doi:10.1111/evo.12237

INTEGRATING LANDSCAPE GENOMICS AND SPATIALLY EXPLICIT APPROACHES TO DETECT LOCI UNDER SELECTION IN CLINAL POPULATIONS

Matthew R. Jones,[1,2,*] Brenna R. Forester,[3,*] Ashley I. Teufel,[4] Rachael V. Adams,[5] Daniel N. Anstett,[6,7] Betsy A. Goodrich,[8] Erin L. Landguth,[9] Stéphane Joost,[10] and Stéphanie Manel[11]

[1]Department of Zoology and Physiology, Berry Biodiversity Conservation Center, University of Wyoming, 1000 E. University Avenue, Dept. 4304, Laramie, WY 82071, USA

[2]E-mail: mjones58@uwyo.edu

[3]University Program in Ecology, Nicholas School of the Environment, Duke University, Durham, NC 27705, USA

[4]Department of Molecular Biology, University of Wyoming, 1000 E. University Avenue, Dept. 3944, Laramie, WY 82071, USA

[5]Department of Biological Sciences, University of Lethbridge, 4401 University Drive, Lethbridge, AB T1K 3M4, Canada

[6]Department of Ecology and Evolutionary Biology, University of Toronto, 25 Willcocks Street, Toronto, ON M5S 3B2, Canada

[7]University of Toronto-Mississauga, Department of Biology, 3359 Mississauga Road N., Mississauga, ON L5L 1C6, Canada

[8]Northern Arizona University, School of Forestry, PO Box 15018, Flagstaff, AZ 86011, USA

[9]University of Montana, Division of Biological Sciences, 32 Campus Drive, Missoula, MT 59846, USA

[10]Laboratory of Geographic Information Systems, School of Architecture, Civil and Environmental Engineering, Ecole Polytechnique Fédérale de Lausanne, Bâtiment GC, Station 18 1015, Lausanne, Switzerland

[11]Laboratoire Population Environnement Développement, Aix-Marseille University Marseille, UMR AMAP, TA A51/PS2 34398 Montpellier Cedex 5, France

Received September 7, 2012
Accepted August 8, 2013
Data Archived: Dryad doi:10.5061/dryad.f3rk4

Uncovering the genetic basis of adaptation hinges on the ability to detect loci under selection. However, population genomics outlier approaches to detect selected loci may be inappropriate for clinal populations or those with unclear population structure because they require that individuals be clustered into populations. An alternate approach, landscape genomics, uses individual-based approaches to detect loci under selection and reveal potential environmental drivers of selection. We tested four landscape genomics methods on a simulated clinal population to determine their effectiveness at identifying a locus under varying selection strengths along an environmental gradient. We found all methods produced very low type I error rates across all selection strengths, but elevated type II error rates under "weak" selection. We then applied these methods to an AFLP genome scan of an alpine plant, *Campanula barbata*, and identified five highly supported candidate loci associated with precipitation variables. These loci

*These authors contributed equally to this work.

EVOLUTION
INTERNATIONAL JOURNAL OF ORGANIC EVOLUTION

also showed spatial autocorrelation and cline patterns indicative of selection along a precipitation gradient. Our results suggest that landscape genomics in combination with other spatial analyses provides a powerful approach for identifying loci potentially under selection and explaining spatially complex interactions between species and their environment.

KEY WORDS: ***Campanula barbata*, computer simulation, landscape genomics, natural selection, spatial statistics.**

The ability to detect loci potentially under selection in natural populations has important implications for understanding the genetic underpinnings of adaptation and reproductive isolation (Storz 2005). Most early population genetics studies of adaptive evolution and speciation were restricted to model organisms and certain types of traits characterized by mutations of large effect (Lewontin 1974). Yet, with advances in DNA sequencing technologies and the increasing accessibility of genomic data sets for nonmodel organisms, uncovering the genetic basis of many ecologically important traits in wild populations is now tenable (Stapley et al. 2010). Common approaches to identify loci under selection, such as surveying candidate genes or quantitative trait locus mapping, are still restricted in their application to the genes they can identify and the organisms with which they can be used (Storz 2005). Genome scans, however, offer an approach to detect regions of the genome under selection without a priori knowledge of their importance (Stinchcombe and Hoekstra 2008). Genome scans consist of DNA polymorphism data distributed across individual genomes. These data can be generated relatively cheaply for nonmodel organisms across large number of individuals (Ekblom and Galindo 2011). However, rigorous analysis of these data to identify loci potentially under selection remains a major obstacle.

One approach is to search for "outlier" loci that deviate from an overall genetic pattern (Luikart et al. 2003). These approaches (henceforth "population genomics" approaches) are based on the theory that genome-wide processes, such as demographic events, genetic drift, and gene flow, can be differentiated from locus-specific processes, such as selection and recombination (Lewontin and Krakauer 1973). For example, a locus under divergent selection (or linked to a locus under divergent selection) across two populations will have a higher pairwise-F_{ST} relative to the genome-wide pairwise-F_{ST} (Storz 2005). Similarly, a recent selective sweep will generally decrease nucleotide diversity adjacent to the loci under selection (Storz 2005) and result in extended regions of high linkage disequilibrium (LD) around the selected locus (Sabeti et al. 2002).

The efficacy of population genomics methods may be affected by a number of factors, including gene flow and population structure (Nielsen et al. 2007). For instance, hierarchical population structure and inclusion of isolated, bottlenecked populations can inflate the number of outlier loci detected (type I errors; Foll and Gaggiotti 2008; Excoffier et al. 2009). A number of population genomics methods have been developed to incorporate the effects of complex demographic history and population structure (e.g., Nielsen et al. 2007; Foll and Gaggiotti 2008; Bazin et al. 2010). For example, simulations reveal that outlier-detection methods based on haplotype LD (Sabeti et al. 2002; Pavlidis et al. 2010) are relatively robust against a number of demographic scenarios (Jensen et al. 2007; Nielsen et al. 2007; Pavlidis et al. 2010). However, many of these approaches require grouping individuals into populations, which may be inappropriate with clinal populations or when population structure is unknown before sampling, as is common with nonmodel organisms (Joost et al. 2013). For example, sampling across unknown populations may increase false positives rates (Jensen et al. 2005) and population clusters in genetic data may actually reflect uneven sampling design across a genetic cline (Serre and Pääbo 2004). Population genomics outlier-detection approaches, therefore, require additional analytical methods that allow for the analysis of individual variation to enhance their utility in identifying potential loci under selection (Schoville et al. 2012).

Landscape genomics attempts to uncover the processes and environmental variables important in natural selection by using correlative methods to link genetic variants to environmental variation (Luikart et al. 2003; Joost et al. 2007; Manel et al. 2010a). Rather than comparing locus-specific patterns to genome-wide patterns, these methods examine associations between allele distributions and predictor variables that are presumed to be important drivers of selection.

However, like population genomics approaches, correlative approaches may also suffer from high type I errors under certain demographic scenarios. For example, spatial bottlenecks may result in false positives if environmental conditions vary between the ancestral and bottlenecked population (Holderegger et al. 2008). In addition, not correcting for population structure or isolation-by-distance (IBD) may lead to false positives (Meirmans 2012). Thus, analyzing loci detected by landscape genomics methods with additional spatially explicit methods may help tease apart selection from demographically derived patterns of genetic variation. Although there are recent landscape genomics approaches that attempt to control for demographic signals (e.g., Bayenv, Coop et al. 2010; Bayenv2, Günther and Coop 2012; latent factor mixed models (LFMMs), Frichot et al. 2013), these important additions to the analytical toolkit require that data be grouped at the population level or require a predefined number of latent factors

(similar to determining a value of populations) to be computationally feasible. Some individual-based landscape genomics approaches have been tested in controlled, simulated environments (De Mita et al. 2013; Frichot et al. 2013); however, additional tests are needed to examine how these methods perform on clinal populations across different selection strengths.

In this study, we used simulations to examine the ability of individual-based landscape genomics approaches to detect a locus under selection across a range of selection coefficients in a clinal population. We tested a commonly used regression approach, generalized linear models (GLMs), in addition to two regression analyses that incorporate spatial autocorrelation in the data, general linear mixed models (GLMMs) and general additive mixed models (GAMMs). We also implemented classification and regression trees (CARTs), a nonparametric procedure of recursive partitioning that distinguishes differences among groups based on a set of predictor variables. We then applied landscape and population genomics approaches to detect loci under selection from an AFLP genome scan of a perennial alpine and subalpine plant, the Bearded Bellflower (*Campanula barbata*). *Campanula barbata*, affiliated with *Nardus stricta*-dominated grasslands, is a late-successional species and nondominant across its range in the European Alps (Meirmans et al. 2011; Scheepens and Stöcklin 2011). The flowers of *C. barbata* are animal pollinated and seeds are gravity dispersed, characteristics associated with strong population structure and small dispersal distances (Meirmans et al. 2011). We supplemented correlative and population genomics methods with spatially explicit analyses, that is cline and spatial autocorrelation analyses, of candidate loci to search for patterns consistent with selection-driven versus neutral processes. Finally, we highlighted the strengths and weaknesses of landscape genomics approaches and discuss the implications of this research for future evolutionary genomics research.

