Refining and defining riverscape genetics: How rivers influence population genetic structure

Chanté D. Davis¹ | Clinton W. Epps² | Rebecca L. Flitcroft³ | Michael A. Banks¹

¹Coastal Oregon Marine Experiment Station, Hatfield Marine Science Center, Department of Fisheries and Wildlife, Oregon State University, Newport, Oregon
²Department of Fisheries and Wildlife, Oregon State University, Corvallis, Oregon
³U. S. Department of Agriculture Forest Service Research Lab, Corvallis, Oregon

Abstract

Traditional analysis in population genetics evaluates differences among groups of individuals and, in some cases, considers the effects of distance or potential barriers to gene flow. Genetic variation of organisms in complex landscapes, seascapes, or riverine systems, however, may be shaped by many forces. Recent research has linked habitat heterogeneity and landscape or seascape configuration to genetic structure by integrating methods from landscape ecology, population genetics, and spatial statistics in approaches known as landscape or seascape genetics. However, functional differences between terrestrial or seascapes systems in comparison to riverscape topography (i.e., movement pathways for aquatic obligate species are constrained to river channels) make translating these approaches into freshwater analyses problematic. Studies that may be described as riverscape genetics (RG) have linked temperature, stream gradient, and confluences to genetic variability. Lack of consistency in methodology, however, has made comparisons across species and scales difficult. We provide a perspective on how RG could be used to provide a more comprehensive conceptual and applied understanding of connectivity and dispersal in freshwater systems. We describe four thematic areas of study representing current and future research opportunities and describe a basic workflow for conducting RG analysis. Although numerous methodological challenges remain, a RG approach can enhance our understanding of habitat heterogeneity in shaping gene flow and spatial genetic structure. These characteristics of populations are critical components for interpreting demographic and evolutionary consequences of habitat loss and fragmentation.

This article is categorized under:
Water and Life > Nature of Freshwater Ecosystems
Water and Life > Methods
Science of Water > Hydrological Processes

1 | INTRODUCTION

In most natural systems, gene flow and species dispersal are key processes with fundamental influences on demography and evolution of spatially structured populations. Gene flow and dispersal are influenced by the interaction of locally adapted life history traits and habitat heterogeneity. Genetic structure is influenced strongly by genetic drift, gene flow, and, in some cases, natural selection (Banks et al., 2013). Maintaining genetic diversity and gene flow may enable populations to respond to environmental change through the spread of locally adapted genes (Kawecki & Ebert, 2004). Yet, the rate and timing of changes to landscapes may be more rapid than the potential adaptive response by organisms, thereby increasing the threat of...
local and regional extirpations (Kawecki & Ebert, 2004). This may be particularly critical for freshwater species whose movement is restricted to connectivity within river networks that are highly vulnerable to fragmentation.

Understanding the role of habitat heterogeneity in shaping spatial genetic structure (Table 1) is necessary to interpret the evolutionary consequences of habitat loss, fragmentation, and environmental change. Yet, only recently have researchers begun to address this systematically. A fundamental component of traditional population genetic analysis is the comparison among groups of samples, relating genetic differences to geographic distance (also referred to as Euclidean distance) and considering “panmixia” (no spatial pattern) as a null hypothesis. The expected correlation between genetic structure and geographic distance is known as isolation by distance, or IBD (Rousset, 1997; Wright, 1943). In rivers, distance along waterways (waterway distance) has been recognized as a more biologically meaningful measure to relate to genetic structure, but even that measure may not capture other key characteristics of the river that may influence movement of a particular species or life stage (Selkoe, Scribner, & Galindo, 2016).

In terrestrial and oceanic systems, increasing numbers of studies have used systematic characterizations of land- or seasca to assess how habitat heterogeneity and configuration influence genetic structure. These approaches are known respectively as “landscape genetics” (LG) (Manel, Schwartz, Luikart, & Taberlet, 2003) and, more recently, “seascape genetics” (SG) (Selkoe, Henzler, & Gaines, 2008). These approaches move beyond explaining genetic structure solely as a function of IBD by generating models that describe how specific habitat enhance or provide resistance to the movement of individuals, and testing whether those models explain genetic differences better than IBD or panmixia. This is accomplished by integrating theory and methods from landscape ecology, population genetics, and spatial statistics (Manel et al., 2003; Selkoe et al., 2016).

Most LG analyses seek to use genetic data to test hypotheses about the influence of landscape on gene flow. Early LG studies often tested for barrier effects (“isolation by barrier,” hereafter IBB), where a specific landscape barrier (e.g., road,

### Table 1: Glossary of terms

| Term                             | Explanation                                                                 |
|----------------------------------|-----------------------------------------------------------------------------|
| Adaptive genetic variation       | Genetic differences with an effect on fitness.                               |
| Allelic richness ($A_t$)         | A measure of genetic diversity based on the average number of alleles per locus, sometimes considered indicative of a population’s long-term potential for adaptability and persistence. |
| Allele                           | A variant at a locus.                                                        |
| Biophysical model                | Spatially explicit modeling that uses mathematical formulations to simulate the interaction of biological and physical factors of a system. |
| Collinearity                     | The nonindependence of predictor variables.                                  |
| DEM                              | Terrain elevation data provided in digital form.                             |
| Dispersal                        | Movement of individuals to different localities that has the potential to lead to gene flow. |
| Edge                             | Connections between nodes in a graph-theoretic network.                     |
| Fixation index ($F_{ST}$)        | Measure of population genetic differentiation that reflects the proportion of allelic variation contained in subpopulations relative to the total genetic variation; may be calculated “pairwise” as a measure of differentiation between any two groups. |
| Gene flow                        | The transfer of genes from one population to another via movement followed by reproduction. |
| Genetic differentiation/structure| A measure of allele frequencies among subpopulations.                        |
| Graph theory                     | A branch of mathematics that deals with statistical descriptions of static networks. |
| Heterozygosity expected, $H_e$; observed $H_o$ | A measure of genetic variation. The proportion of loci expected to be heterozygous ($H_e$) and the observed proportion of heterozygotes, averaged over loci ($H_o$). |
| Hydrologic connectivity          | Connectivity that is mediated by the flow of water.                         |
| Least-cost path                  | Length of a path minimizing the cumulative resistance (distance weighted by the cost of traversing a particular habitat type) between two localities. |
| Locus                            | A specific location on a chromosome, to be informative in a population genetic study, should show variation among individuals. |
| Neutral genetic variation        | Genetic differences with no direct effect on fitness.                       |
| Node                             | An individual element within a network that represents a discrete unit (e.g., a population, spawning site, or sampling location). |
| Raster image                     | A grid image created in GISs.                                                |
| Riverscape                       | A mosaic of freshwater river habitat that is spatially structured and hierarchically organized across multiple scales. |
| Riverscape genetics              | An area of study that evaluates the effect of riverscape features on spatial genetic variation. Shares methodological similarities with seascape and LG. |
| Scale                            | The ratio or relationship between the distance on a map and the corresponding distance on the ground. |

Synthesis of terms from references Smouse and Peakall (1999), Selkoe et al. (2016), Manel and Holderegger (2013), Riginos and Liggins (2013), Hartl and Clark (2007). DEM = digital elevation model; GISs = geographic information systems; LG = landscape genetics.
rivers) is identified and levels of genetic variation among populations separated by the barrier are compared (Holderegger & Wagner, 2008; Storfer et al., 2007). Methods were developed to explore the complex effects of multiple landscape variables (e.g., precipitation, temperature, elevation) on individual movements leading to gene flow; genetic structure resulting from such habitat heterogeneity is often referred to as “isolation by resistance,” hereafter, IBR (McRae, 2006; Shah & McRae, 2008). A common approach to test IBR hypotheses, as pioneered in LG, assigns “resistance” values to different habitats based on their presumed effect on individual movements, which are then mapped across the study area (Cushman, McKelvey, Hayden, & Schwartz, 2006; Epps, Wehausen, Bleich, Torres, & Brashares, 2007). The estimated cumulative “cost” of movement between sampling locations (hereafter, “effective distance”) is then estimated (Shirk, Wallin, Cushman, Rice, & Warheit, 2010). Each set of resistance values can be regarded as a hypothesis to explain genetic structure among sampling locations. Typically, many such hypotheses are generated, and researchers determine which set of effective distance estimates is most strongly correlated with genetic differences (Cushman et al., 2006; Epps et al., 2007; Spear, Balkenhol, Fortin, Mcræ, & Scribner, 2010). For descriptions of a typical workflow for a LG study, see the study of Hall and Beissinger (2014) or Rellstab, Gugerli, Eckert, Hancock, and Holderegger (2015).

