**Detecting exceptional temporal changes in genetic diversity using limited information.**

Julian Wittische, Pierre Legendre, Patrick M. A. James

Département de Sciences Biologiques, CP 6128 Succursale Centre-Ville, Université de Montréal, Pavillon Marie-Victorin, Montréal, QC, Canada, H3C 3J7

Correspondence: Julian Wittische; E-mail: [jwittische@gmail.com](mailto:jwittische@gmail.com)

Running title: Unique spatio-temporal genetic change

**INTRODUCTION**

Global change, including climate change as well as habitat destruction and fragmentation, have caused biodiversity to quickly decline in many parts of the world in the last century (Butchart et al., 2010; Dirzo et al., 2014; Fischer & Lindenmayer, 2007). The future of biodiversity could be bleak (Bellard, Bertelsmeier, Leadley, Thuiller, & Courchamp, 2012), and thus there is an ever-increasing demand from ecosystem managers to evaluate and mitigate biodiversity loss, and to assess current and proposed management plans (Brondizio, Settele, Díaz, & Ngo, 2019). Global change trends in biodiversity and ecosystem functioning, and associated temporal uncertainty and spatial bias, have been closely monitored (Gonzalez et al., 2016; Pereira et al., 2010; Sala et al., 2000; Scholes et al., 2008). Finer scale ecological research is needed to further our ability to predict change at the global scale (Mouquet et al., 2015; Pereira et al., 2010; Randin et al., 2009; Yates et al., 2018). To meet these goals, researchers require tools and methods to improve our understanding of the drivers of biodiversity loss and to make meaningful and reliable forecasts.

Spatial and temporal variation in genetic information can tell us a great deal about demography and movement of populations. Landscape genetics approaches are, and will continue to be, widely used for conservation biology purposes as the approach that is used to translate these genetic data into meaningful inference (Allendorf, Hohenlohe, & Luikart, 2010; Harrisson, Pavlova, Telonis-Scott, & Sunnucks, 2014; Segelbacher et al., 2010). Indeed, landscape genetics bridges an important gap in the field of molecular ecology: providing information about the interaction between micro-evolutionary processes and landscape features (Balkenhol, Cushman, Storfer, & Waits, 2015; Manel & Holderegger, 2013; Manel, Schwartz, Luikart, & Taberlet, 2003; Wagner & Fortin, 2013). Landscape genetics can therefore help us address a wide array of questions, such as how gene flow, and therefore effective dispersal (Bohonak, 1999; Clobert, Le Galliard, Cote, Meylan, & Massot, 2009), is affected by environmental heterogeneity (e.g. Wittische et al. 2019), how local landscape characteristics explain the spatial distribution of neutral and adaptive genetic information (e.g. Janes et al., 2014), or even how to locate genetic boundaries.

One of the main ongoing challenges for biologists, is to detect and predict where and when demographic event events influence the ecological dynamics and the evolution of species. Changes in genetic diversity can be the result of natural or anthropogenic change at any temporal scale, from a local and abrupt change like a wildfire to a global change like climate warming (Manel & Holderegger, 2013). However, it is rarely possible to observe the effects of these events instantaneously and researchers are often left with spatial legacies in xYZ… which may not be readily observable from demographic data alone. When a demographic event does not constitute a selective pressure, alleles are randomly transferred from a generation to the next and genetic drift happens leading to a loss of diversity. Common examples of situations where genetic drift occurs include geographic isolation, population bottleneck and massive migrations from previously isolated populations, which would substantially reduce or alter local genetic variation. The result of such events in a local population tend to alter the genetic distance of this population with surrounding populations (Segelbacher et al., 2010). Detecting changes in the genetic make-up of a population through time, including the nature of those changes, may describe what is happening at the demographic level, and therefore serve as an alarm for managers.

Few methods currently exist for the temporal comparison of multivariate community and genetic data. For example… (*now go through existing tools – Then provide critique*). Very sophisticated methods exist to infer demographic history from genetic data, even from static genetic data. However, those methods are designed for very large genetic datasets with tens of thousands of loci and known ascertainment. It seems to me that you are trying to identify demographic histories using the TBI approach… right?  
Other methods do exist with which one can infer demographic history using (static) genetic data  
<https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1003905> The relevance and performance of traditional approaches (e.g. PCA-based) to test change using temporal genetic datasets, where the objective is to find which population has indeed changed more significantly than others in the landscape, has not been evaluated. The rapid pace of global loss of genetic diversity (Leigh, Hendry, Vázquez‐Domínguez, & Friesen, 2019), is making it increasingly important to move beyond, single sampling/time, snapshot research (Draheim, Moore, Fortin, & Scribner, 2018).