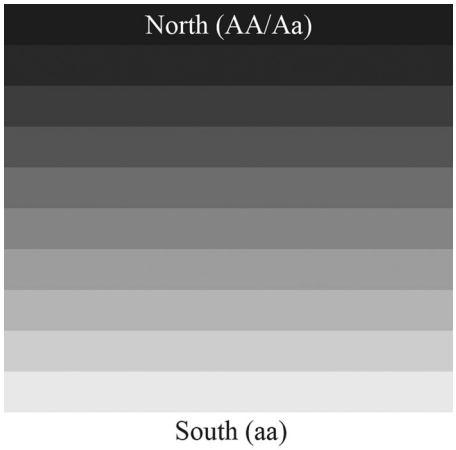

Figure 1. Simulation selection gradient surface representing selection for AA/Aa genotypes in the north and aa genotypes in the south. Color of bars represents strength of selection, increasing from north to south for aa genotypes and increasing from south to north for AA/Aa genotypes.

Methods

SIMULATING SPATIALLY EXPLICIT SELECTION

To assess the effectiveness of landscape genomics approaches to detect loci under selection, we developed spatially explicit simulation scenarios using the program CDPOP version 1.2 (Landguth and Cushman 2010; Landguth et al. 2012). CDPOP produces theoretical changes in allele frequencies for single and double bi-allelic loci under selection (for more details, see Landguth et al. 2012) and yields genetic patterns consistent with Wright–Fisher expectations when parameterized to match Wright–Fisher assumptions (Landguth and Cushman 2010).

For a population ($n = 5000$ individuals) of sexually reproducing individuals distributed randomly across an (x, y) surface, we modeled spatial changes in allele frequencies for 100 bi-allelic, dominant loci, consisting of one locus under directional selection and 99 neutral loci. Initial genotypes were assigned randomly to individuals. Dispersal, mortality, reproduction, and mutation mediated the spatial changes in allele frequencies across the continuous resistance landscape. Spatial changes in allele frequencies at the selection-driven locus were also determined by a selection gradient surface, which governed the viability of an individual along the gradient as a function of its genotype at the locus under selection (Fig. 1). We created selection surfaces for three different selection scenarios: "weak" ($s = 0.01$), "moderate" ($s = 0.1$), and "strong" selection ($s = 0.5$). Our selection gradients ran along the simulated landscape from north to south (Fig. 1), with dominant genotypes (AA, Aa) experiencing 0% mortality in the north end and either 1%, 10%, or 50% mortality in the south end of the landscape for "weak," "moderate," and "strong" selection scenarios, respectively. The recessive genotype (aa) was given the opposite selection gradient surface, with 1%, 10%, or 50% mortality in the north and 0% mortality in the south (Fig. 1). For all selection scenarios, we ran 10 Monte Carlo replicates of 1000 nonoverlapping generations.

Mating and dispersal movements were unbiased for males and females and followed an isolation-by-distance inverse-square function where maximum movement distance was 25% of the maximum Euclidean distance on the landscape. Reproduction began at birth and the number of offspring produced followed a Poisson process ($\lambda = 4$). Thus, a high rate of reproduction maintained a constant population size of 5000 individuals producing an excess number of offspring that were discarded once all 5000 locations were filled through the dispersal process (i.e., forcing individuals out of the simulation study once all available

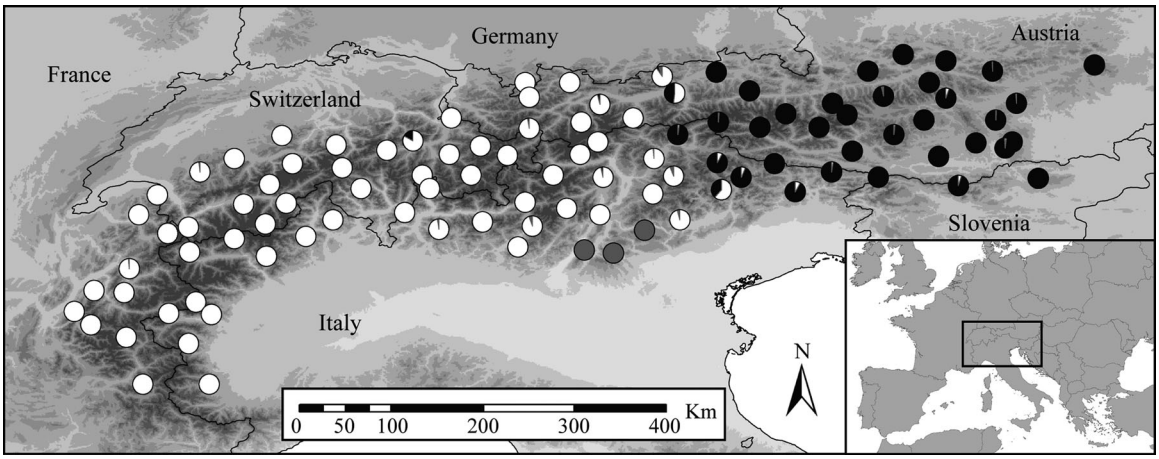

Figure 2. Sampling locations and genetic clusters of *Campanula barbata* defined by Structure. The proportions of colors in each circle reflect the probability (combining all individuals at that site) of membership to the western (white) or eastern (black) population for each sampling site. The three gray locations in the south-central Alps were removed from analyses.

home ranges are occupied; Balloux 2001; Landguth and Cushman 2010). We used a random mutation model with a mutation rate of 10^{-4}, a typical rate previously used for AFLP markers (Wilding et al. 2001; Campbell and Bernatchez 2004). We sampled 300 individuals from each Monte Carlo replicate using a stratified random approach in which the landscape surface was divided into 100 square sites with three individuals randomly sampled from each site.

EMPIRICAL POPULATION SAMPLING

For the empirical analyses, we used population samples of 307 *C. barbata* individuals collected as part of the IntraBioDiv project to assess pan-alpine vascular plant biodiversity across the European Alps and Carpathians (Gugerli et al. 2008). Leaf tissue was sampled and stored in silica gel from three (99 sites) or two (5 sites) individuals at standardized sampling locations throughout the European Alps from June to September 2004 (Fig. 2; Gugerli et al. 2008). At most sites, voucher specimens were collected and deposited in University of Neuchâtel herbarium (Gugerli et al. 2008). All samples were located between 699 and 2806 m elevation and distributed over 171,350 km^2. All individuals were genotyped at 114 AFLP markers (Vos et al. 1995) using three primer–enzyme combinations as described in Gugerli et al. (2008). Sampling locations and AFLP data are available at dryad doi:10.5061/dryad.f3rk4.

GEOGRAPHIC STRUCTURE AND GENETIC VARIATION ANALYSES

We used the program Structure version 2.3.3 (Pritchard et al. 2000) to infer the number of populations sampled in *C. barbata* based on AFLP genotypes. Structure determines the probability that an individual belongs to a genetic cluster (K) by minimizing LD and Hardy–Weinberg disequilibrium within each cluster (Pritchard et al. 2000; Falush et al. 2003). We conducted runs for 10 values of K (ranging from 1 to 10) using the admixture model with correlated allele frequencies. We used 50,000 Markov Chain Monte Carlo (MCMC) repetitions with a burn-in period of 50,000 iterations. We determined the most biologically probable value of K using the log probability of the data (Pritchard et al. 2000) and ΔK statistic (Evanno et al. 2005).

We used AFLP-SURV version 1.0 (Vekemans 2002) to generate genetic summary statistics based on populations defined by Structure. AFLP-SURV estimates genetic diversity using a Lynch and Milligan (1994) approach, which generates unbiased summary statistics for dominant markers. We estimated allele frequencies using a Bayesian method (Zhivotovsky 1999) with nonuniform prior distribution and assuming the Hardy–Weinberg equilibrium. After analyses to detect loci under selection we calculated F_{ST} for loci classified as neutral and for candidate loci.

LANDSCAPE GENOMICS APPROACHES

We analyzed simulated and empirical data sets using GLMs, GLMMs, GAMMs, and CARTs in R version 2.15.1 (R Development Core Team 2012). GLMMs are an extension of GLMs that allow for the inclusion of random effects (Bolker et al. 2009). In the context of our analyses, GLMMs allow for the analysis of nonindependent data, such as individual data points nested within sampling locations, while accommodating the binary response variable. Similarly, GAMMs allow for a comparable extension of generalized additive models (GAMs), which are themselves a nonparametric extension of GLMs (Guisan et al. 2002). In a GAM, the constant regression coefficients of the GLM are replaced with smoothing functions (usually splines) of the predictors, which are fit to local subsets of the data. Like GLMMs, GAMMs extend

Table 1. Environmental variables used to identify loci under selection in *Campanula barbata*.

