

**Title**

Aquatic Landscape Genomics and Environmental Effects on Genetic Variation

**Author's names and affiliations**

Grummer, Jared A.<sup>\*1</sup>, Beheregaray, Luciano B.<sup>2</sup>, Bernatchez, Louis<sup>3</sup>, Hand, Brian K.<sup>4</sup>, Luikart, Gordon<sup>5</sup>, Narum, Shawn R.<sup>6,7</sup>, Taylor, Eric B.<sup>8</sup>

<sup>1</sup>University of British Columbia

Department Zoology, Biodiversity Research Centre and Beaty Biodiversity Museum

6270 University Blvd.

Vancouver, BC, CANADA V6T 1Z4

grummer@zoology.ubc.ca

707-888-1176

604-822-2416 (Fax)

<sup>2</sup>Flinders University

Molecular Ecology Laboratory, College of Science and Engineering

Adelaide, SA, 5001 Australia

luciano.beheregaray@flinders.edu.au

61(8) 82015243

61(8) 82013015 (Fax)

<sup>3</sup>IBIS (Institut de Biologie Intégrative et des Systèmes), Université Laval

1030 Ave de la Médecine, Québec QC G1V 0A6 Canada

louis.bernatchez@bio.ulaval.ca

418 656-3402

418 656-7176 (Fax)

@LouBernatchez

<sup>4</sup>Flathead Lake Biological Station, Division of Biological Sciences, University of Montana

32125 Bio Station Ln, Polson, MT 59860 USA

brian.hand@umontana.edu

(406) 214-9708

(406) 982-3302 (Fax)

<sup>5</sup>Flathead Lake Biological Station, Division of Biological Sciences, University of Montana

32125 Bio Station Ln, Polson, MT 59860 USA

Gordon.luikart@umontana.edu

406-982-3301

406-982-3201 (Fax)

<sup>6</sup>Columbia River Inter-Tribal Fish Commission

3059F National Fish Hatchery Road

Hagerman, Idaho 83332 USA

nars@critfc.org  
(208) 837-9096  
(208) 837-6047 (Fax)

<sup>7</sup>University of Idaho  
Department of Fish and Wildlife Sciences  
975 W. 6th Street  
Moscow, Idaho, 83844 USA

<sup>8</sup>University of British Columbia  
Department Zoology, Biodiversity Research Centre and Beaty Biodiversity Museum  
6270 University Blvd.  
Vancouver, BC, CANADA V6T 1Z4  
etaylor@zoology.ubc.ca  
604-822-9152  
604-822-2416 (Fax)

\*Corresponding author: Grummer, J.A. ([grummer@zoology.ubc.ca](mailto:grummer@zoology.ubc.ca))

# **Keywords**

Riverscape genetics, seascape, gene flow, adaptive capacity, fragmentation, conservation

## 68 **Abstract**

69 Aquatic species represent a vast diversity of metazoans, provide humans with the most  
70 abundant animal protein source, and are of increasing conservation concern, yet landscape  
71 genomics is dominated by research in terrestrial systems. We provide researchers with a  
72 roadmap to plan aquatic landscape genomics projects by aggregating spatial and software  
73 resources and offering recommendations from sampling to data production and analyses, while  
74 cautioning against analytical pitfalls. In consideration of water's unique properties, we discuss  
75 the importance of considering freshwater system structure and marine abiotic properties when  
76 assessing genetic diversity, population connectivity and signals of natural selection. When  
77 possible, genomic datasets should be parsed into neutral, adaptive, and sex-linked datasets to  
78 generate the most accurate inferences of eco-evolutionary processes.

## 79 **Landscape genomics and aquatic organisms**

80 Aquatic species and their ecosystems play fundamental roles in sustaining global  
81 biodiversity and human populations [1]. Marine and freshwater ecosystems alike face  
82 numerous environmental challenges [2], which is alarming considering that they harbour a  
83 tremendous amount of described metazoan fauna. Environmental stressors are the greatest  
84 threat to freshwater habitats, which have caused a 83% decline in species abundances since  
85 1970 [3]. Many marine fisheries are overexploited and on the brink of collapse [4]. Yet, little is  
86 known about how environmental changes are impacting the health and evolutionary potential of  
87 aquatic species, and under what conditions adaptation may occur. **To address these needs,**  
88 **landscape genomics provides a powerful framework for understanding eco-evolutionary**  
89 **processes, assessing the viability of populations, and predicting the future health of species**  
90 **and aquatic ecosystems.**

91 Landscape genetics emerged as a formal discipline 15 years ago as a powerful means  
92 to address problems of understanding how the interaction between ecological, evolutionary,  
93 and geographic factors influence population genetic structure (Box 1; [5]). More recently, the

94 development of high throughput genomic tools [6] made it possible to move from landscape  
95 genetics to landscape genomics - whereby genetic variation can be screened at the scale of  
96 the entire genome - offering greater power to disentangle adaptive from neutral genetic  
97 divergence and identify environmental factors acting as selective agents [7].

98 We define landscape genomics as “The use of genomic technologies to study genome-  
99 wide neutral and adaptive variation of ecologically diverse populations across heterogeneous  
100 landscapes to address novel or previously intractable questions”, such as forecasting of  
101 adaptive capacity under environmental change [8]. Quite clearly, landscape genetic/genomic  
102 studies to-date have been biased towards terrestrial ecosystems (Fig. 1; [9]); of all landscape  
103 genetics papers published since 1991, only 13% were on aquatic systems (9% on freshwater  
104 and 4% in marine systems). This is partly because genomic resources are lacking for most  
105 aquatic species (see [10] for a marine-terrestrial comparison). Substantial differences exist  
106 between terrestrial and aquatic systems (see “Waterscape characteristics” below), questioning  
107 the translatability of terrestrial landscape genomics approaches to aquatic systems.

#### 108 **Waterscape characteristics**

109 Aquatic and terrestrial systems differ in fundamental ways relevant to landscape  
110 genomics. Because water is often flowing with some current, most aquatic organisms need to  
111 spend more energy to stay in place than move. Marine and freshwater systems have many  
112 divergent properties, including differences in patterns of biodiversity, suggesting that processes  
113 generating biodiversity, and potentially tractable through landscape genomics, may differ  
114 between these realms. For instance, ~40% of all named fish taxa are found in freshwater, yet  
115 the total freshwater surface area is miniscule compared to the marine realm (2% vs. 68%,  
116 respectively) [2].

117 Water’s physical properties have created an environment that uniquely affects aquatic  
118 organisms and their eco-evolutionary dynamics. Water is ~800x more dense than air and at  
119 least 40 times as viscous, but provides much greater buoyancy. Water also has a higher

thermal capacity (ability to maintain temperature) and conductivity (ability to transfer heat) than air. Oxygen solubility is inversely related to water temperature, with hypoxic conditions occurring for many aquatic organisms that experience warm temperatures [11], thus the coupling of temperature and oxygen has likely driven adaptations in aquatic ectotherms [12, 13]. Aquatic environments also present particular physiological challenges for **diadromous** (see Glossary) species that move between marine (hypertonic) and freshwater (hypotonic) environments (e.g., [14]).

Aquatic landscapes contain tremendous variation in habitat complexity and physical connectivity that distinguish them from terrestrial habitats. Both marine and freshwater environments are highly dynamic with diel fluctuations in tides and currents in marine systems, or variation in daily discharge, water depth, and temperature in fresh water. Both aquatic environments also see seasonal fluctuations including upwelling in marine environments and flow rates in freshwater systems. In contrast to marine habitats, freshwater habitats are hierarchically organized by relative elevation and connected via headwater streams, reaches, and watersheds. Due to the **dendritic** nature of riverine systems, abiotic characteristics such as river branching extent and confluence position affect genetic variation and population structure (e.g., [15,16]). Further, because predominant river currents are unidirectional, **migration** is expected to be asymmetric. In contrast, marine environments contain discrete yet connected habitat types such as the **pelagic** environment, near-shore (e.g., coral reefs, seagrasses), and estuaries (Fig. 2). Terrestrial habitats, by contrast, are generally characterized by larger diel and annual fluctuations in temperature, particularly in polar and temperate regions, and are typically more structurally complex with steeper climatic gradients.