A permutation-based statistical inference method for the analysis of spatial-temporal changes in community composition have recently been proposed (Legendre & Gauthier, 2014; Shimadzu, Dornelas, & Magurran, 2015). Temporal Beta-diversity Indices (TBI; Legendre 2018) were designed to assess whether there are sites where the difference in community composition between survey times seems exceptionally large. This approach has not yet been tested nor applied to the question of temporal variation in genetic data. The method involves estimating temporal change in each sampling site between two dates using a dissimilarity index/distance, testing the significance of each change through permutations, and partitioning the change into losses and gains. Comparing genetic data at two different dates, whether or not they were separated by an *a priori* known event, may help us understand more about the ecological processes shaping the system. A strong genetic change would also indicate the parts of the landscape where an event had the strongest effects, or highlight which sites should be investigated if managers are not aware of an *a priori* known event.

In this chapter, I develop and describe a method to identify locations that have undergone significant genetic change through time. Identify such locations, and quantifying other locations relative temporal genetic change, is important because… ??? . To demonstrate the effectiveness and applicability of the approach we use both empirical and simulation data. Empricial data represent… Simualtion data were generated using… We simulated scenarios where part of the landscape is affected by non-selective demographic changes mimicking the effects of common demographic event events. We then used TBI to measure changes in genetic make-up of our populations, and evaluated the power and error rates associated with this approach. Finally, we illustrated the possibilities of this approach through applications on two real genetic datasets. We predict that dispersal will affect our ability to detect the genetic legacies of an event, we predict that the higher the number of populations affected by extraordinary events, the lower the performance of the TBI testing procedure, and finally we predict that the longer the time between samplings, the harder it will be to identify where and when a demographic event occurred.

**METHODS**

*Simulation framework*

To simulate changes in genetic information through time, we used the spatially-explicit gene flow simulation software CDMetaPOP (Landguth, Bearlin, Day, & Dunham, 2017). CDMetaPOP simulates dispersal and mating of individuals across a landscape, and allows to define the initial genetic structure, spatial distribution of individuals, dispersal characteristics, and life history traits of the population.

Loci were modelled after single nuclear polymorphism (SNP) and therefore are bi-allelic. The mutation rate was set as to reflect empirically-derived mutation rates found in many taxa (REF I gave to Ryan). Simulated individuals each carried a genome of 100 neutral loci without linkage disequilibrium. Each simulated population in the landscape had a maximum carrying capacity of 50 individuals, and each simulated landscape comprised 25 (a grid of 5 by 5) interconnected such populations with structural connectivity only reflecting geographical distance. That corresponds to a maximum of 1250 individuals in the landscape. Each simulation was run for 100 generations before a demographic event was forced on up to three populations in the landscape. 10 more generations were simulated after the event.

We simulated 180 replicates for each scenario, with the new allocation of allelic frequencies for each replicate. Those parameters were chosen as a compromise between realism and computational time limitations, and we believe they were appropriate to produce the complex evolutionary dynamics necessary to reasonably realistic and useful genetic data.