Yearly climatic layers (1980–1989)	*tmaxavgty*	Mean annual maximum temperature (°C)
	tminavgty	Mean annual minimum temperature (°C)
	prcpavgy	Mean annual precipitation sum (cm)
Seasonal climatic layers (1980–1989)	*prcp0608*	Summer seasonal precipitation (number of rain days from June to August)

GAMs to include random effects, and thus can accommodate both nested and correlated data structures.

Classification and regression trees and their extensions, boosted regression trees and random forests, are an increasingly popular method in ecological modeling (De'ath and Fabricius 2000; Prasad et al. 2006) and have been found to perform equally or better than logistic regression at classification tasks (Vayssières et al. 2000). Classification and regression trees have many advantages over parametric multivariate analyses, including a lack of assumptions about the distribution of the data, the ability to easily handle missing values, and relative insensitivity to correlated predictor variables (De'ath and Fabricius 2000). In addition, CARTs are easily interpreted and can address complex interactions among predictors, including compensatory relationships, context-dependent contingencies, and nonlinear relationships. Classification and regression trees offer an additional nonparametric approach to detect loci under selection while also providing the ability to investigate complex relationships in the response of a selected locus to a set of environmental predictors.

For the simulated data, we used longitude and latitude as the "environmental" predictor variables. As the selection gradients for the simulated populations ran from north to south (Fig. 1), we expected the selected locus to be strongly associated with latitude. For the empirical AFLP data, we selected predictor variables (Table 1) from environmental layers calculated by Zimmermann and Kienast (1999). Four variables related to temperature and precipitation were selected based on their biological relevance to alpine plants (Manel et al. 2012b) and variable screening based on principal components analyses.

We ran GLMs on sampling sites ($n = 104$) using the frequency of presence polymorphisms at each sampling location as the response variable. For GLMMs and GAMMs, performed in the R packages *lme4* (Bates et al. 2011) and *gamm4* (Wood 2011), respectively, we used presence or absence polymorphisms in each individual as a binomially distributed response variable, with sampling site included as the random effect. We used Laplace approximations for GLMMs (Stigler 1986); GAMM models were estimated using a maximum likelihood framework with penalized regression spline smoothers and model comparison based on Laplace approximate log likelihoods. For the three regression approaches, models containing all possible subsets of predictor variables were fitted to each locus. We selected the model with the lowest corrected Akaike Information Criterion (AICc; Hurvich and Tsai 1989) ranks using the multimodel inference R package, *MuMIn* (Barton 2012). We then examined loci with significant effects at 95%, 99%, and 99.5% confidence levels (CIs) after a Bonferonni correction.

We ran CARTs on sampling sites using the *rpart* package in R (Therneau and Atkinson 2012). Classification models were fitted to each locus using all predictors, and the "improve" value for the first split was retained for each model. "Improve" is a measure of the improvement in deviance (a log-likelihood measure based on expected group membership) given by the split (Therneau and Atkinson 2012); we expected loci strongly associated with a predictor variable to show a high improve value, or a large decrease in deviance in response to partitioning the response into two groups based on that variable. We calculated a *P*-value for each locus with 100,000 permutation tests and used significance thresholds corresponding to 95%, 99%, and 99.5% CIs after Bonferroni corrections.

POPULATION GENOMICS OUTLIER-DETECTION APPROACHES

We implemented two population genomics methods on the *C. barbata* data set, using populations defined in Structure ($K = 2$): DFDIST and BayeScan. DFDIST, a variation on the program FDIST (Beaumont and Nichols 1996), was implemented in the program Mcheza (Antao and Beaumont 2011). Mcheza applies a multitest correction based on false discovery rate (FDR) to avoid overestimating outlier loci (Caballero et al. 2008). We used a total of 100,000 iterations and 95%, 99%, and 99.5% CIs. Loci with a significant *P*-value at an FDR threshold of 10% were considered candidate loci; F_{ST} values higher than expected were considered under positive selection.

BayeScan version 2.1 (Foll and Gaggiotti 2008) estimates the posterior probability that a locus is under selection using a reversible-jump MCMC approach. After 20 pilot runs of 5000 iterations and an additional burn-in period of 50,000 iterations, we used 100,000 iterations (sample size = 5000, thinning interval = 10) to identify outlier loci after removing seven monomorphic loci. Selection was evaluated using *q*-values, which are the FDR analog of *P*-values. A *q*-value is calculated for each locus and represents the minimum FDR at which the locus may become significant. Loci with a *q*-value < 0.10 were considered outliers.

CLINE ANALYSES IN *C. BARBATA*

For candidate loci identified in the *C. barbata* genome scan, we tested for relationships between geographic transitions in allele frequencies and environmental variables using the cline-fitting R package *hzar* (Derryberry et al., in review). The package *hzar* fits trait or environmental data to cline models using a Metropolis–Hasting algorithm (Metropolis et al. 1953; Hastings 1970) and calculates cline shape parameters, such as cline center (c), the location along a transect where the frequency of a variable changes most rapidly, and cline width (w), the distance over which the rapid change in frequency occurs (Szymura and Barton 1986). Cline shape parameters are estimated using three equations, which describe the shape of the cline center and the exponential decay of the tails on either side of the cline center (Szymura and Barton 1986).

Summer seasonal precipitation, the variable predominately associated with candidate loci, showed an east–west pattern of variation across the Alps. Therefore, to fit clines to our data, we measured transitions in AFLP band frequencies across a linear transect in distance (km) east from the western-most site (site 1). We transformed environmental data into Bernoulli trials (scales values between 0 and 1) to make data appropriate for cline fitting. We fitted 15 different cline models to observed data that differently describe the exponential decay of the tails on either side of the central cline (none, left tail only, right tail only, mirror tails, or both tails estimated separately) and scaling of minimum ($P_{\min}$) and maximum ($P_{\max}$) values (fixed to 0 and 1, observed values, or estimated values). We used a burn-in period of 10,000 iterations followed by 100,000 iterations (thinning parameter = 100). We determined the optimal cline model by using the lowest model AICc score. If the two log-likelihood support limits for the cline center and cline width overlap between two clines, those clines are said to be coincident and concordant, respectively.

SPATIAL AUTOCORRELATION IN *C. BARBATA*

Spatial autocorrelation of genetic data may reflect limited dispersal capabilities or local adaptation (Durand et al. 2009). To examine patterns of spatial dependency of candidate loci, we measured global and local spatial autocorrelation, respectively, according to sampling sites ($n = 104$) using Moran's I and univariate local indicators of spatial association (LISA; Anselin 1995) in the program OpenGeoDa (Anselin and McCann 2009). Local indicators of spatial association indicators are statistics that measure spatial dependence and evaluate the existence of local clusters in the spatial arrangement of a given variable using the statistical index I. Moran's I values range from -1, indicating perfect dispersion of data, to 1, indicating perfect spatial autocorrelation, with 0 indicating randomly dispersed data. We calculated LISA using a 70 km weighting scheme as 68 km is the minimum distance for which there are no neighborless observations.

Table 2. Type I and type II error rates at a 99.5% confidence level for landscape genomics methods based on simulation data for varying selection strengths.

Error type	Selection strength	Error rates (%)			
		GLM	GLMM	GAMM	CART
Type I	Weak	0.2	0.1	0.1	0.0
	Moderate	0.1	0.1	0.1	0.0
	Strong	0.4	0.4	0.4	0.1
Type II	Weak	100	100	100	100
	Moderate	0	10	10	60
	Strong	0	0	10	0

Results

METHOD PERFORMANCE ON SIMULATED DATA

Under "weak" selection, all landscape genomics methods failed to identify the locus under selection in all runs. In the "moderate" selection scenario, GLMs always detected the selected locus at a 99.5% CI, whereas GLMMs and GAMMs detected the selected locus at a 99.5% CI for 90% of runs, although it was detected at a 99% CI for the other 10% of runs. Classification and regression trees were the least powerful method, failing to detect the locus under selection in 60% of runs at a 99.5% CI and 40% of runs at a 95% CI. Under "strong" selection, landscape genomics methods always detected the locus under selection at a 99.5% CI, with one exception in which GAMMs failed to detect the correct locus in one run (Table 2).

Type I errors were low across selection strengths, ranging from 0.0% to 0.4% at a 99.5% CI (Table 2). The same false-positive loci were often identified by the three linear and additive modeling approaches (GLM, GLMM, GAMM). In both "weak" and "moderate" selection simulations, a single neutral locus was incorrectly identified by linear and additive models. Under "strong" selection, across all runs linear and additive models falsely detected four loci, all but one associated with longitude. Classification and regression trees had the lowest type I error rate, producing only one false positive in the "strong" selection scenario, which was not identified by the GLMs, GLMMs, or GAMMs.

POPULATION STRUCTURE AND GENETIC DIVERSITY OF *C. BARBATA*

We identified two populations using Structure, occupying the western and eastern portions of the study area (Fig. 2). Pairwise F_{ST} between these populations and global F_{ST} both indicated moderate genetic differentiation (pairwise $F_{ST} = 0.144$; global $F_{ST} = 0.139$; Table 3). A small, peripheral population in the southern Alps identified by Structure, consisting of three contiguous sampling locations and nine individuals, was removed from

Table 3. Population genetic diversity for western and eastern populations of *Campanula barbata*. We generated summary statistics for total gene diversity (H_T), average within-population diversity (H_W), average among-population diversity (H_B), expected heterozygosity (H_E, Nei's gene diversity), the proportion of polymorphic loci (P_P, polymorphic if band is present in less than 95% of all individuals), and pairwise and global F_{ST}.

