	Y	F	M	N	c
		53	4.56	8.91	N	2012	2013	AHY	HY	Y	N	F	M	N	b
		54	4.36	2.12	N	2009	2011	AHY	AHY	Y	Y	F	F	N	e
		55	4.16	21.95	N	2011	2012	AHY	AHY	Y	Y	M	M	N	e
		56	4.07	1.31	N	2010	2012	HY	HY	N	N	F	M	N	g
		57	3.95	1.26	N	2009	2013	AHY	AHY	Y	Y	F	M	N	e
		58	3.79	7.88	N	2012	2013	AHY	HY	Y	N	F	M	N	b
		59	3.65	1.21	N	2009	2012	AHY	AHY	Y	Y	F	M	N	e
		60	3.30	8.43	N	2012	2013	AHY	AHY	Y	Y	M	F	N	e
		61	2.72	44.31	N	2009	2011	HY	AHY	Y	Y	U	U	N	e
		62	1.86	1.30	N	2009	2013	AHY	AHY	Y	Y	F	M	N	e
		63	1.65	80.07	N	2009	2012	AHY	AHY	Y	Y	F	M	N	e
		64	1.03	81.38	N	2010	2012	HY	AHY	N	Y	F	M	N	e
	1.00	65	0.56	8.89	N	2009	2009	AHY	AHY	Y	Y	U	M	N	e
VIRA		1	13.14	0.00	Y	2011	2011	HY	AHY	N	Y	M	M	Y	d
		2	10.87	0.04	Y	2012	2012	AHY	HY	Y	Y	U	U	Y	d
		3	10.35	0.06	Y	2011	2012	HY	HY	Y	Y	F	U	Y	a
		4	10.30	0.01	Y	2009	2009	HY	AHY	N	Y	U	F	Y	d
		5	10.28	0.13	Y	2011	2011	HY	HY	Y	Y	F	F	Y	a
		6	9.77	0.01	Y	2009	2009	HY	AHY	Y	Y	U	F	Y	d
		7	9.00	2.12	N	2011	2012	HY	AHY	Y	Y	F	M	N	e
		8	8.27	1.47	N	2012	2013	HY	HY	Y	Y	F	U	N	g
		9	8.06	0.14	Y	2011	2012	HY	HY	Y	N	F	U	Y	a
		10	8.02	0.00	Y	2010	2010	HY	AHY	N	Y	U	U	Y	d

	11	7.54	0.00	Y	2009	2009	HY	HY	Y	N	U	U	Y	a
	12	7.23	13.76	N	2012	2013	HY	AHY	Y	Y	F	U	N	e
	13	7.05	1.31	N	2010	2013	HY	HY	Y	Y	F	U	N	g
0.05	14	6.72	0.05	Y	2012	2012	AHY	HY	Y	N	F	M	Y	d
	15	6.50	0.05	Y	2012	2012	AHY	HY	Y	N	M	M	Y	d
	16	5.90	0.00	Y	2010	2010	HY	AHY	N	Y	U	U	Y	d
	17	5.82	63.08	N	2011	2011	HY	AHY	Y	Y	U	F	N	e
0.10	18	5.77	13.48	N	2012	2012	AHY	HY	Y	Y	F	M	N	e
	19	5.40	1.38	N	2010	2013	AHY	HY	Y	Y	M	U	N	b
	20	5.11	10.83	N	2008	2013	HY	U	Y	Y	U	U	N	e
	21	4.92	0.06	Y	2010	2010	HY	AHY	Y	Y	F	M	Y	d
	22	4.86	23.05	N	2009	2013	HY	AHY	N	Y	U	U	N	e
	23	4.44	1.55	N	2010	2013	HY	AHY	N	Y	U	U	N	e
	24	4.44	0.15	Y	2011	2012	HY	HY	Y	N	F	U	Y	a
1.00	25	4.25	1.79	N	2010	2011	HY	HY	N	Y	U	F	N	g

a Full siblings captured in their natal patch

b Parent-offspring pair where the adult dispersed after capture to a new patch where it hatched an offspring that was captured in a subsequent year

