Posted on Authorea 28 Aug 2020 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.159863198.86187354 — This a preprint and has not been peer reviewed. Data may be preliminary.

Experimental support for genomic prediction of climate maladaptation using the machine learning approach Gradient Forests

Matthew Fitzpatrick[1], Vikram Chhatre[2], Raju Soolanayakanahally[3], and Stephen Keller[4,4]

[1]University of Maryland Center for Environmental Science
[2]University of Wyoming
[3]Agri-Environment Services Branch -Agriculture and Agri-Food Canada
[4]University of Vermont

August 28, 2020

Abstract

Gradient Forests is a machine learning algorithm that is gaining in popularity for studying the environmental drivers of genomic variation and for incorporating genomic information into climate change impact assessments. Here we provide the first experimental evaluation of the ability of 'genomic offsets' - a metric of climate maladaptation derived from Gradient Forests - to predict organismal responses to environmental change. We used high-throughput sequencing, genome scans, and several methods (including Gradient Forests) to identify candidate loci associated with climate adaptation in balsam poplar (Populus balsamifera L.). Individuals collected throughout balsam poplar's range also were planted in two common garden experiments. We used Gradient Forests to relate candidate loci to environmental gradients and to predict the expected magnitude of response (i.e., the genetic offset) of populations when transplanted from their "home" environment to the new environments in the common gardens. We then compared the predicted genetic offsets to measurements of population performance in the common gardens. We found the expected inverse relationship between genetic offset and performance in the common gardens: populations with larger predicted genetic offsets performed worse in the common gardens than populations with smaller offsets. Also, genetic offset better predicted performance in the common gardens than did 'naive' climate distances. Our results provide preliminary evidence that genomic offsets may provide a first order estimate of the degree of expected maladaptation of populations exposed to rapid environmental change.

KEYWORDS

climate adaptation, climate change, forests, species distributions, single nucleotide polymorphism, intraspecific variation

INTRODUCTION

Climate change is expected to become a major threat to biodiversity this century (Sala et al., 2000; Urban, 2015), with cascading impacts on human well-being and ecosystem function (Pecl et al., 2017). Anticipating and mitigating these impacts requires actionable predictions of expected biological responses, which are expected to become increasingly difficult to anticipate under novel climates of the future (Fitzpatrick, Blois, et al., 2018; Urban et al., 2016). The adaptive capacity of species represents an important component of climate change vulnerability (Dawson, Jackson, House, Prentice, & Mace, 2011), yet few studies incorporate local adaptation into forecasting models, while even fewer have attempted to compare genomic predictions to actual organismal responses.

Posted on Authorea 28 Aug 2020 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.159863198.86187354 — This a preprint and has not been peer reviewed. Data may be preliminary.

Recent technological advancements now provide access to massive quantities of data pertinent to biodiversity science and conservation (e.g., species occurrence databases, genome-scale DNA sequencing, high-resolution projections of future climate; Wüest et al., 2020). At the same time, new sophisticated machine learning methods have emerged that can take advantage of these data to identify conservation risks and opportunities under a changing climate. In particular, the application of machine learning to genomic studies of local adaptation represents an especially promising frontier for improving our understanding of biotic responses to climate change and the potential to consider climate vulnerability at the population level (Fitzpatrick & Keller, 2015; Gougherty, Keller, Chhatre, & Fitzpatrick, 2020; Savolainen, Lascoux, & Merilä, 2013).

Fitzpatrick & Keller (2015) described how a machine learning method known as Gradient Forests (GF; Ellis, Smith, & Pitcher, 2012) can be used to (1) analyze and map spatial variation in allele frequencies as a function of environmental gradients and (2) project patterns of genomic variation under future climate. GF derives monotonic, nonlinear functions that characterize compositional turnover in allele frequencies along each fitted environmental gradient. In addition to identifying the primary environmental drivers associated with genomic variation, these turnover functions provide unique insights into the nature of how genomic patterns vary along multiple environmental gradients, including where changes in allele frequencies are rapid or slow across space. The turnover functions from GF also can be used to transform (or rescale) the fitted environmental predictors from their arbitrary anthropogenic measurement units (e.g., of temperature or mm of precipitation) to common biological units of compositional turnover (Ellis et al., 2012). By transforming each of the predictor variables using its associated turnover function, the multidimensional environmental space can be converted into a multidimensional genomic space that characterizes differences in the expected genetic makeup between populations in different environments. By applying the turnover functions to scenarios of environmental change, one can project expected genomic patterns under future climate. The Euclidean distance between the locations of each population in the current and future genomic spaces characterizes the magnitude of expected change in genetic composition for each population given the pattern of climate change in each location. Fitzpatrick & Keller (2015) termed this distance the "genetic offset", which can be viewed as a metric of the degree of expected maladaptation when a population is exposed to rapid climate change, assuming no adaptive evolution *in situ* or migration to allow adaptive alleles to track climate change. Gougherty et al. (2020) recently extended the genetic offset concept to consider the contributions of climate maladaptation, migration, and the potential for future novel gene-climate associations to the vulnerability of climatically adapted populations.

Since the publication of Fitzpatrick & Keller (2015), a growing number of studies have used genetic offsets to estimate climate maladaptation in a variety of species, including trees (Gugger, Liang, Sork, Hodgskiss, & Wright, 2018; Ingvarsson & Bernhardsson, 2020; Jia et al., 2020; Martins et al., 2018), birds (Bay et al., 2018; Ruegg et al., 2018), and agricultural crops such as maize landraces in Mexico (Aguirre-Liguori, Ramirez-Barahona, Tiffin, & Eguiarte, 2019). However, like projections of species-level responses to climate change from species distribution models, genetic offsets are in essence derived from a correlative, space-for-time substitution approach (Blois, Williams, Fitzpatrick, Jackson, & Ferrier, 2013) that ignores the enormous complexities underlying actual evolutionary responses of populations to environmental change, including interactions between selection, effective population size, and evolutionary processes shaping adaptive variation (e.g. migration, mutation, recombination). Instead, the use of genetic offsets assumes that, after correcting for neutral population structure, correlations between allele frequencies and environmental gradients reflect current patterns of local selection and relative fitness and that these existing gene-environment associations*across space* can be used to project the magnitude of change in allele frequencies expected *through time* to maintain gene-environment associations at their current status quo. Very few studies have tried to relate local adaptation analyses and associated predictions to actual organismal responses. As such, genetic offsets lack empirical validation, and it remains unknown what if any utility the concept has for predicting the actual performance of populations in novel environments.

Here we use machine learning, population genomic data, and common garden experiments to provide an empirical space-for-time test of the extent to which genetic offsets predict performance of populations in new environments. We measured growth performance of trees collected from climatically diverse populations

Posted on Authorea 28 Aug 2020 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.159863198.86187354 — This a preprint and has not been peer reviewed. Data may be preliminary.

which were clonally propagated in two common gardens. For these same populations, we also obtained genome-wide single nucleotide polymorphisms (SNPs) which were used in a series of genome scans for local adaptation employing multiple methods to determine outlier loci associated with climate. We then fit GF to the different sets of candidate SNPs determined using the different outlier detection methods and used these models to (1) identify the primary environmental variables driving the signals of local climate adaptation in the genome, (2) fit flexible functions describing how genetic patterns vary along the gradients, and (3) predict genetic offsets associated with transplanting individuals from their home climatic environment to the climates they experienced at the common garden sites. Specifically, we aim to address the following questions:

1. How do GF models fit to different sets of statistical outlier SNPs differ in terms of variable importance, turnover functions, and predicted spatial patterns?
2. How well do genetic offsets predict responses of populations transplanted to new common garden environments and do genetic offsets outperform naive 'climate-only' transfer distances?
3. How sensitive is the predictive ability of genetic offsets to the composition of SNP panels derived from different outlier detection methods, or when randomly sampled from the genomic background?

MATERIALS AND METHODS

Study species: Balsam poplar (*Populus balsamifera* L., Salicaceae) is a common deciduous tree in northern temperate and boreal forest ecosystems across North America. Its expansive geographic range encompasses broad climatic gradients in growing season length, and populations exhibit abundant clinal variation and adaptive population divergence in allele frequencies and phenotypic traits, particularly those related to phenology and ecophysiology (Fitzpatrick & Keller, 2015; Keller, Levsen, Ingvarsson, Olson, & Tiffin, 2011; Keller, Levsen, Olson, & Tiffin, 2012; Keller, Soolanayakanahally, et al., 2011; Olson et al., 2013; Soolanayakanahally, Guy, Silim, Drewes, & Schroeder, 2009; Soolanayakanahally, Guy, Silim, & Song, 2013).

Based on previous work, balsam poplar is known to exhibit relatively strong regional population structure, with multiple studies recognizing a clearly divergent subpopulation in the eastern part of its range in Atlantic Canada and the New England States (Chhatre et al., 2019; Keller, Olson, Silim, Schroeder, & Tiffin, 2010; Meirmans, Godbout, Lamothe, Thompson, & Isabel, 2017). The rest of the range consists of a large, relatively homogenous genetic subpopulation in the middle core of its range that is weakly differentiated ($F_{ST} = 0.008$) but with signatures of isolation-by-distance along an axis of longitude (Chhatre et al., 2019; Meirmans et al., 2017). Lastly, some studies suggest a third major subpopulation in northwestern Canada and Alaska (Keller et al., 2010).

In the current study, we focused our sampling and analysis on the widespread but weakly structured "Core" region in the middle of the range (Chhatre et al., 2019). We did this to minimize the known confounding introduced by regional population structure to genome scans of selection (Lotterhos & Whitlock, 2014), and because the Core region still captures large variation in the climatic environment, making it ideal for studying adaptive contributions of genomic variation while reducing the likelihood of false positives due to demographic history (Chhatre et al., 2019). Our selection of individuals and populations was based on retaining individuals with ADMIXTURE ancestry scores >0.9 in the Core region based on Chhatre et al. (2019). This resulted in 336 individuals from 42 populations for analysis of genomic variation at 107,309 biallelic SNPs, with populations distributed across north central US and Canada, from Vermont, New York, and Quebec in the east to Manitoba and Saskatchewan in the west, and south as far as WY (Table S1).