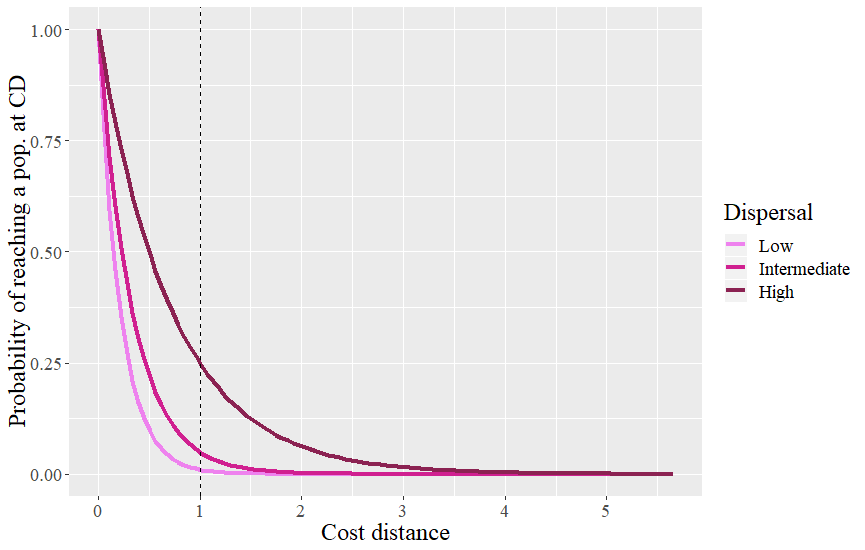
We examined the influence of dispersal, demographic event type, and demographic event spatial extent on the persistence of genetic spatial legacies using this simulation model. With 3 dispersal regimes, 2 different demographic event types and 3 different numbers of populations affected, we have 18 different scenarios giving us a total of 3240 (18 x 180) simulations. In the next sections, we detail how we modelled the aforementioned three factors. Additional material concerning the use of the simulator can be found in Supplementary Materials.

*Dispersal regimes*

The dispersal of individuals was controlled through a dispersal kernel based on a negative exponential distribution from which the distance realised by an individual () is sampled. This distribution transforms the cost distance of travel (CD) between cells according to a single parameter (*B*): . Cost distances used here are simply the geographical distances between the centroids of the populations. The values created through the use of the negative exponential distribution can then be rescaled using the maximum and the minimum distance possible in the landscape, which gives us the probability that an individual reaches a population beyond a certain CD:

When is randomly sampled as being higher than 1, the target population to which an individual travels, was selected randomly from the set of populations available at the distance selected in the previous step. Otherwise, the individual stays within its original population. We chose this way of modelling dispersal so that most individuals stay within their original population, that is more individuals randomly travel a distance below 1 than higher, while keeping opportunities for occasional long distance dispersal. This holds advantages compared to simpler approaches such as nearest neighbours or linear probability (REF).

In order to investigate the effect of different levels of dispersal, we changed the dispersal kernel by choosing values of *B* which would give us low, intermediate and high dispersal (Fig.1). We considered the % of individuals reaching an adjacent population as an indicator of the intensity of dispersal. We therefore respectively chose 1% (*B* = 2), 5% (*B* = 1.301), and 25% (*B* = 0.6015).



**Fig.1:** Probability of dispersal of an individual at all possible distances in the landscape, for three different dispersal scenarios. If an individual disperse below a distance of 1 (dashed line), then it does not leave its original population.

*Demographic events design*

The first demographic event we considered involves modelling a massive extraneous immigration from a previously isolated population otherwise sharing the same characteristics as other populations. This population was simulated during the same number of generations and the cost distance from the isolated population to the target population(s) and was set to 0 between the 100th and 101st generations, mimicking a mass immigration event between the two. The cost distances are then set back to normal.

The second scenario involves a demographic bottleneck through massive mortality. To do that, the carrying capacity of the disturbed population was set to 20% of its original value between the 100th and 101st generations.

*Number and position of target populations*

Beyond the dispersal intensity and the demographic event type, we wanted to evaluate how the number of target populations affected the performance of our testing procedure. To achieve this, we disturbed from 1 to 3 populations among the 25. When only 1 population was disturbed we partitioned the 180 replicates of that scenario equally among 6 populations in the landscape. Because our landscape is homogenous and symmetric, only 6 positions need to be assessed (Fig YYY). When several (*k*) populations were disturbed, we randomly sampled 1 position among the 6 previously described and randomly picked 1 or 2 additional populations directly adjacent (when possible) to it. We did this 6 times (30 replicates for each set of targeted populations). We choose to pick target populations this way to respect a degree of spatial autocorrelation often exhibited in demographic events.

*Genetic dissimilarity*

The Chord distance has been commonly used in both community ecology (Orlóci 1967; Legendre & Borcard 2018) and population genetics (Cavalli-Sforza & Edwards 1967; Balkenhol et al. 2016). We chose chord distance because it has already been tested for use with TBI with non-genetic data (Legendre 2019) and because it may be more appropriate than other indices of genetic dissimilarity when most of the variation among populations is due to recent changes (Takezaki & Nei 1996; Kalinowski 2002) as it does not assume populations are in drift-mutation equilibrium. Here we use the Chord distance to calculate genetic dissimilarity of a single site sampled at two different points in (simulated) time.