Studies using SG face challenges different from LG. Life histories of many marine organisms are characterized by large population sizes, high fecundity, external fertilization, and planktonic larvae. Ocean currents transport organisms during their planktonic larval phase leading to populations that are spatially distributed across considerable distances. The time and cost associated with adequate environmental and genetic sampling of such broadly distributed populations is a difficulty when conducting SG analysis (Lal, Southgate, Jerry, Bosserelle, & Zenger, 2017; Riginos & Liggins, 2013). As a result, empirical data describing dispersal for the majority of marine organisms are limited (Cowen & Sponaugle, 2009; Liggins, Treml, & Riginos, 2013). Due to the variable and complex nature of ocean currents, a single oceanographic graphic value cannot be attributed to a specific location, therefore LG-type approaches based on resistance surfaces have rarely if ever been employed (Hansen & Hemmer-Hansen, 2007; Riginos & Liggins, 2013). Instead, SG studies have revealed the influence of hydrogeomorphic properties on spatial genetic structure by combining genetic data with a variety of nongenetic information derived from complex models (e.g., sea surface temperature), current speed and direction, chlorophyll, or coupled biological-physical models) summarized over an appropriate time scale (Banks et al., 2007; Hansen & Hemmer-Hansen, 2007; Liggins et al., 2013; Riginos & Liggins, 2013; Selkoe et al., 2010). With increased computing power, simulations have provided a way to circumvent the lack of empirical data. For instance, larval dispersal models can be created by compiling data from physical oceanographic conditions and particle tracking models. Those simulated data can be used in a graph-theoretic approach, discussed in a later section, where migration probabilities derived from biophysical models define “edges” that connect patches of genetically distinct populations, or “nodes” to create a network (Johansson et al., 2015; Urban, Minor, Treml, & Schick, 2009). Node clusters and individual nodes or edges defined within the network reflect complex processes that contribute to spatial genetic structure. Each alteration to modeled migration probabilities reflects a hypothesized explanation of how functional connectivity affects genetic structure. Multiple hypotheses can be tested allowing researchers to determine which model generated patterns are most similar to observed genetic structure.

Recently, LG and SG approaches have been extended to studies of riverine organisms in research that can be called “riverscape genetics” (RG). Defined as an aquatic counterpart to LG, by Kanno, Vokoun, and Letcher (2011), RG analysis is becoming more common. In the context of RG (Table 1), our definition of “riverscape” bears resemblance to definitions for the terms landscape and seascape in LG and SG (Ward & Stanford, 1983; Wiens, 2002a). We define the riverscape as a hierarchically structured mosaic of differentially distributed habitat within freshwater environments (Benda et al., 2004; Carbonneau, Fonstad, Marcus, & Dugdale, 2012; Frissell, Liss, Warren, & Hurley, 1986). Our definition considers riverscapes as continuous entities where habitat patches have permeable boundaries (Fausch, Torgersen, Baxter, & Li, 2002; Wipfl, Richardson, & Naiman, 2007) but also acknowledges that variation among patches form discontinuities along the river continuum that are functionally important for riverine organisms (Benda et al., 2004; Thorp, Thoms, & Delong, 2006; Ward & Stanford, 1983). In this definition, processes that occur within a riverscape patch affect processes in neighbouring patches, forming complex interactions within and among populations of aquatic organisms at various spatial scales (Frissell et al., 1986; Miller, Luce, & Benda, 2003). Thus, studies that can be described as RG have sampled biota using spatially continuous and discrete sampling designs. Analysis conducted using spatially continuous sampling at the river network scale (e.g., Kanno et al., 2011) has identified waterway distance and physical barriers such as waterfalls and confluence number to affect spatial genetic variation (Earnest, Scott, Schaefer, & Duvernell, 2014; Kanno et al., 2011). Studies that have utilized spatially discrete sampling of biota (Bowlby, Fleming, & Gibson, 2016; Ozerov, Veselov, Lumme, & Primmer, 2012) similarly identified physical barriers but also revealed significant correlations between spatial genetic variation and other riverscape variables (e.g., seasonal water flow, gradient, high flow events, precipitation; Cook, Kennard, Real, Pusey, & Hughes, 2011; Dionne, Caron, Dodson, & Bernatchez, 2008; Hand et al., 2016; Tortonrot, Perrier, Bergeron, & Bernatchez, 2014). Therefore, fully extending land- and seascape approaches into the riverscape will require consideration of the dynamic features of rivers:
notably, the directional movement of water flow and physical network structure that is highly constrained and hierarchically organized across multiple spatial scales (Isaak et al., 2014).

1.1 Implications of water flow and rivers as networks in the development of methodologies for RG

Riverscapes include features that are rarely encountered in landscapes, and hydrologic processes that influence dispersal in aquatic environments differ from those in marine environments. For example, freshwater hydrologic connectivity is driven by downstream water flow (Poff et al., 1997); while in marine systems, ocean currents are largely responsible for this interaction (Hughes, Schmidt, & Finn, 2009). Seascapes lack strong ties to terrestrial landscapes, while many river systems exhibit complex branching patterns formed by iterative tributary junctions that are embedded in terrestrial habitat. This branching pattern is often called the “river network.” (Benda et al., 2004; Thorp et al., 2006) and is much more physically constrained with respect to potential connections among locations as compared to seascapes (Rodriguez-Iturbe, Muneep醚a-kul, Bertuzzo, Levin, & Rinaldo, 2009). Changes in hydrology add variability that is rarely observed in terrestrial landscapes. Directional water flow and river network pattern underline movement pathways for aquatic species, directly affecting patterns of gene flow, dispersal, and increasing the potential for isolation. Existing analytical methods developed for LG and SG may not adequately account for the effect of these physical differences on expectations of gene flow, dispersal, and connectivity.

The spatial configuration of the network is another important characteristic of riverscapes, but fitting spatial statistical models to stream networks is a particularly challenging task because of the flow-connected nature of freshwater habitats (Ver Hoef & Peterson, 2010). The pressing need to understand how connectivity, flow, and stream hierarchy affect dispersal or movement has led to development and adaptation of spatial statistical models in freshwater ecology (e.g., variograms and graph theory; for reviews see Dale & Fortin, 2010; Isaak et al., 2014). Development of more refined tools that incorporate geophysical properties and better reflect the spatial configuration of the river is an important step toward understanding connectivity (Ganio, Torgersen, & Gresswell, 2005; Peterson et al., 2013; Ver Hoef, Peterson, & Theobald, 2006). These methods have not been fully incorporated into analysis of spatial genetic variation of freshwater organisms. Therefore, developments of spatial statistical tools and models that expressly use the physical stream network in ecological studies have potential to be foundational for novel genetic analysis in freshwater systems (Brauer, Hammer, & Beheregaray, 2016).

Understanding how changes to rivers influence dispersal and genetic structure is essential for appropriate management and attempts to restore connectivity. Therefore, continued development of the RG approach could provide important tools for conservation, management, ecology, and evolutionary biology of species in these ecosystems. In this paper, we provide a perspective on the utility of such an approach for providing a more comprehensive and holistic understanding of dispersal and connectivity in river networks. We evaluate studies that self-identify, or could be described, as using a RG approach. Based on those studies as well as directions for future work, we describe four broad themes: (a) detecting the impacts of anthropogenic and natural barriers on dispersal, connectivity, and genetic structure, (b) identifying riverscape factors that affect the scale and pattern of spatial genetic structure in a stream, (c) separating effects of historical, and contemporary riverscapes on genetic structure, and (d) linking spatial adaptive genetic variation to the heterogeneous riverscape. Within each theme, we summarize what current research has contributed to understanding correlations between gene flow and environmental variation while also identifying avenues for continued exploration. Then, we describe three methods commonly used to quantify the effect of environmental variables on genetic structure. We provide a perspective on how to expand analysis so that each method may be more appropriate for RG and discuss the implications of such methodologies for study design.

