**Detecting temporal changes in genetic diversity using limited information: a new tool for molecular ecology studies with repeated samples**

# INTRODUCTION

Global biodiversity at the gene, species, population, and ecosystem scales is being lost at an increasing rate, with significant consequences for ecosystem functioning and the long-term viability of the biosphere [1–3]. Novel monitoring techniques are needed to track these losses and inform conservation efforts. Furthermore, conservation biologists are increasingly recognizing that it is no longer sufficient to study spatial patterns in biodiversity loss at a single point in time. Instead, trends in biodiversity must be observed across both space and time [4,5].

Information about spatiotemporal variation in genetic diversity can provide important insights into the connectivity and demographic history of different populations [5,6]. Indeed, population genetics has proven essential for translating observed genetic variation into meaningful inferences that can inform conservation efforts [7–9], and the causes and consequences of temporal variation in genetic diversity are at the crux of many conservation and public health issues [10–12]. Researchers commonly explore patterns in spatiotemporal population genetic data [13] to quantify isolation-by-distance [14,15], time since population bottlenecks [16,17], rates of migration between isolated populations [18,19], and the timing and extent of outbreak expansions [20,21].

However, new approaches for detecting temporal variation in genetic diversity must be developed to further elucidate the processes that govern demographically dynamic events such as insect outbreaks, invasions, disease spread, and species declines [4,5,8]. Such tools are especially needed to identify populations that have experienced important historical demographic events, including major weather events, invasions of predators or competitors, disease outbreaks, or other disturbances such as a wildfire [22–26]. These analyses could similarly identify which populations, among a set of previously sampled populations, received migrants from a long-distance dispersal event [27]. Because temporal genetic variation reflects the evolutionary potential of a population and the probability of its persistence [28–30], relating temporal genetic variation to landscape change can provide important insights about the eco-evolutionary dynamics of a species and be used to inform conservation strategies (e.g. Landguth, Holden, Mahalovich, & Cushman, 2017).

There are currently two general approaches for investigating temporal genetic variation. The first suite of approaches uses complex statistical models to infer demographic history from genetic data obtained at a single timepoint [32–34]. This approach is often computationally intensive, requires high-quality microsatellite datasets encompassing a minimum of tens of thousands of loci, and depends on extensive knowledge of the biological system, including information on recombination processes [17] and ascertainment bias [35–37]. The second suite of approaches compares genetic diversity between samples taken from the same site over extended time. These repeated-sample approaches are more readily usable in systems where less information is available, such as non-model species. They can also be used in systems that were sampled historically, with a goal of comparing contemporary patterns with these older data [10].

Despite our ability to compare genetic diversities evaluated at two points in time, there are still several technical and conceptual challenges associated with the analysis of repeated samples. One such challenge is determining how to meaningfully quantify and detect temporal changes. Some studies have used genetic differentiation metrics such as Jost’s D or FST or its analogs [21,38,39] to evaluate temporal changes between genetic datasets. However, translating our spatial understanding of these genetic differentiation indices to the temporal dimension is not straightforward [40]. An additional challenge for temporal genetic analyses is disentangling spatial from temporal effects, because the additivity of genetic drift means that genetic differentiation can be associated with both space and time [41,42]. Analyses that can distinguish natural temporal variation in genetic structure due to recombination, mutation, and demographically-induced drift from the changes caused by external forces remain elusive.

Temporal beta-diversity indices (TBI; Legendre 2019) are used to quantify and assess temporal changes in ecological community composition using a dissimilarity index calculated between samples taken at different times. The significance of these dissimilarities is then tested using a permutational procedure. The TBI approach has been extensively tested on simulated community composition data [43], but the potential of a TBI-inspired tool to detect meaningful temporal changes in genetic diversity has not yet been examined. Given the conceptual similarity between species diversity in multi-species community composition data and genetic diversity in multi-locus genotype data, we suspected that TBI could be modified to identify significant variation in spatiotemporal genotypic data.

In this paper, we propose and evaluate a method for extending the TBI framework to spatiotemporal population genetic data. Our new framework, which we call temporal genetic diversity indices (TGI), is designed to identify significant temporal variation in spatial genetic diversity using relatively information-poor genetic data while accounting for confounding forces such as drift. We demonstrate the effectiveness and applicability of the TGI approach using simulated genetic data, where each simulation included multiple scenarios in which portions of a landscape were affected by a non-selective demographic change. We additionally determined how the ability of TGI to detect significant, atypical temporal variation in genetic diversity was affected by three different demographic contexts: population dispersal ability, the number of populations affected by a demographic event (i.e., spatial extent of the event), and the time between two sampling efforts. With respect to these demographic contexts, we predicted that our ability to detect temporal genetic changes would 1) decrease in populations with higher dispersal capacity because of the homogenizing effect of higher gene flow; 2) \_\_\_\_\_ when the spatial extent of the demographic event was larger because \_\_\_\_\_\_; and 3) decrease when the time between successive sampling events increased, regardless of when an event occurred between samples. Finally, we illustrate how TGI provides a functional testing framework by applying it to a real genetic dataset representing a large landscape with many populations of a threatened vertebrate.

# METHODS

## Adapting TBI for genetic data

Temporal beta-diversity indices (TBI) are calculated by computing dissimilarities in species composition between data sampled at two different times at all sampling sites. The significance of these indices is then tested through simultaneous permutations of the two site-by-species input matrices. To extend TBI to TGI, we substituted community dissimilarities with genetic distances calculated from site-level allele frequencies in order to compare two different temporal samples (see Fig. 1A for a simple example showing how we transformed two temporal samples into a genetic distance, for a two-site landscape). The null hypothesis in this case is that genetic composition between the two time points does not differ more than would be expected due to background processes typical to the landscape.

