**Detecting exceptional neutral genetic changes in resampled landscapes.**

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

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

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

Running title: Unique spatio-temporal genetic change

INTRODUCTION

Global change, including climate change but also habitat destruction and fragmentation, have caused biodiversity to quickly decline in many parts of the world in the last century (Fischer & Lindenmayer, 2007; Butchart et al., 2010; Dirzo et al., 2014). The future of biodiversity is likely bleak (Bellard, Bertelsmeier, Leadley, Thuiller, & Courchamp, 2012), and thus there is an ever-increasing demand from ecosystem managers to evaluate and mitigate biodiversity loss, but also to assess current and proposed management plans (Brondizio, Settele, Díaz, & Ngo, 2019).

Global change ecology is still an emerging field and improvements are being made at a fast pace (Rands et al., 2010). Among the main axes of improvement followed by researchers, are increasing our understanding of synergies between drivers of ecosystem services and biodiversity loss, covering previously poorly described biodiversity (including genetic diversity), and increasing our predictive abilities, notably through model validation and development. Although global change trends, and associated uncertainty, have long been closely monitored and described (Sala et al., 2000; IPCC, 2014), more local information about within-landscape temporal changes is needed to further our ability to predict change (Randin et al., 2009; Potter, Arthur Woods, & Pincebourde, 2013; Yates et al., 2018).

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

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

While the development of tools to identify aberrant loci (Frichot & François, 2015; Luu, Bazin, & Blum, 2017; Duruz et al., 2019) or classify population samples in genetically coherent clusters (Pritchard, Stephens, & Donnelly, 2000; Jombart, Devillard, & Balloux, 2010; Caye, Deist, Martins, Michel, & François, 2016; Verity & Nichols, 2016) is thriving, options offering to test whether a sample has truly changed relative to others, are less visible. Indeed, although one could use existing clustering methods, their relevance and performance on temporal datasets, where the objective is to find which population has indeed changed more significantly than others in the landscape, has not been evaluated.

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). A number of applications of the method on composition data have been made on a range of systems (Legendre & Salvat, 2015; Winegardner, Legendre, Beisner, & Gregory-Eaves, 2017; Kuczynski, Legendre, & Grenouillet, 2018; Legendre & Condit, 2019) and its power and type 1 errors have recently been thoroughly tested through simulations (Legendre, 2018). Temporal Beta-diversity Indices (TBI; Legendre 2018) were designed to asses whether there are sites where the difference in community composition between survey times seems exceptionally large and it was suggested that one could use TBI on gene frequency data. The method involves estimating temporal change in each sampling site between two dates using a dissimilarity index/distance, testing the significance of each change through permutations, and partitioning the change into losses and gains. Comparing genetic data at two different dates separated by a known event may help us understand more about the ecological processes shaping the system, indicate the parts of the landscape where an event had the strongest effects, or highlight which sites should be investigated in the absence of an *a priori* known event.

Our paper seeks to describe how to find out find out which parts of a landscape have undergone atypical and substantial genetic change, for example after a disturbance event. We simulated scenarios where part of the landscape is affected by non-selective demographic changes mimicking the effects of common disturbance events. We then used TBI to measure changes in the gene pool of our populations, and evaluated the power and error rates associated with this approach. Finally, we illustrated the possibilities of this approach through applications on two real genetic datasets.

METHODS

*Simulation framework*

To model the variation of 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 and therefore are bi-allelic. The mutation rate was set as to reflect empirically-derived mutation rates found in many taxa (REF I gave to Ryan). There was no selective pressure in the virtual landscape and so all loci were considered neutral. Simulated individuals each carried a genome of 500 loci without linkage disequilibrium. We simulated 1000 replicates per scenario, with randomly allocated initial conditions for each replicate. Each simulated population had a maximum carrying capacity of 100 individuals and each simulated landscape comprised 25 (5 by 5) interconnected populations with structural connectivity only reflecting geographical distance.

The dispersion of individuals was controlled through a dispersal kernel(% of individuals within of X% (5 vs 50 atm) of individuals within a population may migrate at each generation. 220 generations. Dispersal kernel equation.

The simulation landscape was modelled as a homogeneous rectangular grid of 20 by 80 cells with absorbing boundary conditions. Mirroring the geographic extent of mountain pine beetle range expansion from northwestern British Columbia to northern Alberta within 10 years of the current outbreak, the simulated landscape can be interpreted as an area of 160 by 640 km with cells of 8 by 8 km (6 400 ha). The carrying capacity of each cell was set to 30 individuals, which translates into a maximum possible population of 48 000 individuals on the simulated landscape. These parameter values were selected as a compromise between realism and time-related computational limitations. While the size of the simulated population is much smaller than that of the actual mountain pine beetle outbreak, it was large enough to generate the complex eco-evolutionary dynamics (i.e., allele surfing and neutral clines) that are the subject of this study.

Three main processes determined the spatio-temporal population and genomic dynamics of this simulated population: 1) dispersal; 2) selection; and 3) reproduction. Below, we summarize each process and describe the associated parameters. Additional details regarding parameters and CDMetaPOP architecture are included in the Supplementary Materials.

*Experimental Design*

We examined model sensitivity to three levels of dispersal (weak, intermediate, and strong), three strengths of selection (weak, intermediate, and strong) in a crossed design for a total of 9 experimental scenarios (Table 1). We also examined the effect of the timing of sampling (*i.e.*, time since the beginning of the expansion) within each scenario. Simulations were run for a time horizon of 150 non-overlapping, univoltine generations (*i.e.*, 1 generation per year). The choice of 150 generations was motivated by a desire to allow the simulation grid to be filled or near filled in all scenarios while avoiding unnecessary simulation of additional time steps in scenarios that filled the grid quickly. Simulations were sampled at six different points in time: generations 25, 50, 75, 100, 125 and 150. Within each selected generation, we sampled all individuals from each of 200 regularly spaced sample locations. These sample locations were consistent for all scenarios and generations. Each scenario was replicated 50 times to capture a range of stochastic variation in the simulated spatial processes (Table 1).

*Simulation experiments*

a) Simulation parameters

We modeled the effects of disturbance on the genetic diversity using the spatially-explicit gene flow simulation software CDMetaPOP (Landguth et al. 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. For each scenario we simulated 1000 replicates, with 25 interconnected populations, 100 bi-allelic loci and maximum carrying capacity of 100 individuals per population. X% (5 vs 50 atm) of individuals within a population may migrate at each generation. 220 generations. Dispersal kernel equation.

b) Simulation scenarios

The first scenario involves modelling a massive extraneous migration from a previously isolated 26th population. This population was simulated during the same number of generations and the cost distance between the 13th (central) and the 26th (isolated) populations is set to 0 between the 200th and 201st generations, mimicking a mass migration event between the two. The 26th population is then isolated again by resetting the cost distance to an unreachable number.

The second scenario involves modelling a demographic bottleneck through massive mortality. To do that, the carrying capacity of the 13th population (central), was set to 10% of its original value between the 200th and 201st generations.

Massive extraneous migration/Bottleneck

*Genetic dissimilarity*

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

*Estimating type I and type II errors*

a) Permutation approach

b) Two-step criterion

RESULTS

DISCUSSION

A few systems are consistently monitored through time (e.g. REFs LTER...) and/or exhaustively sampled in space (e.g. REFs Fushan...).

REFERENCES