Population	n	H_T	H_W	H_B	H_E	P_P	F_{ST}
Western	198	–	–	–	0.115	0.29	0.144 (pairwise)
Eastern	100	–	–	–	0.183	0.533	0.144 (pairwise)
Average	298	0.174	0.149	0.025	0.138	0.372	0.139 (global)

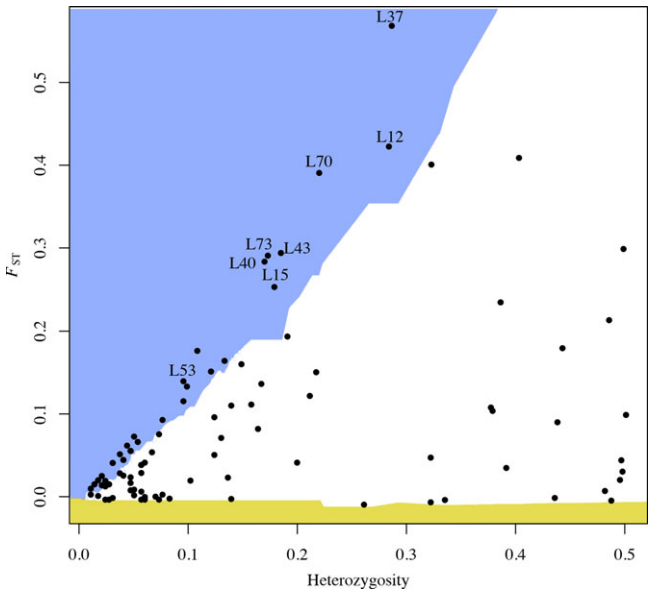

Figure 3. Distribution of F_{ST} values for each locus as a function of locus heterozygosity based on *Campanula barbata* AFLP data. Candidate loci identified by DFDIST are located in the blue (positive selection) and yellow (balancing selection) regions with neutral loci in the white region. Labeled loci are those identified by at least one landscape genomics method at a 99.5% confidence level.

population genetic analyses to avoid type I errors (Foll and Gaggiotti 2008; Excoffier et al. 2009). When included in analyses, this small population significantly increased global F_{ST} (F_{ST} = 0.348, including this population). "Highly supported" candidate loci ($n = 5$) had a high global F_{ST} ($F_{ST} = 0.321$) in comparison with neutral loci ($n = 102$), which showed a moderate global F_{ST} ($F_{ST} = 0.116$).

DETECTION OF LOCI UNDER SELECTION IN *C. BARBATA*

Generalized linear models detected the most loci associated with environmental variation (19 loci) at a 99.5% CI, followed by GLMMs (five loci), GAMMs (four loci), and CARTs (two loci; Table 4). The majority of these loci were most strongly associated with summer seasonal precipitation (*prcp0608*); only GLMs identified loci associated with other predictor variables at a 99.5% CI. These included L41 and L94, which were correlated with mean annual minimum temperature.

DFDIST detected 36 outlier loci at a 99.5% threshold, all under positive selection (Fig. 3). BayeScan detected no loci under selection at a q-value threshold of 0.10. Loci with the three lowest q-value scores (q-value = 0.89; mean $F_{ST} = 0.24$) were not detected by any landscape genomics methods or DFDIST. L37, which was identified by three landscape genomics methods at varying CIs and DFDIST (Table 4), had the fourth lowest q-value (q-value = 0.89) and the highest F_{ST} (0.26). Running BayeScan on three Structure populations yielded L37 with the lowest q-value (0.41) and the highest F_{ST} (0.36), however still not at a significant outlier threshold.

We considered "highly supported" candidate loci those that were identified in at least half of the detection approaches at a 99.5% CI. Based on these criteria, we identified five "highly supported" candidate loci: L26, L40, L45, L69, and L70 (Table 4).

CLINE AND SPATIAL AUTOCORRELATION ESTIMATES

We observed a sharp transition in summer seasonal precipitation along longitude occurring at 440.7 km east of site 1 (c; 439.0–447.4) with a cline width of 0.53 km (0.00–13.6 km; Table 5 and Fig. 4). Cline centers of highly supported candidate loci fell within 171 km of the precipitation cline with the western most cline (L45) at 418.5 km and the eastern most cline (L69) at 614.8 km east of site 1. One locus (L45) possessed a cline center coincidence with summer seasonal precipitation, although cline center confidence intervals of three other loci (L26, L40, L70) fell within 27.5 km of the summer precipitation cline center confidence interval. In most cases, the cline center for precipitation was shifted slightly west of cline centers for candidate loci (Fig. 4). Cline widths varied significantly from 14.4 km at the narrowest (L40) to 557.9 km at the widest (L69) with only L40 having a cline width concordant with summer precipitation.

We found significant positive global spatial autocorrelation (range: 0.369–0.697; pseudo P-value < 0.001) for all highly supported candidate loci for a 70 km weighting scheme (Fig. 5). The western and eastern Alps, respectively, are generally defined by positive spatial autocorrelation for low band frequency with low band frequency and high band frequency with high band frequency, excluding L45, which shows the opposite pattern. Interestingly, we found a consistent neutral corridor (no spatial dependence) located between the western and eastern populations.

Table 4. Candidate loci for *Campanula barbata* identified by two or more approaches. Loci are sorted by the number of asterisks (* = 95% CI, ** = 99% CI, *** = 99.5% CI) then the number of methods identifying the locus. Environmental variable(s) associated with each locus are listed as well as the type of selection for outlier methods. "Highly supported" loci are L40–L26.

Locus	GLM	GLMM	GAMM	CART	DFDIST	BayeScan
L40	*prcp0608***	*prcp0608***	*prcp0608***	*prcp0608**	Positive***	–
L45	*prcp0608***	*prcp0608***	*prcp0608***	*prcp0608****	–	–
L70	*prcp0608***	*prcp0608**	–	*prcp0608****	Positive***	–
L69	*prcp0608***	*prcp0608***	*prcp0608***	–	–	–
L26	*prcp0608***	*prcp0608***	*prcp0608***	–	–	–
L12	*prcp0608***	*prcp0608**	*prcp0608**	–	Positive***	–
L15	*prcp0608***	–	*prcp-0608**	*prcp0608**	Positive***	–
L37	*prcp0608***	*prcp0608**	–	*prcp0608**	Positive***	–
L88	*prcp0608***	*prcp0608***	*prcp0608***	–	–	–
L43	*prcp0608***	–	–	–	Positive***	–
L53	*prcp0608***	–	–	–	Positive***	–
L62	*prcp0608***	*prcp0608****	–	–	–	–
L73	*prcp0608***	–	–	–	Positive***	–
L55	*prcp0608***	*prcp0608**	*prcp0608**	–	–	–
L25	*prcp0608***	–	–	–	Positive***	–
L81	*prcp0608**	–	–	–	Positive***	–
L89	*prcp0608**	–	–	–	Positive***	–
L100	*prcp0608****	–	–	*prcp0608**	–	–

In L26 and L69, an additional neutral corridor is found in the western portion of the study area ~200 km east of site 1.

Discussion

METHOD PERFORMANCES

Landscape genomics methods require rigorous vetting to determine how they perform under different scenarios. In particular, given their ability for individual level analysis, assessing the effectiveness of correlative methods on clinal populations is important for understanding the contexts in which they can be implemented most effectively. To this end, we simulated a clinal population under varying selection along an environmental gradient to determine the effectiveness of four landscape genomics methods. Our results suggest that in clinal populations: (1) landscape genomics methods produce few type I errors across varying selection strengths; (2) under "moderate" selection, linear and additive regressions may outperform CARTs, which have higher type II errors; and (3) landscape genomics approaches may completely fail to detect loci under "weak" selection. The strikingly low type I errors were consistent at different levels of significance and for different methods, although CARTs had the lowest type I

Table 5. Parameter estimates for the genetic and environmental clines using HZAR for *Campanula barbata*. Two log-likelihood unit support limits for cline center (*c*) and cline width (*w*) are presented in parentheses.

