Land, river, & seascape genomics

What is landscape genomics?

(landscape = any habitat)

> "At the heart of spatial and space–time analysis of population genetics is the connection between observed spatial patterns and the space–time processes that generate them." - Epperson 2003
>
> "the interaction between landscape features and microevolutionary processes, such as gene flow, genetic drift and selection." - Manel et al 2003
>
> Landscape genetics tests the model that $G \sim f(E)$ - Dyer 2015, Molecular Ecology

Is landscape *genomics* fundamentally different from landscape *genetics*?

No, just swap "genetics" for "genomics" - Balkenhol et al 2016

Yes - "Whereas landscape genetics studies primarily focus on testing the effects of landscape variables on gene flow and genetic population structure, landscape genomics studies focus on detecting candidate genes under selection that indicate possible local adaptation." - Storfer et al 2018

What do *you* think?

In this course, we will focus on a few topics in landscape genomics aiming to give you a solid foundation in the field. We will not be comprehensive but will try to point out important topics when we encounter them. There is a stronger emphasis on population genomics and various landscapes (especially marine) in this course than might be found in other landscape genomic courses. Look at the optional readings as suggestions on how to learn more about various topics.

What are we covering and why?

- Making maps and using spatial data

TABLE 1 | General differences between landscape genetics and landscape genomics studies.

	Questions	Scale of study	Sampling design	Analysis methods
Landscape genetics	Influence of landscape on gene flow	Among populations	**Stratified random**, opportunistic, clumped, individual-level	Mantel tests, *Assignment tests* (spatial and aspatial; e.g., Structure, Tess, Geneland), *Ordination (dbRDA, sPCA. MDS)*, Least cost paths (multiple regression, MLPE), Spatial autocorrelation, Spatial regression, EEMS*
	Influences of landscape on at-site variation	Within populations	**Across ecological gradients**, stratified	Graph models (e.g, Popgraph), GDMs, Structural equation models
	Barriers	Among populations	**Across hypothesized barrier(s)**	Wombling, Monmonier's maximum difference algorithm, spatial assignment tests (e.g., Geneland)
	Species' ecology	Within and among populations	**Across ecological gradients** (stratified)	Ordination, Least cost paths, Spatial autocorrelation, Spatial regression
	Source-sink dynamics	Among populations	**Across populations of different sizes or fragmentation levels**	Mantel tests, genetic diversity estimates (e.g., F-statistics, bottleneck tests)
Landscape genomics	Spatial patterns of selection	Among populations	**Paired sampling**, transect sampling	Outlier differentiation methods (eg., Bayescan, FLK, X^TX); Genotype-environment associations (e.g., Bayenv2, PC Adapt, LFMM, sGLMM, Samβada), *Ordination*, *Assignment tests* (e.g., FASTSTRUCTURE, Admixture, Tess3)
	Influence of landscape on local adaptation	Among populations	Transect sampling, paired sampling, stratified sampling	Outlier differentiation methods; Genotype-environment associations, *Ordination*, *Assignment tests*, Genomic cline analysis*, GDM*, EEMS*

*Note that, when conducting a landscape genomics study, that when loci under selection are removed and putatively neutral loci remain, that landscape genetics questions and analyses can then be conducted. Nonetheless, sampling designs generally differ between landscape genetics and landscape genomics studies, so some landscape genetics questions may not be addressable in studies with landscape genetics goals. Bolded sampling designs indicate preferred designs for that particular question. Not all analysis methods under each study type are listed, just those that are most commonly used or best suited to address the goals of the study. Note also that assignment test methods generally differ between landscape genetics and landscape genomics studies. Italicized words under analysis type indicate those commonly used in both landscape genetics studies of gene flow and landscape genomics studies of loci involved in adaptation. dbRDA, distance-based redundancy analyses; sPCA, spatial principal components analysis; MDS, multidimensional scaling; MLPE, maximum likelihood of population effects (Clarke et al., 2002); LFMM, latent factor mixed models; sGLMM, spatial generalized linear mixed models; EEMS, Estimated Effective Migration Surface (Petkova et al., 2016). Software names include: Geneland (Guillot et al., 2005), Structure (Pritchard et al., 2000), Tess (Durand et al., 2009), Popgraph (Dyer and Nason, 2004); Bayescan Foll and Gaggiotti, 2008, FLK (Bonhomme et al., 2010), Bayenv2 (Günther and Coop, 2013), PCadapt (Duforet-Frebourg et al., 2014) Faststructure (Raj et al., 2014), Admixture (Alexander et al., 2009), Tess3 (Caye et al., 2016). *indicates methods not yet widely used but show promise–see Sections Generalized Dissimilarity Modeling (GDM)–Clinal Analyses.*

Figure 1: Comparing landscape genetics and genomics - Storfer et al 2018

- Describing genetic variation and genetic structuring
- RDA as a flexible tool
- Simulations and demographic analyses
- Resistance surfaces
- Projecting into the future with generalized dissimilarity modelling, gradient forests
- Biophysical models of dispersal
- More genotype-environment associations and genomic offsets
- (Time permitting) Landscape genomics and genetic architectures

Some important topics we will not cover very much/at all

- Sampling design
- Tests of selection

To discuss: Pattern vs. Process, many ways to have different allele frequencies

---

Activity 1

Make a poster of your study system - no one will be judged for their artistry!

Your poster needs to include information on your organism and your landscape.

- What are the important spatial aspects of your landscape?
- Are there critical genomic elements to your study?
- What is the overarching question or hypothesis?
- (Don't forget to add your name!)

Unifying elements of landscape genomics

Fundamentally there are three main steps to any landscape genetic/genomic study:

1. Describe spatial variability
2. Describe genetic variability
3. Use statistics to look for correlations between spatial and genetic attributes.

(Recommended by rarely done: independent corroboration/validation)

(Optional - very popular recently, predict adaptive matching to future environments, genomic offsets)

Spatial variable attributes

(that often violate statistical assumptions)

Other factors to consider -

- What is the grain size of your spatial variables?
- Are spatial variables site specific or gridded (remote sensing)?

Activity 2

Consider three basic habitat types: terrestrial, marine, freshwater

- What spatial factors affect dispersal, either blocking or facilitating?
- What spatial factors affect population sizes?
- What environmental factors are likely to have been important agents of selection within species' ranges over evolutionary time?
- Under historical conditions, what is the time frame of the above spatial factors? (thousands of years, tens of years, months, days…)
- How have human activities modified any of the above factors (spatially, temporally)?
- (For all of these questions you might decide to break their effects down by different taxonomic groups)
- Nominate one person to report back to the class

