Phenotypic and genomic insights into population differentiation, introgression, and selection in *Quercus rubra* across a narrow but steep environmental gradient

María José Gómez Quijano[1], Briana Gross[2], and Julie Etterson[2]

[1]Queen's University
[2]University of Minnesota Duluth

July 7, 2023

Abstract

Adaptive differentiation in functional traits and their underlying loci can occur across a small geographic area if natural selection is stronger than the countervailing effects of gene flow and drift. We investigated this hypothesis in a long-lived, wind-pollinated species, *Quercus rubra,* across a fine spatial scale with a steep climate gradient. We examined phenotypic differentiation in a common garden study with eight populations sampled 0-160 km from the coast of Lake Superior. We estimated genomic differentiation for these and 22 additional populations from the same region, along with two populations of a congener, *Quercus ellipsoidalis*, using RAD-seq. We found a strong signal of population differentiation associated with climate in the common garden study, and differentiation was significantly associated with at least one climate factor for nine of ten measured traits. At the genomic level, we discovered widespread introgression from *Q. ellipsoidalis* into *Q. rubra* that increased with distance from the lake. Pairwise F_{ST} among *Q. rubra* populations was low, but both distance-based and environmental association analyses identified loci under selection, with one locus in common across all analyses *(CalS10/GSL8)*. This locus was associated with the precipitation of the driest month, a climate factor that was also significant in the common garden analyses. In sum, this study reveals signatures of selection at the phenotypic and genomic level consistent with climate adaptation, a pattern that is usually seen across a much broader geographic scale.

Hosted file

`MEC_Oak_Manuscript_Text_2023.doc` available at https://authorea.com/users/636738/articles/653464-phenotypic-and-genomic-insights-into-population-differentiation-introgression-and-selection-in-quercus-rubra-across-a-narrow-but-steep-environmental-gradient

Hosted file

`MEC_Manuscript_Tables_Quercusrubra.docx` available at https://authorea.com/users/636738/articles/653464-phenotypic-and-genomic-insights-into-population-differentiation-introgression-and-selection-in-quercus-rubra-across-a-narrow-but-steep-environmental-gradient