Indeed, background genetic processes, such as drift, can also produce temporal differences in genetic structure, and so the challenge in designing TGI was to identify temporal changes that are significantly different from what would be expected under a drift-only scenario. Because there are no reference distributions for what constitutes a significant temporal genetic change, we used a permutation-based approach to generate a distribution of values to which an observed value can be compared. For each of the two input genotypic matrices, which were obtained from two temporally distinct samples, we permuted the genotypes at each locus (see Fig. 1B for a simple example showing how we conducted the permutations, for a two-site landscape). Permutations were performed using the *poppr* R package (see *Software*) to maintain allelic structure and heterozygosity [46]. We used 999 permutations for all analyses.

## Genetic distance

Genetic distances between time points for a given location were calculated using Rogers’ genetic distance [47], which is similar to the Euclidean genetic distance (see Annex A). Rogers’ distance makes no assumptions about base-pair substitutions or time since separation and is therefore suitable to study short-term dynamics. It has most recently been used to investigate spatial genetic structure in a pond turtle [48] and a fungus [49]. We computed Rogers’ distance using the *dist.genpop* function from the *adegenet* R package (see *Software*).

## Simulation framework

To simulate the dynamics of population genetic changes through time and test the performance of TGI, we used the spatially-explicit gene flow simulator *CDMetaPOP* [44]. *CDMetaPOP* simulates dispersal and mating of individuals across a landscape and allows the user to define the initial genetic structure, spatial distribution of individuals, dispersal characteristics, and life history traits of the population. The physical landscape we simulated was a homogeneous, interconnected 5 × 5 square grid, with each of the 25 cells representing a population. Each population had a maximum carrying capacity of 50 individuals, and t, (25 × 50) Distance between populations was set as the Euclidean geographic distance. The genotypic information of each individual consisted of 100 neutral, unlinked, bi-allelic single nucleotide polymorphism (SNP) loci. The mutation rate was set at 10-8 [50].

Multiple demographic processes affect the spatial apportionment of genetic variation. For this study, we investigated the influence of a single demographic event: immigration from a previously isolated population. We simulated immigration from a population that was separate from our 5 × 5 grid (i.e., population #26), with the goal of applying the TGI approach to detect historical changes in population genetic data due to immigration. This independent source population shared the same attributes as other populations in our simulated landscape and was only allowed to disperse into the 5 × 5 simulation grid during simulated demographic events.

Using this model, we examined the influence of dispersal (movement among our 25 populations) and the spatial extent of demographic events (number of populations that received immigrants from population #26) on the persistence of spatial genetic legacies. We examined three levels of dispersal capacity (described below) and three different numbers of affected populations (1, 2, or 3) for a total of nine unique scenarios (Table 1). Each scenario was then simulated 180 times, for a total of 1,620 (9 × 180) unique replicates for this experiment, excluding the control simulations (Table 1). In the next sections, we detail how we modeled dispersal and spatial extent.

For each replicate, we initialized the simulation with random, unique allocations of alleles among individuals, therefore approaching maximum genetic diversity [51]. Those parameters were chosen as a compromise between realistic allele distributions and computational limitations, and were appropriate for producing simulated genetic data that could reasonably recreate the complex evolutionary dynamics in real populations. Each simulation was run for 100 generations before the demographic event was imposed on up to three populations in the landscape. Ten additional generations were simulated after the event. performed

## Dispersal

To model dispersal, we weighted the geographic distances between populations using a power law function, , where *B* represents the difficulty of dispersal. High values of *B* correspond to low dispersal capacity. We then rescaled the values, using the maximum and the minimum (0) distances possible in this virtual landscape, as described in the *CDMetaPOP* [44] user manual (p.63). This produced values in the [0,1] range. Rescaled values were considered to represent probabilities that an individual disperses to a cell located at that distance (Fig.1). We chose this way of modeling dispersal to allow for both within-population movement and long-distance dispersal.

The population to which an individual dispersed was selected randomly from the set of populations available at the given distance. Individuals always stayed within our simulated landscape, and any individual could disperse to any one of the 25 populations at each generation. To investigate the effect of different levels of dispersal, we ran separate simulations using three different values of *B*: low (*B* = 2), moderate (*B* = 1.301) and high (*B* = 0.6015) dispersal capacity (Fig.1; Table 1).

## Spatial extent

We also wanted to evaluate how the spatial extent of the simulated immigration event affected the performance of our TGI method. To do this, we allowed individuals from population #26 to immigrate into one, two, or three populations that were randomly selected from the original 25. We decided to vary the position of where the demographic event occurred in the landscape because deme topology may influence the outcomes of population genetic analyses [52]. For scenarios in which only one population was affected, we partitioned the 180 simulations equally among six populations in the landscape. The positions of these six populations were randomly selected once and were identical across runs. Indeed, because our landscape is square and homogeneously resistant to movement, it is completely symmetric, and therefore only six unique positions need to be assessed. When multiple (two or three) populations underwent a demographic event, we randomly chose one of these six geographically unique populations and randomly picked one or two additional populations directly adjacent (when possible) to it. We chose to pick adjacent populations to respect the spatial autocorrelation often exhibited in demographic events. For each of the two- and three-population simulations, we repeated this population selection procedure six times and ran 30 replicate simulations for each set of populations.