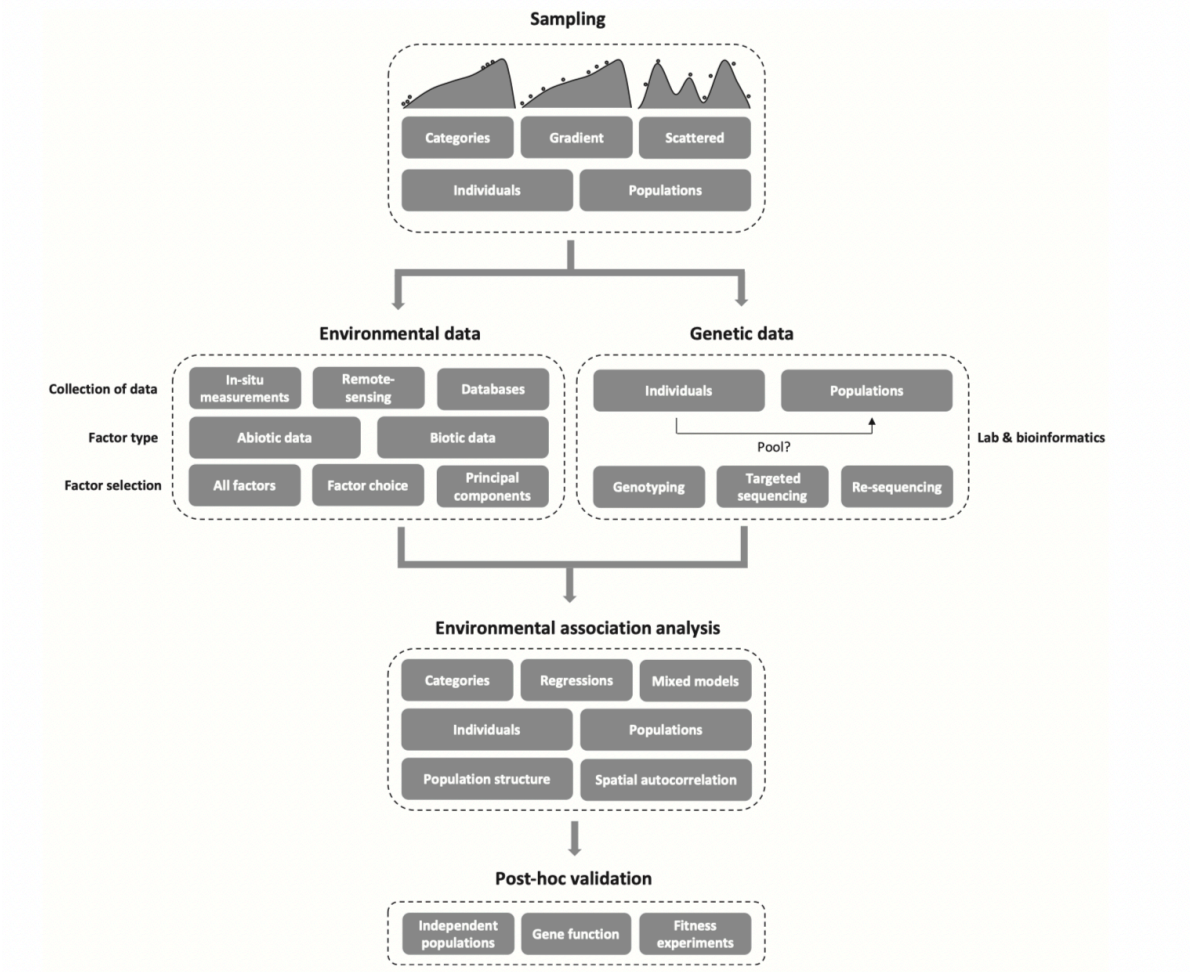


Fig. 1 A typical workflow in environmental association analysis (EAA). The three most important options per step are horizontally aligned. Genetic and environmental data are collected at the same sampling locations, processed separately and jointly analysed in EAA. The results can subsequently be validated with complementary approaches. All steps and options are described in detail in the manuscript.

Figure 2: LG overview - Rellstab et al. 2015

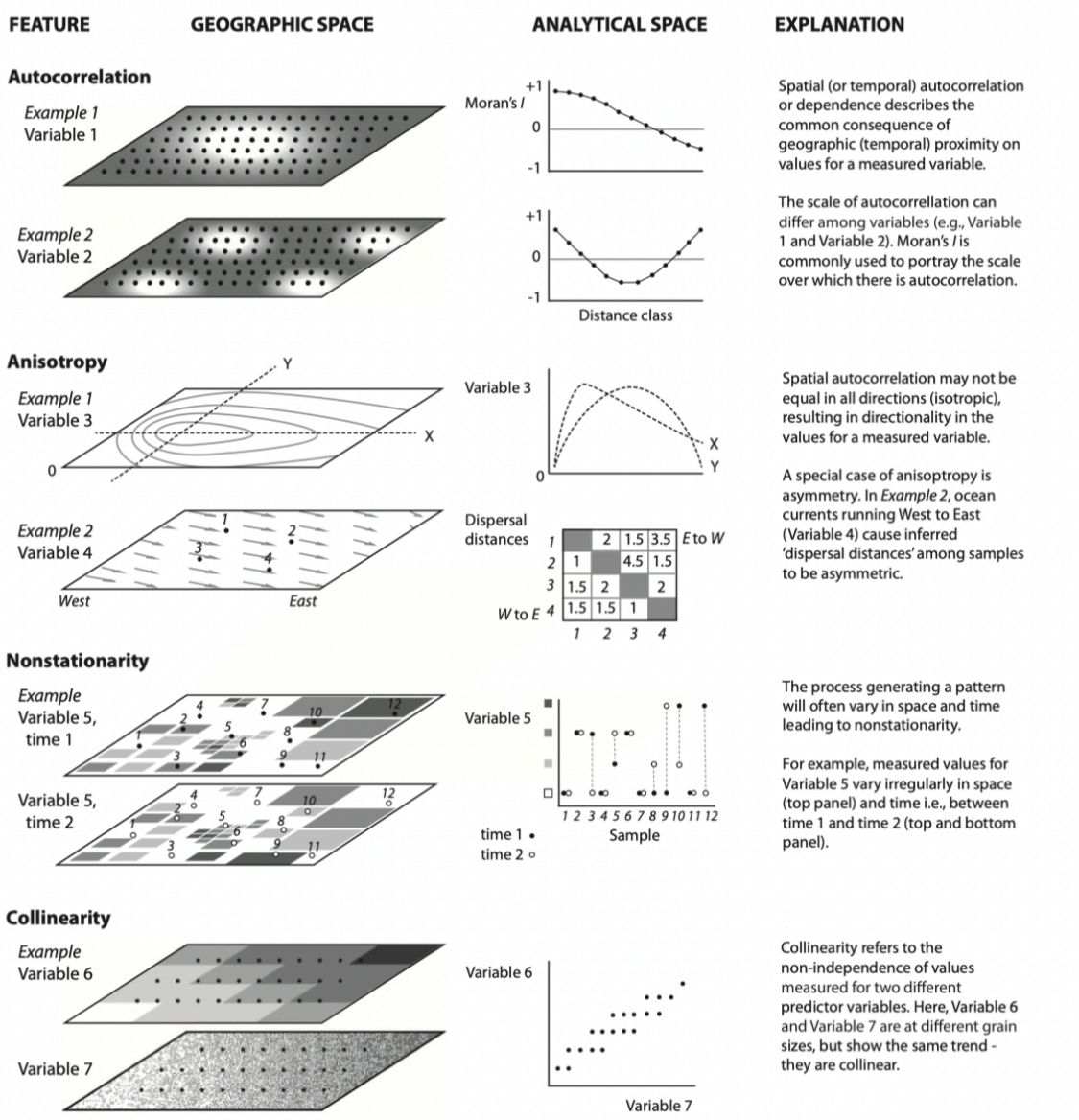


Figure 1. Key concepts relevant to the properties of spatial and environmental variables used in seascape genomic analyses. These properties should be considered during the project design as they will influence which variables and what representative values of variables may be used. Moreover, these properties will help determine what methods are appropriate for analysis. The figure displays examples of the concepts in geographic space and their manifestations in analytical space. Points in the geographic space depict the location of sampling, and dashed lines represent a transect (Anisotropy, Example 1 only).

