Title

Genomic responses to parallel temperature gradients in the eelgrass Zostera marina in adjacent bays

Journal

Molecular Ecology, 32(11)

ISSN

0962-1083

Authors

Schiebelhut, Lauren M
Grosberg, Richard K
Stachowicz, John J
et al.

Publication Date

2023-06-01

DOI

10.1111/mec.16899

MOLECULAR ECOLOGY

Genomic responses to parallel temperature gradients in the eelgrass *Zostera marina* in adjacent bays

Journal:	*Molecular Ecology*
Manuscript ID	MEC-22-1114.R1
Manuscript Type:	Original Article
Date Submitted by the Author:	n/a
Complete List of Authors:	Schiebelhut, Lauren; University of California Merced, Life and Environmental Sciences Grosberg, Richard ; University of California Davis Stachowicz, Jay; University of California Davis Bay, Rachael; UC Davis,
Keywords:	non-parallel evolution, population genomics, Zostera marina, Population Genetics - Empirical

Genomic responses to parallel temperature gradients in the eelgrass *Zostera marina* in adjacent bays

Lauren M. Schiebelhut[1], Richard K. Grosberg[2], & John J. Stachowicz[2], Rachael A. Bay[2]

[1]Life and Environmental Sciences, University of California, Merced, California

[2]Department of Evolution and Ecology, University of California, Davis, California

Abstract

The extent of parallel genomic responses to similar selective pressures depends on a complex array of environmental, demographic, and evolutionary forces. Laboratory experiments with replicated selective pressures yield mixed outcomes under controlled conditions and our understanding of genomic parallelism in the wild is limited to a few well-established systems. Here, we examine genomic signals of selection in the eelgrass *Zostera marina* across temperature gradients in adjacent embayments. Although we find many genomic regions with signals of selection within each bay. However, there is very little overlap in signals of selection at the SNP level, despite most polymorphisms being shared across bays. We do find overlap at the gene level, potentially suggesting multiple mutational pathways to the same phenotype. Using polygenic models we find that some sets of candidate SNPs are able to predict temperature across both bays, suggesting that small but parallel shifts in allele frequencies may be missed by independent genome scans. Together, these results highlight the continuous rather than binary nature of parallel evolution in polygenic traits and the complexity of evolutionary predictability.

Keywords: non-parallel evolution, population genomics, *Zostera marina*

Introduction

The question of whether patterns of trait and underlying genomic and developmental variation show parallel (or convergent) responses across similar selection gradients is fundamental to our understanding of how evolution operates and whether evolutionary outcomes are predictable (reviewed in Bolnick, Barrett, Oke, Rennison, & Stuart, 2018). All else being equal (and notwithstanding the stochastic effects of genetic drift), independent populations experiencing comparable selective regimes should exhibit parallel responses, at least at the trait level. The degree to which such populations share similar genetic histories should also govern the extent to which they exhibit parallel responses at the genomic level (Härer, Bolnick, & Rennison, 2021; Rennison, Delmore, Samuk, Owens, & Miller, 2020). Thus, populations most likely to evolve along parallel evolutionary pathways should have recently diverged, sharing a similar pool of ancestral genetic variation, and independently facing comparable forces of selection (Bohutínská et al., 2021; Conte, Arnegard, Peichel, & Schluter, 2012).

The extent of parallelism in populations experiencing nominally comparable selective regimes may also depend on the level of organization under consideration, from the traits themselves, to underlying variation at the level of nucleotides, genes, and developmental pathways (Bolnick et al., 2018; Conte et al., 2012; Stuart, 2019). For example, many selectively important traits are highly polygenic, reducing the likelihood of parallelism at the genetic/genomic level, with different genetic combinations producing functionally equivalent phenotypes (Arendt & Reznick, 2008). In addition, there are often multiple phenotypic solutions, some based on historical contingency, to the same selective challenges, with different combinations of traits leading to a suite of phenotypes with similar performance

(Bolnick et al., 2018). Finally, non-parallel mutations in particular genes, or mutations within the same class of genes (but not the same genes), may produce parallel phenotypic or functional effects (e.g., Cassin-Sackett, Callicrate, & Fleischer, 2019; Rosenblum, Römpler, Schöneberg, & Hoekstra, 2010).

In principle, the strongest tests for parallel evolution would use highly replicated populations with known evolutionary histories inhabiting comparable selective regimes in which patterns can be assayed at both the trait and genomic levels. For these reasons, the majority of studies exploring parallelism of both genomes and traits have involved long-term laboratory studies of microbes, often initiated from a single clone, and allowed to evolve over thousands of generations under varying selective regimes (reviewed in Blount, Lenski, & Losos, 2018; also see Pickersgill (2018) for analysis of crop plants). Consistent with predictions, parallel outcomes are most likely in recently established populations experiencing equivalent selective regimes (e.g., Blount et al., 2018). Nevertheless, responses are often strikingly inconsistent among replicate lines, even those initiated from the same founder and subjected to presumably identical selective regimes (Tenaillon et al., 2012). This suggests that stochastic processes such as genetic drift, and the often complex relationships among genotypes, phenotypes and fitness may also be shaping evolutionary responses at the genomic and trait levels (Conte et al., 2012).

Far less is known about evolutionary patterns in natural populations of multicellular organisms (Blount et al., 2018; Bolnick et al., 2018). The majority of the data that we do have from a few well-characterized model systems suggests that parallelism is equally inconsistent in these systems (Stuart et al., 2017). However, recent advances in the accessibility of genomic tools make it possible to simultaneously assess population structure and evolutionary history

while identifying specific mutations associated with parallel selection pressures. This allows for studies of parallel evolution in "replicate" populations with unknown/uncertain evolutionary histories and longer generation times and thereby extends these experimental studies to natural populations of more complex non-model organisms. Several recent studies comparing natural populations identified substantial degrees of parallelism both in terms of traits and underlying genetic patterns (e.g., Bohutínská et al., 2021; Roda et al., 2013; Stuart et al., 2017); however, as with several laboratory studies on microbes, in many wild populations multiple factors, including genetic drift (Szendro, Franke, de Visser, & Krug, 2013), environmental heterogeneity (Magalhaes et al., 2021; Stuart et al., 2017), and different histories of fluctuating selection (Liu, Ferchaud, Grønkjaer, Nygaard, & Hansen, 2018) often contribute to varying degrees of decoupling of parallelism between genes and traits (e.g., Rivas et al., 2018).

As the most widely distributed marine angiosperm in the northern hemisphere, the seagrass *Zostera marina* occurs from the Arctic to the subtropics and in the Pacific, Atlantic and the Mediterranean. The broad distribution of *Z. marina* encompasses a wide range of environmental conditions (with respect to light, salinity, and temperature) that vary both seasonally and geographically. *Z. marina* undergoes both vegetative and sexual reproduction, with varying proportions across sites (Reusch et al., 1999; Olsen et al., 2004). In most populations, individual shoots (ramets) derived from a sexually produced individual are intermingled to form the seagrass meadow (see map in Kollars et al. 2022), with only a few extreme exceptions (Yu et al. 2022). Clonal genotypes, even those collected from the same bed, differ in fitness-related traits, including shoot production, biomass, photosynthesis, and nutrient uptake (Abbott, DuBois, Grosberg, Williams, & Stachowicz, 2018; Hughes, Stachowicz,

& Williams, 2009; Reynolds, DuBois, Abbott, Williams, & Stachowicz, 2016; Salo, Reusch, & Boström, 2015). However, the expression of these traits also depends on environmental context, leading to differential fitness of genotypes at particular sites or seasons (DuBois, Abbott, Williams, & Stachowicz, 2019; DuBois, Williams, & Stachowicz, 2021). This standing genetic variation provides the foundation for local adaptation and reciprocal transplants have demonstrated home-site advantage even at spatial scales on the order of a few km (DuBois, Pollard, Kauffman, Williams, & Stachowicz, 2022; Hämmerli & Reusch, 2002). Moreover, population structure in *Z. marina* tends to be high with significant divergence at all spatial scales, including across tidal heights (Campanella, Bologna, Smith, Rosenzweig, & Smalley, 2010; DuBois et al., 2022; Kamel, Hughes, Grosberg, & Stachowicz, 2012; J. H. Kim et al., 2017), so limited gene flow, even at the scale of meters, may facilitate local adaptation.

In this paper we simultaneously characterize patterns of genomic and functional variation in multiple populations of *Z. marina*, across adjacent bays with overlapping thermal gradients. We focus on populations of *Z. marina* inhabiting multiple locations along temperature gradients in the adjacent Tomales Bay and Bodega Harbor in north-central California, USA. These two bays are just 10 kilometers apart, making it likely that populations within each bay arose from the same ancestral population. This, along with the overlapping selective gradient (i.e., temperature) represents a set of conditions under which parallel responses might be most likely. Previous work has demonstrated strong genetic structure in this region, as well as local adaptation to temperature; reciprocal transplants in Tomales Bay show home-site advantage and plants from cooler sites have decreased growth under experimental warming (DuBois et al., 2022). Because of this strong population structure and

111 because warm sites in each embayment are farthest from the open sea and thus the most
112 distant from one another of all our sites, we can interpret parallel shifts in allele frequencies as
113 signatures of selection, relative to random processes such as genetic drift. We build on this
114 trait-based evidence of local adaptation to investigate the genomic basis of adaptation and the
115 potential for parallel evolution, focusing on three questions. We first ask whether genotype-
116 temperature associations exist across each temperature gradient independently and whether
117 there are parallel changes in allele frequencies across tidal heights. We then determine
118 whether parallelism at the genomic level exists across the two gradients, assessing overlap at
119 three levels of organization: the mutation, the gene, and the functional pathway. Finally, to
120 assess the predictability of detected genotype-temperature relationships, we test whether
121 genetic variation across many SNPs can be used to predict thermal environment from individual
122 genotypes, both within and between bays.
123
124 **Methods**
125
126 ***Sampling, DNA extraction, and sequencing***
127 For population genomic analyses, we collected 2-3 shoots attached by a rhizome from fifteen
128 putative genets (separated by approximately 5-10 m) at 14 sites across Tomales Bay and
129 Bodega Harbor in California (Table 1) from a height below 0.0 mean lower low water (MLLW)
130 (i.e., not sampling the uppermost or lowermost vertical distribution of *Z. marina*). At our study
131 sites (and throughout much of the Pacific Ocean) genets cover a relatively small area such that
132 ramets sampled 1 - 2m apart rarely include the same multi-locus genotype (Kamel et al. 2012;

Reynolds et al. 2017; Duffy et al. 2022; Kollars et al. 2022). For two of the sites sampled in Bodega Harbor (Mason's Marina and Westside Park) we also collected a deeper set of specimens (at least –0.6m below MLLW) to test for genetic differences between shallower versus deeper plants. We transported plants back to the University of California, Davis in a cooler with ice packs, and stored them for no more than one day in a recirculating seawater table before dissecting out the tissue from within the leaf sheath of all shoots within a genet and then flash-froze them in liquid nitrogen and stored at -80°C. Using the inner leaf sheath tissue allowed us to minimize the amount of non-eelgrass DNA by selecting tissue that was free of epibionts and had lower chloroplast concentrations.

We extracted DNA from up to 200 mg of frozen tissue by grinding with a plastic pestle and liquid nitrogen in a 1.5 ml tube until powdered and then by using a modified CTAB chloroform extraction (Doyle & Doyle, 1987). Briefly, tissue was resuspended in 800 ul CTAB (0.1 M Tris-HCl [pH 8.0], 0.02 M EDTA [pH 8.0], 3% CTAB, 1.4 M NaCl, 0.2% β-mercaptoethanol), after the first chloroform-isoamyl alcohol step, the upper aqueous phase was transferred to a new tube and treated with 2µl of RNAse A at 37°C for one hour, followed by an additional chloroform-isoamyl alcohol step before completing the remaining steps. We quantified DNA on a Qubit fluorometer and adjusted the concentration to ~13 ng/µl; in cases where the concentration was lower than 17 ng/µl the concentration was not adjusted. DNA quality was visually assessed on a 2% agarose gel. We submitted genomic DNA for 240 individuals to the Genomics and Bioinformatics Services Texas A&M Agrilife Research centre (College Station, TX) for library preparation using the high-throughput PerkinElmer NEXTFLEX® Rapid XP DNA-Seq Kit

and paired-end 150bp sequencing (targeting 10X coverage with ~5.8Gb/sample) on two lanes of a NovaSeq 6000 S4 X.

Alignment and SNP calling

For the whole genome sequences we used bbduk from the BBTools suite v38.73 for adapter trimming and quality and length filtering (Bushnell, 2021; see code for specific parameters). We aligned whole genome sequences to the *Zostera marina* 3.1 genome (NCBI accession number PRJNA701932; Ma et al., 2021; Olsen et al., 2016) with bwa-mem in bwa v0.7.13 (Li & Durbin, 2009) and called SNPs using GATK v. 4.1.0.0 (McKenna et al., 2010). Briefly, we converted sam files to bam format and sorted using samtools v1.9 (Li et al., 2009). Using GATK, we marked duplicates, called haplotypes, combined g.vcf files and then genotyped individuals across batches of 50 (following GATK recommendations for working with large cohort sizes). After retaining only SNPs, we applied additional hard filters for mapping quality, strand bias, variant confidence, and variants with excessive depth following the best practices guidelines in (Van der Auwera et al., 2013; see code for specific parameters). Genotypes for individuals were recoded to missing if they did not have a minimum depth of 10 and a minimum genotype quality score of 30. Although we sampled plants 5-10m apart to limit sampling multiple clonal shoots from the same genet, we also used the SNP data to identify and remove clones, which could confound downstream population genetic analyses. To filter clones from the data set, we used Rclone v1.0.2 (Bailleul, Stoeckel, & Arnaud-Haond, 2016) in R v4.0.3 (R Core Team, 2020) with a reduced set of SNPs without any missing data (as required by the program) for each geographic location separately. After removing clones we used vcftools v0.1.16 (Danecek et al.,

2011) to filter the final data set to include only bi-allelic SNPs with a minor allele frequency of at least 0.01 and a genotype call rate of at least 85% of individuals.

