**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 Forestry, 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

**INTRODUCTION**

Globally, biodiversity at multiple levels of organization is being lost at an increasing rate with significant consequences for ecosystem functioning and long term viability of the biosphere. (Leigh, Hendry, Vázquez‐Domínguez, & Friesen, 2019). Novel monitoring techniques are needed to track these losses and to inform conservation efforts. Further, it is recognized that it is no longer sufficient to study spatial patterns in biodiversity loss at a single point in time. Instead, one must examine trends and patterns in biodiversity through both space and time.

Spatial and temporal variation in genetic diversity can tell us a great deal about demography and population connectivity (Bradburd & Ralph, 2019; Lowe & Allendorf, 2010). Indeed, population genetics have proven 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). The field of landscape genetics examines interactions between micro-evolutionary processes and landscape features (Manel & Holderegger, 2013; Manel, Schwartz, Luikart, & Taberlet, 2003; Wagner & Fortin, 2013) in order to improve understanding of how spatial heterogeneity influences population genetic processes. The historically spatial focus of landscape genetics, where sampling and/or analysis is done at a single point in time (*e.g.* Wittische, Janes, & James, 2019), may limit the quality and usefulness of inference (Anderson et al., 2010; Draheim, Moore, Fortin, & Scribner, 2018; Martensen, Saura, & Fortin, 2017; Sun & Hedgecock, 2017). Demographically dynamic systems, such as outbreaks, invasions and species declines especially require both a spatial and a temporal perspective.

Temporal variation in genetic diversity, and its drivers, are at the crux of many conservation and public health issues. For example, spatio-temporal genetic studies have led to a better understanding of the invasion history of a major diseases vector species (Maynard et al., 2017) and to the impacts of landscape fragmentation on a food web (Nair, Fountain, Ikonen, Ojanen, & Van Nouhuys, 2016). Temporal genetic variation reflects the evolutionary potential of a population and the probability of its persistence (Aeschbacher, Selby, Willis, & Coop, 2016; Bolnick & Nosil, 2007; Kremer et al., 2012). However, assessing change in spatial genetic variation through time is challenging because population genetic diversity is under the combined influences of recombination, mutation, and demographically-induced genetic drift. Nonetheless, it remains important to develop the capacity to identify changes in genetic diversity through time, specifically when searching for signals of recent demographic changes in the context of ongoing worldwide biodiversity loss.

It is unfortunately rarely possible to directly observe the effects of landscape and climate change on spatial and temporal genetic variation. We can, however, observe these effects through their population genetic legacies (Banks et al., 2013). Although genetic legacies may not be detectable as rapidly as the demographic consequences of change they can persist 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 et al., 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, but they are generally purpose-built for information-rich genetic datasets, which span great sections or the whole genome or are the result of deep sequencing, and are collected at a single point in time. For example, simulation-based frameworks may be used to infer demographic history from at least tens of thousands of loci, based on different demographic scenarios (Excoffier, Dupanloup, Huerta-Sánchez, Sousa, & Foll, 2013; Gutenkunst, Hernandez, Williamson, & Bustamante, 2009; Kamm, Terhorst, Durbin, & Song, 2019). In general, these methods require input beyond simple allele counts, such as information about recombination processes (Gattepaille et al., 2013) and ascertainment bias (Albrechtsen, Nielsen, & Nielsen, 2010; Clark, Hubisz, Bustamante, Williamson, & Nielsen, 2005; Marth, Czabarka, Murvai, & Sherry, 2004) to estimate demographic parameters and history. 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). Detecting significant population genetic changes, relative to what would be expected due to drift, based on limited time series of genetic data remains a challenge, but is highly necessary.