Figure 3: Attributes of spatial variables - Riginos et al. 2016

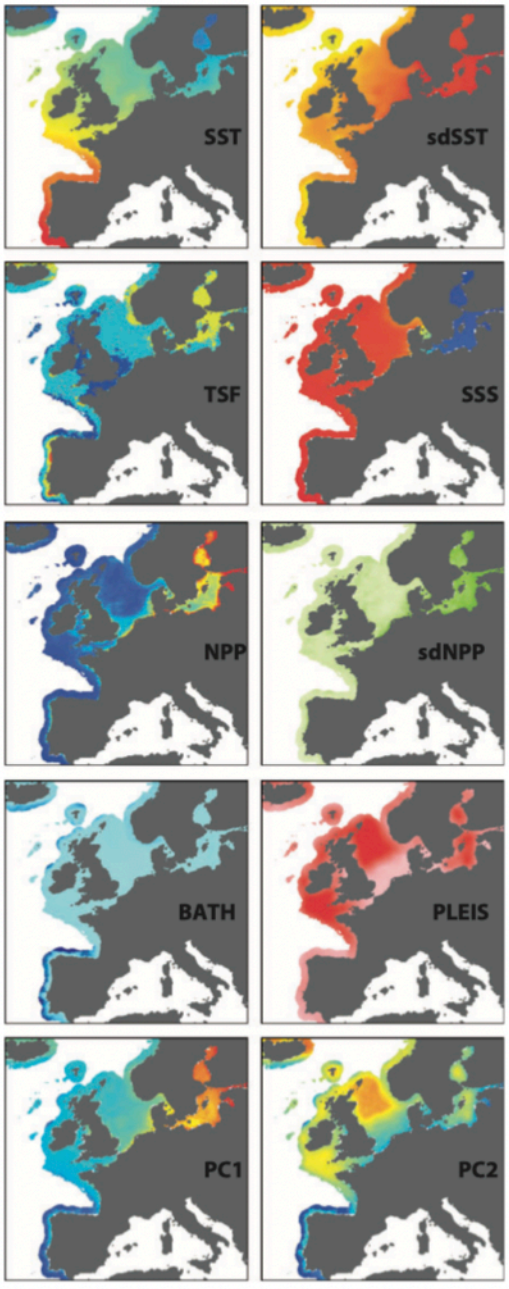


Figure 2. Spatial patterns in environmental variables of Atlantic European coastal waters. Eight select coastal seascape variables are shown including mean sea surface temperature (SST), standard deviation of sea surface temperature (sdSST), mean thermal stress frequency (TSF), mean sea surface salinity (SSS), mean net primary productivity (NPP), standard deviation of net primary productivity (sdNPP), bathymetry (BATH), and Pleistocene habitat suitability (PLEIS). In addition, the values for principal components 1 and 2 describing the eight coastal variables are also shown. PC1 and PC2 account for 47.1% and 17.9% of the variance among variables, respectively.

Figure 4: Various spatial variables for Northern Europe - Riginos et al 2016

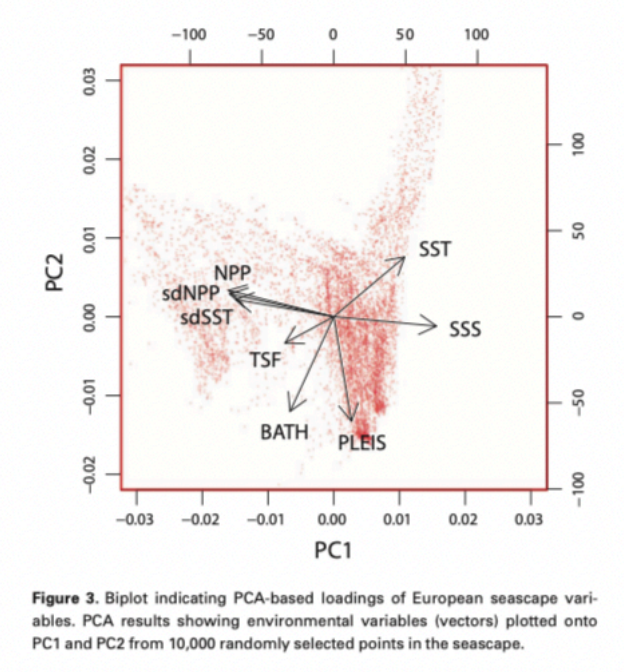


Figure 3. Biplot indicating PCA-based loadings of European seascape variables. PCA results showing environmental variables (vectors) plotted onto PC1 and PC2 from 10,000 randomly selected points in the seascape.

Figure 5: Using PCA biplot to visualise correlations - Riginos et al 2016

Further reading on related ideas for aquatic habitats:

- Blanchet, S., Prunier, J. G., Paz-Vinas, I., Saint-Pe, K., Rey, O., Raffard, A., Mathieu-Begne, E., Loot, G., Fourtune, L., & Dubut, V. (2020). A river runs through it: The causes, consequences, and management of intraspecific diversity in river networks. Evolutionary Applications, 13(6), 1195-1213.
- Grummer, J. A., Beheregaray, L. B., Bernatchez, L., Hand, B. K., Luikart, G., Narum, S. R., & Taylor, E. B. (2019). Aquatic landscape genomics and environmental effects on genetic variation. Trends in Ecology and Evolution, 1-14.
- Riginos, C., & Liggins, L. (2013). Seascape genetics: populations, individuals, and genes marooned and adrift. Geography Compass, 7(3), 197-216.
- Selkoe, K. A., Henzler, C. M., & Gaines, S. D. (2008). Seascape genetics and the spatial ecology of marine populations. Fish and Fisheries, 9(4), 363-377.

The "Matrix" and how it differs on land and in the sea

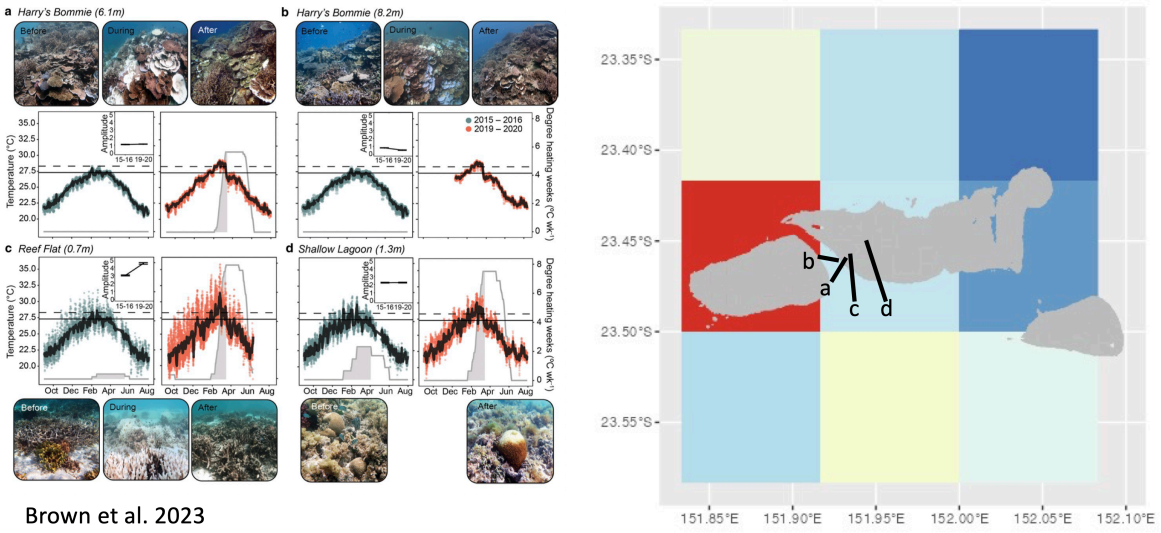
McRae 2006 – Isolation by resistance:

