**Supporting Information for:**

**Population genomics and Lagrangian modeling shed light on dispersal events in the Mediterranean endemic *Ericaria zosteroides* (= *Cystoseira zosteroide*s) (Fucales)**

**LEGENDS OF TABLES**

**Table S1.** Quality criteria applied to build the neutral and outlier dataset of *Ericaria zosteroides*. For each filtering step, the number of SNPs passing the filter (SNPs count), as well as the number of individuals (individuals count), are reported. The catalog of 314 538 loci was initially built using the Stacks v.2.4 bioinformatics pipeline, with the draft genome of *E. zosteroides* used as a reference.

**Table S2.** Pairwise correlations of the initial dataset of 11 environmental predictors used in the Distance-based RDA (db-RDA) framework.

**Table S3.** Comparisons in the distribution of the size of the main axis between populations using the Kolmorov-Sirmov pairwise test. Non-significant P-values are in bold type (p>0.05).

**Table S4.** Comparisons in genetic diversity (expected heterozygosity among SNPs) among populations using the Pairwise Wilcoxon Test. P-values were adjusted for multiple testing with the Bonferroni method. Non-significant P-values are in bold type (p>0.001).

**LEGENDS OF FIGURES**

**Figure S1.** variation in cross-entropy according to the number K of ancestral populations in sNMF.

**Figure S2.** distribution of the size of the main axis (cm) in the sampled populations; the sizes were distributed into fourteen discrete classes.

**Figure S3**. Additional axis from PC3 to PC8 (PC3: 12%; PC4: 11%; PC5: 10%; PC6: 9%; PC7: 9%; PC8: 7%) of the PCA conducted on the neutral set of 10 755 SNPs and 229 individuals.

**Figure S4**. Assignment tests over the different scales of the study. For each population (Sampled population, y-axis), the percentage of individuals assigned (Inferred population, x-axis) to their respective sources (green circle) or other putative sources (red circle). The size of circles is proportional to the percentage of assignment, (A) assignment tests restricted to the top 100 ranked SNPs (decreasing overall FST), (B) assignment tests for all neutral markers (10 755 SNPs).

**Figure S5.** Assignment success (%) at the population level and over populations (Overall) as a function of the number of SNPs ranked based on decreasing FST values.

**APPENDIX**

**Appendix S1.** Species delimitation analyses

**Appendix S2.** Genetic structure on the basis of outlier SNPs

**Table S1**

|  |  |  |
| --- | --- | --- |
| **Filtering SNP candidates** | **SNP counts** | **Individual counts** |
| Initial catalog of 314 538 loci | 721 374 | 271 |
| **Locus genotyped** |  |  |
| > 80% of individuals per population | 184 723 | 271 |
| > 80% of study sites |
| **SNP threshold** |
| heterozygosity observed < 50% |
| **Per-individual quality control filters** | 21 606 | 248 |
| missing genotypes < 25% |
| heterozygosity observed < 3 SD from the mean |
| **Per-SNP quality control filters** |
| missing genotypes < 5% |
| Minor Allele Frequency (MAF) > 1% |
| **Individual relatedness filter** | 21 606 | 229 |
| KING-robust estimates = 0.354 |
| **Linkage Disequilibrium (LD) between SNPs** | 11 067 | 229 |
| LD threshold for clumping (on MAF); r2 > 0.3 |
| **Genome scan filter** |  | 229 |
| Putatively neutral | 10 755 |
| Putatively under divergent selection | 12 |

**Table S2**

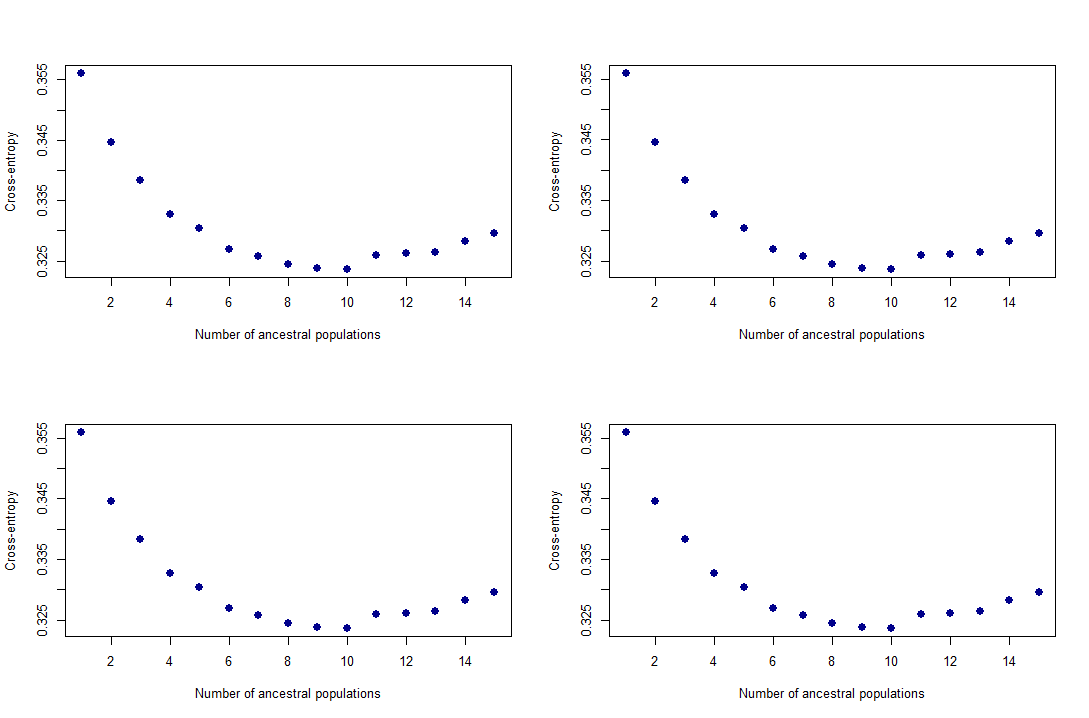
|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Colonne1** | **AEM1** | **AEM2** | **AEM3** | **AEM4** | **AEM5** | **dbMEM1** | **Density** | **Min\_T** | **Max\_T** | **Mean\_T** | **Var\_T** |
| **AEM1** | 1 | -0.04 | -0.01 | -0.02 | 0.02 | -0.61 | -0.14 | -0.8 | 0.57 | 0.16 | 0.23 |
| **AEM2** | -0.04 | 1 | 0.05 | 0.01 | -0.01 | 0.24 | -0.6 | 0.5 | -0.49 | -0.6 | -0.62 |
| **AEM3** | -0.01 | 0.05 | 1 | 0.01 | -0.01 | 0.08 | 0.01 | 0.35 | -0.62 | -0.78 | -0.78 |
| **AEM4** | -0.02 | 0.01 | 0.01 | 1 | -0.02 | -0.39 | 0.2 | 0.04 | 0.28 | 0.22 | 0.07 |
| **AEM5** | 0.02 | -0.01 | -0.01 | -0.02 | 1 | 0.64 | 0.75 | 0.14 | -0.03 | -0.01 | 0.01 |
| **dbMEM1** | -0.61 | 0.24 | 0.08 | -0.39 | 0.64 | 1 | 0.36 | 0.71 | -0.64 | -0.38 | -0.35 |
| **Density** | -0.14 | -0.6 | 0.01 | 0.2 | 0.75 | 0.36 | 1 | -0.03 | 0.21 | 0.34 | 0.33 |
| **Min\_T** | -0.8 | 0.5 | 0.35 | 0.04 | 0.14 | 0.71 | -0.03 | 1 | -0.88 | -0.65 | -0.71 |
| **Max\_T** | 0.57 | -0.49 | -0.62 | 0.28 | -0.03 | -0.64 | 0.21 | -0.88 | 1 | 0.89 | 0.9 |
| **Mean\_T** | 0.16 | -0.6 | -0.78 | 0.22 | -0.01 | -0.38 | 0.34 | -0.65 | 0.89 | 1 | 0.99 |
| **Var\_T** | 0.23 | -0.62 | -0.78 | 0.07 | 0.01 | -0.35 | 0.33 | -0.71 | 0.9 | 0.99 | 1 |