c Parent-offspring pair where the juvenile remained within its natal patch

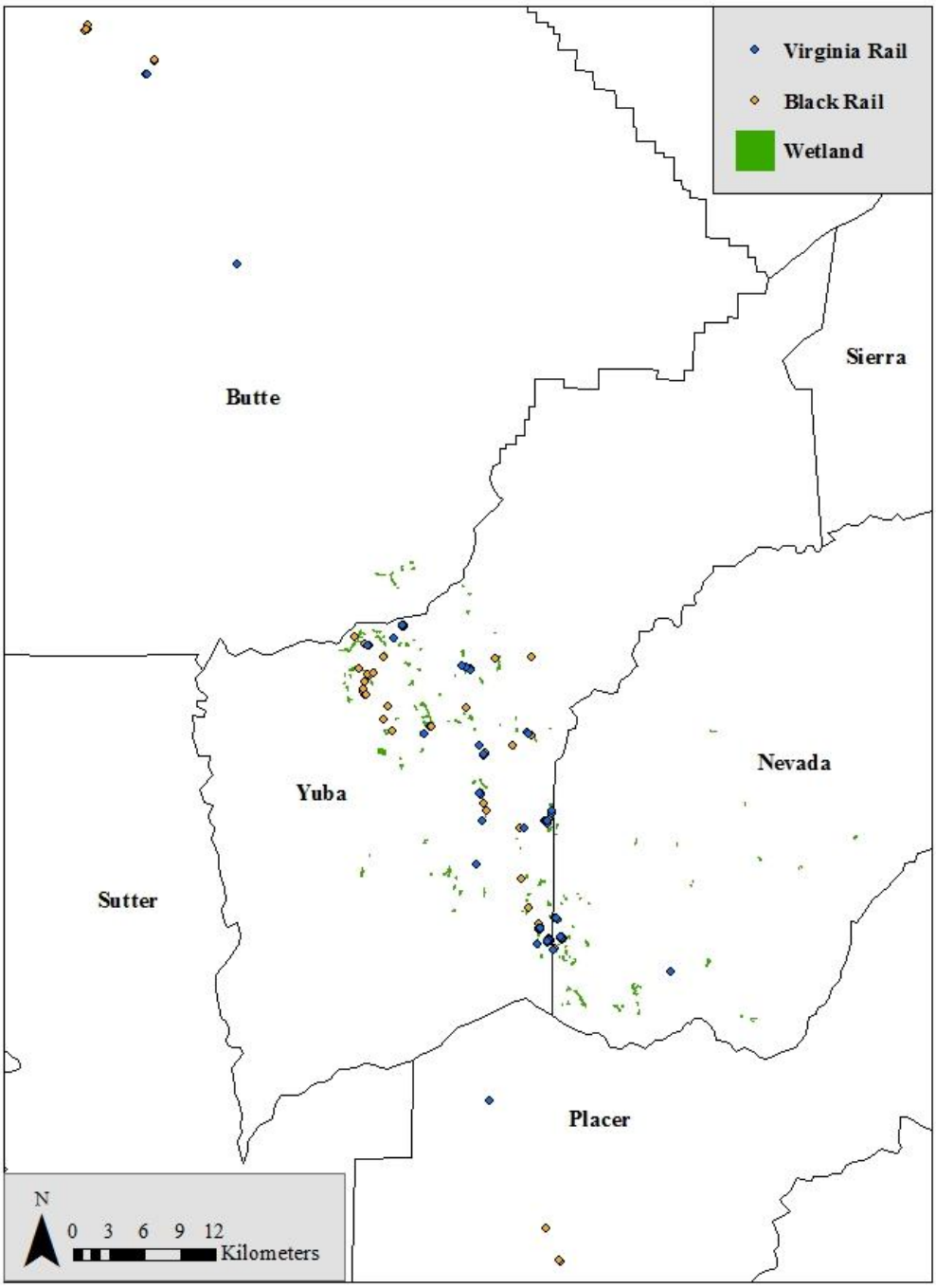
d Parent-offspring pair where the juvenile was captured in its natal patch and had not yet dispersed

e Parent-offspring pair where the juvenile or adult (or both) dispersed from the juvenile's natal patch

f Parent-offspring pair where the juvenile was captured, then the adult dispersed from its patch and was captured after dispersal

g Parent-offspring pair where the juvenile dispersed after capture, then matured and hatched an offspring that was captured in a subsequent year

Figure A1. Collection locations of genetic samples from California black rails (yellow; n = 213; *Laterallus jamaicensis coturniculus*) and Virginia rails (blue; n = 98; *Rallus limicola*) in the Sierra Foothills of California. Wetland patches (n = 192) surveyed for rail occupancy during the 2007 to 2013 breeding seasons are shown in green.