Locus	*c*	*w*	P_{min}	P_{max}
prcp0608	440.7	0.71	0.13	0.70
	(439.0–447.4)	(0.00–15.4)	(0.08–0.20)	(0.64–0.76)
L40	458.3	14.4	0.0001	0.50
	(450.6–472.9)	(0.1–53.68)	(0.00–0.01)	(0.41–0.59)
L45	418.5	190.6	0	1
	(344.0–440.3)	(120.5–388.1)	(Fixed)	(Fixed)
L70	470.8	60.9	0.01	0.68
	(451.6–489.4)	(28.5–109.1)	(0.00–0.02)	(0.57–0.77)
L26	493.5	116.4	0	1
	(474.9–518.7)	(67.0–229.9)	(Fixed)	(Fixed)
L69	614.8	557.9	0.03	0.49
	(567.0–682.4)	(431.2–762.0)	(Fixed)	(Fixed)

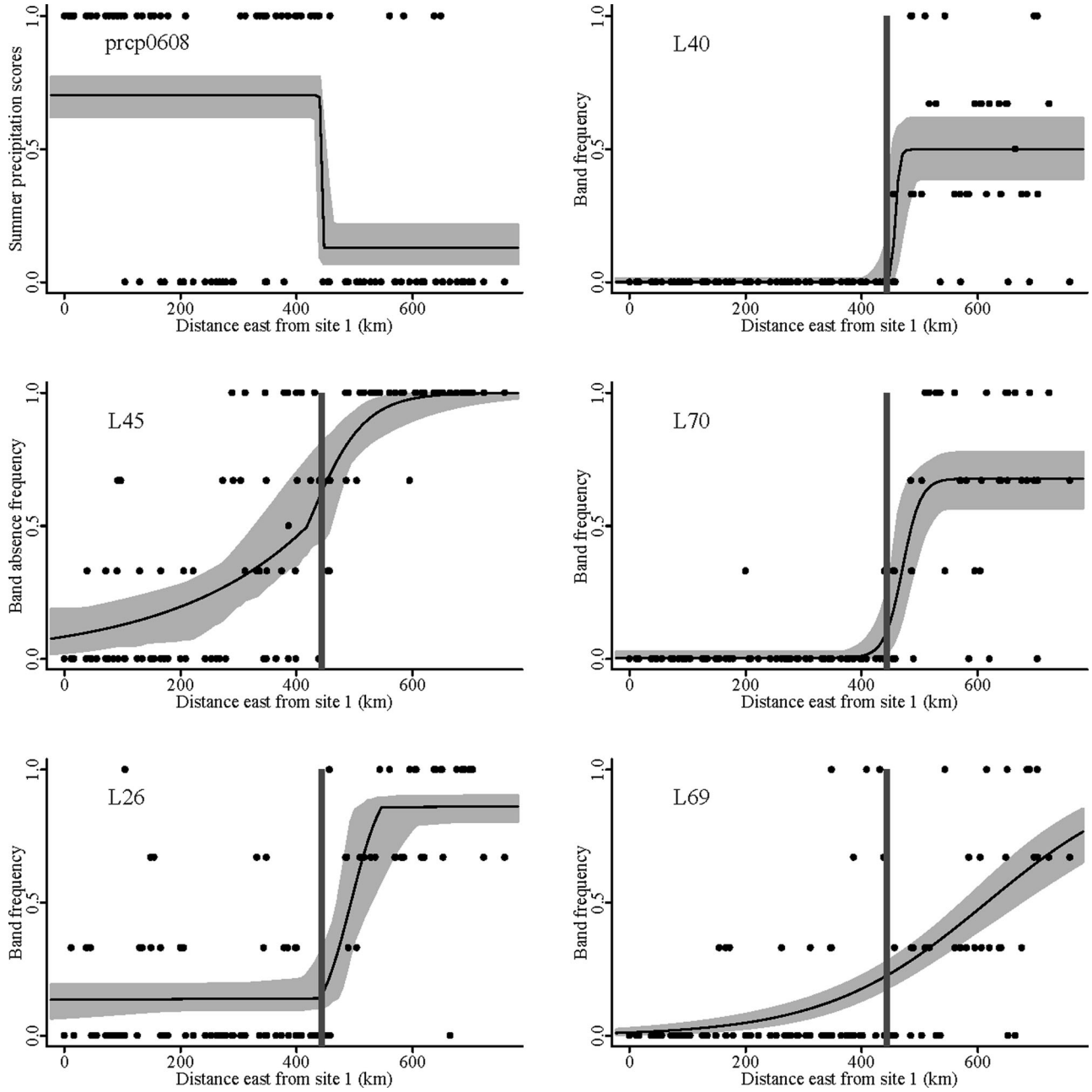

Figure 4. **Cline shapes for summer seasonal precipitation and highly supported candidate loci in *Campanula barbata*. The 95% credible cline region is shaded in light gray. The vertical gray rectangle corresponds to the two log-likelihood unit support limits for summer seasonal precipitation cline center. Cline shape parameters are presented in Table 5.**

error rates. Low type I errors were likely a result of minimal IBD patterns in the simulated populations due to high dispersal capability of the simulated individuals (maximum dispersal distance = 25% of landscape). We performed a Mantel test to examine the association between genetic and Euclidean distance and found a weak, yet significant, pattern of IBD that seemed to arise as a result of the selection-driven locus. For example, under weak selection, the strength of IBD was lower ($r = 0.038$, $P < 0.05$) than under strong selection ($r = 0.081$, $P < 0.05$). When present, strong IBD can confound landscape and population genomics outlier detection techniques, resulting in higher type I error rates (Meirmans 2012). Thus, landscape genomics methods may be particularly well suited to detect loci under selection in clinal populations where strong population structure is not present. We also found that type II error rates were low except for "weak" selection simulations where all methods failed to detect the locus under selection. Also, under "moderate" selection GLMs, GLMMs, and GAMMs were much more robust that CARTs. In a similar

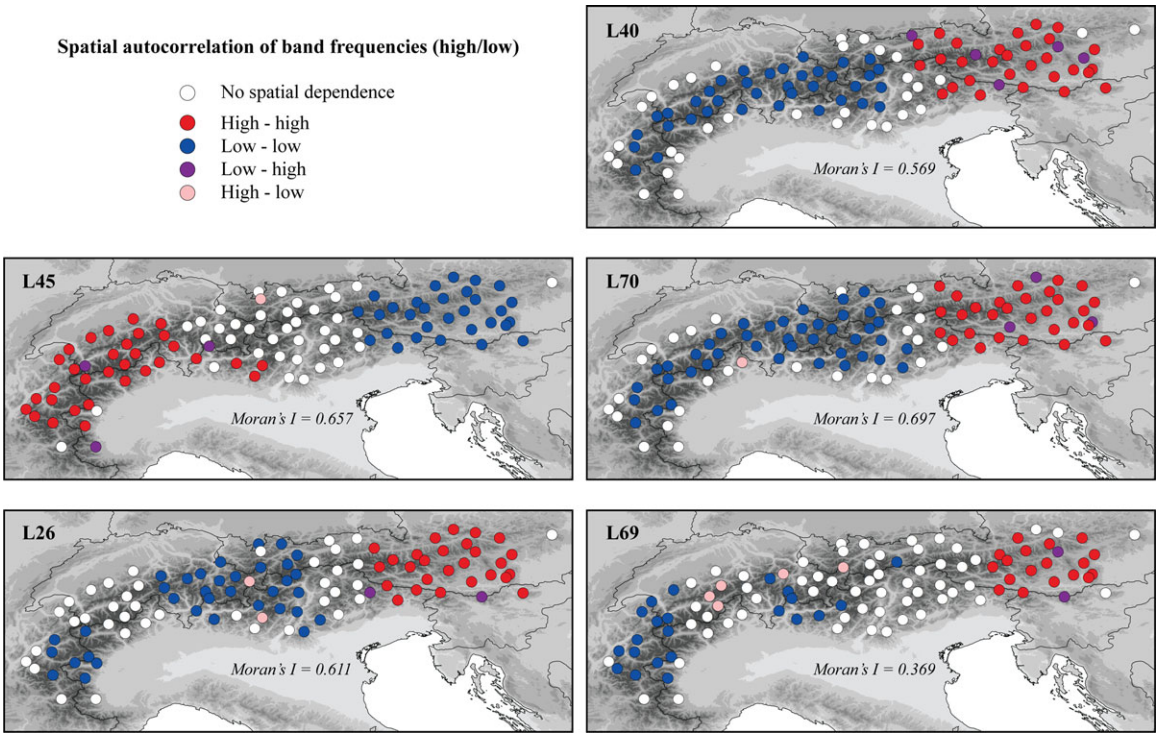

Figure 5. **Local indicators of spatial association (LISA) of AFLP band frequencies for highly supported candidate loci in *Campanula barbata*. Shown in red are locations where high band frequencies are correlated with a high mean of band frequencies measured at the neighboring sampling sites located within a radius of 70 km (spatial weighting scheme). Shown in blue are locations where low band frequencies are correlated with a low mean of band frequencies in neighboring sampling sites using the same weighting scheme. In purple are locations where low band frequencies are correlated with a high mean of band frequencies. Shown in pale red are locations where high band frequencies are correlated with a low mean of band frequencies. Finally, locations with band frequencies showing no spatial dependence are displayed in white.**

simulation study, De Mita et al. (2013) found that GLMs were most robust under highest migration rates and when sampling occurred at the individual level. Thus, in clinal populations with gradual genetic gradients, these methods may outperform CARTs. Additional simulation testing of correlative methods is needed to assess their sensitivity to different levels of IBD and population structure, as well as more complex spatial selection and genic (e.g., mutation strength, mutational models, or recombination) scenarios.