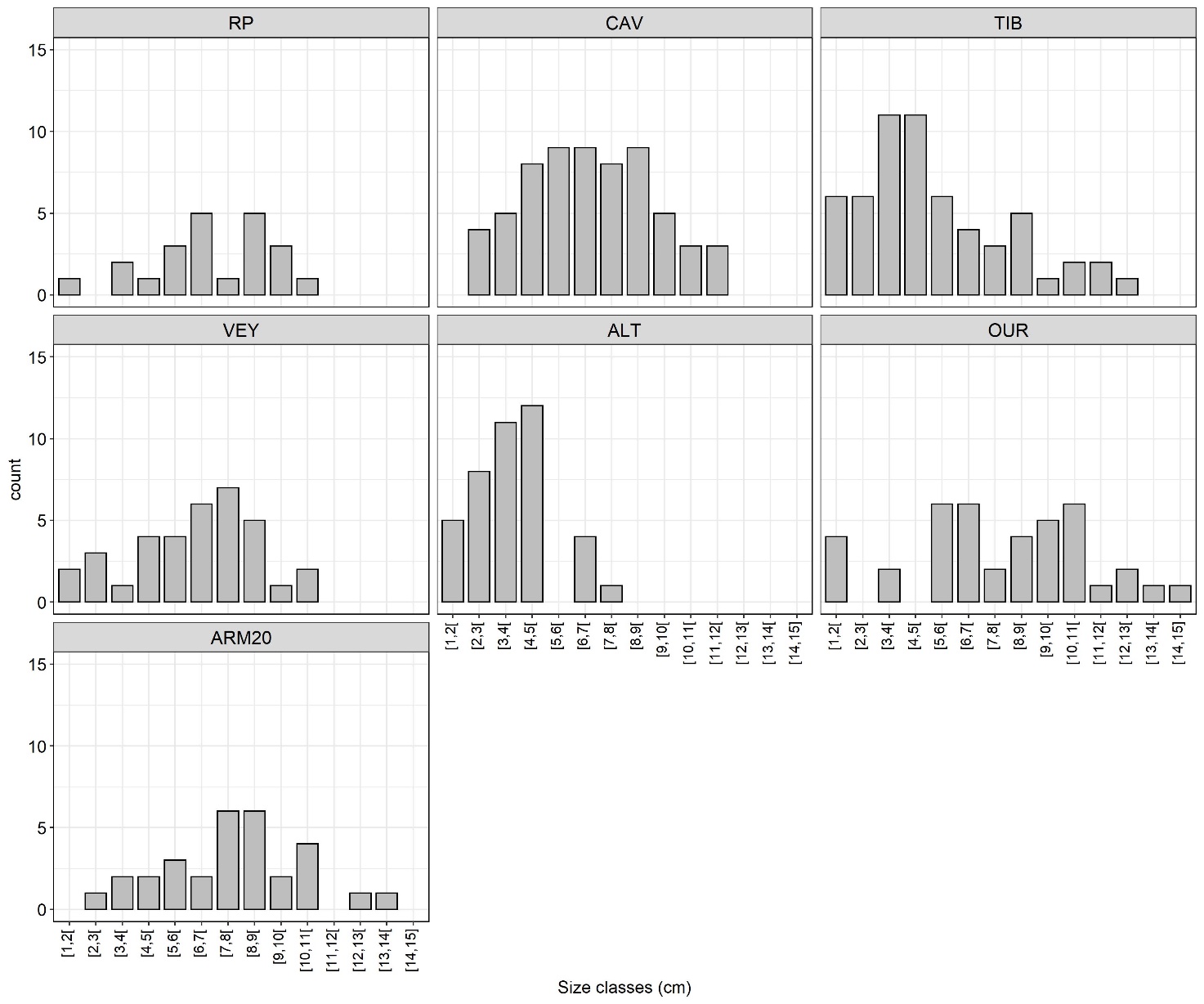
|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | CAV | VEY | ALT | RP | TIB | OUR |
| VEY | **9.6e-01** |  |  |  |  |  |
| ALT | 1.4e-08 | 3.5e-06 |  |  |  |  |
| RP | **1.0e+00** | **7.7e-01** | 1.9e-06 |  |  |  |
| TIB | 3.5e-04 | 2.9e-02 | 3.3e-02 | 1.1e-02 |  |  |
| OUR | **1.7e-01** | 5.0e-02 | 9.5e-10 | 3.7e-01 | 1.5e-04 |  |
| ARM20 | **2.7e-01** | **3.0e-01** | 4.9e-08 | 6.2e-01 | 1.6e-03 | **7.8e-01** |