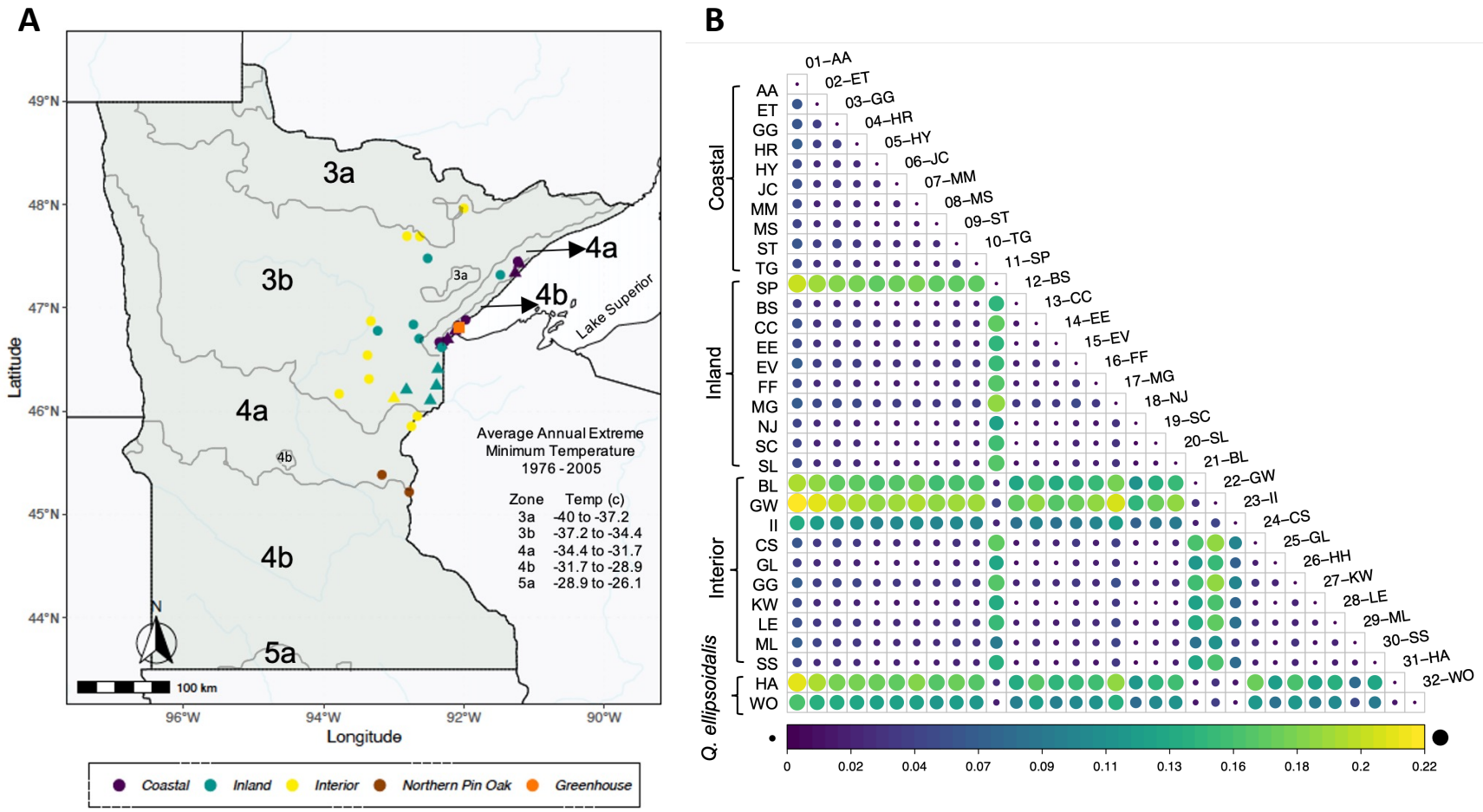

Figure 1. A) Map with USDA hardiness zones (USDA, 2019) of *Q. rubra* populations from coastal (0–16 km) (purple), inland (17–80 km) (teal), interior (81–160 km) (yellow) regions and two *Q. ellipsoidalis* populations (red). Circles represent sites where only leaf tissue for the molecular marker study was collected, while triangles represent sites where both leaf tissue and acorns were collected for the greenhouse study. The orange circle represents greenhouse seedling samples included in the molecular analysis. **B)** Pairwise F_{ST} matrix for *Q. rubra* and *Q. ellipsoidalis* populations based on dataset 1. Circle size and color are correlated and represent the magnitude of the pairwise F_{ST} value for each set of populations. Because the F_{ST} values are small, the matrix is scaled from a minimum of 0 to a maximum of 0.22.