There are many situations where such as detection would prove precious. For example, we could use such a framework to identify which previously sampled populations have undergone significant genetic changes, after a major weather event could have led to higher mortality (Poff et al., 2018; Suárez, Betancor, Fregel, Rodríguez, & Pestano, 2012) or immigration from distantly related populations (Apodaca, Trexler, Jue, Schrader, & Travis, 2013), through long distance dispersal events. Another example would be the monitoring of the genetic diversity of a pest throughout the landscape during an outbreak or a large spawning event, to have a more accurate understanding of when and where populations undergo drastic genetic changes (*e.g.* Larroque et al., 2019; Segura-García et al., 2019). Finally, yet another example could be the evaluation of how the genetic diversities of populations have changed after facing intense and heterogeneous anthropogenic pressure, such as habitat fragmentation and alteration (*e.g.* Baker et al., 2018; Nair et al., 2016). Detecting atypical change in the genetic diversity of populations in all those examples could help with the better management, including the prioritization of conservation or mitigation efforts.

Temporal Beta-diversity Indices (TBI; Legendre 2019) have been used to assess the significance of changes in community composition through time. Given the conceptual similarity between the question of how multi-species communities change through time and that of how genetic diversity changes through time, we expect that TBI can be applied to spatial-temporal multi-locus genotypic data. The TBI approach quantifies temporal change in sampling sites between two points in time using a dissimilarity index. The significance of these dissimilarities is then tested using permutation. The TBI approach has been extensively tested on community composition data (Legendre, 2019b), but its ability to detect meaningful changes in genetic diversity has not yet been examined.

In this study, we expand the TBI framework to be applicable to spatial temporal population genetic data. The objective of our new method, Temporal Genetic diversity Indices (TGI), is 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 and to provide sensible information to policy-makers and managers. Persisting spatial legacies in genetic diversity can also be used to identify sites that were most strongly impacted by previous demographic events. To demonstrate the effectiveness and applicability of the approach, we used a spatially-explicit gene flow simulator (Landguth, Bearlin, Day, & Dunham, 2017). We simulated scenarios in which portions of a landscape are affected by different non-selective demographic changes. We then used TGI to measure changes, losses or gains, in the genetic diversity of our populations under different demographic contexts. Specifically, 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. Performance was quantified using standard false positive/negative rates binary classification. We predict that our ability to detect historical demographic changes would be lower with increasing dispersal ability because of the homogenizing effect of a higher gene flow. We also 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 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. In extending TBI to TGI we considered population-level genotype frequency matrices as input, used genetic distances as dissimilarity, and the null hypothesis became that genetic diversity did not differ between the two points in time that were sampled.

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). Testing the significance of TGI involved permuting the input genotypic matrices.

There are several different ways that one can permute spatial-temporal genetic data and choosing the right way to permute can be important to make the correct inference (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. As it is not known which type of permutation would produce the best performance with genetic data we tested the performance of each of these three permutation approaches in identifying identify statistically significant temporal changes 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 otherwise specified.

*Genetic distance*

Genetic distance between points in time for a given location were calculated using the chord distance. We chose the Rogers’ genetic distance (Avise, 1994; Rogers, 1972), which is very similar to the Euclidean genetic distance. It makes no assumptions about base-pair substitutions or time since separation, and is suitable to study short-term dynamics. We computed the distance using the *dist.genpop* from the *adegenet* R package (see *Software*).

*Simulation framework*

To simulate the dynamics of population genetic changes 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 (see below) 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. Sampling was done before and after the event unless specified.

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 chose to simulate immigration from a diverged population to our landscape, as a demographic event. Our goal was to apply the TGI approach to detect these historical population changes using genetic data. In simulating immigration, we allowed individuals from a 26th separate population to be added to our study area. This independent source population otherwise shared the same attributes as other populations in our simulated landscape. Only during the demographic event, were individuals from the 26th isolated population allowed to disperse to *a priori* defined parts of the landscape by reducing dispersal distance between the 26th and those parts.