## Statistical performance

We assessed the statistical performance of our TGI testing procedure using the false positive rate (FPR) and false negative rate (FNR). In our study, a false positive was a population that we knew did not undergo the demographic change we imposed but was found to have done so using the TGI test, whereas a false negative was a population that experienced the demographic event but exhibited no significant change in the TGI test. The FPR is expressed as the ratio of false positives to the total number of negative tests (i.e., true negatives and false positives), and the FNR is expressed as the ratio of false negatives to the total number of positive tests (i.e., true positives and false negatives).

A high FPR would indicate that our TGI measure often selected the wrong population(s) as having changed substantially and that our testing procedure was less selective. Researchers generally want to minimize the FPR when there are, for example, limited resources available for conservation efforts. In contrast, a high FNR would mean that we often failed to identify the population(s) that were actually affected and that our testing procedure had less discriminatory power. 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 conservation action. In addition, selecting a proper significance threshold for the p-value calculated from the TGI test permutations, and therefore defining which changes in genetic diversity are significant or not, is important for balancing selectivity (1 – FPR) and power (1 – FNR). To characterize this compromise, we evaluated the statistical performance of TGI using a range of significance thresholds for calculating FPR and FNR: 0.0001, 0.001, 0.005, 0.01, 0.015, 0.020, 0.025, 0.030, 0.035, 0.040, 0.045, 0.050, 0.055, 0.060, 0.065, 0.070, 0.075, 0.080, 0.085, 0.090, 0.095, and 0.1

## Time

To assess how the time since the simulated demographic event affected our ability to detect genetic changes in each of the nine dispersal/spatial extent scenarios, we calculated the TGI for simulated data collected up to nine years before and after the event and compared it to the TGI calculated from data collected immediately before and after the event year. We chose nine years as the maximum time between samplings because this timeline is longer than most “before/after” population genetic studies in the literature and most long-term ecological research programs monitor at a shorter time interval. Comparisons between TGI results were based on the FPR and FNR calculated at a significance threshold of p < 0.05, as this threshold was a good compromise between different performance metrics in our earlier results.

## Controls

We additionally ran control simulations in which populations were not affected by demographic events and were therefore only subject to the processes of gene flow, drift, and mutation. Dispersal was the only parameter that varied among the control simulations, resulting in three control scenarios (Table 1). We only evaluated the FPR of these control scenarios; because there were no true positives or false negatives for populations affected by the demographic event, the FNR was always equal to 0. The performance of experimental scenarios was always compared to the control scenario with the same dispersal capacity.

## Software

*CDMetaPOP* runs on *Python 2.7* [44]. We used the *R* software [53] in the *RStudio* IDE [54] for all analyses and illustrations. We used the *adegenet* [55,56], *pegas* [57], *poppr* [58,59] and *adespatial* [60] *R* packages for calculations. Our *TGI* function is available in the appendix as an *R* script.

## Applied example

To demonstrate that our TGI measure provides valuable information about temporal change in a real system with conservation implications, we applied it to real genetic data from a study of a threatened vertebrate, the Northern tidewater goby [61]. We chose this example because it uses a different type of genetic data than we used for our simulations, thus demonstrating that TGI is applicable to a variety of genetic markers. In addition, the study authors suggested that one goby population had undergone more genetic change than the other, more stable local populations, allowing us to test a real hypothesis and go beyond a simple illustration of our method [61]. The dataset was downloaded from DRYAD (doi: 10.5061/dryad.871db), and we used 9,999 permutations for this analysis.

# RESULTS

We were able to translate the TBI framework to TGI by adapting it to the specific structure of genetic data. Results from a wide array of simulations suggested that TGI could consistently and accurately identify populations that have experienced a demographic event. Although our results support the general efficacy of TGI and warrant its use on empirical datasets, we noted that the accuracy of TGI varied in relation to the dispersal capacity of the population, the spatial extent of the demographic event producing the genetic change, and the time difference between sampling and the demographic event.

## Dispersal

Dispersal capacity influenced our ability to detect temporal changes in genetic diversity, as the FNR generally increased with dispersal intensity (Fig. 3). However, only one scenario (H3; Table 1) exhibited FNR values above a very conservative limit of 1%, regardless of the p-value threshold used (Fig. 3). Of the four scenarios that did not achieve an average FNR of 0 (L3, M3, H2, and H3), two involved high dispersal. When we averaged the FNR values calculated at the traditional p < 0.05 threshold across scenarios sharing the same dispersal parameters (e.g.one averaged FNR value for L1, L2, and L3 grouped together), the mean FNRs were 0.0037 (0.0007 - 0.0066; 95% confidence interval [CI]) for low dispersal, 0.0049 (0.0015 - 0.0083; 95% CI) for moderate dispersal, and 0.0108 (0.0055 - 0.0161; 95% CI) for high dispersal. FNR values also decreased with the chosen significance threshold, with a sharp decrease (most notable for H3) before 0.025 followed by a slower decrease until 0.1.

In contrast, dispersal capacity did not substantially affect the FPR, as low dispersal did not consistently result in a higher FPR than moderate or high dispersal (Fig. 4). Indeed, there were no consistent trends when comparing scenarios with the same number of affected populations: L1 had slightly higher values than M1 and H1; L2 had slightly lower values than M2 and H2; and L3 had intermediate values between those of M3 and H3. Average FPR values for scenarios sharing the same dispersal parameters, calculated using FPRs at the p < 0.05 threshold as before, were 0.0599 (0.0558 - 0.0641; 95% CI) for low dispersal, 0.0621 (0.0580 - 0.0662; 95% CI) for moderate dispersal, and 0.0600 (0.0562 - 0.0638; 95% CI) for high dispersal. FPR values also increased with the chosen significance threshold, with a sharp increase at low thresholds followed by a continued but saturating increase until p < 0.1.