Due to the connected nature of aquatic systems, many aquatic organisms can encounter a broad range of habitats over their lifetime. For instance, reproductively mature adults of many species occupy dynamic intertidal and rocky nearshore habitats where temperature and solar radiation go through diel fluctuations, whereas their larval forms are

often found in the more homogeneous pelagic zone (e.g., giant green anemone [*Anthopleura xanthogrammica*]). Furthermore, in fresh water, some species may be **adfluvial** where juveniles born in streams move to lakes to mature before returning to streams as adults for spawning (e.g., bull trout [*Salvelinus confluentus*]).

## **Landscape connectivity and gene flow**

Although aquatic systems have often been overlooked in favor of terrestrial systems for developing genetic **connectivity model** theory (e.g., least-cost path, circuitscape, etc.; Box 2), they provide a range of conditions and challenges to test methods and models [17-20].

Because of water's physical properties, dispersal energetics are distinct in aquatic versus terrestrial environments. Consequently, aquatic organisms have evolved a myriad of behavioural, morphological, and life history traits that impact connectivity [10]. In freshwater systems, streams and rivers can often be represented in a one-dimensional **cost surface**.

Conversely, marine environments often provide the ultimate challenge in connectivity modeling because of weak population structuring ( $F_{ST}$  is typically less than 0.01), a vast three-dimensional environment where species interactions and survival are not well understood, and partially/poorly defined barriers [21].

### **Measuring genetic connectivity in aquatic systems**

In freshwater systems, migration can be easier to measure than in terrestrial ones, specifically in streams and rivers whose **dendritic structure typically results in well-defined migratory paths**. "Link-based" approaches are often applicable to these systems by using a least-cost path [17] or "stream tree" approach where pairwise genetic distances ( $F_{ST}$ ) are fit to non-overlapping stream segments to assess population connectivity (Box 2; [19]). A challenge that remains, however, is **determining "resistance" along migratory paths** ([22], but see [20]).

Delineating discrete populations and connectivity among them pose distinct challenges in marine systems. A knowledge of physiological requirements such as thermal and osmoregulatory tolerance can be supplemented with predominant oceanic currents to generate

Lagrangian dispersal models to inform circuitscape models [23]. In these cases, circuitscape methods can be applied to model gene flow, or an **isolation-by-environment (IBE)** type approach might also be suitable [24]. Along these lines, Xuereb et al. [25] determined that genetic connectivity of two populations of the giant California sea cucumber (*Apostichopus californicus*) was driven by local asymmetric currents as opposed to distance alone (**isolation-by-distance, IBD**). Duranton et al. [26] recently used haplotype length information in European sea bass (*Dicentrarchus labrax*) to estimate timing, directionality, and amount of gene flow. Finally, a landscape community genomics approach may help elucidate ecological and evolutionary processes important in structuring populations in particularly challenging systems ([27]; Box 3).

#### *Defining discrete populations and identifying barriers to gene flow in marine species*

Most barriers in the marine realm are porous or represent spatial clines (e.g., thermal or haline gradients, currents). Marine species are often assumed to have **panmictic population structure** (random mating resulting in high gene flow) due to the lack of potential barriers to **movement**. Recent studies, however, have demonstrated that high dispersal ability does not always mean that spatial genetic structure is unresolvable. Indeed, cryptic population structure exists within multiple marine species and is driven by environmental clines [28]. For instance, Benestan et al. [29] used a seascape genomics framework that allowed quantifying the relative importance of spatial distribution, ocean currents and sea temperature on connectivity among American lobster (*Homarus americanus*) populations.

#### *Measuring population structure at neutral, adaptive, and sex-linked loci*

The **increased resolution of genomic data** allows investigation of **functionally distinct groups such as neutral, adaptive, and sex-linked** (in genetically determined sex systems) loci. However, identifying sex-linked markers is difficult for many aquatic species because they are not sexually dimorphic and/or lack the genomic resources to do so. When possible, it is important to organize genomic data in this way because the relative strengths of mutation,

migration, selection, and drift differ among these groups [30], which may lead to misleading patterns if analysed in aggregate. For instance, Benestan et al. [31] showed that relatively few sex-linked markers (12 and 94, respectively), rather than genome-wide drift and gene flow, were driving genetic structure in both American lobster and Arctic char (*Salvelinus alpinus*). Similarly, adaptive markers associated with phenotype or particular environmental variables under selection often show a different pattern than neutral loci. In redband trout (*Oncorhynchus mykiss gairdneri*), Chen et al. [12] demonstrated that 5,890 neutral loci revealed genetic differentiation as expected under IBD, whereas 13 outlier loci associated with cardiac and physiological function differentiated desert from montane populations irrespective of geographic distance.

Important advances of understanding gene flow and landscape connectivity could be made within the explicit incorporation of candidate adaptive markers into a landscape resistance modeling framework (e.g., [16]). The addition of adaptive gene flow into connectivity modeling theory could improve understanding of adaptive capacity, as influenced by movement of adaptive alleles among populations [32], or by environmental variables driving selection along a migratory path [20]. Despite being computationally less challenging than terrestrial environments, freshwater systems have not been fully explored for theory purposes and in development of genetic connectivity models (but see [33]). For example, the influence of population topology (the spatial arrangement of populations throughout a landscape) on gene flow and population connectivity is often neglected in fresh waters, but could improve this type of research in terrestrial systems [34].

### **Genome scans and association studies for detecting local adaptation**

Recent advances in sequencing technology, computational approaches, and genomic resources have enabled high density genome scans to detect local adaptation, as well as Genotype-Environment-Associations (GEA) in natural populations [6,35]. In aquatic species, studies have discovered the genetic basis for specific phenotypic traits [36], broad signals of



224 local adaptation across landscapes [12,16,29], and candidate genes for conservation  
225 monitoring [37]. Genome scans and GEA tests have become routine and offer immense  
226 potential to investigate adaptive variation [38].

227         Researchers can now address critical questions related to evolutionary adaptation and  
228 resilience in aquatic ecosystems (e.g., [39]). Yet, study design for genome scans and GEA  
229 tests in aquatic systems requires careful consideration of many factors, some of which are  
230 distinct in marine vs. freshwater systems. These include (1) sampling strategies, (2) candidate  
231 environmental variables, (3) marker density across the genome, and (4) statistical approaches  
232 to detect drivers of selection, the type and strength of selection, and candidate genes involved.  
233 We focus on genome scans in an **association mapping** framework because non-model  
234 aquatic organisms are often not well-suited for **QTL mapping**, salmonid fishes being the  
235 exception rather than the rule [40].

236         Sampling strategy to adequately represent organisms across time and space (and to  
237 achieve statistical power) is a crucial component for both marine and freshwater landscape  
238 genomics studies, with temporal and spatial replicates needed to rigorously test the stability of  
239 selection signals [27,41]. In complex marine systems, additional layers of spatial dimensions  
240 must be considered [9,21]. For instance, many species are often broadly distributed across  
241 porous dispersal barriers, but population connectivity in the sea can be influenced by climatic  
242 gradients [28], spatially and temporally variable recruitment associated to dynamic local  
243 oceanography [42], and multifarious ecological requirements of adults that utilize various  
244 niches across daily or seasonal timeframes [43]. Freshwater species show more limited  
245 dispersal, but often occupy different components of habitat based on temporal cycles and  
246 resource availability [44].