Population genetic analyses

For population genetic analysis, we first thinned the SNPs based on linkage disequilibrium (LD) using SNPRelate (Zheng et al., 2012) with an LD threshold of 0.5 based on SNPRelate's "composite" measure of LD. This thinned set was used for clustering analysis, PCA, and F_{ST}. We used SNPRelate to conduct principal components analysis and to calculate pairwise F_{ST} (Weir & Cockerham, 1984) between all pairs of sites. We conducted clustering analysis and estimated ancestry proportions using the R package tess3r (Caye, Deist, Martins, Michel, & François, 2016). For each K value (1-6) we ran five replicate runs with the lambda parameter set to 0 to ignore priors based on the spatial distribution of samples.

Sampling water temperature

To record water temperature we deployed HOBO Pendant® MX2201 loggers (fastened to PVC pipe) in the area at each of the sites from which we collected genetic samples. The pipe was driven into the sediment until the logger was approximately <15 cm above the sediment surface, positioned to rarely be emersed except during low spring tides. We recorded water temperature at 15-minute intervals during a two-week period at all sites from August 16th to 29th in 2019 and for 14 weeks (Aug 1st – Nov 11th, 2019) for a reduced set of sites that excluded Mason's Marina and one logger at Westside Park in Bodega and Pelican Point and Pita Beach in Tomales Bay. Because our main interest was to determine the relative differences in

198 temperatures between sites, we calculated the mean and first and third quartiles for each
199 location and used Spearman's correlation coefficient to evaluate whether the two-week
200 temperature interval correlated with the longer 14-week interval.
201
202 ***Environmental association analyses***
203 We used two approaches to search for SNPs associated with environmental gradients in
204 Tomales Bay and Bodega Harbor separately: (1) latent factor mixed models (LFMM), correlating
205 individual genotypes at each SNP with mean temperature at the sampling location (Figure 1)
206 using lfmm2 in LEA v4.0.0 (Frichot & François, 2015) and (2) F_{ST} outliers using OutFLANK v0.2
207 (Whitlock & Lotterhos, 2015), grouping warmer vs. cooler sites. For LFMM analysis, genotypes
208 were first imputed in LEA using ancestry coefficients estimated by LEA (K=3). The temperature
209 gradient in Tomales Bay reaches much higher temperatures than that in Bodega Harbor, so to
210 capture a similar range of temperatures for comparisons between bays, we also repeated the
211 analyses with a subset of four sites in Tomales Bay (Lawson's Landing, Pita Beach, Nick's Cove,
212 and Sacramento Landing; Figure 1) that more closely matched the temperature range of
213 Bodega Harbor. We hereafter refer to this reduced Tomales sample as "Tomales (cool)" We also
214 used OutFLANK to identify SNPs associated with higher versus lower intertidal habitat in
215 Bodega. All selection scans were performed after first filtering for SNPs with a minor allele
216 frequency greater than 5% for samples within a given subset of sampling locations. This filter
217 reduces the number of tests performed and discards SNPs unlikely to result in true positives
218 (Caye et al., 2019). Additionally, we assessed outlier SNPs at two different significance cut-offs:
219 $p < 0.001$, a more liberal cutoff without multiple test correction and $q < 0.05$, adjusted for false

discovery rate. Because false negatives can be high especially for moderate sample sizes and complex traits, the opportunity to compare across these two significant thresholds provides information about the sensitivity of our conclusions to these decisions.

To characterize parallel associations with temperature at the SNP level between the two bays, we directly compared the genomic positions of outlier SNPs from the LFMM and OutFLANK analyses. At the gene level, we used LD-Annot v0.4 with $r^2 = 0.9$ (Prunier et al., 2019) to first identify the genes in linkage disequilibrium with candidate SNPs and then compare lists to detect overlapping genes. Finally, to determine whether there was overlap at the functional level, we tested whether the outlier-associated gene sets were enriched for particular gene ontology (GO) terms using TopGO v2.40.0 (Alexa & Rahnenführer, 2009) in R. We used a possible gene universe of all genes within linkage disequilibrium ($r^2 = 0.9$) of the full set of SNPs, rather than the full set of annotated genes. We used a Fisher's exact test to identify significantly enriched GO terms and required that more than two genes be significant for a particular GO term. We used a significance threshold of $p<0.05$ and did not adjust for multiple testing, as suggested in (Alexa & Rahnenführer, 2009), as our goal was to summarize the functional categories based on our SNPs rather than to make strong conclusions about any particular GO term.

To test predictability of genotype-environment associations within and across bays, we created polygenic scores using the R-package randomForest v4.6-14 (Liaw & Wiener, 2002). We used the mean temperature for each site as the response variable and candidate SNPs derived from LFMM or OutFLANK analyses as predictors. To reduce redundancy and computational time due to extensive linkage across our candidate SNPs, we first thinned the candidate SNP

sets based on linkage disequilibrium in SNPRelate (ld.threshold = 0.5). Because randomForest cannot accept missing data, we used genotypes that were imputed with LEA. We ran separate random forest models with the sets of candidate SNPs from LFMM and OutFLANK for each bay. For each SNP set, we conducted runs using either all samples to train the model, or only samples from the bay from which the candidate set was derived (e.g., when candidate SNPs were identified using analysis of Bodega Harbor samples, we trained using either all samples or Bodega Harbor samples only). For validation, we used either all samples, Bodega Harbor only, or Tomales Bay only. To limit bias in the training set, we conducted cross-validation at the level of the sampling site, training the random forest on the dataset with a single sampling site removed, then predicting temperature for that site. For each random forest run, this procedure gave us a random-forest predicted temperature for each individual sample based on candidate SNPs. We then compared predicted and observed temperatures using Spearman's correlation coefficient. For each run, we report the percent variance explained in the training model and the correlation coefficient when comparing observed and predicted temperatures.

Because isolation-by-distance is strong in *Z. marina* at this scale in this region (Kamel et al. 2012), we used two approaches to reduce the impacts of IBD driving our predictive power. First, we included in our random forest predictors the sample loadings on the first two PC axes. These first two PC axes represent the major geographical patterns in the system. Second, we ran both training and prediction with 100 random sets of SNPs (in addition to the first two PC axes). We used these randomizations to estimate a 90% confidence interval for a null distribution, showing how well our predictions performed when based on neutral genetic

variation alone. When observed variance explained by the random forest model or Spearman's correlation coefficients fell outside this range, we considered them significant.

Results

Genetic variation

After filtering and removing 14 individuals due to low read count, we retained 446,718 SNPs. Where putatively separate samples were identified as clonemates, we retained only one individual from each genet, resulting in 42 additional samples excluded (Table 1, Figure S1). In Tomales Bay, a greater fraction of our samples were identified as clonemates (Figure S1). In the final set of 182 individuals and 446,718 SNPs, 91% of individuals had less than 20% missing data; and all individuals had less than 29% missing genotypes. After thinning for linkage disequilibrium, we were left with 46,166 SNPs for analysis of population structure. Both within and among bays, we found strong isolation by distance (IBD), with pairwise F_{ST} correlated with geographic distance between sites (Figure S3; Spearman's ρ=0.65). Differentiation between bays was highest (mean pairwise F_{ST} = 0.058, SD = 0.020), and sites within Tomales Bay were more differentiated (mean pairwise F_{ST} = 0.047, SD = 0.021) than sites within Bodega Harbor (mean pairwise F_{ST} = 0.002, SD = 0.005). Both PCA and clustering analyses supported the IBD pattern (Figure 1), with PC1 and PC2 explaining 4.06% and 1.98%, respectively. Bodega Harbor and Tomales Bay represented distinct clusters on the PCA plot, separated along PC1. PC2 mirrored the geography of Tomales Bay, with higher values for sites closer to the mouth of the bay. In the clustering analysis, we observed decreasing cross-validation scores across all values

of K, suggesting no clear 'best' K value within the range we tested. We present K=3 (Figure 1), as K=4 did not show further geographic differentiation but rather separated five Bodega Harbor individuals. However, at higher K values (K=5,6) additional structure within Tomales Bay became apparent, further highlighting the strong IBD signal (Figure S2).

Environmental association analyses

We documented clear temperature gradients within both Bodega Harbor and Tomales Bay. Summer mean temperatures recorded from 16-29 August in 2019 in Bodega Harbor and Tomales Bay ranged from 16.1–18.5°C and 16.1–22.2°C, respectively, with the mouth of each bay cooler than the back of the bay (Figure 1*C*; Figure S4). Mean temperatures over this two-week period were highly correlated with an extended period (1 August to 11 November; ρ=0.95, p<2.2e-16) for which we had logger coverage at all but three sites in Bodega Harbor (two at Mason's Marina and one at Westside Park) due to logger failure, and all but two sites in Tomales Bay, because the Pelican Point and Pita Beach loggers were not deployed until mid-August (Figure S4). The ordinal ranking of sites by summer temperature reported here is identical to that in other studies of subsets of these sites with longer temperature records (Aoki et al., 2022; DuBois et al., 2022).

We identified SNP candidates for selection within each of the two embayments. After filtering for minor allele frequency >5%, 357,010 SNPs remained for environmental association analysis. The vast majority of genetic variation is shared across bays, with only 7,984 SNPs (2.2%) being fixed in one of the bays. Table 2 summarizes the number of SNP outliers at each significance threshold for all analyses. In this discussion and in downstream analyses, unless

stated otherwise, we use the false discovery rate-corrected (q<0.05). Bodega Harbor, 8560 candidate SNPs were associated with site mean temperature in the LFMM analysis and 314 SNPs were significant outliers when comparing the warmest vs. coolest sites in OutFLANK (Figure 2; Table 2). Of the 8472 SNPs identified by LFMM, 8183 (96%) are largely located in a single linked block on Chromosome 1 (Figure 2). Twenty-nine SNPs overlapped between these two analyses. We found 711 genes in linkage disequilibrium with candidate SNPs from the LFMM analysis and 62 from the OutFLANK analysis (see supplementary files), of which 17 overlapped between both analyses. Gene enrichment analyses of the candidate genes identified 33 GO terms from LFMM (15 after removing the highly linked region) and 9 from OutFLANK, though none overlapped between the two analyses (Table 2; supplementary files). In the LFMM analysis, significant GO terms included categories broadly involved in cell wall modification, nucleotide metabolism, and enzyme activity while the OutFLANK analysis highlighted categories related to lipid metabolism and heme binding, among others.

For Tomales Bay, there were no SNPs significant (q<0.05) when including all Tomales locations. At the unadjusted significance cutoff, 659 candidate SNPs ($p < 0.001$) were associated with mean site temperature across all 10 sampling sites (Table 2; Figure 2). In the LFMM analysis that was restricted to the four coolest sites that most closely matched the temperature gradient of Bodega Harbor, LFMM identified 6710 SNPs (q<0.05) (Figure S5). As with the Bodega Harbor LFMM analysis, the large number of significant SNPs in the reduced Tomales LFMM analysis is largely due to a single highly linked block of SNPs on Chromosome 1; with 6256 of 6710 (93%) in that region. The OutFLANK analyses comparing warmer versus cooler sites in Tomales did not yield any candidate SNPs, for either the full 10 sites or the reduced set

of four sites. We found 662 genes in linkage disequilibrium with candidate SNPs from the LFMM analysis restricted to only the cooler sites. Twenty GO terms were enriched in the Tomales (cool) LFMM analysis, including nucleic acid and amino acid metabolism functions and enzyme activities. When the linked section of Chromosome 1 was excluded, 21 terms were enriched including functions associated with oxidative stress, carbohydrate metabolism, and organic compound binding.

Tidal height associations

The OutFLANK analysis comparing samples from upper and lower tidal heights in Bodega Harbor identified 461 SNPs differentiating the two groups (Table 2, Figure S6, supplementary file). Only one of these SNPs overlapped with other environmental association analysis at the SNP, gene, or functional levels. The unique SNPs differentiating upper and lower intertidal groups were in linkage disequilibrium with 108 genes (Table 2). Gene enrichment analysis found 49 enriched GO-terms (Table 2; supplementary files). Most enriched GO terms were generally involved in transcriptional regulation. None of these GO terms overlapped with any of the environment association analyses.