Table 1. Descriptive statistics for eight select seascape variables for the northeast Atlantic region

Layer	Abbrev	Min	Max	Mean	Standard deviation	Units	Moran's *I* by distance (km)[a] 25	50	100	200	500	Data source
Mean sea surface temperature	SST	2.197	19.859	10.600	2.916	°C	0.66	0.60	0.50	0.39	0.26	NOAA
Standard deviation of sea surface temperature	sdSST	1.201	8.211	3.587	1.595	°C	**0.74**	0.69	0.63	0.58	0.47	NOAA
Mean thermal stress frequency	TSF	0.00	22.00	1.10	0.84	freqency[b]	0.45	0.39	0.30	0.23	0.15	CoRTAD
Mean sea surface salinity	SSS	2.108	36.524	29.928	10.607	unitless[c]	**0.72**	0.64	0.57	0.51	0.37	World Ocean Atlas 2013 v2
Mean net primary productivity	NPP	478	12,788	2,063	1697	C m^{-2} day^{-1}	**0.73**	0.62	0.52	0.44	0.35	Ocean Productivity web
Standard deviation of net primary productivity	sdNPP	276	15,945	3,521	3,080	C m^{-2} day^{-2}	**0.76**	0.67	0.59	0.52	0.39	Ocean Productivity web
Bathymetry	BATH	−5,029	839	−266	654	Metres	**0.86**	**0.71**	0.45	0.30	0.17	ETOPO1
Habitat exposure during Pleistocene low sea level stands	PLEIS	0.000	1.000	0.398	0.297	unitless[c]	**1.00**	**0.99**	**0.98**	**0.87**	0.35	Derived from ETOPO1

[a]Moran's *I* is a measure of spatial autocorrelation and can range from −1 (complete negative spatial autocorrelation) to +1 (complete positive spatial autocorrelation). Values were estimated from 10,000 random points and values above 0.70 (high spatial autocorrelation) are in bold.
[b]Mean frequency of thermal stress anomalies ≥ 1 °C over the previous 52 weeks.
[c]SSS: g/kg seawater; PLEIS: proportion.

Figure 6: Scales of autocorrelation with Moran's I - Riginos et al 2016



Brown et al. 2023

Figure 7: Remote sensing products may have a biologically irrelevant grain size

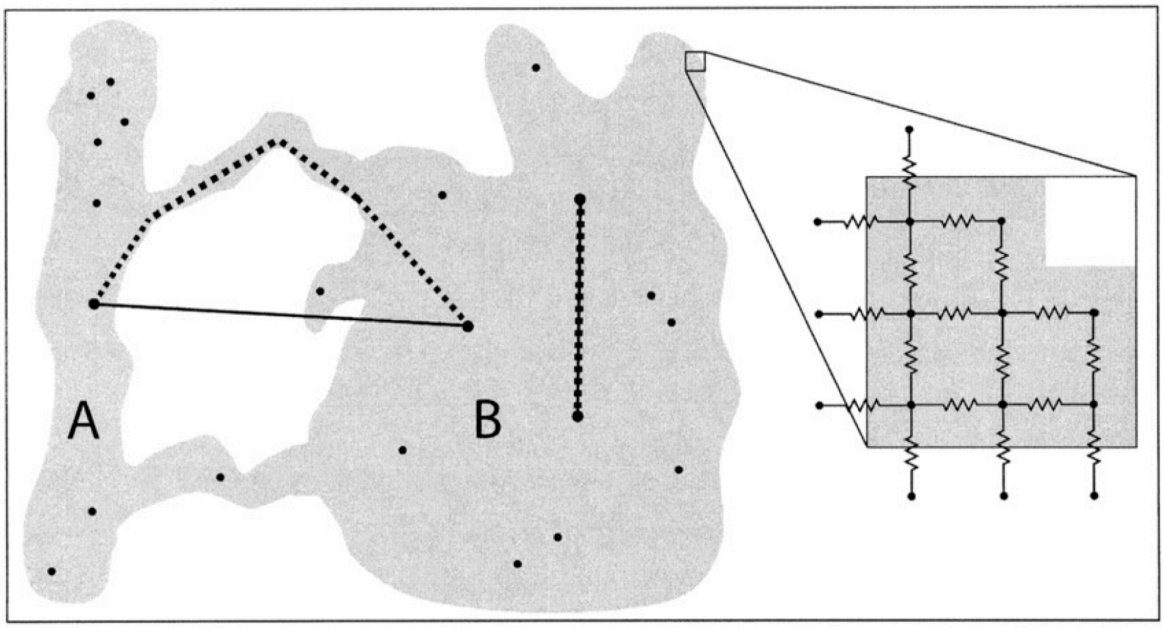


Figure 8: Static terrestrial matrix - McRae 2003

Temporal variability in planktonic dispersal - Watson et al 2012

Describing relationships between sampling sites

Relationship of LG to landscape ecology

Many of the methods used in landscape genetics/genomics have their origins in spatial (landscape) ecology.

- Methods papers and supporting documentation for analyses are likely to have species as the unit of inference - usually you can replace "species" with "loci"
- Also, get used to thinking about genetic diversity in terms of alpha and beta diversity
- Methodological inspirations for landscape genomics often come from landscape ecology, especially for describing spatial structure. Searching this literature for solutions and inspiration can be fruitful.
- Tools borrowed from landscape ecology help move between different types of data and analyses

