Genetic differentiation across a steep and narrow environmental gradient: Quantitative genetic and genomic insights into Lake Superior populations of *Quercus rubra*

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

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

August 24, 2024

Abstract

Adaptive differentiation of traits and underlying loci can occur at a small geographic scale if natural selection is stronger than gene flow and drift. We investigated this hypothesis using coupled quantitative genetic and genomic approaches for a wind-pollinated tree species, *Quercus rubra*, along the steep, narrow gradient of the Lake Superior coast that encompasses four USDA Hardiness Zones within 100 km. For the quantitative genetic component, we examined phenotypic differentiation among eight populations in a common garden. For the genomic component, we quantified genetic differentiation for 26 populations using RAD-seq. Because hybridization with the congener *Quercus ellipsoidalis* has been documented elsewhere, we also included two populations of this species for comparison. In the common garden study, we found a strong signal of population differentiation that was significantly associated with at least one climate factor for nine of ten measured traits. In contrast, we found no evidence of genomic differentiation among populations based on *FST* or other measures. However, distance-based, and genotype-environment association analyses identified loci showing the signature of selection, with one locus in common across five analyses. This locus was associated with the minimum temperature of the coldest month, a climate factor was also significant in the common garden analyses. In addition, we documented introgression from *Q. ellipsoidalis* into *Q. rubra*, with rates of introgression correlated with the climate gradient. In sum, this study reveals signatures of selection at the quantitative trait and genomic level consistent with climate adaptation occurring at a small geographic scale.

Hosted file

`MEC-23-0687_Resubmission_Manuscript_Text.docx` available at https://authorea.com/users/636738/articles/1217823-genetic-differentiation-across-a-steep-and-narrow-environmental-gradient-quantitative-genetic-and-genomic-insights-into-lake-superior-populations-of-quercus-rubra

Hosted file

`MEC-23-0687_Resubmission_TrackChanges_Manuscript_Text.docx` available at https://authorea.com/users/636738/articles/1217823-genetic-differentiation-across-a-steep-and-narrow-environmental-gradient-quantitative-genetic-and-genomic-insights-into-lake-superior-populations-of-quercus-rubra

Hosted file

`MEC-23-0687_Resubmission_TrackChanges_Manuscript_Figures_and_Tables.docx` available at https://authorea.com/users/636738/articles/1217823-genetic-differentiation-across-a-steep-and-narrow-environmental-gradient-quantitative-genetic-and-genomic-insights-into-lake-superior-populations-of-quercus-rubra

Manuscript Figures and Tables

Table 1. A) Analysis of variance test statistics from the region-based model, used to analyze genetic differences based on quantitative traits of *Q. rubra* populations from two regions (coastal and noncoastal) grown in a common environment. **B)** Analysis of variance test statistics from the climate-based model, used to analyze genetic differences based on quantitative traits of *Q. rubra* populations using temperature and precipitation factors as predictors.