247         **Anadromous** or **catadromous** species that migrate between freshwater and marine  
248 environments are exposed to a broad range of conditions throughout their life cycle that may  
249 require additional sampling considerations to resolve adaptive variation related to each

environment (e.g., [20]). Sampling at different life stages (e.g., larva vs adult) is crucial to confirm whether signals of selection reflect long-term local adaptation among genetically distinct populations (e.g., divergent selection), or short-term selection within the lifespan of individuals in a panmictic population representing spatially varying balancing selection [45].

Additional sampling considerations include sex ratio of collections (when sex can be identified either phenotypically or genetically) because sex-linked variation could be falsely interpreted as [31], and consideration of specific phenotypes within populations that may be controlled by genes of major effect [46]. Finally, detailed phenotyping (phenomics) may provide insight into specific morphology, behaviour, and development related to adaptive ecological processes [40].

A second factor to be considered relates to the choice of candidate environmental variables. Natural history provides the best source of information for developing *a priori* hypotheses about which variables might be ecologically relevant for the study species. Considerations about how environmental heterogeneity impacts habitat composition and structural and functional connectivity are nonetheless difficult to make ([47]; see Box 2). This is particularly true in marine systems because of their asymmetric physical flows and dynamics, inherent non-stationarity, and size of habitats [21]. Landscape mapping that maximizes environmental variance is comparatively easier in fresh water than marine systems, where a large number of observational, modelled and remotely sensed variables have recently become available for various scales [33]. Genome scans and GEA tests are bound to benefit from the increase in resolution and extent of spatial resources (examples in Table II) driven by pressing human needs, such as fresh water availability for consumption and irrigation, fisheries resources through biophysical modelling, and tracking plastic in our seas through customisable simulations. These developments are expected to extend our options beyond the traditional candidate variables (e.g., temperature, salinity, rainfall) and towards environmental mapping

capable of informing on natural and anthropogenic disturbances, resource availability, range shifts, and biotic interactions.

Adjusting the density of markers to the research question, particularly in relation to linkage disequilibrium (LD), is a third important aspect when planning genome scan or GEA studies, with specific considerations for freshwater and marine species alike that often have limited genomic resources. It is ill-advised to draw strong inferences regarding candidate adaptive loci in cases where marker density is low and LD is high. As a reference point, LD estimates in wild fish populations have been reported from ~1kb in zebrafish (*Danio rerio*) and threespine stickleback (*Gasterosteus aculeatus*) to ~10-20kb in the European eel (*Anguilla anguilla*) and up to 1 MB in lake whitefish (*Coregonus clupeaformis*) [48-50]. For species with large effective population sizes, as is the case with many marine taxa, recombination may cause rapid linkage decay requiring high marker density to provide multiple SNPs per linkage block to achieve sufficient power for detecting candidate adaptive genes [51]. In systems where LD is high, such as small, isolated freshwater populations, lower density markers may be adequate to detect signals of adaptive variation, especially in inverted regions with extended LD [52,53].

In either case, a quality reference genome assembly that is well annotated is a powerful resource to characterize the genomic architecture of adaptation that includes identification of candidate genes, genomic position, and putative biological function (Supplemental Material; [38]). Although many aquatic species lack genomic resources, community efforts aimed at developing reference genomes across many taxa are expected to lead to tremendous improvements (e.g., Earth BioGenome Project seeks to sequence ~1.5 million eukaryotic species [54]). Researchers can capitalize on these resources while also seeking to enhance them by contributing data to improve genome assemblies for target species (e.g., linkage maps, Hi-C libraries, optical maps).

A fourth consideration for genome scans and GEA tests is choice of statistical analyses well-suited to address the study question and intricacies of aquatic systems. Genome scans are susceptible to detection of false positive signals of adaptation, particularly in freshwater species comprising small, isolated populations prone to pronounced drift [55]. On the other end, detecting local adaptation and genomic outliers can be a challenge in marine species with large and well connected populations. Fortunately, several studies have provided guidance to balance **Type I and II errors** [35,41]. Statistical analyses that combine multiple approaches such as outlier tests, genome-wide association mapping, transcriptomics, and GEA offer corroborating evidence for local adaptation in aquatic systems [12]. Significance testing that accounts for multiple SNPs in linkage disequilibrium provides stronger evidence than single marker tests, as does multivariate testing for polygenic effects [40]. Recent simulations suggest that multivariate GEA methods such as redundancy analysis (RDA) provide the best balance of low false positive and high true-positive rates across a range of demographic histories, sampling designs, sample sizes, and selection levels [35]. Current statistical models used for association mapping typically correct for population structure, but this may come with the caveat of reducing power to detect candidate loci if selection gradients follow the same direction as neutral structure [7]. Background selection combined with genetic hitchhiking can also generate correlation between local recombination rates and genetic diversity which could falsely be interpreted as signal of divergent selection between populations [56].

#### **Adaptive capacity, conservation and management of wild populations**

Landscape genomics may advance conservation management and recovery of threatened and exploited populations by helping to understand their **adaptive capacity** to evolve under environmental change. Under climate change, ectothermic species face particular stresses to their preferred thermal niches, highlighting the importance of predicting adaptive capacities of aquatic populations [57]. Any intrinsic or extrinsic factors that will affect the strength of the four evolutionary forces can influence adaptive capacity. These include mutation

rate and generation time, species life-history, amount and architecture of genetic variation, effective population size and thus genetic drift, biotic and abiotic factors impacting the strength and mode of selection, and gene flow from ecologically distinct populations.

Using landscape genomic analyses to identify genotype-environment associations is an obvious first step for assessing selection in wild populations and integrating adaptive capacity into predictive models of vulnerability to environmental change [11,58]. On one end of the spectrum, landscape genomics can help assess adaptive potential of declining populations known to have persisted in variable and often degraded habitats, a topic of increasing importance and debate [45]. For example, in a range-wide study of a poorly dispersing and endangered fish, GEA tests that consider the effects of dendritic riverine structure recovered signals of adaptive diversity associated with a hydroclimatic gradient and human impacts [59]. The possibility that these small populations are responding to selection was further supported through comparative ecological transcriptomics [60]. Yet, other studies have suggested limited adaptation in small, geographically isolated populations that experience high inbreeding [55].

On the other end of the spectrum, landscape genomics can assess the influence of environmental heterogeneity and disturbance on local adaptation in abundant and exploited species with high gene flow. Such studies indicated that heterogeneous environments may drive and maintain adaptive divergence among connected populations of marine [13,29], anadromous [20], and freshwater species [61]. These species may have the potential for tracking future environments because their individuals are capable of rapidly spreading alleles that affect fitness over vast distances.

The adaptive potential of a population is likely related to its “genomic vulnerability”, a metric defined within a landscape genomics framework as the “mismatch between current and predicted genomic variation based on genotype-environment correlations modelled upon contemporary populations” [62]. **Environment-associated SNPs** can also be used to predict the putative environmental range for individuals with known genotypes [63]. This approach can

help predict genetically-mediated environmental limits across taxa and compare environmental ranges among multiple species over the same landscape [63]. Landscape genomics can also predict the spatio-temporal spread of adaptive alleles and resistance to spread of mal-adaptive alleles across space [32]. Frameworks for evidence-based genetic management decisions and policies exist (e.g., [64,65]), and in spite of the challenges associated with their implementation, genomic data have been used in many conservation-based decisions (see [66] for examples). When fueled with information about adaptive capacity, these frameworks should improve management plans targeting (1) the recovery of exploited populations, (2) *in situ* and *ex situ* efforts of **evolutionary rescue** (e.g., captive breeding, translocations, reintroductions) and (3) the anticipated re-design of climate-ready populations.

#### **Concluding remarks and a roadmap for future research**

The lack of aquatic landscape genomics studies compared to the number of terrestrial landscape genomic studies is surprising (Fig. 1). This is partly because genomic resources are lacking for aquatic species [10]. Therefore, a pressing need exists to develop resources to improve aquatic landscape genomics studies (e.g. reference genomes, transcriptomes, sex-linked markers, large SNP catalogs). Freely available environmental databases are increasing for both marine and freshwater ecosystems, as well as geospatial tools and computer programs that help meet the particular challenges that aquatic landscape genomics studies face (Table II).