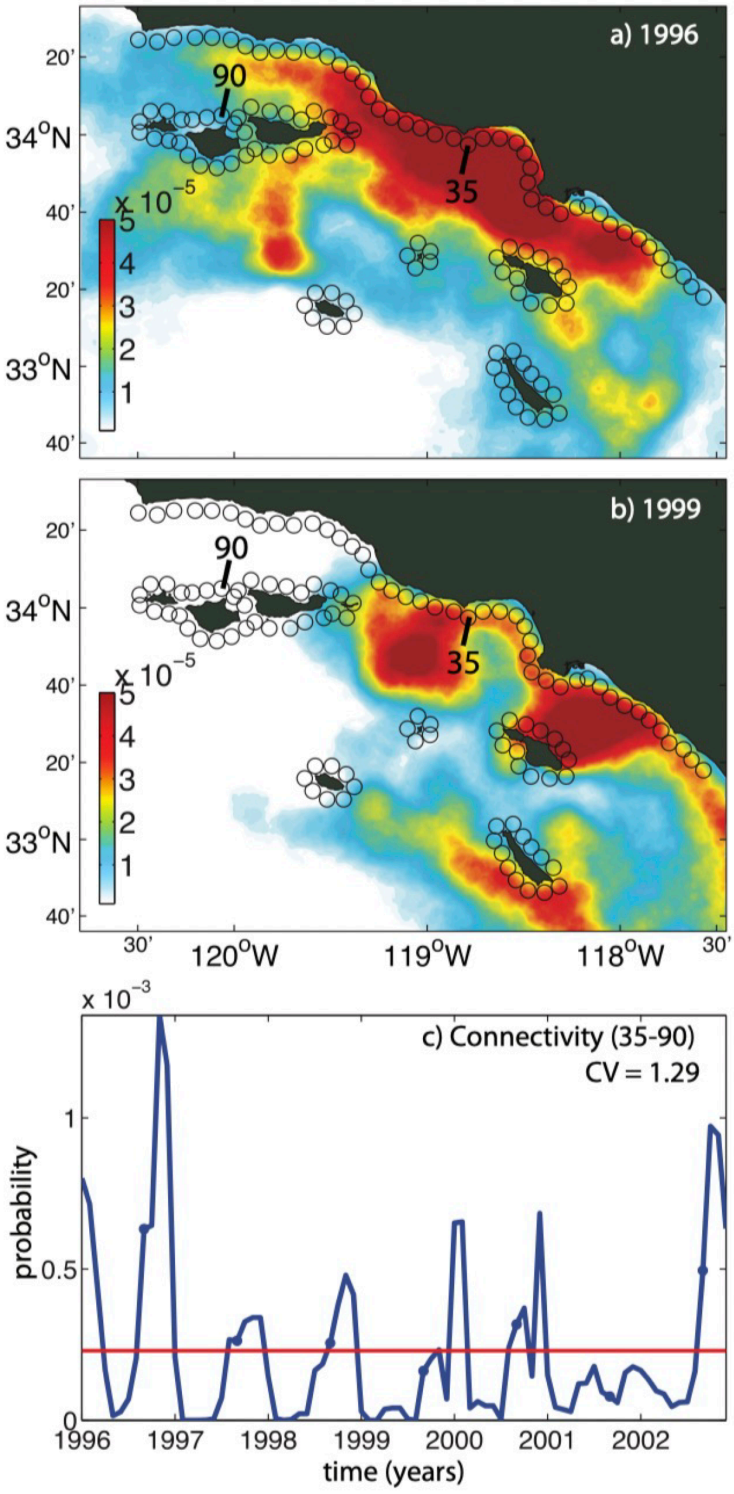


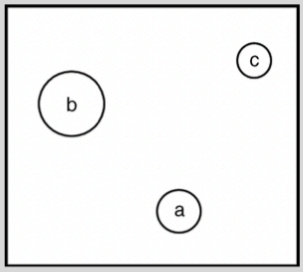
Figure 9: Temporally variable dispersal through the marine matrix

Box 5.3 Analytical levels

The main approaches of landscape genetics studies can be classified into four analytical levels. The following illustrations are adapted from Wagner and Fortin (2013).

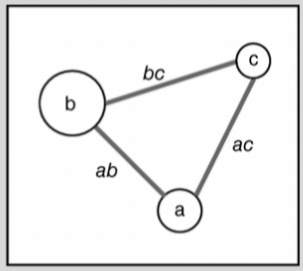
1 Node-level analysis

This relates adaptive variation to local landscape factors at sites *a*, *b*, and *c* while accounting for isolation-by-distance (Schoville et al. 2012). The node-level methods include multivariate ordination methods (e.g., RDA; Dray et al. 2012; Manel et al. 2012) and general linear models (Bolker 2008).



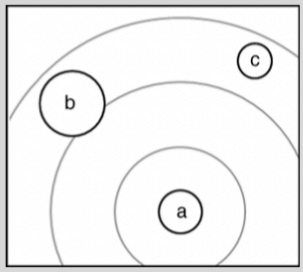
2 Link-level analysis

This relates neutral variation between sites *a*, *b*, and *c* to between-site landscape factors observed along links *ab*, *ac*, and *bc* to test hypotheses on isolation-by-distance (IBD), isolation-by-resistance (IBR), or isolation-by-barrier (IBB). The most commonly used link-level method is the Mantel test (Mantel 1967; Smouse et al. 1986; Cushman & Landguth 2010), which for multiple predictors extends to multiple regression on distance matrices (MRMs) (Smouse et al. 1986). Partial Mantel tests (Smouse et al. 1986) and causal modeling (Cushman et al. 2006) have been used to account for one process (e.g., IBD) while testing for another process (e.g., IBR). However, several studies have showed the relative lower power of the Mantel test to detect significant relationships and other inferential problems (Dutilleul et al. 2000: Legendre & Fortin 2010: Guillot & Rousset 2013).



3 Neighborhood-level analysis

This relates the relative contribution of all neighboring sampled locations (here *b* and *c*) to the genetic variation observed at a given sampling location *a*. Connectivity measures (Keyghobadi et al. 2005; James et al. 2011) and gravity models (Murphy et al. 2010) can be used in neighborhood-level analyses to assess neighborhood effects on spatial genetic structure.



4 Boundary-level analysis

This relates genetic groups *a*, *b*, and *c* to landscape barriers. Once spatial groups are identified based on either Bayesian clustering algorithms or edge-detection techniques (see Chapter 7; Guillot et al. 2005; François & Durand 2010; Safner et al. 2011), the next step is to relate these genetic barriers to environmental and landscape barriers using spatial boundary overlap methods (Fortin et al. 1996) or POPS (Prediction of Population genetic Structure Program) (Jay 2011).

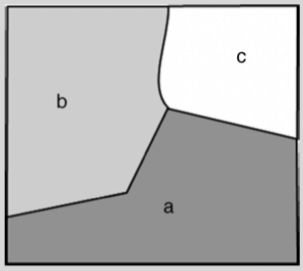


Figure 10: Describing relationships between populations (or individuals) - Wagner & Fortin 2016

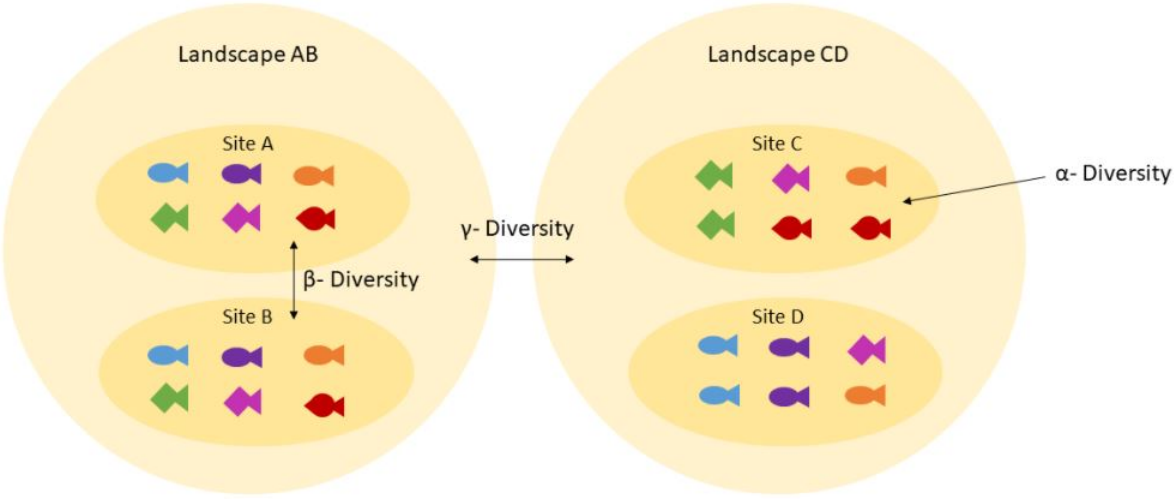


Figure 11: "7: Alpha, Beta, and Gamma Diversity." Biology LibreTexts, Libretexts, 10 Sept. 2021.