Empirical results further suggest that CARTs are the least powerful approach for identifying candidate loci under selection. A CART model identified the fewest loci under selection in *C. barbata* ($n = 2$; 99.5% CI), which is consistent with simulation results. Given their higher type II error rate under "moderate" selection strengths, loci not detected by CARTs, but found by multiple alternative approaches (i.e., landscape or population genomics methods) should be considered potentially under selection. In contrast to simulations, we found GLMs identified far more loci compared to other approaches ($n = 19$; 99.5% CI), as observed by De Mita et al. (2013). The relatively low dispersal capabilities of *C. barbata* likely increases IBD patterns for this species, which can boost false positive rates in GLMs (De Mita et al. 2013; Frichot et al. 2013).

We find striking differences with respect to the detection levels of the two population genomics methods we employed on our empirical data. DFDIST was the least conservative method, identifying 36 loci under selection. Several studies indicate DFDIST is prone to high false detection rates (Caballero et al. 2008; Pérez-Figueroa et al. 2010) and these authors advocate the use of conservative significance levels to decrease false positives. Even when using a very conservative significance level (99.5%), we found a large number of loci detected as under selection using this method. In contrast, BayeScan identified no loci under selection. Furthermore, most loci that had the lowest q-values using BayeScan did not overlap with any loci identified by other methods.

GENETIC PATTERNS IN *C. BARBATA*

Ecological gradients can drive adaptation and population differentiation by imposing differential selection pressures on populations at gradient extremes (Cheviron and Brumfield 2009; Freedman et al. 2010), resulting in steep genetic clines at loci under selection. We found genetic clines across the Alps in several candidate

loci are broadly coincident and concordant with clines in summer precipitation (Fig. 4 and Table 5). In addition, we found that F_{ST} in highly supported candidate loci is significantly elevated relative to "neutral" F_{ST}, indicative of diversifying selection. Although precipitation is generally not considered a limiting variable for most alpine plant species (Körner 2003), winter and spring precipitation play an important role in snowpack development, which both limit the growing season while providing protection from early season frost events (Inouye 2008), as well as summer soil moisture. Given the different cline shapes observed in candidate loci (e.g., broad cline widths in L45 and L69 and narrow cline widths in L40 and L70) selection on these loci may in fact be driven by different environmental variables that covary with summer precipitation. The narrow clines of L40 and L70 are more coincident with summer precipitation clines, although the loss of information due to the Bernoulli transformation of the precipitation data adds uncertainty to our cline shape estimates and assessments of coincidence and concordance. Furthermore, the similar spatial autocorrelation clusters of L26 and L69, showing an additional neutral corridor in the western portion of the study area, suggests that similar environmental variables drive selection at these loci.

The steep clines in genetic data may also form by purely neutral processes if differentiated populations have recently come into secondary contact (Haldane 1948; Endler 1977). For example, post-glacial recolonization of the Alps has frequently been implicated in the formation of secondary contact zones between divergent lineages (Taberlet et al. 1998). In *C. barbata*, the two distinct eastern and western lineages match phylogeographic break zones corresponding with other, codistributed silicicolous alpine plants (Thiel-Egenter et al. 2011). These phylogeographic break zones are presumed to have formed by expansion from distinct glacial refugia for silicicolous species in combination with topographically mediated dispersal barriers (Thiel-Egenter et al. 2011). The neutral corridors of no spatial autocorrelation at the zone of contact between eastern and western populations (Fig. 5) are consistent with increased genetic variance as a result of admixture at contact zones between diverged lineages (Scheepens and Stöcklin 2011). Assuming secondary contact occurred between western and eastern populations after glacial retreats at the end of the Pleistocene about 9000 generation ago, we would expect cline widths of approximately 600 m (assuming generation time of two years and mean dispersal distance of 2.4 m; Barton and Gale 1993; Engler et al. 2009; Meirmans et al. 2011). However, cline widths for candidate loci are much larger than expected due to secondary contact (Table 5). Furthermore, undetected loci show random patterns of variation across longitude; we would expect these 'neutral' loci to show similar neutral corridors if demographic history was driving the patterns. Thus, secondary contact of differentiated western and eastern populations after the retreat of the Pleistocene ice sheets does not appear to drive cline shapes of candidate loci.

UTILITY OF LANDSCAPE GENOMICS METHODS

Landscape genomics offers practical advantages to detect signatures of selection, but these advantages may be dependent on the population structure of the study system and the sampling scheme employed (De Mita et al. 2013). Thus, it is the responsibility of the researcher to choose appropriate methods based on a consideration of the biology of their study system.

Importantly, landscape genomics methods allow for inferences about the nature of selective forces operating on natural populations. Environmental variables associated with genetic variants may be important drivers of selection at detected loci. Thus, landscape genomics provides a priori hypotheses for follow-up functional genomics experiments, especially important for non-model organisms. However, the ability of landscape genomics methods to detect environmental drivers of selection may also be a drawback in some cases. To identify loci that are under environmentally driven selection, judicious choice of biologically appropriate and functionally relevant predictor variables is required; yet, this task may be difficult (e.g., Joost et al. 2010). For example, sampling design must be carefully planned to capture the appropriate scales of environmental variation for the species or population under consideration (Schoville et al. 2012). This includes sampling homogenously across the landscape with the aim of maximizing environmental variation (Manel et al. 2012a). Furthermore, when important variables for selection are not included in analyses, models are expected to have poor explanatory power (Manel et al. 2010a). However, Moran's eigenvector maps (Borcard et al. 2004) may be appropriate to deal with this problem because they can be included as explanatory variables in the regression analysis as proxies for unmeasured variables (Manel et al. 2010b). Landscape genomics approaches also rely on the assumption that selection has occurred over a long enough period of time to establish a detectable relationship between genes and the environment (Joost et al., in press). If selection is recent in a population then landscape genomics methods may have reduced power in detecting selection.

Until recently, AFLP genome scans, like the one used in this study, were one of the most commonly used genetic data sets for non-model organisms. However, because AFLP markers are dominant (decreasing their information content) and generally have low marker density, these data sets are being used less frequently (Manel et al. 2010a). In addition, high-throughput sequencing technologies are generating comprehensive genomic data for many natural populations at lower costs. These data offer incredible opportunities to uncover genetic variation in populations but place a premium on bioinformatics tools and analytical methods capable of handling them. Correlative approaches must

be used cautiously when analyzing high-throughput data, given that these methods are sensitive to false positives as a result of LD or spurious correlations with environmental variables (Manel et al. 2010a; De Mita et al. 2013). However, the fast processing capacities of regression and classification-based landscape genomics methods make them well suited to analysis of next-generation genomic data (Ekblom and Galindo 2011; De Mita et al. 2013). Current landscape genomics research is already heading towards using larger genomic and environmental data sets to uncover the genetic basis of adaptation (Eckert et al. 2010; Frichot et al. 2013; Vincent et al. 2013).

In addition, landscape genomics offers a flexible framework for investigating genetic variation. By identifying associations between genes and environmental variables, the genetic architecture of selection can be investigated at multiple scales (e.g., individual, population, metapopulation, subspecies) while incorporating the effect of spatial heterogeneity of the landscape on patterns of allelic variation (reviewed in Schoville et al. 2012). This approach is useful when sampled populations have weak or clinal population structure or unknown structure. Because population differentiation methods may not be appropriate for such data sets, individual-based landscape genomics analyses provide an excellent alternative. Our analyses of simulated clinal populations indicate that these methods perform well under such circumstances.

Again, the flexibility of landscape genomics approaches may also be a weakness if they are used without appropriate data screening. If population structure is present, correlated allele frequencies between populations can substantially increase type I errors in certain landscape genomics approaches (De Mita et al. 2013). Interestingly, in their simulation study De Mita et al. (2013) found that correlative approaches actually produced more false positives than population genomics approaches under patterns of IBD. Thus, we advocate supplementing landscape genomics analyses with additional spatial analyses to more thoroughly investigate candidate loci to tease out spurious correlations. Latent factor mixed models (Frichot et al. 2013) may be an appropriate landscape genomics method to investigate selection while accounting for population structure and correlated allele frequencies. Latent factor mixed models account for population structure through unobserved variables and thus can estimate effects of residual population structure or IBD; however, this method requires an a priori definition of the number of latent factors, which may be challenging in nonmodel species. Population-based landscape genomics methods such as Bayenv (Coop et al. 2010) or Bayenv2 (Günther and Coop 2012) that incorporate allele frequency correlation between populations may also be a suitable alternative to individual-based approaches to account for population structure. Ideally, correlated allele frequencies and population structure can be calculated from an independent genetic data set and translated into a matrix, which is then accounted for in further analyses with different genetic data sets.

Finally, we urge caution in interpreting results from landscape genomics studies. Results from these approaches cannot be interpreted in a population genetic context and are not to be taken as proof of selection. Exploratory landscape genomics or population genomics analyses should be followed with sequencing of candidate genes and functional assays to ascertain important selective processes and the functional significance, if any, of the candidate locus. In this sense, the use of landscape genomics methods on next generation sequencing data will allow for more rigorous inferences about functional relevance of detected loci using a comparative genomics approach (e.g., Eckert et al. 2010).