Another encouraging indicator of the efficacy of TGI was that experimental FPR values were consistently lower than control FPR values, regardless of dispersal parameters (Fig. 4). This suggests that, following an atypical demographic event, we were always more likely to correctly identify a population as having been affected. Among the control simulations, simulations with higher dispersal capacity had a lower FPR (Fig. 4). Control FPR values were generally at least twice as high as the maximum experimental FPR values (L1, M1), regardless of the significance threshold used. This means that, even for the lowest-performing scenarios in our simulations, TGI was much more effective at avoiding false positives in the presence of an event than in the absence of one.

## Spatial extent

The number of populations affected by a demographic event also influenced our ability to detect meaningful temporal change. Scenarios in which fewer populations were affected exhibited a reduced FNR and an increased FPR (Figs. 3, 4). Scenarios in which a single population was affected (i.e., L1, M1, H1) had a perfect FNR (0; Fig. 3), while scenarios L2 and M2 only reached this perfect FNR at more liberal significance thresholds (i.e., above p < 0.03; Fig 3). The mean FNRs at p < 0.05, averaged across scenarios sharing the same number of affected populations (e.g.one averaged value for L1, M1, and H1 grouped together), were zero for scenarios with one affected population, 0.0028 (0 - 0.0059; 95% CI) for scenarios with two affected populations, and 0.0167 (0.0105 – 0.0228; 95% CI) for scenarios with three affected populations.

The number of affected populations also appeared to influence the FPR performance more than dispersal, at least for the dispersal capacities and spatial extents used in our simulations, as FPR values were consistent across scenarios with different dispersal but the same number of affected populations rather than across scenarios with similar dispersal but different numbers of affected populations (Figs. 3, 4). The average FPRs from scenarios with the same number of affected populations, determined at the p < 0.05 significance threshold, were 0.0820 (0.0778 - 0.0863; 95% CI) for scenarios with one affected population, 0.0553 (0.0516 - 0.0591; 95% CI) for scenarios with two affected populations, and 0.0447 (0.0413 - 0.0481; 95% CI) for scenarios with three affected populations.

## Time

We found that the genetic signal of the demographic event decayed over time, but that TGI was still able to identify significant changes in genetic diversity at a time scale of 1-10 years. However, as the time interval between pre- and post-event sampling increased, the ability of TGI to detect the demographic event decreased, evidenced by the increase in false positives and false negatives for several demographic scenarios (Fig. 5, 6). The effect of time between sampling periods on the sensitivity of TGI was strongly affected by dispersal capacity and the extent of the event.

The timing of sampling prior to a simulated event was, as expected, generally less important than the timing of the post-event sampling. The decrease in genetic signal over time – which would be found with any comparative method, not just TGI – was purposely strong in our simulations. For example, if the second (post-event) sample was taken nine years after the first (pre-event) sample, we observed high FNR values that approached 75-90% in high- and moderate-dispersal scenarios (Fig. 5). The FNR also increased with the time lag in low-dispersal scenarios, but the increase was more linear and values never reached 30%, even after nine years (Fig. 5). One interesting observation was that the number of affected populations was the main factor driving increasing FNR values when the pre-event sample was older (3>2>1; left side of Fig. 5), while dispersal capacity was the main factor driving increasing FNR values when the post-event sample was older (H>M>L; right side of Fig. 5). For scenarios with the same number of affected populations, moderate-dispersal scenarios showed the worst performance with pre-event sampling time lags, whereas high-dispersal scenarios generally showed the worst performance with post-event sampling time lags (Fig .5). Over our nine-year sampling window, the FNR changed the least for the L1 scenario and the most for the H3 scenario (Fig. 5).

While the relative differences in FPR performance given different time lags were not as high as for FNR, FPR nonetheless increased with the sampling time lag. For example, the FPR for scenario L1 increased to more than twice the value of the threshold for pre-event samplings, and the FPR was even higher as the post-event sampling lag approached nine years (Fig. 6). There were no clear patterns for whether dispersal or the number of affected populations most influenced the change in FPR associated with pre-event sampling time (Fig. 6); however, dispersal was the main factor driving FPR for time gaps associated with post-event sampling (Fig. 6). The strong relationship that we observed between FPR and the number of populations affected by the demographic event therefore became less pronounced as dispersal became more influential. As with the FNR, the FPR did not change much for the L1 scenario and changed the most dramatically for the H3 scenario (Fig. 6), indicating differences in how time affects our two most extreme scenarios that could be a useful consideration for potential TGI users.

The simulation that was most likely to preserve the signal of the demographic event, was the low-dispersal scenario with a single affected population (L1). In this scenario, the TGI approach was still able to keep false negatives below 15% and false positives below 10%, even when the second sampling was done nine years after the event (Fig. 5, 6) and regardless of whether the first or second sampling was responsible for the time lag.