Common gardens: We measured performance of poplar clones in two common garden locations with contrasting climates, "IH" near Indian Head, Saskatchewan, Canada (50.52 degN, -103.68 degW), and "VT" near Burlington, Vermont, USA (44.44 degN, -73.19 degW). Establishment of both common gardens was from clonally propagated stem cuttings (6-9 cm with at least 2 vegetative buds) taken from natural populations while dormant during winter, rooted in potting media in the greenhouse and grown for 1 year, and then outplanted directly into the ground at the respective garden site. Details of the IH site are provided in Soolanayakanahally et al. (2013), which describes the original garden established in 2005 consisting of 15

Posted on Authorea 28 Aug 2020 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.159863198.86187354 — This a preprint and has not been peer reviewed. Data may be preliminary.

genotypes from each population planted in aggregate with 2x2 meter spacing. Each population group was then clonally replicated 5 times in random locations throughout the garden. In 2011, a second planting at IH was established adjacent to the original garden, consisting of additional populations planted with the same specifications as the original planting. At the VT site, planting consisted of a completely randomized design with a target replication of 3 ramets per genotype planted with 2x2 meter spacing. At both sites, plants were not fertilized, but were watered as needed during the year they were established and then received no supplemental water thereafter. Additional details of the VT site can be found in Fetter, Nelson, & Keller (2019).

After losses due to mortality during establishment, plants available for phenotyping totaled 357 unique genotypes (N=117 in IH; 337 in VT) representing 41 of the 42 target populations (N=9 in IH; 41 in VT; population 'NIC' was absent from both; Table S1). Most of these genotypes with phenotype data were the same individuals sequenced with GBS for the 107,309 SNPs (297 of 336 individuals with GBS; representing 41 of the 42 populations with GBS). As an overall metric of growth performance, we measured the yearly height increment (cm) gain between the apical bud and the previous year's bud scar on the most dominant stem. Height increment was measured at the end of the 2015 growing season for both sites, after all plants had finished setting bud. This gives an overall measure of growth achieved during the 2015 growing season that integrates effects of genotypic variability in phenology and relative growth rate (Soolanayakanahally et al., 2013). Genotypes with height increment data were represented by a mean (SD) of 2.6 (1.5) ramets.

Environmental data for field locations : To characterize environmental conditions at the sampling locations of the source populations, we used a set of seven variables at 30-arcsecond resolution (˜1 km x 1 km). We selected this set of seven variables from a larger (n =21) initial set composed of 19 bioclimatic variables and elevation from WorldClim v1.4 (representing averages for the period 1960-1990; www.worldclim.org; Hijmans, Cameron, Parra, Jones, & Jarvis, 2005), plus latitude. We used an iterative process of GF model fitting to assess variable importance combined with variance inflation factor (VIF) analysis to assess collinearity, with the goal of retaining a final set of the most important and interpretable and least collinear variables to the greatest extent possible. VIF estimates the potential impact of multicollinearity in regression-type models by quantifying the extent to which standard errors are inflated due to collinearity compared to when uncorrelated variables are used. To calculate VIFs, we used the 'vifcor' function with a correlation threshold of 0.75 in the '*usdm* ' package (Naimi, Hamm, Groen, Skidmore, & Toxopeus, 2014) in R (R Core Team, 2018). The final set of seven variables included latitude (y), elevation (elev) and five bioclimatic variables: mean diurnal range (bio2), mean summer (bio10) and winter (bio11) temperature, and summer (bio18) and winter (bio19) precipitation. Latitude was included to characterize the gradient in daylength not otherwise captured by the bioclimatic variables. Of our seven retained variables, mean winter temperature (bio11) and latitude were found to have a VIF greater than 10 (10.31 and 17.00 respectively), which is considered an approximate rule-of-thumb cutoff VIF (Guisan, Thuiller, & Zimmermann, 2017). However, given that GF has some ability to accommodate correlated variables (Ellis et al., 2012) and daylength is an important determinant of phenology and height growth cessation in poplars (Soolanayakanahally et al. 2013), we opted to retain both of these variables in our models. We aggregated the individual-level environmental data to the level of populations (42 total) by taking the mean across individuals within each population (mean = 8 individuals per population, min=2, max=13).

Environmental data for common garden locations : To characterize climatic conditions at the two common garden experiments, we used DayMet (Thornton et al., 2014) data for 2014 and 2015, which spanned the relevant period from planting and establishment of the clonal replicates to when the growth measurements were collected. The Daymet dataset provides gridded estimates of temperature and precipitation for North America on a daily time step at ˜1 km x 1 km spatial resolution. We aggregated the daily data for 2014 and 2015 to produce monthly averages for maximum and minimum temperature and monthly sums for precipitation. We used these monthly summaries and the 'biovars' function in the '*dismo* ' package (Chamberlain, 2017) to calculate the same set of bioclimatic predictor variables used to characterize climate for the field collections.

Posted on Authorea 28 Aug 2020 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.159863198.86187354 — This a preprint and has not been peer reviewed. Data may be preliminary.

Univariate outlier detection methods: Full details of the GEA selection scans are described in Chhatre et al. (2019), and recapped again briefly here. Namely, we used two complementary approaches: (1) the population-based method *bayenv2*(Gunther & Coop, 2013) that corrects for population relatedness using a variance-covariance matrix of allele frequencies before testing for a correlation between SNP allele frequencies and an environmental predictor; and (2) the individual-based method *lfmm*(Frichot, Schoville, Bouchard, & Francois, 2013) that controls for background genetic structure by introducing latent factors into a linear mixed models while testing for associations between the genotypic state at each SNP and an environmental predictor. Both *bayenv2* and *lfmm* are inherently univariate, testing one environmental predictor's association with one SNP at a time. Therefore, to avoid inflating type I error rates, we used as environmental predictors in each approach the first 2 PC's from a principal components analysis on 19 bioclim variables (Hijmans et al., 2005) plus source latitude.

We generated empirical P -values for assessing significance in GEA tests, as advocated by Lotterhos & Whitlock (2014), using a subset (n =1,353) of non-coding intergenic SNPs based on the *Populus trichocarpa* v3.0 genome annotation. We used these intergenic SNPs to generate empirical null distributions of the test statistics used in our selection scans (Bayes Factors for*bayenv2* and ?-adjusted z-scores for *lfmm*), based on ranking each test SNP within the empirical null distribution and determining its quantile *sensu*Lotterhos & Whitlock (2014). Candidate SNPs with a test statistic equal to or exceeding that of the empirical null distribution were designated as selection outliers with a P [?] 0.000739 (=1/[1353+1]).

GF background: Ellis et al. (2012) provide details regarding the Gradient Forest algorithm and Fitzpatrick & Keller (2015) describe the application of GF to modeling genomic data and predicting genetic offsets. In brief, GF is a flexible, non-parametric extension of the machine learning approach known as Random Forests (Breiman, 2001). GF uses Random Forest to fit an ensemble of regression trees to model change in allele frequencies across sites and derive monotonic, nonlinear functions of environmental predictors. GF stands on the shoulders of Random Forests and inherits its robust statistical measures of model performance and variable importance. The predictive performance of the GF model for each SNP is quantified using the proportion of out-of-bag data variance explained (R^2), which is a robust estimate of generalization error. These robust goodness-of-fitR^2 values allow ranking of SNPs by how well the environmental gradients explain changes in SNP allele frequencies and can inform the detection of statistical outliers by selecting SNPs withR^2 values that exceed an empirical threshold (see below). The accuracy importance of predictors is quantified as the decrease in performance when each predictor is randomly permuted. For correlated variables, a conditional approach can be used (see Ellis et al., 2012). Lastly, from the ensemble of regression models, GF determines how well partitions distributed at numerous "split values" along each environmental variable explain changes in allele frequencies on either side of a split. The amount of variation explained by each split is the 'raw split importance'. The empirical, nonlinear turnover functions are constructed by distributing the R^2 values from all SNPs among the predictor gradients in proportion to their accuracy importance and along each gradient according to the density of the raw split importance values. The split importance values for all modeled SNPs also are aggregated to an overall, genome-wide turnover function for each variable using weightings based on predictor importance and the goodness-of-fit for each SNP model. Portions of gradients where split importance is high emerge as thresholds where genetic change is rapid (as might be expected between genetic groups). Gradients strongly associated with genetic variation will have more important splits therefore greater overall cumulative importance than gradients with little biological relevance. Important gradients will also have greater contribution to predicted genetic offsets.

GF outlier detection – Simulations: In previous work (Fitzpatrick & Keller, 2015; Fitzpatrick, Keller, & Lotterhos, 2018), we have advocated the use of GF to calculate genetic offsets on sets of outlier loci pre-ascertained using established methods for outlier detection that have been shown to have low false positive rates for the presumed demographic history of the sample. Here, we were interested in assessing the behavior of GF for outlier detection, given its strengths of incorporating multivariate predictors, interactions between predictors, and non-linear allele frequency gradients, but also its weakness of not controlling for demographic history. Rather than producing a comprehensive test of GF on an array of different demographic histories, we instead focused on quantifying statistical power and the rate of false positives under a demographic model

Posted on Authorea 28 Aug 2020 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.159863198.86187354 — This a preprint and has not been peer reviewed. Data may be preliminary.

of isolation-by-distance. We chose this based on our sample of Core region populations that show minimal genetic structure ($F_{ST} = 0.008$) but weak isolation by distance along an east-west gradient (Chhatre et al., 2019; Meirmans et al., 2017). We simulated population genetic data using CDPOP v1.2.20 (Landguth & Cushman, 2010) under a simple linear stepping stone model of 30 equal sized demes arrayed along a 1-dimensional spatial gradient, with each deme consisting of 20 unisexual individuals of equal sex ratio (total size = 600 inds). Each diploid individual consisted of 1,000 unlinked biallelic loci. To generate starting allele frequencies, we used the coalescent to simulate a site frequency spectrum (SFS) under a neutral Wright-Fisher model in *ms*(Hudson, 2002), removed loci with minor allele frequencies <0.1, and then randomly drew 1,000 loci from the SFS to seed the initial frequencies for each deme. Gene flow was simulated with a migration function that allowed adults to disperse to nearby demes prior to mating each generation. Probability of migration was a negative linear function of distance from the source deme, with probability declining to zero at a number of demes (u) distant from the source. To evaluate different levels of migration, we chose u values of 2, 4 or 8 demes away, which resulted in meanF_{ST} levels of 0.15, 0.04, and 0.008, respectively. After migration, mating occurred randomly among members within a deme, and females produced a Poisson distributed number of offspring with a mean and variance (? value) = 5. Offspring genotypes were assembled based on Mendelian inheritance at each locus and linkage equilibrium between loci, with a probability of *de-novo* mutation = 1e-8. For each migration scenario, we designated *locus 1* as experiencing one of three strengths of selection (weak: $s = 0.01$; moderate: $s = 0.1$; or strong: $s = 0.2$). Linear viability selection was applied whereby the probability of mortality decreased linearly from s to 0 for the dominant homozygote and heterozygote and increased linearly from 0 to s for the recessive homozygote across the 1D landscape. Under the non-linear scenario, the fitness of genotypes decayed or increased by the same magnitude, but followed an exponential function defined by (a) slope of mortality decay and (b) the$y$ -intercept (Supplementary Figure S1). We also included a neutral scenario for each migration level in which *locus 1* experienced no selection ($s = 0$). Thus, there were a total of 12 simulated scenarios (3 migration rates x 4 selection strengths), which we replicated independently 100 times per scenario. For each simulation replicate, we obtained allele frequencies per deme from the 100^{th} simulated generation and used these in GF to identify outlier status using the same settings applied to empirical data (see below). For scenarios where $s > 0$, we calculated power as the proportion out of 100 simulation replicates where *locus 1* was identified as an outlier in the distribution ofR^2 values from GF for a given significance level (range of ?: 0.001 - 0.1). For the neutral scenario ($s = 0$), we calculated the false positive rate (? = 0.05) as the proportion out of 100 simulation replicates where *locus 1* was in the 95% quantile of R^2 values from GF.