A. Regional Model	**Region**			**Population (Reg.)**			**Seed Mass**			**Radicle Presence**			**Block**		
	df	***F / χ^2***	***P***	***df***	***F / χ^2***	***P***	***df***	***F / χ^2***	***P***	***df***	***F / χ^2***	***P***	***df***	***F / χ^2***	***P***
Seed Mass	1, 1423	1.87	0.17	6, 1423	15.87	**<0.0001**	-	-	-	1, 1424	20.56	**<0.0001**	-	-	-
Radicle Presence	1, 1423	46.19	**<0.0001**	6, 1423	168.34	**<0.0001**	1, 1423	23.33	**<0.0001**	-	-	-	-	-	-
Germination	1, 1423	1.46	0.23	6, 1423	81.19	**<0.0001**	1, 1423	12.12	**<0.0001**	1, 1424	292.66	**<0.0001**	3	43.87	**<0.0001**
Germination date	1, 1107	0.08	0.76	6, 1107	2.67	**0.01**	1, 1107	6.47	**0.01**	1, 1107	113.8	**<0.0001**	3	19.52	**<0.0001**
Early growth rate	1, 712	7.87	**0.005**	6, 712	5.36	**<0.0001**	1, 712	4.02	**0.04**	1, 712	3.8	**0.05**	3	8.23	**<0.0001**
Late growth rate	1, 997	0.01	0.92	6, 997	0.67	0.67	1, 997	3.92	**0.04**	1, 997	0.26	0.61	3	24.97	**<0.0001**
Stem Height	1, 1089	17.15	**<0.0001**	6, 1089	5.49	**<0.0001**	1, 1089	48.22	**<0.0001**	1, 1089	2.91	0.09	3	12.47	**<0.0001**
Stem Diameter	1, 1091	3.10	0.08	6, 1091	5.08	**<0.0001**	1, 1091	215.31	**<0.0001**	1, 1091	21.74	**<0.0001**	3	127.35	**<0.0001**
Leaf Number	1, 1078	1.52	0.22	6, 1078	3.62	**0.001**	1, 1078	14.91	**0.001**	1, 1078	0.23	0.63	3	4.88	**0.002**
SLA (log)	1, 1011	0..70	0.4	6, 1011	6.51	**<0.0001**	1, 1011	26.25	**<0.0001**	1, 1011	10.86	**0.001**	3	10.6	**<0.0001**
Survival	1, 1423	4.75	**0.03**	6, 1423	73.22	**<0.0001**	1, 1423	1.41	0.23	1, 1423	221.17	**<0.0001**	3	40.41	**<0.0001**

B. Climate Model	**Mean Temp**			**Min Temp**			**Precip of Driest Month**			**Seed Mass**			**Radicle Presence**			**Block**		
	df	***F / χ^2***	***P***	***df***	***F / χ^2***	***P***	***df***	***F / χ^2***	***P***	***df***	***F / χ^2***	***P***	***df***	***F / χ^2***	***P***	***df***	***F / χ^2***	***P***
Seed Mass	1, 1427	20.11	**<0.0001**	1, 1427	7.78	**0.0053**	1, 1427	2.75	0.097	-	-	-	1, 1427	20.18	**<0.0001**	-	-	-
Radicle Presence	1, 1427	192.6	**<0.0001**	1, 1427	0.0326	0.86	1, 1427	40.648	**<0.0001**	1, 1427	22.76	**<0.0001**	-	-	-	-	-	-
Germination	1, 1427	14.64	**0.0001**	1, 1427	14.64	**0.0001**	1, 1427	4.95	**0.026**	1, 1427	5.99	**0.014**	1, 1427	269.15	**<0.0001**	3	49.48	**<0.0001**
Germination date	1, 1111	5.35	**0.021**	1, 1111	2.19	0.13	1, 1111	0.33	0.56	1, 1427	7.04	**0.008**	1, 1111	115.85	**<0.0001**	3	19.37	**<0.0001**
SLA	1, 1015	13.02	**0.0003**	1, 1015	11.15	**0.0009**	1, 1015	9.48	**0.0021**	1, 1015	23.61	**<0.0001**	1, 1015	6.18	**0.013**	3	10.06	**<0.0001**
Early growth rate	1, 716	48.92	**<0.0001**	1, 716	1.36	0.24	1, 716	4.04	0.044	1, 716	2.81	0.094	1, 716	5.26	**0.022**	3	7.72	**<0.0001**
Late growth rate	1, 1001	0.089	0.75	1, 1001	0.0014	0.97	1, 1001	0.76	0.38	1, 1001	4.19	**0.041**	1, 1001	0.19	0.65	3	24.87	**<0.0001**
Stem Height	1, 1081	47.26	**<0.0001**	1, 1081	25.65	**<0.0001**	1, 1081	4.31	**0.038**	1, 1081	45.65	**<0.0001**	1, 1081	1.64	0.19	3	11.71	**<0.0001**
Leaf Number	1, 1082	16.77	**<0.0001**	1, 1082	0.47	0.49	1, 1082	0.13	0.71	1, 1082	17.71	**<0.0001**	1, 1082	0.24	0.62	3	4.91	**0.0021**
Stem Diameter	1, 1095	2.85	0.091	1, 1095	15.8	**<0.0001**	1, 1095	6.59	**0.01**	1, 1095	219.32	**<0.0001**	1, 1095	18.13	**<0.0001**	3	124.53	**<0.0001**
Survival	1, 1427	25.57	**<0.0001**	1, 1427	14.56	**0.0001**	1, 1427	4.26	**0.038**	1, 1427	0.14	0.7	1, 1427	200.39	**<0.0001**	3	42.49	**<0.0001**