(Non)overlapping associations

There was no overlap in candidate SNPs between Tomales and Bodega Harbor association analyses (Figure 3), with the exception of one highly linked region. Even at a more lenient significance cutoff (unadjusted $p<0.001$) only 114 candidate SNPs overlapped between embayments, a small fraction (0.8%) of the total 13,360 candidates across all analyses,

351 excluding the highly linked region (Figure S7). There were 5265 significant SNPs that overlapped
352 between the Tomales LFMM analysis restricted to cooler sites and the Bodega LFMM analysis.
353 These SNPs were all located on a single linked region of Chromosome 1 (Figure 2, Figure S5) and
354 so likely represent a single locus. We found 549 genes linked (at LD $r^2 = 0.9$) to candidate SNPs
355 in the Tomales LFMM analysis that was restricted to four cooler sites overlapped with the
356 Bodega OutFlank analysis. After excluding SNPs from the highly linked region, there was still
357 overlap in four genes, despite the absence of overlap at the SNP level. These included a
358 phytochrome interacting factor (PIF1) and endoglucanase, as well as two genes that also were
359 also significant in the more lenient ($p<0.001$) Tomales LFMM analyses that included all 10
360 locations: UDP-glucuronate:xylan and protein disulfide isomerase.
361 Gene ontology (GO) enrichment analysis for outlier-associated gene sets revealed 14 GO
362 categories that overlapped between the Bodega and Tomales analyses when all SNPs were
363 included. These categories include a range of functions associated with enzyme activity, RNA
364 metabolism, and binding of sugars and phosphates. However, when the large linked region is
365 removed, there are still five functional categories that overlap between Bodega and Tomales
366 (cool) analyses: carbohydrate biosynthetic process, heme binding, UDP-glycosyltransferase
367 activity, transferase activity, and tetrapyrrole binding. Although the large linked block of SNPs
368 represents a prime candidate for follow-up studies, it was not significant in the analysis of all
369 Tomales locations and was only slightly above the significance threshold in the reduced
370 Tomales set. Additionally, both the Bodega and reduced Tomales analyses contain few sites (6
371 in Bodega including high/low intertidal and 4 in Tomales), so further information is needed to
372 test a role for this region in parallel evolution.

Polygenic analyses using random forest successfully predicted temperature variation from SNP variation within bays. We used four sets of candidate SNPs that yielded significant candidates for SNPs associated with temperature: Candidates from the three analyses that yielded significant SNPs at q<0.05 (Bodega LFMM, Bodega OutFLANK, and Tomales (cool) LFMM) and an additional set from the full Tomales LFMM at p<0.001, which was included so that we had representation from SNPs that might vary with temperature across the full length of Tomales Bay. We trimmed each SNP set for LD, yielding 175, 61, 121, and 86 SNPs respectively, and used these thinned sets as predictors in the random forest alongside the first two PC axes (to reduce the impact of population structure). Figure 4*A*-*C* shows an example run, in which we built the random forest model using SNPs significant in the Bodega LFMM analysis and used all sites in both training and prediction (leaving out one site at a time). In this analysis, the candidate SNPs result in a model with a higher fraction of variance explained (R^2=0.84) than the same number of random SNPs (90% confidence interval 0.76-0.80). Additionally, the correlation between observed temperature and that predicted by the random forest analysis (Figure 4*B*) is higher than with a random set of SNPs (Figure 4*C*; ρ=0.80; 90% CI=0.62-0.69). In all cases except the Tomales (cool) LFMM, the percent variation explained by the random forest model was significantly higher than with a random set of SNPs. When Tomales Bay sites were used to train the random forest, the null distribution also explained a large percentage of the variance, likely because we included PC axes in the model, which represent the strong isolation by distance in Tomales Bay rather than temperature per se. Still, candidate SNPs rather than random SNPs explained a significantly higher proportion of the variance.

Although random forest models performed well when candidate SNP discovery and prediction included only locations within a single embayment, the models did not necessarily perform well when predicting across embayments (Figure 4*D*). Temperature-associated SNPs discovered in Tomales Bay using LFMM did not predict the temperature at the sites from which Bodega Harbor individuals were collected. Interestingly, random SNPs trained and predicted on Bodega Harbor individuals had a high but negative correlation between observed and predicted temperature (mean LFMM ρ=-0.72; OutFLANK ρ=-0.69). These negative correlations have been seen in previous studies (Exposito-Alonso et al., 2019), and may reflect an unmeasured environmental variable that is negatively correlated with temperature and with overall population structure. Notably, however, models based on SNPs discovered in Bodega Harbor (both LFMM and OutFLANK) show significantly higher predictive capability than random SNPs, even when predicting the temperature of Tomales Bay sites and regardless of training population.

Although the overall success of prediction within embayments and mixed performance across bays is consistent with the lack of parallelism at the SNP level, the significant predictive power of SNPs discovered within the Bodega Harbor candidate SNPs for plants in both embayments suggests that our genome scans may have missed more subtle parallelism. We used the variable importance rankings in the random forest to identify SNPs that contributed most to models trained on Bodega Bay samples, which ultimately resulted in improved predictions of temperature across both bays. Two measures of importance are available within the random forest framework: increase in mean squared error and increase in node purity. Both measures identified three SNPs (one from OutFLANK and two from LFMM) with

substantially higher importance than other SNPs or even PC axes (Figure S8), though none overlapped directly with other candidate SNPs from Tomales Bay (Figure 3) or showed linkage with the nearest candidate SNPs (with the range of LD r^2 from 0.001–0.089 at 0.11–6.7 Mb away from the SNPs). Allele frequencies for these SNPs show that they are associated with temperature across both bays and standard linear models indicate a significant correlation with temperature ($p<0.05$), even when embayment is included as a covariate (Figure S9). Therefore, while the overwhelming conclusion from the genome scans was a lack of parallelism, it appears that the polygenic framework revealed subtle parallel signals.

Discussion

Many factors influence the degree of parallel change in parallel selective environments, such as population size, population history (demographic and genetic), gene flow, environmental heterogeneity, and the fact that there are multiple pathways to the same functional endpoint (Bolnick et al., 2018; Conte et al., 2012; Ralph & Coop, 2015). That all of these vary widely in natural populations at least in part explains the mixed results observed in analyses of parallelism in the wild (Kern & Langerhans, 2018; Oke, Rolshausen, LeBlond, & Hendry, 2017; Stuart et al., 2017). In our study we examine a system in which parallelism would be expected to be likely due to overlapping selective gradients in geographically proximate populations. Counter to this expectation and despite a large degree of shared genetic variation overall, we find that genomic signatures of selection in *Zostera marina* are largely non-overlapping at the SNP level. However, there is detectable overlap in the genes and functional categories associated with temperature variation, and some predictability of temperature across gradients

using polygenic scores. Together, these data suggest a complex pattern of subtle parallelism within a mostly non-parallel signal. This supports the emerging view of parallelism as continuous rather than binary (Bolnick et al., 2018).

The degree of local adaptation and parallelism can be influenced by the magnitude of gene flow among populations, with migration from nearby populations providing sources of both adaptive and maladaptive genetic variation. We find strong differentiation and isolation by distance among our sample sites, both within and between bays. This is likely due to limited pollen and seed dispersal in *Z. marina* (Ruckelshaus, 1996). Microsatellite studies from Bodega Harbor and Tomales Bay, as well as several other regions, show consistently strong genetic differentiation even at small spatial scales (Campanella et al., 2010; DuBois et al., 2022; Kamel et al., 2012; J. H. Kim et al., 2017; Muñiz-Salazar, Talbot, Sage, Ward, & Cabello-Pasini, 2005; Ort, Cohen, Boyer, & Wyllie-Echeverria, 2012). With limited gene flow to disrupt the effects of selection, local adaptation might occur and be maintained very quickly even at microgeographic scales (Richardson, Urban, Bolnick, & Skelly, 2014).

Within each bay, genotype-environment associations were strong, consistent with previous work establishing local adaptation among populations in Tomales Bay. Dubois et al. (DuBois et al., 2022), in a reciprocal transplant that included three of the sites in the present study, documented home-site advantage in survival and growth. Laboratory experiments identified temperature (along with light and grazing pressure) as one of the drivers of local adaptation, with plants from the cool end of the gradient showing decreased performance under elevated temperatures. In Bodega Harbor, temperature-associated SNPs were linked to genes enriched in metabolic processes (e.g., lipid metabolism, oxidation reduction) and cell wall

synthesis (e.g., pectinesterase activity, cell wall modification). These perhaps reflect genotype and environment-specific tradeoffs in the use of carbon for growth and metabolism. In Tomales Bay, gene ontology enrichments involved both categories potentially related to growth (e.g. carbohydrate biosynthetic process, cellular carbohydrate metabolic process), but also stress response (e.g. response to oxidative stress, antioxidant activity).

Despite evidence for selection across each temperature gradient, there was little overlap in signals of selection at the SNP level. The lack of a pervasive signal of parallelism at the genomic level in the two populations inhabiting Tomales Bay and Bodega Harbor could reflect the impacts of several historical and contemporary evolutionary processes. First, despite their proximity, the two embayments may not be recently derived from the same ancestral population and could instead originate from different sources that had already differentiated to a greater or lesser extent. Second, contemporary gene flow is low between bays (based on our population structure analyses), and even within bays population structure is quite high, likely because of overall limited pollen and seed dispersal (Ruckelshaus, 1996). Beneficial (and deleterious) alleles will thus be slower to spread between bays. Using simulations, Ralph and Coop (2015) demonstrated that in patchy habitats the threshold at which patches are more likely to evolve independent beneficial mutations is a function of dispersal distance and selective advantage of the mutations. In the patchy eelgrass system characteristic of both Tomales Bay and Bodega Harbor, with limited dispersal of pollen and seeds, very strong and persistent selection may be required to enhance the spread of shared adaptive alleles. This is potentially exacerbated by the fact that warmer sites, which exert the strongest selection pressure, are farthest from the mouths of the bays so that when a beneficial allele arises it

must pass through less favorable habitat, impeding gene flow to other populations, even those nearby (Ralph & Coop, 2010, 2015).

Finally, the range of temperatures within Tomales Bay is much greater than in Bodega Harbor, although we did attempt to reduce the impact of these differences by conducting a reduced analysis based only on overlapping temperature ranges. However, adaptation to other factors besides temperature, whose relative importance may differ between the two embayments, could disrupt parallel signals of selection. For example, seasonal upwelling and phytoplankton and macroalgae blooms are associated with spatially and temporally fluctuating temperature, salinity, chlorophyll, and dissolved oxygen (Hollarsmith et al., 2020; Kimbro, Largier, & Grosholz, 2009). Two of our sites with very similar temperature conditions still exhibited local adaptation to differences in light availability as a result of differences in macroalgal abundance between the sites (DuBois et al., 2021). Although temperature may still impose strong selection, these other sources of selection may constrain the degree of parallelism through, for example through tradeoffs maintained by pleiotropy or epistasis. For example, an analysis of lake-stream pairs of stickleback, although all had adapted across the same primary axis (flow), parallelism was higher when environmental differences were generally more similar (Stuart et al., 2017), highlighting the complexity of multivariate adaptation in wild populations.

Even in highly controlled laboratory experiments, parallelism at the genomic level is far from pervasive, suggesting that in natural populations, where ancestry, selective landscapes, and historical and contemporary patterns of gene flow are far more complex, deep parallelism is even less likely (Bailey, Blanquart, Bataillon, & Kassen, 2017; Lenski, 2017). Indeed, a growing

number of studies examining adaptation across parallel selection gradients reveal a mix of parallel and non-parallel signals (e.g., Liu et al., 2018; Rivas et al., 2018; Stuart et al., 2017). In our data, despite the overall lack of overlap of signals of selection between the two embayments at the SNP level, there was a small amount of overlap at the gene level and some power for predicting temperature with polygenic scores. Although candidate SNPs from each bay analysed separately had little overlap, the random forest model could still predict Tomales Bay temperatures based on Bodega Harbor candidate SNPs, an unlikely relationship in the absence of any parallel signal.

Notably, the most informative SNPs in the random forest model showed correlations between allele frequencies and temperature. These SNPs were perhaps undetected in the genome scan because their effect in Tomales Bay was below the significance threshold; we did not have the power in our study to detect their effects. This is likely a larger issue for detecting parallelism in highly polygenic traits, where we expect small shifts in allele frequencies across many SNPs. Despite the lack of overlap in candidate SNPs, four genes overlapped between Bodega and Tomales analyses, even after we excluded the linked region on Chromosome 1. Since some of the comparisons had no SNP overlap, this may be a case where different mutations in the same gene can lead to comparable phenotypic effects. A similar case was seen in a comparison of small and large morphs of Arctic Charr across multiple lakes in Labrador, where outlier SNPs were largely non-overlapping, but there was overlap at the gene and paralog levels, suggesting multiple mutational pathways to the same functional outcome (Salisbury et al., 2020). In our study, the four genes overlapping between Bodega and Tomales were UDP-glucurorate, protein disulfide isomerase (PDI), an endoglucanase, and PIF1. All of

these genes have known functions in plants, many related to growth. Both the endoglucanase and UDP-glucurorate function in cell wall synthesis (Kuang et al., 2016), and PDI and PIF1 are involved in chloroplast regulation and chlorophyll biosynthesis (J. Kim & Mayfield, 1997; Moon, Zhu, Shen, & Huq, 2008). This perhaps points to different genotypes having different light harvesting abilities, which can result in differences in growth and thermal performance. This hypothesis is in line with previous work in *Z. marina* showing genotypic differences in growth under different temperature and seasonal (light) conditions (DuBois et al., 2021). At the functional level, overlapping enriched GO terms between Bodega and Tomales may also be interpreted as associated with growth, including "carbohydrate biosynthesis process" and UDP-glycosyltransferase activity while others, such as "transferase activity" and "tetrapyrrole binding" may be involved in signaling. Additionally, "heme binding" was enriched in both Bodega and Tomales analyses. Hemes play a number of roles in eukaryotic cells, including respiration, transcription, and protein degradation (Severance & Hamza, 2009).