GF outlier detection – Empirical: In addition to*bayenv2* and *lfmm* , we also used GF to select outlier SNPs using two types of allele frequency estimates: (1) "raw" or uncorrected allele frequencies (hereafter termed "GF-Raw"), and (2) standardized allele frequencies (X) output from *bayenv2*(Gunther & Coop, 2013) corrected using the estimated population genetic (co)variance matrix (hereafter termed "GF-X "). GF-Raw used direct estimates of population minor allele frequencies based on our sample of N individuals per population (mean N = 8; range: 2 - 13). The GF-X approach was included to explore a more robust approach to allele frequency estimation that corrects for finite sampling and population relatedness prior to feeding into GF for outlier detection. To select GF-Raw outliers, we fit GF to the raw minor allele frequencies for the 107,309 SNPs and obtained an R^2 for each locus. We compared the distribution of the resulting R^2values to the empirical P -value derived from the 1,353 intergenic SNPs. To select GF-X outliers, we first obtained standardized allele frequencies (X) estimated using the fitted omega matrix in*bayenv2* as described above (*cf* . Univariate outlier detection methods) and using the '-f' flag to output 190 MCMC draws of$X$ for each candidate SNP. Following advice given in the*bayenv2* manual, we then fit GF models separately to each of the 190 MCMC draws and extracted the resulting 190R^2 values for each of the 107,309 SNPs. From these values, we calculated the median R^2 for each candidate SNP, and determined empirical P- values by determining their rank within the distribution of the medianR^2 values from the 1,353 intergenic SNPs.

Post-outlier GF modeling: We fit GF to each of the outlier SNP data sets (Bayenv, LFMM, GF-Raw, and GF-X) and to the set of SNPs identified as outliers by both Bayenv and LFMM (Bayenv-LFMM). For each SNP dataset, we fit GF using 500 regression trees per SNP and a variable correlation threshold of 0.5

Posted on Authorea 28 Aug 2020 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.159863198.86187354 — This a preprint and has not been peer reviewed. Data may be preliminary.

to invoke conditional importance estimates (Ellis et al., 2012). We used default values for the number of predictor variables randomly sampled as candidates at each split and for the proportion of samples used for training and testing each tree. For comparison, we also fit GF to 999 sets of 500 SNPs selected at random from the full set of 107,309 SNPs, which we combined into a single model using the 'combinedGradientForest' function. All models were fit using the '*gradientForest* ' library (Ellis et al., 2012) in R (R Core Team, 2018).

GF genetic offsets: We used each of the GF models and the 'predict.GradientForest' function to transform environmental conditions described by the seven predictor variables into common units of compositional turnover (1) throughout the geographic range of balsam poplar and for (2) each population and (3) the two common gardens. To quantify genetic offsets resulting from transplanting each population from its home environment to the common gardens, we calculated the Euclidean distance between each population and the common gardens in the resulting multidimensional transformed environmental space from GF. For comparison, we also calculated the Mahalanobis distance (Mahalanobis, 1936) between each population and the common gardens using the raw (untransformed) environmental predictors, which serves as a 'naive' climate transfer distance uninformed by genomic patterns.

To visualize and compare genetic patterns in geographic and biological space, we used Principal Components Analysis (PCA) to reduce the seven transformed environmental variables into three factors. The PCA was centered but not scale transformed to preserve differences in the magnitude of genetic importance among the environmental variables. Variation in genetic composition was visualized as (1) a bi-plot of the first two principal components with labeled vectors indicating the direction and magnitude of major environmental correlates and (2) by mapping the patterns back to geographic space. Variation in genetic composition was visualized by assigning the first three principal components to an RGB color palette. The resulting color similarity corresponds to the similarity of expected patterns of genetic composition. For comparison, we repeated this process using the raw (untransformed) environmental variables.

GF models fit to different sets of SNPs are expected to produce different predicted patterns of genomic variation. To estimate and visualize differences in expected geographic patterns for each of the six sets of modeled SNPs, we used Procrustes superimposition on the PCA ordinations, where the matrices were rotated to minimize the sum of squares of the distances between the sites in genetic space (Peres-Neto & Jackson, 2001). The Procrustes residuals, which in this case measure the absolute distance between sites in genetic space and the rotated ordination space, were mapped to visualize differences in the predicted genetic composition patterns between all pairwise comparisons of the six different GF models fit using the different sets of SNPs. For all visualizations, we constrained predictions to within the geographic range of balsam poplar as defined by (Little, 1971).

Genetic offset predictions of growth in common gardens: To experimentally test how well genetic offset predicts performance in the field when populations experience a novel climate, we used genetic offset values to predict population mean height growth increment in the two common gardens. We first generated population BLUPs of height growth separately for each garden (VT and IH) using linear mixed-effects models of the form:

$$Y_{\mathrm{ijk}} \;=\; \mu \;+ B_i + \;P_j + P(G)_{\mathrm{jk}} + \varepsilon_{\mathrm{ijk}}$$

where height growth (Y) is modeled as a function of the overall mean (μ), the fixed effect of block or position within the garden (B_i), the random effect of population (P_j), the random effect of genotype nested within population ($P(G)_{\mathrm{jk}}$), and a normally distributed residual error ($\varepsilon_{\mathrm{ijk}}$). The resulting population-level BLUPs from each model were then merged across gardens and combined with their corresponding garden-specific genetic offset predictions for use in a second model in which height growth was predicted as a function of genetic offset while controlling for overall differences between gardens. To allow for potential non-linearity in the response, we included a quadratic term, giving a final model of the form:

$$Y_{\mathrm{ij}} \;=\; \mu \;+\; Go_i + \;\mathrm{Go}_i^2 + \mathrm{Gd}_j + \varepsilon_{\mathrm{ij}}$$

Posted on Authorea 28 Aug 2020 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.159863198.86187354 — This a preprint and has not been peer reviewed. Data may be preliminary.

where Go and Go^2 are the linear and quadratic genetic offset terms, respectively, and Gd is the effect of the j^{th} garden (VT or IH).

We evaluated height growth for different estimates of genetic offset derived from (a) a combined GF model of all outlier loci, (b) outliers identified by GF-X , (c) the average of 999 random draws of 500 random SNPs from the genomic background, and (d) Mahalanobis distances based on climate-only. When not significant, we dropped the quadratic term in favor of the simple linear model. We report the percentage of variance explained (R^2) as a metric of model performance.

RESULTS

GF outlier detection – Simulations: Testing GF without any correction for population structure (equivalent to GF-Raw) against simulated scenarios of 1D isolation-by-distance showed that under linear selection, GF had good power (>0.8 at ? = 0.05) to detect loci under moderate to strong selection for most migration scenarios, although power was reduced somewhat under moderate selection with high migration ($s = 0.1$, $u = 8$) (Fig. S2a). Under weak selection ($s = 0.01$), GF was under-powered to detect selection under all migration scenarios. Under non-linear selection (Fig. S2b), power was also generally low (<0.5) for all but strong selection ($s = 0.2$) and low to moderate migration ($u = 2$ or 4). However, the false positive rate was well calibrated between 0.04-0.06 for ? = 0.05 (Fig. S2c). Thus, under this specific scenario of 1D isolation by distance, GF had good power to detect moderate to strong linear selection or strong nonlinear selection with low frequencies of false positives.

GF outlier detection – Empirical : Out of 107,309 high-quality SNPs, 23 (0.02%) were identified as statistical outliers by all four outlier detection methods (Fig. 1). GF-X detected the fewest number of outliers (120), had the smallest number of outliers unique to that method (22), and therefore shared the largest proportion (98/120=81.67%) of statistical outliers with one or more of the other outlier detection methods. In contrast, Bayenv detected the largest number of statistical outliers (320) and GF-Raw had the largest proportion (234/291=80.41%) of detected outliers unique to that method.

GF modeling of SNP outliers - Of the 320 outlier SNPs detected using Bayenv, 242 (75.62%) had an R^2greater than zero in the GF model. This compares to 146 of 310 (47.1%) outlier SNPs for LFMM and 42 of 71 (59.2%) outlier SNPs for Bayenv-LFMM (note that by definition all (100%) GF-Raw and GF-X outliers had an R^2 greater than zero). On average 49.64 of 500 (9.93%) SNPs had an R^2 greater than zero in the 999 GF models fitted to randomly selected SNPs.