*Permutation approaches*

xxx. Most calculations used in this paper are based on the *TBI()* function available in the *R* package *adespatial* (Dray et al., 2019). Three permutation approaches were considered to test the significance of TBI, but only one was kept in the final version of the *TBI()* function (Legendre, 2019). Because they were tested on very a different type of data, we used an older version of the *TBI()* function to tests which one should be kept for genetic data. The first permutation approach consisted in permuting a locus in the same way in both (original sampling and resampling) gene frequency data frames. The second permutation approach consisted in permuting a locus independently in both data frames. The third permutation approach consisted in permuting sampling sites in both data frames. We summarized statistical performance per permutation approach, and used the best approach to answer all other questions.

*Statistical performance*

We used the False Positive Rate (FPR) and False Negative Rate (FNR) frameworks to assess statistical performance of the TBI testing procedure and to evaluate which of the permutation procedures, and permutation p.value thresholds, is most appropriate. A false positive is a population that we *a priori* know did not undergo any specific demographic event, but has been classified as having experienced one of the two simulated demographic events by the testing procedure. A false negative is a population that we had set as target for demographic event but that was not classified as having been disturbed by the testing procedure. FPR represents the number of false positives over the total number of negatives, and FNR represents the number of false negative over the total number of positives. A high FPR means that we often select the wrong population(s). A high FNR means that we often miss the right population(s). The higher the FNR, the lower the power of our testing procedure. Because choosing a proper threshold for the TBI permutation tests is important in order to find a compromise between power and selectivity, we evaluated statistical performance across a range of thresholds: from 0.001 to 0.15.

**RESULTS**

Table 1. Mean and standard deviation of FPR and FNR associated with TBI permutation tests using three different permutation approaches, at two adjusted p-values thresholds.

From Table 1 we can see that only the first permutation approach, that is permuting a locus in the same way in matrices from both samplings, gives usable FNR and FPR. This is true regardless of the threshold value chosen for the adjusted p-value. The first approach is the one suitable for genetic data and was therefore chosen for the rest of the analyses.

Table 2. Mean and standard deviation of FPR and FNR associated with TBI permutation tests at different adjusted p-values thresholds. The first permutation approach was used.

Stricter values for the adjusted p-value threshold (0.001 and 0.01) expectedly bring a better FPR but also bring a pathological FNR (no power). From 0.025 and up, the FPR increases, eventually reaching the value of the higher adjusted p-value thresholds (0.1 and 0.15), whereas the power increases continuously. The increase in mean power is accompanied by a decrease of the associated standard variation, whereas the increase in mean FPR concurrent with an increasing of its variation. Depending on the relative costs and benefits or FP and FN, thresholds above 0.025 are suitable.

Table 3. Mean and standard deviation of FPR and FNR associated with TBI permutation tests using four different post-event time lags, at an adjusted p-values threshold of 0.1. The first permutation approach was used.

As shown in Table 2, it seems that so far, in our high-dispersion bottleneck simulations, we miss the right population where the event occurred after 1 generation of random mating. In other words, demographic processes are quickly diluting the signal by transferring the initial effect on genetic diversity to other populations.

Table 4. Mean and standard deviation of FPR and FNR associated with TBI permutation tests using six different time lags between samplings, at an adjusted p-values threshold of 0.025. The first permutation approach was used.

We can see in Table 3 that the longest the pre-event sampling is from the event, the less power and the more false positives we get. In 6 years the performance as measured by FPR and power has almost been halved. The increase in mean FPR does not seem to be associated with a similar increase in variation, whereas the increase of FNR is associated with an increase in variation.

**DISCUSSION**

*(given the limited [only one scenario/one dispersal] set of simulations I have)*

**TBI is applicable to genetic data under certain conditions.**

* The first permutation approach (same line swap across both genetic distance matrices) is the only suitable approach when using gene frequency data.
* Although signal of a past demographic event can be kept in richer genomic data (e.g. probability of mutational configurations in sequence blocks), gene frequency data in a high-dispersion species and connected landscape will not keep the signal beyond a year. To be investigated for non-SNP gene frequency data.
* Provided the landscape was resampled the year of the event, the closer the date of the first sampling, the better performance-wise. However, reasonable performance can be expected even if the first sampling was a few years before the event.
* Liberal thresholds should be used in order to obtain reasonable statistical power, while keeping the FPR low.