The most striking exception to the lack of parallelism was a large region of completely linked SNPs, covering approximately 5Mb on Chromosome 1. These SNPs were significant in both LFMM analyses of Bodega Harbor sites and the four cooler sites in Tomales Bay. We hypothesize that these linked sites represent a chromosomal rearrangement, such as an inversion, that is polymorphic in the sampled populations. Because of the suppressed recombination between alleles, inversions may preserve the integrity of co-adapted gene complexes, which could be especially beneficial when there is gene flow across the selective gradient (Thompson & Jiggins, 2014; Wellenreuther & Bernatchez, 2018). With the increasing accessibility of whole genome sequencing, large inversions appear to be more common than

548 previously thought and may be maintained by selection for a long period (Wellenreuther &
549 Bernatchez, 2018). An interesting case study is that of the seaweed fly *Coelopa frigida*, where
550 inversions are associated with life-history trait variation and show strong and parallel
551 segregation across bioclimatic gradients on two continents (Mérot et al., 2018, 2021). However,
552 links between genotype and environment in our study must be cautiously interpreted because
553 of limited sampling replication across parallel thermal gradients. Further sequencing with long
554 reads will be required to characterize the putative rearrangement. Trait-mapping studies and
555 sampling with broader geographic scope are essential for understanding the fitness
556 consequences of the inversion across different environments.
557 The distribution of climate-associated standing genetic variation in contemporary
558 populations will determine their capacity to adapt to future environmental challenges.
559 However, predictions of future adaptive responses will be aided by an understanding of the
560 genetic architecture of local adaptation (Bay et al., 2017; Capblancq, Fitzpatrick, Bay, Exposito-
561 Alonso, & Keller, 2020). In *Zostera marina* there is mounting evidence that marine heatwaves
562 can reduce growth and reproduction (Ehlers, Worm, & Reusch, 2008; Qin et al., 2020; Saha et
563 al., 2020; Smale et al., 2019), but also that these impacts vary across individuals and
564 populations (Bergmann et al., 2010; DuBois et al., 2019; DuBois, Williams, & Stachowicz, 2020).
565 This study shows strong but largely non-parallel temperature-associated genetic variation
566 across adjacent bays. Whether this is due to a low overall dispersal distance, slowing the spread
567 of adaptive alleles, polygenic adaptation with little overlap in genetic architecture, or
568 populations are adapted to a different suite of complex environmental conditions remains
569 unclear. The distinction is important when considering the loss of genetic variation due to

disturbance or when choosing genotypes for restoration: do populations from comparable thermal regimes offer comparably effective sources for restoration of impacted habitats? Our results additionally set expectations for the degree of parallelism we might expect range-wide in *Z. marina*. Our study, which represents conditions likely to generate parallelism (i.e., geographic proximity, overlapping gradients, contemporary gene flow) shows weakly parallel signals at best, which suggests that parallelism at the genomic level across global *Z. marina* populations is unlikely. Further understanding of the phenotypic effects of candidate SNPs under a range of environmental conditions is essential for understanding how parallel selection shapes the evolutionary trajectories of geographically distinct populations, and for developing a more general framework for predicting individual and population response to warming temperatures.

Authors' contributions

L.M.S.: conceptualization, data curation, formal analysis, investigation, methodology, project administration, resources, validation, visualization, writing - original draft, writing - review & editing. R.A.B.: conceptualization, data curation, formal analysis, funding acquisition, investigation, methodology, project administration, resources, supervision, validation, visualization, writing - original draft, writing - review & editing. R.K.G.: conceptualization, funding acquisition, investigation, project administration, resources, supervision, writing - review & editing. J.J.S.: conceptualization, funding acquisition, investigation, methodology, project administration, resources, supervision, writing - review & editing

Conflict of interest

The authors have no conflict of interests.

Acknowledgements

Research was supported by a grant from the U.S. National Science Foundation (OCE-1829976) awarded to R.A.B, R.K.G., and J.J.S. Specimens were collected under scientific collecting permit no. SC-4874 and Specific Use Permit ID: S-211340004-21134-001 from the California Department of Fish & Wildlife and PORE-20I 7-SCI-003 from the National Parks Service. We thank Benjamin Rubinoff for sharing boat time to access Pita Beach and Pelican Point in Tomales Bay. Library prep and sequencing was conducted by Texas A&M AgriLife Research: Genomics and Bioinformatics Service. Sequence processing used the Extreme Science and Engineering Discovery Environment (XSEDE) bridges through allocation and TG-BIO150074 to R.A.B. and TG-OCE200004 to L.M.S., which is supported by National Science Foundation grant number ACI-1548562.

Data accessibility statement

All code, supplementary files, and temperature data are available from the Dryad Digital Repository: https://doi.org/10.6071/M3DD4F (Schiebelhut, Grosberg, Stachowicz, & Bay, 2022b). Sequence data are available at the NCBI Sequence Read Archive (SRA) under BioProject PRJNA887384 (Schiebelhut, Grosberg, Stachowicz, & Bay, 2022a).

Benefit-sharing statement

614 Benefits Generated: Benefits from this research accrue from the sharing of our data and results

615 on public databases as described above.

616
617

Tables

Table 1. List of sample sites, latitude, longitude, sample sizes for genetic analyses, and mean temperature for *Z. marina* intertidal collections. Sample sizes indicate the final numbers of individuals included in analyses after filtering out individuals with high missing data and those that came from the same genet. Parentheses indicate values for lower intertidal samples at Mason's Marina and Westside Park.

Bay	Site name	Site code	Collection Date	Latitude	Longitude	n	Mean temperature (°C)
Bodega	Mason's Marina	MM	16/Jul/2019; 29/Sep/2019	38.3334	-123.0595	11 (9)	17.4 (17.3)
	Doran Beach	DB	16/Jul/2019; 30/Sep/2019	38.3209	-123.0455	14	18.5
	Westside Park	WP	16/Jul/2019; 29/Sep/2019	38.3195	-123.0538	13 (10)	16.8 (16.4)
	Campbell Cove	CC	16/Jul/2019; 29/Sep/2019	38.3097	-123.0584	12	16.1
Tomales	Lawson's Landing	LL	18/Jul/2019; 27/Sep/2019	38.2303	-122.9588	14	17.3
	Pita Beach	PB	16/Aug/2019	38.2049	-122.9495	7	16.1
	Nick's Cove	NC	01/Aug/2019; 27/Sep/2019	38.2048	-122.9272	14	20.3
	Pelican Point	PP	16/Aug/2019	38.1874	-122.9324	10	16.8
	Blake's Landing	BL	01/Aug/2019; 27/Sep/2019	38.1785	-122.9091	14	20.4
	Sacramento Landing	SL	30/Jul/2019; 26/Sep/2019	38.1496	-122.9056	10	19.8
	Marshall Store	MS	30/Jul/2019; 26/Sep/2019	38.1522	-122.8889	13	20.9
	Heart's Desire	HD	26/Sep/2019; 30/Jul/2019	38.1328	-122.8918	12	20.9
	Teacher's Beach	TB	30/Jul/2019; 26/Sep/2019	38.1141	-122.8694	8	22.0
	Millerton Point	MP	31/Jul/2019; 26/Sep/2019	38.1050	-122.8464	13	22.2

Table 2. Summary of candidate SNPs from environmental association analyses and linked genes. Each analysis was conducted for each bay separately. For Tomales Bay, in addition to using all sites in the LFMM and OutFLANK analyses we also performed the analyses on a reduced set of four locations that had a temperature profile more closely matching Bodega Harbor. The number of candidate SNPs are reported for $p < 0.001$ and false discovery rate adjusted $q < 0.05$. Numbers in parentheses exclude the large linked region on Chromosome 1 (positions 32,368,298–37,501,531). Numbers of candidate genes and functions were evaluated for SNPs with $q < 0.05$.

Bay	**Analysis**	**Candidate SNPs** (*p*-value<0.001)	**Candidate SNPs** (qvalue<0.05)	**Candidate Genes**	**Candidate Function**
Bodega	LFMM	8472 (289)	8560 (377)	711 (138)	33 (15)
	OutFLANK	4704 (4682)	314	62	9
Tomales	LFMM - all sites	659	0	NA	NA
	OutFLANK - all sites	0	0	NA	NA
	LFMM - cooler sites	6704 (449)	6710 (454)	662 (97)	20 (21)
	OutFLANK - cooler sites	3439 (3411)	0	NA	NA
Bodega	OutFLANK - tidal	4285	461	108	49

Figures

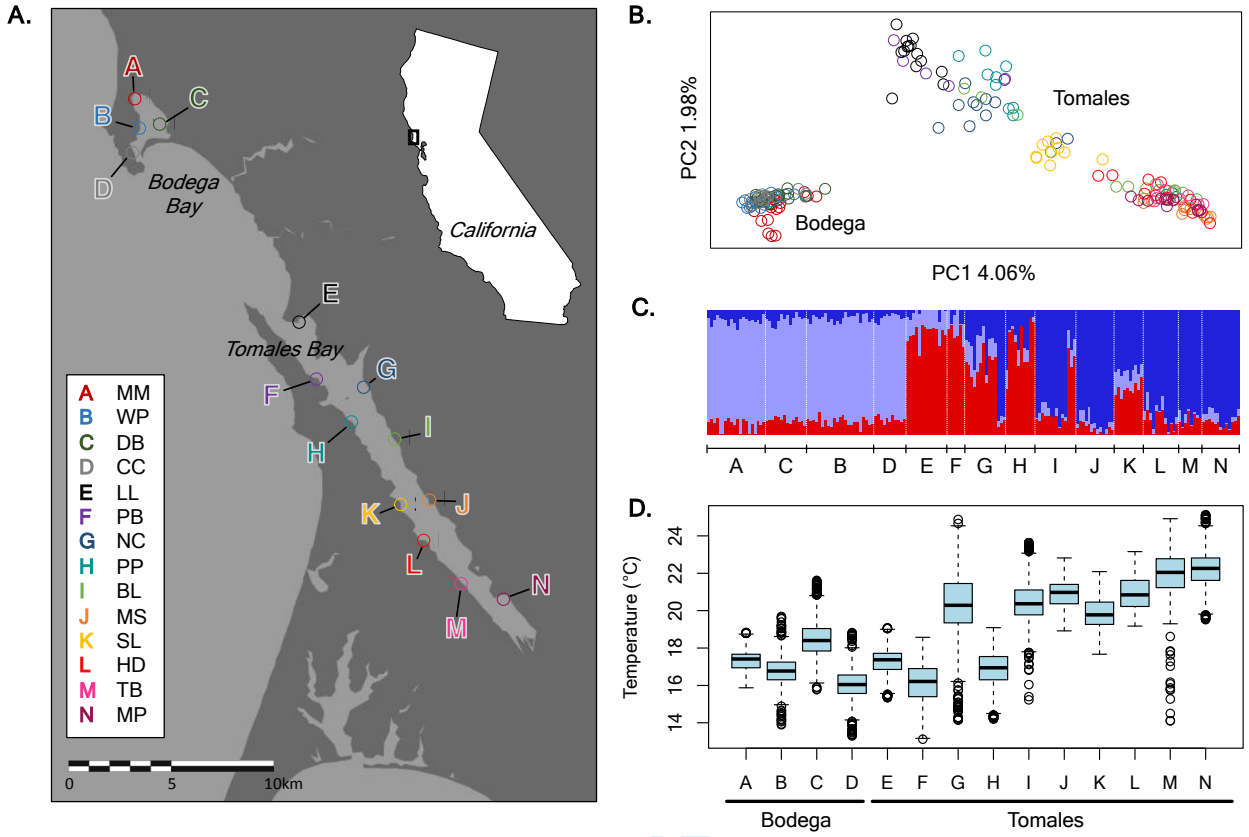


Figure 1. Sampling locations and population genetic structure in *Z. marina*. **(A)** The map indicates the geographic locations where *Z. marina* were collected and temperature loggers deployed (see Table 1 for site details and sample sizes). Colored points and letters correspond to points in the **(B)** PCA. **(C)** Estimated ancestry proportions from clustering analysis (K=3) for each individual, organized by sample site **(D)** Boxplots show the median temperature and first and third quartiles for each site, as measured over a two-week period in August 2019.