Latitude was the most important predictor for all sets of SNPs (both outliers and random), followed by winter temperature (bio11), whereas elevation and diurnal range (bio2) were the least important variables (Supplementary Fig. S3). GF-Raw had the strongest associations (highestR^2 of all models) with all variables, and therefore the aggregate turnover functions for GF-Raw attained the greatest maximum height for all variables (Fig. 2). Random SNPs had the weakest associations (lowest R^2 of all models) for all variables except elevation and diurnal range (bio2), for which Bayenv-LFMM had a lower R^2. Although the aggregate turnover functions differed in their maximum height, reflecting differences in variable importance, most of the aggregate turnover functions based on outlier SNPs had a similar shape, with thresholds falling in the same general region of the gradients (Fig. 2). The GF-Raw and GF-Random turnover functions were notable exceptions to this pattern. Unlike the aggregate turnover functions for the five sets of outlier SNPs, which exhibited pronounced thresholds, the turnover functions for SNPs selected at random largely lacked thresholds and instead turnover tended to be relatively constant along the seven environmental gradients. For GF-Raw, SNP turnover was more rapid at the colder and drier portions of the temperature and precipitation gradients than other sets of outlier SNPs, reflecting the substantial differences in patterns of turnover in the individual outlier SNPs uniquely detected by GF-Raw (Fig. 3, Supplementary Fig. S4). Integrating across all environmental predictors, the total R^2distribution across SNPs showed marked differences among different outlier detection methods (1-way ANOVA: $F = 152.18$; df = 3, 823;$P < 0.0001$), with the highestR^2 values coming from GF-Raw and GF-$X$and lower R^2 values from outliers detected by*bayenv2* and *lfmm* (Supplementary Fig. S5).

Posted on Authorea 28 Aug 2020 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.159863198.86187354 — This a preprint and has not been peer reviewed. Data may be preliminary.

Spatial patterns of genomic variation - The GF models fit to different sets of SNPs produced different predicted patterns of genomic variation (Fig. 4). The most similar mapped predictions were between GF models fitted to outlier SNPs from Bayenv, LFMM, Bayenv-LFMM, and GF-X . Differences in predicted spatial patterns were greatest between GF-Raw and all other sets of outlier SNPs, followed by GF fit to SNPs selected at random, with the largest range-wide differences being between GF-Raw and GF-X . Differences in mapped patterns were generally greatest in the southern third of balsam poplar's range and for most comparisons reached a maximum in a latitudinal band centered near 50° N and in trailing range edge populations in the Rocky Mountains.

Genetic offsets & climatic transfer distances - Because GF-X had the largest proportion of outliers that overlapped with other detection methods (and conversely, the smallest proportion of unique SNPs), here we report results for GF-X only. Northwesternmost populations, most distant from VT were predicted to have the largest genetic offsets associated with transplanting populations from their home environment to the VT common garden (Fig. 5a). The pattern of predicted genetic offsets was largely reversed for transplanting populations to the IH common garden: populations in the southeasternmost portion of the range, farthest from IH, were predicted to have the largest genetic offsets (Fig. 5c). This resulted in a highly significant negative correlation for the genetic offsets between the two garden sites (r = 0.897, df = 40, P <0.0001; Supplementary Fig. S6). In contrast, climate-only transfer distances (i.e., genetically-naive climate distances based on Mahalanobis distance) showed no clear cline with distance from the common gardens (Fig. 5b, d), and in fact climate-only distances showed a weak but positive correlation across gardens (r = 0.380, df = 40, P= 0.013; Supplementary Fig. S6).

Plotting the populations and the common gardens in the transformed multidimensional genomic space and the untransformed multidimensional environmental space reveals the locations of populations relative to the common gardens in terms of expected genomic similarity (Fig. 6a) and climatic similarity (Fig. 6b), thereby providing a means to conceptualize genetic offsets and climate transfer distances (though in only two of the seven dimensions as variation along additional axes is not shown). Consistent with variable importance ranking, latitude (y) and winter temperature (bio11) have the strongest contribution to variation in the multidimensional genomic space (as indicated by the length of the vectors in Fig. 6a). Shading indicates the degree of expected similarity of genetic patterns, with locations with similar shading being expected to have similar genomic composition. Numerous populations are predicted to have similar genomic patterns as those for the climate of the VT common garden. These populations plot near the VT common garden in the transformed genomic space and therefore have lower predicted genetic offsets for movement to VT common garden climate. In contrast, all seven variables have roughly equal contribution to variation in the untransformed environmental space (Fig. 6b) and the locations of populations and their distances from the common gardens reflects climatic similarity rather than underlying genomic patterns. For example, SSR is located within the unique higher elevation climate space (Fig. 6b), despite having predicted genetic composition similar to some eastern populations (Fig. 6a).

Genetic offset prediction of common garden performance - Genetic offset was significantly associated with the realized performance of populations transferred to the novel environments of the common gardens. Genetic offset models explained >60% of the variation in height increment growth (Table 1). Consistent with predictions, height growth was highest for populations experiencing the lowest values of genetic offset and declined with larger values of offset (Fig. 7). The shape of the height-offset relationship was non-linear, represented by a significant quadratic effect (Table 1), and exhibited the steepest decline as offset increased above zero followed by a flattening out at larger genetic offset values. Surprisingly, the estimates of genetic offset made from the random selection of SNPs from the genomic background were just as good or slightly better (R^2 = 0.66) than genetic offsets based on outlier loci (R^2 = 0.61-0.63). Climate-only distance had a negative linear association with height growth, but was a weaker predictor overall, explaining a bit more than half the variance in growth compared to genetic offset models (R^2= 0.34).

For the subset of populations that were phenotyped in both VT and IH (N=9 of 41), we observed a clear rank order change and crossing reaction norms in the genetic offset predictions, indicative of a tradeoff in the

Posted on Authorea 28 Aug 2020 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.159863198.86187354 — This a preprint and has not been peer reviewed. Data may be preliminary.

locally adaptive gene-climate relationship across sites (Fig. 8). Consistent with the prediction of a tradeoff, the height growth of populations tended to increase or decrease in a trend that was inverse to the change in genetic offset across sites, although without consistent change in the rank-order of populations. Accordingly, the per-population difference in height growth between sites (VT minus IH) was negatively correlated with difference in offset (Spearman's rho = -0.6, $P_{\text{1-tailed}} = 0.048$), although with only 9 populations statistical power was limited.

DISCUSSION

The primary goal of this study was to provide an experimental test of the extent to which genetic offsets, a correlative space-for-time approach, can predict performance of populations exposed to new environments. By transplanting individuals from their home environment to the novel climate of the common gardens, we substituted space for time as a proxy for rapid climate change. We found that genetic offsets based on existing gene-environment relationships work well to predict performance of populations experiencing new environments - and much better than climate differences alone (Table 1). We view this finding as encouraging preliminary evidence that genetic offsets may represent a first order estimate of the degree of expected maladaptation of populations exposed to novel environments. While our study considered climate differences across geographic space, in principle our findings should be relevant to temporal changes in climate as well. As such, genetic offsets could provide a means to estimate aspects of population-level vulnerability to climate change. Additional research is warranted to determine the extent to which our findings are generalizable to other systems and populations growing in natural environments.

That genetic offsets outperformed naive climate distances is not surprising and can be best understood by considering the similarities and distinctions between these two methods. In many ways, genetic offset share a conceptual foundation with climate transfer distances long used in forestry (Mátyás, 1996). The establishment of provenance trials, in which tree seed from multiple locations are collected and grown in multiple sites, has allowed for evaluation of tree performance as a function of differences in climate between sources and planting sites (i.e., response functions derived from climate transfer distances; Wang, Hamann, Yanchuk, O'Neill, & Aitken, 2006; Wang, O'Neill, & Aitken, 2010). These experiments provide excellent insight into the climate variables that best predict phenotypic performance upon transfer to a new site, but are time and labor-intensive, and not practical for most study systems. A simpler approach is to delineate climate-based seed zones from which seeds should be selected for restoration under the hypothesis that maladaptation of seedlings is minimized (and production is maximized) when movement of seeds is restricted to other sites with similar climate (e.g., Bower, St Clair, & Erickson, 2014; Pike et al., 2020). The distinction between the "traditional" climate transfer distances used for seed zone delineation and genetic offsets is simply that genetic offsets use re-scaled climate distances based on the modeled associations with (adaptive) genomic variation, whereas climate distances typically weigh the included variables equally despite potential variation in their adaptive importance. Existing gene-environment relationships described by the fitted turnover functions from GF provide the mechanism that allows proper weighting of different climate variables, based on how allele frequencies are aligned with climate gradients. Gradients strongly associated with genomic variation (and portions of these gradients where genetic patterns change most rapidly) will have greater contribution to genetic offsets than will unimportant variables (or portions of gradients where allele frequencies generally are constant; Capblancq, Fitzpatrick, Bay, Exposito-Alonso, & Keller, 2020; Fitzpatrick & Keller, 2015). This also fits well with a recent study in lodgepole pine, (*Pinus contorta*), in which the climate variables identified as important in GEA models were strongly correlated ($r = 0.9$) with the climate variables associated with phenotypic performance in a 20-year provenance trial (Mahony et al., 2020). This suggests that one of the realized benefits of GEA may be in identifying which among a set of climate variables are most predictive of local adaptation, which is the same principle being employed by GF to weight different climate variables based on the strength of the genomic association when calculating genetic offsets. The use of GEA plus genetic offsets may prove useful for conservation planning in long-lived species or those for which phenotypic information from experimental assessment of field performance is lacking.

The finding that genetic offsets had good predictive power regardless of whether they were based on sets of

Posted on Authorea 28 Aug 2020 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.159863198.86187354 — This a preprint and has not been peer reviewed. Data may be preliminary.

outlier SNPs or simply SNPs selected at random from the genome (which surprisingly slightly outperformed genetic offsets based on outlier SNPs) is harder to explain. One explanation is if allele frequencies of the genome as a whole tend to be aligned with the same environmental gradients that are important to local adaptation (i.e., the gradients of adaptive and neutral genomic background are parallel or proportional), then one could serve as an adequate proxy for the other. If this is the case, then SNPs selected at random should provide the same rank weighting of the climate gradients as would outlier SNPs, which was generally the case in our study (Fig. 2, Supplementary Fig. S3). However, as mentioned above, the shapes of the turnover functions also will influence genetic offsets. All else being equal, larger genetic offsets will occur for populations transferred between environments on either side of a threshold as compared to populations transferred along flat portions of allele turnover gradients. Assuming these nonlinearities reflect true signals of local adaptation, we would then expect genetic offsets that incorporate these patterns to outperform linear methods that do not. Our findings do not support this expectation. In this study, the turnover functions based on outlier SNPs often showed pronounced nonlinearities, whereas those based on randomly sampling SNPs from the genomic background tended to be more linear (Figs. 2, 3), yet genetic offsets based on outliers tended to be strongly correlated with those from random SNPs (Supplementary Figs. 6). Further, random SNPs slightly outperformed outlier SNPs in explaining height growth in the common gardens. Additional research is required to determine whether this result is an artefact of our study or a more general pattern.