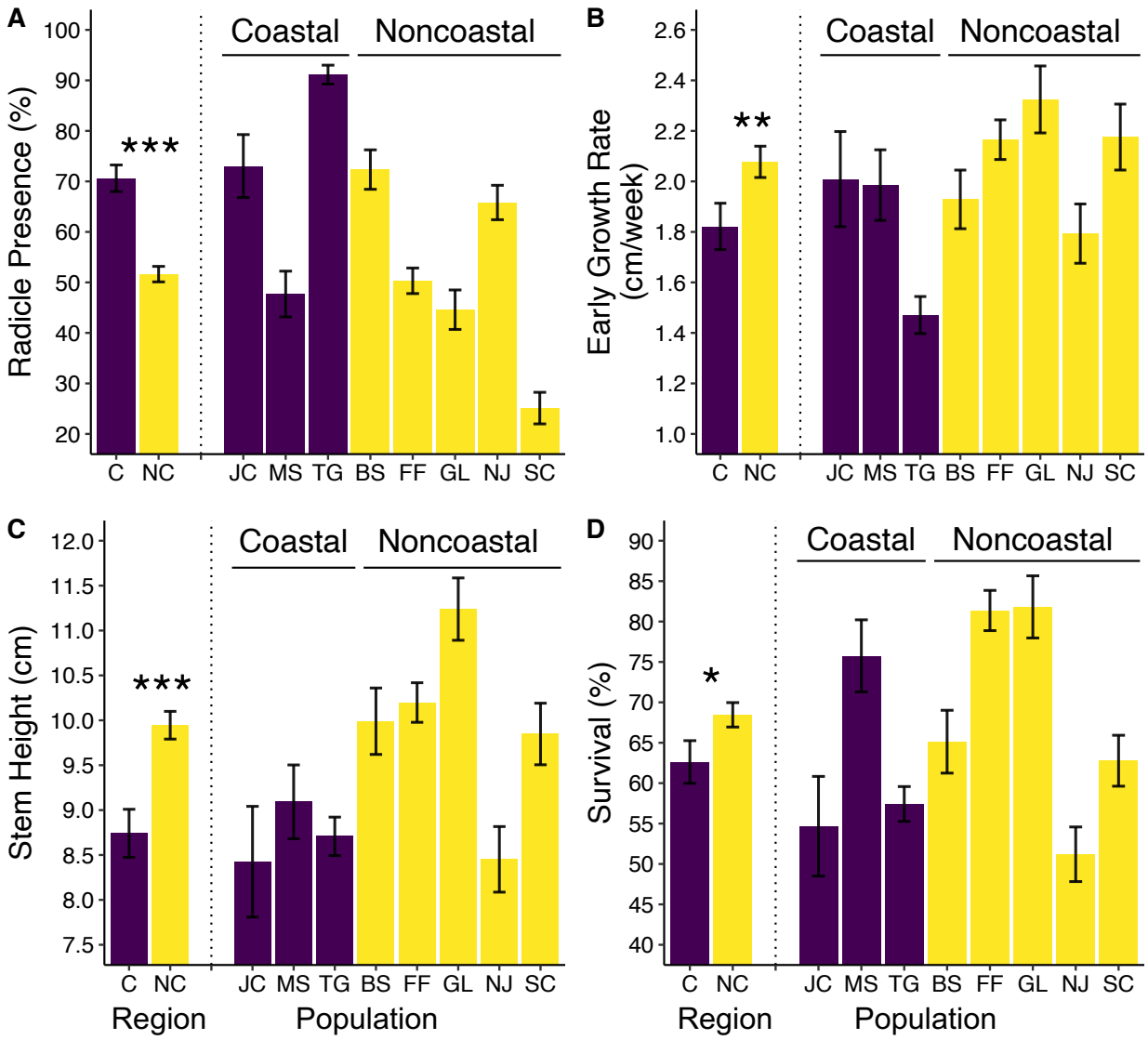

Figure 2. Bar graphs showing estimated means and SE of response factors in trait analyses in which region (coastal or noncoastal) and population within region predictors were significant. **A)** Presence of a radicle at the time of planting, **B)** Growth rate based on early height measurements, **C)** Stem height measured at week four and, **D)** Survival. *** indicates a *P*-value < 0.0001, ** indicates a *P*-value = 0.005, and * indicates a *P*-value = 0.03. The y-axis is scaled for each panel to better show the difference between groups.