Alpha and beta diversity:

Statistical models for representing relationships among and between locations:

Sampling strategies and goals

Key questions:

- Is your study exploratory or testing specific hypotheses?

Exploratory

- Try to sample across the landscape
  - Random
  - Stratified (break up spatial correlations among environmental variables)
  - Across "hotspots" of environmental turnover

Hypothesis testing

- Interested in one environmental variable?

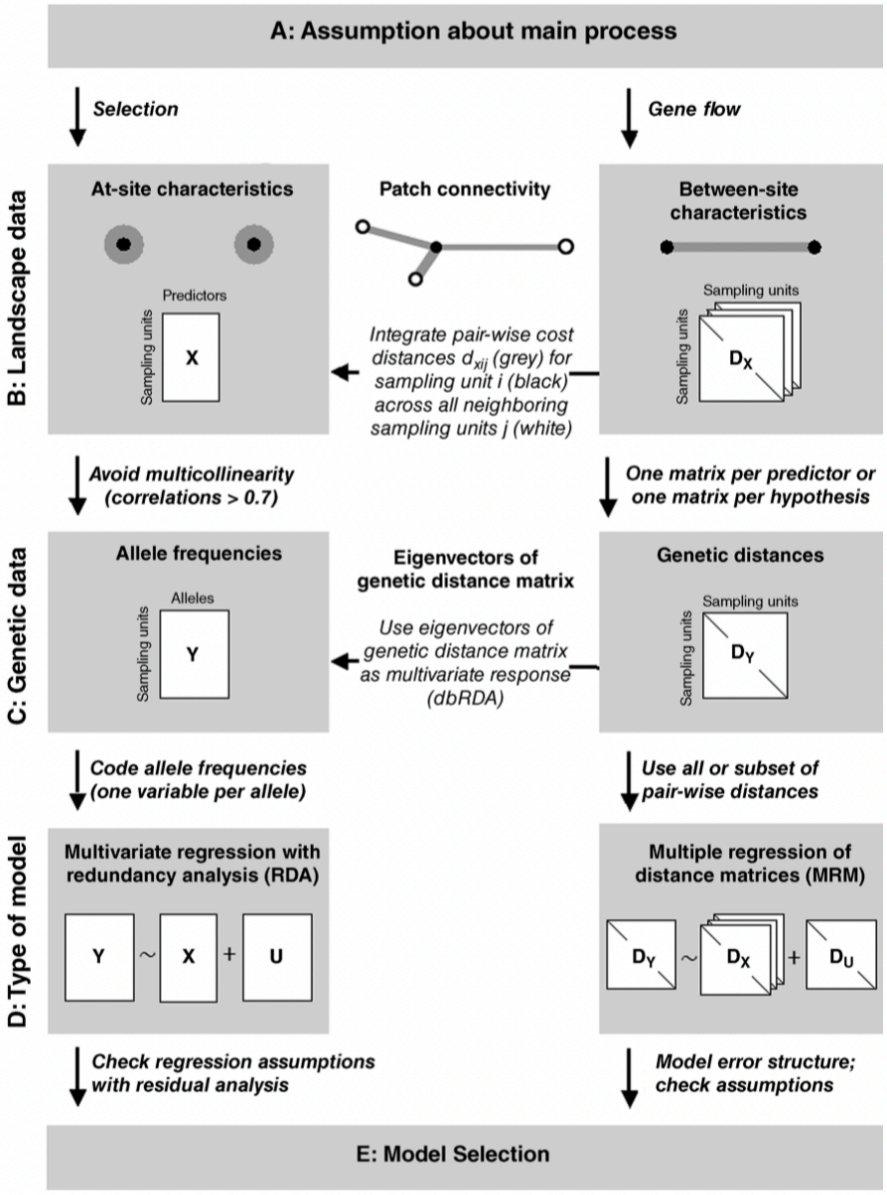


Fig. 5.1 Flowchart of the statistical model that can be used to relate genetic to landscape data depending on whether one assumes selection or gene flow to be the main underlying evolutionary process. In either case five steps are needed. (A) Determining implicitly or explicitly the main assumptions of the processes. (B) Determining how the landscape data will be analyzed. (C) Determining how the genetic data will be analyzed. (D) Selecting the appropriate regression framework. (E) Selecting the appropriate model.

Figure 12: Statistical models - Wagner & Fortin 2016

  - Pairs or gradients (but look at co-varying environmental factors, apply stratification concepts)

- Candidate loci have been independently identified?
  - What is the relevant environmental variable and can you design sampling to break up correlations with other variables (stratification, again)?

(In reality, most landscape genomic studies sample opportunistically and try to deal with spatial correlations at the analytical stage.)

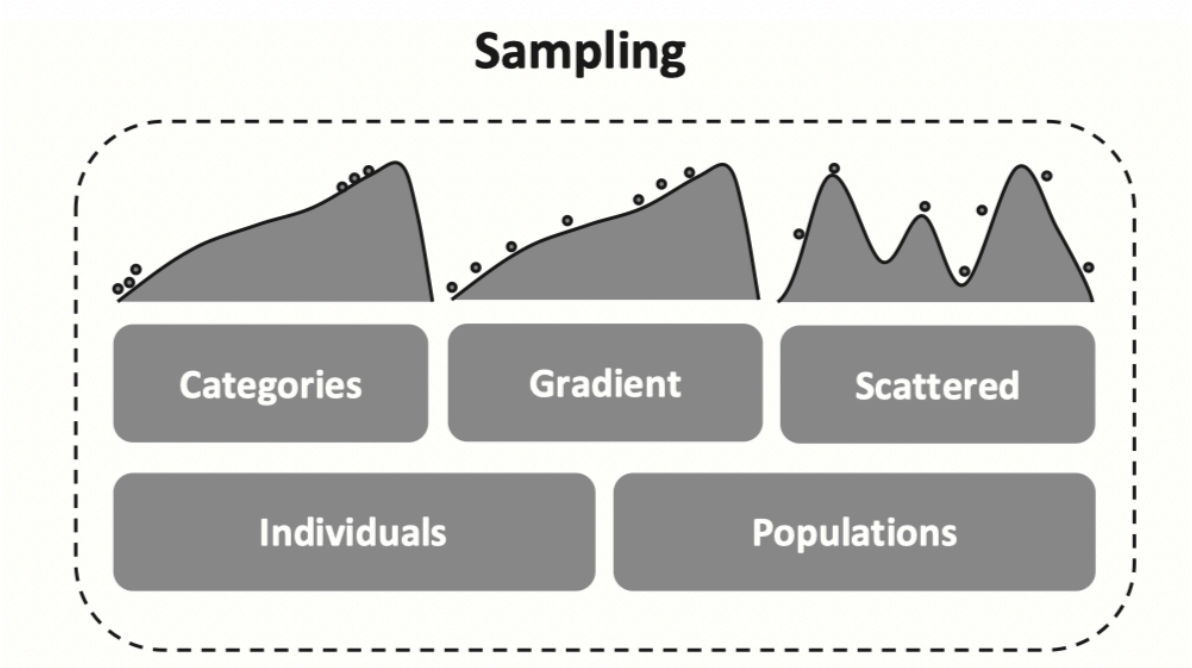


Figure 13: Sampling for different purposes - Rellstab et al 2015

Only 3.4% of studies have used paired sampling! (Dauphin et al 2023)