Another primary goal of our study was to explore differences between GF models fit to different sets of outlier SNPs. There are numerous ways to detect statistical outlier SNPs, and, as was the case in this study, it is not uncommon for different methods to identify different SNPs as outliers, leaving some uncertainty regarding which SNPs are false vs. true positives, and therefore which SNPs truly are associated with climate adaptation and thus most informative from a predictive standpoint. By fitting GF models to different sets of outlier SNPs, we can ask: To what extent do different sets of outlier SNPs produce different inferences? We found that although the different outlier methods detected different sets outlier SNPs (Fig. 1), GF models fit to different sets of outliers from *bayenv2* , *lfmm* , and GF-X were similar in terms of variable importance ranking (thoughR^2 values differed, Supplementary Figs. 3 and 5), the general shapes of the turnover functions (Figs. 2, 3), and therefore, the predicted spatial patterns of genetic variation (Fig. 4), and by extension, the predicted genetic offsets (Supplementary Fig. 6). GF models fit to SNPs selected at random or those selected using allele frequencies uncorrected for population structure (GF-Raw) also generally followed the same pattern of variable importance ranking as other outlier detection methods, but given that these methods selected a large proportion of unique SNPS, they produced turnover functions and predicted spatial patterns that differed from each other and from*bayenv2* , *lfmm* , and GF-X . The similarity in variable importance ranking and predictions from different sets of outlier SNPs would arise if (1) the outlier SNPs they shared in common tended to have strong relationships with climate (and therefore would have greater contribution to the fitted turnover functions from GF; (Ellis et al., 2012) and/or (2) the outlier SNPs unique to each method tended to have similar relationships (i.e., shapes of turnover functions) with climate. We have evidence for both possibilities. The shapes and cumulative importance of the turnover functions for the outlier SNPs unique to *bayenv2* ,*lfmm* , and GF-X were similar (Fig. 3) and the totalR^2 from GF models increased for SNPs as their outlier status was shared among an increasing number of detection methods (Supplementary Fig. S7). Outliers unique to a single method likely represent a mix of false positive SNPs along with some true positives that may be better detected by one method over another, although these are difficult to separate in real data. Our experimental design and sampling strategy were specifically chosen to minimize false positives arising from demographic history, and our simulations testing GF-Raw suggested a low type I error rate under a simple scenario of isolation by distance. However, under more complex demographic histories we would expect GF-Raw to be prone to false positives because it does not have an internal control for neutral population structure. Given this, and the observed reduction in unique outliers identified by GF before and after correcting for population relatedness (i.e., GF-Raw vs. GF-X), we advocate fitting GF only to allele frequencies that have been properly corrected for demographic history.

In terms of outlier detection, it is notable that GF-X detected the fewest outliers overall and the fewest

Posted on Authorea 28 Aug 2020 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.159863198.86187354 — This a preprint and has not been peer reviewed. Data may be preliminary.

outliers unique to that method (Fig. 1). Unlike *bayenv2* and *lfmm* , GF is multivariate, can accommodate interactions between variables, and assumes no parametric form of the allele frequency ~ environment relationship (although it does assume monotonicity). Therefore, GF-X may be less prone to the multiple testing problem inherent in univariate methods, or to outlier loci arising due to departures from the assumed linear model. As such, the combination of GF run on standardized allele frequencies produced by *bayenv2* as done in this study (GF-X) could provide a more holistic approach to multivariate outlier detection that is robust to the shape of the allele frequency ~ environment relationship, while also correcting for finite sampling and population structure. Because GF reports an R^2 for each predictor variable in the model as well as for the model as a whole, it also provides a means to consider outlier status from the context of individual climate gradients as well as more comprehensively. Taken together, we feel GF warrants further study as a useful outlier detection method under simple demographic histories, or when provided with allele frequencies that have been corrected for population relatedness; especially for systems under strong, linear selection and intermediate migration (supplementary Figure S2).

While still a new and largely untested method, GF is increasingly being applied to genomic studies, including quantifying population-level climate change vulnerability. However, concerns have been raised about the application of genetic offsets in this capacity, especially for mobile organisms with short generation times (Fitzpatrick, Keller, et al., 2018). Common garden experiments are not perfect proxies for climate change or organisms in natural environments, but our results suggest that existing genetic patterns across space and associated genetic offsets may be informative for predictions across time as well - even if these predictions are based on neutral genetic patterns. Given the inherent complexities, for most any organism it will be challenging to predict the exact *nature* of genomic change in response to environmental change. However, for some organisms, it may be possible to use existing gene-environment relationships to develop adequate assessments of the *magnitude* of expected genomic change based on genetic offsets, which can provide a proxy for population-level exposure to climate change.

ACKNOWLEDGEMENTS

We appreciate the contributions of Ron Zalesny, Andrew Elmore, and Steven Guinn for help with sample collections, Áki Láruson, Ben Haller, and Katie Lotterhos for discussions about testing genetic offset with common gardens, and Erin Landguth for help with CDPOP. This work was supported by National Science Foundation Plant Genome Research Program award #1461868 to SRK and MCF.

DATA ACCESSIBILITY

Upon acceptance, all data supporting the results in the paper will be archived in appropriate public repositories (e.g., Dryad, GenBank, etc).

AUTHOR CONTRIBUTIONS

MCF and SRK designed the research. VC prepared the genomic data and performed the selection scans and simulations. MCF performed the GF modeling and associated analyses. SRK oversaw the genomic and simulation analyses, managed the VT common garden and analyzed the common garden data. RS managed the IH common garden, provided data, and edited the manuscript. MCF and SRK led the writing of the manuscript with contributions from VC.

Table 1. Model estimates (SE) relating common garden height increment growth to genetic offset predictions from Gradient Forests. Offset predictions were tested for different sets of outlier loci: (a) All outliers presented in Fig. 1; (b) GF-X outliers; (c) random sampling of SNPs from the genomic background; and (d) the climate-only Mahalanobis distance based between source and garden sites.

	(a) All outliers	(b) GF-X	(c) Random-SNPs	(d) Climate-only
Intercept	46.33 (3.66)****	39.81 (4.68)****	50.63 (6.93)****	37.23 (2.79)***
Offset	-6.18e2 (1.48e2)***	-3.94e2 (1.07e2)***	-2.37e3 6.16e2)***	-0.28 (0.09)**
Offset^2	3.50e2 (1.11e2)**	1.55e2 (6.12e2)*	3.80e4 (1.4e4)**	–

Posted on Authorea 28 Aug 2020 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.159863198.86187354 — This a preprint and has not been peer reviewed. Data may be preliminary.

Garden	2.39 (0.90)*	5.34 (2.29)*	5.76 (2.23)*	8.74 (2.77)**
N SNPs	823	120	500[a]	n/a
Model R^2	0.61	0.63	0.66	0.34

[ns]not significant, * $P < 0.05$, **$P < 0.01$, *** $P < 0.001$, **** $P < 0.0001$.

[a]Average genetic offset from 999 random draws of 500 SNPs from the genomic background.

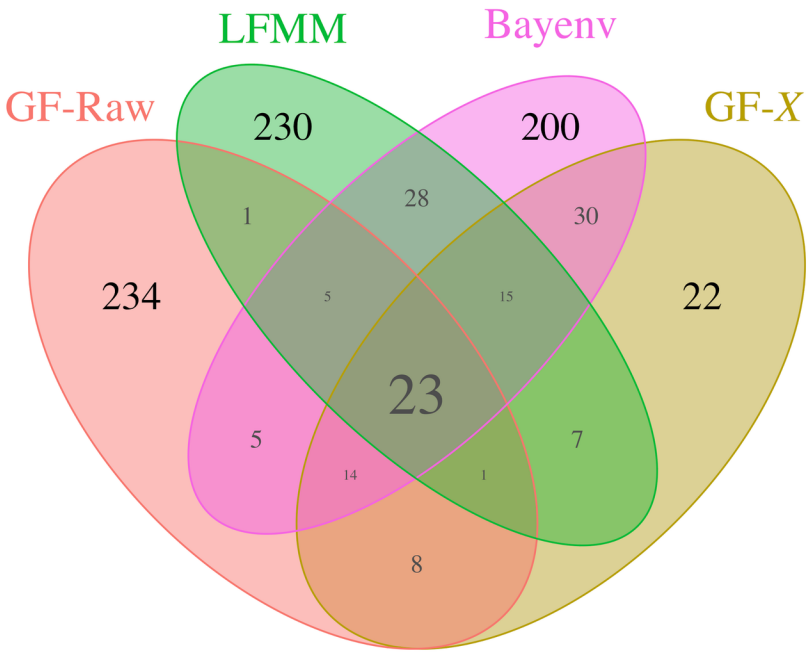

FIGURE 1 - Venn diagram showing the intersection between four outlier detection analyses. Each ellipse represents a different detection method and numbers correspond to either (black) the number of unique outliers from each method or (gray) the number of shared outliers among methods.

Posted on Authorea 28 Aug 2020 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.159863198.86187354 — This a preprint and has not been peer reviewed. Data may be preliminary.