2 LINKING GENETIC STRUCTURE WITH THE GEOPHYSICAL TEMPLATE OF THE RIVER SYSTEM

Our review of the literature has demonstrated that studies have linked dispersal barriers (Theme 1) and environmental variables (Theme 2) to contemporary gene flow and genetic structure within and among populations of riverine species. While the gains in understanding linkages between the geomorphological riverscape and genetic structure have been illuminating, substantial opportunities for growth remain. For example, incorporation of ancient riverscapes (Theme 3) would provide a control for effects of historic events on observed gene flow, thereby allowing for the correct correlation of contemporary genetic structure with effects from modern riverscape features. We found that the majority of studies have used markers with no effect on fitness (neutral genetic variation) but markers that are experiencing selection may bear more direct link to the environment. Therefore, a more visionary application of gene–environment associations (GEAs) may codevelop with RG and other leading edges of inquiry (e.g., evolutionary and molecular ecology, phylogeography, or epigenetics) into the
genetic basis of local adaptation using markers that are associated with direct effects on fitness (adaptive genetic variation; Theme 4). The following themes provide a glimpse of the breadth of knowledge that a RG perspective can bring to our collective understanding of dispersal and spatial genetic variation.

2.1 Theme 1: Detecting the impacts of anthropogenic and natural barriers on dispersal, connectivity, and genetic structure

The structure of the river network is inherently vulnerable to fragmentation that alters hydrologic connectivity (Fagan, 2002). Long-term population persistence is specifically related to connectivity. The physical distribution of habitat patches (structural connectivity) and the ways organisms navigate the river network to access specific habitat (functional connectivity) are components of connectivity that influence genetic structure in different ways. As organisms migrate among habitat patches at different places in the network, they experience environmental variation (e.g., temperature, stream gradient, or waterfalls) that produce variable resistances to movement. Naturally occurring disturbances (e.g., debris flows or fires) fragment habitat at seasonal and intermittent time scales, presenting barriers to dispersal that may erode over time, changing patterns of structural and functional connectivity throughout the river network. Anthropogenic barriers (e.g., dams, culverts, dikes) fragment the riverscape over longer timescales and exacerbate the effects of ongoing natural disturbance events (Benda et al., 2004; Reeves, Benda, Burnett, Bisson, & Sedell, 1995). Additionally, river network structure rarely allows alternative dispersal pathways; therefore, changes in connectivity caused by either type of fragmentation have potential to greatly influence dispersal and genetic structure (Hughes et al., 2009; Yamamoto, Morita, Koizumi, & Maekawa, 2004).

Barrier effects are especially evident for upstream passage of migratory species that encounter dams and fish ladder operations. In some cases, extinguition of fish has occurred above such barriers, while in other river networks fish are able to persist (Lindley et al., 2004). For example, Torterotot et al. (2014) evaluated effects of fragmentation by natural and non-natural barriers on Salvelinus fontinalis and found that the cumulative number of barriers correlated significantly with patterns of genetic diversity. Genetic structure is often identified among populations that are up- or downstream of these types of physical barriers for species including Chinook, brook trout, and chum salmon (Neville, Dunham, & Peacock, 2006; Torterotot et al., 2014). Natural barriers also affect connectivity of habitat within the river network by altering genetic structure for migratory and non-migratory fishes. Castric, Bonney, and Bernatchez (2001) associated waterfalls with decreased heterozygosity among populations of brook charr but the authors were unable to find significant genetic structure that was related to IBD or IBB. Leclerc, Mailhot, Mingelbier, and Bernatchez (2008) evaluated spatial genetic structure of yellow perch (Perca flavescens), finding distinct populations that were separated by a dam or zones of high velocity water flow that prevented migration.

Although the effect of barriers on genetic diversity was commonly investigated among studies we reviewed, distance (IBD) and barriers (IBB) alone rarely explained the majority of genetic variation in freshwater systems (Cook et al., 2011; Dionne et al., 2008; Earnest et al., 2014; Sprehn, Blum, Quinn, & Heins, 2015; Torterotot et al., 2014). Kanno et al. (2011) used eight microsatellite loci (Box 1) to calculate genetic diversity measures of resident fish (S. fontinalis) and associated weak structure with barriers that partitioned individuals among three tributaries. However, one tributary within the study lacked a physical barrier but still supported genetically structured populations. To better understand how genetic structure is shaped within the river we should look beyond barrier and distance hypotheses to include IBR.

2.2 Theme 2: Quantifying the effect of the riverscape on scale and pattern of spatial genetic structure

The effects of hydrologic connectivity on fish ecology have been studied in freshwater systems because hydrology is one of the primary contributors to the spatial configuration of habitat within the river network (Fagan, 2002; Fullerton et al., 2010). For example, Flitcroft, Burnett, Snyder, Reeves, and Ganio (2014) determined that the spatial distance between habitats and the availability of habitats for specific life histories are determinants in the pattern of juvenile coho salmon distribution throughout river networks. The authors hypothesized that stream flow contributed to observed spatial patterns but were unable to test this hypothesis with their dataset. Studies using multiple genetic and riverscape datasets to test hypothesis of IBR have demonstrated correlations between spatial genetic structure and temperature, stream gradient, number of confluences, drainage basin, seasonal precipitation, seasonal water flow, and high flow events (Cook et al., 2011; Kanno et al., 2011; Olsen, Beacham, et al., 2010; Olsen, Crane, et al., 2010). For example, Castric et al. (2001) quantified spatial genetic patterns in Mogurnda mogurnda sampled at 17 sites within multiple watersheds. They tested if genetic variation was correlated with linear distance, maximum stream gradient, elevation, or discharge. IBR models have successfully quantified genetic structure for a wide variety of fish species including salmonids, (Hand et al., 2016; Olsen, Beacham, et al., 2010; Olsen, Crane, et al., 2010; Ozerov et al., 2012) topminnows (Earnest et al., 2014), tropical freshwater fish (Brauer et al., 2016; Cook et al., 2011), headwater chub (Pilger, Gido, Propst, Whitney, & Turner, 2015), and electric fish (Cooke,
Landguth, & Beheregaray, 2014), and in a more limited capacity, freshwater mussels (Galbraith, Zanatta, & Wilson, 2015) and parasites (Sprehn et al., 2015). Collectively, these studies provide evidence that complex interaction between climate variation and habitat heterogeneity has shaped elements of spatial genetic structure in freshwater species, and further emphasize limitations of IBD used alone to explain genetic variation in streams.

2.3 Theme 3: Separating effects of ancient and contemporary riverscapes on genetic structure

Genetic diversity is the result of cumulative environmental and geological processes that have occurred at varying temporal and spatial scales, each leaving interpretable marks in the genome. As addressed above, dispersal barriers resulting from riverscape fragmentation may strongly influence spatial genetic structure of populations. Traditionally, species or populations that occupy high-elevation, dendritic tributaries are expected to be more physically isolated because of human alteration to downstream reaches. The increased probability of barriers increases resistance to gene flow with upstream populations. However, ancient historical events (e.g., the last glacial maximum) have shaped genetic structure and species diversification (Avise, 2000; Hickerson et al., 2010). Geological processes that restructured the range and distribution of biota across entire river networks included ancient mega-flood events and Pleistocene glacial cycles that dramatically changed connections.

BOX 1
DEFINING MOLECULAR MARKERS FOR GENETIC ANALYSIS

Mitochondrial DNA

A maternally inherited, small, circular strand of DNA is found in the mitochondria of cells. The molecule consists of a coding region, the majority of the molecule, and a control region (D-loop), responsible for regulating the production of gene products from the coding region. Mitochondrial DNA (mtDNA) is stable over time because it is present in multiple copies within the cell and the circular form resists degradation. However, it is maternally inherited and therefore haploid which provides less information than biparentally inherited nuclear DNA. Amplification is accomplished through polymerase chain reaction (PCR) and visualized by gel electrophoresis. Newer automated platforms have improved the ability to sequence mtDNA, resulting in the rapid sequencing of hypervariable regions and decreasing the time it takes to sequence the whole genome. See Rand (2001) for a review of Mitochondrial DNA.

Microsatellites

Short repeats of nucleotides (i.e., guanine [G], thymine [T], cytosine [C], adenine [A]) that are found throughout the nuclear genome. A microsatellite locus repeat consists of two (dinucleotide repeat), three (trinucleotide repeat), or four (tetranucleotide repeat) nucleotides, although more are possible (Schlötterer & Harr (2001)). A locus has variable repeat lengths alleles) that will vary by locus, species, and population. For instance, a heterozygous individual may show one allele with four GT repeats (GTGTGTGT), while the second allele has seven repeats (GTGTGTGTGTGTGT). This difference in length can be assessed by gel electrophoresis or by using automated sequencing platforms after PCR amplification (Schlötterer (2000)). Marker isolation and optimization are time consuming but once identified they are easy to amplify at relatively low costs. Microsatellites that are associated with neutral genetic variation have been used to identify bottlenecks, parentage analysis, gene flow, population structure, and many other evolutionary effects (Selkoe & Toonen (2006)). Although the advent and growth of next generation sequencing contributed to the decline in use of microsatellites, they are still a useful marker. For a review of Microsatellites and their use see Selkoe & Toonen (2006) and references therein.