## Thresholds

## Applied example

The Northern tidewater goby (*Eucyclogobius newberryi*) is a small endangered fish that lives in brackish estuaries and lagoons along the coast of California. This species represents an interesting model for population genetic studies because dispersal between suitable habitat patches only occurs during rare, discrete events. A previous study investigated extinction–colonization dynamics in the tidewater goby by evaluating genetic diversity across the landscape at several points in time [61]. Those authors suggested that the Elk River goby population had experienced unexpected temporal genetic change between 2006 and 2011 [61]. We used the TGI to re-analyze this dataset, as the TGI provided a rigorous test of whether significant temporal genetic change had indeed occurred in any population in this landscape.

Using our TGI measure, we found that the genetic structure of the Elk River population of Northern tidewater goby [61] has indeed changed significantly relative to the other populations sampled in the study area (permutation p-value = 0.0005), even after using strict p-value adjustments (Holm-Bonferroni adjusted permutation p-value = 0.004). Our results thus confirm the purely descriptive results of the previous study, that there was a loss of genetic diversity in the Elk River population, with a robust statistical framework.

# DISCUSSION

## A new and useful framework

TGI fills a gap: it provides a framework for comparing temporally-repeated samples and testing whether an observed change in genetic diversity is significant relative to the ubiquitous and landscape-wide changes associated with genetic drift. This approach is robust even when the number of available markers is limited. Our results suggest that we successfully adapted TBI to genetic data, which required translating a species-by-site approach to a genotype-by-site approach and changing the permutation algorithm to accommodate the specific structure of various genetic data formats. In addition to describing and found that \_\_\_\_\_\_\_\_\_One main contrast between our new TGI approach and previous investigations of the performance of TBI, which used community composition data, is that we also examined how the timing of sampling may affect the downstream conclusions. Although community composition data generally varies across a longer time scale than genetic data, we encourage future investigations of how sample timing affects TBI performance.

In this study, we specifically investigated how dispersal, the spatial extent of a demographic event, and the timing of sampling affected our ability to identify populations that have experienced significant changes in genetic diversity. By modifying these parameters in various demographic simulations and evaluating whether TGI could correctly identify populations affected by a significant immigration event, we were able to identify the range of systems in which our TGI procedure can be successfully applied.

## Dispersal and spatial extent

The ability of our method to detect temporal genetic changes was sensitive to the dispersal capacity of the organism of interest: false negatives increased with dispersal capacity, though false positives did not show a clear trend (Fig. 3, 4). The influence of dispersal capacity on the FNR was also affected by the time lag between an event and the subsequent sampling effort; the effects of different dispersal capacities were evident even when samples were separated by only one generation (i.e., samples were collected immediately before and after the event) and were magnified as the time between samplings increased. Dispersal capacity is also functionally linked to landscape connectivity, as connectivity represents the degree to which a landscape constrains dispersal [62]. Considering that we used dispersal ability as a proxy for landscape connectivity, the effects of sampling time and dispersal capacity on the FNR suggest that well-connected landscapes might require more frequent sampling to overcome the negative effect of connectivity on our ability to correctly identify affected populations.

The spatial extent of a demographic event increased our ability to correctly identify populations that have not truly changed (lower FPR), but it also decreased our ability to correctly identify populations that have (higher FNR). The magnitude of this trade-off varied with dispersal capacity. Although a broader spatial extent may help researchers detect an event, as the chance of sampling an affected population increases, it may also increase the risk of not identifying the genetic legacy of the event at all, especially in high-dispersal landscapes. Given the way it was designed, TGI is better suited to analyze systems in which disturbances are discrete and affect a limited part of the landscape. It is less effective for analyzing gradual, landscape-wide disturbances. In addition, when multiple populations were affected in our simulations, we always chose to affect adjacent populations; we did not investigate whether lowering the degree of spatial autocorrelation in the spatial genetic legacy (i.e., targeting populations not necessarily adjacent to each other) influenced our ability to detect the event. Spatial autocorrelation may greatly affect many genetic analyses, and new analyses are being developed to account for it [63]. We believe that explicitly accounting for spatial autocorrelation in temporal analyses of genetic diversity [5] represents a promising and challenging avenue for future research.

Overall, our results on how the dispersal capacity of an organism and the spatial extent of a demographic event affect the TGI indicate that different performances are to be expected for demographically different systems, and that researchers will greatly benefit from considering the nature of their system when interpreting TGI and choosing an appropriate significance threshold in the context of their specific objectives. For example,some researchers may wish to prioritize avoiding false negatives rather than false positives – thus ensuring that we detect all the affected populations no matter the cost of detecting, and therefore monitoring and preserving, some populations which do not need preservation.

## Time between sampling efforts

As expected, spatial genetic legacies decayed over time, though we found that TGI was generally suitable for identifying changes over 1-10 years given our landscape and demographic parameters. However, two main points emerged from our analysis of how the timing of sampling affected the detection of significant genetic changes. First, when comparing an old sample to a sample collected soon after a potential disturbance, the spatial extent of the disturbance affected the power of TGI, with smaller spatial extents preserving high power even with large time gaps. This result should be reassuring to potential users planning to use TGI in similar systems. Second, when comparing a sample collected immediately before a disturbance to one collected several years after, dispersal was the most important factor driving the performance of TGI, with low-dispersal scenarios better preserving the performance of TGI against the decaying effects of time and genetic drift. Researchers in high-dispersal systems could find as many as 10% of false positives even when sampling only a few years after an event. This has serious implications: if researchers are unaware of the dynamics of their system and arbitrarily choose an inappropriate significance threshold, they might systematically spend a larger part of their resources monitoring or treating unaffected populations while missing some affected populations.

