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

Julian Wittische1, Pierre Legendre1, Patrick M. A. James1,2

1 Département de Sciences Biologiques, Université de Montréal, Pavillon Marie-Victorin, Montréal, QC, Canada, H3C 3J7

2 John H. Daniels Faculty of Architecture, Landscape, and Design, University of Toronto, 33 Willcocks St., Toronto, ON, Canada, M5S 2J5

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

Running title: Testing spatio-temporal genetic change

**ABSTRACT**

**INTRODUCTION**

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). Spatial and temporal genetic variation can tell us a great deal about demography and population connectivity (Bradburd & Ralph, 2019; Lowe & Allendorf, 2010). Indeed, population genetics have proven to be essential to translating observed genetic variation into meaningful inferences regarding connectivity and demography that are necessary for conservation efforts. (Allendorf, Hohenlohe, & Luikart, 2010; Harrisson, Pavlova, Telonis-Scott, & Sunnucks, 2014; Segelbacher et al., 2010). In particular, landscape genetics provides 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). The field of landscape genetics has thus made enormous contributions to our understanding of how spatial heterogeneity influences population genetic processes. However, many situations such as outbreaks, invasions and species declines require researchers and managers to also look through a temporal lens.

Temporal variation in genetic diversity, and its drivers, are indeed at the crux of many conservation issues because they influence the evolution and persistence of a species through processes such as local adaptation, maladaptation, or divergent natural selection (Aeschbacher, Selby, Willis, & Coop, 2016; Bolnick & Nosil, 2007; Kremer et al., 2012). Comparing genetic data at two different dates, whether or not they were separated by an a priori known event, is a challenge because the genetic diversity of a population is dynamic, as beyond recombination and mutation constantly introducing new diversity, every demographic change creates genetic drift. Although it is rarely possible to directly observe the effects of landscape and climate change on spatial and temporal genetic variation, we may observe these effects through their population genetic legacies. Genetic legacies may not be detectable as rapidly as the demographic consequences of a landscape change for example but they may carry a signal of its effects for several generations (Bolliger, Lander, & Balkenhol, 2014; Epps & Keyghobadi, 2015). Researchers commonly use spatio-temporal population genetic legacies to study isolation-by-distance (Rousset, 1997; Wright, 1943), population bottlenecks (Gattepaille, Jakobsson, & Blum, 2013; Maruyama & Fuerstt, 1985), migration between isolated populations (Bezemer, Krauss, Roberts, & Hopper, 2019; Buschbom, Yanbaev, & Degen, 2011), and outbreak expansions (Larroque et al., 2019; Wittische, Janes, & James, 2019). Identifying meaningful and statistically significant relationships between temporal landscape-change and the spatial apportionment of genetic variation can give us important insights about the eco-evolutionary dynamics of a species, and be used to inform conservation strategies (e.g. Landguth, Holden, Mahalovich, & Cushman, 2017).

Spatio-temporal population genetics methods to detect significant past demographic events exist (Excoffier, Dupanloup, Huerta-Sánchez, Sousa, & Foll, 2013; Günther & Coop, 2013; Gutenkunst, Hernandez, Williamson, & Bustamante, 2009), but they are generally purpose-built for large genetic datasets, which span great sections or the whole genome, collected at a single point in time. Those methods often need additional input such as information about recombination processes (Gattepaille et al., 2013) and ascertainment bias (Marth, Czabarka, Murvai, & Sherry, 2004). The lack of phase information also limits the use of some methods because it is needed to account for parent of origin or identity by descent for each allele (Howie, Donnelly, & Marchini, 2009; Kong et al., 2013). Some other studies have directly used genetic differentiation metrics such as Fst, to evaluate temporal change between genetic datasets (e.g. Larroque et al 2019b; Segura-García et al., 2019). However, translating our spatial understanding of Fst-based results to the temporal dimension is not always straightforward. Indeed, appropriate use and interpretation of pairwise Fst requires that certain assumptions such as equal amounts of drift in both populations be respected (Bhatia, Patterson, Sankararaman, & Price, 2013) and translated in a temporal context. Additionally, disentangling spatial from temporal effects is a challenge because the additivity of genetic drift, means than genetic differentiation can be associated with both temporal structure or population divergence (Murray et al., 2016; Skoglund, Sjödin, Skoglund, Lascoux, & Jakobsson, 2014). Testing whether significant change, relative to the expected variation mostly associated with genetic drift, has occurred in a population based on limited time series genetic data remains a challenge.

One of the most crucial steps in this comparison is to evaluate the significance of the change. Indeed, without a mean to determine adequate significance thresholds for their analyses, decision makers and researchers would be left to arbitrarily set thresholds for what constitute change for their specific genetic dataset. Permutation-based approaches can be used to generate a distribution of values against which an observed value (here temporal change in genetic diversity) can be compared . Such 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 2019) assess the significance of changes in community composition through time. Given the conceptual similarity between the question of how multi-species communities might change through time and our question of monitoring genetic change through time, we expect that TBI can be applied/modified for the analysis of multi-locus genotypic data. The method involves estimating temporal change in each sampling site between two dates using a dissimilarity index/distance, and testing the significance of each change through permutations. Although other permutation-based methods have been successful in answering other genetics questions (e.g. Churchill & Doerge, 1994; Fourtune et al., 2018; Prunier et al., 2013), and TBI has been extensively tested on community composition data (Legendre, 2019b), there is uncertainty about the performance of TBI, when applied to genetic data.

In this study, we build upon the temporal beta diversity indices framework to develop and apply a method, Temporal Genetic diversity Indices (TGI), to quantify and statistically assess temporal variation in spatial genetic diversity. Quantifying relative temporal genetic change among locations will allow us to infer the existence of past demographic events. Persisting spatial legacies in genetic diversity can also be used to identify sites that were most strongly impacted by previous demographic events. Such spatial legacies could also highlight which sites should be investigated if managers are not aware of an a priori known demographic event. To demonstrate the effectiveness and applicability of the approach, we used a spatially-explicit gene flow simulator (Landguth, Bearlin, Day, & Dunham, 2017). We simulated multiple scenarios in which portions of a landscape are affected by different non-selective demographic changes. We then used TGI to measure changes in the genetic make-up of our populations, and evaluated the power and error rates associated with this approach. The goal of our approach is to help researchers with limited time series of genetic data, to identify whether substantial change has occurred in one of the population they studied.