Single nucleotide polymorphism

Individuals of the same species share many DNA sequences that are almost identical and differ only at a few nucleotide positions within the sequence. At these sites, the two copies of a gene in a heterozygous individual show different nucleotides, whereas a homozygous individual shows only a single nucleotide. Finding single nucleotide polymorphisms (SNPs) first requires the sequencing of many genes or regions of a genome, a process that has decreased in cost and time as more automated platforms have become available. A single SNP offers little power to distinguish genetic structure among populations, but automated sequencing platforms now enable discovery of 10–100 K’s SNPs with relative ease. SNPs are also used in similar contexts as microsatellite markers. For a review of SNPs see Morin, Luikart, Wayne, and the S. N. P. workshop Group (2004) and references therein.

Many journals require that molecular markers (mtDNA, microsatellites, and SNPs) are published and offer specific journals and databases for the purpose (i.e., Molecular ecology notes, National Center for Biotechnology Information [NCBI] provides a searchable link for GenBank [www.ncbi.nlm.nih.gov/genbank/]).
among populations (Hickerson et al., 2010). Without accounting for the effects of ancient riverscapes on contemporary spatial genetic structure, genetic structure from such ancient legacies could be incorrectly attributed to a modern riverscape feature, complicating efforts to understand and predict changes in dispersal or genetic structure.

Molecular techniques and computer simulation have aided in the ability to detect legacies of ancient geographic, geologic, and climatic events on extant populations (Hickerson et al., 2010). Identifying evidence of bottlenecks, decreased genetic variation associated with long periods of isolation, and other genetic signatures have helped to illuminate ancient colonization events (Swatdipong, Primmer, & Vasemägi, 2009). For instance, Waples (2001) describes the contemporary patterns of diversity among Pacific salmon lineages as resulting from interspecific diversification following the last glacial maximum and evolution during the Holocene. High-magnitude but low-frequency disturbance regimes around the last glacial maximum caused massive extinction events and reshaped entire river networks enabling species diversification in Pacific Northwest salmonids. Similar historic relationships exist between the present day diversification and historical geologic events within the lineages of marine organisms. For instance, genetic discontinuities among the greenshell mussel (Perna canaliculus) have been attributed to sea level fluctuations that occurred throughout the Pleistocene when dynamic geological and hydrological processes established Cook Strait (Wei, Wood, & Gardner, 2013). Without accounting for this legacy, it would not be possible to distinguish between current and past processes that are influencing genetic variation.

LG and SG studies offer several examples of such approaches (Epps & Keyghobadi, 2015). For instance, in a LG study of a desert-dwelling plant, Dyer, Nason, and Garrick (2010) analyzed the correlation of genetic distances among sampling locations to a distance matrix summarizing phylogeographic variation. Then, from the residuals of that relationship, they evaluated effects of IBR using effective distance matrices estimated from the current landscape. Similarly, based on recognition that greenshell mussel divergence among the north and south islands of New Zealand had been attributed to historical events, Wei et al. (2013) used an SG approach to quantify contemporary genetic diversity within the two historically diverged populations correlated with regional environmental data. They explored the relationship between pairwise genetic differentiation and distance matrices of environmental and geological data by using regression analysis for each island separately. In a freshwater example, Osborne, Perkin, Gido, and Turner (2014) investigated landscape-scale spatial genetic structure among three Great Plains fish species using a simple linear regression of allelic richness ($A_r$) and latitude. The authors found that increasing genetic variation correlated with increasing latitude, reflecting a postglacial colonization history. Then, after accounting for the effect of latitude, the authors investigated whether species-specific genetic diversity ($A_r$) and structure ($F_{ST}$) was influenced by modern fragmentation. They found that genetic variation among the sites did not reflect the influence of contemporary barriers on gene flow, highlighting the importance of first controlling for influences of historical processes (Gouin, Souty-Grosset, Borquez, Bertin, & Grandjean, 2011; Hanffling, Hellemans, Volckaert, & Carvalho, 2002).

### 2.4 Theme 4: Linking spatial adaptive genetic variation to the heterogeneous riverscape

Individuals that occupy the same local riverscape experience selective pressures that act to maximize individual fitness within a specific habitat. However, local adaptation is not a guaranteed outcome for individuals in all populations because it is mediated by the life-history of the organism, evolutionary processes, and environmental interaction. For instance, gene flow will counter effects of selection by maintaining frequencies of alleles, genetic drift balances selection by buffering increased frequency of adaptive loci (especially in small populations), and the underlying genetics of traits (e.g., plasticity and epigenetic effects) can constrain adaptation (Hoban et al., 2016; Kawecki & Ebert, 2004). Furthermore, selective pressure from the heterogeneous riverscape will vary spatially and temporally, favoring genotypes differentially over time and space (Hoban et al., 2016). Nevertheless, many GEA studies find convincing evidence of local adaptation. For instance, Xu et al. (2013) identified loci that were related to stress response of a cyprinid fish, and demonstrated adaptation to alkalinity by sequencing RNA. Torres-Dowdall, Handelsman, Reznick, and Ghelmanb (2012) found significant correlations between genetic differentiations among multiple traits among Trinidadian guppies reared in high- and low-predation conditions. Whitehead, Galvez, Zhang, Williams, and Oleksiak (2011) studied plasticity in killifish (Fundulus heteroclitus) and found genes related to osmotic shock correlated with local adaptation to pollution tolerance.

Evidence of local adaptation in natural populations is challenging to acquire because demonstrating these types of effects typically requires that a GEA leads to higher fitness by an organism in one environment relative to another environment (Hereford, 2009). This can be accomplished by translocation or common garden experiments, which provide estimates of heritability and genetic effects from replicated experimentation. These types of studies are difficult to conduct on wild populations because of difficulties associated with rearing in laboratory conditions or limitations associated with listed species status, among other complications. In the absence of experimentation, computer simulations have provided the means to identify molecular markers that may be experiencing selection (Gunther & Coop, 2013; Lotterhos & Whitlock, 2014, 2015; Narum & Hess, 2011). For example, the program LOSITAN produces an estimate of $F_{ST}$ for each locus from an empirical dataset and then simulates the expected distributions of $F_{ST}$ and heterozygosity ($H_e$) under neutral processes. Loci that fall
outside of the distribution are “outliers” and may be experiencing selection (Antao, Lopes, Lopes, Beja-Pereira, & Luikart, 2008). This method was used by Chang et al. (2013) to test if genetic differentiation among cyprinid fish (*Leuciscus waleckii*) sampled from alkaline and freshwater environments was caused by selection pressure from the increase in alkalinity. Using microsatellite loci and mtDNA sequences the authors detected a single outlier locus that may play a role in local adaptation to alkalinity for this species. Continued discovery of markers that are potentially under selection will improve the quantification of patterns of adaptive genetic differentiation (Forester, Jones, Joost, Landguth, & Lasky, 2016; Lotterhos & Whitlock, 2014, 2015), advancing research in evolutionary ecology and epigenetics that seeks to understand adaptive evolution in natural populations.

The adaptability of spatially structured populations is affected by dispersal of organisms across their spatially heterogeneous environment. Therefore, understanding how dispersal is affected by functional and structural connectivity would provide useful information applicable for the study of adaptive evolution. The movement or redistribution of local adaptive genetic variation among subpopulations has fitness consequences. Evaluating adaptive genetic variation as well as neutral genetic variation in a RG context may clarify processes influencing genetic structure and provide predictive capabilities for interpreting the influence of new challenges to native species (e.g., biological invasions, hybridization resulting in decreased fitness, disease vectors).

### 3 APPLYING THE MECHANICS OF LG TO RIVERSCAPES

The themes that were identified above illuminate the importance of incorporating a RG perspective in analysis of freshwater organisms. To achieve this goal, a robust study design and novel analytical methods that account for directionality of flowing water and spatial layout of the river network river system are needed. Although there is not a unifying or “one-size-fits-all” approach in LG or SG, shared similarities can be used as the basis for a unifying set of approaches. Therefore, we recommend RG practitioners carefully consider reviews of LG (e.g., Balkenhol, Gugerli et al., 2009; Segelbacher et al., 2010) and SG (e.g., Liggins et al., 2013; Selkoe et al., 2016). In this section, we discuss the similarities in approaches with regard to application of dissimilarity matrices, resistance surfaces, and network and graph theories (Dyer et al., 2010; Murphy, Evans, Cushman, & Storfer, 2008; Proulx, Promislow, & Phillips, 2005).