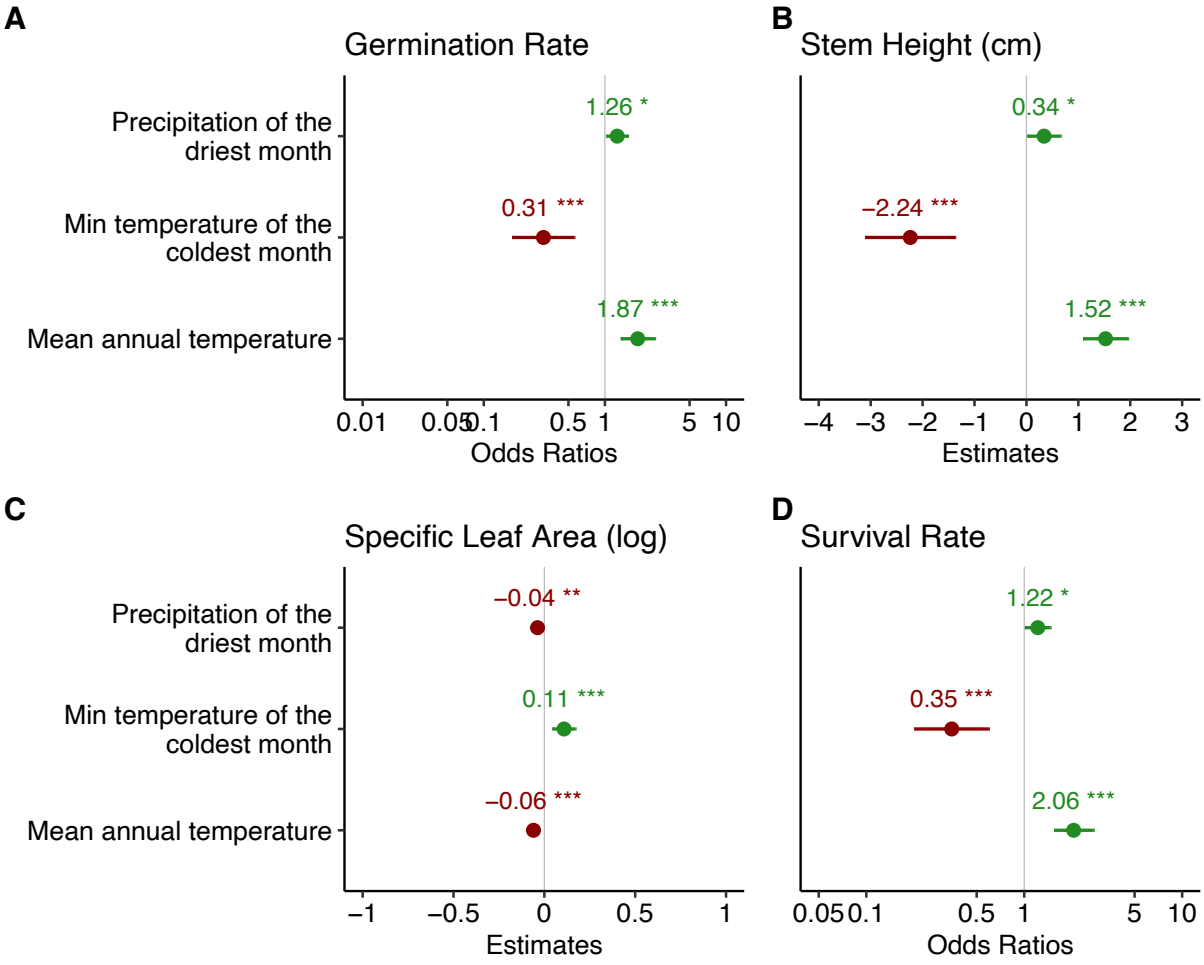

Figure 3. Forest plots of estimates and odds ratios for quantitative traits in the climate model where all climate variables had a significant effect. Effects sizes are only shown for the three climate variables (precipitation of the driest month, minimum temperature of the coolest month and mean annual temperature). Positive effects are in green and negative effects in red. Significance of the factor in the climate model (Table 1b) is represented by asterisk. *** indicates a *P*-value < 0.0001, ** indicates a *P*-value = 0.005, and * indicates a *P*-value = 0.03.