## Empirical application of the method

We successfully applied TGI to an empirical dataset from an endangered vertebrate, the Northern tidewater goby, for which temporal genetic change had been described but not tested. The authors of the original publication hypothesized that one goby population had undergone atypical genetic change, relative to the rest of the landscape; our application of TGI supported this hypothesis. We therefore demonstrated that the straightforward TGI testing procedure can be used to strengthen the results from temporal genetic studies that use repeated samples. Different empirical datasets and research objectives may require TGI users to tweak our procedure, but the TGI function is transparent and flexible, and different permutation and genetic distance algorithms could easily be used by simply changing a few lines in the annotated TGI function provided in the supplementary material.

## Considerations about the use of TGI

Our TGI procedure has a demonstrated place in a researcher’s arsenal for studying genetic change, but there are still several considerations for its effective use. For example, stricter (lower) values for the TGI p-value threshold expectedly result in a lower FPR but may also result in a higher FNR (lower power). Identifying the most sensible threshold for a chosen objective would be valuable to better understand the trade-offs of different sampling schemes in specific systems. TGI can also readily be used on other types of genetic data, such as microsatellites.

TGI provides a robust statistical framework beyond arbitrarily comparing pairwise genetic differentiation, or node-based genetic diversity values. TGI was not designed as an alternative to the sophisticated and well-established methods for inferring demographic history from large genetic datasets collected at a single time. Instead, it was designed to help research teams collecting repeated samples from non-model organisms with limited genotypic information, and especially for teams wanting to compare new samples to older ones. Nonetheless, more work is needed to explore how the performance of TGI varies with factors that were not tested in our simulations, including 1) the chosen genetic distance algorithm; 2) spatial autocorrelation in genetic legacies; 3) effective population size; and 4) spatial heterogeneity in landscape resistance to movement.

The implementation of TGI in new systems will ultimately be more successful if researchers have an *a priori* understanding of the population dynamics of their system and the nature and scale of possible disturbances in their study area. Indeed, this prior knowledge could help researchers choose a range of significance thresholds appropriate to their study system. These values ultimately represent trade-offs in potential conservation costs, and it is therefore essential that researchers grasp their importance and choose these values deliberately.

As we have demonstrated in this study, simulation is a powerful tool for investigating how demography and spatial context influence population genetic dynamics [64], and simulations can be used to help identify appropriate threshold values. Several programs, including *CDMetaPOP* [44], *Nemo* [65], *SPLATCHE* [66], and *SLIM* [67], provide flexible and sophisticated ways to implement such simulations. We expect greater sensitivity to threshold selection in systems that exhibit dramatic demographic fluctuations, as is the case for outbreaking or invasive species. Those systems would therefore benefit from a simulation study designed to systematically identify the most appropriate threshold for the analysis of empirical data, rather than reliance on an arbitrary threshold.