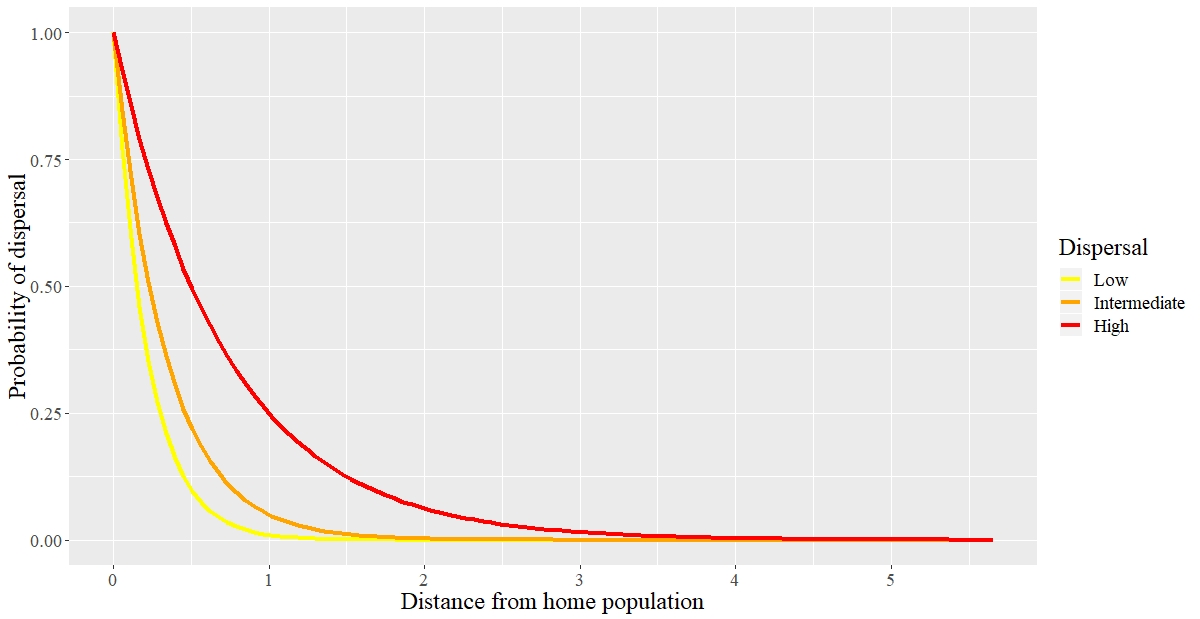
*Dispersal*

To model dispersal, we simply transformed distances between populations by using , where *B* represents how hard it is to disperse with high values of B correspond to low dispersal capacity (elaborated below). We 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. Individuals stay within our simulated landscape, and any individual may disperse to one of our 25 populations at each generation. 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; Table 1).

**Table 1:** Two-factor simulation experiment with scenario abbreviations used throughout the manuscript. The numbers in parentheses indicate the number of unique simulations ran for each factor level or combination of factor levels. We executed 2160 simulations in total.

|  |  |  |  |
| --- | --- | --- | --- |
| *Pop. number \ Dispersal* | **Low** (720) | **Moderate** (720) | **High** (720) |
| **1** (540) | L1 (180) | M1 (180) | H1 (180) |
| **2** (540) | L2 (180) | M2 (180) | H2 (180) |
| **3** (540) | L3 (180) | M3 (180) | H3 (180) |
| **0: control** (540) | CL (180) | CM (180) | CH (180) |



**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 9 years after the event, and compared them with data from the event year. We did the same with the earliest sampling period, that is how far back an earlier sampling can be compared with a sampling done after the event. We chose nine years as the maximum time between samplings as this time gap would represent most of the “before/after” population genetic studies we encountered, and because most long-term ecological research programs monitor at a shorter interval. 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.