**Table S3**

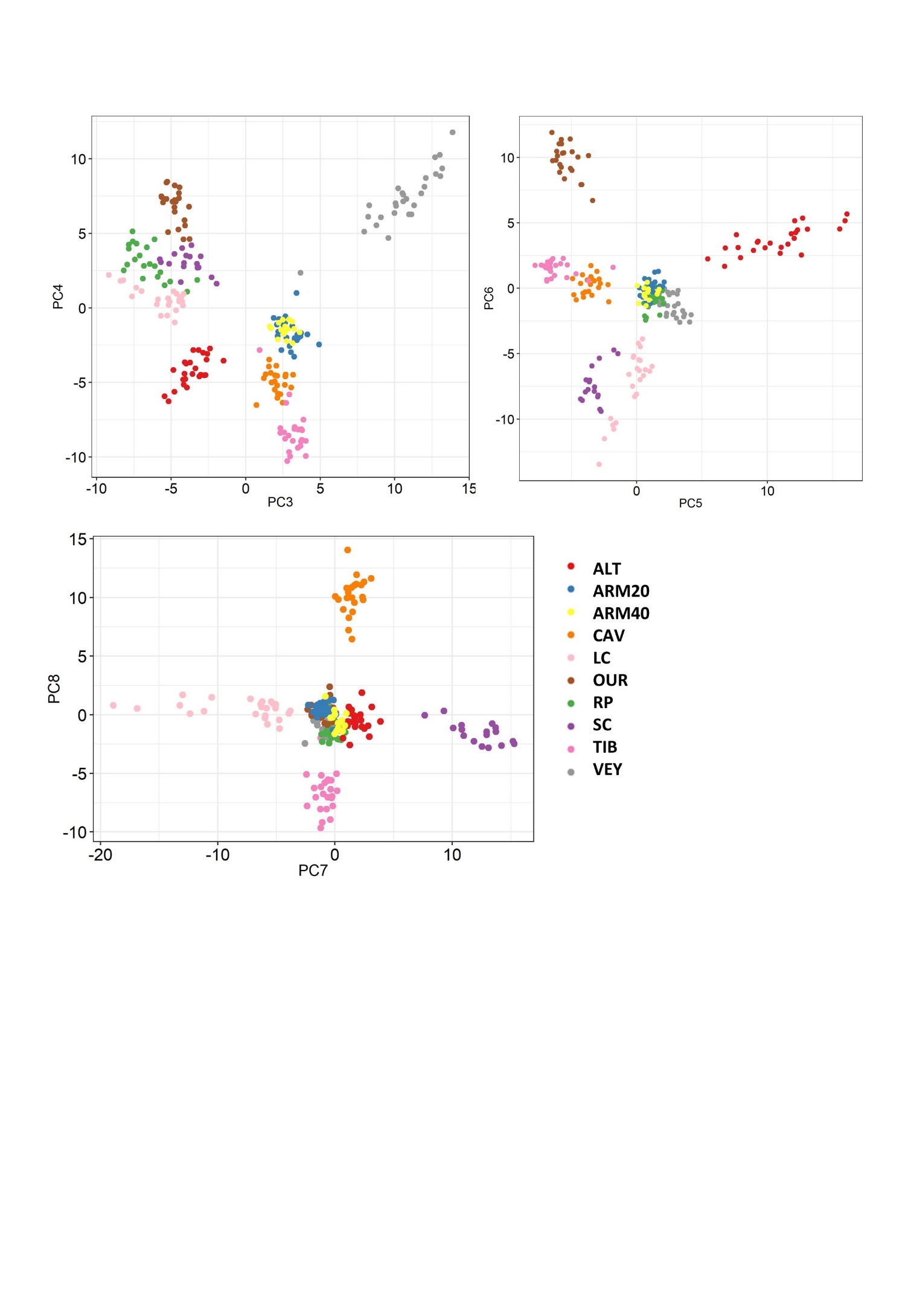
**Table S4**

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **ALT** | **ARM20** | **ARM40** | **CAV** | **LC** | **OUR** | **RP** | **SC** | **TIB** |
| **ARM20** | 5.7e-44 |  |  |  |  |  |  |  |  |
| **ARM40** | 6.1e-54 | **3.3e-02** |  |  |  |  |  |  |  |
| **CAV** | **8.6e-01** | 6.5e-48 | 2.6e-62 |  |  |  |  |  |  |
| **LC** | 2.8e-14 | 1.4e-93 | 1.5e-102 | **1.0e+00** |  |  |  |  |  |
| **OUR** | **3.6e-03** | 2.5e-54 | 3.5e-70 | **1.0e+00** | **1.0e+00** |  |  |  |  |
| **RP** | 5.6e-178 | 4.7e-64 | 1.0e-36 | 1.4e-186 | 2.8e-259 | 5.4e-195 |  |  |  |
| **SC** | 1.0e-06 | 3.9e-69 | 1.5e-75 | 2.0e-05 | **1.0e+00** | **2.0e-02** | 1.9e-262 |  |  |
| **TIB** | **1.0e+00** | 5.3e-28 | 1.9e-38 | 2.4e-08 | 9.3e-13 | 6.8e-09 | 2.1e-147 | 1.9e-16 |  |
| **VEY** | **3.5e-03** | 8.8e-30 | 1.3e-41 | 7.9e-04 | 2.4e-26 | 4.0e-07 | 2.0e-152 | 3.1e-14 | **1.0e+00** |