**Paragraph discussing the influence of dispersal on the detectability of demographic event from gene frequency data.**

**Paragraph discussing the influence of the fraction of the landscape affected by a demographic event on the detectability of demographic event from gene frequency data.**

**Paragraph discussing how we were able to use TBI to test and illustrate temporal change** (results for empirical data not ready yet)**.**

**Paragraph discussing the limits of TBI use on genetic data, including the fact that it may need to be parameterized (e.g. choosing a threshold) based on landscape or taxa characteristics.**

**Paragraph discussing further investigation of the relative importance of genetic drift, gene flow and other forces, in shaping temporal variation.**

**Paragraph discussing the importance of LTER, exhaustive sampling, and the need to move beyond single-time snapshot studies of landscape genetics.**

**REFERENCES**

Allendorf, F. W., Hohenlohe, P. A., & Luikart, G. (2010). Genomics and the future of conservation genetics. *Nature Reviews. Genetics*, *11*(10), 697–709. https://doi.org/10.1038/nrg2844

Balkenhol, N., Cushman, S., Storfer, A., & Waits, L. (2015). *Landscape Genetics: Concepts, Methods, Applications*. Wiley-Blackwell.

Bellard, C., Bertelsmeier, C., Leadley, P., Thuiller, W., & Courchamp, F. (2012). Impacts of climate change on the future of biodiversity. *Ecology Letters*, *15*(4), 365–377. https://doi.org/10.1111/j.1461-0248.2011.01736.x

Bohonak, A. J. (1999). Dispersal, gene flow, and population structure. *Quarterly Review of Biology*, Vol. 74, pp. 21–45. https://doi.org/10.1086/392950

Brondizio, E. S., Settele, J., Díaz, S., & Ngo, H. T. (editors). (2019). *Global assessment report on biodiversity and ecosystem services of the Intergovernmental Science- Policy Platform on Biodiversity and Ecosystem Services. , Bonn, Germany.IPBES. 2019.* Bonn, Germany.

Butchart, S. H. M., Walpole, M., Collen, B., von Strien, A., Scharlemann, J. P. W., Almond, R. E. A., … Carr, G. M. (2010). Global Biodiversity : Indicators of Recent Declines. *Science*, *328*(1164). https://doi.org/10.1126/science.1187512

Clobert, J., Le Galliard, J. F., Cote, J., Meylan, S., & Massot, M. (2009). Informed dispersal, heterogeneity in animal dispersal syndromes and the dynamics of spatially structured populations. *Ecology Letters*, *12*(3), 197–209. https://doi.org/10.1111/j.1461-0248.2008.01267.x

Dirzo, R., Young, H. S., Galetti, M., Ceballos, G., Isaac, N. J. B., & Collen, B. (2014). Defaunation in the Antrhopocene. *Science*, *401*(6195), 401–406. https://doi.org/10.1126/science.1251817

Draheim, H. M., Moore, J. A., Fortin, M. J., & Scribner, K. T. (2018). Beyond the snapshot: Landscape genetic analysis of time series data reveal responses of American black bears to landscape change. *Evolutionary Applications*, *11*(8), 1219–1230. https://doi.org/10.1111/eva.12617

Fischer, J., & Lindenmayer, D. B. (2007). Landscape modification and habitat fragmentation: a synthesis Joern Fischer\* and David B. Lindenmayer Centre. *Global Ecology and Biogeography*, *15*, 55–66. https://doi.org/10.1111/j.1466-8238.2006.00287.x

Gonzalez, A., Cardinale, B. J., Allington, G. R. H., Byrnes, J., Endsley, K. A., Brown, D. G., … LoreAu, M. (2016). Estimating local biodiversity change : a critique of papers claiming no net loss of local diversity. *Ecology*, *97*(8), 1949–1960.