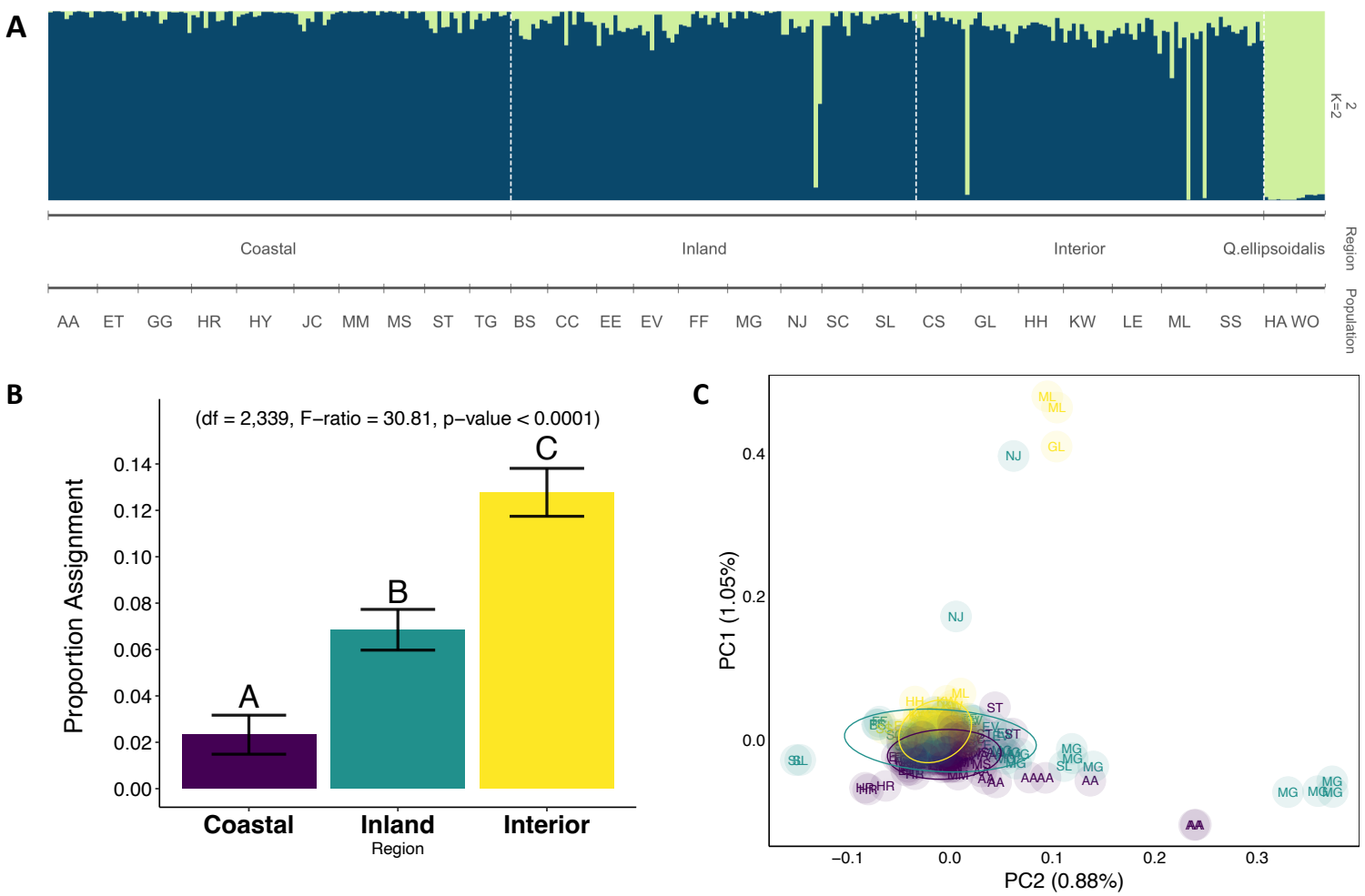

Figure 4. Population structure and individual assignment of field-collected *Q. rubra* and *Q. ellipsoidalis*. **A)** STRUCTURE plot of *Q. rubra* and *Q. ellipsoidalis* at $K = 2$ based on dataset 1 with misclassified *Q. rubra* populations removed. **B)** Proportion assignment of field-collected *Q. rubra* to *Q. ellipsoidalis* by region based on dataset 1; different letters indicate significant differences between regions ($P < 0.05$). **C)** Principal component analysis field-collected *Q. rubra* populations based on dataset 2. Each region is represented by a color (as shown in panel B) and each population is represented by a two-letter code (Supplementary Table 1). Ellipses represent the 95% confidence intervals for each region.