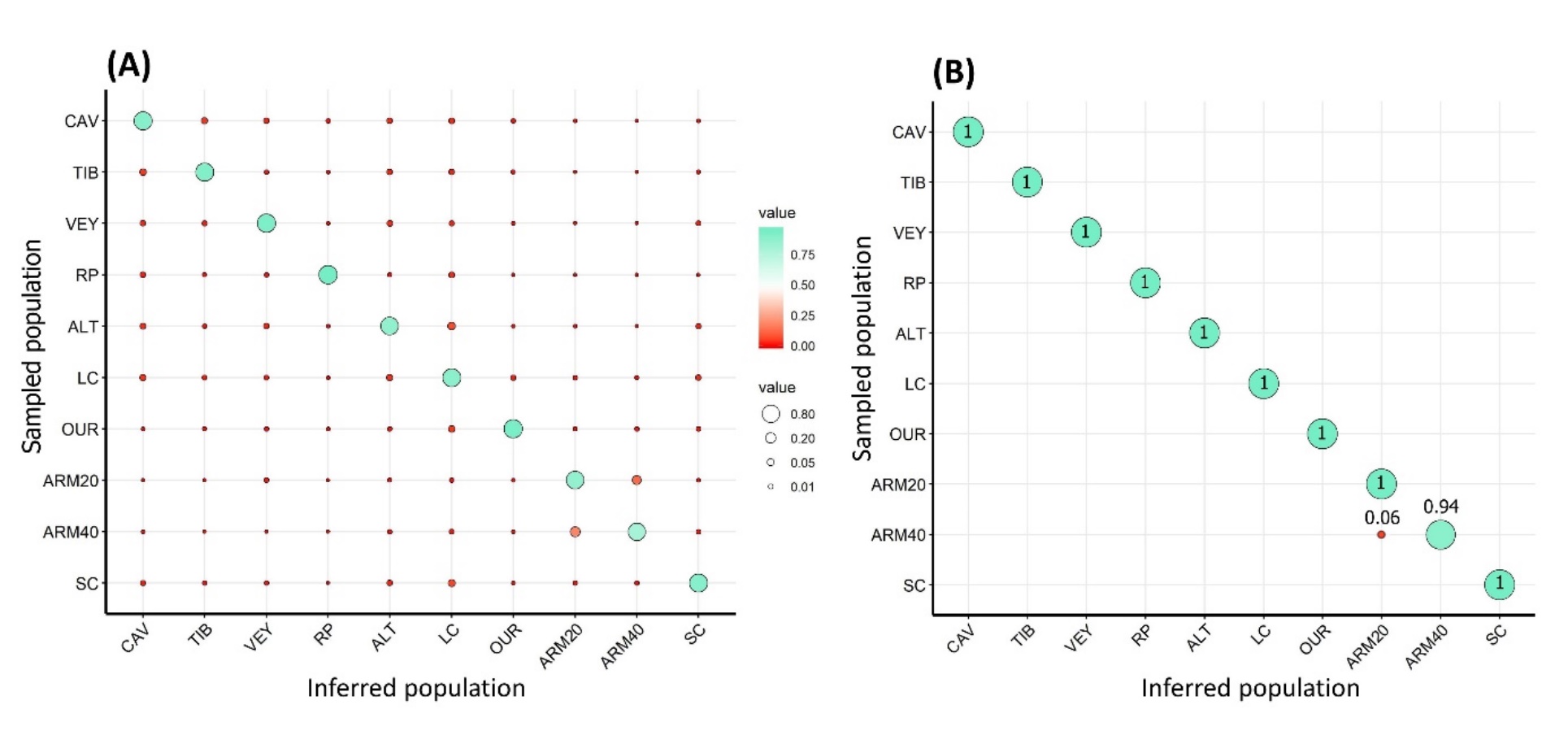
**Figure S1**

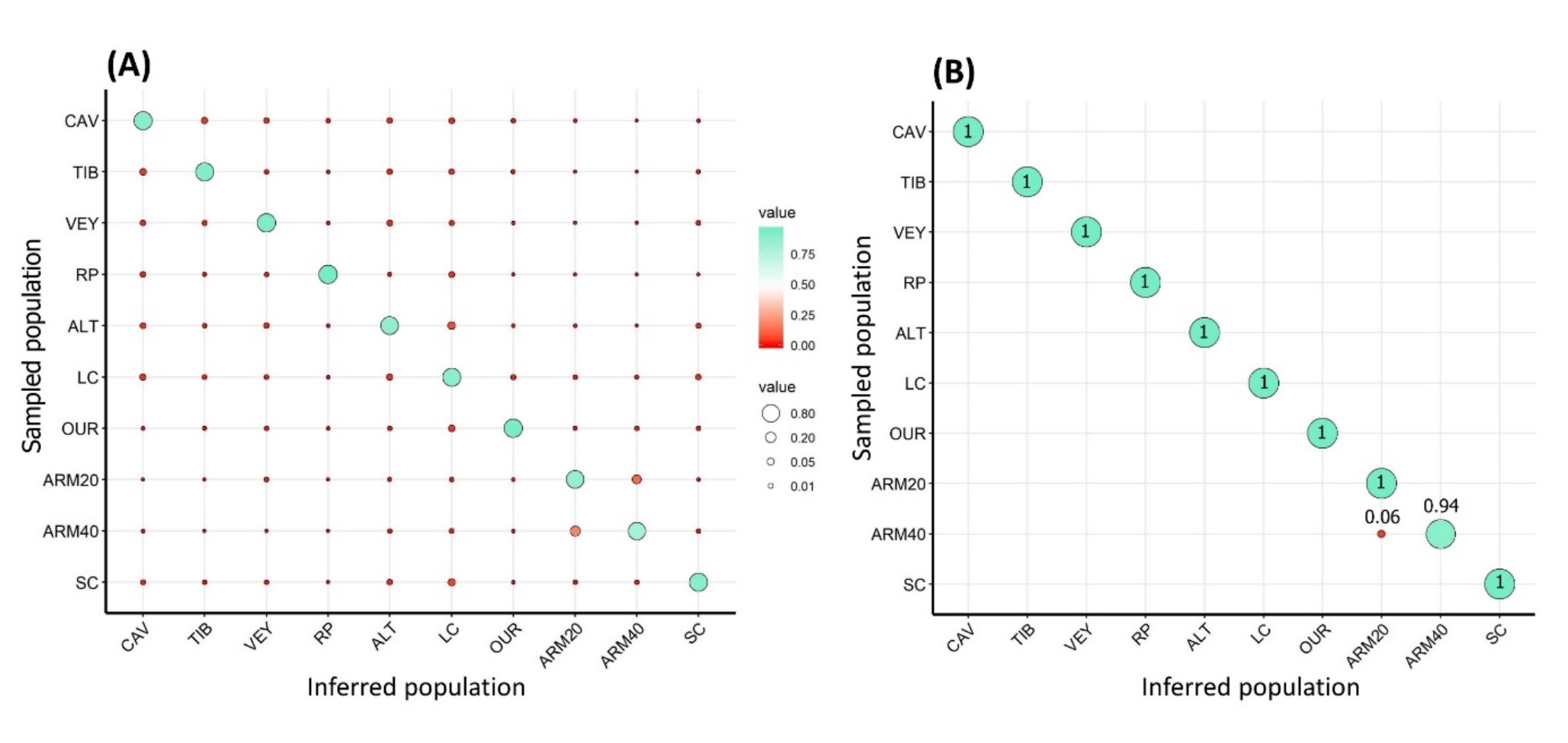
 **Figure S2**

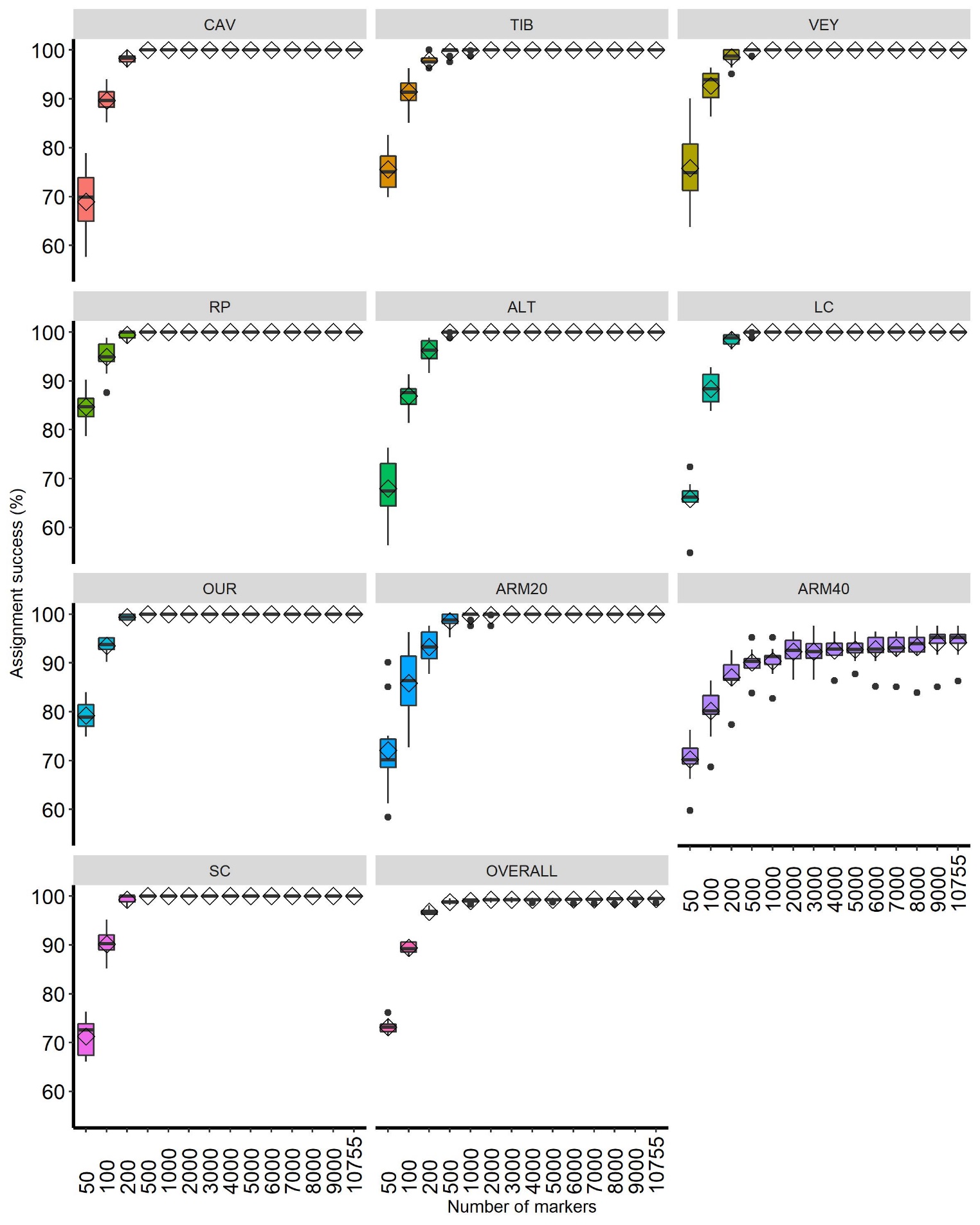
**Figure S3**



**Figure S4**

****

****

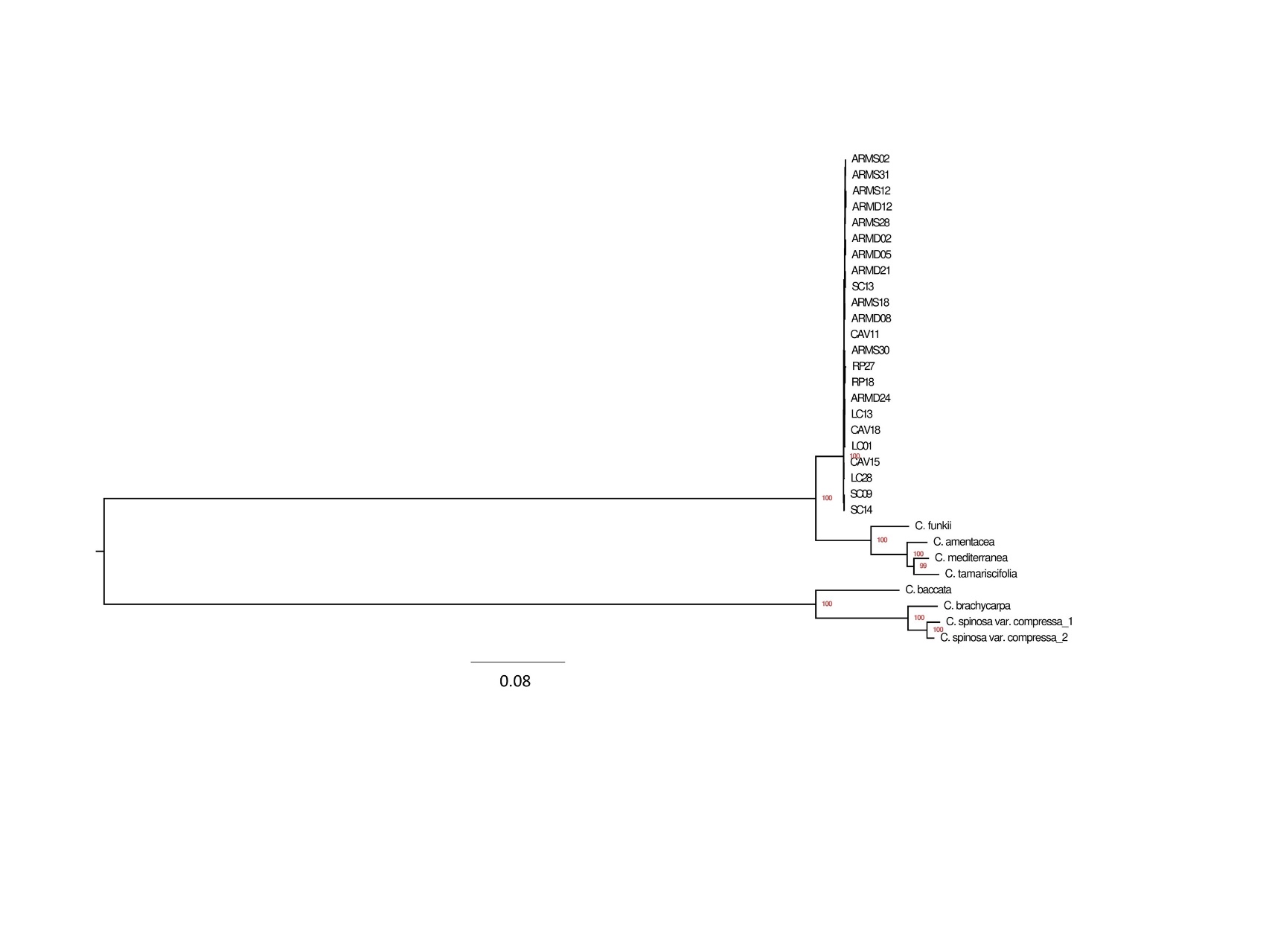
**Figure S5**

**Appendix S1**

**Species delimitation analyses**

We built an additional dataset consisting of twenty-three individuals of *E. zosteroides* from several geographical locations and eight individuals belonging to other species of *Cystoseira sensu lato* including *C. spinosa var. compressa*, *C. funkii*, *C. amentacea*, *C. tamariscifolia*, *C. mediterranea*, *C. baccata,* and *C. brachycarpa.* The DNA extraction, dd-RAD-seq library preparation, sequencing and sequence quality filtering were executed in an identical way to *E. zosteroides* (see main text for details). The reads were mapped to the *E. zosteroides* genome with the BWA-mem algorithm in BWA v.0.7.17 (Li and Durbin, 2009), and the catalog of loci was built with the reference mode of the Stacks v2.4 pipeline (Rochette et al., 2019). Using the population module of Stacks, we selected loci exclusively shared by all samples (--min-samples-overall = 1). Following this step, 11 010 SNPs (25 587 loci) were filtered from the raw catalog of 569 242 SNPs (166 493 loci). Finally, each variable site was exported in PHYLIP format for downstream analyses. Phylogenetic maximum-likelihood (ML) analyses