In testing the performance of our TGI approach, we explored how dispersal ability, the number of populations affected a demographic event, and time between two sampling efforts, affected temporal variation in genetic diversity. We also explored how different permutation algorithms in our framework affected our ability to identify statistically significant deviation from neutral expectations, based on simulated processes such as genetic drift. Performance was quantified using standard false positive/negative rates binary classification. We predicted that performance would be lower with increasing dispersal ability because of the homogenizing effect of a higher gene flow. We predict that the longer the time between samplings, regardless of when an event occurred between them, the harder it will be to identify where and when a demographic event occurred. Finally, we briefly showed that TGI testing works on microsatellite data.

**METHODS**

*Adapting Temporal Beta diversity Indices for genetic data*

Calculating TBI (Legendre, 2019b) involves computing dissimilarities in species composition between temporal surveys of the same sites, and testing their significance through permutations of the site-species input matrices. With TGI we considered population-level genotype frequency matrices as input, used genetic distances as dissimilarity, and the null hypothesis became that genetic diversity has not exceptionally changed between both sampling efforts, relative to other genetic diversity changes in the landscape. Testing the significance of TGI involved permuting the input matrices so that the observed genetic distance of a population with its former self, can be evaluated against a number of other genetic distances obtained through permutations. There are several different ways that one can permute spatial-temporal genetic data and choosing the right way to permute can be important to avoid errors (Adams & Collyer, 2015). For example, one can permute a locus with another in the same way in both temporal datasets, or one can permute loci independently in each dataset. An alternative way to permute genetic data is to permute sampling sites instead of loci. It was uncertain which type of permutation approach would produce the best performance with genetic data. Here, we explicitly tested the performance of these three permutation approaches to identify statistically significant temporal change in genetic diversity. We summarized the statistical performance of each permutation approach, and used the best approach to answer all other questions. We used 999 permutations in all analyses, unless specified.

For the genetic distance, we used the chord distance which has been commonly used in both community ecology (Orlóci 1967; Legendre & Borcard 2018) and population genetics (Cavalli-Sforza & Edwards 1967; Balkenhol et al. 2016). Here we use the chord distance to calculate genetic dissimilarity of a single site sampled at two different points in (simulated) time. We chose the 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.

*Simulation framework*

To simulate changes in genetic information through time, we used the spatially-explicit gene flow simulation software CDMetaPOP (Landguth, Bearlin, et al., 2017). CDMetaPOP simulates dispersal and mating of individuals across a landscape, and allows the user to define initial genetic structure, spatial distribution of individuals, dispersal characteristics, and life history traits of the population. The physical landscape we simulated was modelled as a homogeneous and interconnected square grid of 5 by 5 cells, with each cell representing a population. Each population had a maximum carrying capacity of 50 individuals. Structural connectivity between populations was modelled following geographical distance alone. The populated landscape therefore represents to a maximum of 1250 individuals. Each simulation was run for 100 generations before a demographic event was imposed on up to three populations in the landscape. 10 more generations were simulated after this event. The mutation rate was set at 10-8 to reflect empirically-derived mutation rates found in many taxa. The genotypic information of each individual was recorded, and consisted of 100 neutral, unlinked, bi-allelic SNP loci.

*Demographic events*

We simulated 180 replicates for each scenario. For each replicate, we initialized the simulation with a random and unique allocation of alleles among individuals, therefore reaching maximum diversity (Landguth, Bearlin, Day, & Dunham, 2016). 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 produce reasonably realistic and useful simulated genetic data.

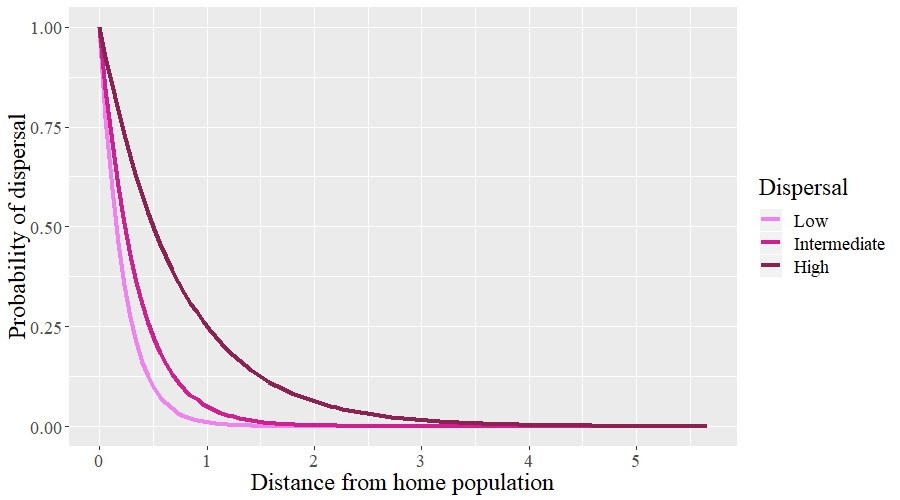
We examined the influence of dispersal and demographic event spatial extent (number of populations) on the persistence of genetic spatial legacies using this simulation model. We examined three levels of dispersal, two demographic event types, and three different numbers of populations affected for a total of 18 unique scenarios, each of which was replicated 180 times, for a total of 3240 (18 × 180) simulations. In the next sections, we detail how we modelled these three experimental factors.

We simulated two different demographic events within our simulation framework: population immigration, and a population bottleneck. Our goal through these simulated events was to test the capacity of the TGI approach to detect these changes in population density. In simulating immigration, we allowed individuals from a 26th separate population to be added to our study area. This immigrating population, otherwise sharing the same characteristics as other populations in our simulated landscape, had developed in isolation and never exchanged individuals before. Only during the demographic event, was the isolation broken by allowing individuals from the 26th isolated population to randomly disperse to *a priori* defined parts of the landscape by setting dispersal distance to the minimum. 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 during the demographic event.

*Dispersal*

To model dispersal, we simply transformed distances between populations by using and then rescaled the values, using the maximum and the minimum (0) distances, possible in this virtual landscape, as described in the CDMetaPOP (Landguth, Bearlin, et al., 2017) user manual (p.63). This gave us a probability that an individual disperses at a distance (Fig.1). We chose this way of modelling dispersal to allow both within population movement and long distance dispersal.