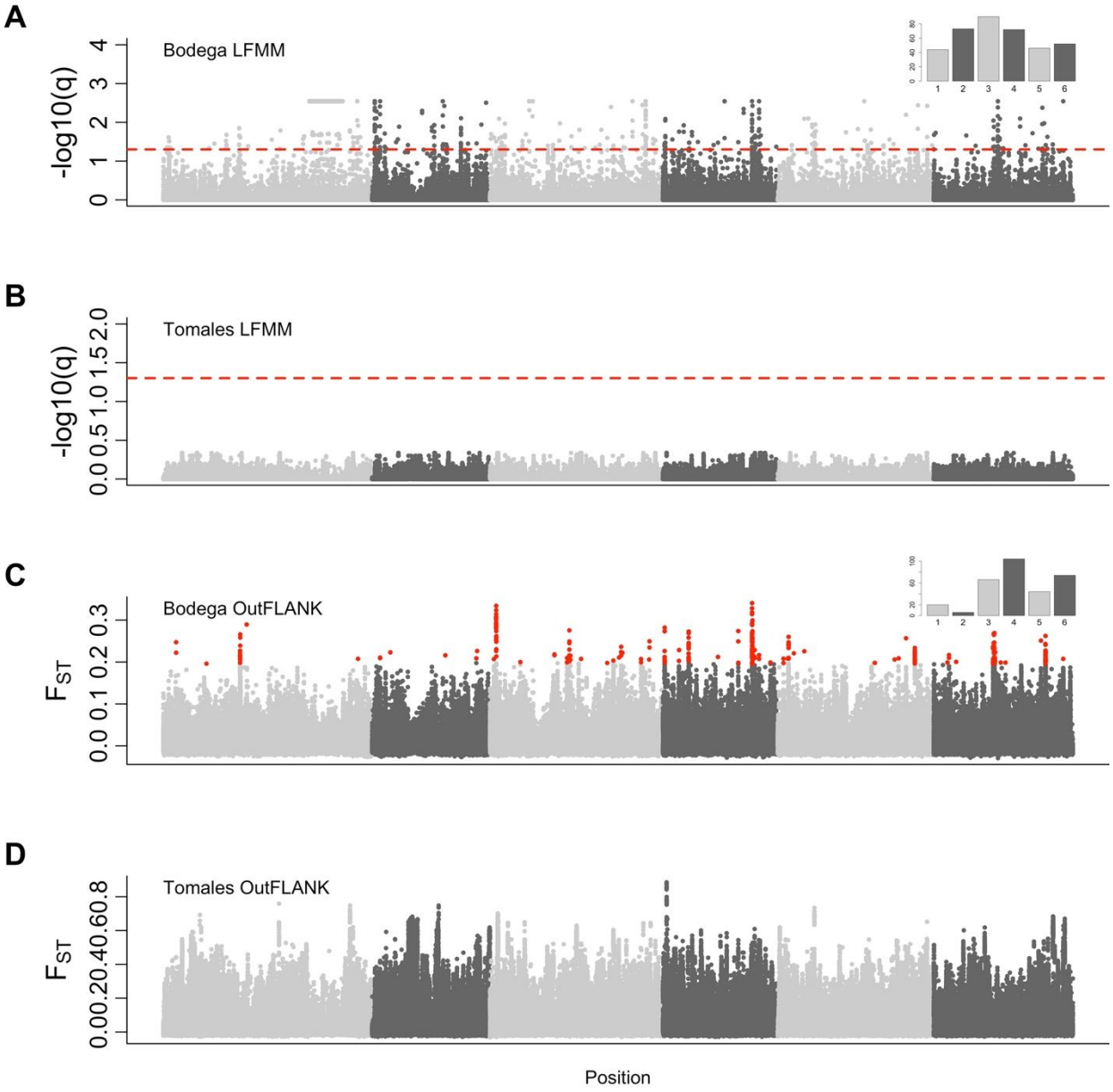
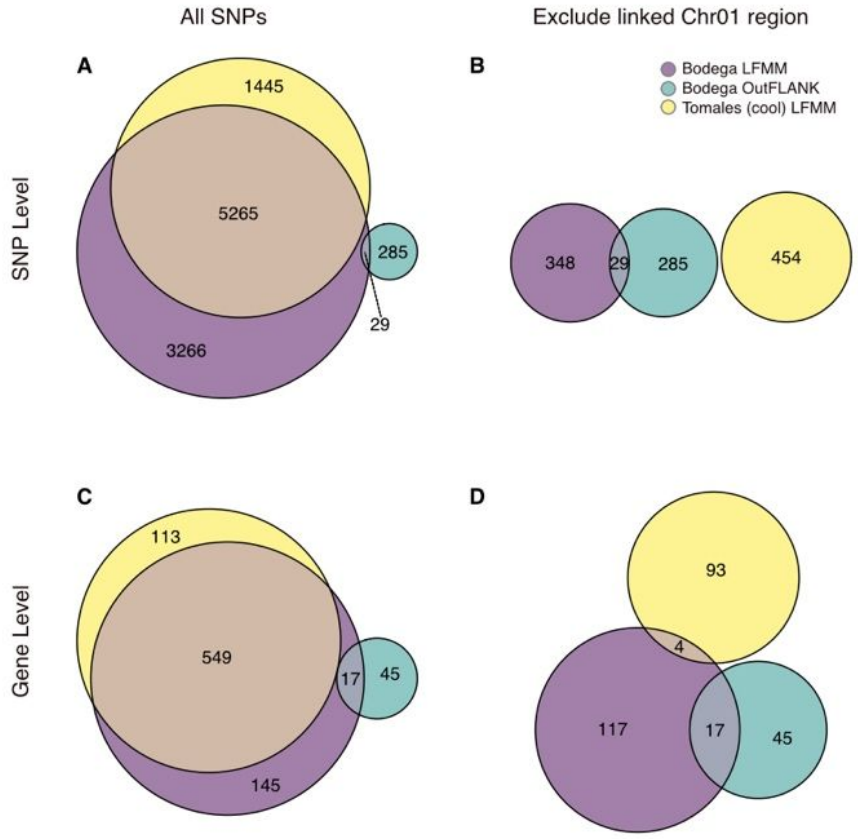


Figure 2. Manhattan plots show putative regions of interest from environmental association analyses using LFMM with average site temperature and OutFLANK comparing warmer versus cooler sites for Bodega and Tomales separately. For LFMM plots, the red line indicates *p*=0.001 and for the OutFLANK points with q<0.05 are highlighted in red. Insets show the number of outlier SNPs by chromosome, excluding the outliers contained in the large linked region on Chromosome 1.

675



676
677

678 **Figure 3.** Overlap among analyses for detecting SNP associations with temperature across two
679 adjacent embayments. Three sets of candidate SNPs are represented, those detected in Bodega
680 by LFMM, in Bodega by OutFLANK, and detected in Tomales (cool) by LFMM. Other analyses did
681 not contain significant SNPs at a threshold of q<0.05. (**A,B**) Overlap at the SNP level for all SNPs
682 (**A**) and excluding the highly linked region on Chromosome 1 (**B**). (**C,D**) Overlap in genes linked
683 to all candidate SNPs (**C**) and excluding the highly linked region on Chromosome 1 (**D**).

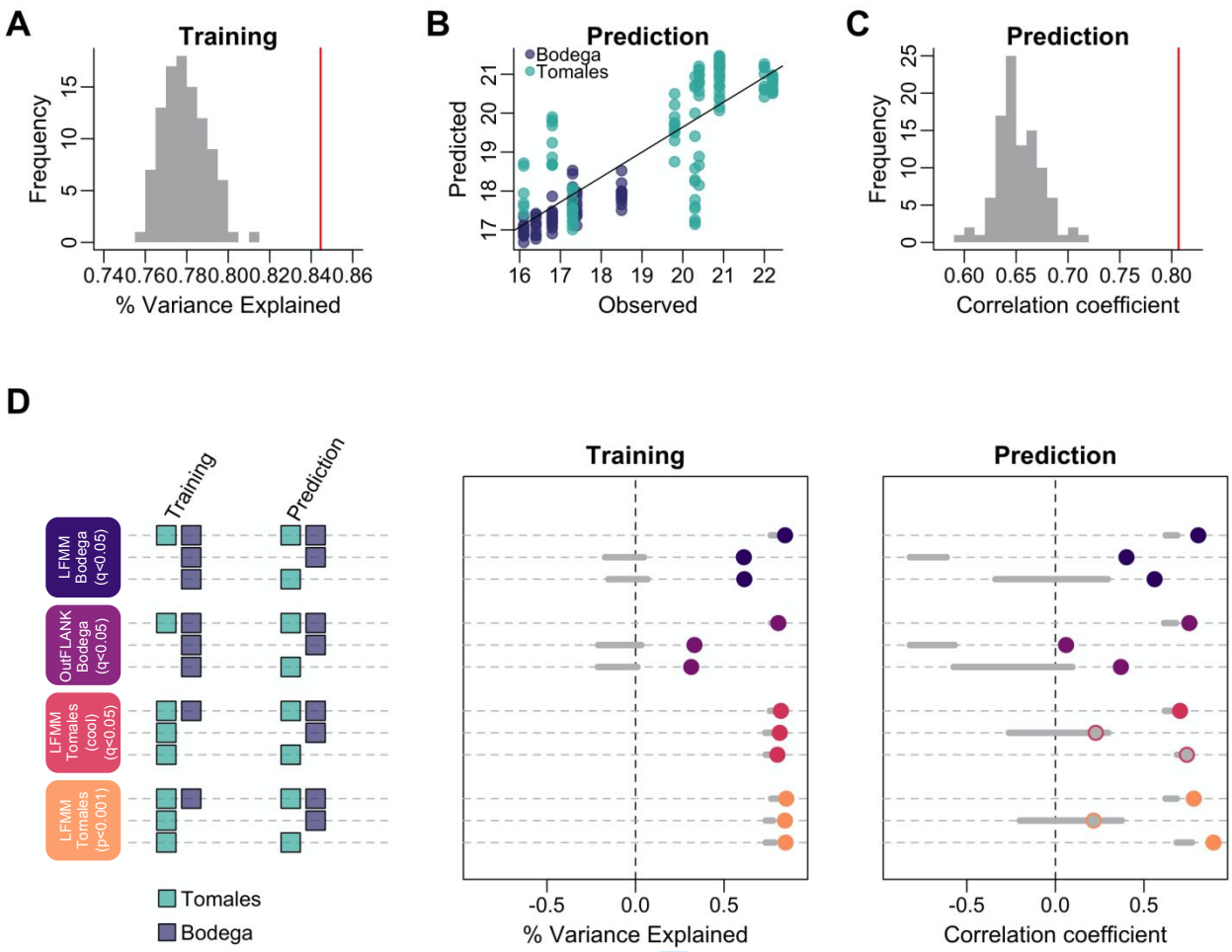


Figure 4. Polygenic analysis with random forest suggests limited predictability of thermal environment across bays. **(A-C)** An example run showing candidate SNPs from the LFMM analysis in Bodega Bay. **(A)** Variance explained by random forest using candidate SNPs as predictors (red line) vs. 100 runs with random SNPs (grey histogram). **(B)** Random forest predicted temperatures plotted against observed temperatures for both Bodega and Tomales samples. **(C)** Correlation coefficients between observed and random forest predicted temperature using candidate SNPs as predictors (red line) vs. 100 runs with random SNPs (grey histogram). **(D)** All combinations of SNP sets and training/prediction populations used in the random forest models. The first panel summarizes which SNP set was used (from top to bottom: LFMM Bodega, OutFLANK Bodega, LFMM Tomales (cool), LFMM Tomales; there is no OutFLANK for Tomales because this analysis did not yield any significant outliers) and which groups were used for training and prediction (i.e., trained using both embayments or only the local embayment to predict Bodega, Tomales, or both). Grey lines are 90% confidence intervals from the randomizations, filled dots are outside that (i.e., 'significant') and open dots are not significant.

References

Abbott, J. M., DuBois, K., Grosberg, R. K., Williams, S. L., & Stachowicz, J. J. (2018). Genetic distance predicts trait differentiation at the subpopulation but not the individual level in eelgrass, *Zostera marina*. *Ecology and Evolution*, *8*(15), 7476–7489. doi: 10.1002/ece3.4260

Alexa, A., & Rahnenführer, J. (2009). Gene set enrichment analysis with topGO. *Bioconductor Improv*, *27*, 1–26.

Aoki, L. R., Rappazzo, B., Beatty, D. S., Domke, L. K., Eckert, G. L., Eisenlord, M. E., … Harvell, C. D. (2022). Disease surveillance by artificial intelligence links eelgrass wasting disease to ocean warming across latitudes. *Limnology and Oceanography*, *67*(7), 1577–1589. doi: 10.1002/lno.12152

Arendt, J., & Reznick, D. (2008). Convergence and parallelism reconsidered: what have we learned about the genetics of adaptation? *Trends in Ecology & Evolution*, *23*(1), 26–32. doi: 10.1016/j.tree.2007.09.011

Bailey, S. F., Blanquart, F., Bataillon, T., & Kassen, R. (2017). What drives parallel evolution?: How population size and mutational variation contribute to repeated evolution. *BioEssays: News and Reviews in Molecular, Cellular and Developmental Biology*, *39*(1), 1–9. doi: 10.1002/bies.201600176

Bailleul, D., Stoeckel, S., & Arnaud-Haond, S. (2016). RClone: a package to identify MultiLocus Clonal Lineages and handle clonal data sets in r. *Methods in Ecology and Evolution / British Ecological Society*, *7*(8), 966–970. doi: 10.1111/2041-210x.12550

Bay, R. A., Rose, N., Barrett, R., Bernatchez, L., Ghalambor, C. K., Lasky, J. R., … Ralph, P. (2017). Predicting Responses to Contemporary Environmental Change Using Evolutionary Response Architectures. *The American Naturalist*, *189*(5), 463–473. doi: 10.1086/691233

Bergmann, N., Winters, G., Rauch, G., Eizaguirre, C., Gu, J., Nelle, P., … Reusch, T. B. H. (2010). Population-specificity of heat stress gene induction in northern and southern eelgrass *Zostera marina* populations under simulated global warming. *Molecular Ecology*, *19*(14), 2870–2883. doi: 10.1111/j.1365-294X.2010.04731.x

Blount, Z. D., Lenski, R. E., & Losos, J. B. (2018). Contingency and determinism in evolution: Replaying life's tape. *Science*, *362*(6415). doi: 10.1126/science.aam5979

Bohutínská, M., Vlček, J., Yair, S., Laenen, B., Konečná, V., Fracassetti, M., … Kolář, F. (2021). Genomic basis of parallel adaptation varies with divergence in *Arabidopsis* and its relatives. *Proceedings of the National Academy of Sciences of the United States of America*, *118*(21). doi: 10.1073/pnas.2022713118

Bolnick, D. I., Barrett, R. D. H., Oke, K. B., Rennison, D. J., & Stuart, Y. E. (2018). (Non)Parallel Evolution. *Annual Review of Ecology, Evolution, and Systematics*, *49*(1), 303–330. doi: 10.1146/annurev-ecolsys-110617-062240