were performed with IQ-TREE v.1.6.1 (Nguyen et al., 2015) using ModelFinder (Kalyaanamoorthy et al., 2017) and 1000 ultrafast bootstraps (Hoang et al., 2018). The best-fit model of nucleotide substitution according to Bayesian information criterion (BIC) was TVM+F.

The percentage of average mapped reads per individual ranged from 50.10% in *C. brachycarpa* to 89.67% in *C. mediterranea*. We reported that individuals of *E. zosteroides* belonged to the same clade of the phylogenetic tree, whatever the initial geographical origin, while the eight other species of *Cystoseira sensu lato* belonged to evolutionally distinct clades (Figure S1.1). All of these clades were supported by high ultrafast bootstrap values (99 or 100%). The genome-wide nuclear distance, as well as the mitochondrial genetic distance (e.g. Draisma et al., 2010; Bruno de Sousa et al., 2019; Pérez et al., 2020), were large enough among *E. zosteroides*, *C. spinosa var. compressa, and C. funkii* to discriminate species.

**Figure S1.1.** ML phylogeny inferred using 11 010 SNPs in IQ-TREE. The tree was rooted using *Cystoseira baccata* *sensu lato* as an outgroup. The values in red indicate the percentages of ultrafast bootstrap resampling (n = 1000); the values below 90% are not shown.

REFERENCES