The population to which an individual disperses was selected randomly from the set of populations available at the distance which was itself randomly sampled in the previous step. We made the assumption that individuals cannot disperse beyond our landscape. While this may sacrifice some realism for organisms that disperse passively, it should appropriately model organisms which use landscape cues to find suitable habitat. It also reduced computational time. In order to investigate the effect of different levels of dispersal, we used three different values of *B*: low (*B* = 2), intermediate (*B* = 1.301) and high (*B* = 0.6015) dispersal levels (Fig.1).



**Fig.1:** Probability of dispersal of an individual in three different dispersal scenarios.

*Number and position of populations with spatial legacies*

We also wanted to evaluate how the number of populations bearing spatio-temporal population genetic legacies, influenced the performance of our testing procedure. To achieve this, we triggered demographic events from 1 to 3 populations among the 25. When only 1 population was affected we partitioned the 180 replicates of that scenario equally among 6 populations in the landscape. Because our landscape is square and homogenous, and therefore symmetric, only 6 positions need to be assessed. When several (*k*) populations underwent a demographic event, 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 different set of populations). We chose to pick populations this way to respect the spatial autocorrelation often exhibited in demographic events.

*Time since demographic change*

To assess how the time since the simulated demographic event affects our ability to detect genetic change, we used TGI on simulation data collected each year, up to five years after the event, and compared them with data from the event year. We did the same with the earliest sampling period, that is we used simulation data dating 5 years before the event, as well as each year until the event. We used two out of the eighteen scenarios, and chose the most extreme according to the previous results, to represent the “easiest” and “hardest” contexts to detect change. We showed results for the 0.05 *p*-value threshold as it was a good compromise between decent FPR and FNR in our initial results.

*Microsatellites*

Although we investigated several aspects of TBI application on genetic data on SNP, we also simulated one scenario modelling microsatellites markers (low dispersal, one affected population, bottleneck). We chose to do this because microsatellites are still relevant in molecular ecology in the age of whole genome sequencing (e.g. Bezemer, Krauss, Roberts, & Hopper, 2019), and because technology keeps being developed and improved for them (e.g. Lepais et al., 2019). We changed the simulation parameters to have 10 microsatellite loci, with 10 alleles each. We also had to change the way we calculate the genetic dissimilarities. For that matter we created a new TGI function dedicated to microsatellite data (*TGImicro*), and used *dist.genpop* from the *adegenet* R package (see *Software*) to calculate dissimilarities. Among the metrics it offers, we chose the Roger’s distance because it is a Euclidean genetic dissimilarity metric which does not make biological assumptions and therefore would apply to many empirical cases.

*Statistical performance*

We used the False Positive Rate (FPR) and False Negative Rate (FNR) frameworks to assess statistical performance of the TGI 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 affected 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 TGI permutation tests is important in order to find a compromise between power and selectivity, we showcased it to potential TGI users by evaluating statistical performance across a range of thresholds: 0.0001, 0.00025, 0.0005, 0.00075, 0.001, 0.0025, 0.005, 0.0075, 0.01, 0.025, 0.05, 0.075, 0.1.

*Controls*

Control populations are never affected by any event and therefore only display other sources of genetic variation such as gene flow, drift, and mutation. Dispersal ability was therefore the only parameter to change for the controls, giving us 3 control scenarios. We evaluated the FPR of those three control scenarios (no need for FNR because there are no true positives/false negatives so it was always equal to 0). When describing the performance of other scenarios with similar dispersal parameters, we always put control values as a reference.

*Software*CDMetaPOP runs on Python 2.7 (Landguth, Bearlin, et al., 2017). We used the R software (R Core Team, 2019) in the RStudio IDE (RStudio Team, 2018) for all analyses and illustration. We used the *adegenet* (Jombart, 2008; Jombart & Ahmed, 2011), *pegas* (Paradis, 2010), and *adespatial* (Dray et al., 2019) R packages for the calculations.

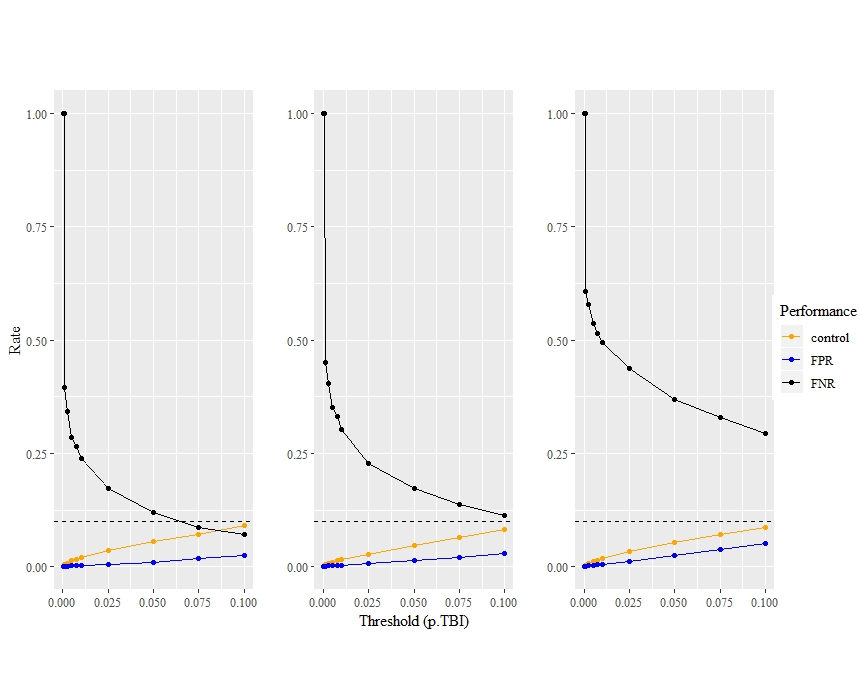
**RESULTS**

*Permutation approach*

The first permutation approach (permuting loci in the same way in both samples) is the only one that is functional with genetic data. Indeed, the second and third approaches most often failed to find any significant change. This means that they never found any false positive (FPR = 0), which is great, but also that they very rarely found any true positive (FNR > 0.9), regardless of the scenario or the *p*-value threshold we used. Because only the first approach was suitable to study simulation outputs, we used it for the rest of the analyses.