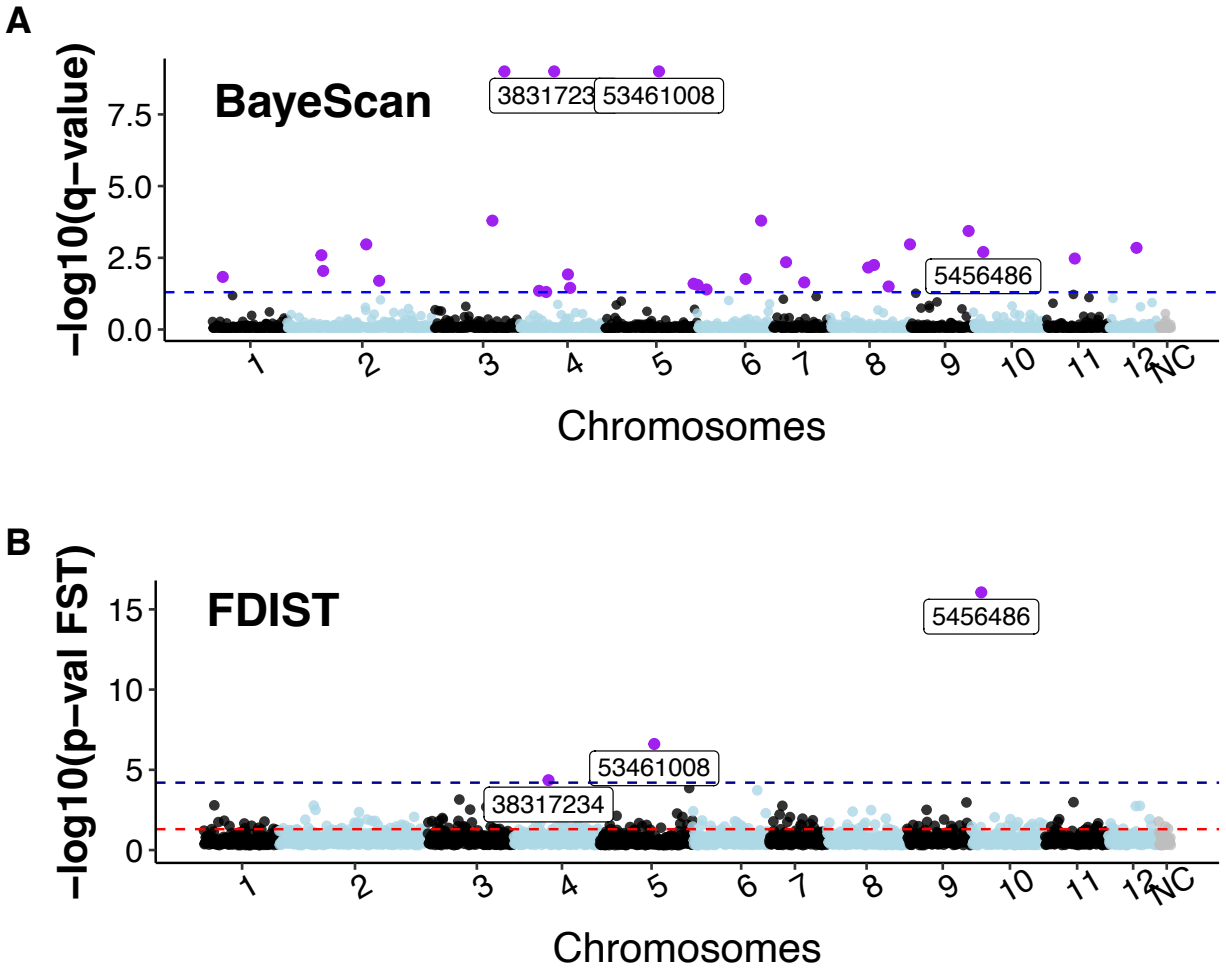

Figure 5. Manhattan plot of outlier loci from two distance-based approaches: **A)** BayeScan and **B)** FDIST using dataset 2. Each point in the graph represents the –log of the F_{ST} value for each locus within its respective chromosome based on the *Q. lobata* reference genome. NC are non-characterized chromosomal regions. Points above the dotted blue line are loci that have been identified as outliers after a 0.05% FDR correction. The red dotted line in the FDIST approach represents an uncorrected *P*-value of 0.05.