Bruno de Sousa, C., Cox, C.J., Brito, L., Pavão, M.M., Pereira, H., Ferreira, A., Ginja, C., Campino, L., Bermejo, R., Parente, M., and Varela, J. (2019). Improved phylogeny of brown algae *Cystoseira* (Fucales) from the Atlantic-Mediterranean region based on mitochondrial sequences. *PLOS ONE*, *14*(1), e0210143. doi: 10.1371/journal.pone.0210143

Draisma, S.G.A., Ballesteros, E., Rousseau, F., and Thibaut, T. (2010). Dna sequence data demonstrate the polyphyly of the genus *Cystoseira* and other Sargassaceae genera (Phaeophyceae). *J. Phycol.* *46*(6), 1329–1345. doi: 10.1111/j.1529-8817.2010.00891.x

Hoang, D.T., Chernomor, O., von Haeseler, A., Minh, B.Q., and Vinh, L.S. (2018). UFBoot2: Improving the ultrafast bootstrap approximation. *Mol. Biol. Evol. 35*(2), 518–522. doi: 10.1093/molbev/msx281

Kalyaanamoorthy, S., Minh, B. Q., Wong, T.K.F., von Haeseler, A., and Jermiin, L S. (2017). ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nat. Methods*. *14*(6), 587–589. doi: 10.1038/nmeth.4285

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics.* *25*(14), 1754–1760. doi: 10.1093/bioinformatics/btp324

Nguyen, L.T., Schmidt, H.A., von Haeseler, A., and Minh, B.Q. (2015). IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol.Evol. 32*(1), 268–274. doi: 10.1093/molbev/msu300

Pérez, A.B.J., Fernández, M.T., López-Moya, F., Berbegal, L.A., and López-Llorca, L.V. (2020). Phylogeny of Cystoseira sensu lato from SE Spain suggests a complex colonization of the Mediterranean. *J. Mar. Sci. Eng*. 8, 961. doi: 10.3390/jmse8120961