*Dispersal ability*

As hypothesized, the dispersal ability of an organism, relative to its landscape, greatly affects our ability to detect exceptional temporal changes from limited genetic datasets. Indeed, when we group scenarios with the same dispersal parameters (low, intermediate, high) together, FNR and FPR substantially increase with dispersal intensity (Fig. 2). This is true regardless of the threshold used, and the bigger the threshold, the larger the difference between average values of FPR of the three scenarios. For example, at the ubiquitous 0.05 threshold, which here seems to be a decent compromise between low FNR and FPR, average FNR values are 0.1210, 0.1727 and 0.3702, for the low, intermediate and high dispersal scenarios respectively. At this threshold and for the same scenario groups, FPR also increases, from 0.0107 to 0.0138 and 0.0244.



C

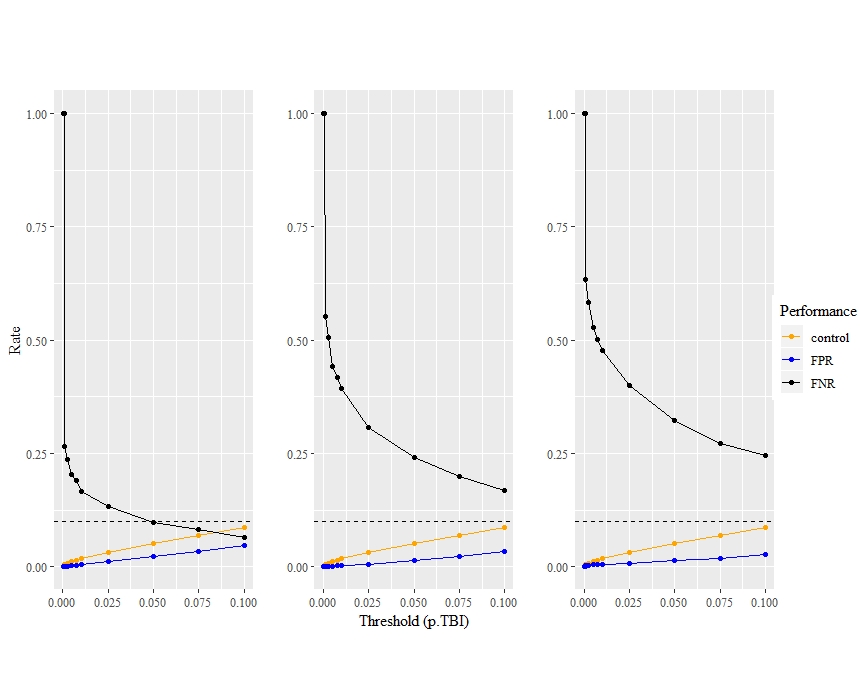
B

A

**Fig. 2.** Influence of dispersal ability on our ability to detect exceptional change. FPR and FNR values at 13 different p.TGI thresholds for low (A), intermediate (B), and high (C) dispersal scenarios. Control FPR values, from scenarios with identical dispersal parameters, are also featured. The dashed horizontal line indicates 0.1 which is the maximum threshold value used, for comparison with FPR values.

*Number of populations affected*

The number of populations affected by an event also affects our ability to detect exceptional temporal change. When looking at groups of scenarios with the same number of affected populations (1, 2, and 3 populations), we can see that FNR increases with additional affected populations, regardless of which threshold is considered (Fig. 3). FPR values from scenarios with 2 affected populations are consistently higher than values from one affected population scenarios. FPR values from scenarios with 3 affected populations are on average lower than values from other scenarios up to a threshold of 0.01, and are on average higher for thresholds above 0.05, therefore indicating an interaction between the number of affected populations and the threshold used in the permutation procedure. However, for thresholds that would be considered suitable regarding power (*e.g.* power > 50%), a higher number of populations always leads to a lower FPR.



A

C

B

**Fig. 3.** Influence of the number of affected populations on our ability to detect exceptional change. FPR and FNR values at 13 different p.TGI thresholds for 1 (A), 2 (B), and 3 (C) affected populations scenarios. Control FPR values, from scenarios with identical dispersal parameters, are also featured. The dashed horizontal line indicates 0.1 which is the maximum threshold value used, for comparison with FPR values.

*Lag time between pre-event-sampling and event*

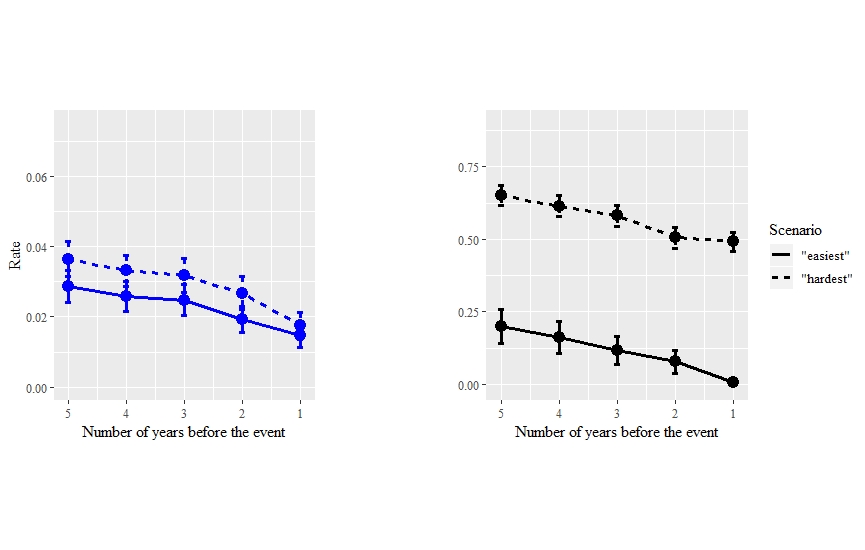
We can see in Fig. 4 A & B, that the longest the pre-event sampling is from the event, the less power and the more false positives we get. Sampling done 5 years before the event led to about twice as much false positives as sampling done the year before the event. The effect of time on FPR or FNR is similar regardless of the scenarios (Fig. 4 A & B), however FNR variation increased with time for the “easiest” scenario (immigration event, 1 population, low dispersal) whereas FPR variation did not increase for “easiest” or the “hardest” (immigration event, 3 populations, high dispersal) scenarios. The difference in performance between scenario sharply changed between 1 and 2 years, and then it stayed about the same for longer periods between samplings.