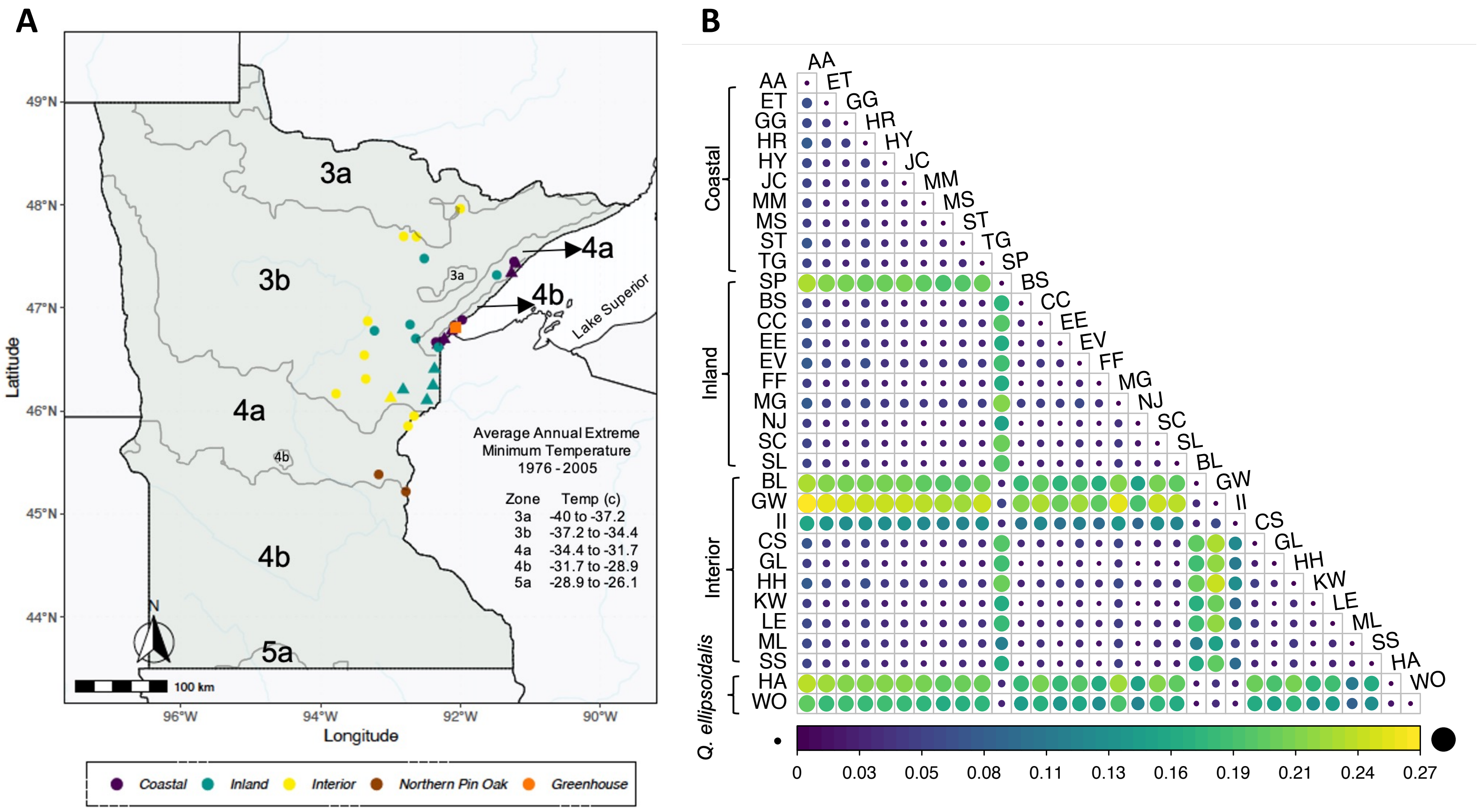

Figure 1. A) Map with USDA hardiness zones (USDA, 2019) of *Q. rubra* populations from coastal (0–16 km) (purple), inland (17–80 km) (teal), interior (81–160 km) (yellow) regions and two *Q. ellipsoidalis* populations (red). Circles represent sites where only leaf tissue for the genomic study was collected, while triangles represent sites where both leaf tissue (genomic study) and acorns (quantitative genetic study) were collected. The orange circle represents the location of the greenhouse for the quantitative genetic study. **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 0.27.