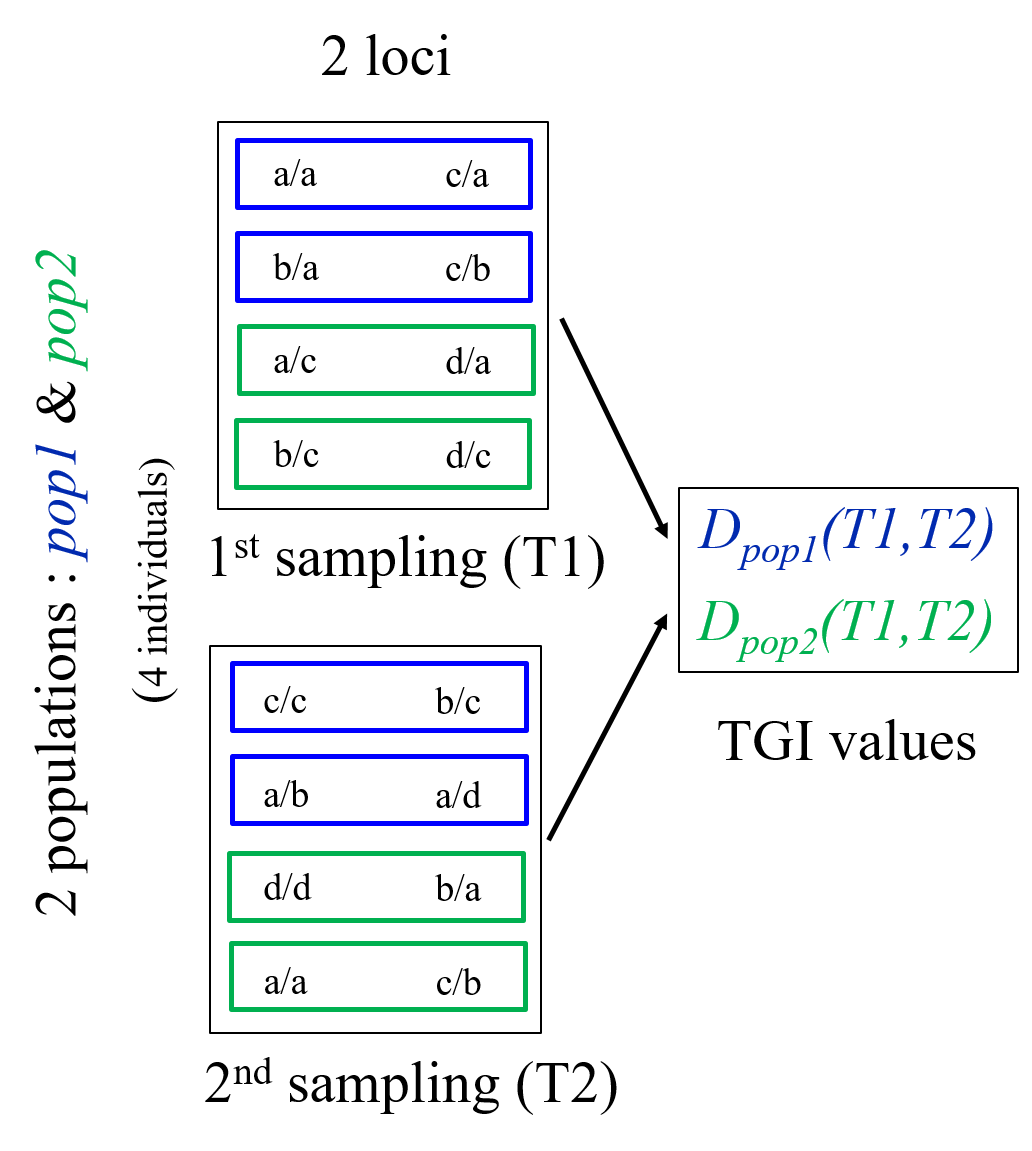
## Conclusions

Identifying changes in genetic diversity, beyond the expected changes due to background micro-evolutionary processes, can help researchers and conservation managers identify locations or populations that have experienced important past demographic events. These events could be detrimental (e.g., loss of diversity or maladaptation) or beneficial (e.g.,higher effective population size or genetic rescue). Such locations and populations could then be prioritized for increased monitoring and further investigation into the origin of these changes. As shown in our application of TGI to empirical data from the endangered Northern tidewater goby, our method provides a framework for directly testing hypotheses about exceptional temporal genetic change. Our approach to detecting temporal genetic differentiation does not require extensive genomic information and can therefore be used to explore the temporal dynamics of genetic diversity changes using relatively small genetic datasets (e.g., hundreds of SNPs). We believe that the TGI approach is a promising tool for the spatiotemporal analysis of wild, non-model organisms for which extensive genomic resources are yet to be developed.

# TABLES AND FIGURES

**Table 1**

|  |  |  |  |
| --- | --- | --- | --- |
| *Dispersal (B)*  *No. populations* | **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) |

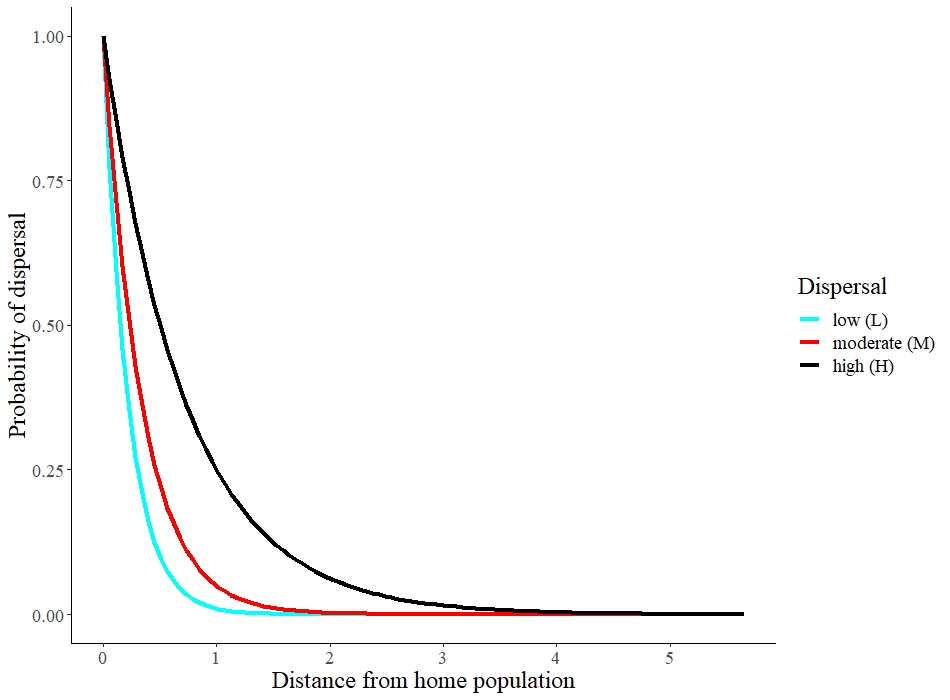
****

A)