*Statistical performance*

We used the False Positive Rate (FPR) and False Negative Rate (FNR) 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 know *a priori* did not undergo the atypical demographic change we forced on it during the simulation, 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), and researchers generally want to keep it as low as possible when there are, for example, heavy costs to focusing on wrong populations such as limited money to invest in a conservation action. A different aspect of performance, a high FNR means that we often miss the right population(s). The higher the FNR, the lower the power of our testing procedure. Researchers may want to minimize the FNR in situations where finding the right population is the most important aspect, for example, if there is limited time to take a conservation action. Because choosing a proper threshold for the TGI permutation tests is important in order to find a compromise between power (1- FNR) and selectivity (1 – FPR), we showcased it to potential TGI users by evaluating statistical performance across a range of thresholds: 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 punctual demographic event and therefore are only subject to the processes of gene flow, drift, and mutation. Dispersal ability was the only parameter varied for the controls, resulting in 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 use control values as a reference.

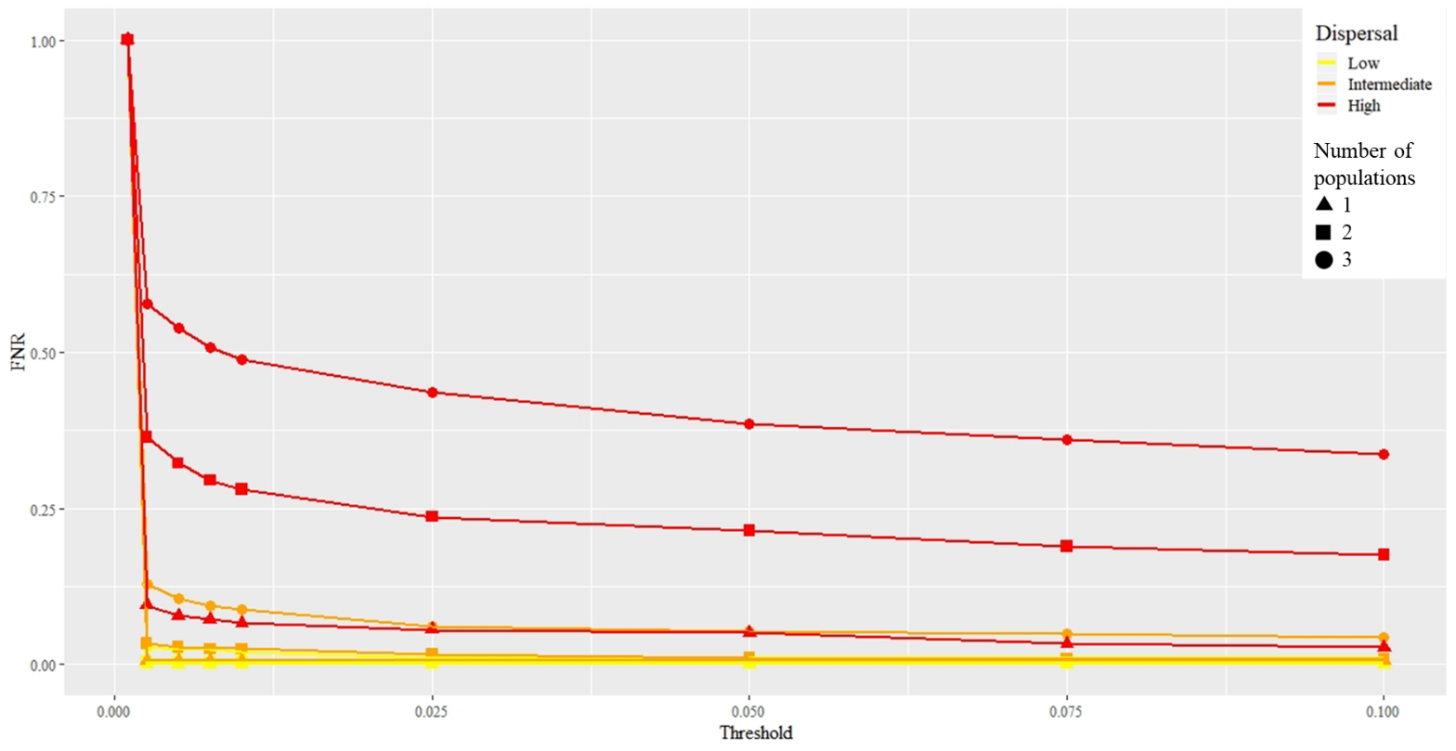
*Software*CDMetaPOP runs on Python 2.7 (Landguth, Bearlin, et al., 2017). We used the R software (R Core Team, 2019) in the RStudio IDE (RStudio Team, 2018) for all analyses and illustration. We used the *adegenet* (Jombart, 2008; Jombart & Ahmed, 2011), *pegas* (Paradis, 2010), and *adespatial* (Dray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