Bushnell, B. (2021). *BBMap Short Read aligner, and Other Bioinformatic Tools* (Version 38.73). Retrieved from sourceforge.net/projects/bbmap/

Campanella, J. J., Bologna, P. A. X., Smith, S. M., Rosenzweig, E. B., & Smalley, J. V. (2010). *Zostera marina* population genetics in Barnegat Bay, New Jersey, and implications for grass bed restoration. *Population Ecology*, *52*(1), 181–190. doi: 10.1007/s10144-009-0170-4

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*, *51*(1), 245–269. doi: 10.1146/annurev-ecolsys-020720-042553

Cassin-Sackett, L., Callicrate, T. E., & Fleischer, R. C. (2019). Parallel evolution of gene classes, but not genes: Evidence from Hawai'ian honeycreeper populations exposed to avian malaria. *Molecular Ecology*, *28*(3), 568–583. doi: 10.1111/mec.14891

Caye, K., Deist, T. M., Martins, H., Michel, O., & François, O. (2016). TESS3: fast inference of spatial population structure and genome scans for selection. *Molecular Ecology Resources*, *16*(2), 540–548. doi: 10.1111/1755-0998.12471

Caye, K., Jumentier, B., Lepeule, J., François, O. (2019) LFMM 2: Fast and Accurate Inference of Gene-Environment Associations in Genome-Wide Studies. *Molecular Biology and Evolution*, 36(4), 852–860. https://doi.org/10.1093/molbev/msz008

Conte, G. L., Arnegard, M. E., Peichel, C. L., & Schluter, D. (2012). The probability of genetic parallelism and convergence in natural populations. *Proceedings. Biological Sciences / The Royal Society*, *279*(1749), 5039–5047. doi: 10.1098/rspb.2012.2146

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., … 1000 Genomes Project Analysis Group. (2011). The variant call format and VCFtools. *Bioinformatics* , *27*(15), 2156–2158. doi: 10.1093/bioinformatics/btr330

Doyle, J. J., & Doyle, J. L. (1987). A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin*, *19*(1), 11–15. Retrieved from https://worldveg.tind.io/record/33886/

DuBois, K., Abbott, J. M., Williams, S. L., & Stachowicz, J. J. (2019). Relative performance of eelgrass genotypes shifts during an extreme warming event: disentangling the roles of multiple traits. *Marine Ecology Progress Series*, *615*, 67–77. doi: 10.3354/meps12914

DuBois, K., Pollard, K. N., Kauffman, B. J., Williams, S. L., & Stachowicz, J. J. (2022). Local adaptation in a marine foundation species: Implications for resilience to future global change. *Global Change Biology*, *28*(8), 2596–2610. doi: 10.1111/gcb.16080

DuBois, K., Williams, S. L., & Stachowicz, J. J. (2020). Previous exposure mediates the response of eelgrass to future warming via clonal transgenerational plasticity. *Ecology*, *101*(12), e03169. doi: 10.1002/ecy.3169

DuBois, K., Williams, S. L., & Stachowicz, J. J. (2021). Experimental Warming Enhances Effects of Eelgrass Genetic Diversity Via Temperature-Induced Niche Differentiation. *Estuaries and Coasts*, *44*(2), 545–557. doi: 10.1007/s12237-020-00827-9

Duffy, J. E., Stachowicz, J. J., Reynolds, P. L., Hovel, K. A., Jahnke, M., Sotka, E. E., ... & Olsen, J. L. (2022). A Pleistocene legacy structures variation in modern seagrass ecosystems. *Proceedings of the National Academy of Sciences*, *119*(32), e2121425119.

Ehlers, A., Worm, B., & Reusch, T. B. H. (2008). Importance of genetic diversity in eelgrass *Zostera marina* for its resilience to global warming. *Marine Ecology Progress Series*, *355*, 1–7. doi: 10.3354/meps07369

Exposito-Alonso, M., 500 Genomes Field Experiment Team, Burbano, H. A., Bossdorf, O., Nielsen, R., & Weigel, D. (2019). Natural selection on the *Arabidopsis thaliana* genome in present and future climates. *Nature*, *573*(7772), 126–129. doi: 10.1038/s41586-019-1520-9

Frichot, E., & François, O. (2015). LEA: An R package for landscape and ecological association

studies. *Methods in Ecology and Evolution*, *6*(8), 925–929. doi: 10.1111/2041-210X.12382
Hämmerli, A., & Reusch, T. B. H. (2002). Local adaptation and transplant dominance in genets of the marine clonal plant *Zostera marina*. *Marine Ecology Progress Series*, *242*, 111–118. doi: 10.3354/meps242111
Härer, A., Bolnick, D. I., & Rennison, D. J. (2021). The genomic signature of ecological divergence along the benthic-limnetic axis in allopatric and sympatric threespine stickleback. *Molecular Ecology*, *30*(2), 451–463. doi: 10.1111/mec.15746
Hollarsmith, J. A., Sadowski, J. S., Picard, M. M. M., Cheng, B., Farlin, J., Russell, A., & Grosholz, E. D. (2020). Effects of seasonal upwelling and runoff on water chemistry and growth and survival of native and commercial oysters. *Limnology and Oceanography*, *65*(2), 224–235. doi: 10.1002/lno.11293
Hughes, A. R., Stachowicz, J. J., & Williams, S. L. (2009). Morphological and physiological variation among seagrass (*Zostera marina*) genotypes. *Oecologia*, *159*(4), 725–733. doi: 10.1007/s00442-008-1251-3
Kamel, S. J., Hughes, A. R., Grosberg, R. K., & Stachowicz, J. J. (2012). Fine-scale genetic structure and relatedness in the eelgrass *Zostera marina*. *Marine Ecology Progress Series*, *447*, 127–137. doi: 10.3354/meps09447
Kern, E. M. A., & Langerhans, R. B. (2018). Urbanization drives contemporary evolution in stream fish. *Global Change Biology*, *24*(8), 3791–3803. doi: 10.1111/gcb.14115
Kimbro, D. L., Largier, J., & Grosholz, E. D. (2009). Coastal oceanographic processes influence the growth and size of a key estuarine species, the Olympia oyster. *Limnology and Oceanography*, *54*(5), 1425–1437. doi: 10.4319/lo.2009.54.5.1425
Kim, J. H., Kang, J. H., Jang, J. E., Choi, S. K., Kim, M. J., Park, S. R., & Lee, H. J. (2017). Population genetic structure of eelgrass (*Zostera marina*) on the Korean coast: Current status and conservation implications for future management. *PloS One*, *12*(3), e0174105. doi: 10.1371/journal.pone.0174105
Kim, J., & Mayfield, S. P. (1997). Protein disulfide isomerase as a regulator of chloroplast translational activation. *Science*, *278*(5345), 1954–1957. doi: 10.1126/science.278.5345.1954
Kollars, N.M., J.M. Abbott, and J.J. Stachowicz. 2022. Hidden biodiversity: Spatial mosaics of eelgrass genotypic diversity at the centimeter to meadow scale. *Ecology*. doi: 10.1002/ecy.3813
Kuang, B., Zhao, X., Zhou, C., Zeng, W., Ren, J., Ebert, B., … Wu, A.-M. (2016). Role of UDP-Glucuronic Acid Decarboxylase in Xylan Biosynthesis in *Arabidopsis*. *Molecular Plant*, *9*(8), 1119–1131. doi: 10.1016/j.molp.2016.04.013
Lenski, R. E. (2017). Experimental evolution and the dynamics of adaptation and genome evolution in microbial populations. *The ISME Journal*, *11*(10), 2181–2194. doi: 10.1038/ismej.2017.69
Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest. *R News*, *2*(3), 18–22. Retrieved from https://www.researchgate.net/file.PostFileLoader.html?id=57d984d196b7e421f8398993&assetKey=AS%3A406267404275720%401473873105695
Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* , *25*(14), 1754–1760. doi: 10.1093/bioinformatics/btp324

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., … 1000 Genome Project Data Processing Subgroup. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics* , *25*(16), 2078–2079. doi: 10.1093/bioinformatics/btp352

Liu, S., Ferchaud, A.-L., Grønkjaer, P., Nygaard, R., & Hansen, M. M. (2018). Genomic parallelism and lack thereof in contrasting systems of three-spined sticklebacks. *Molecular Ecology*, *27*(23), 4725–4743. doi: 10.1111/mec.14782

Magalhaes, I. S., Whiting, J. R., D'Agostino, D., Hohenlohe, P. A., Mahmud, M., Bell, M. A., … MacColl, A. D. C. (2021). Intercontinental genomic parallelism in multiple three-spined stickleback adaptive radiations. *Nature Ecology & Evolution*, *5*(2), 251–261. doi: 10.1038/s41559-020-01341-8

Ma, X., Olsen, J. L., Reusch, T. B. H., Procaccini, G., Kudrna, D., Williams, M., … Van de Peer, Y. (2021). Improved chromosome-level genome assembly and annotation of the seagrass, *Zostera marina* (eelgrass) [version 1; peer review: 2 approved]. *F1000Research*, *10*, 289. doi: 10.12688/f1000research.38156.1

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., … DePristo, M. A. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, *20*(9), 1297–1303. doi: 10.1101/gr.107524.110

Mérot, C., Berdan, E. L., Babin, C., Normandeau, E., Wellenreuther, M., & Bernatchez, L. (2018). Intercontinental karyotype-environment parallelism supports a role for a chromosomal inversion in local adaptation in a seaweed fly. *Proceedings. Biological Sciences / The Royal Society*, *285*(1881). doi: 10.1098/rspb.2018.0519

Mérot, C., Berdan, E. L., Cayuela, H., Djambazian, H., Ferchaud, A.-L., Laporte, M., … Bernatchez, L. (2021). Locally Adaptive Inversions Modulate Genetic Variation at Different Geographic Scales in a Seaweed Fly. *Molecular Biology and Evolution*, *38*(9), 3953–3971. doi: 10.1093/molbev/msab143

Moon, J., Zhu, L., Shen, H., & Huq, E. (2008). PIF1 directly and indirectly regulates chlorophyll biosynthesis to optimize the greening process in *Arabidopsis*. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(27), 9433–9438. doi: 10.1073/pnas.0803611105

Muñiz-Salazar, R., Talbot, S. L., Sage, G. K., Ward, D. H., & Cabello-Pasini, A. (2005). Population genetic structure of annual and perennial populations of *Zostera marina* L. along the Pacific coast of Baja California and the Gulf of California. *Molecular Ecology*, *14*(3), 711–722. doi: 10.1111/j.1365-294X.2005.02454.x

Oke, K. B., Rolshausen, G., LeBlond, C., & Hendry, A. P. (2017). How Parallel Is Parallel Evolution? A Comparative Analysis in Fishes. *The American Naturalist*, *190*(1), 1–16. doi: 10.1086/691989

Olsen, J. L., Rouzé, P., Verhelst, B., Lin, Y.-C., Bayer, T., Collen, J., … Van de Peer, Y. (2016). The genome of the seagrass *Zostera marina* reveals angiosperm adaptation to the sea. *Nature*, *530*(7590), 331–335. doi: 10.1038/nature16548

Olsen, J. L., Stam, W. T., Coyer, J. A., Reusch, T. B., Billingham, M., Boström, C., ... & Wyllie-Echeverria, S. (2004). North Atlantic phylogeography and large-scale population differentiation of the seagrass *Zostera marina* L. *Molecular Ecology*, *13*(7), 1923-1941.

Ort, B. S., Cohen, C. S., Boyer, K. E., & Wyllie-Echeverria, S. (2012). Population structure and

genetic diversity among eelgrass (*Zostera marina*) beds and depths in San Francisco Bay. *The Journal of Heredity*, *103*(4), 533–546. doi: 10.1093/jhered/ess022

Pickersgill, B. (2018). Parallel vs. Convergent Evolution in Domestication and Diversification of Crops in the Americas. *Frontiers in Ecology and Evolution*, *6*. doi: 10.3389/fevo.2018.00056

Prunier, J., Lemaçon, A., Bastien, A., Jafarikia, M., Porth, I., Robert, C., & Droit, A. (2019). LD-annot: A Bioinformatics Tool to Automatically Provide Candidate SNPs With Annotations for Genetically Linked Genes. *Frontiers in Genetics*, *10*, 1192. doi: 10.3389/fgene.2019.01192

Qin, L.-Z., Kim, S. H., Song, H.-J., Kim, H. G., Suonan, Z., Kwon, O., … Lee, K.-S. (2020). Long-term variability in the flowering phenology and intensity of the temperate seagrass *Zostera marina* in response to regional sea warming. *Ecological Indicators*, *119*, 106821. doi: 10.1016/j.ecolind.2020.106821

Ralph, P., & Coop, G. (2010). Parallel adaptation: one or many waves of advance of an advantageous allele? *Genetics*, *186*(2), 647–668. doi: 10.1534/genetics.110.119594

Ralph, P., & Coop, G. (2015). Convergent Evolution During Local Adaptation to Patchy Landscapes. *PLoS Genetics*, *11*(11), e1005630. doi: 10.1371/journal.pgen.1005630

R Core Team. (2020). *R: A language and environment for statistical computing*. Vienna, Austria.

Rennison, D. J., Delmore, K. E., Samuk, K., Owens, G. L., & Miller, S. E. (2020). Shared Patterns of Genome-Wide Differentiation Are More Strongly Predicted by Geography Than by Ecology. *The American Naturalist*, *195*(2), 192–200. doi: 10.1086/706476

Reusch, T. B. H., Stam, W. T. & Olsen, J. L. (1999) Microsatellite loci in eelgrass *Zostera marina* reveal marked polymorphism within and among populations. *Molecular Ecology*, 8, 317–322.