Rochette, N.C., Rivera‐Colón, A.G., and Catchen, J.M. (2019). Stacks 2: Analytical methods for paired-end sequencing improve RADseq-based population genomics. *Mol. Ecol.* 28(21), 4737–4754. doi: 10.1111/mec.15253

**Appendix S2**

**Genetic structure on the basis of outlier SNPs**

This appendix presents details of the response of outlier SNPs to geographical and environmental predictors of genetic structure. We present the results of the distance-based RDA (db-RDA) according to the set of the 12 outlier SNPs reported in the main text, then we describe the spatial distribution of allelic frequencies for these outlier SNPs and we discuss our findings. According to the forward selection procedure, the minimum seawater temperature predictor (Min\_T) explained 10% of the variance (R2adj = 0.10, P < 0.001) in genetic structure while oceanographic connectivity with AEM-4, AEM-5, AEM-2, AEM-1 predictors explained 39% (R2adj = 0.39, P < 0.001) of the variance (Figure S2.1). Individuals seemed to be mainly differentiated along the first significant axis of the db-RDA (74.85% of the total variance), enabling the distinction of the deepest population (LC, depth = 45 m) from other ones that correspond to depths ranging from 25 to 36 m. The positive coordinate of individuals on this axis was driven by the minimum seawater temperature and was mainly related to the population of LC. Conversely, the negative coordinate of the axis was driven by oceanographic connectivity predictors such as AEM1 and AEM4, without specific differentiation of a population.

The allelic frequencies of the 12 putatively outlier SNPs were different between the deepest water population of LC and the other ones in the Bay of Marseille (Figure S2.2), which is congruent with the results of the db-RDA. Among these outlier SNPs, five of them (i.e. 112036\_64, 136434\_44, 1521\_81, 211417\_31, and 57060\_94) were particularly interesting since one of the two alleles tended to be absent in the Bay of Marseille and overrepresented in the population of LC. These allele frequencies ranged from 0.34 to 0.72 in the population of LC and from 0.03 to 0.06 within the Bay of Marseille (Figure S1.2). The alleles of the population of LC are not specific to depth, in that they are observed at high frequencies in populations of Eastern-Provence and Corsica, whatever their depth. However, three outlier loci attracted much attention (i.e. 1521\_81, 57060\_94, and 211417\_31) since major alleles reported at high frequency in the deepest water population were relatively less frequent in shallow populations. Also, although these alleles are too infrequent at the isolated sites of Armoire, they appeared to be more preserved in the deepest ARM40 than the shallowest ARM20 populations (Figure S1.2).

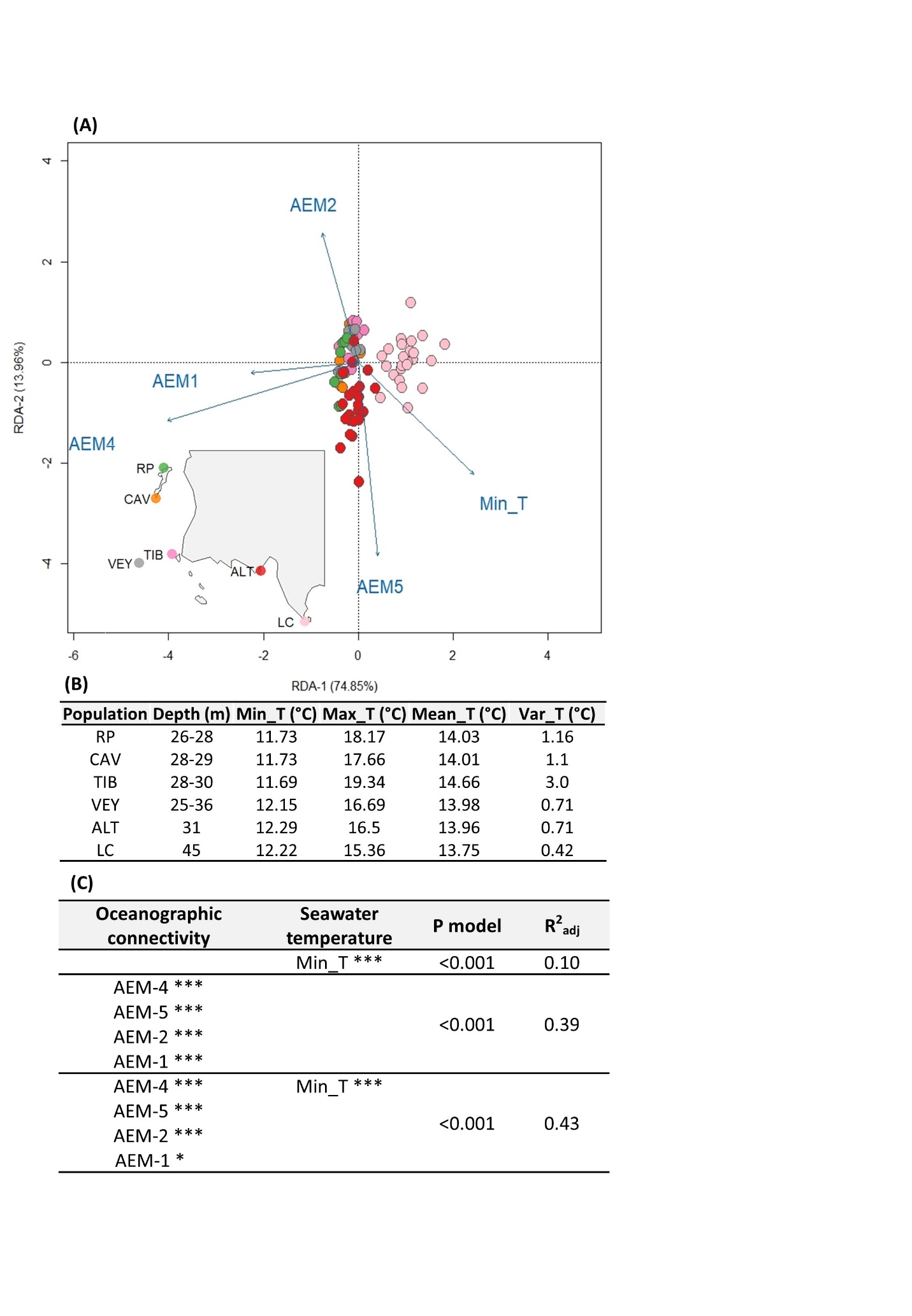
We reported from the neutral set of SNPs that the isolated bank populations of Armoire had one of the lowest levels of genetic variation reported in our study. This is discussed in the main manuscript as the effects of restricted gene flow from other regions and strong intensity of genetic drift, which potentially limit the strength of divergent selection with depth.