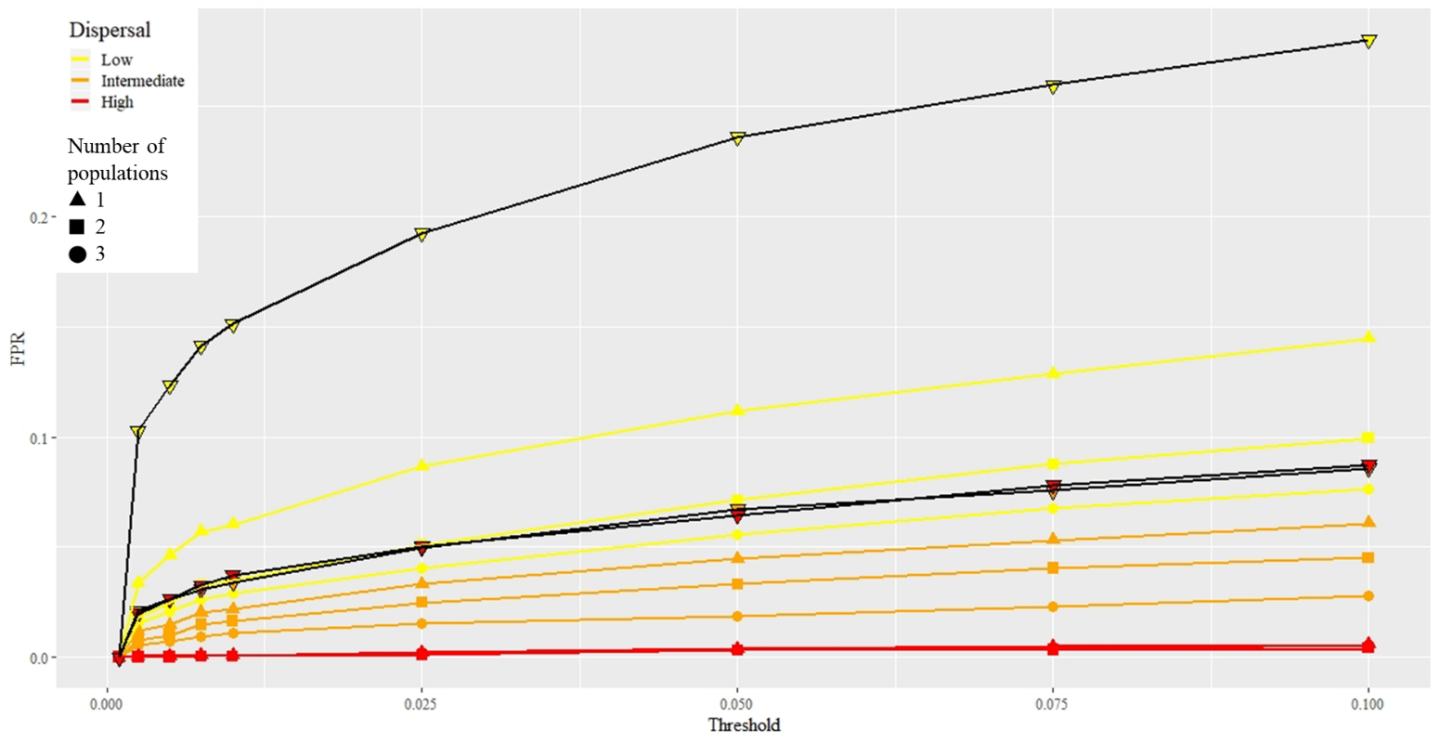
**RESULTS***Dispersal*

As hypothesized, the dispersal capacity influences our ability to detect temporal changes in genetic diversity.

