**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**

Spatial and temporal variation in genetic information can tell us a great deal about demography and movement of populations. Population genetics approaches are, and will continue to be, widely used for conservation purposes as the approach is used to translate genetic data into meaningful inference (Allendorf, Hohenlohe, & Luikart, 2010; Harrisson, Pavlova, Telonis-Scott, & Sunnucks, 2014; Segelbacher et al., 2010). Among population genetic approaches, landscape genetics approaches bridge an important gap in the field of molecular ecology by 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). The vast majority of studies using these approaches focus on explaining the spatial variation rather than the temporal variation in genetic diversity. However, temporal gains and losses of genetic diversity are 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.

One of the main ongoing challenges for biologists is therefore to detect both when and where in the landscape, a significant gain or loss of genetic diversity occurs. Once detected, those changes in genetic diversity may be associated with natural or anthropogenic landscape changes, from local and abrupt like a wildfire, to global and long-term like climate warming. Such associations may describe what is happening at the demographic level, and therefore serve as an alarm for managers. However, it is rarely possible to observe the effects of these events instantaneously and researchers are often left with spatial legacies in genetic diversity, as these effects may not be readily observable from demographic data alone. Common examples of situations where spatial legacies may be left in the genetic make-up of populations include geographic isolation, population bottleneck and massive migrations from previously isolated populations, which would substantially alter local genetic variation.

Few methods currently exist for the temporal comparison of genetic data. For example, very sophisticated methods exist to infer demographic history from genetic data, even from static, i.e. single time genetic data (Excoffier, Dupanloup, Huerta-Sánchez, Sousa, & Foll, 2013; Günther & Coop, 2013; Gutenkunst, Hernandez, Williamson, & Bustamante, 2009). However useful, those methods are usually designed for very large genetic datasets, which span large sections or the whole genome, and need additional input such as information about recombination processes (Gattepaille, Jakobsson, & Blum, 2013) and ascertainment bias (Marth, Czabarka, Murvai, & Sherry, 2004). The conceptual relevance and performance of other commonly used approaches focusing on population structure (e.g. ordination-based) in answering temporal questions, where one objective is to find which population has indeed changed more significantly than others in the landscape, has not been evaluated to our knowledge. Some studies have directly used genetic differentiation metrics such as Fst, to evaluate temporal change (e.g. Larroque et al 2019b; Segura-García et al., 2019). However, translating biological assumptions and our spatial understanding of Fst-based results to the temporal dimension is not straightforward, as disentangling spatial from temporal effects is a challenge (Skoglund et al 2014). Nevertheless, 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). Testing whether significant and atypical change, relative to the metapopulation, has occurred in a population based on limited time series genetic data, such as microsatellites or unphased and patchy SNP datasets, remains a challenge.

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) 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, and could potentially help us achieve the goal of detecting genetic change from limited time series datasets. 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. 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.

In this study, we described and tested a method to identify locations that have undergone significant genetic change through time. Identifying such locations, and quantifying other locations relative temporal genetic change, is important because significant genetic change could unearth the existence of a past event from spatial legacies in genetic diversity, and would also indicate the parts of the landscape where an event had the strongest effects. It could also highlight which sites should be investigated if managers are not aware of an *a priori* known event. To demonstrate the effectiveness and applicability of the approach we used a large panel of simulations which were generated from a spatially-explicit gene flow simulator. We simulated scenarios where part of the landscape is affected by different non-selective demographic changes. 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. This paper does not aim to infer demographic histories, rather it aims to help researchers with subsequent sets of limited genetic data, to identify whether substantial change has occurred in one of the population they studied.

We notably explored how dispersal ability, the number of affected populations, time between two sampling efforts, and permutation algorithms, influence our ability to detect exceptional temporal change. We predict performance will lower with increasing dispersal ability because of the homogenizing effect of a higher gene flow. We predict performance will also lower with the number of affected populations, because it would make permutations less effective. 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. We also investigated the performance of our approach on microsatellite data. Finally, we illustrated the possibilities of this approach through applications on two real genetic datasets.

**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 at 10-8 to reflect empirically-derived mutation rates found in many taxa. 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 square 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 produce reasonably realistic and useful genetic data.