RG analysis requires a dataset representing genetic variation (Box 2) among samples at an individual or a population level, and a suite of measured environmental variables (Box 3) that are hypothesized to affect genetic variation. The more complex task, once datasets are available, is to select an appropriate suite of analytical tools with the power to detect effects while minimizing Type I and Type II error. Although many methods have been used to assess these relationships, there is ongoing discussion in the LG literature regarding benefits and limitations of these tests. We refer readers to Legendre and Fortin (2010) and Legendre, Fortin, Borcard, and Peres-Neto (2015) for reviews on the discussion. Typical methods include multiple-regression on distance matrices (MRDM; Bowlby et al., 2016; Diniz-Filho, Nabout, Telles, Soares, & Rangel, 2009b; Kanno et al., 2011), causal modeling (Cushman et al., 2006; Cushman, Wasserman, Landguth, & Shirk, 2013), ordination techniques (Kierepka & Latch, 2015), and regression modeling (Hand et al., 2016), but the most common tools are the Mantel test and its derivatives (Mantel, 1967; Smouse, Lone, & Sokal, 1986; Storfer et al., 2007; Zeller et al., 2016).

#### BOX 2

**QUANTIFYING GENETIC VARIATION IN POPULATION GENETIC ANALYSIS**

DNA is extracted from tissue samples following an extraction protocol. PCR amplifies specific portions of the genome, based on the molecular marker selected for use (Box 1). Amplified DNA fragments are visualized using gel-electrophoresis and genotypes (e.g., microsatellites [Selkoe & Toonen, 2006] or SNPs [Morin et al., 2004]), or whole sequences are identified. Data are evaluated for quality and evidence of genotyping or sampling errors (Morin et al., 2010). The resulting dataset is used to calculate genetic diversity values that describe the diversity of gene variants among samples (*allelic richness, A*; Kalinowski, 2004), differences in variation of each population compared to the total population (*F*<sub>ST</sub> or other measures; Merimas & Hedrick, 2011; Weir & Cockerham, 1984) and *heterozygosity* (*H*<sub>e</sub>). Individual assignment analysis and clustering algorithms (e.g., STRUCTURE [Pritchard, Stephens, & Donnelly, 2000], BAPS [Corander, Marttinen, Siren, & Tang, 2008]) can be used to identify patterns among individuals that may reflect spatial structure.
BOX 3

QUANTIFYING THE RIVERSCAPE IN ECOLOGICAL ANALYSIS

The selection of environmental features for analysis that characterize the riverscape is predicated on the expectation that they have potential influence on target species (Steel et al., 2010). Therefore, a priori use of expert opinion or the scientific literature can identify potential factors that are expected to influence movement. Data are then collected following an appropriate sampling design (discussed in text) targeting a relevant spatial and temporal scale for the phenomena investigated (Carbonneau et al., 2012; Steel et al., 2010). As remote sensing has increased in accuracy and quality, providing data that are widely available at multiple spatial scales (e.g., LiDAR [light detection and ranging] or drone imagery), geoprocessing needs also increase (Carbonneau et al., 2012; Hugue, Lapointe, Eaton, & Lepoutre, 2016; Lane & Chandler, 2003). In response, technical toolboxes have been and continue to be developed for use in geographic information system (GIS) applications. For example, to investigate physical and chemical parameters, STARS (Peterson & Ver Hoef, 2014), SSN (Ver Hoef, Peterson, Clifford, & Shah, 2014), and FLoWS (Theobald, Peterson, Ferraz, Wade, & Sherburne, 2006) provide tools that will predict catchment-scale information.

Once extracted, spatial data should be evaluated for the presence of heteroskedasticity or bimodal distributions and transformed if necessary to avoid violating assumptions when applying regression analysis (Pardoe, 2012). The relationships among data are also evaluated for collinearity to identify evidence of spatial autocorrelation. In rivers, autocorrelation among some variables is likely when sites are flow-connected (i.e., temperature, pH, or dissolved O2) (Peterson et al., 2013). Additional correlation between variables occurs regardless of flow-connectedness. For example, elevation typically increases with increasing distance from the river mouth, while stream width often decreases. In general, only one of the correlated variables is retained, although including multiple variables may be more realistic of natural processes. Finally, the variable list may be further transformed into a smaller number of uncorrelated variables by reducing the dimensions of the dataset and identifying, which variables account for the majority of the variability in the data (e.g., principal components analysis [PCA] or principal coordinates analysis [PCoA]), although the use of synthetic variables that may be developed using these tools may complicate biological interpretation (Cox & Cox, 2001).

FIGURE 1 A hypothetical watershed that includes a forest, dendritic river network, and land that has been developed for industry, agriculture, and urban use. Sampling locations are represented by black circles.
3.1 Dissimilarity matrix: Comparisons among sites across large spatial distances

The simplest comparison that can be made is to test if genetic differences among individuals or populations are related to differences in factors (i.e., elevation, temperature, or stream depth) that have been measured at each of the sampling sites (Figure 1). Although such an approach may not adequately describe the riverscape between sampling sites it is a common approach in riverine systems. Pairwise differences are calculated for each factor, forming a dissimilarity matrix (Legendre & Legendre, 2012). Genetic variation is measured at each site and pairwise differences between sites are calculated. The datasets are compared to identify if changes in any variables correlate with identified genetic differences. In an example of this approach, Castric et al. (2001) hypothesized that a greater occurrence of permanent waterfalls existed in areas with high altitudinal differences, and therefore hypothesized that genetic structure would exist across the permanent barriers. To test this, the authors formed a dissimilarity matrix of altitude variation following the shortest waterway distance between 30 Brook Charr (S. fontinalis) populations and compared these data to a pairwise genetic distance matrix generated from a suite of six microsatellite markers. Similarly, Kanno et al. (2011) evaluated if mean stream temperature, mean stream gradient, waterway distance, number of seasonal barriers, and number of confluences were related to pairwise genetic variation among populations of S. fontinalis. To test for IBR, the authors created a dissimilarity matrix of each variable by reach using regression analysis to determine which predictor explained genetic structure.

While the flexibility provided by the dissimilarity matrix makes it a versatile tool for addressing a variety of questions, the method does not allow for accurate interpretation of processes between sampling sites and may not capture a full riverscape perspective. For instance, the above example where Castric et al. (2001) used the difference in elevation among sites as a measure of potential waterfall barriers could have been strengthened by a detailed estimation of such barriers along each section of waterway using a fine-scaled digital elevation model (DEM). Thus, dissimilarity matrices provide some context of the differences among patches that contribute to overall genetic variation, but cannot easily capture the continuous exposure to environmental selection and resistance experienced by organisms moving through the riverscape. The following sections present methods that were developed as a way to quantify the contribution of these “en-route” effects on genetic variation in terrestrial and marine environments.

3.2 Resistance surface and path-based analysis to calculate effective distance

The resistance surface transforms the landscape into numerical values that depict different habitats, substrate, vegetation, or other features of interest so that hypotheses about the cost or presumed influence of those features on movement (i.e., effective distance) can be tested (Zeller, McGarigal, & Whiteley, 2012). Analytical GIS tools have enabled widespread use of resistance surfaces to calculate effective distances for species in heterogeneous environments. A raster image is a grid of cells (pixels) representing individual or multiple variables (Figure 2). Each cell in the raster is assigned a numerical value. In the simplest raster map, values can reflect the presence or absence of a variable (e.g., road or barrier). In more complex models, weights reflect the presumed influence of each variable on species dispersal, movement, or gene flow (Epps et al., 2007; Shirk et al., 2010). Cells in the final raster represent cumulative weights that are derived by independently summing each cell across all variables. The values or weights on the resistance surface allow researchers to quantify the influence of covariates (e.g., elevation, gradient, temperature) on some response variable of interest (e.g., movement, genetic differentiation, gene flow). Between two points, for example, Epps et al. (2007) employed resistance surfaces to understand the role of slope and distance in genetic structure of 26 bighorn sheep populations using 14 microsatellite loci. To test for IBR, Epps et al. (2007) created raster maps from DEMs and established 18 topographic resistance models that represented a range of weighting schemes based on slope. Then, using least cost path analysis (Table 1), they estimated the cumulative cost along the least costly path between each pair of populations for each resistance model. The resulting matrices of effective distance were tested to see, which was most strongly correlated with matrices of pairwise estimates of gene flow.