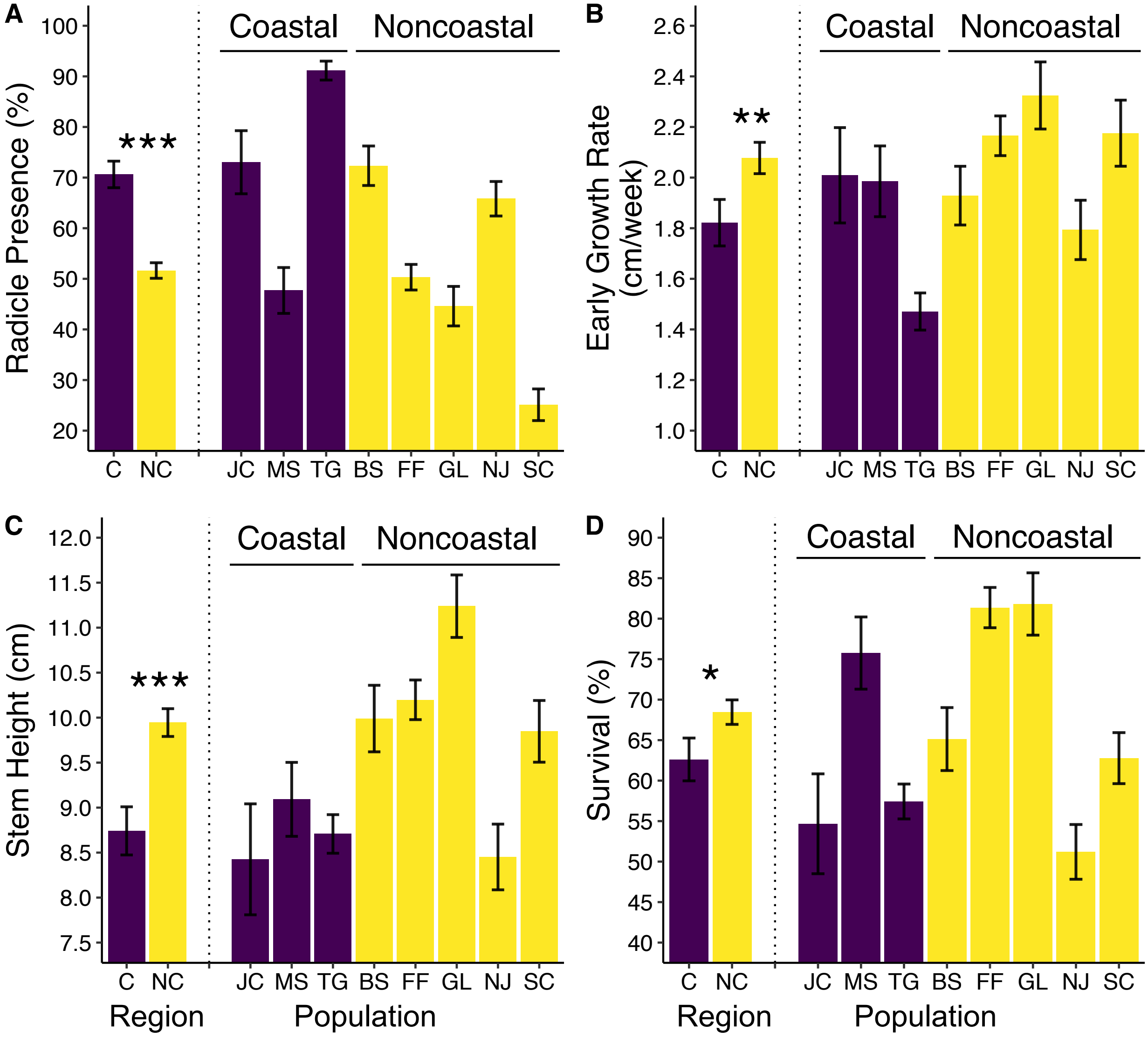

Figure 2. Bar graphs showing estimated means and SE of quantitative trait response factors in which region (coastal or noncoastal) and population within region predictors were significant. **A)** Presence of a radicle at the time of planting, **B)** Growth rate based on early height measurements, **C)** 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 differences between groups.