Harrisson, K. A., Pavlova, A., Telonis-Scott, M., & Sunnucks, P. (2014). Using genomics to characterize evolutionary potential for conservation of wild populations. *Evolutionary Applications*, *7*(9), 1008–1025. https://doi.org/10.1111/eva.12149

Janes, J. K., Li, Y., Keeling, C. I., Yuen, M. M. S. S., Boone, C. K., Cooke, J. E. K. K., … Sperling, F. A. H. H. (2014). How the mountain pine beetle (Dendroctonus ponderosae) breached the canadian rocky mountains. *Molecular Biology and Evolution*, *31*(7), 1803–1815. https://doi.org/10.1093/molbev/msu135

Landguth, E. L., Bearlin, A., Day, C. C., & Dunham, J. (2017). CDMetaPOP: an individual-based, eco-evolutionary model for spatially explicit simulation of landscape demogenetics. *Methods in Ecology and Evolution*, *8*(1), 4–11. https://doi.org/10.1111/2041-210X.12608

Legendre, P. (2019). A temporal beta-diversity index to identify sites that have changed in exceptional ways in space–time surveys. *Ecology and Evolution*, *9*(6), 3500–3514. https://doi.org/10.1002/ece3.4984

Legendre, P., & Gauthier, O. (2014). Statistical methods for temporal and space-time analysis of community composition data. *Proceedings of the Royal Society B: Biological Sciences*, *281*(1778). https://doi.org/10.1098/rspb.2013.2728

Leigh, D. M., Hendry, A. P., Vázquez‐Domínguez, E., & Friesen, V. L. (2019). Estimated six percent loss of genetic variation in wild populations since the industrial revolution. *Evolutionary Applications*, (April), 1–8. https://doi.org/10.1111/eva.12810

Manel, S., & Holderegger, R. (2013). Ten years of landscape genetics. *Trends in Ecology & Evolution*, *28*(10), 614–621. https://doi.org/10.1016/j.tree.2013.05.012

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. https://doi.org/10.1016/S0169-5347(03)00008-9

Mouquet, N., Lagadeuc, Y., Devictor, V., Doyen, L., Duputié, A., Eveillard, D., … Loreau, M. (2015). Predictive ecology in a changing world. *Journal of Applied Ecology*, *52*(5), 1293–1310. https://doi.org/10.1111/1365-2664.12482

Pereira, H. M., Leadley, P. W., Proença, V., Alkemade, R., Scharlemann, J. P. W., Fernandez-Manjarrés, J. F., … Walpole, M. (2010). Scenarios for Global Biodiversity in the 21st Century. *Science*, *330*(6010), 1496–1501. https://doi.org/10.1126/science.1196624

Randin, C. F., Engler, R., Normand, S., Zappa, M., Zimmermann, N. E., Pearman, P. B., … Guisan, A. (2009). Climate change and plant distribution: Local models predict high-elevation persistence. *Global Change Biology*, *15*(6), 1557–1569. https://doi.org/10.1111/j.1365-2486.2008.01766.x

Sala, O. E., Huber-Sanwald, E., Mooney, H. A., III, S. C., Laura F. Huenneke, ´n Oesterheld, Juan J. Armesto, N. L. P., … David M. Lodge, M. W. (2000). Global Biodiversity Scenarios for the Year 2100. *Science*, *287*(5459), 1770–1774. https://doi.org/10.1126/science.287.5459.1770

Scholes, A. R. J., Mace, G. M., Turner, W., Geller, G. N., Jürgens, N., Larigauderie, A., … Mooney, H. A. (2008). *Toward a Global Biodiversity*. *321*(5892), 1044–1045.

Segelbacher, G., Cushman, S. A., Epperson, B. K., Fortin, M. J., Francois, O., Hardy, O. J., … Manel, S. (2010). Applications of landscape genetics in conservation biology: Concepts and challenges. *Conservation Genetics*, *11*(2), 375–385. https://doi.org/10.1007/s10592-009-0044-5

Shimadzu, H., Dornelas, M., & Magurran, A. E. (2015). Measuring temporal turnover in ecological communities. *Methods in Ecology and Evolution*, *6*(12), 1384–1394. https://doi.org/10.1111/2041-210X.12438

Wagner, H. H., & Fortin, M.-J. (2013). A conceptual framework for the spatial analysis of landscape genetic data. *Conservation Genetics*, *14*(2), 253–261. https://doi.org/10.1007/s10592-012-0391-5

Yates, K. L., Bouchet, P. J., Caley, M. J., Mengersen, K., Randin, C. F., Parnell, S., … Sequeira, A. M. M. (2018). Outstanding Challenges in the Transferability of Ecological Models. *Trends in Ecology and Evolution*, *33*(10), 790–802. https://doi.org/10.1016/j.tree.2018.08.001