*Lag time between event and post-event sampling*

As hypothesized from the nature of genetic processes in connected populations, the genetic signal of the demographic event inflicted upon populations disappears gradually over time. When considering the scenario most likely to preserve the signal according to earlier results on FNR and FPR, the TGI approach was still able to avoid false adequately two years after the event (Fig. 4 C) but average FPR sharply increased at the three years’ mark, then increased linearly again in the following years. For the harder scenario FPR increased much faster with the years, following a slightly saturated curve, and reaching 5% of false positives after only two years (Fig. 4 C). Average FNR, and the width of its confidence intervals, increased linearly for the easier scenario, but (Fig. 4 D). Beyond the fact that its starting FNR at 0.05 was much higher for the harder scenario (Fig. 2; Fig. 4 D), it also increased much faster with time, reaching a plateau at unacceptable power values. With the harder scenario, almost 25% of power is lost as the result of only two generations.

FPR

FNR



A

B

C

D

**Fig. 4.** Influence of number years between the event and a pre-event sampling (A, B) or a post-event sampling (C, D) on averages and confidence intervals of FPR (A, C) and FNR (B, D), for two extreme scenarios with the 0.05 *p*-value threshold.

*Threshold and general performance*

Stricter values (lower values) for the TGI *p*-value threshold expectedly bring a better FPR but also bring a pathological FNR (low power). Indeed, across all scenarios, the FNR decreases exponentially when threshold values increase, while the FPR increases linearly (e.g. Fig. 3; Fig. 4). Notably, FPR values never surpassed 0.1, which was the maximum threshold chosen in our testing, indicating that they may be acceptable (Legendre & Legendre, 2012). The decrease in average FNR across all scenarios associated with an increase in the threshold value, is accompanied by a decrease of the associated standard variation, as soon as variation exists (FNR not equal to 1): from 0.3749 (0.001) to 0.2471 (0.1), considering all scenarios. In contrast, the increase in average FPR is concurrent with an increasing of its variation: from 0 (0.0001) to 0.0377 (0.1). Given the large variation in performance, along each parameter we considered, we believe that the parameters we chose to define different scenarios produced sufficiently complex, and useful simulations.

*Control simulations*

Experimental FPR values consistently stayed below control FPR values, also the difference generally diminished with the intensity of dispersal (Fig. 2). This means that in the presence of an actual event, we were less likely to wrongfully choose a population as having been affected. Control FPR values did not vary between scenario groups (ANOVA; *p*-value = 0.353), which means that dispersal does not affect the selection of a random population as a positive. Finally, control FPR values never passed 0.1, which was the maximum threshold chosen in our testing.

*Microsatellite*

At a threshold of 0.05, FNR is equal to 0.0500 and FPR is equal to 0.0007, which both indicate very good performances of *TGImicro()* in detecting significant changes, when using microsatellite data. Similarly to the simulations with biallelic data, FNR decreases and FPR decreases with increasing threshold values. From 0 (0.0001) to 0.0021 (0.1) for FPR, and from 1 (0.0001) to 0.0278 (0.1). The method’s performance differed between microsatellite and biallelic genetic data, for the same number of alleles (100), and for an otherwise identical scenario (low dispersal, one affected population, bottleneck). Indeed, at a threshold of 0.05 for example, the average FPR value is higher (0.0208 vs 0.0007), whereas the average FNR value is lower (0.0111 vs 0.0500) for SNP than microsatellite respectively. Please note however that the distance metric we used for both differed and that this could influence this comparison.

**DISCUSSION**

Being able to detect which populations have changed significantly over time, from genetic data, has always been a challenge for researchers. When genetic data is available at several points in time, we believe the aforementioned challenge is within reach for biologists, even if given relatively poor genetic datasets (*e.g.* hundreds of SNPs), as our permutation approach was generally able to achieve this goal, under certain conditions.

Detecting exceptional change is harder in landscapes with strong functional connectivity. Indeed, we found a general decrease in performance, whatever the performance focus was, with an increase of dispersal ability (Fig. 2). This decrease exists even with only one generation separating two temporal samples, which suggests that studying highly connected systems might require more frequent sampling, or at least that higher uncertainty should be acknowledged. High dispersal, and higher gene flow through it (Cayuela et al., 2018), is implicated is many short-term or long-term mechanisms which lower our ability to understand the eco-evolutionary dynamics of species. For example, high dispersal during range expansion lowers our ability to correctly detect loci under natural selection (Mayrand, Filotas, Wittische, & James, 2019), and high gene flow may not always be associated with a strong decrease in measured structure (Landguth, Cushman, Murphy, & Luikart, 2010) or early detections of barriers to gene flow (Landguth, Cushman, Schwartz, et al., 2010).

The spatial extent of an event, e.g. the number of populations it affected in the landscape in our study, also decreases our ability to correctly identify which populations have truly changed. Although the spatial extent of a legacy may help researchers detect it because it increases the chance of the legacy being sampled, it also greatly increased the risk of missing the legacy when using our permutation approach (Fig.3). Whether this could be offset by a lesser degree of spatial autocorrelation in the spatial genetic legacy has not been investigated in our paper. Spatial autocorrelation may greatly affect many genetic analyses, and solutions are being developed to integrate it within them (Rousset & Ferdy, 2014). We believe explicitly taking spatial autocorrelation into account in temporal analyses of genetic diversity (Bradburd & Ralph, 2019) represents a promising and challenging avenue of research .

Demographic processes quickly dilute the signal in spatial genetic legacies, by transferring the initial effect of an event on genetic diversity, to other populations (Fig. 4). Although the spatial legacy of a past demographic event could be kept in richer genomic data (*e.g.* probability of mutational configurations in sequence blocks), limited biallelic gene frequency data will not keep most of the signal beyond a few years, even in the best situations (Fig. 4). Microsatellite data seem to have a different behavior in retaining information from our results, and their retention should be investigated further. Reasonable performance can be expected if the first sampling was a few years before the event (Fig. 4 A B), which makes past sampling, which purpose was not necessarily to study temporal change, still useful. Regardless, the closer the date of the first or last sampling is to the date of the event, the better performance-wise. This is especially true for the numbers of years passed since the event, in situations where other factors lower performance as well (Fig. 4 C).