Rasterized maps work well for depicting difference in terrain across a terrestrial landscape. The hypothetical landscape shown in Figure 2a includes roads, river, forested, and urban areas. For a terrestrial organism, the weighted rasterized grid may represent features such as roads, nonforested habitat, or wide high-flow river sections as somewhat or highly resistant to passage (e.g., weight = 3). However, there are multiple routes of travel that would avoid these impediments. Identification of the most likely route that will be travelled is based on knowledge of species-specific movement and biological needs of the organism, and the assumption that organisms will tend to move through the habitat in a pattern that represents the least costly path of resistance (e.g., least-cost path analysis, Wang, Savage, & Shaffer, 2009). The increase in use of resistance surfaces and path-based analysis for LG reflects, in part, the potential for such analyses to inform conservation efforts by clarifying how complex landscapes influence movement.

Although resistance surfaces have worked well in a terrestrial landscape, this technique may behave differently in river networks. In a freshwater river context (Figure 2b), weights for the grid cells could reflect areas with a dam, culvert, or...
waterfall as somewhat or highly resistant to passage, but different velocities of free-flowing water could also vary in resistance. In contrast to Figure 2a, the dendritic river network depicted lacks alternative pathways (Figure 2b). Path-based analysis in this situation can be used to describe costs of travel among sites, different weighting schemes will not result in different travel paths between sample sites (as is usually the case for a landscape [Castillo, Epps, Davis, & Cushman, 2014; Cushman et al., 2006; Zeller et al., 2012] or seascape), but instead reflects cumulative costs along the path of travel (Landguth, Bearlin, Day, Dunham, & Travis, 2017). Thus, discriminating between models based on linear multipliers of resistance is not possible, as correlation with genetic distances among sites will not change. Alternative transformations (i.e., exponential or power functions) of habitat variables may provide resistance surfaces with contrastable differences. A further complication with resistance surfaces in RG is that existing methods for acquiring accurate measurements of the riverscape at fine spatial scales and across broad extents within these dynamic systems are limited (Fausch et al., 2002; Mertes, 2002).

3.3 | Spatial graphs and network theory model connectivity

Theoretical models available in graph-theoretic analysis are well suited to model a dendritic environment such as freshwater rivers. In a graph-theoretic approach, the user is able to interpret structural or functional connectivity of a population by assigning habitat patches or populations as “nodes” (e.g., spawning or breeding sites, reef structure or discrete habitat) that are connected by “edges” representing any measure of connectivity (e.g., pairwise genetic distance, dispersal rates, or migration patterns; Hines & Borrett, 2014; Urban et al., 2009). The resulting graph or network aids visual depiction of levels of clustering or metapopulation structure among the nodes, revealing relationships that are not easily identifiable in other ways (Proulx et al., 2005; Urban et al., 2009). More importantly, this method can provide an efficient characterization of connectivity at multiple spatial scales. Graph topology analysis provides additional interpretation of network structure (Urban et al.,

---

**FIGURE 2** The hypothetical watershed presented in Figure 1 has been transformed into a raster image. The grid cells are weighted to reflect the costs of travel for a terrestrial organism (a) and a freshwater organism (b).
In 2009, the sampling sites from the hypothesized riverscape shown in Figure 1 were transformed into two hypothetical networks representing different life histories. In each, nodes represent spawning habitats that are connected by dispersal. Examining node “degree,” the number of edges connected to a node, may identify populations that have possible genetic isolation (i.e., fewer or weak edges). Sequences of node-edge pairs that are oriented in closed loop “cycles” provide information about potential subpopulations that may experience gene flow. For example, site 3 (Figure 3b) is in a reservoir created by a dam. Few edges connect it to the network, suggesting the potential for isolation that may be reflected by increased genetic structure and lower heterozygosity than other sites. In this example, a mobile aquatic organism may experience limited connectivity with the rest of the network if migration into and out of the reservoir was not completely restricted. In Figure 3, a directed network was used to reflect life history of a sessile organism that experiences dispersal during the larval phase. The direction of water flow drives patterns of dispersal. In a directed network, sites upstream connect to all other flow-connected sites that are downstream. Site 2 may therefore reflect similarity with both upstream sites and have greater heterozygosity, while upstream sites are increasingly dissimilar from each other. Thus, in a distinctly different approach than the distance-based statistics commonly employed in LG, network statistics can be used to summarize testable hypotheses about the influence of riverscape on the characteristics of nodes, such as genetic diversity.

In an example of a weighted network analysis applied in rivers, Schick and Lindley (2007) depicted source-sink dynamics of spring-run Chinook salmon (Oncorhynchus tshawytscha) with a historical perspective. Nodes reflected the size of spawning populations; edges that linked nodes were constructed from a migration matrix calculated as a function of distance, dispersal probability, and population size. The authors tested if changes in topology reflected cumulative effects of historical migration barriers by creating multiple networks from which they identified extirpations, loss of source nodes, and changes in structure of node cycles. If applying a RG approach to this analysis, the Schick and Lindley network could be used to establish predictions that could be tested using empirical genetic data. For example, a prediction that cycles reflect potential subpopulations within the network could be evaluated from genetic samples collected at all sites (nodes). Significant pairwise measures of genetic structure or identification of genetically distinct clusters could provide support for the presence of subpopulations that may be meaningful for conservation and management of the species.

SG demonstrates the utility and flexibility of networks and graph-theoretic analysis to incorporate genetic, biophysical, and hydrogeological model into analysis. Johansson et al. (2015) investigated connectivity (IBD) of giant kelp (Macrocystis pyrifera) throughout their northeast Pacific distribution to evaluate if biogeographic population structure was explained by ocean transport. A network built from empirical data consisted of nodes, genetically distinct populations that were identified by individual assignment tests (box 2), and edges that were weighted by pairwise values of genetic differentiation ($F_{ST}$). To test if clustering observed within the network resulted from propagule dispersal by ocean transport, the authors constructed a network where edges reflected transport time between nodes based on seasonal oceanographic transport distance as modeled by Lagrangian particle simulations. Network theory as applied by SG has great potential for development of RG where similar relationships between geographic space, organism life history, and hydrology also exist.

### 3.4 A model for riverscape genetics

Although limitations are associated with each of the methods currently employed by LG and SG, we view effective distance and network theory to be pertinent methods for application in riverscape genetics. The flexibility of directed networks within graph theory allows researchers to weight edges and assign direction thereby accommodating a greater variety of connectivity scenarios that exist in riverscapes. The effective distance model employed by LG provides a useful means for accounting
for resistance to movement that occurs between habitat patches thereby accounting for costs that also contribute to observed genetic variation. By combining these methods (hereafter, “effective distance network”) we propose that the river may be modeled as a network of habitat patches (nodes) that are linked by a measure of effective distance (edges).

4 | USING COMPUTER SIMULATIONS IN RIVERSCAPE GENETICS

Computer simulations are used in a variety of scientific fields to validate findings, make predictions, build scientific theories, and test hypotheses. Computer simulations offer tools to explore natural systems and provide insights about system functions when it may be impractical to do so with empirical data (Epperson et al., 2010; Hoban, Bertorelle, & Gaggiotti, 2011). For example, larval dispersal in marine systems results from the physical movement of water; therefore, larval settlement and resulting gene flow are connected to oceanographic conditions. Empirical data describing larval dispersal for marine organisms are sparse but computer simulations have provided a way to estimate these processes. The utility of simulations is vast and varied; they can be used to evaluate current conditions, predict future conditions, and even recreate past conditions (Epperson et al., 2010). A spatially explicit individual modeling program (Landguth & Cushman, 2010) was used by Castillo et al. (2014) to simulate mating and dispersal of American Pika at Crater Lake National Park in Oregon as a function of environmental resistance. The authors also collected empirical data, and using resistance surfaces to analyze effective distance they determined that gene flow was affected by topographic complexity, water, and aspect. The modeled predictions of genetic variation were compared against empirical data to evaluate if resistance values could generate the observed genetic variation, and more importantly, whether the analytical methods used to pick the “best” model were likely to select the correct explanatory model. Simulations provide a useful method to test multiple hypotheses about the evolutionary mechanisms that underlie the observed spatial patterns. User-friendly programs have made simulation more accessible to researchers who conduct such analysis in aquatic environments (e.g., CDMETAPOP [Landguth et al., 2017], AQUASPLATCHE [Neuenschwander, 2006], SPLATCHE2 [Ray, Currat, Foll, & Excoffier, 2010]).