Appendix B: Additional methods and results for Chapter 3.

Table of Contents

Table B1. Polymerase chain reaction annealing temperatures (T_m) and expected (H_e) and observed (H_o) heterozygosities for microsatellite loci amplified from California black rails (*Laterallus jamaicensis coturniculus*). One locus from each pair of loci that deviated significantly from linkage equilibrium (LE) and all loci that deviated significantly from Hardy-Weinberg Equilibrium (HWE) were excluded from genetic population assignments (shown in bold). Deviations from HWE resulted from heterozygote deficiencies. The possible causes for a heterozygote deficit in a small number of loci include: sex-linkage, assortative mating, non-random sampling, or null alleles (Waples 2015). The loci that deviated from HWE were not sex-linked, and it was unlikely that they affected rail phenotypes in a manner that would cause non-random mating or sampling. Therefore, the presence of null alleles was the most reasonable explanation for deviations from HWE.

		Foothills				**Bay Area**				
Locus	**T_m**	**H_o**	**H_e**	***p***	**Linked**	**H_o**	**H_e**	***p***	**Linked**	**Excluded**
C2	**52**	**0.6901**	**0.7967**	**0.0050**	**A122**	**0.7236**	**0.8390**	**0.0976**	**-**	**Y**
C4	52	0.8028	0.8188	0.0381	-	0.7886	0.8035	0.2247	-	N
B10	52	0.7793	0.7467	0.0404	-	0.7724	0.7749	0.4743	-	N
D8	54	0.7840	0.7716	0.0791	-	0.8130	0.8214	0.8559	-	N
A10	56	0.7606	0.7759	0.1405	-	0.8455	0.8359	0.6322	-	N
D105	**54**	**0.8169**	**0.8243**	**0.1424**	**-**	**0.6179**	**0.8546**	**< 0.0001**	**-**	**Y**
A122	52	0.6667	0.6989	0.2677	C2	0.6667	0.7544	0.0186	-	N
A121	56	0.4836	0.4501	0.2894	-	0.5203	0.4681	0.3245	-	N
B116	56	0.3803	0.3857	0.3723	-	0.4959	0.4686	0.2411	-	N
B106	52	0.7700	0.7718	0.4522	-	0.6423	0.7258	0.0232	-	N
D112	54	0.6526	0.7205	0.4787	-	0.5691	0.6296	0.1070	-	N
C102	56	0.8545	0.8578	0.5071	-	0.9350	0.8948	0.8538	-	N
A117	56	0.5540	0.5793	0.5786	-	0.5935	0.5834	0.1865	-	N
A116	52	0.1972	0.2011	0.5840	-	0.3984	0.5062	0.0022	-	N
D11	52	0.7746	0.7850	0.9283	-	0.8374	0.8260	0.1248	-	N
C11	56	0.8310	0.8464	0.9736	-	0.9024	0.8923	0.8622	-	N
B121	52	0.6948	0.6808	0.9780	-	0.6504	0.6704	0.4764	-	N

Table B2. Mean (± standard deviation) delta values of stable carbon ($\delta^{13}C$), nitrogen ($\delta^{15}N$), and sulfur ($\delta^{34}S$) isotopes measured in feather standards from common murre (*Uria aalge*; COMU), pacific loon (*Gavia pacifica*; PALO), and wild turkey (*Meleagris gallopavo*; WITU).

	Mean ± SD		
Species	$\delta^{13}C$	$\delta^{15}N$	$\delta^{34}S$
COMU	-15.48 ± 0.15	16.55 ± 0.08	17.55 ± 0.38
PALO	-18.15 ± 0.14	11.80 ± 0.03	5.79 ± 0.40
WITU	-23.93 ± 0.14	4.54 ± 0.04	6.61 ± 0.31

Figure B1. Frequency distributions of delta values of stable carbon ($\delta^{13}C$), nitrogen ($\delta^{15}N$), and sulfur ($\delta^{34}S$) isotopes reported as parts per thousand measured in feathers of California black rails (Laterallus jamaicensis coturniculus) captured in the Sierra Foothills (FH; green) and San Francisco Bay Area (BA; blue).