Our analyses have shown that TGI testing is functional, under certain conditions. First, only one permutation algorithm (permutations done locus by locus, and in the same way for both samples) is suitable when using gene frequency data. The other permutation approaches were incontrovertibly poor in their ability to pick up on the genetic legacy signal left by the demographic events occurring in some populations, as they almost never select any. This result may not come as a surprise as the same permutation approach was also selected as the best for community composition data (Legendre, 2019a). Given these results, we want to warn readers that other permutation algorithms should be extensively tested with the help of varied simulations, before being considered for use on genetic data.

Simulations provide a very useful tool for the planning researchers who would want to investigate change in their study landscape. Simulations have been used with much success in a variety of applications (Epperson et al., 2010), from investigating a species evolutionary ecology to protect it (Creech et al., 2017; Landguth, Holden, et al., 2017), to showcasing the performance of various approach to extract valuable information from genetic data (Cubry, Vigouroux, & François, 2017; Forester, Jones, Joost, Landguth, & Lasky, 2016; Mayrand et al., 2019). We do not advise future users of TGI, or other permutation approaches, to arbitrarily choose a *p*-value threshold to pick which populations display significant changes, or to base the timing of their sampling based on default simulations, or our simulations. Instead, we encourage them to run simulations with a reasonable realism, that is by inputting demographic parameters, such as reproduction parameters, available in the literature (if any) and by carefully creating a virtual landscape resembling their study area. If accurate demographic parameters are not available, we encourage them to simulate scenarios with wide ranging parameters values as we did in this study. In order to test more complex and competing hypotheses for specific phenomenon using spatio-temporal data, adequate process-based null models should be created. Such spatial null models, can be generated by simulations by modelling major phenomena that are not generating the pattern of interest, so that tests can be better calibrated to reliably identify significance (Gardner & Urban, 2007; James, Fleming, & Fortin, 2010; Paz-Vinas, Loot, Stevens, & Blanchet, 2015). This increased realism, and evaluation of uncertainty, would provide more accurate tests, to pick the best *p*-value threshold, as well as understand when is it still adequate to sample, to get the best out of spatial genetic legacies. A number of programs such as CDMetaPOP (Landguth, Bearlin, et al., 2017), Nemo (Guillaume & Rougemont, 2006), SPLATCHE (Currat, Ray, & Excoffier, 2004), or SLIM (Haller & Messer, 2019) provide very flexible and sophisticated ways to implement such simulations.

We believe future work ought to explore how the performance of TGI, as well as other methods depending on the quantity and quality of the genetic data available, in detecting spatio-temporal genetic change, varies with different factors. Some of the factors we think are relevant include the choice of the genetic distance used in the algorithm, the influence of the degree of spatial autocorrelation in genetic legacies, and varying effective population sizes.

**DATA AND SOFTWARE AVAILABILITY**

All simulation data used for this paper will be deposited online. Functions used to analyze the simulations will be available on a public repository on GitHub. *TGImicro()* which is the function that would be most useful to potential users of our approach, will continue to be maintained and developed and may be contributed to a package.

**ACKNOWLEDGEMENTS**

This research was supported by a grant to PMAJ and the TRIA Network from the Natural Sciences and Engineering Research Council of Canada (grant no. NET GP 434810-12), with contributions from Alberta Agriculture and Forestry, fRI Research, Manitoba Conservation and Water Stewardship, Canadian Forest Service (Natural Resources Canada), Northwest Territories Environment and Natural Resources, Ontario Ministry of Natural Resources and Forestry, Saskatchewan Ministry of Environment, West Fraser, and Weyerhaeuser. JW was also supported by a scholarship from the Forest Complexity Modelling (FCM) NSERC CREATE. Finally, we thank Jeremy Larroque, Hinatea Ariey and others, for their comments on an earlier version of the manuscript.

**REFERENCES**

Adams, D. C., & Collyer, M. L. (2015). Permutation tests for phylogenetic comparative analyses of high-dimensional shape data: What you shuffle matters. *Evolution*, *69*(3), 823–829. doi: 10.1111/evo.12596

Aeschbacher, S., Selby, J. P., Willis, J. H., & Coop, G. (2016). *Population-genomic inference of the strength and timing of selection against gene flow*. (18), 1–6. doi: 10.1101/072736

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

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

Bezemer, N., Krauss, S. L., Roberts, D. G., & Hopper, S. D. (2019). Conservation of old individual trees and small populations is integral to maintain species’ genetic diversity of a historically fragmented woody perennial. *Molecular Ecology*, (January), 3339–3357. doi: 10.1111/mec.15164

Bhatia, G., Patterson, N., Sankararaman, S., & Price, A. L. (2013). Estimating and interpreting F. *Genome Research*, (2), 1–9. doi: 10.1101/gr.154831.113.23

Bolliger, J., Lander, T., & Balkenhol, N. (2014). Landscape genetics since 2003: Status, challenges and future directions. *Landscape Ecology*, *29*(3), 361–366. doi: 10.1007/s10980-013-9982-x

Bolnick, D. I., & Nosil, P. (2007). Natural selection in populations subject to a migration load. *Evolution*, *61*(9), 2229–2243. doi: 10.1111/j.1558-5646.2007.00179.x

Bradburd, G. S., & Ralph, P. L. (2019). Spatial Population Genetics: It’s About Time. *Annual Review of Ecology, Evolution, and Systematics*, *50*(1), 427–449. doi: 10.1146/annurev-ecolsys-110316-022659

Buschbom, J., Yanbaev, Y., & Degen, B. (2011). Efficient long-distance gene flow into an isolated relict oak stand. *Journal of Heredity*, *102*(4), 464–472. doi: 10.1093/jhered/esr023

Cayuela, H., Rougemont, Q., Prunier, J. G., Moore, J. S., Clobert, J., Besnard, A., & Bernatchez, L. (2018). Demographic and genetic approaches to study dispersal in wild animal populations: A methodological review. *Molecular Ecology*, *27*(20), 3976–4010. doi: 10.1111/mec.14848

Churchill, G. A., & Doerge, R. W. (1994). Empirical threshold values for quantitative trait mapping. *Genetics*, *138*(3), 963–971.