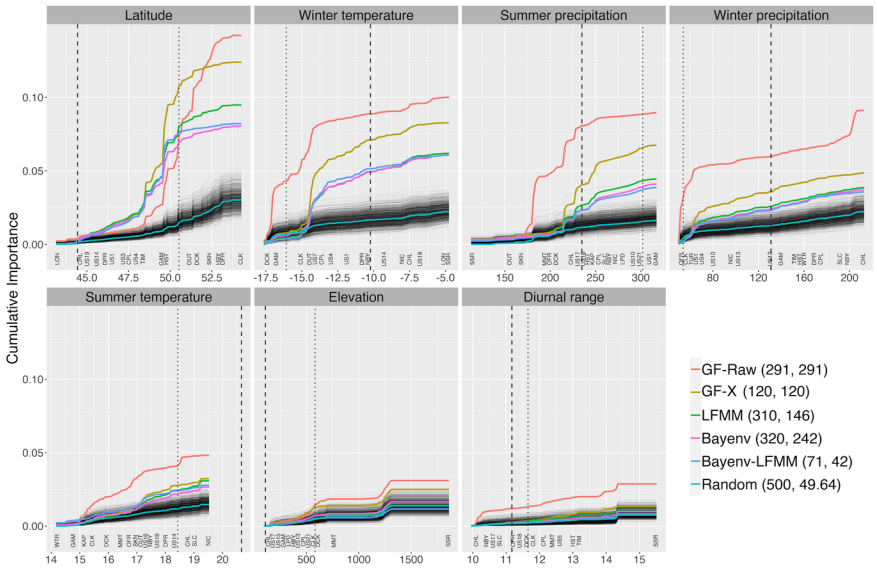

FIGURE 2 - Compositional turnover functions from GF models fit to different sets of outlier and randomly selected SNPs. Each bold, colored curve in each panel is an aggregate turnover function from a different GF model representing the R^2-weighted average of the individual turnover functions (shown in Figure 3) fitted to the number of SNPs indicated in the legend. For example, the green line indicates the overall turnover function from a GF model fit to minor allele frequencies of 310 significant outliers identified by *lfmm* analyses, 146 of which had an R^2 greater than zero in the GF model and therefore contributed to the aggregate function for this model. The thin black lines indicate aggregate turnover functions from GF models fit to 999 different sets of 500 randomly selected SNPs and the teal line represents the combined aggregate of these 999 models. The location of where each population occurs on each gradient is labeled along the x-axes (for clarity only a subset of populations are shown). The vertical dashed and dotted lines indicate where the VT and IH common gardens respectively fall along each gradient given the local conditions during the common garden experiments in 2014-2015.

Posted on Authorea 28 Aug 2020 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.159863198.86187354 — This a preprint and has not been peer reviewed. Data may be preliminary.

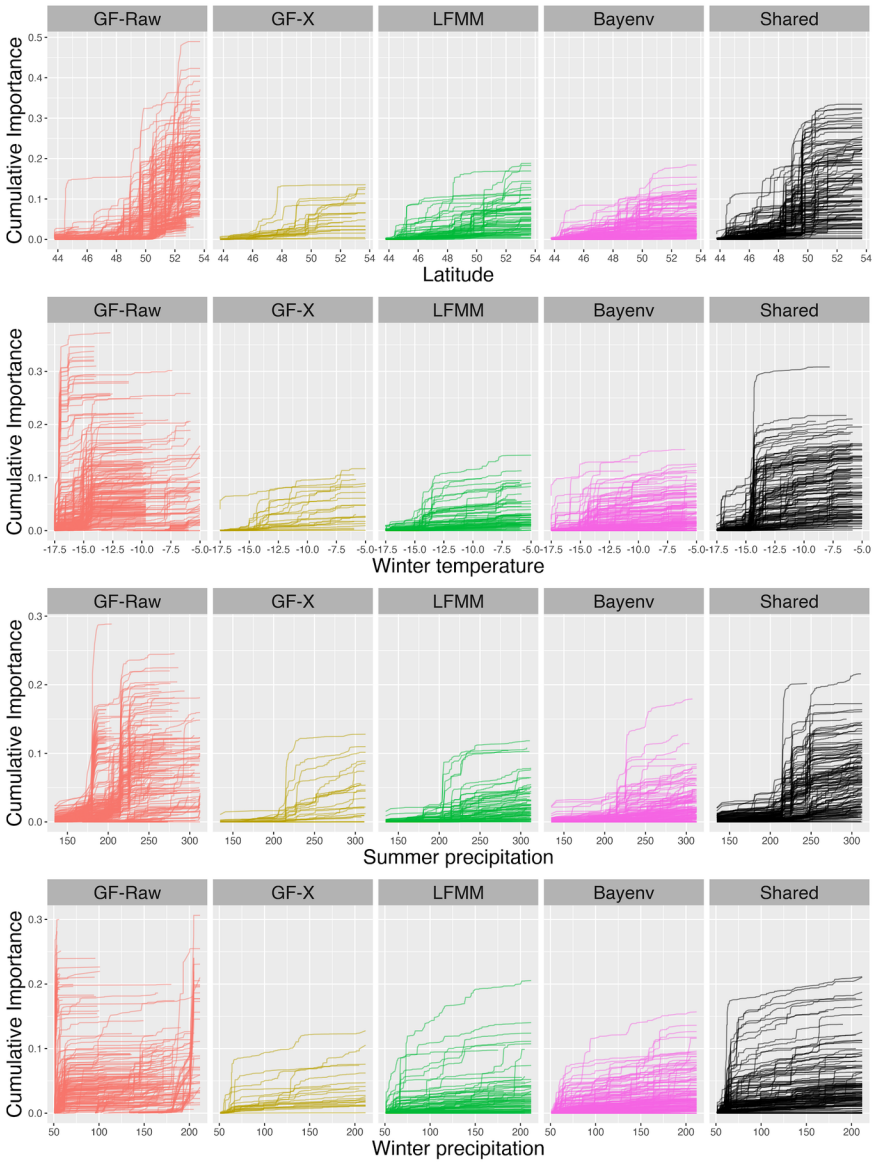

FIGURE 3 - SNP-level compositional turnover functions for the four outlier detection methods (columns) and variables (rows; only the first four most important variables are shown, see Supplementary Figure S4 for the remaining three variables). The first four columns show the fitted SNP-level turnover functions for outlier SNPs unique to that method, whereas the last column ('Shared') plots the fitted SNP-level turnover functions for SNPs detected by two or more outlier detection methods.

Posted on Authorea 28 Aug 2020 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.159863198.86187354 — This a preprint and has not been peer reviewed. Data may be preliminary.

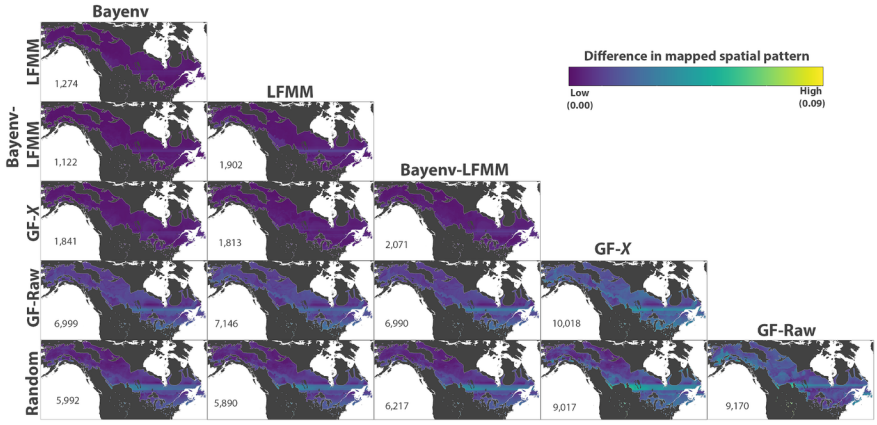

FIGURE 4 - Mapped pairwise comparisons of predicted patterns of spatial turnover in genetic composition from six GF models fit to different sets of outlier and randomly selected SNPs. The number in the lower left of each panel is the sum of the Procrustes residuals across all pixels within the range of balsam poplar for each pairwise model comparison, with higher values and lighter shading indicating greater differences in predicted geographic patterns between the model pairs.

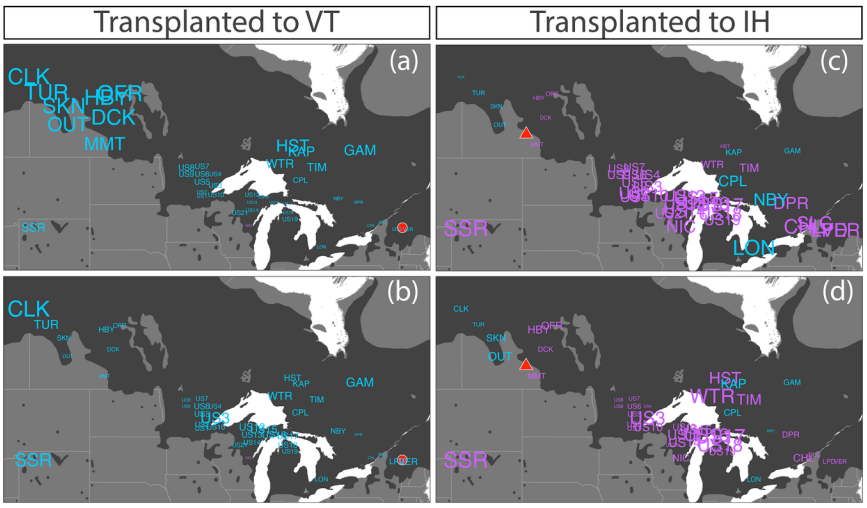

FIGURE 5 - Maps of population locations and common gardens, with label sizes scaled to (a, c) predicted genetic offsets from GF-*X* or (b, d) climatic transfer distances for transplanting populations from their home environment to either the (red dot) VT or (red triangle) IH common garden. Only populations with cyan labels were planted in the respective garden in each column. Dark shading is the range of balsam poplar from (Little, 1971).

Posted on Authorea 28 Aug 2020 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.159863198.86187354 — This a preprint and has not been peer reviewed. Data may be preliminary.