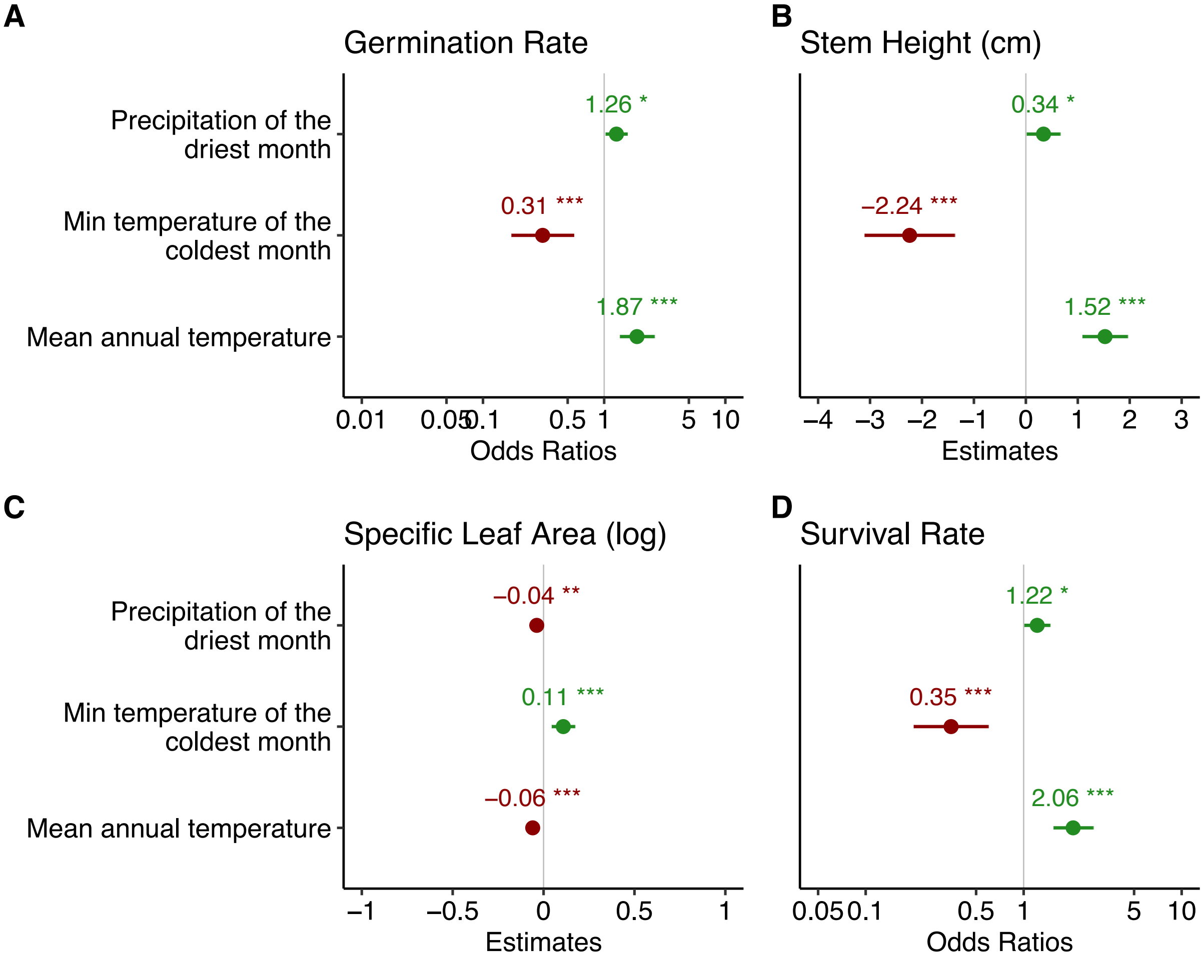

Figure 3. Forest plots of estimates and odds ratios for quantitative traits in the climate model where all climate variables had a significant effect. Effects sizes are shown for the three climate variables (precipitation of the driest month, minimum temperature of the coolest month and mean annual temperature). 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