FNR substantially increased with dispersal intensity (Fig. 2). Two scenarios (Table 1) with high dispersal (H2, H3) were the only ones with FNR values above 10% regardless of which threshold is used. FNR values for all scenarios with lower dispersal and two scenarios with moderate dispersal (M1, M2) stayed below 5% except for the lowest threshold which has a value of 1 for scenarios as it is so conservative that it never selects a population as positive. Taking the average from scenario sharing the same dispersal parameters, for the ubiquitous 0.05 threshold, we had FNRs of 0.0046 for low dispersal, 0.0235 for moderate dispersal, and 0.2164 for high dispersal. FNR values overall decrease with threshold, with a sharp decrease before 0.025 followed by a slower decrease until 0.1 (Fig. 2).

FPR substantially decreased with dispersal intensity (Fig. 3). Low dispersal scenarios systematically had higher FPR values than moderate dispersal scenarios, which in turn had higher FPR values than higher dispersal scenario (Fig. 3). If we conservatively define the appropriateness of a FPR value by whether it is below the threshold used in the test, then higher dispersal scenarios more often offered appropriate FPR values (Table 2). Only the high dispersal scenarios (H1, H2, H3) presented appropriate FPR values across all thresholds (Table 2). Conversely, one low dispersal scenario (L1) never satisfied the condition with FPR values consistently higher than the threshold, except for the first one which value is always 0 across all scenarios. FPR averages from scenario sharing the same dispersal parameters, for 0.05 threshold, were 0.0796 for low dispersal, 0.0322 for moderate dispersal, and 0.0035 for high dispersal. FNR values overall decrease with threshold, with a sharp decrease before 0.025 followed by a slower decrease until 0.1 (Fig. 2). Similarly to the FNR, the performance changes more sharply between low thresholds (Fig. 2).



**Fig 2.** FNR across all threshold and scenarios.

**Fig 3.** FPR across all threshold and scenarios. Controls are shown with black lines and triangles.

**Table 2.** Are FPR values staying below the thresholds used in the TGI tests? T stands for “True” and F stands for “False”.

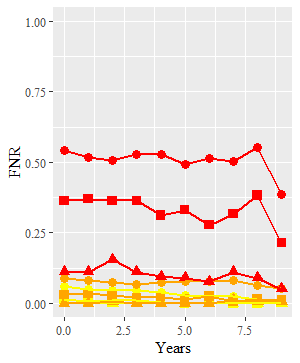
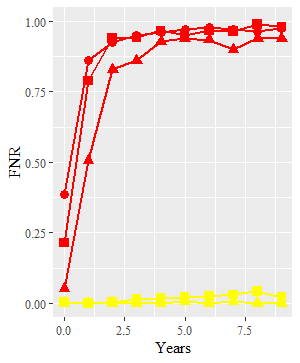
|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| *Scenario*  *--- Threshold* | **L1** | **L2** | **L3** | **M1** | **M2** | **M3** | **H1** | **H2** | **H3** |
| **0.0010** | **T** | **T** | **T** | **T** | **T** | **T** | **T** | **T** | **T** |
| **0.0025** | F | F | F | F | F | F | **T** | **T** | **T** |
| **0.0050** | F | F | F | F | F | F | **T** | **T** | **T** |
| **0.0075** | F | F | F | F | F | F | **T** | **T** | **T** |
| **0.0100** | F | F | F | F | F | F | **T** | **T** | **T** |
| **0.0250** | F | F | F | F | **T** | **T** | **T** | **T** | **T** |
| **0.0500** | F | F | F | **T** | **T** | **T** | **T** | **T** | **T** |
| **0.0750** | F | F | **T** | **T** | **T** | **T** | **T** | **T** | **T** |
| **0.1000** | F | **T** | **T** | **T** | **T** | **T** | **T** | **T** | **T** |