Such challenges include more rigorously defining population structure and quantifying genetic and demographic connectivity in the marine realm, and gaining an understanding of landscape genomic patterns of species from non-developed regions (see Outstanding Questions). Along these lines, the inter-annual variability of abiotic conditions in many aquatic systems and their population-level effects, particularly in the marine realm, must be recognized; although field sampling is admittedly difficult, future studies would benefit from temporal replicates for understanding landscape genetic processes. Another major challenge is that

strong inferences about GEAs may be constrained by false positives [67]. Arguably, associations remain an indirect demonstration of an actual functional relationship under the influence of natural selection. Consequently, future studies should rigorously test hypotheses derived from GEAs via gene functional analyses (e.g., comparative physiological studies), and perform experimental tests of natural selection [12].

Landscape genomics modeling can help predict population viability by facilitating modeling of complex interactions between biotic and abiotic factors that influence individual vital rates and control population distribution, abundance, growth rates, and species interactions (e.g., Box 3; [27,32]). Yet, this has been rarely applied in aquatic ecosystems. Consequently, another potentially fruitful research area would be to apply recently developed landscape genomics and meta-models to test models' reliability in forecasting changes of population sizes and community composition [32,68]. Aquatic landscape genomics research should also increasingly consider the role of differential gene expression and epigenetic inheritance as a source for rapid adaptation [69,70], for instance in the face of new stressors [12,60]. Similarly, recent genomics studies are revealing the important role of structural genetic variants (SGV) in eco-evolutionary processes [53]. Catanach et al. [71] recently showed that in the Australasian snapper (*Chrysophrys auratus*), the number of base pairs affected by SGV variants was almost three times higher compared to other SNPs, with a sizeable portion of these located in regions under putative selection.

In summary, although further work is needed to improve a quantitative and predictive theory of the genetic basis of adaptation and to validate recent approaches, knowledge derived from landscape genomics studies already provides a foundation to address real-world problems in conservation management of aquatic biodiversity.

## **Acknowledgments**

This work was supported in part by Genome Canada and Genome British Columbia (project code 242RTE). BKH and GL were supported in part by funds provided by National Science Foundation grant DEB-1639014 and NASA grant NNX14AB84G, and we thank the Australian Research Council for a Future Fellowship (FT130101068) to LBB. We also thank C. Brauer and M. Whitlock for providing comments on an earlier version of the manuscript, and F. Allendorf for sending helpful publications and ideas on the origins of landscape genetic approaches.

## **Author Contributions**

SRN conceived the study, JAG organized and led the writing, and all authors contributed equally to writing and editing the text. After JAG, authors are listed alphabetically to reflect equal contributions.



## REFERENCES

1. Béné, C. *et al.* (2016) Contribution of fisheries and aquaculture to food security and poverty reduction: assessing the current evidence. *World Dev.* 79, 177-196.
2. Dudgeon, D. *et al.* (2006) Freshwater biodiversity: importance, threats, status and conservation challenges. *Biol. Rev.* 81, 163–182.
3. World Wildlife Federation. (2018) Living Planet Report - 2018: Aiming Higher. Grooten, M. and Almond, R.E.A. (Eds). WWF, Gland, Switzerland, 144p.
4. Pauly, D. and Zeller, D. (2016) Catch reconstructions reveal that global marine fisheries catches are higher than reported and declining. *Nature Comm.* 7, 10244.
5. Manel, S. *et al.* (2003) Landscape genetics: combining landscape ecology and population genetics. *Trends Ecol. Evol.* 18, 189-197.
6. Fuentes-Pardo, A.P. and Ruzzante, D.E. (2017) Whole-genome sequencing approaches for conservation biology: Advantages, limitations and practical recommendations. *Mol. Ecol.* 26, 5369-5406.
7. Yeaman, S. *et al.* (2016) Convergent local adaptation to climate in distantly related conifers. *Science* 353, 1431-1433.
8. Schwartz, M.K. *et al.* (2010) Landscape genomics: a brief perspective, In *Spatial complexity, informatics, and wildlife conservation*. Cushman, S.A., and F. Huettmann (eds.), pp. 165-174, Springer.
9. Selkoe, K.A. *et al.* (2015) Waterscape genetics—applications of landscape genetics to rivers, lakes, and seas. *Landscape Genetics: Concepts, Methods, Applications*, pp. 220-246, John Wiley & Sons Ltd.
10. Kelley, J.L. *et al.* (2016) The life aquatic: advances in marine vertebrate genomics. *Nature Rev. Gen.* 17, 523.
11. Harrison, J.F. *et al.* (2018) Functional Hypoxia in Insects: Definition, Assessment, and Consequences for Physiology, Ecology, and Evolution. *Ann. Rev. Ent.* 2018, 303-325.

12. Chen, Z. *et al.* (2018) Mechanisms of thermal adaptation and evolutionary potential in conspecific populations to changing environments. *Mol. Ecol.* 27, 659-674.
13. Sandoval-Castillo, J. *et al.* (2018) Seascape genomics reveals adaptive divergence in a connected and commercially important mollusc, the greenlip abalone (*Haliotis laevis*), along a longitudinal environmental gradient. *Mol. Ecol.* 27, 1603-1620.
14. Whitehead, A. *et al.* (2011) Genomic mechanisms of evolved physiological plasticity in killifish distributed along an environmental salinity gradient. *Proc. Nat. Acad. Sci.* 108, 6193-6198.
15. Thomaz, A.T. *et al.* (2016) The architecture of river networks can drive the evolutionary dynamics of aquatic populations. *Evol.* 70, 731-739.
16. Brauer, C.J. *et al.* (2018) On the roles of landscape heterogeneity and environmental variation in determining population genomic structure in a dendritic system. *Mol. Ecol.* 27, 3484–3497.
17. Adriaensen, F. *et al.* (2003) The application of ‘least-cost’ modelling as a functional landscape model. *Landsc. Urban. Plan.* 64, 233–47.
18. McRae, B.H. and Beier, P. (2007) Circuit theory predicts gene flow in plant and animal populations. *Proc. Natl. Acad. Sci.* 104, 19885–19890.
19. Kalinowski, S.T. *et al.* (2008) Stream trees: a statistical method for mapping genetic differences between populations of freshwater organisms to the sections of streams that connect them. *Can. J. Fish. Aquat. Sci.* 65, 2752-2760.
20. Micheletti, S.J. *et al.* (2018) Landscape features along migratory routes influence adaptive genomic variation in anadromous steelhead (*Oncorhynchus mykiss*). *Mol. Ecol.* 27, 128-145.
21. Riginos, C. *et al.* (2016) Navigating the currents of seascape genomics: how spatial analyses can augment population genomic studies. *Curr. Zool.* 62, 581-601.