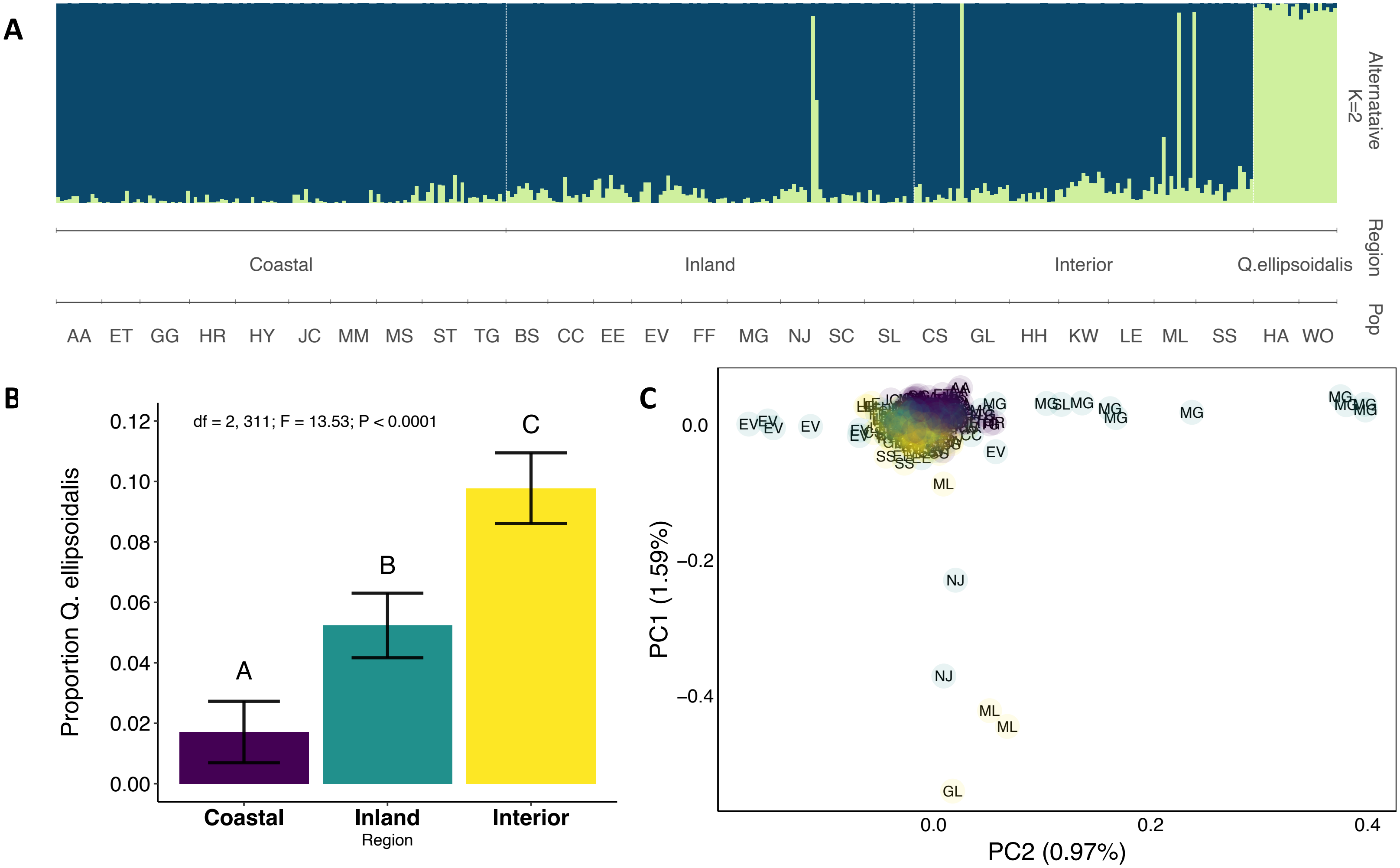

Figure 4. Population structure and individual assignment of field-collected *Q. rubra* and *Q. ellipsoidalis*. **A)** STRUCTURE plot of *Q. rubra* and *Q. ellipsoidalis* at $K = 2$ based on dataset 1 with misclassified *Q. rubra* populations removed. **B)** Proportion assignment of field-collected *Q. 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).