Conceptual example of stratified sampling

Sampling across shifts in multidimensional environments for exploratory studies

Ubiquitous problems in LG

- Methods are biased to find few genes of large effect and yet most traits are likely polygenic

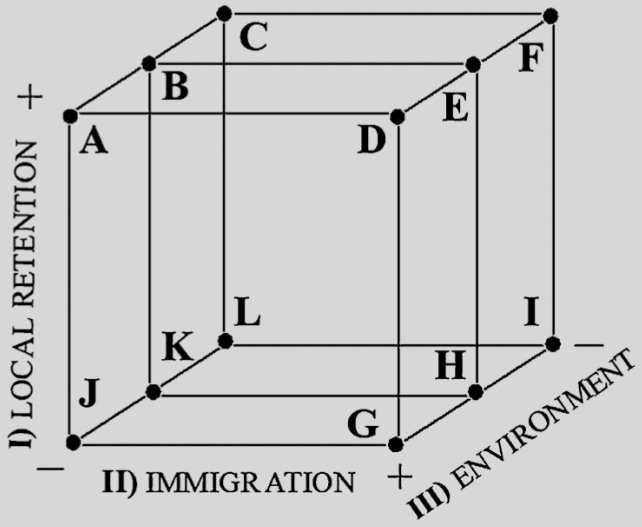


Fig. 2 Demographic-environment multivariate parameter space. A to L are our "strata" and represent strategic contrasts in the parameter values for the drivers we are interested in: i.e. (I) local retention, (II) immigration, and (III) environmental change

Figure 14: Stratified sampling - Liggins et al 2019

- Outlier methods are biased to find false positives when there is underlying population structure
- Collinearity of environmental variables makes moving from association to causation impossible without experiments

A more detailed workflow:

Activity 3

Update your poster in light of the class discussions.

Is your study exploratory or hypothesis testing?

Could you modify your study design to align to your goals?

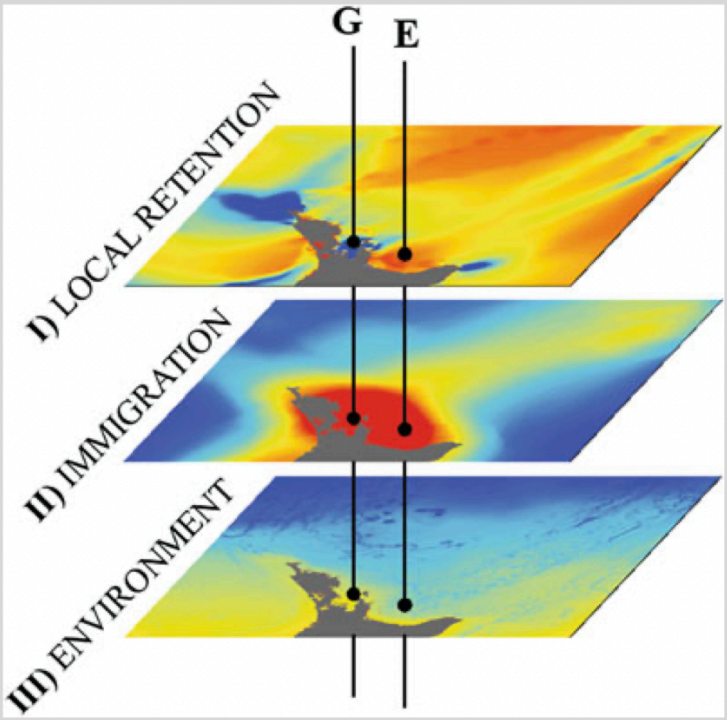


Fig. 3 Hypothetical samples of our strata (i.e. demographic-environment contrasts). With reference to Fig. 2, G represents a patch where local retention is low (I), immigration is high (II), and environmental change has been high (III). In contrast, E represents a patch where both local retention and immigration are high, and environmental change has been moderate. Replicate samples would be taken for each stratum (A to L)