Creech, T. G., Epps, C. W., Landguth, E. L., Wehausen, J. D., Crowhurst, R. S., Holton, B., & Monello, R. J. (2017). Simulating the spread of selection-driven genotypes using landscape resistance models for desert bighorn sheep. *PLoS ONE*, *12*(5), 1–26. doi: 10.1371/journal.pone.0176960

Cubry, P., Vigouroux, Y., & François, O. (2017). The Empirical Distribution of Singletons for Geographic Samples of DNA Sequences. *Frontiers in Genetics*, *8*(September), 1–10. doi: 10.3389/fgene.2017.00139

Currat, M., Ray, N., & Excoffier, L. (2004). SPLATCHE: A program to simulate genetic diversity taking into account environmental heterogeneity. *Molecular Ecology Notes*, *4*(1), 139–142. doi: 10.1046/j.1471-8286.2003.00582.x

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. doi: 10.1111/eva.12617

Dray, S., Bauman, D., Blanchet, G., Borcard, D., Clappe, S., Guenard, G., … Wagner, H. H. (2019). *adespatial: Multivariate Multiscale Spatial Analysis.* Retrieved from https://cran.r-project.org/package=adespatial

Epperson, B. K., McRae, B. H., Scribner, K., Cushman, S. a, Rosenberg, M. S., Fortin, M.-J., … Dale, M. R. T. (2010). Utility of computer simulations in landscape genetics. *Molecular Ecology*, *19*(17), 3549–3564. doi: 10.1111/j.1365-294X.2010.04678.x

Epps, C. W., & Keyghobadi, N. (2015). Landscape genetics in a changing world: Disentangling historical and contemporary influences and inferring change. *Molecular Ecology*, *24*(24), 6021–6040. doi: 10.1111/mec.13454

Excoffier, L., Dupanloup, I., Huerta-Sánchez, E., Sousa, V. C., & Foll, M. (2013). Robust Demographic Inference from Genomic and SNP Data. *PLoS Genetics*, *9*(10). doi: 10.1371/journal.pgen.1003905

Forester, B. R., Jones, M. R., Joost, S., Landguth, E. L., & Lasky, J. R. (2016). Detecting spatial genetic signatures of local adaptation in heterogeneous landscapes. *Molecular Ecology*, *25*(1), 104–120. doi: 10.1111/mec.13476

Fourtune, L., Prunier, J. G., Paz-Vinas, I., Loot, G., Veyssière, C., & Blanchet, S. (2018). Inferring Causalities in Landscape Genetics: An Extension of Wright’s Causal Modeling to Distance Matrices. *The American Naturalist*, *191*(4), 000–000. doi: 10.1086/696233

Gardner, R. H., & Urban, D. L. (2007). Neutral models for testing landscape hypotheses. *Landscape Ecology*, *22*(1), 15–29. doi: 10.1007/s10980-006-9011-4

Gattepaille, L. M., Jakobsson, M., & Blum, M. G. B. (2013). Inferring population size changes with sequence and SNP data: Lessons from human bottlenecks. *Heredity*, *110*(5), 409–419. doi: 10.1038/hdy.2012.120

Guillaume, F., & Rougemont, J. (2006). Nemo: An evolutionary and population genetics programming framework. *Bioinformatics*, *22*(20), 2556–2557. doi: 10.1093/bioinformatics/btl415

Günther, T., & Coop, G. (2013). Robust identification of local adaptation from allele frequencies. *Genetics*, *195*(1), 205–220. doi: 10.1534/genetics.113.152462

Gutenkunst, R. N., Hernandez, R. D., Williamson, S. H., & Bustamante, C. D. (2009). Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *PLoS Genetics*, *5*(10). doi: 10.1371/journal.pgen.1000695

Haller, B. C., & Messer, P. W. (2019). SLiM 3: Forward Genetic Simulations Beyond the Wright-Fisher Model. *Molecular Biology and Evolution*, *36*(3), 632–637. doi: 10.1093/molbev/msy228

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. doi: 10.1111/eva.12149

Howie, B. N., Donnelly, P., & Marchini, J. (2009). A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genetics*, *5*(6). doi: 10.1371/journal.pgen.1000529

James, P. M. a., Fleming, R. a., & Fortin, M. J. (2010). Identifying significant scale-specific spatial boundaries using wavelets and null models: Spruce budworm defoliation in Ontario, Canada as a case study. *Landscape Ecology*, *25*, 873–887. doi: 10.1007/s10980-010-9465-2

Jombart, T. (2008). Adegenet: A R package for the multivariate analysis of genetic markers. *Bioinformatics*, *24*(11), 1403–1405. doi: 10.1093/bioinformatics/btn129

Jombart, T., & Ahmed, I. (2011). adegenet 1.3-1: New tools for the analysis of genome-wide SNP data. *Bioinformatics*, *27*(21), 3070–3071. doi: 10.1093/bioinformatics/btr521

Kong, A., Steinthorsdottir, V., Masson, G., Thorleifsson, G., Sulem, P., Besenbacher, S., … Jonsson, T. (2013). *Europe PMC Funders Group Parental origin of sequence variants associated with complex diseases*. *462*(7275), 868–874. doi: 10.1038/nature08625.Parental

Kremer, A., Ronce, O., Robledo-Arnuncio, J. J., Guillaume, F., Bohrer, G., Nathan, R., … Schueler, S. (2012). Long-distance gene flow and adaptation of forest trees to rapid climate change. *Ecology Letters*, *15*(4), 378–392. doi: 10.1111/j.1461-0248.2012.01746.x

Landguth, E. L., Bearlin, A., Day, C. C., & Dunham, J. (2016). CDMetaPOP: an individual-based, eco-evolutionary model for spatially-explicit simulation of landscape demogenetics. *Methods in Ecology and Evolution*. doi: 10.1111/2041-210X.12608

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. doi: 10.1111/2041-210X.12608