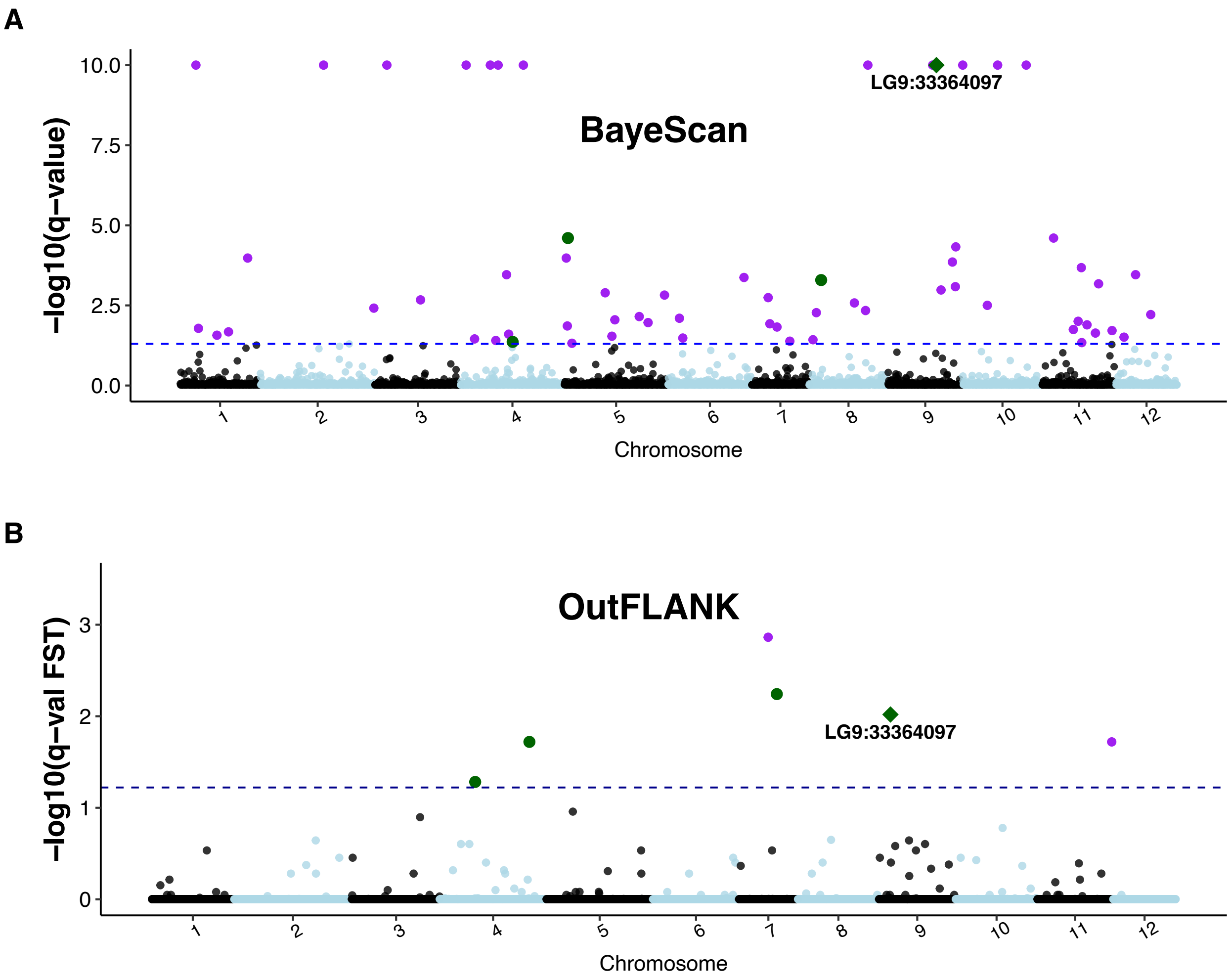

Figure 5. Manhattan plot of outlier loci from two distance-based approaches: **A)** BayeScan and **B)** OutFLANK using dataset 2. Each point in the graph represents the –log of the F_{ST} q-value for each locus within its respective chromosome based on the *Q. rubra* reference genome. Points above the dotted blue line are loci that have been identified as outliers after a 0.05% FDR correction. Green points represent overlapping outlier loci for both approaches. Green diamond indicates the outlier locus 33364907 in chromosome 9 that was identified as an outlier for all the genomic signatures of selection approaches.