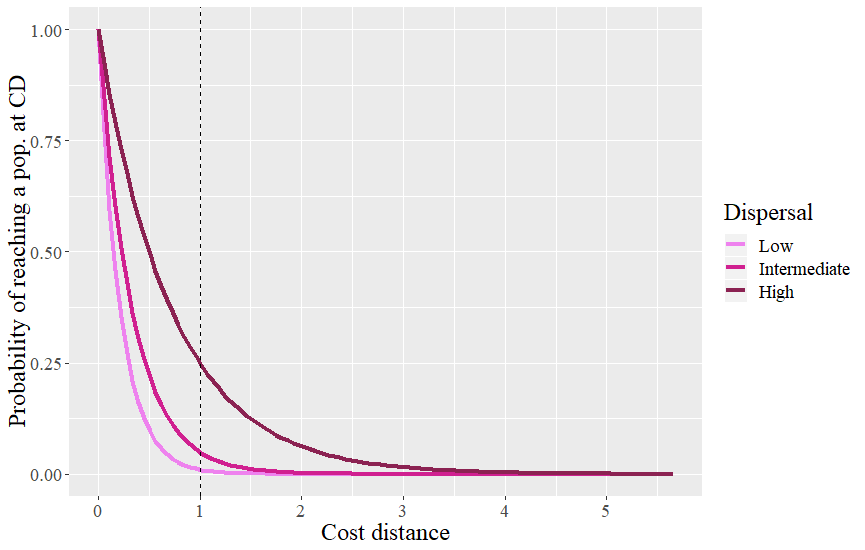
We examined the influence of dispersal, and demographic event spatial extent (number of populations) on the persistence of genetic spatial legacies using this simulation model. With 3 dispersal regimes, 2 different demographic event types and 3 different numbers of populations affected, we have 18 different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

*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.

In order to investigate the effect of different levels of dispersal, we changed the dispersal model 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 an exogeneous 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 square and homogenous, and therefore symmetric, only 6 positions need to be assessed. 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 the spatial autocorrelation often exhibited in demographic events.

*Controls*

To further the quality and transparency of our simulation experiments, we used simulations designed to serve as controls for the rest of the scenarios. Those 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.

*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), because it was readily available in the function, 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.

*Timing*

To assess the influence of the timing of sampling on our ability to detect significant temporal change, we used TBI on simulation data collected each year, up to 5 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, in order to save computation time, 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.

*Permutation approaches*

One of the most crucial steps in describing change is to evaluate the significance of the change. Indeed, without a mean to distinguish typical variation from truly atypical change, decision makers and researchers would be left to arbitrarily set thresholds for what constitute change. Permutation-based approaches may help us to achieve this by creating a distribution of values which can then be compared to the measured value of change. Most calculations used in this paper are based on the TBI function (*TBI()*)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 *TBI()* (Legendre, 2019). Because they were tested on very a different type of data, we used an older version of *TBI()* (*TBIold()*) 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 loci 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. We used 999 permutations in all analyses, unless specified.

*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 TBI function dedicated to microsatellite data (*TBImicro()*), 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. Because *TBImicro()* is slower than *TBI()*, we limited the number of permutations to 99.

*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: 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.

*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.

*Illustration*

To briefly illustrate the use of TBI on genetic data, we use spruce budworm (*Choristoneura fumiferana*) SNP data from 2012 and 2013 (Larroque et al., 2019). Spruce budworm is an irruptive moth species which cyclically defoliates huge areas of spruce and fir forests in Canada. 8 sites from Quebec were sampled in both years, and 3562 loci were extracted from 370 individuals (Larroque et al., 2019).

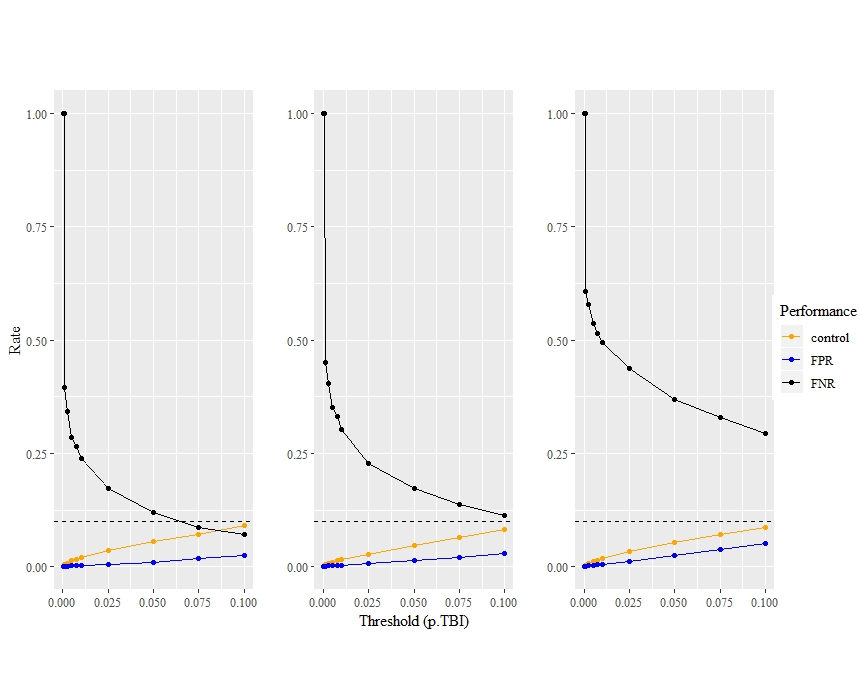
**RESULTS**

*Permutation approach*

The first permutation approach 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