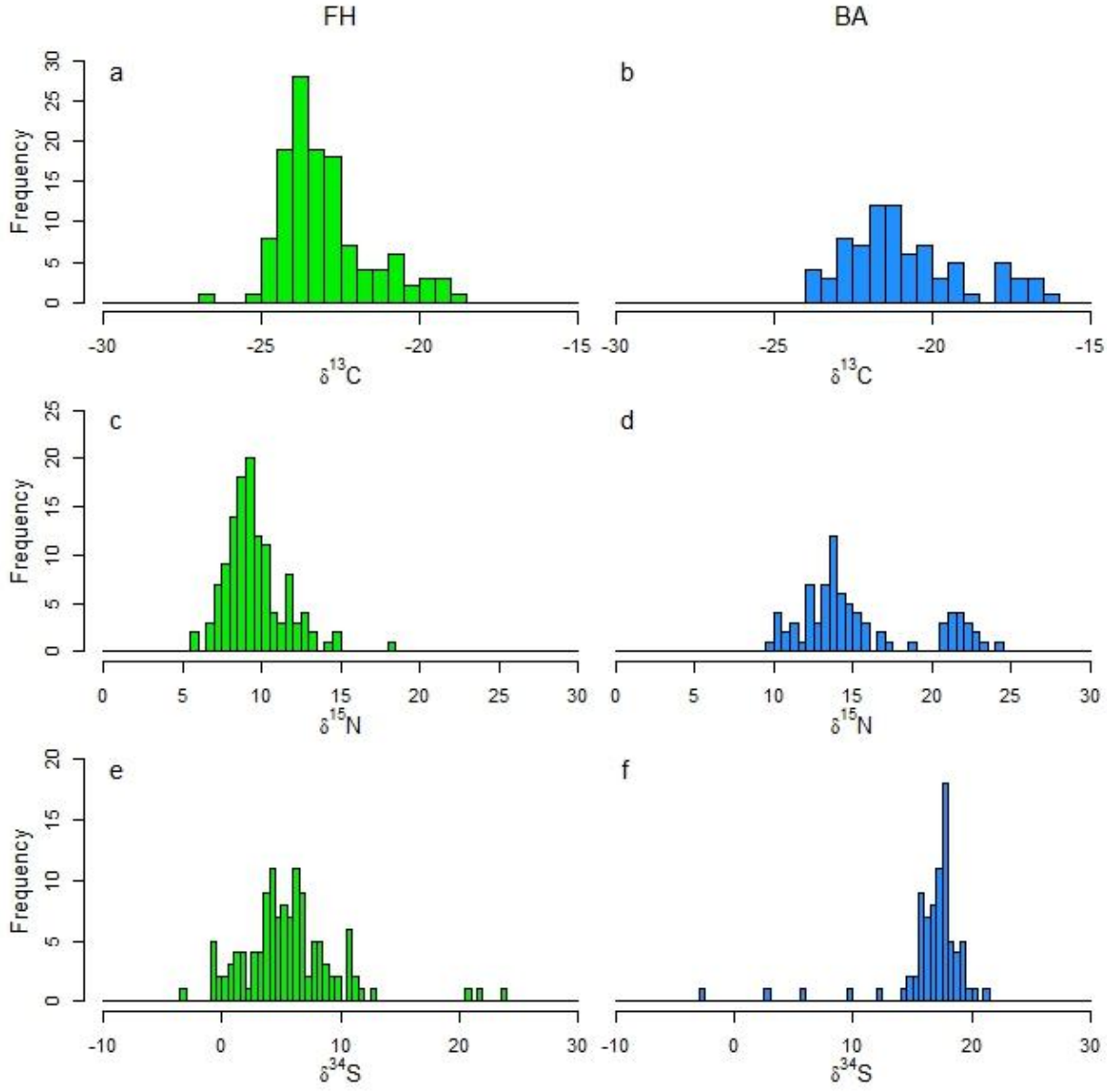


Figure B2. Box plots of median (with upper 75% and lower 25% quartiles) delta values of stable carbon ($\delta^{13}C$), nitrogen ($\delta^{15}N$), and sulfur ($\delta^{34}S$) isotopes reported as parts per thousand measured in feathers of hatch-year (HY) and after-hatch-year (AHY) California black rails (*Laterallus jamaicensis coturniculus*) captured in the Sierra Foothills (FH) and San Francisco Bay Area (BA).