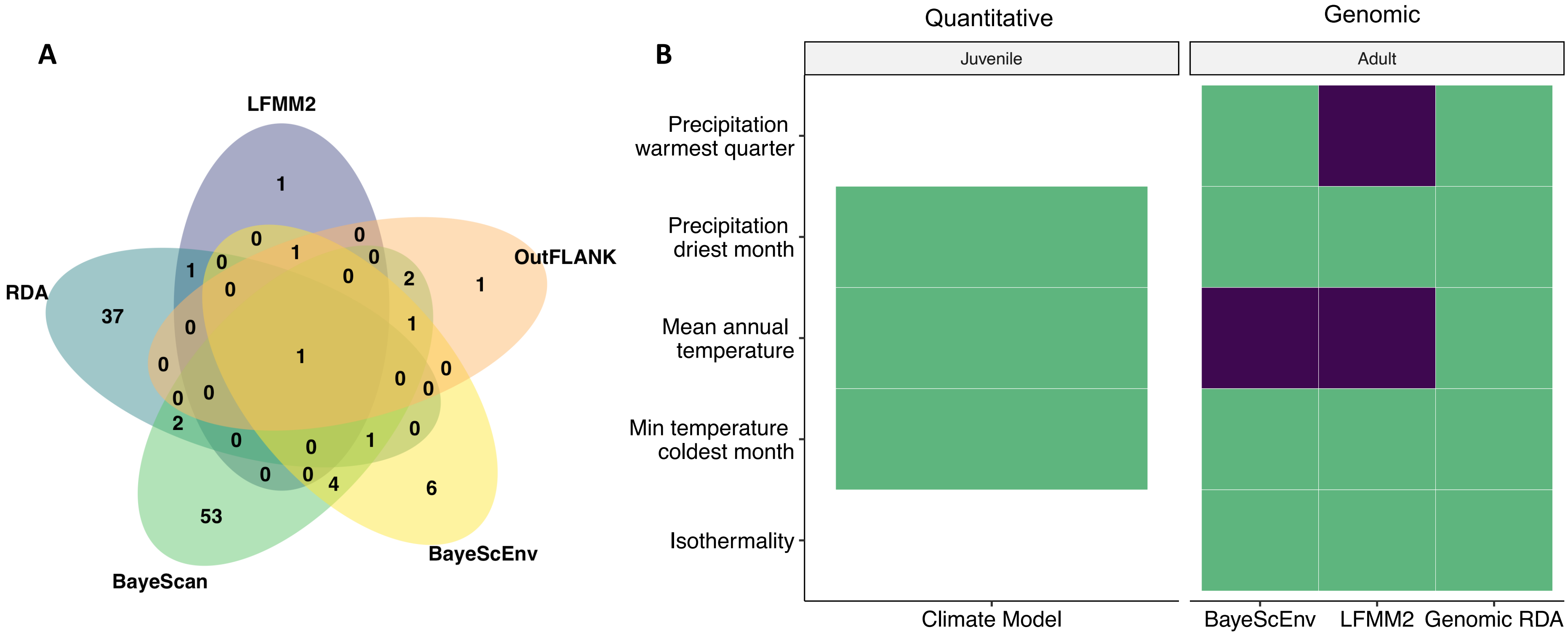

Figure 6. **A)** Venn diagram of loci identified as outliers by distance-based methods (OutFLANK and BayeScan) and loci associated with environmental variables (BayeScEnv, LFMM2 and RDA) using field-collected samples from dataset 2. Only one locus was found to be under selection and associated with an 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. Green squares represent climatic variables that had significant associations in the analyses, purple 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.