The predictive capability of computer simulation is another useful function with direct application to SG and RG. Computer simulations are used to build hydrologic models that predict flow and runoff in freshwater rivers or current fluctuations and ocean circulation in marine environments. Typically, models predict processes that are not easily acquired through physical sampling of the environment. For instance, Galindo et al. (2010) modeled larval dispersal of the intertidal acorn barnacle Balanus glandula to predict larval settlement. The authors coupled a regional ocean modeling system (ROMS) with an ecosystem model (carbon, silicate, and nitrogen ecosystem [CoSINE]) as well as models of larval development and particle-tracking to simulate larval dispersal. Genetic structure was calculated for the settlement sites that were predicted by the simulated data. Predictions of genetic structure were compared against empirical data to determine which model best described genetic structure. Computer simulations can be powerful and cost-effective tools that enable researchers to understand the effects of complex processes on evolution and demography but should be used with care. Simulated models are essentially detailed hypotheses about processes; therefore, without empirical data to confirm or deny findings and without quality inputs (model parameters), modeled predictions may differ quite considerably from reality.

5 | CONSIDERATIONS FOR BUILDING AN APPROPRIATE SAMPLE DESIGN

In the following section, we highlight considerations for sample design in the context of riverscape genetics. Of the many decisions associated with design and implementation for a robust sampling plan, sample location, sample number, marker type, and loci number (subsection: Genetic data) are often the most difficult (Oyler-McCance, Fedy, & Landguth, 2012). We identify, when appropriate, LG studies that have addressed these questions but maintain that these issues must be reevaluated for riverscapes.

5.1 | Sample collection

Sampling design is a fundamental component of robust research but it is unclear how different sampling strategies affect interpretation of spatial genetic variation based on riverscape data. Sampling is often associated with time, space, and cost limitations that ultimately may dictate overall feasibility of a project, but other factors are sometimes overlooked. For example, the extent of the distribution of an organism and its mobility dictate the scale of habitat patches that should be considered within the scope of study. Therefore, collecting an appropriate number of samples relates to whether sampling will occur at the level of individuals or populations and with continuous or discrete sampling units (Fausch et al., 2002; Le Pichon, Tales, Gorges, Baudry, & Boet, 2016; Wiens, 2002b). LG has embraced individual level sampling and analysis,
while both SG and RG studies are primarily reliant on population level sampling. For example, LG researchers have compared spatially explicit individual-based models with population-based sampling designs typical of classic population genetics to understand the difference between sampling individuals and populations in interpreting population structure (Landguth & Cushman, 2010). Simulation tools are now available that allow researchers to evaluate interactions between gene flow and selection in terrestrial systems but are limited in flexibility for modeling diverse reproductive and dispersal strategies of marine and freshwater organisms (Landguth, Muhlfeld, & Luikary, 2012). A number of papers in LG also have investigated the power of genetic tools to detect spatial genetic structure based on variation in sample design, statistical methodology, and complexity of the landscape, using continuous and discrete population distributions (Balkenhol, Waits, & Dezzani, 2009; Cushman et al., 2013; Diniz-Filho, Nabout, Telles, Soares, & Rangel, 2009a; Jones et al., 2013; Landguth, Cushman, & Johnson, 2012; Landguth, Fedy, et al., 2012; Lotterhos & Whitlock, 2015; Murphy et al., 2008). While the gains in understanding linkages between sample design and genetic structure have been illuminating for terrestrial habitats, the unique features of riverscape genetics will require reevaluation of all these sampling questions in that context.

### 5.2 Genetic data

Selection of an appropriate molecular marker requires understanding the capabilities and limitations of each marker and what type of genetic information is necessary to test the hypothesis (Box 1; Schlötterer, 2004). For example, SNPs are easy to generate and many sequencing platforms are available for SNP discovery, but single SNPs have limited information content because there are few polymorphisms per locus (Morin et al., 2004). Therefore, hundreds may be needed for appropriate power to detect genetic structure (Schlüterer, 2004). In comparison, microsatellites are more time consuming to isolate but tend to have more polymorphisms per locus; therefore, fewer are needed to garner the necessary power to detect genetic differentiation (Chambers & MacAvoy, 2000; Li, Korol, Fahima, Beiles, & Nevo, 2002; Selkoe & Toonen, 2006). The mtDNA would be useful marker choice when tissue samples are degraded because its structure resists degradation, but because inheritance of mtDNA originates from a single parent, the applicability of this marker is limited (Box 1). Next generation sequencing has improved the ability to acquire genetic data of all marker types at increasingly affordable costs (De Wit et al., 2012; Helyar et al., 2011). Increasing interest in the development of novel genetic tools will continue. Deciding on the appropriate marker will require trade-offs between costs of acquiring the number of loci that have statistical power to identify differences, and the number of individual samples to collect (DeFaveri, Viitaniemi, Leder, & Merila, 2013; Hess, Matala, & Narum, 2011; Morin et al., 2004). As demonstrated by Landguth, Fedy, et al. (2012), a greater number of polymorphic loci, rather than increased sample size, resulted in greater power to detect LG relationships. To identify this correlation, the authors employed a spatially explicit individual assignment program, CDPOP, to simulate genetic differentiation and modeled scenarios that varied in number of loci, number of alleles per locus, and the number of sampled individuals. Beyond conducting analysis similar to what has been accomplished in LG for selection of molecular markers, it would be pertinent to also understand the consequences of variation in hierarchical river structure, continuous sampling, and network configuration on the power of a molecular marker to detect correlations between riverscape features and spatial genetic variation.

### 5.3 Environmental data

Field-based data collection at fine-spatial scales that encompass the extent of a species range has often been cost prohibitive or limited by restricted accessibility to the watershed for data collection (Fausch et al., 2002). A locally intensive map may include a few river reaches within the study extent (e.g., paired watershed study) while research at broad spatial scales would include relatively low sampling densities (e.g., cross sections; Marcus & Fonstad, 2008). Newer technologies utilize remote sensing to measure riverscape features, and expanded the ability to sample more broad spatial extents with high-resolution and greater sampling densities (Lane & Chandler, 2003; Mertes, 2002). The technological advance provided increased automation and resulted in development of continuous watershed maps at 1 m or better resolution of in stream habitat (Bergeron & Carbonneau, 2012). This improvement to data collection was accompanied by decreased cost, collection, processing, and analysis times paving the way for researchers to assess both the continuous and discontinuous structure of rivers with greater power (Carbonneau et al., 2012).

A few of these tools include LiDAR (Williams, Olsen, Roe, & Glennie, 2013), thermal infrared (TIR) remote sensing (Dugdale, 2016), passive optical remote sensing or ultra-high-resolution optical imagery (Legleiter, Roberts, Marcus, & Fonstad, 2004), and Doppler current profiling (ADCP; Parsons et al., 2013; see Carbonneau et al., 2012 and references therein). The mobility of ADCP technologies (termed moving vessel measurements) propelled data collection forward to include sediment transport, bed shear stress and soon will expand into other fluvial metrics. Using LiDAR (e.g., near infrared or green wavelengths) researchers have acquired image resolutions of $10^{-2}$–$10^0$ m across study extents of 10’s of meters to 100’s of kilometers (Marcus & Fonstad, 2010 and references therein). Research with ADCP has gained traction for use in
rivers because of the rapid real time predictions of velocity and discharge measurements that closely match traditional stream gauge data. Deployment of sensors has also experienced an increase in performance ability and is aiding automation of data collection. The use of unmanned aerial vehicle (UAV) platforms increased accessibility to isolated habitat giving researchers an alternative to conventional platforms (i.e., manned areal surveys) for remote sensing. UAVs are able to carry small payloads (e.g., LiDAR, go-pros, or TIR sensors); therefore, they offer flexible data collection opportunities. Analysis conducted at the resolutions and extents enabled by these tools described above have provided researchers with the means to assess river biotypes (e.g., pools, riffles, etc.), map rivers of shallow depths (Pan et al., 2015), improve mapping of braided or turbulent waters, and calculate primary fluvial variables (Carbonneau, Bergeron, & Lane, 2005; Hugue et al., 2016; Petrie, Diplas, Gutierrez, & Nam, 2013) (i.e., mean flow velocity, water depth, grain size, stream width, etc.).

6 | DISCUSSION

This review focuses on the utility of RG to address ecological questions, and explores methodology to accomplish such a task. Although the literature demonstrates that a variety of studies have used methods that can be considered RG to test hypotheses, several fundamental concepts need to be addressed before structural frameworks or “best methodological practices” can be developed (Table 2). LG studies have used simulation modeling to address many of these questions, which lead to the development, and design of conceptual frameworks for conducting robust LG studies (Hall & Beissinger, 2014; Landguth, Fedy, et al., 2012). This work has not begun for RG and therefore many methodological questions about the effects of sample design on the detection of spatial genetic variation still remain.