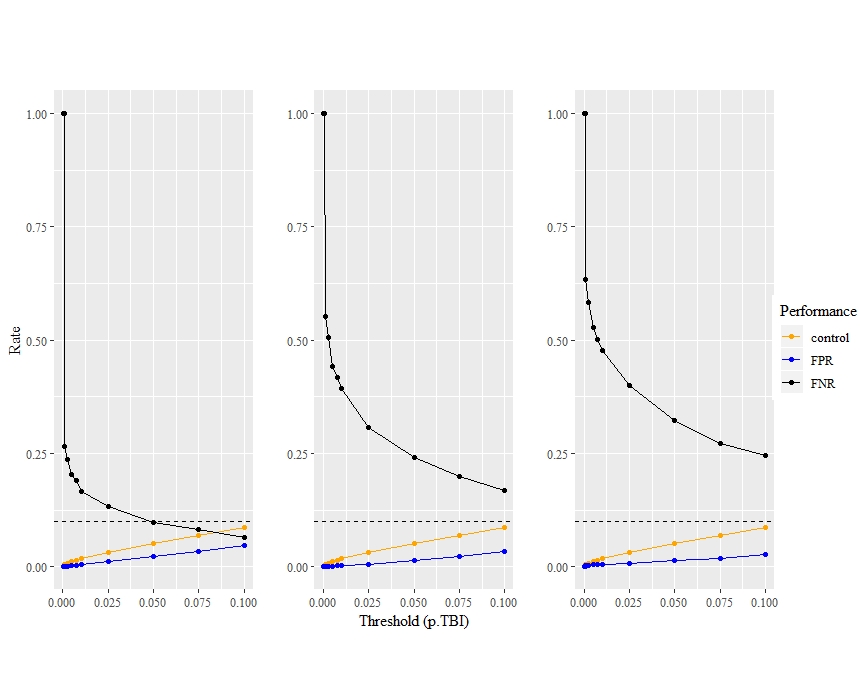
B

A

**Fig. 2.** Influence of dispersal ability on our ability to detect exceptional change. FPR and FNR values at 13 different p.TBI 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 population 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.TBI 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*

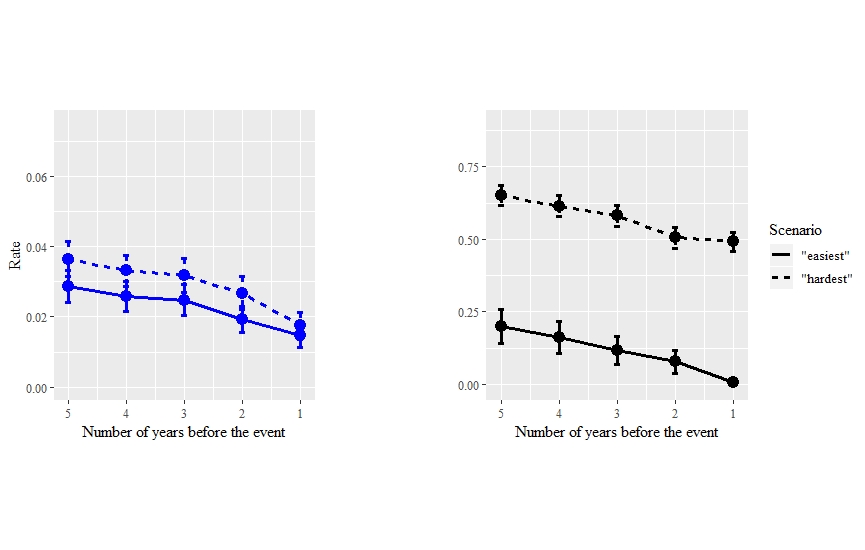
We can see in Fig. 4 A & B, that the longest the pre-event sampling is from the event, the less power and the more false positives we get. Sampling done 5 years before the event led to about twice as much false positives as sampling done the year before the event. The effect of time on FPR or FNR is 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 TBI 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 TBI 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).

*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 *TBImicro()* 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.

*Spruce budworm*

Our method did not identify any Quebec spruce budworm population as having undergone a significant change in genetic diversity, relative to the general change between the years 2012 and 2013. Indeed, the lowest p-value associated with a population-specific TBI was only 0.927 (calculated with 999 permutations), for the easternmost site among the eight: T020 (Larroque et al., 2019).

**DISCUSSION**

Being able to detect which populations have changed significantly, 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, as our permutation approach was generally able to achieve this goal.

Detecting exceptional change is harder in landscape 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 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, 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 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), 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 (REF baleines). 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 TBI is applicable to genetic data 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, 2019). 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, Mahalovich, & Cushman, 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 TBI, 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. 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. This increased realism 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 ways to implement such simulations.

* **Some limits: how would population size or amount of genomic information affect results; magnitude of demographic event**
* **Story about why our results make sense given the highly connected SBW system**
* **Paragraph discussing the importance of LTER, exhaustive sampling, and the need to move beyond single-time snapshot studies of landscape genetics.**

**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. *TBImicro()* 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. We thank Hinatea Ariey for help with the creation of the first figure. Finally, we thank Jeremy Larroque for his comments on an earlier version of the manuscript.

**REFERENCES**

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

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

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

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

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

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

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. (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. 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

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

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

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

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

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

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