Landguth, E. L., Cushman, S. a., Schwartz, M. K., McKelvey, K. S., Murphy, M., & Luikart, G. (2010). Quantifying the lag time to detect barriers in landscape genetics. *Molecular Ecology*, *19*, 4179–4191. doi: 10.1111/j.1365-294X.2010.04808.x

Landguth, E. L., Cushman, S. a, Murphy, M. a, & Luikart, G. (2010). Relationships between migration rates and landscape resistance assessed using individual-based simulations. *Molecular Ecology Resources*, *10*(5), 854–862. doi: 10.1111/j.1755-0998.2010.02867.x

Landguth, E. L., Holden, Z. A., Mahalovich, M. F., & Cushman, S. A. (2017). Using landscape genetics simulations for planting blister rust resistant whitebark pine in the US Northern Rocky Mountains. *Frontiers in Genetics*, *8*(FEB), 1–12. doi: 10.3389/fgene.2017.00009

Larroque, J., Legault, S., Johns, R., Lumley, L., Cusson, M., Renaut, S., … James, P. M. A. (2019). Temporal variation in spatial genetic structure during population outbreaks: Distinguishing among different potential drivers of spatial synchrony. *Evolutionary Applications*, (July), 1–15. doi: 10.1111/eva.12852

Legendre, P. (2019a). 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. doi: 10.1002/ece3.4984

Legendre, P. (2019b). A temporal beta‐diversity index to identify sites that have changed in exceptional ways in space-time surveys. *Ecology and Evolution*, *9*, 3500–3514. doi: 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). doi: 10.1098/rspb.2013.2728

Legendre, P., & Legendre, L. (2012). *Numerical Ecology* (Third Engl). Amsterdam: Elsevier.

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. doi: 10.1111/eva.12810

Lepais, O., Chancerel, E., Boury, C., Salin, F., Manicki, A., Taillebois, L., … Guichoux, E. (2019). Fast sequence-based microsatellite genotyping development workflow for any non-model species. *BioRxiv*, 649772. doi: 10.1101/649772

Lowe, W. H., & Allendorf, F. W. (2010). What can genetics tell us about population connectivity? *Molecular Ecology*, *19*(15), 3038–3051. doi: 10.1111/j.1365-294X.2010.04688.x

Manel, S., & Holderegger, R. (2013). Ten years of landscape genetics. *Trends in Ecology & Evolution*, *28*(10), 614–621. doi: 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. doi: 10.1016/S0169-5347(03)00008-9

Marth, G. T., Czabarka, E., Murvai, J., & Sherry, S. T. (2004). The Allele Frequency Spectrum in Genome-Wide Human Variation Data Reveals Signals of Differential Demographic History in Three Large World Populations. *Genetics*, *166*(1), 351–372. doi: 10.1534/genetics.166.1.351

Maruyama, T., & Fuerstt, P. A. (1985). POPULATION BOTTLENECKS AND NONEQUILIBRIUM MODELS IN POPULATION GENETICS. 11. NUMBER O F ALLELES IN A SMALL POPULATION T H A T WAS FORMED BY A RECENT BOTTLENECK HE fate of genes in a population that experiences a sudden reduction in. *Genetics*, *111*(3), 675–689. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1202664/pdf/675.pdf

Mayrand, P., Filotas, E., Wittische, J., & James, P. M. A. (2019). The role of dispersal, selection, and timing of sampling on the false discovery rate of loci under selection during geographic range expansion. *Genome*, *13*(July), 1–13. doi: 10.1139/gen-2019-0004

Murray, G. G. R., Wang, F., Harrison, E. M., Paterson, G. K., Mather, A. E., Harris, S. R., … Welch, J. J. (2016). The effect of genetic structure on molecular dating and tests for temporal signal. *Methods in Ecology and Evolution*, *7*(1), 80–89. doi: 10.1111/2041-210X.12466

Paradis, E. (2010). Pegas: An R package for population genetics with an integrated-modular approach. *Bioinformatics*, *26*(3), 419–420. doi: 10.1093/bioinformatics/btp696

Paz-Vinas, I., Loot, G., Stevens, V. M., & Blanchet, S. (2015). Evolutionary processes driving spatial patterns of intraspecific genetic diversity in river ecosystems. *Molecular Ecology*, *24*(18), 4586–4604. doi: 10.1111/mec.13345

Prunier, J. G., Kaufmann, B., Fenet, S., Picard, D., Pompanon, F., Joly, P., & Lena, J. P. (2013). Optimizing the trade-off between spatial and genetic sampling efforts in patchy populations: Towards a better assessment of functional connectivity using an individual-based sampling scheme. *Molecular Ecology*, *22*(22), 5516–5530. doi: 10.1111/mec.12499

R Core Team. (2019). *R: A language and environment for statistical computing*. Retrieved from https://www.r-project.org/

Rousset, F. (1997). Genetic Differentiation and Estimation of Gene Flow from FStatistics Under Isolation by Distance. *Genetics*, *145*(4), 1219–1228.

Rousset, F., & Ferdy, J.-B. (2014). Testing environmental and genetic effects in the presence of spatial autocorrelation. *Ecography*, *37*(December 2013), 781–790. doi: 10.1111/ecog.00566

RStudio Team. (2018). *RStudio: Integrated Development for R*. Retrieved from http://www.rstudio.com/

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. doi: 10.1007/s10592-009-0044-5

Segura-García, I., Garavelli, L., Tringali, M., Matthews, T., Chérubin, L. M., Hunt, J., & Box, S. J. (2019). Reconstruction of larval origins based on genetic relatedness and biophysical modeling. *Scientific Reports*, *9*(1), 1–9. doi: 10.1038/s41598-019-43435-9

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

Skoglund, P., Sjödin, P., Skoglund, T., Lascoux, M., & Jakobsson, M. (2014). Investigating population history using temporal genetic differentiation. *Molecular Biology and Evolution*, *31*(9), 2516–2527. doi: 10.1093/molbev/msu192

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

Wittische, J., Janes, J. K., & James, P. M. A. (2019). Modelling landscape genetic connectivity of the mountain pine beetle in western Canada. *Canadian Journal of Forest Research*, *1348*(September), 1339–1348. doi: 10.1139/cjfr-2018-0417

Wright, S. (1943). Isolation by Distance. *Genetics*, *28*(2), 114–138.