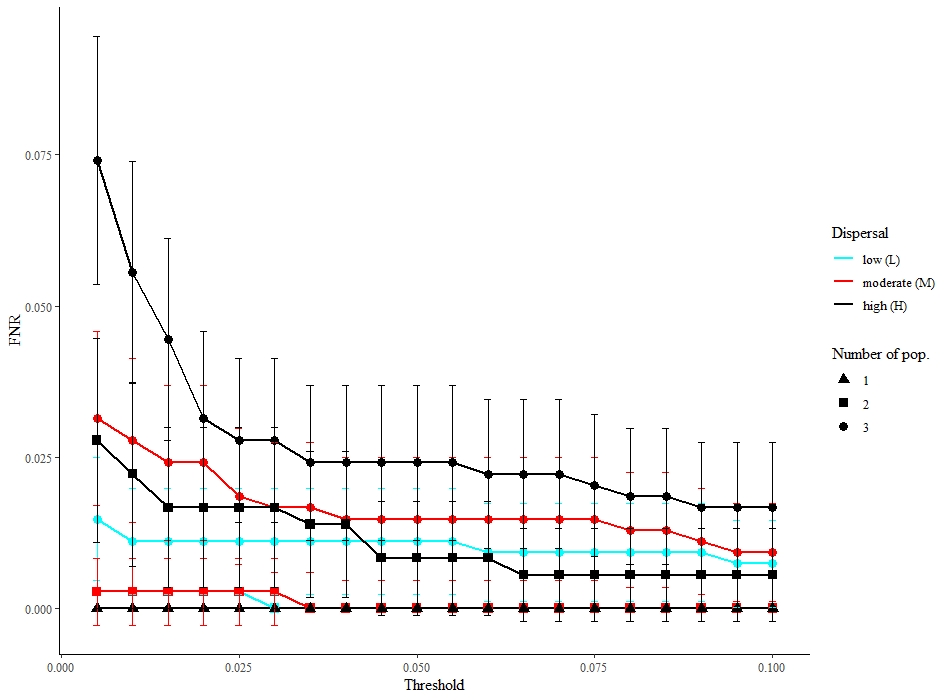
****

B)

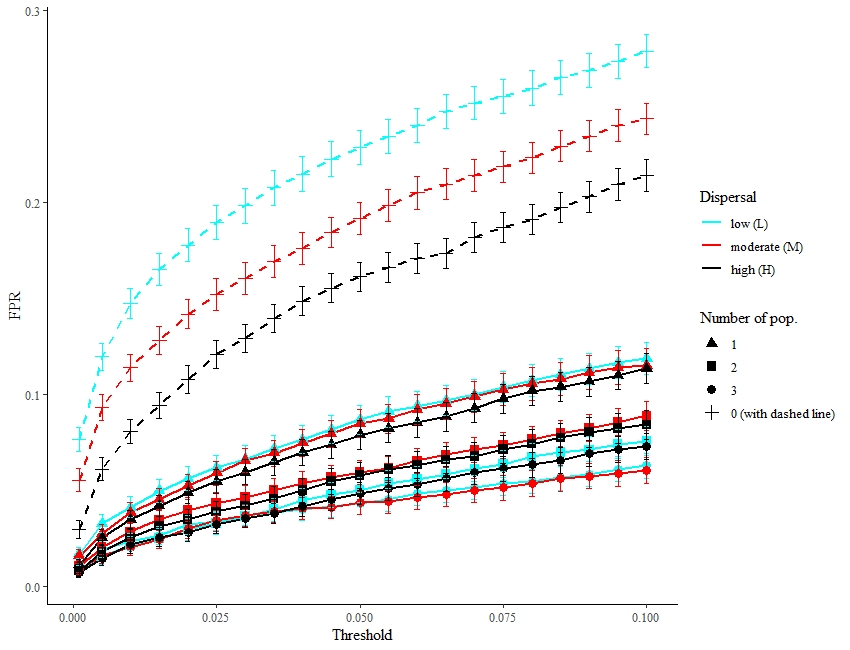
**Figure 1.**



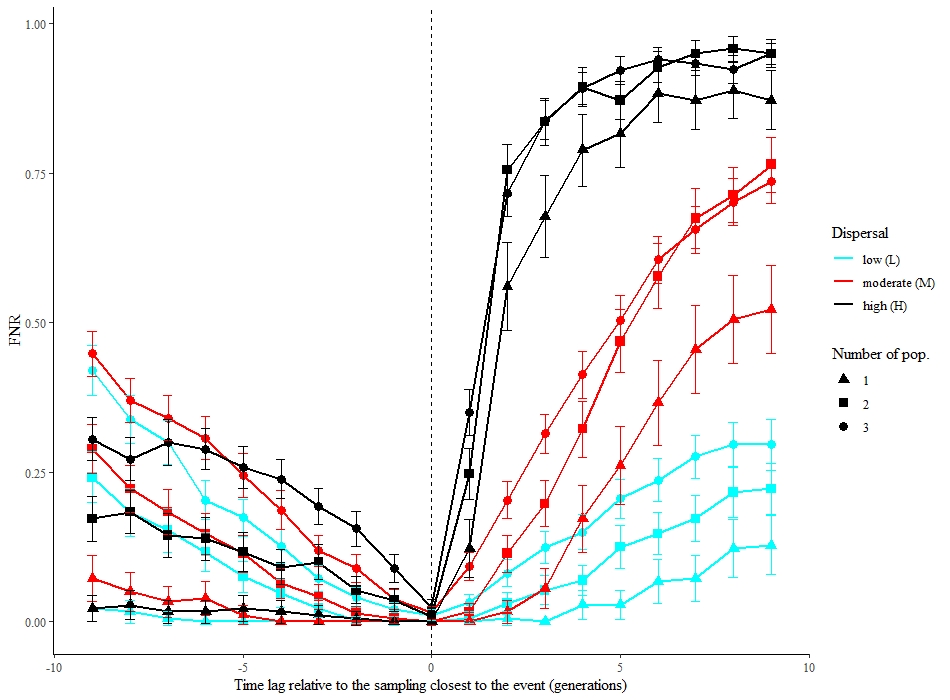
**Figure 2.**



**Figure 3.**

****

**Figure 4.**

****

Influence of the timing of the posterior sampling

(0 represents sampling right after the event)

Influence of the timing of the prior sampling

(0 represents sampling right before the event)

**Figure 5**.

****

Influence of the timing of the posterior sampling

(0 represents sampling right after the event)

Influence of the timing of the prior sampling

(0 represents sampling right before the event)

**Figure 6.**