Reynolds, L. K., DuBois, K., Abbott, J. M., Williams, S. L., & Stachowicz, J. J. (2016). Response of a Habitat-Forming Marine Plant to a Simulated Warming Event Is Delayed, Genotype Specific, and Varies with Phenology. *PloS One*, *11*(6), e0154532. doi: 10.1371/journal.pone.0154532

Reynolds, L.K., Stachowicz, J.J., Hughes, A.R., Kamel, S.J., Ort, B. and R.H. Grosberg. (2017). Temporal stability in patterns of genetic diversity and structure of a marine foundation species (*Zostera marina*). *Heredity*, *118*:404–412.

Richardson, J. L., Urban, M. C., Bolnick, D. I., & Skelly, D. K. (2014). Microgeographic adaptation and the spatial scale of evolution. *Trends in Ecology & Evolution*, *29*(3), 165–176. doi: 10.1016/j.tree.2014.01.002

Rivas, M. J., Saura, M., Pérez-Figueroa, A., Panova, M., Johansson, T., André, C., … Quesada, H. (2018). Population genomics of parallel evolution in gene expression and gene sequence during ecological adaptation. *Scientific Reports*, *8*(1), 16147. doi: 10.1038/s41598-018-33897-8

Roda, F., Ambrose, L., Walter, G. M., Liu, H. L., Schaul, A., Lowe, A., … Ortiz-Barrientos, D. (2013). Genomic evidence for the parallel evolution of coastal forms in the *Senecio lautus* complex. *Molecular Ecology*, *22*(11), 2941–2952. doi: 10.1111/mec.12311

Rosenblum, E. B., Römpler, H., Schöneberg, T., & Hoekstra, H. E. (2010). Molecular and functional basis of phenotypic convergence in white lizards at White Sands. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(5), 2113–2117. doi: 10.1073/pnas.0911042107

Ruckelshaus, M. H. (1996). Estimation of genetic neighborhood parameters from pollen and seed dispersal in the marine angiosperm *Zostera marina* L. *Evolution; International Journal of Organic Evolution*, *50*(2), 856–864. doi: 10.1111/j.1558-5646.1996.tb03894.x

Saha, M., Barboza, F. R., Somerfield, P. J., Al-Janabi, B., Beck, M., Brakel, J., … Sawall, Y. (2020). Response of foundation macrophytes to near-natural simulated marine heatwaves. *Global Change Biology*, *26*(2), 417–430. doi: 10.1111/gcb.14801

Salisbury, S. J., McCracken, G. R., Perry, R., Keefe, D., Layton, K. K. S., Kess, T., … Ruzzante, D. E. (2020). Limited genetic parallelism underlies recent, repeated incipient speciation in geographically proximate populations of an Arctic fish (*Salvelinus alpinus*). *Molecular Ecology*, *29*(22), 4280–4294. doi: 10.1111/mec.15634

Salo, T., Reusch, T. B. H., & Boström, C. (2015). Genotype-specific responses to light stress in eelgrass *Zostera marina*, a marine foundation plant. *Marine Ecology Progress Series*, *519*, 129–140. doi: 10.3354/meps11083

Schiebelhut, L. M., Grosberg, R. K., Stachowicz, J. J., & Bay, R. A. (2022a). *BioProject PRJNA887384. National Center for Biotechnology Information Sequence Read Archive* [Data set]. Retrieved from http://www.ncbi.nlm.nih.gov/bioproject/887384

Schiebelhut, L. M., Grosberg, R. K., Stachowicz, J. J., & Bay, R. A. (2022b). *Data and code associated with "Genomic responses to parallel selection in the eelgrass Zostera marina in adjacent bays". Dryad Digital Repository* [Data set]. Dryad Digital Repository. https://doi.org/10.6071/M3DD4F

Severance, S., & Hamza, I. (2009). Trafficking of heme and porphyrins in metazoa. *Chemical Reviews*, *109*(10), 4596–4616. doi: 10.1021/cr9001116

Smale, D. A., Wernberg, T., Oliver, E. C. J., Thomsen, M., Harvey, B. P., Straub, S. C., … Moore, P. J. (2019). Marine heatwaves threaten global biodiversity and the provision of ecosystem services. *Nature Climate Change*, *9*(4), 306–312. doi: 10.1038/s41558-019-0412-1

Stuart, Y. E. (2019). Divergent Uses of "Parallel Evolution" during the History of The American Naturalist. *The American Naturalist*, *193*(1), 11–19. doi: 10.1086/700718

Stuart, Y. E., Veen, T., Weber, J. N., Hanson, D., Ravinet, M., Lohman, B. K., … Bolnick, D. I. (2017). Contrasting effects of environment and genetics generate a continuum of parallel evolution. *Nature Ecology & Evolution*, *1*(6), 158. doi: 10.1038/s41559-017-0158

Szendro, I. G., Franke, J., de Visser, J. A. G. M., & Krug, J. (2013). Predictability of evolution depends nonmonotonically on population size. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(2), 571–576. doi: 10.1073/pnas.1213613110

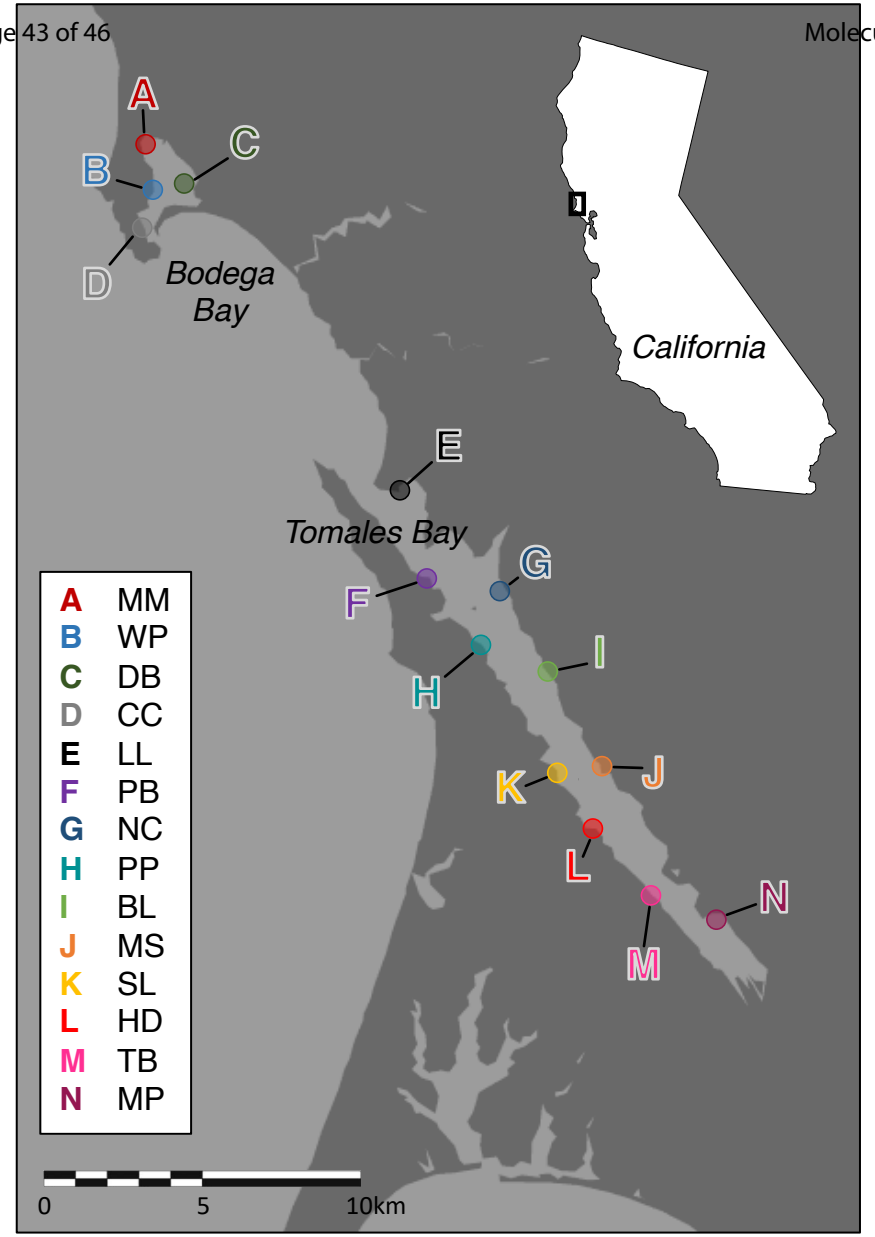
Tenaillon, O., Rodríguez-Verdugo, A., Gaut, R. L., McDonald, P., Bennett, A. F., Long, A. D., & Gaut, B. S. (2012). The molecular diversity of adaptive convergence. *Science*, *335*(6067), 457–461. doi: 10.1126/science.1212986

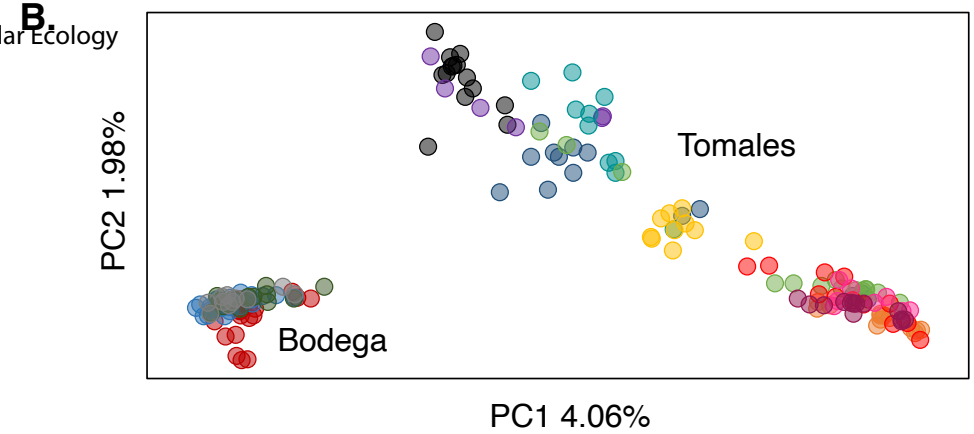
Thompson, M. J., & Jiggins, C. D. (2014). Supergenes and their role in evolution. *Heredity*, *113*(1), 1–8. doi: 10.1038/hdy.2014.20

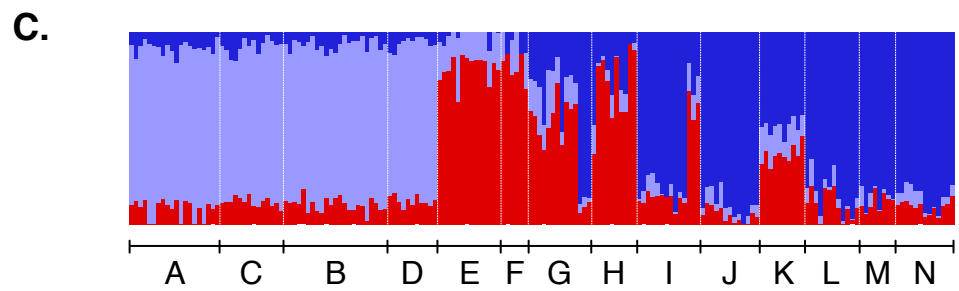
Van der Auwera, G. A., Carneiro, M. O., Hartl, C., Poplin, R., Del Angel, G., Levy-Moonshine, A., … DePristo, M. A. (2013). From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Current Protocols in Bioinformatics / Editoral Board, Andreas D. Baxevanis ... [et Al.]*, *43*, 11.10.1–11.10.33. doi: 10.1002/0471250953.bi1110s43

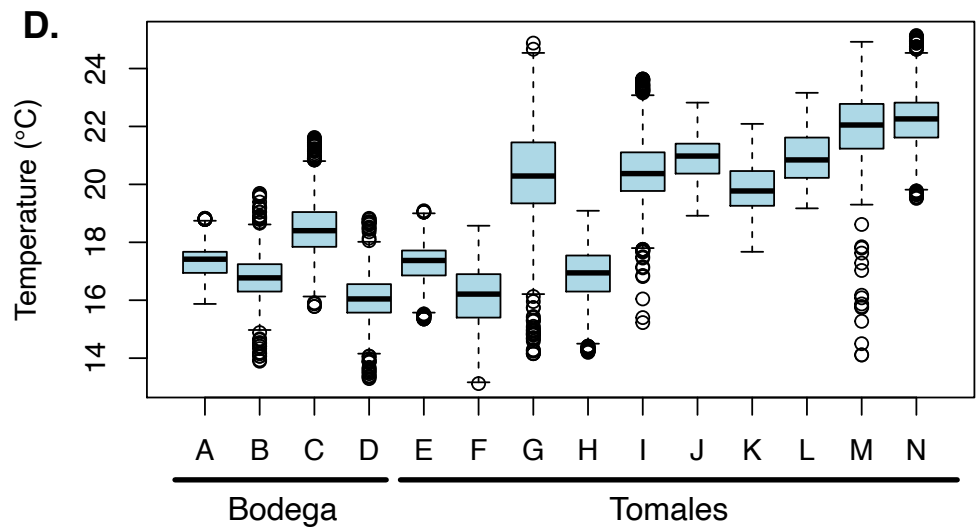
Weir, B. S., & Cockerham, C. C. (1984). Estimating F-Statistics for the Analysis of Population