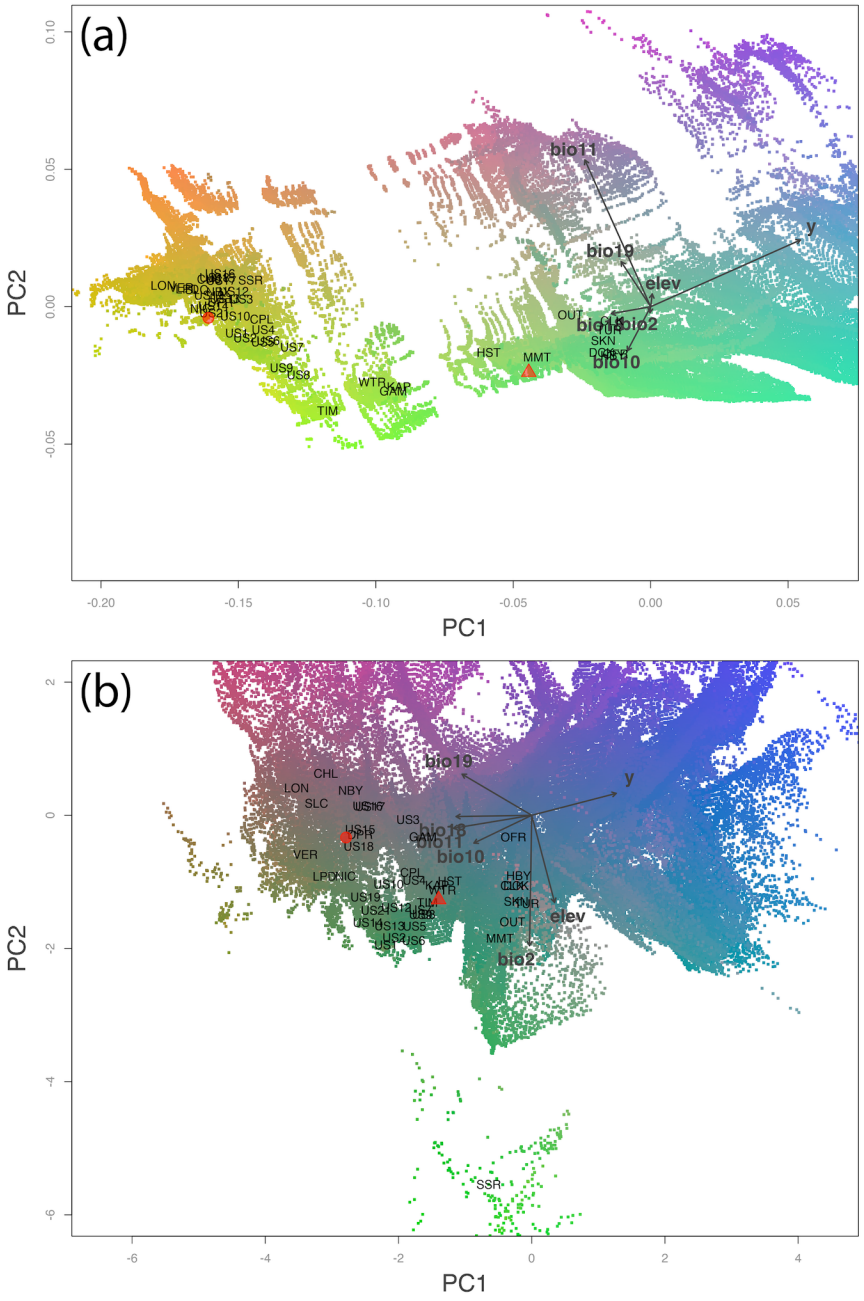

FIGURE 6 - Bi-plots showing (a) predicted variation in genetic composition from GF-X and (b) variation in untransformed environmental conditions (each colored point is a climate grid cell within a subset of the range of balsam poplar containing the sample populations). Locations of the 42 populations within each of these spaces are indicated by three letter population codes. The VT and IH common gardens are indicated by a red dot and red triangle respectively. Colors in (a) represent gradients in genetic turnover derived from transformed environmental predictors; locations with similar colors are expected to harbor populations with similar genetic composition. In (b), color similarity corresponds to environmental similarity uninformed by genetic patterns. Labeled vectors indicate the direction and magnitude of gradients with greatest contribution (y=latitude, elev=elevation, bio2=mean diurnal range, bio10/bio11=mean summer / winter temperature,

bio18/bio19=summer/winter precipitation). Genetic offsets are approximately the Euclidean distance from each population to the common gardens in (a), whereas climate transfer distances are approximately the Euclidean distance from each population to the common gardens in (b), though note that variation in higher dimensions is not shown.

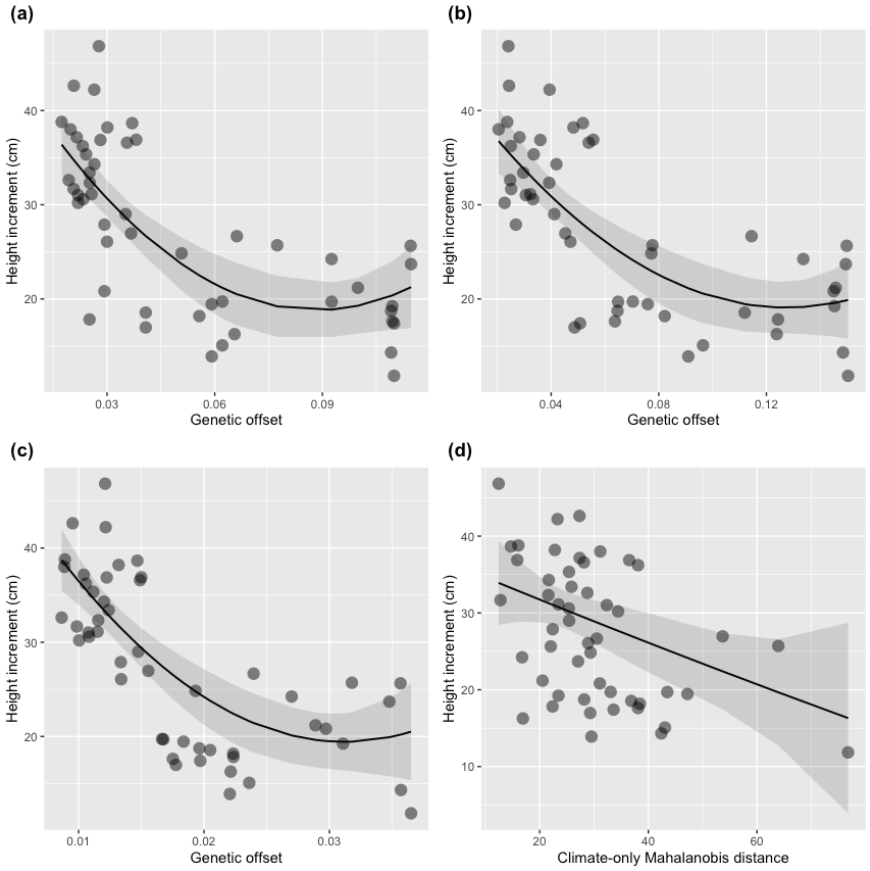

FIGURE 7 - Predicted genetic offset from Gradient Forest versus height growth increment measured in common gardens. The GF turnover functions in Figure 2 were used to predict the genetic offset of populations when transplanted from their source climate to the climate of the common garden. The fitted lines and confidence intervals show the modeled relationships between genetic offset and height growth increment based on a quadratic fit (panels a-c) or linear fit (panel d). Different offsets are plotted based on (a) all outlier loci, (b) GF-X , (c) average of the 999 random draws of 500 background SNPs, and (d) climate-only Mahalanobis distances.

FIGURE 8

Posted on Authorea 28 Aug 2020 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.159863198.86187354 — This a preprint and has not been peer reviewed. Data may be preliminary.

Posted on Authorea 28 Aug 2020 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.159863198.86187354 — This a preprint and has not been peer reviewed. Data may be preliminary.

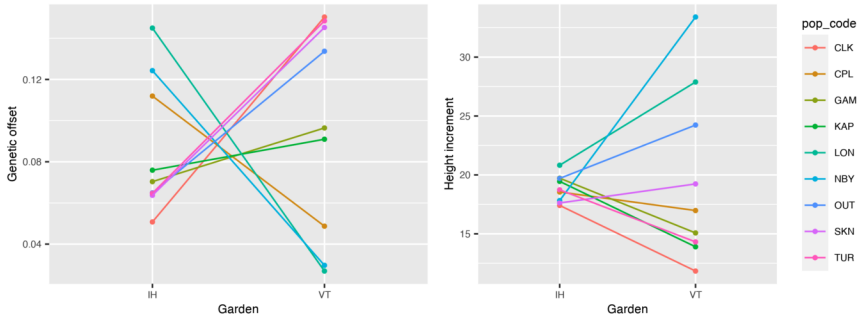

FIGURE 8 - Norm of reaction for height increment and genetic offset (GF-X outliers) for the subset of nine populations represented in both gardens.

LITERATURE CITED

Aguirre-Liguori, J. A., Ramírez-Barahona, S., Tiffin, P., & Eguiarte, L. E. (2019). Climate change is predicted to disrupt patterns of local adaptation in wild and cultivated maize. *Proceedings of the Royal Society B: Biological Sciences*, *286*(1906), 20190486.

Bay, R. A., Harrigan, R. J., Underwood, V. L., Gibbs, H. L., Smith, T. B., & Ruegg, K. (2018). Genomic signals of selection predict climate-driven population declines in a migratory bird. *Science*, *359*(6371), 83–86.

Blois, J. L., Williams, J. W., Fitzpatrick, M. C., Jackson, S. T., & Ferrier, S. (2013). Space can substitute for time in predicting climate-change effects on biodiversity. *Proceedings of the National Academy of Sciences*,*110*(23), 9374–9379.

Bower, A. D., St Clair, J. B., & Erickson, V. (2014). Generalized provisional seed zones for native plants. *Ecological Applications: A Publication of the Ecological Society of America*, *24*(5), 913–919.

Breiman, L. (2001). Random Forests. *Machine Learning*, *45*(1), 5–32.

Capblancq, T., Fitzpatrick, M. C., Bay, R. A., Exposito-Alonso, M., & Keller, S. R. (2020). Genomic Prediction of (Mal)Adaptation Across Current and Future Climatic Landscapes. *Annual Review of Ecology, Evolution, and Systematics*.

Chamberlain, S. (2017). rnoaa: "NOAA" Weather Data from R. R package version 0.7.0 (Version 0.7.0). Retrieved from https://CRAN.R-project.org/package=rnoaa

Chhatre, V. E., Fetter, K. C., Gougherty, A. V., Fitzpatrick, M. C., Soolanayakanahally, R. Y., Zalesny, R. S., & Keller, S. R. (2019). Climatic niche predicts the landscape structure of locally adaptive standing genetic variation (p. 817411). doi: 10.1101/817411

Dawson, T. P., Jackson, S. T., House, J. I., Prentice, I. C., & Mace, G. M. (2011). Beyond predictions: Biodiversity conservation in a changing climate.*Science*, *332*(6025), 53–58.

Ellis, N., Smith, S. J., & Pitcher, C. R. (2012). Gradient forests: calculating importance gradients on physical predictors. *Ecology*, *93*(1), 156–168.