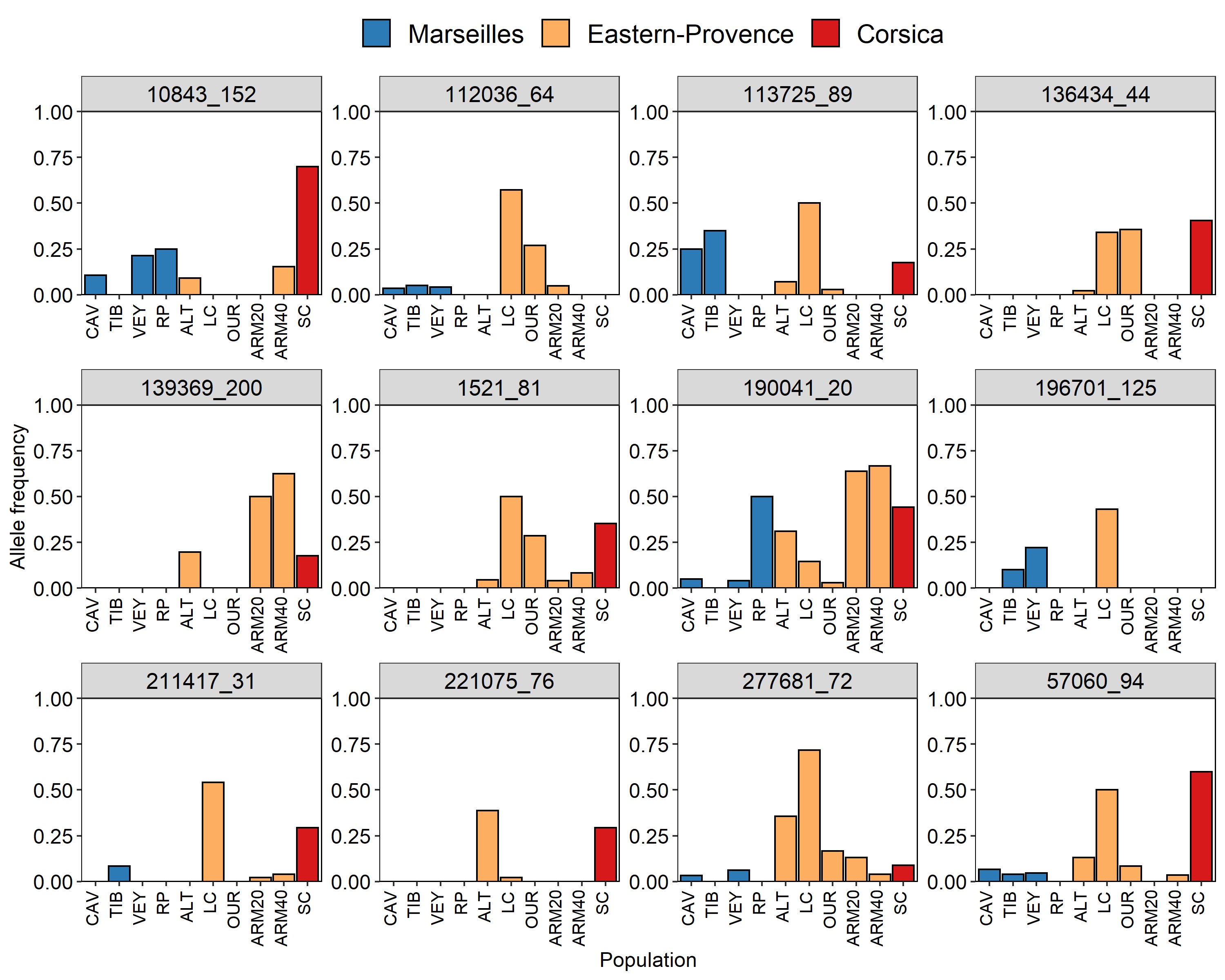
Furthermore, our results indicated that directional gene flow from shallow to deep populations occurred at Armoire. Gene frequency change by natural selection can be slowed by homogenizing gene flow, imposing a limit of local adaptation (Lenormand, 2002). Hence, the lack of shift in allele frequency from shallow to deep individuals at Armoire for the outlier loci could be the consequence of gene flow, which would potentially limit local adaptation along this depth gradient. Further studies will be necessary to conclude whether the pattern of differentiation by depth reported in the population of LC can be interpreted as a result of local adaptation. Nevertheless, the fitness of *Ericaria zosteroides* was significantly impacted by increasing seawater temperature from 16°C to 24°C (Capdevila et al., 2019), suggesting that seawater temperature may involve changes in metabolic traits that may be under selection depending on the local thermal regime.

**Figure S2.1** Distance-based RDA (db-RDA) using the set of 12 putatively outlier SNPs. (A) db-RDA biplot showing the significant predictors. The dots represent individuals and are colored according to their populations. (B) seawater temperature predictors averaged over seven years of simulations. (C) significant predictors retained following the forward selection procedure using both oceanographic connectivity and seawater temperature predictors simultaneously and independently from each other. Geographical distance was not retained as a significant predictor.

**Figure S2.2** Spatial distribution of the minor allele frequencies (MAF) of the 12 putatively outlier SNPs at the scales of the bay of Marseille (blue), Eastern-Provence (yellow), and Corsica (red).

**Figure S2.1**



**Figure S2.2**

REFERENCES

Capdevila, P., Hereu, B., Salguero‐Gómez, R., Rovira, G., Medrano, A., Cebrian, E., Garrabou, J., Kersting, D.K., and Linares, C. (2019). Warming impacts on early life stages increase the vulnerability and delay the population recovery of a long-lived habitat-forming macroalga. *J. Ecol.* 107(3), 1129–1140. doi: https://doi.org/10.1111/1365-2745.13090

Lenormand, T. (2002). Gene flow and the limits to natural selection. *Trends Ecol. Evol*. 17(4), 183-189.