Conclusions

Genome scans are an appealing method for identifying candidate loci under selection because they can be easily used in nonmodel organisms without prior knowledge about the forces driving selection (Stinchcombe and Hoekstra 2008). However, different analytical methods may be preferable under different types of population structure. In particular, landscape genomics, which combines spatially explicit statistical methods and genetic data to elucidate complex evolutionary responses of species to their environment, may be most useful when individual-based analysis is warranted, for example in clinal populations. Landscape genomics methods provide the most information when used in an ensemble context (to rank candidate loci based on agreement or disagreement across methods), in concert with complementary methods that can independently verify potential loci under selection (Manel et al. 2009). We advocate the use of landscape genomics, population genomics (when applicable), and additional spatial analyses to maximize information content of candidate loci and disentangle demographic from selective signals. Given their fast processing capacities, landscape genomics is a promising approach to analyze large next-generation single nucleotide polymorphism (SNP) data sets. Landscape genomics methods will likely become increasingly useful given the burgeoning of genomic and environmental data sets that require statistical tools to detect loci under selection.

ACKNOWLEDGMENTS

The authors thank M. Carling, G. Derryberry, M.-J. Fortin, J. Maley, D. Urban, H. Wagner, and the 2012 Canadian Institute of Ecology and Evolution Graduate minicourse participants for assistance on various components throughout this project. SM was funded by the Institut Universitaire de France.

LITERATURE CITED

Anselin, L. 1995. Local indicators of spatial association – LISA. Geog. Anal. 27:93–115.

Anselin, L., and M. McCann. 2009. OpenGeoDa, open source software for the exploration and visualization of geospatial data. Pp. 550–551 *in* Proceedings of the 17th ACM SIGSPATIAL international conference on advances in geographic information systems, Seattle, Washington.

Antao, T., and M. A. Beaumont. 2011. Mcheza: a workbench to detect selection using dominant markers. Bioinformatics 27:1717–1718.

Balloux, F. 2001. EASYPOP (Version 1.7): a computer program for population genetic simulations. J. Hered. 92:301–302.

Barton, K. 2012. MuMIn: multi-model inference. R package version 1.7.11. Available at http://CRAN.R-project.org/package=MuMIn. Accessed April 1, 2012.

Barton, N. H., and K. S. Gale. 1993. Genetic analysis of hybrid zones. Pp. 13–45 *in* R. G. Harrison, ed. Hybrid zones and the evolutionary process. Oxford Univ. Press, Oxford, U.K.

Bates, D., M. Maechler, and B. Bolker. 2011. lme4: linear mixed-effects models using S4 classes. R package version 0.999375-42. Available at http://CRAN.R-project.org/package=lme4. Accessed May 1, 2012.

Bazin, E., K. J. Dawson, and M. A. Beaumont. 2010. Likelihood-free inference of population structure and local adaptation in a Bayesian hierarchical model. Genetics 185:587–602.

Beaumont, M. A., and R. A. Nichols. 1996. Evaluating loci for use in the genetic analysis of population structure. Proc. R. Soc. B 263:1619–1626.

Bolker, B. M., M. E. Brooks, C. J. Clark, S. W. Geange, J. R. Poulsen, M. H. H. Stevens, and J.-S. S. White. 2009. Generalized linear mixed models: a practical guide for ecology and evolution. Trends Ecol. Evol. 24:127–135.

Borcard, D., P. Legendre, C. Avois-Jacquet, and H. Tuomisto. 2004. Dissecting the spatial structure of ecological data at multiple scales. Ecology 85:1826–1832.

Caballero, A., H. Quesada, and E. Rolán-Alvarez. 2008. Impact of amplified fragment length polymorphism size homoplasy on the estimation of population genetic diversity and the detection of selective loci. Genetics 179:539–554.

Campbell, D., and L. Bernatchez. 2004. Generic scan using AFLP markers as a means to assess the role of directional selection in the divergence of sympatric whitefish ecotypes. Mol. Biol. Evol. 21:945–956.

Cheviron, Z. A., and R. T. Brumfield. 2009. Migration-selection balance and local adaptation of mitochondrial haplotypes in rufous-collared sparrows (Zonotrichia capensis) along an elevational gradient. Evolution 63:1593–1605.

Coop, G., D. Witonsky, A. D. Rienzo, and J. K. Pritchard. 2010. Using environmental correlations to identify loci underlying local adaptation. Genetics 185:1411–1423.

De'ath, G., and K. E. Fabricius. 2000. Classification and regression trees: a powerful yet simple technique for ecological data analysis. Ecology 81:3178–3192.

De Mita, S., A.-C. Thuillet, L. Gay, N. Ahmadi, S. Manel, J. Ronfort, and Y. Vigouroux. 2013. Detecting selection along environmental gradients: analysis of eight methods and their effectiveness for outbreeding and selfing populations. Mol. Ecol. 22:1383–1399.

Derryberry, G. E., E. P. Derryberry, J. M. Maley, and R. T. Brumfield. HZAR: Hybrid Zone Analysis using an R software package: *In review*.

Durand, E., F. Jay, O. E. Gaggiotti, and O. François. 2009. Spatial inference of admixture proportions and secondary contact zones. Mol. Biol. Evol. 26:1963–1973.

Eckert, A. J., A. D. Bower, S. C. González-Martínez, J. L. Wegrzyn, G. Coop, and D. B. Neale. 2010. Back to nature: ecological genomics of loblolly pine (*Pinus taeda*, Pinaceae). Mol. Ecol. 19:3789–3805.

Ekblom, R., and J. Galindo. 2011. Applications of next generation sequencing in molecular ecology of non-model organisms. Heredity 107:1–15.

Endler, J. A. 1977. Geographic variation, speciation, and clines. Princeton Univ. Press, Princeton, NJ.

Engler, R., C. F. Randin, P. Vittoz, T. Czáka, M. Beniston, N. E. Zimmermann, and A. Guisan. 2009. Predicting future distributions of mountain plants under climate change: does dispersal capacity matter? Ecography 32: 34–45.

Evanno, G., S. Regnaut, and J. Goudet. 2005. Detecting the number of clusters of individuals using the software structure: a simulation study. Mol. Ecol. 14:2611–2620.

Excoffier, L., M. Foll, and R. J. Petit. 2009. Genetic consequences of range expansions. Annu. Rev. Ecol. Evol. Syst. 40:481–501.

Falush, D., M. Stephens, and J. K. Pritchard. 2003. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. Genetics 164:1567–1587.

Foll, M., and O. Gaggiotti. 2008. A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. Genetics 180:977–993.

Freedman, A. H., H. A. Thomassen, W. Buermann, and T. B. Smith. 2010. Genomic signals of diversification along ecological gradients in a tropical lizard. Mol. Ecol. 19: 3773–3788.

Frichot, E., S. D. Schoville, G. Bouchard, and O. François. 2013. Testing for associations between loci and environmental gradients using latent factor mixed models. Mol. Biol. Evol. 30:1687–1699.

Gugerli, F., T. Englisch, H. Niklfeld, A. Tribsch, Z. Mirek, M. Ronikier, N. E. Zimmermann, R. Holderegger, and P. Taberlet. 2008. Relationships among levels of biodiversity and the relevance of intraspecific diversity in conservation—a project synopsis. Perspect. Plant Ecol. Evol. Syst. 10:259–281.

Guisan, A., T. C. Edwards Jr, and T. Hastie. 2002. Generalized linear and generalized additive models in studies of species distributions: setting the scene. Ecol. Model. 157:89–100.

Günther, T., and G. Coop. 2012. Robust identification of local adaptation from allele frequencies. Quant. Biol. In: eprint arXiv: 1209.3029.

Haldane, J. B. S. 1948. The theory of a cline. J. Genet. 48:277–284.

Hastings, W. K. 1970. Monte Carlo sampling methods using Markov chains and their applications. Biometrika 57:97–109.

Holderegger, R., D. Herrmann, B. Poncet, F. Gugerli, W. Thuiller, P. Taberlet, L. Gielly, D. Rioux, S. Brodbeck, S. Aubert, et al. 2008. Land ahead: using genome scane to identify markers of adaptive relevance. Plant Ecol. Divers. 1:273–283.

Hurvich, C. M., and C. Tsai. 1989. Regression and time series model selection in small samples. Biometrika 76:297–307.

Inouye, D. W. 2008. Effects of climate change on phenology, frost damage, and floral abundance of montane wildflowers. Ecology 89:353–362.

Jensen, J. D., Y. Kim, V. B. DuMont, C. F. Aquadro, and C. D. Bustamante. 2005. Distinguishing between selective sweeps and demography using DNA polymorphism data. Genetics 170:1401–1410.

Jensen, J. D., K. R. Thornton, C. D. Bustamante, and C. F. Aquadro. 2007. On the utility of linkage disequilibrium as a statistic for identifying targets of positive selection in nonequilibrium populations. Genetics 176: 2371–2379.