*Number of populations affected*

The number of populations affected by an event also affects our ability to detect exceptional temporal change. Scenarios with a lower number of populations consistently performed better according to FNR or FPR (Fig.2 and 3). While the effect of the number of populations did not affect the performance as much as dispersal, with the levels we used (Table 2; Fig. 2 and 3), a higher number of populations consistently produced higher FNR (Fig. 2) and lower FPR (Fig. 3) for scenarios with the same dispersal parameters. The effect of the number of populations was generally the most substantial on performance, for the lowest performing scenarios in either FNR (high dispersal; Fig. 2) or FPR (low dispersal; Fig. 3).

*Lag time between samplings*

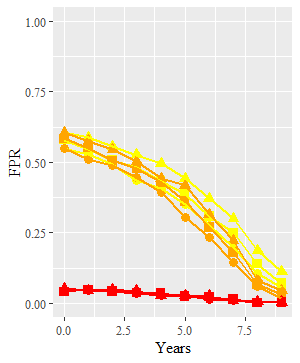
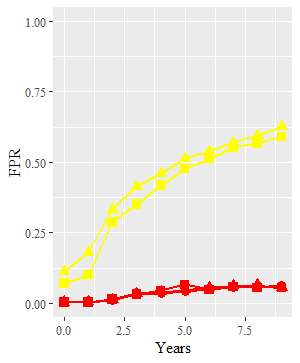
As hypothesized from the nature of genetic processes in connected populations, the genetic signal of the demographic event inflicted upon populations disappears gradually over time. Generally, the longer the interval between a pre-event sampling and post-event sampling, the less power we have to detect it, as evidenced by the increase in false positives and false negatives for several scenarios (Fig. 5 and 6). However, this is strongly affected by the dispersal level present in the landscape. For example, for low and moderate dispersal scenarios, sampling undertaken 9 years before the punctual event led to more than five times the FPR as sampling undertaken the year immediately before (Fig. 6). However, for scenarios with high dispersal, the absolute difference in FPR performance between old samplings and recent samplings is negligible (Fig. 6). A near symmetric relationship between time lag, FPR and dispersal level exists with regards to sampling after the event (Fig. 6). Regarding FNR, this one major difference compared to FPR. Indeed, the relationship between FPR and time is not symmetric in that the age of a sampling conducted before the event matters much less than the time between the event and the second sampling for high dispersal scenarios (Fig. 5). The FNR became prohibitive for second samplings done after two years (Fig .5) while the FPR changed sharply for time lags of 4 years of less and then more slowly (Fig. 6). When considering the scenario most likely to preserve the signal according to earlier results on FNR and FPR (M1, M2, M3), the TGI approach was still able to avoid false negatives reasonably (Fig. 5) but average FPR sharply increased, reached more than 10% of false positives after only two years (Fig. 6) regardless of whether the first or second sampling is responsible for the time lag. 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.



-9 0 9

Years

**Fig 5**. FNR from TGI tests performed between sampling executed up to 9 years before or after the event (arrow) when compared with sampling done the year after the event for prior samplings, or the year before the event for posterior samplings.



-9 0 9

Years

**Fig 6**. FPR from TGI tests performed between sampling executed up to 9 years before or after the event (arrow) when compared with sampling done the year after the event for prior samplings, or the year before the event for posterior samplings.

*Control simulations*

Experimental FPR values consistently stayed below control FPR values, which performance according to their dispersal parameters, followed the same order as experimental FPR values (Fig. 3). This means that in the presence of an actual event, we were always less likely to wrongfully identify a population as having been affected. Finally, control FPR values for scenarios with low dispersal reached a high value, which was about twice as high as the maximum experimental FPR values (L1).