22. Milanese, P. *et al.* (2017) Expert-based versus habitat-suitability models to develop resistance surfaces in landscape genetics. *Oecologia*, 183, 67-79.
23. Paris, C.B. *et al.* (2013) Connectivity Modeling System: A probabilistic modeling tool for the multi-scale tracking of biotic and abiotic variability in the ocean. *Env. Mod. Soft.* 42, 47-54.
24. Wang, I.J. and Bradburd, G.S. (2014) Isolation by environment. *Mol. Ecol.* 23, 5649-5662.
25. Xuereb, A. *et al.* (2018) Asymmetric oceanographic processes mediate connectivity and population genetic structure, as revealed by RAD seq, in a highly dispersive marine invertebrate (*Parastichopus californicus*). *Mol. Ecol.* 27, 2347-2364.
26. Duranton M. *et al.* (2018) The origin and remolding of genomic islands of differentiation in the European sea bass. *Nat Commun.* 9, 2518.
27. Hand, B.H. *et al.* (2015) Landscape community genomics: understanding eco-evolutionary processes in complex environments. *Trends Ecol. Evol.* 30, 161-168.
28. Stanley, R.R. *et al.* (2018) A climate-associated multispecies cryptic cline in the northwest Atlantic. *Science Adv.* 4, eaaq0929.
29. Benestan, L. *et al.* (2016) Seascape genomics provides evidence for thermal adaptation and current-mediated population structure in American lobster (*Homarus americanus*). *Mol. Ecol.* 25, 5073–5092.
30. Luikart, G. *et al.* (2003) The power and promise of population genomics: from genotyping to genome typing. *Nature Rev. Gen.* 4, 981-994.
31. Benestan, L. *et al.* (2017) Sex matters in massive parallel sequencing: Evidence for biases in genetic parameter estimation and investigation of sex determination systems. *Mol. Ecol.* 26, 6767-6783.
32. Landguth, E.L. *et al.* (2017) CDMetaPOP: an individual-based, eco-evolutionary model for spatially explicit simulation of landscape demogenetics. *Meth. Ecol. Evol.* 8, 4-11.
33. Davis, C.D. *et al.* (2018) Refining and defining riverscape genetics: How rivers influence population genetic structure. <https://doi.org/10.1002/wat2.1269>

34. van Strien, MJ (2017) Consequences of population topology for studying gene flow using link-based landscape genetic methods. *Ecol. Evol.* 7, 5070–5081.
35. Forester, B.R. *et al.* (2018) Comparing methods for detecting multilocus adaptation with multivariate genotype-environment associations. *Mol. Ecol.* 27, 2215–2233.
36. Prince, D.J. *et al.* (2017) The evolutionary basis of premature migration in Pacific salmon highlights the utility of genomics for informing conservation. *Sci. Adv.* 3, e1603198.
37. Waples, R.S. and Lindley, S.T. (2018) Genomics and conservation units: The genetic basis of adult migration timing in Pacific salmonids. *Evol. App.* 11, 1518–1526.
38. Manel, S. *et al.* (2016) Genomic resources and their influence on the detection of the signal of positive selection in genome scans. *Mol. Ecol.* 25, 170–184.
39. Thompson, T.Q. *et al.* (2018) Anthropogenic habitat alteration leads to rapid loss of adaptive variation and restoration potential in wild salmon populations. *Proc. Nat. Acad. Sci.*, <https://doi.org/10.1073/pnas.1811559115>.
40. Wellenreuther, M. and Hansson, B. (2016) Detecting polygenic evolution: problems, pitfalls, and promises. *Trends Gen.* 32, 155–164.
41. Lotterhos, K.E. and Whitlock, M.C. (2015) The relative power of genome scans to detect local adaptation depends on sampling design and statistical method. *Mol. Ecol.* 24, 1031–1046.
42. Teske, P.R. *et al.* (2015) On-shelf larval retention limits population connectivity in a coastal broadcast spawner. *Mar. Ecol. Prog. Ser.* 532, 1–12.
43. Rodríguez-Zárate, C.J. *et al.* (2018) Isolation by environment in the highly mobile olive ridley turtle (*Lepidochelys olivacea*) in the eastern Pacific. *Proc. R. Soc. B* 285, 20180264.
44. Fitzgerald, D.B. *et al.* (2017) Seasonal changes in the assembly mechanisms structuring tropical fish communities. *Ecology*, 98, 21–31.

45. Bernatchez, L. (2016) On the maintenance of genetic variation and adaptation to environmental change: considerations from population genomics in fishes. *J. Fish Bio.* 89, 2519-2556.
46. Narum, S.R. *et al.* (2018) Genomic variation underlying complex life-history traits revealed by genome sequencing in Chinook salmon. *Proc. Roy. Soc. B* 285, 20180935.
47. Storfer, A. *et al.* (2007) Putting the 'landscape' in landscape genetics. *Heredity* 98, 128-142.
48. Whiteley, A.R. *et al.* (2011) Population genomics of wild and laboratory zebrafish (*Danio rerio*). *Mol. Ecol.* 20, 4259-4276.
49. Hemmer-Hansen, J. *et al.* (2014) Population genomics of marine fishes: next-generation prospects and challenges. *Biol. Bull.* 227, 117-132.
50. Roesti, M. *et al.* (2015) The genomics of ecological vicariance in threespine stickleback fish. *Nature Comm.* 6, 8767. DOI: 10.1038/ncomms9767.
51. Hoban, S. *et al.* (2016) Finding the genomic basis of local adaptation: pitfalls, practical solutions, and future directions. *Am. Nat.* 188, 379-397.
52. Pearse, D.E. *et al.* (2014) Rapid parallel evolution of standing variation in a single, complex, genomic region is associated with life history in steelhead/rainbow trout. *Proc. R. Soc. Biol. Sci.* 281, 20140012
53. Wellenreuther, M. and Bernatchez, L. 2018. Eco-evolutionary genomics of chromosomal inversions. *Trends Ecol. Evol.* 33, 427-440.
54. Lewin, H.A. *et al.* (2018) Earth BioGenome Project: Sequencing life for the future of life. *Proc. Nat. Acad. Sci.* 115, 4325-4333.
55. Perrier C. *et al.* (2017) Do tremendous genetic drift and accumulation of deleterious mutations preclude adaptation to temperature in a northern lacustrine fish? *Mol. Ecol.* 26, 6317–6335.

56. Tine M. *et al.* (2014) European sea bass genome and its variation provide insights into adaptation to euryhalinity and speciation. *Nature Comm.* 5, 5770.
57. Pearson, R.G. *et al.* (2014) Life history and spatial traits predict extinction risk due to climate change. *Nature Clim. Ch.* 4, 217–221.
58. Harrisson, K. A. *et al.* (2014) Using genomics to characterize evolutionary potential for conservation of wild populations. *Evolut. Appl.* 7, 1008–1025.
59. Brauer C.J. *et al.* (2016) Riverscape genomics of a threatened fish across a hydroclimatically heterogeneous river basin. *Mol. Ecol.* 25, 5093–5113.
60. Brauer C.J. *et al.* (2017) Comparative ecological transcriptomics and the contribution of gene expression to the evolutionary potential of a threatened fish. *Mol. Ecol.* 26, 6841–6856.
61. Attard, C.R.M. *et al.* (2018) Ecological disturbance influences adaptive divergence despite high gene flow in golden perch (*Macquaria ambigua*): implications for management and resilience to climate change. *Mol. Eco.* 27, 196–215.
62. Bay, R.A. *et al.* (2018) Genomic signals of selection predict climate-driven population declines in a migratory bird. *Science* 359, 83–86.
63. Manel, S. *et al.* (2018) Predicting genotype environmental range from genome–environment associations. *Mol. Ecol.* 27, 2823–2833.
64. Bernatchez, L. *et al.* (2017) Harnessing the power of genomics to secure the future of seafood. *Trends Ecol. Evol.* 32, 665–680.
65. Ralls, K. *et al.* (2018) Call for a Paradigm Shift in the Genetic Management of Fragmented Populations. *Con. Lett.* 11, e12412.
66. Garner, B.A. *et al.* (2016) Genomics in conservation: case studies and bridging the gap between data and application. *Trends Ecol. Evol.* 31, 81–83.
67. Rellstab, C., *et al.* (2015) A practical guide to environmental association analysis in landscape genomics. *Mol. Ecol.* 24, 4348–4370.

570 68. Pollak, J.P. and Lacy, R.C. (2017) Metamodel Manager (v.1.0.4). Chicago Zoological  
571 Society, Brookfield, Illinois, USA.

572 69. Jeremias G. *et al.* (2018) Synthesizing the role of epigenetics in the response and  
573 adaptation of species to climate change in freshwater ecosystems. *Mol Ecol.* 27, 2790–  
574 2806.