Joost, S., A. Bonin, M. W. Bruford, L. Després, C. Conord, G. Erhardt, and P. Taberlet. 2007. A spatial analysis method (SAM) to detect candidate loci for selection: towards a landscape genomics approach to adaptation. Mol. Ecol. 16:3955–3969.

Joost, S., L. Colli, P. V. Baret, J. F. Garcia, P. J. Boettcher, M. Tixier-Boichard, P. Ajmone-Marsan, and The GLOBALDIV Consortium. 2010. Integrating geo-referenced multiscale and multidisciplinary data for the management of biodiversity in livestock genetic resources. Anim. Genet. 41:47–63.

Joost, S., S. Vuilleumier, J. D. Jensen, S. Schoville, K. Leempoel, S. Stucki, C. Melo de Lima, J. Rolland, I. Widmer, and S. Manel, 2013. Uncovering the genetic basis of adaptive change: on the intersection of landscape genomics and theoretical population genetics. Mol. Ecol. 22:3659–3665.

Körner, C. 2003. Alpine plant life: functional plant ecology of high mountain ecosystems. 2nd ed. Springer, Berlin, Germany.

Landguth, E. L., and S. A. Cushman. 2010. cdpop: a spatially-explicit cost distance population genetics program. Mol. Ecol. Res. 10:156–161.

Landguth, E. L., S. A. Cushman, and N. A. Johnson. 2012. Simulating natural selection in landscape genetics. Mol. Ecol. Res. 12:363–368.

Lewontin, R. C. 1974. The genetic basis of evolutionary change. Columbia Univ. Press, New York.

Lewontin, R. C., and J. Krakauer. 1973. Distribution of gene frequency as a test of the theory of the selective neutrality of polymorphisms. Genetics 74:175–195.

Luikart, G., P. R. England, D. Tallmon, S. Jordan, and P. Taberlet. 2003. The power and promise of population genomics: from genotyping to genome typing. Nat. Rev. Genet. 4:981–994.

Lynch, M., and B. G. Milligan. 1994. Analysis of population genetic structure with RAPD markers. Mol. Ecol. 3:91–99.

Manel, S., C. Conord, and L. Després. 2009. Genome scan to assess the respective role of host-plant and environmental constraints on the adaptation of a widespread insect. BMC Evol. Biol. 9:288.

Manel, S., S. Joost, B. K. Epperson, R. Holderegger, A. Storfer, M. S. Rosenberg, K. T. Scribner, A. Bonin, and M. Fortin. 2010a. Perspectives on the use of landscape genetics to detect genetic adaptive variation in the field. Mol. Ecol. 19:3760–3772.

Manel, S., B. N. Poncet, P. Legendre, F. Gugerli, and R. Holderegger. 2010b. Common factors drive adaptive genetic variation at different spatial scales in Arabis alpina. Mol. Ecol. 19:3824–3835.

Manel, S., C. Albert, and N. G. Yoccoz. 2012a. Sampling in landscape genomics. Pp. 93–112 *in* A. Bonin and F. Pompanon, eds. Methods in molecular biology series. Humana Press, New York.

Manel, S., F. Gugerli, W. Thuiller, N. Alvarez, P. Legendre, R. Holderegger, L. Gielly, P. Taberlet, and IntraBioDiv Consortium. 2012b. Broad-scale adaptive genetic variation in alpine plants is driven by temperature and precipitation. Mol. Ecol. 21:3729–3738.

Meirmans, P. G. 2012. The trouble with isolation by distance. Mol. Ecol. 21:2839–2846.

Meirmans, P. G., J. Goudet, and O. E. Gaggiotti. 2011. Ecology and life history affect different aspects of the population structure of 27 high-alpine plants. Mol. Ecol. 20:3144–3155.

Metropolis, N., A. Rosenbluth, M. Rosenbluth, A. Teller, and E. Teller. 1953. Equation of state calculations by fast computing machines. J. Chem. Phys. 21:1087–1092.

Nielsen, R., I. Hellmann, M. Hubisz, C. Bustamante, and A. G. Clark. 2007. Recent and ongoing selection in the human genome. Nat. Rev. Genet. 8:857–868.

Pavlidis, P., J. D. Jensen, and W. Stephan. 2010. Searching for footprints of positive selection in whole-genome SNP data from nonequilibrium populations. Genetics 194:907–922.

Pérez-Figueroa, A., M. J. García-Pereira, M. Saura, E. Rolán-Alvarez, and A. Caballero. 2010. Comparing three different methods to detect selective loci using dominant markers. J. Evol. Biol. 23:2267–2276.

Prasad, A., L. Iverson, and A. Liaw. 2006. Newer classification and regression tree techniques: bagging and random forests for ecological prediction. Ecosystems 9:181–199.

Pritchard, J. K., M. Stephens, and P. Donnelly. 2000. Inference of population structure using multilocus genotype data. Genetics 155:945–959.

R Development Core Team. 2012. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Sabeti, P., D. E. Reich, J. M. Higgins, H. Z. P. Levine, D. J. Richter, S. F. Schaffner, S. B. Gabriel, J. V. Platko, N. J. Patterson, G. J. McDonald, et al. 2002. Detecting recent positive selection in the human genomce from haplotype structure. Nature 419:832–837.

Scheepens, J. F., and J. Stöcklin. 2011. Glacial history and local adaptation explain differentiation in phenotypic traits in the Alpine grassland herb Campanula barbata. Plant Ecol. Divers. 4:403–413.

Schoville, S. D., A. Bonin, O. Francois, S. Lobreaux, C. Melodelima, and S. Manel. 2012. Adaptive genetic variation on the landscape: methods and cases. Annu. Rev. Ecol. Evol. Syst. 43:23–43.

Serre, D., and S. Pääbo. 2004. Evidence for gradients of human genetic diversity within and among continents. Genome Res. 14:1679–1685.

Stapley, J., J. Reger, P. G. D. Feulner, C. Smadja, J. Galindo, R. Ekblom, C. Bennison, A. D. Ball, A. P. Beckerman, and J. Slate. 2010. Adaptation genomics: the next generation. Trends Ecol. Evol. 25:705–712.

Stigler, S. M. 1986. Laplace's 1774 memoir on inverse probability. Stat. Sci. 1:359–378.

Stinchcombe, J. R., and H. E. Hoekstra. 2008. Combining population genomics and quantitative genetics: finding the genes underlying ecologically important traits. Heredity 100:158–170.

Storz, J. F. 2005. Using genome scans of DNA polymorphism to infer adaptive population divergence. Mol. Ecol. 14:671–688.

Szymura, J. M., and N. H. Barton. 1986. Genetic analysis of a hybrid zone between the fire-bellied toads, *Bombina bombina* and *B. variegata*, near Cracow in southern Poland. Evolution 40:1141–1159.

Taberlet, P., L. Fumagalli, A. G. Wust-Saucy, and J. F. Cosson. 1998. Comparative phylogeography and postglacial colonization routes in Europe. Mol. Ecol. 7:453–464.

Therneau, T. M., and B. Atkinson. 2012. rpart: recursive partitioning. R package version 3.1–52. Available at http://mayoresearch.mayo.edu/mayo/research/biostat/splusfunctions.cfm. Accessed June 1, 2012.

Thiel-Egenter, C., N. Alvarez, R. Holderegger, A. Tribsch, T. Englisch, T. Wohlgemuth, L. Colli, M. Gaudeul, L. Gielly, N. Jogan, et al. 2011. Break zones in the distributions of alleles and species in alpine plants. J. Biogeogr. 38:772–782.

Vayssières, M. P., R. E. Plant, and B. H. Allen-Diaz. 2000. Classification trees: an alternative non-parametric approach for predicting species distributions. J. Veg. Sci. 11:679–694.

Vekemans, X. 2002. AFLP-SURV, version 1.0. Distributed by the author, Laboratoire de Génétique et Ecologie Végétale, Université Libre de Bruxelles, Belgium.

Vincent, B. M. Dionne, M. P. Kent, S. Lien, and L. Bernatchez. 2013. Landscape genomics in Atlantic Salmon (*Salmo salar*): searching for gene-environment interactions driving local adaptation. Evolution: *In press*.

Vos, P., R. Hogers, M. Bleeker, M. Reijans, T. van de Lee, M. Hornes, A. Frijters, J. Pot, J. Peleman, and M. Kuiper. 1995. AFLP: a new technique for DNA fingerprinting. Nucleic Acids Res. 23:4407–4414.

Wilding, C. S., R. K. Butlin, and J. Grahame. 2001. Differential gene exchange between parapatric morphs of *Littorina saxatilis* detected using AFLP markers. J. Evol. Biol. 14:611–619.

Wood, S. 2011. gamm4: generalized additive mixed models using mgcv and lme4. R package version 0.1–5. Available at http://CRAN.R-project.org/package=gamm4. Accessed May 15, 2012.

Zhivotovsky, L. A. 1999. Estimating population structure in diploids with multilocus dominant DNA markers. Mol. Ecol. 8:907–913.

Zimmermann, N. E., and F. Kienast. 1999. Predictive mapping of alpine grasslands in Switzerland: species versus community approach. J. Veg. Sci. 10:469–482.

Associate Editor: K. Petren