963 Structure. *Evolution; International Journal of Organic Evolution*, *38*(6), 1358–1370. doi:
964 10.2307/2408641
965 Wellenreuther, M., & Bernatchez, L. (2018). Eco-Evolutionary Genomics of Chromosomal
966 Inversions. *Trends in Ecology & Evolution*, *33*(6), 427–440. doi: 10.1016/j.tree.2018.04.002
967 Whitlock, M. C., & Lotterhos, K. E. (2015). Reliable Detection of Loci Responsible for Local
968 Adaptation: Inference of a Null Model through Trimming the Distribution of F(ST). *The*
969 *American Naturalist*, *186 Suppl 1*, S24–S36. doi: 10.1086/682949
970 Yu, L., Stachowicz, J.J., Dubous, K. and Reusch, T.B.H. (2022). Detecting clonemate pairs in
971 multicellular diploid clonal species based on a shared heterozygosity index. *Molecular*
972 *Ecology Resources*, doi: 10.1111/1755-0998.13736
973 Zheng, X., Levine, D., Shen, J., Gogarten, S. M., Laurie, C., & Weir, B. S. (2012). A high-
974 performance computing toolset for relatedness and principal component analysis of SNP
975 data. *Bioinformatics* , *28*(24), 3326–3328. doi: 10.1093/bioinformatics/bts606


A.
A
B
C
D
Bodega Bay
California
E
Tomales Bay
F
G
H
I
J
K
L
M
N
A MM
B WP
C DB
D CC
E LL
F PB
G NC
H PP
I BL
J MS
K SL
L HD
M TB
N MP
0
5
10km


B.
PC2 1.98%
Tomales
Bodega
PC1 4.06%


C.
A C B D E F G H I J K L M N


D.
Temperature (°C)
14 16 18 20 22 24
A B C D E F G H I J K L M N
Bodega
Tomales

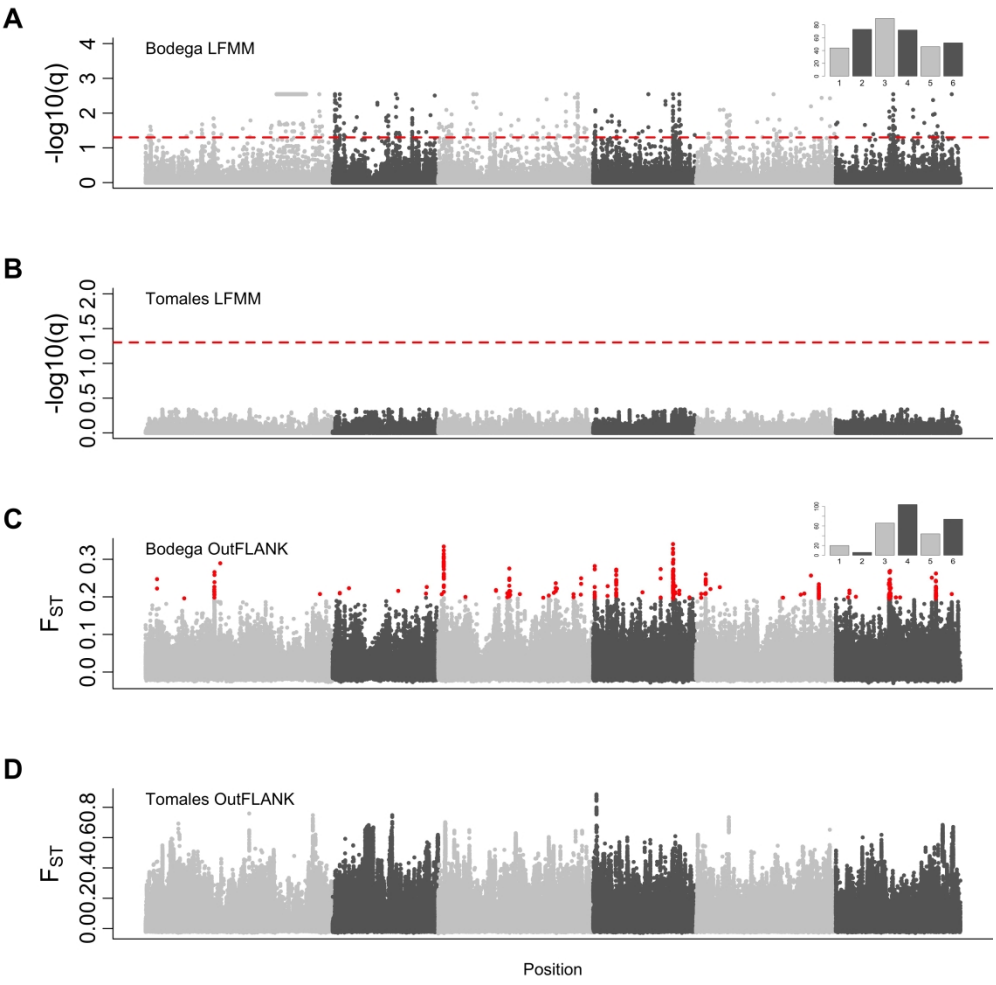


Figure 2. Manhattan plots show putative regions of interest from environmental association analyses using LFMM with average site temperature and OutFLANK comparing warmer versus cooler sites for Bodega and Tomales separately. For LFMM plots, the red line indicates p=0.001 and for the OutFLANK points with q<0.05 are highlighted in red. Insets show the number of outlier SNPs by chromosome, excluding the outliers contained in the large linked region on Chromosome 1.

1763x1763mm (72 x 72 DPI)

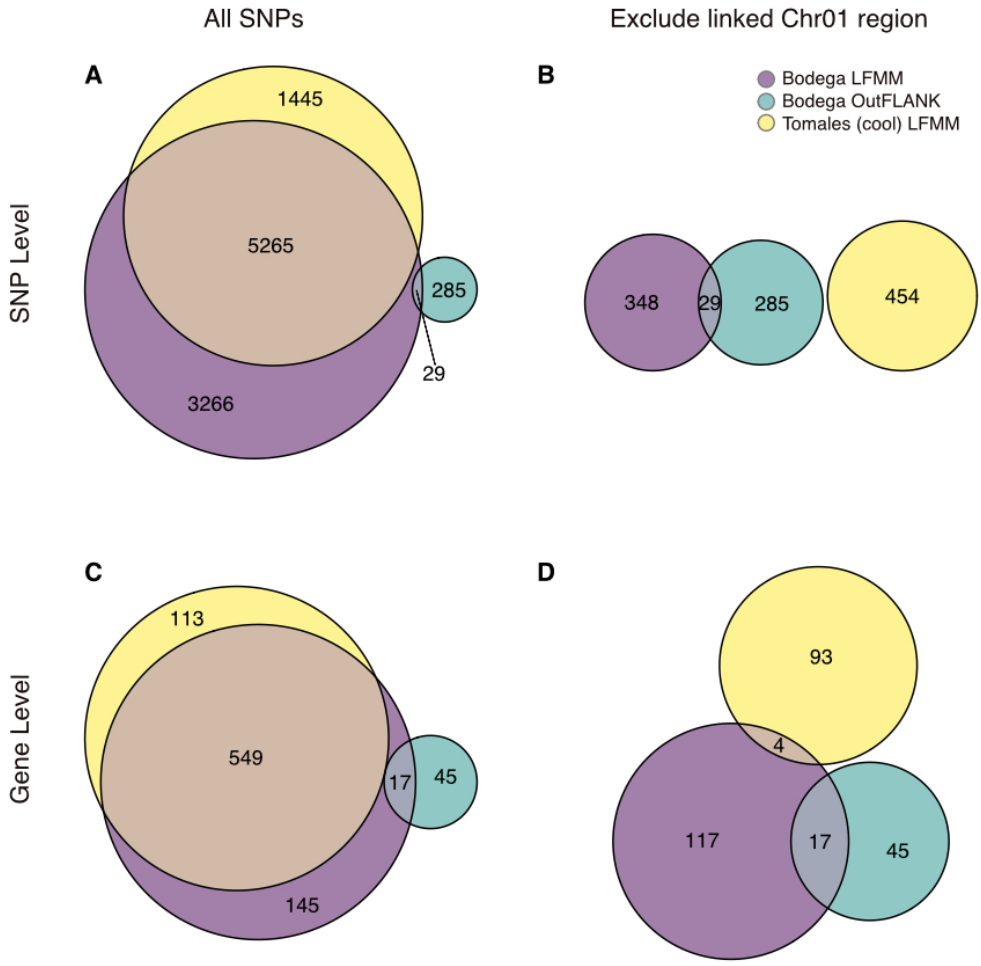
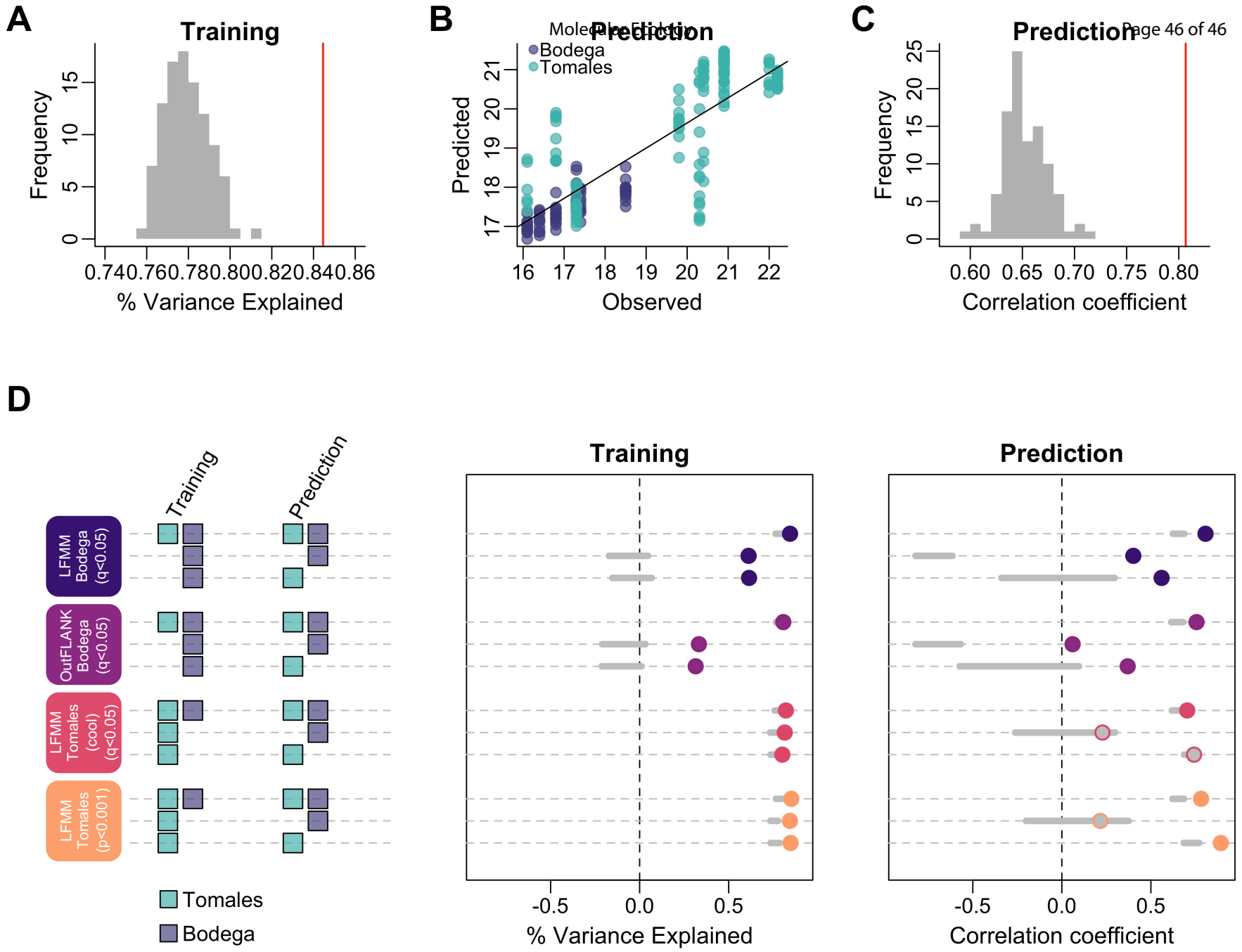


Figure 3. Overlap among analyses for detecting SNP associations with temperature across two adjacent embayments. Three sets of candidate SNPs are represented, those detected in Bodega by LFMM, in Bodega by OutFLANK, and detected in Tomales (cool) by LFMM. Other analyses did not contain significant SNPs at a threshold of $q<0.05$. (A,B) Overlap at the SNP level for all SNPs (A) and excluding the highly linked region on Chromosome 1 (B). (C,D) Overlap in genes linked to all candidate SNPs (C) and excluding the highly linked region on Chromosome 1 (D).

304x304mm (72 x 72 DPI)


A
Training
Frequency
0
5
10
15
0.74
0.76
0.78
0.80
0.82
0.84
0.86
% Variance Explained
B
Prediction
Bodega
Tomales
Predicted
17
18
19
20
21
16
17
18
19
20
21
22
Observed
C
Prediction
Frequency
0
5
10
15
20
25
0.60
0.65
0.70
0.75
0.80
Correlation coefficient
D
Training
Prediction
LFMM Bodega (q<0.05)
OutFLANK Bodega (q<0.05)
LFMM Tomales (cool) (q<0.05)
LFMM Tomales (p<0.001)
Tomales
Bodega
Training
-0.5
0.0
0.5
% Variance Explained
Prediction
-0.5
0.0
0.5
Correlation coefficient