Figure 15: Stratified sampling - Liggins et al 2019

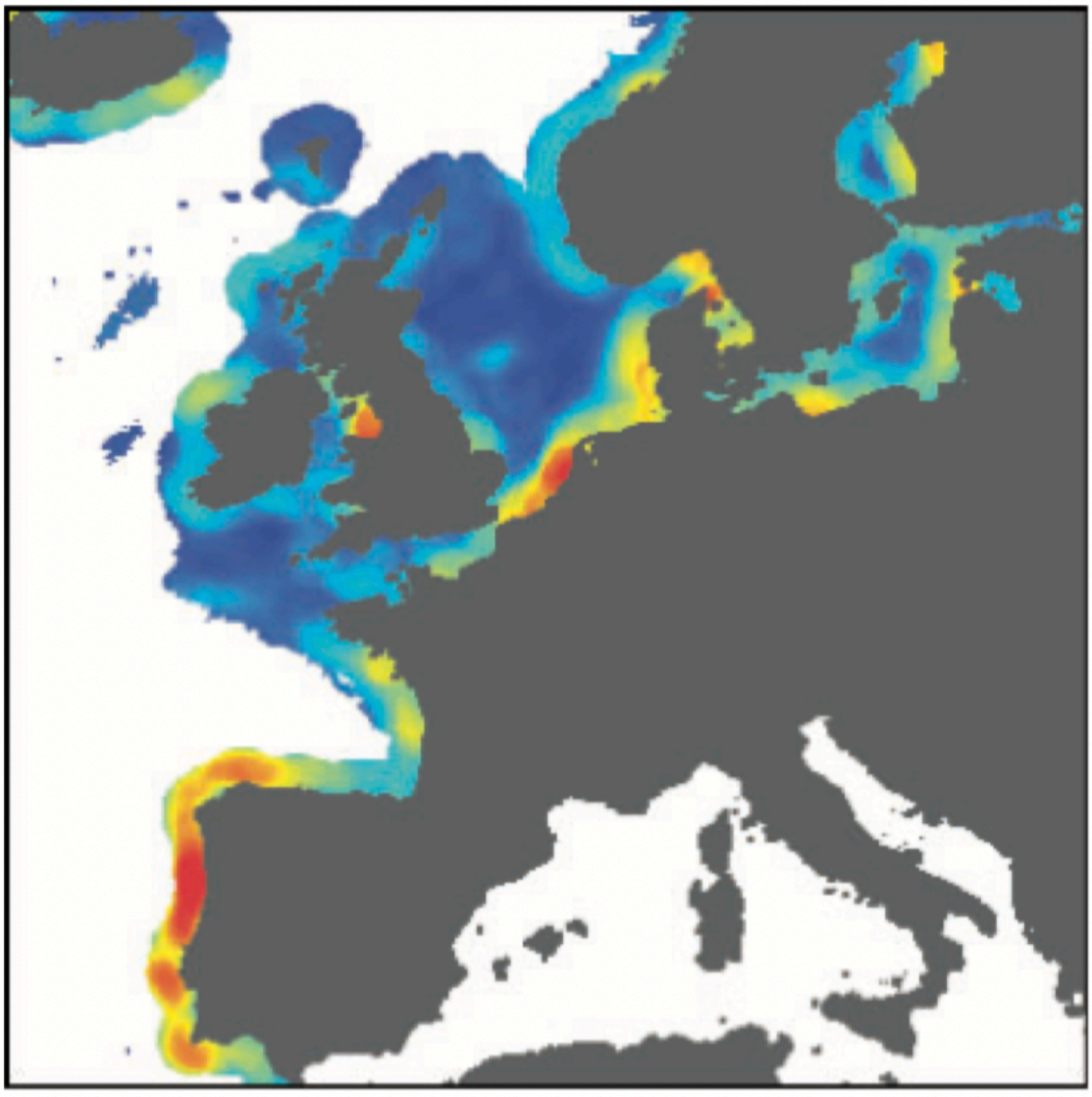


Figure 16: “Hotspots” of environmental change - Riginos et al 2016

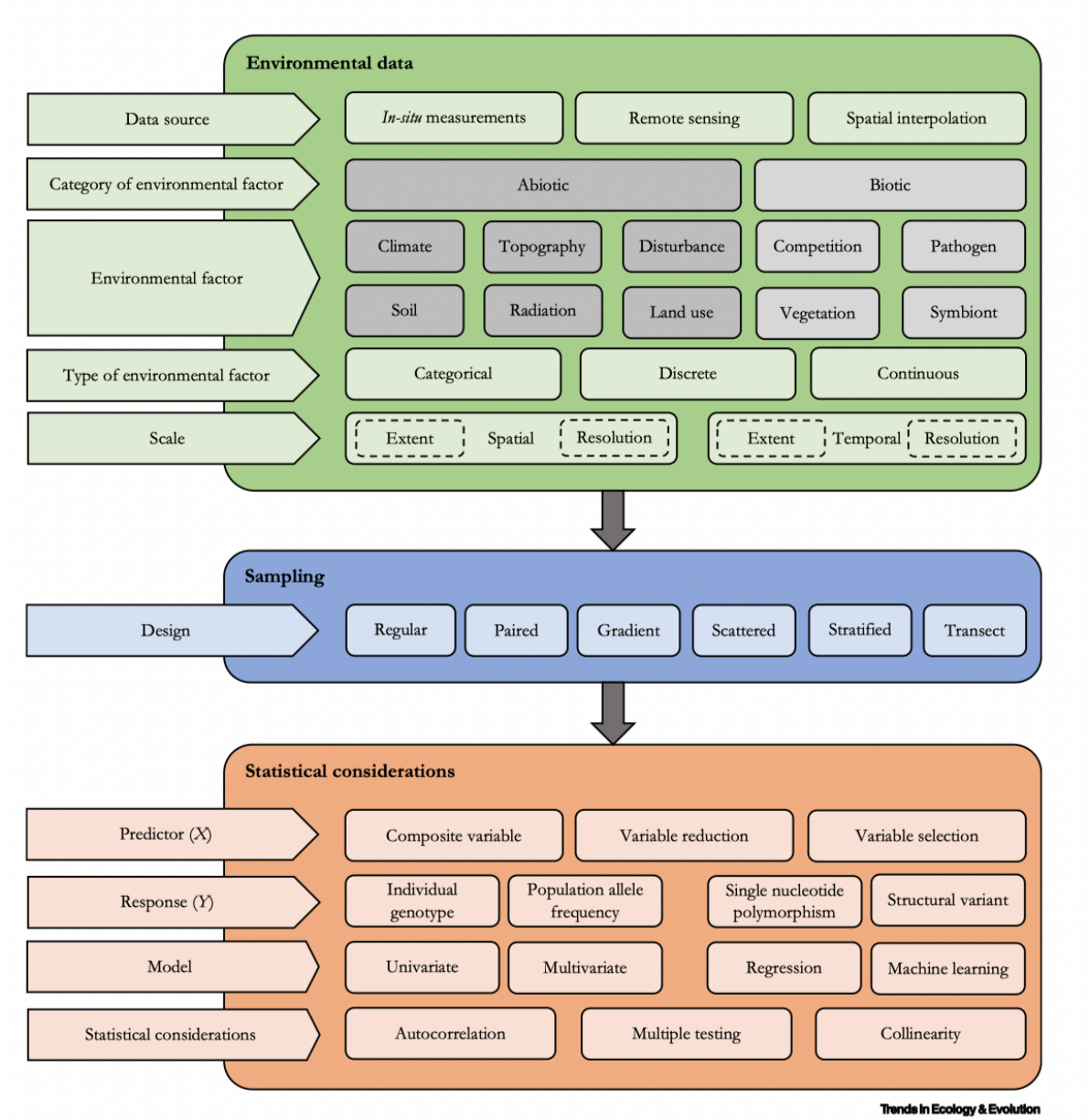


Figure 2. Overview of the workflow and important decision steps in the use of environmental data in landscape genomic studies. Note that for each step, the suggestions are not exhaustive, but only common examples are given. Steps and options are described in more detail in the main text.

Figure 17: Workflow for GEA and landscape genomics - Dauphin et al 2023

References cited in this tutorial

- Balkenhol, N., Dudaniec, R. Y., Krutovsky, K. V., Johnson, J. S., Cairns, D. M., Segelbacher, G., Selkoe, K. A., von der Heyden, S., Wang, I. J., Selmoni, O., & Joost, S. (2016). Landscape Genomics: Understanding Relationships Between Environmental Heterogeneity and Genomic Characteristics of Populations. In *Population Genomics* (Vol. 54, pp. 261-322). Cham: Springer International Publishing.
- Dauphin, B., Rellstab, C., Wuest, R. O., Karger, D. N., Holderegger, R., Gugerli, F., & Manel, S. (2023). Re-thinking the environment in landscape genomics. Trends Ecol Evol, 38(3), 261-274. doi:10.1016/j.tree.2022.10.010
- Dyer, R. J. (2015). Is there such a thing as landscape genetics? *Molecular Ecology, 24*(14), 3518-3528.
- Epperson, B.K. (2003) Geographical Genetics, Princeton University Press as origin of landscape genetics
- Liggins, L., Treml, E. A., & Riginos, C. (2019). Seascape genomics: contextualizing adaptive and neutral genomic variation in the ocean environment.
- Manel, S., Schwartz, M. K., Luikart, G., & Taberlet, P. (2003). Landscape genetics: combining landscape ecology and population genetics. *Trends in ecology & evolution, 18*(4), 189-197.
- Riginos, C., Crandall, E. D., Liggins, L., Bongaerts, P., & Treml, E. A. (2016). Navigating the currents of seascape genomics: how spatial analyses can augment population genomic studies. Current Zoology, 62(6), 581-601.
- Rellstab, C., Gugerli, F., Eckert, A. J., Hancock, A. M., & Holderegger, R. (2015). A practical guide to environmental association analysis in landscape genomics. Molecular Ecology, 24(17), 4348-4370.
- Storfer, A., Patton, A., & Fraik, A. K. (2018). Navigating the interface between landscape genetics and landscape genomics. *Front Genet, 9*, 68.
- Wagner, H. H., & Fortin, P. D. M.-J. (2016). Basics of spatial data analysis: linking landscape and genetic data for landscape genetic studies. In N. Balkenhol, S. A. Cushman, A. Storfer, & L. P. Waits (Eds.), Landscape Genetics: Concepts, Methods, Applications (pp. 1-22): John Wiley & Sons, Ltd.
- Wang, I. J., & Bradburd, G. S. (2014). Isolation by environment. Molecular Ecology, 23(23), 5649-5662.

Further readings - books

- Landscape Genetics: Concepts, Methods and Applications, 2016. Edited by Balkenhol, Cushman, Storfer & Waits. Wiley Blackwell.
- Population genomics: Marine organisms. 2020. Oleksiak, Marjorie F., and Om P. Rajora, eds. Springer.