The challenge facing growth in RG is development of methods that adequately describe the contribution of physical and geomorphological processes in shaping genetic variation. LG methods may not be easily transferrable into RG because of limitations related to the branched physical structure of riverscapes that form constrained pathways for dispersal and migration, directional movement of water, and sparser availability of continuous spatial data for rivers, but the concept of effective distance offers opportunities for application in riverscapes. In dynamic seascapes with complex directional movement of water, researchers have incorporated simulations of larval dispersal using oceanographic processes and then applied network theory to test alternate models of connectivity. Additionally, the ability to detect the effects of riverscape resistance on dispersal is also contingent on directional movement of water (e.g., flow connected and flow unconnected habitat; Peterson et al., 2013) and the asymmetrical effects of resistance that may result for aquatic organisms with varying motilities.

We outlined several thematic areas of research that would benefit from use of RG approach throughout this overview and highlighted several specific ecological questions pertaining to many of those themes in Table 2. For example, how does temporal variation in riverscape heterogeneity (e.g., stream flow velocity, temperature, etc.) influence dispersal capability of individuals and how will habitat restoration (e.g., dam removal, large wood, stream temperature regulation) does temporal variation in riverscape heterogeneity (e.g., stream flow velocity, temperature, etc.) influence dispersal capability of individuals or populations?

| Category            | Relevant questions                                                                 |
|---------------------|-------------------------------------------------------------------------------------|
| Methodological      | What is the effect of field sampling design (discrete vs. continuous sampling strategies), genetic index (i.e., FST, Gst, Dst), number of samples, loci number (including molecular marker type) on the ability to detect the effects of riverscape features on gene flow? |
|                     | How do Mantel, partial Mantel, MRDM tests or other analytical methods perform when detecting effects of riverscape features on gene flow? |
|                     | How do resistance surfaces perform as a tool to interpret effects of riverscape features on genetic structure in watersheds with braided or lattice topographies? |
|                     | What model(s) best captures the effects of asymmetric riverscape features (e.g., complete vs. incomplete barriers) on functional connectivity? |
| Ecological          | What are the effects of effective distance on gene flow as measured by neutral genetic variation and genetic variation resulting from effects of selective pressure (adaptive)? |
|                     | What habitat patches (nodes) or effective distances are central to the maintenance of long-term gene flow/dispersal? |
|                     | Under climate change scenarios? |
|                     | How does temporal variation in stream flow velocity influence movement or dispersal capability of individuals or populations? |
|                     | How does habitat restoration (e.g., dam removal, large wood, stream temperature regulation) influence population and genetic variation? |
|                     | At what spatial scale(s) does riverscape resistance (IBR) better explain genetic structure over geographic distance (IBD) or permanent dispersal barriers (IBB)? |
|                     | To what degree have historical legacies (e.g., sea level, ice age, flooding, landslides) impacted contemporary genetic structure? |
|                     | At what scales do specific riverscape variables serve as functionally important indicators of dispersal resistance for species of varying mobilities? |

FST = fixation index; IBB = isolation by barrier; IBD = isolation by distance; IBR = isolation by resistance; MRDM = multiple-regression on distance matrices.
influence population persistence are examples of pressing questions that face local and regional managers. Data collection at either site-specific or broad spatial scales currently informs policy decisions; however, answers to these questions would be more informed by continuous fine-scale sampling across the watershed to capture the fine-scale variation that is present within dynamic riverscapes. In the future, increased automation for collection of continuous habitat data will continue to improve spatial maps and molecular datasets will become increasingly cost-effective to generate making this scale of spatial data easier to acquire. Therefore, our ability to sample the watershed at scales that are needed to address these ecological questions from a riverscape perspective will be possible. In the future development of methods (e.g., effective distance network) within RG will enhance or ability to incorporate datasets from multiple fields (e.g., genetic data and ecological data) and test for relationships that exist between organisms and their environment at relevant spatial scales.

ACKNOWLEDGMENTS
This overview was supported by the Coastal Oregon Marine Experimental Station (COMES) project no. ASF501, the Living Marine Resource Cooperative Science Center (LMRSCSC) and NOAA Educational Partnership Program with Minority Serving Institutions Graduate Research and Training Scholarship Program. We thank Kathryn Ronnenberg for the creation of figures used throughout the manuscript. We are also grateful to the two reviewers who provided insightful comments that helped to improve the manuscript.

CONFLICT OF INTEREST
The author has declared no conflicts of interest for this article.

RELATED WIREs ARTICLES
Applications of spatial statistical network models to stream data

ORCID
Chanté D. Davis http://orcid.org/0000-0001-9869-7819

REFERENCES
Antao, T., Lopes, A., Lopes, R. J., Reja-Pereira, A., & Luikart, G. (2008). LOSITAN: A workbench to detect molecular adaptation based on a Fst-outlier method. BMC Bioinformatics, 9, 323. https://doi.org/10.1186/1471-2105-9-323
Avise, J. C. (2000). Phylogeography: The history and formation of species. Cambridge, MA: Harvard University Press.
Balkenhol, N., Gugerli, F., Cushman, S. A., Waits, L. P., Coulon, A., Arntzen, J. W. H., … Wagner, H. H. (2009). Identify future research needs in landscape genetics: Where to from here? Landscape Ecology, 24, 8.
Balkenhol, N., Waits, L. P., & Dezzani, R. J. (2009). Statistical approaches in landscape genetics: An evaluation of methods for linking landscape and genetic data. Ecology, 90(5), 818–830. https://doi.org/10.1111/j.1600-0587.2009.05807.x
Banks, S. C., Cary, G. J., Smith, A. L., Davies, I. D., Driscoll, D. A., Gill, A. M., … Peakall, R. (2013). How does ecological disturbance influence genetic diversity? Trends in Ecology & Evolution, 28(11), 670–679. https://doi.org/10.1016/j.tree.2013.08.005
Banks, S. C., Pigott, M. P., Williamson, J. E., Bové, U., Holbrook, N. J., & Beheregaray, L. B. (2007). Ocean variability and coastal topography shape genetic structure in a long-dispersing sea urchin. Ecology, 88(12), 3055–3064.
Benda, L., Poff, L., Miller, D., Dunne, T., Reeves, G., Pess, G., & Pollock, M. (2004). The network dynamics hypothesis: How channel networks structure riverine habitats. Bioscience, 54(5), 413–427.
Bergeron, N. E., & Carbonneau, P. E. (2012). Geosalar: Innovative remote sensing methods for spatially continuous mapping of fluvial habitat at riverscape scale. In P. E. Carbonneau & H. Piégay (Eds.), Fluvial remote sensing for science and management, Chichester, England: John Wiley & Sons.
Bowley, H. D., Fleming, I. A., & Gibson, A. J. F. (2016). Applying landscape genetics to evaluate threats affecting endangered Atlantic salmon populations. Conservation Genetics, 17(4), 823–838. https://doi.org/10.1007/s10592-016-0824-7
Brauer, C. J., Hammer, M. P., & Beheregaray, L. B. (2016). Riverscape genomics of a threatened fish across a hydroclimatically heterogeneous river basin. Molecular Ecology, 25, 5093–5113. https://doi.org/10.1111/mec.13830
Carbonneau, P., Fonstad, M. A., Marcus, W. A., & Dagdale, S. J. (2012). Making riverscapes real. Geomorphology, 137(1), 74–86. https://doi.org/10.1016/j.geomorph.2010.09.030
Carbonneau, P. E., Bergeron, N., & Lane, S. N. (2005). Automated grain size measurements from airborne remote sensing for long profile measurements of fluvial grain sizes. Water Resources Research, 41(11), W11426. https://doi.org/10.1029/2005WR003994
Castillo, J. A., Epps, C. W., Davis, A. R., & Cushman, S. A. (2014). Landscape effects on gene flow for a climate-sensitive montane species, the American pika. Molecular Ecology, 23(4), 843–856. https://doi.org/10.1111/mec.12650
Castric, V., Bonney, F., & Bernatchez, L. (2001). Landscape structure and hierarchical genetic diversity in the brook charr, Salvelinus fontinalis. Evolution, 55(5), 1016–1028. https://doi.org/10.1111/j.0016-805X.2001.tb09618.x
Chambers, G. K., & MacAvoy, E. S. (2000). Microsatellites: Consensus and controversy. Comparative Biochemistry and Physiology Part B Biochemistry and Molecular Biology, 126, 21.