575 70. Luikart, G. *et al.* (2018) Population genomics: advancing understanding of nature. In  
576 Population genomics: concepts, approaches and applications (ed. Rajora, O.P.) Springer  
577 International Publishing AG.

578 71. Catanach, A. *et al.* (2019) The genomic pool of standing structural variation outnumbers  
579 single nucleotide polymorphism by more than three-fold in the marine teleost *Chrysophrys*  
580 *auratus*. *Mol. Ecol.* (in press)

581 72. Günther, T. and Coop, G. (2013) Robust identification of local adaptation from allele  
582 frequencies. *Genetics* 195, 205–220.

583 73. de Villemereuil, P. and Gaggiotti, O.E. (2015) A new FST-based method to uncover local  
584 adaptation using environmental variables. *Methods Ecol. Evol.*, 6, 1248-1258.

585 74. Foll, M. and Gaggiotti, O. (2008) A genome-scan method to identify selected loci  
586 appropriate for both dominant and codominant markers: A Bayesian perspective. *Genetics*  
587 180, 977–993

588 75. Manion, G. *et al.* (2016) gdm: Functions for generalized dissimilarity modeling. R package.

589 76. Guillot, G. *et al.* (2005) A spatial statistical model for landscape genetics. *Genetics* 170,  
590 1261–1280.

591 77. Foll, M., and Gaggiotti, O. (2006) Identifying the environmental factors that determine the  
592 genetic structure of populations. *Genetics* 174, 875–891.

593 78. Guillot, G. *et al.* (2014) Detecting correlation between allele frequencies and environmental  
594 variables as a signature of selection. A fast computational approach for genome-wide  
595 studies. *Spat. Stat.* 8, 145–155.

596 79. Frichot, E *et al.* (2015) Detecting adaptive evolution based on association with ecological  
597 gradients: Orientation matters!. *Heredity* 115, 22–28.

598 80. Luu, K. *et al.* (2017) pcadapt: an R package to perform genome scans for selection based  
599 on principal component analysis. *Mol. Ecol. Resour.* 17, 67-77.

600 81. Micheletti, S.J. and Narum, S.R. 2018. Utility of pooled sequencing for association mapping  
601 in nonmodel organisms. *Mol. Ecol. Resour.* 2018, 825–837.

602 82. Breiman, L. (2001). *Random Forests*, *Machine Learning* 45, 5-32.

603 83. Legendre, P. and Legendre, L. (2012) *Numerical Ecology*. 3rd English ed. Elsevier.

604 84. Wright, S. (1943) Isolation by distance. *Genetics* 28, 114-128.

605 85. Krimbas, C.N. (1967) The genetics of *Drosophila subobscura* populations. III. Inversion  
606 polymorphism and climatic factors. *Mol. Gen. Genetics* 99, 133-150.

607 86. Lewontin, R.C. and Hubby, J.L. (1966) A molecular approach to the study of genic  
608 heterozygosity in natural populations. II. Amount of variation and degree of heterozygosity  
609 in natural populations of *Drosophila pseudoobscura*. *Genetics* 54, 595-609.

610 87. Balkenhol, N. *et al.* (2016). *Landscape genetics: concepts, methods, applications*. J. Wiley  
611 and Sons Ltd.

612 88. Epperson, B.K. (2003) *Geographical genetics*. Princeton University Press.

613 89. Holderegger, R. and Wagner, H.H. (2006) A brief guide to landscape genetics. *Landscape*  
614 *Ecol.* 21, 793-796.

615 90. Segelbacher, G. *et al.* (2010) Applications of landscape genetics in conservation biology:  
616 concepts and challenges. *Conserv. Genet.*, 11, 375-385.

617 91. Holderegger, R. *et al.* (2010) Landscape genetics of plants. *Trends Plant Sci.* 15, 675-683.

618 92. Biek, R. and Real, L.A. (2010) The landscape genetics of infectious disease emergence  
619 and spread. *Mol. Ecol.* 19, 3515-3531.

620 93. Hanski, I. (1998) Metapopulation dynamics. *Nature* 396, 41.



94. Taylor, P. *et al.* (1993) Connectivity is a vital element of landscape structure. *Oikos* 68, 571–573.
95. Graves TA, Beier P, and Royle JA. 2013. Current approaches using genetic distances produce poor estimates of landscape resistance to interindividual dispersal. *Mol Ecol*: 1–16.
96. van Rees, C.B. *et al.* (2018) Landscape genetics identifies streams and drainage infrastructure as dispersal corridors for an endangered wetland bird. *Ecol. Evol.* 8, 8328–8343.
97. Hotaling, S. *et al.* (2018) Demographic modelling reveals a history of divergence with gene flow for a glacially tied stonefly in a changing post-Pleistocene landscape. *J. Biog.* 45, 304–17.
98. Olsen, J.B. *et al.* (2010) Comparative landscape genetic analysis of three Pacific salmon species from subarctic North America. *Con. Gen.* 12, 223–41.
99. Kremer, C.S. *et al.* (2017) Watershed characteristics shape the landscape genetics of brook stickleback (*Culaea inconstans*) in shallow prairie lakes. *Ecol. Evol.* 7, 3067–79.
100. De Kort, H. *et al.* (2018) Interacting grassland species under threat of multiple global change drivers. *J. Bio.* DOI: 10.1111/jbi.13397
101. Raeymaekers, J.A.M. *et al.* (2017) Adaptive and non-adaptive divergence in a common landscape. *Nature Comm.* 8, 267.
102. Stat, M. *et al.* (2017) Ecosystem biomonitoring with eDNA: metabarcoding across the tree of life in a tropical marine environment. *Sci. Rep.* 7, 12240.
103. Page, T.J. *et al.* (2019) Multiple molecular markers reinforce the systematic framework of unique Australian cave fishes (Milyeringa: Gobioidae). *Austr. J. Zoo.*, 66, 115-127.
104. Deiner, K. *et al.* (2017) Environmental DNA metabarcoding: transforming how we survey animal and plant communities. *Mol. Ecol.* 26, 5872–5895.
105. Lacoursière-Roussel A. *et al.* (2018) eDNA metabarcoding as a new surveillance tool for coastal Arctic biodiversity. *Ecol. Evol.* 8, 7763-7777.

- 647 106. Bálint, M. *et al.* (2018) Environmental DNA Time Series in Ecology. Trends Ecol. Evol. 33,  
648 945-957.
- 649 107. Meier, K. *et al.* (2011) An assessment of the spatial scale of local adaptation in brown  
650 trout (*Salmo trutta* L.): footprints of selection at microsatellite DNA loci. Heredity 106, 488-  
651 499.

## Glossary

**Adaptive capacity:** ability of a population to evolve in response to changing environments such that the mean population fitness is maintained or increases following the change

**Adfluvial:** aquatic organisms that breed and develop in streams and subsequently enter nearby lakes to reach sexual maturity

**Anadromy/anadromous:** migration strategy where an individual is born in fresh water, subsequently migrates to the marine environment where it develops as an adult, then returns to fresh water to spawn

**Association mapping:** uses natural populations, as opposed to controlled breeding lines, to associate genomic regions with a trait (phenotypic or environmentally related) of interest (also known as “linkage disequilibrium mapping”)

**Catadromy/catadromous:** migration strategy where an individual is born in the marine environment and subsequently migrates to fresh water to rear and develop, then returns to the marine environment to spawn

**Cost surface:** representation of the fitness cost associated with features of a landscape/waterscape for a given species as a set of spatially discrete weights (also known as a resistance map)

**Connectivity modeling:** the application of a computational model (e.g., least-cost path, circuit theory, dispersal kernel, etc.) on a cost surface (resistance) map

**Dendritic network:** the spatial arrangement of river basins in hierarchic units such as reaches, streams, subcatchments, and catchments, where two water segments join at confluence points and become a single segment

**Diadromy/diadromous:** an organism that spends part of its life in fresh water and part in saltwater; see anadromy and catadromy for examples

**Environment-associated SNPs:** SNPs with allele frequencies significantly associated with variation in one or more environmental variables of interest. Often identified via GEA analyses. Associated SNPs can be validated as adaptive SNPs through experiments or mapped to genes of functional relevance.