Fetter, K. C., Nelson, D. M., & Keller, S. R. (2019). Trade-offs and selection conflicts in hybrid poplars indicate the stomatal ratio as an important trait regulating disease resistance. *BioRxiv*. Retrieved from https://www.biorxiv.org/content/10.1101/814046v1.abstract

Fitzpatrick, M. C., Blois, J. L., Williams, J. W., Nieto-Lugilde, D., Maguire, K. C., & Lorenz, D. J. (2018). How will climate novelty influence ecological forecasts? Using the Quaternary to assess future reliability. *Global Change Biology*, *24*(8), 3575–3586.

Posted on Authorea 28 Aug 2020 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.159863198.86187354 — This a preprint and has not been peer reviewed. Data may be preliminary.

Fitzpatrick, M. C., & Keller, S. R. (2015). Ecological genomics meets community-level modelling of biodiversity: mapping the genomic landscape of current and future environmental adaptation. *Ecology Letters*, *18*(1), 1–16.

Fitzpatrick, M. C., Keller, S. R., & Lotterhos, K. E. (2018). Comment on "Genomic signals of selection predict climate-driven population declines in a migratory bird." *Science*, *361*(6401). doi: 10.1126/science.aat7279

Frichot, E., Schoville, S. D., Bouchard, G., & Francois, O. (2013). Testing for associations between loci and environmental gradients using latent factor mixed models.*Molecular Biology and Evolution*, *30*(7), 1687–1699.

Gougherty, A. V., Keller, S. R., Chhatre, V. E., & Fitzpatrick, M. C. (2020). Future climate change promotes novel gene-climate associations in balsam poplar (Populus balsamifera L.), a forest tree species (p. 2020.02.28.961060). doi: 10.1101/2020.02.28.961060

Gugger, P. F., Liang, C. T., Sork, V. L., Hodgskiss, P., & Wright, J. W. (2018). Applying landscape genomic tools to forest management and restoration of Hawaiian koa (Acacia koa) in a changing environment. *Evolutionary Applications*, *11*(2), 231–242.

Guisan, A., Thuiller, W., & Zimmermann, N. E. (2017). *Habitat Suitability and Distribution Models: with Applications in R*. Cambridge University Press.

Gunther, T., & Coop, G. (2013). Robust identification of local adaptation from allele frequencies. *Genetics*, *195*(1), 205–220.

Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G., & Jarvis, A. (2005). Very high resolution interpolated climate surfaces for global land areas.*International Journal of Climatology*, *25*(15), 1965–1978.

Hudson, R. R. (2002). Generating samples under a Wright–Fisher neutral model of genetic variation. *Bioinformatics* , *18*(2), 337–338.

Ingvarsson, P. K., & Bernhardsson, C. (2020). Genome-wide signatures of environmental adaptation in European aspen (Populus tremula) under current and future climate conditions. *Evolutionary Applications*,*13*(1), 132–142.

Jia, K., Zhao, W., Maier, P. A., Hu, X., Jin, Y., Zhou, S., ... Mao, J. (2020). Landscape genomics predicts climate change-related genetic offset for the widespread Platycladus orientalis (Cupressaceae). *Evolutionary Applications*, *13*(4), 665–676.

Keller, S. R., Levsen, N., Ingvarsson, P. K., Olson, M. S., & Tiffin, P. (2011). Local selection across a latitudinal gradient shapes nucleotide diversity in Balsam Poplar, Populus balsamifera L. *Genetics*, *188*(4), 941–952.

Keller, S. R., Levsen, N., Olson, M. S., & Tiffin, P. (2012). Local adaptation in the flowering-time gene network of balsam poplar, Populus balsamifera L.*Molecular Biology and Evolution*, *29*(10), 3143–3152.

Keller, S. R., Olson, M. S., Silim, S., Schroeder, W., & Tiffin, P. (2010). Genomic diversity, population structure, and migration following rapid range expansion in the Balsam Poplar, Populus balsamifera. *Molecular Ecology*,*19*(6), 1212–1226.

Keller, S. R., Soolanayakanahally, R. Y., Guy, R. D., Silim, S. N., Olson, M. S., & Tiffin, P. (2011). Climate-driven local adaptation of ecophysiology and phenology in balsam poplar, Populus balsamifera L.(Salicaceae).*American Journal of Botany*, *98*(1), 99–108.

Landguth, E. L., & Cushman, S. A. (2010). cdpop: A spatially explicit cost distance population genetics program. *Molecular Ecology Resources*, *10*(1), 156–161.

Posted on Authorea 28 Aug 2020 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.159863198.86187354 — This a preprint and has not been peer reviewed. Data may be preliminary.

Little, E. L. (1971). Atlas of United States trees. Volume 1. Conifers and important hardwoods.*Miscellaneous Publications. United States Department of Agriculture*, (1146.).

Lotterhos, K. E., & Whitlock, M. C. (2014). Evaluation of demographic history and neutral parameterization on the performance of FST outlier tests.*Molecular Ecology*, *23*(9), 2178–2192.

Mahalanobis, P. C. (1936). On the generalized distance in statistics. *Proceedings of the National Institute of Sciences of India*, *2*, 49–55. New Delhi.

Mahony, C. R., MacLachlan, I. R., Lind, B. M., Yoder, J. B., Wang, T., & Aitken, S. N. (2020). Evaluating genomic data for management of local adaptation in a changing climate: A lodgepole pine case study. *Evolutionary Applications*,*13*(1), 116–131.

Martins, K., Gugger, P. F., Llanderal-Mendoza, J., Gonzalez-Rodriguez, A., Fitz-Gibbon, S. T., Zhao, J.-L., ... Sork, V. L. (2018). Landscape genomics provides evidence of climate-associated genetic variation in Mexican populations of Quercus rugosa. *Evolutionary Applications*, *11*(10), 1842–1858.

Matyas, C. (1996). Climatic adaptation of trees: rediscovering provenance tests. *Euphytica*, Vol. 92, pp. 45–54. doi: 10.1007/bf00022827

Meirmans, P. G., Godbout, J., Lamothe, M., Thompson, S. L., & Isabel, N. (2017). History rather than hybridization determines population structure and adaptation in Populus balsamifera. *Journal of Evolutionary Biology*, *30*(11), 2044–2058.

Naimi, B., Hamm, N. A. S., Groen, T. A., Skidmore, A. K., & Toxopeus, A. G. (2014). Where is positional uncertainty a problem for species distribution modelling?*Ecography*, *37*(2), 191–203.

Olson, M. S., Levsen, N., Soolanayakanahally, R. Y., Guy, R. D., Schroeder, W. R., Keller, S. R., & Tiffin, P. (2013). The adaptive potential of Populus balsamifera L. to phenology requirements in a warmer global climate. *Molecular Ecology*, *22*(5), 1214–1230.

Pecl, G. T., Araujo, M. B., Bell, J. D., Blanchard, J., Bonebrake, T. C., Chen, I.-C., ... Williams, S. E. (2017). Biodiversity redistribution under climate change: Impacts on ecosystems and human well-being. *Science*,*355*(6332). doi: 10.1126/science.aai9214

Peres-Neto, P. R., & Jackson, D. A. (2001). How well do multivariate data sets match? The advantages of a Procrustean superimposition approach over the Mantel test.*Oecologia*, *129*(2), 169–178.

Pike, C., Potter, K. M., Berrang, P., Crane, B., Baggs, J., Leites, L., & Luther, T. (2020). New Seed-Collection Zones for the Eastern United States: The Eastern Seed Zone Forum. *Journal of Forestry*, Vol. 118, pp. 444–451. doi: 10.1093/jofore/fvaa013

R Core Team. (2018). *R: A language and environment for statistical computing. R Foundation for Statistical Computing. Austria: Vienna.*

Ruegg, K., Bay, R. A., Anderson, E. C., Saracco, J. F., Harrigan, R. J., Whitfield, M., ... Smith, T. B. (2018). Ecological genomics predicts climate vulnerability in an endangered southwestern songbird. *Ecology Letters*, *21*(7), 1085–1096.

Sala, O. E., Chapin, F. S., Armesto, J. J., Berlow, E., Bloomfield, J., Dirzo, R., ... Wall, D. H. (2000). Biodiversity - Global biodiversity scenarios for the year 2100. *Science*, *287*, 1770–1774.

Savolainen, O., Lascoux, M., & Merila, J. (2013). Ecological genomics of local adaptation.*Nature Reviews. Genetics*, *14*(11), 807–820.

Soolanayakanahally, R. Y., Guy, R. D., Silim, S. N., Drewes, E. C., & Schroeder, W. R. (2009). Enhanced assimilation rate and water use efficiency with latitude through increased photosynthetic capacity and internal conductance in balsam poplar (Populus balsamifera L.). *Plant, Cell & Environment*, *32*(12), 1821–1832.

Posted on Authorea 28 Aug 2020 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.159863198.86187354 — This a preprint and has not been peer reviewed. Data may be preliminary.

Soolanayakanahally, R. Y., Guy, R. D., Silim, S. N., & Song, M. (2013). Timing of photoperiodic competency causes phenological mismatch in balsam poplar (Populus balsamifera L.). *Plant, Cell & Environment*, *36*(1), 116–127.

Thornton, P. E., Thornton, M. M., Mayer, B. W., Wilhelmi, N., Wei, Y., Devarakonda, R., & Cook, R. B. (2014). *Daymet: Daily Surface Weather Data on a 1-km Grid for North America, Version 2.* Oak Ridge National Laboratory (ORNL).

Urban, M. C. (2015). Accelerating extinction risk from climate change. *Science*,*348*(6234), 571–573.

Urban, M. C., Bocedi, G., Hendry, A. P., Mihoub, J.-B., Pe'er, G., Singer, A., ... Travis, J. M. J. (2016). Improving the forecast for biodiversity under climate change. *Science*, *353*(6304). doi: 10.1126/science.aad8466

Wang, T., Hamann, A., Yanchuk, A., O'Neill, G. A., & Aitken, S. N. (2006). Use of response functions in selecting lodgepole pine populations for future climates. *Global Change Biology*, Vol. 12, pp. 2404–2416. doi: 10.1111/j.1365-2486.2006.01271.x

Wang, T., O'Neill, G. A., & Aitken, S. N. (2010). Integrating environmental and genetic effects to predict responses of tree populations to climate. *Ecological Applications: A Publication of the Ecological Society of America*,*20*(1), 153–163.

Wuest, R. O., Zimmermann, N. E., Zurell, D., Alexander, J. M., Fritz, S. A., Hof, C., ... Others. (2020). Macroecology in the age of Big Data–Where to go from here? *Journal of Biogeography*, *47*(1), 1–12.