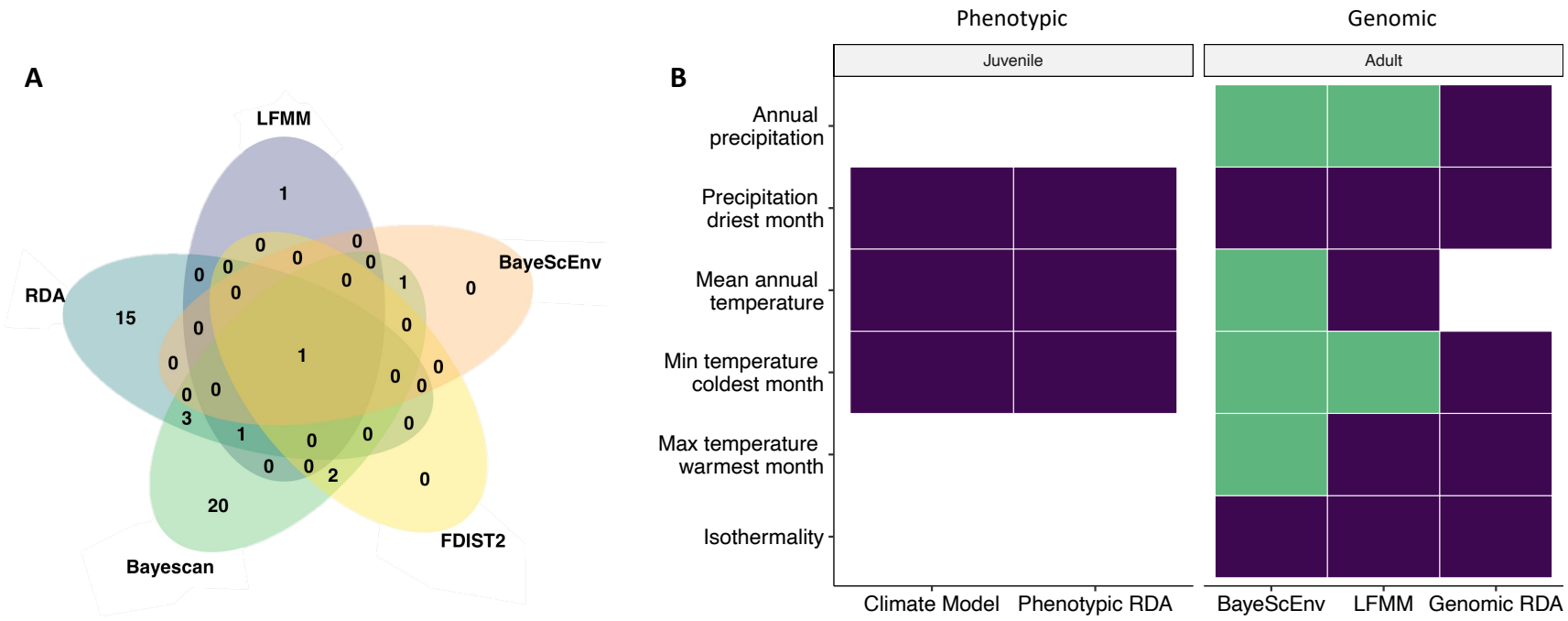

Figure 6. **A)** Venn diagram of loci identified as outliers by distance-based methods (FDIST and BayeScan) and loci associated with environmental variables (BayeScenv, LFM and RDA) using field-collected samples from dataset 2. Only one locus was found to be under selection and associated with an environmental variable between the five different approaches. **B)** Worldclim climatic variables used in quantitative genetic and population genomic environmental analyses of field-collected samples from dataset 2. Variable inclusion was based on individual model stepwise selection and VIF values. Purple squares represent climatic variables that had significant associations in the analyses, green squares represent variables that were included in the analysis but were not found to be associated with any traits or loci, and empty squares represent variables that we not included for a specific analysis.