**Evolutionary rescue:** the recovery of a population from environmental perturbation via genetic adaptation

**Genotype-environment association (GEA) analyses:** uni or multivariate analyses used to identify candidate adaptive SNPs by testing for direct associations between variation in allele frequencies and environmental variables.

**IBD:** isolation-by-distance, a pattern where genetic similarity decays with increasing geographic distance between individuals/populations

**IBE:** isolation-by-environment, a pattern where genetic similarity decays with increasing ecological distance between individuals/populations

**Linkage disequilibrium (LD):** the non-random association between alleles at different regions in the genome, often caused by physical genomic proximity

**Migration:** dispersal of an individual followed by successful reproduction (also referred to as “effective migration”)

**Panmictic/panmixia:** interbreeding between populations leading to no population genetic structure

**Pelagic:** open water in lakes, oceans and seas not near the bottom or shore

696 **QTL**: quantitative trait locus, a genomic region associated with the variation of a quantitative  
697 (often phenotypic) trait  
698 **SNP**: single nucleotide polymorphism, a variant position in the genome  
699 **Type I error**: “false positive” rate - the null hypothesis is rejected when it is actually true  
700 **Type II error**: “false negative” rate - failing to reject the null hypothesis when the alternative  
701 hypothesis is true

702 Table II. A list of databases and software for researchers to utilize in aquatic landscape genomic studies.  
703

Spatial		
Resource Name	Description	Website
Bio-Oracle	Marine data base for > 20 environmental parameters for present and projected future conditions	<a href="http://www.bio-oracle.org/downloads-to-email.php">http://www.bio-oracle.org/downloads-to-email.php</a>
BioClim 2.0	Global climate layers for mapping and spatial modeling	<a href="http://worldclim.org/version2">http://worldclim.org/version2</a>
BioSim	Simulation of climate-driven models to forecast future events	<a href="https://cfs.nrcan.gc.ca/projects/133">https://cfs.nrcan.gc.ca/projects/133</a>
Copernicus Global Land Service	Bio-geophysical data for European and Global ecosystems	<a href="https://land.copernicus.eu">https://land.copernicus.eu</a>
Coriolis	Real time geo-physical marine data for Western Europe	<a href="http://www.coriolis-cotier.fr">http://www.coriolis-cotier.fr</a>
Geoscience Australia	Geospatial datasets for Australia, including multiple online tools for data analysis	<a href="http://www.ga.gov.au/">http://www.ga.gov.au/</a>
Global Biodiversity Information Facility (GBIF)	Geographic distribution data for a multitude of species	<a href="http://www.gbif.org">http://www.gbif.org</a>
National Hydrographic Network	Geospatial data for Canada's inland surface waters	<a href="https://www.nrcan.gc.ca/earth-sciences/geography/topographic-information/geobase-surface-water-program/21361">https://www.nrcan.gc.ca/earth-sciences/geography/topographic-information/geobase-surface-water-program/21361</a>
Marspec	High resolution contemporary and paleo marine spatial ecology data	<a href="http://www.marspec.org">http://www.marspec.org</a>
Natural Earth Geographic	Public domain map dataset for map making and GIS usage	<a href="http://www.naturalearthdata.com/">http://www.naturalearthdata.com/</a>
NOAA WOD (World Ocean Database)	Oceanic datasets from million-year-old sediment records to near real-time	<a href="https://www.nodc.noaa.gov/OC5/WOD/pr_wod.html">https://www.nodc.noaa.gov/OC5/WOD/pr_wod.html</a>

	satellite images	
OceanParcels	Lagrangian framework to create customisable particle tracking simulations	<a href="http://oceanparcels.org/">http://oceanparcels.org/</a>
Ocean Surface Current Analyses Real-time (OSCAR)	Near-real-time global ocean surface data	<a href="https://www.esr.org/research/oscar/">https://www.esr.org/research/oscar/</a>
Genomics		
Software Name	Description	Reference
BayEnv	Outlier loci and local adaptation identification via allelic frequencies and environmental variables	[72]
BayeScEnv	Local adaptation detection via genotypic and environmental data	[73]
BayeScan	Outlier detection, no environmental data	[74]
gdm	Generalised dissimilarity modelling and gradient forests	[75]
Geneland	Identification of populations and their boundaries with genomic and geographic data	[76]
GESTE	Identification of environmental factors contributing to population structure	[77]
gINLAnd	Univariate method for local adaptation identification via allelic frequencies and environmental variables	[78]
LEA* (LFMM)	Local adaptation detection via genotypic and ecological data	[79]
PCAdapt*	Outlier detection, no environmental	[80]

	data	
PoolParty	Pipeline to identify genes associated with adaptation & phenotypic traits from whole genome resequencing	[81]
randomForest*	A powerful machine-learning algorithm to discern loci underlying phenotypic traits of environment association	[82]
vegan*	Implementation of RDA, local adaptation identification via allelic frequencies and environmental variables	[83]
*denotes R packages		

### **Box 1. A perspective on the history of landscape genetics/genomics**

The roots of landscape genetics may trace to 19<sup>th</sup> Century biogeographers who noted variable community composition and species traits across the landscape [5]. The first theoretical articulation of spatial variation for neutral traits was Wright's "isolation-by-distance" [84]. Cline analysis was an early analytical framework for landscape genetics because clines are associated with local adaptation and gene flow between populations [85]. Landscape genetics emerged as a discipline following: (1) development of methods to resolve genetic or protein (allozyme) variation at multiple loci in the 1960s and the realization that natural populations housed variation associated with environmental factors [86], (2) the "DNA revolution" beginning in the 1980s and associated growth of conservation genetics and molecular ecology, and (3) the realization that human alterations to habitats could impact genetic variation, along with (4) the subsequent founding of landscape ecology as a discipline in the 1990s [87]. All led to the notion that the concepts and tools of population genetics and landscape ecology could be combined to understand environmental heterogeneity and its impacts on genetic diversity, divergence, and microevolutionary processes. These ideas coalesced in Epperson's *Geographical Genetics* [88] and the first definition of "landscape genetics" [5].

A novel aspect of landscape genetics was the use of individual-based approaches to assess fine-scale variation and more precisely localize barriers relative to population-level approaches. The landscape genetic approach initially focused on genetic assays and analytical methods available at the time with, understandably, little ability to draw broad inferences about pattern or process [5]. Luikart et al. [30] advocated a "population genomic approach" to studying associations between genetic and environmental variation, i.e., simultaneously examining neutral and adaptive variation across putative selection gradients at thousands of loci *across the genome*. Later, landscape genetics was expanded by explicitly including "adaptive and neutral variation" and specifying the study of "landscape composition and



729 configuration, including matrix quality” [89]. This idea was further extended by calling for explicit  
730 quantification of landscape effects on genetic variation [47]. Other reviews highlighted: the  
731 formal recognition of a “landscape genomic” approach [8,67], landscape genetics in  
732 conservation [90], plants [91], infectious diseases [92], that neutral and selective factors  
733 impacting the genome may include species interactions, i.e., “landscape community genomics”  
734 (Box 3; [27]), and the first textbook on the subject [87].

## Box 2: **Measuring and modeling genetic connectivity**

Modeling gene flow within a connectivity framework is rooted in metapopulation and spatial ecology, where migration is described between habitat patches ([93]; also see Box 1). Taylor et al. [94] advocated for the importance of understanding landscape connectivity as the degree to which the landscape facilitates or impedes movement among resource patches. This is the definition most often used for functional connectivity, which can further represent the response of individuals (physiological and behavioral) to the structural landscape and can disrupt or modify dispersal patterns that is realized through immediate or deferred mortality costs and risks [87]. Finally, functional connectivity is often measured in terms of the “effective distance” that represents the cost of a path between suitable habitat patches or across heterogeneous landscapes that is the Euclidean distance weighted by the cumulative resistance of all landscape types traversed [17].

A common approach to measure genetic connectivity in landscape genetics is to statistically compare the effective distance to some measure of genetic distance (often  $F_{ST}$  or individual-based genetic differentiation metric, or allele frequencies). The most challenging part of connectivity modeling remains model selection, and there have been multiple simulation-based studies on model selection tests, with a perhaps overemphasis on Mantel tests (Table I; [95]). Despite extensive testing using simulation-based approaches, there remains to be a consensus on the most appropriate (or most correct) model selection test, and development of new approaches (and testing of older ones) is still ongoing.

Recent approaches to measure connectivity have used a mixed-model maximum-likelihood population-effects framework to identify linear water features (e.g., streams, canals and ditches) as potentially important in the dispersal of a wetland bird [96]. Additionally, tools like StreamTree [19] might offer improved granularity in dendritic (or dendritic-like) systems where a specific  $F_{ST}$  can be associated with each branch segment rather than each pair of populations; this could therefore be useful in identifying barriers to gene flow between

populations. Brauer et al. [16] further used StreamTree with multiple matrix regression with randomization to integrate genetic connectivity model results into a GEA framework for rainbowfish (*Melanotaenia fluviatilis*).

**Table I. Model selection approaches for assessing population connectivity.** A non-exhaustive list of model selection approaches used to assess population connectivity in aquatic landscape genetics, including potential weaknesses for each method.

Statistical approach	Notes	Potential weaknesses	Example Reference
Mantel tests	Most common test for testing IBD in genetic structure	High type I error rates [95]	[97]
Partial Mantel tests	Common test for IBR by partially out Euclidean distance	High type I error rates [95]	[98]
Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC)	Commonly used in many genetic analyses. BIC more heavily penalizes model overparameterization	Not appropriate for all mixed-model approaches, or multiple regression on distance matrices	[99]
Distance-based Moran's eigenvector map (dbMEMs)	Capable of detecting spatial structure at several scales to help control for spatial correlation in tests of y~x relationships	None yet determined	[29]
Mixed-model maximum-likelihood population-effects framework with (MLPE)	Can be used with AIC, BIC or $R_{\beta}^2$ *	None yet determined	[96]
Multiple matrix regression with randomization (MMRR)	Assesses the relative effects of IBD and IBE	Difficulties in estimating relative importance of correlated variables, as well as choosing best model selection method	[16]

\*The  $R_{\beta}^2$  statistic measures the proportion of observed variation explained by the fixed effects of the model

### Box 3. **Landscape community genomics**

Landscape genomics investigates interactions between population genetic diversity and environmental variation, whereas community genetics investigates interactions between genetic diversity and species interactions; landscape community genomics (LCG) is the merging of these two approaches [27]. To fully understand processes of eco-evolutionary change, researchers should consider simultaneously the effects of abiotic (environmental) and biotic (community) factors on demography, evolution, and genomic variation within and among populations.

To design a landscape community genomic study, researchers ideally include multiple strongly-interacting species distributed across environmental (selection) gradients and both candidate adaptive and genome-wide (high density) neutral loci. Here we discuss three informative LCG examples: a terrestrial LCG study, an aquatic study lacking strong species-interactions, and community environmental DNA (eDNA) studies lacking intraspecific population genetic markers.

An exemplary LCG study [100] involved the specialized Alcon butterfly (*Phengaris alcon*), which is sensitive to grassland habitat configuration and requires the presence of the rare marsh gentian plant (*Getiana pneumonanthe*) and an ant species (*Myrmica spp.*). RAD sequencing was used to assess relations between genetic diversity, connectivity, habitat suitability, grazing (by livestock), and altitude. Climate warming and seasonal grazing abandonment strongly affected the distribution of the Alcon butterfly because grazing and climate affect availability of the gentian host plant.

Raeymaekers et al. [101] used a comparative framework to test if two stickleback species differ in neutral and adaptive divergence along an environmental (salinity) gradient. Phenotypic and neutral marker differentiation along with genomic signatures of adaptation were stronger in the three-spined (*Gasterosteus aculeatus*) than in the nine-spined (*Pungitus pungitus*) stickleback. Signatures of adaptation involved different genomic regions in the two

species, and thus were non-parallel. Such multispecies studies provide insight into mechanisms underlying evolutionary change and adaptive strategies within landscapes. Future studies that include strongly-interacting species (e.g., competitors, predator-prey, host-pathogen) could prove to be especially informative.

Environmental DNA (eDNA) metabarcoding will allow for genotyping or microhaplotyping of eDNA fragments from each of multiple species, simultaneously. It thus offers a potentially powerful means for population genetic/genomics studies, although few multilocus studies have been published (e.g., [102,103]). This approach will eventually allow for inferences about biotic and abiotic factors shaping population genetic structure and also community structure [27,104-106]. It is exciting to consider that future eDNA metabarcoding studies (including many loci) will eventually allow for LCG studies.

## Highlights

- Proliferation of genome-scale studies on aquatic species have resulted from the decreasing costs of high throughput sequencing combined with novel computational approaches
- Our increasing understanding of aquatic species' genomes has enabled the annotation of loci that are adaptive, sex-linked, and associated with phenotype, allowing the inference of evolutionary and demo-genetic processes from spatio-temporal genetic patterns
- Recent improvements in climate and habitat data for aquatic systems provide a more precise characterization of aquatic niches, facilitating landscape genomics
- Many landscape genetic analytical methods have recently been developed specifically for aquatic systems
- We provide a list of spatial and genomic resources as part of a "roadmap" to guide future aquatic landscape genomic studies

## Outstanding Questions

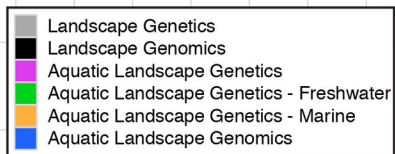
1. How do we better integrate epigenetics and transcriptomics into aquatic landscape genomics to understand eco-evolutionary processes and improve biodiversity conservation?
2. What is the role of structural genetic variants (SGV) in affecting ecological and evolutionary processes on the landscape?
3. How can landscape genomic approaches be used to monitor, model, control, and inform policy regarding the spread of adaptive and mal-adaptive alleles between natural and genetically-manipulated populations?
4. How well do genomic regions identified using GEA match results from experimental functional analyses?
5. How can landscape genomic modeling approaches improve prediction of population viability and community vulnerability?

## Figure captions

Figure 1. **Aquatic landscape genomics studies are on the rise.** Results from a literature search in the ISI Web of Science on the six topics listed in the legend. “Aquatic landscape genomics” was first referenced in the literature in by Meier et al. [107], and although still underrepresented, has been increasing since then. See Supplemental Table S1 to determine how the literature search was conducted.

Figure 2. **Environmental and demographic features affect landscape genetic patterns and processes.** Conceptual summary highlighting key points of aquatic landscape genomics illustrating headwaters (near glaciers in white), lakes, large rivers, and marine environments. Many aquatic systems are characterized by sharp environmental gradients including temperature (headwaters, lakes, oceans), pH (lakes), and salinity (oceans), all of which create adaptive selective pressures. Many populations in marine environments are characterized by large effective population sizes ( $N_e$ ) and high rates of gene flow that are often asymmetrically affected by prevailing currents. Conversely, many inland and alpine lake populations show small population sizes with low rates of migration between lakes; riverine environments represent a mix of these extremes and often have impediments to gene flow including anthropogenic (e.g., dams) and natural (e.g., waterfalls) barriers.

Number of Publications



300

200

100

0

1991

1994

1997

2000

2003

2006

2009

2012

2015

2018

Year Published



