MOLECULAR ECOLOGY WILEY

From tides to nucleotides: Genomic signatures of adaptation to environmental heterogeneity in barnacles

Joaquin C. B. Nunez[1] | Stephen Rong[1,2] | David A. Ferranti[1] |
Alejandro Damian-Serrano[3] | Kimberly B. Neil[1] | Henrik Glenner[4,7] |
Rebecca G. Elyanow[1,2] | Bianca R. P. Brown[1] | Magnus Alm Rosenblad[5] |
Anders Blomberg[5] | Kerstin Johannesson[6] | David M. Rand[1,2]

[1]Department of Ecology and Evolutionary Biology, Brown University, Providence, RI, USA

[2]Center for Computational Molecular Biology, Brown University, Providence, RI, USA

[3]Department of Ecology and Evolutionary Biology, Yale University, New Haven, CT, USA

[4]Department of Biological Sciences, University of Bergen, Bergen, Norway

[5]Department of Chemistry and Molecular Biology, University of Gothenburg, Lundberg Laboratory, Göteborg, Sweden

[6]Department of Marine Sciences, University of Gothenburg, Tjärnö Marine Laboratory, Strömstad, Sweden

[7]Center of Macroecology and Climate, GLOBE, University of Copenhagen, Copenhagen, Denmark

Correspondence
Joaquin C. B. Nunez and David M. Rand, Department of Ecology and Evolutionary Biology, Brown University, 80 Waterman Street, 02912 Providence, RI, USA.
Emails: joaquin_nunez@brown.edu (JCBN); david_rand@brown.edu (DMR)

Present address
Joaquin C. B. Nunez, Department of Biology, University of Virginia, Charlottesville, VA, USA

Abstract

The northern acorn barnacle (*Semibalanus balanoides*) is a robust system to study the genetic basis of adaptations to highly heterogeneous environments. Adult barnacles may be exposed to highly dissimilar levels of thermal stress depending on where they settle in the intertidal (i.e., closer to the upper or lower tidal boundary). For instance, barnacles near the upper tidal limit experience episodic summer temperatures above recorded heat coma levels. This differential stress at the microhabitat level is also dependent on the aspect of sun exposure. In the present study, we used pool-seq approaches to conduct a genome wide screen for loci responding to intertidal zonation across the North Atlantic basin (Maine, Rhode Island, and Norway). Our analysis discovered 382 genomic regions containing SNPs which are consistently zonated (i.e., SNPs whose frequencies vary depending on their position in the rocky intertidal) across all surveyed habitats. Notably, most zonated SNPs are young and private to the North Atlantic. These regions show high levels of genetic differentiation across ecologically extreme microhabitats concomitant with elevated levels of genetic variation and Tajima's *D*, suggesting the action of non-neutral processes. Overall, these findings support the hypothesis that spatially heterogeneous selection is a general and repeatable feature for this species, and that natural selection can maintain functional genetic variation in heterogeneous environments.

1 | INTRODUCTION

Understanding how natural populations evolve and adapt to heterogeneous environments is a longstanding goal in evolutionary genetics. Classical theory posits that populations should be locally adapted to their environments in order to maximize fitness (Williams, 1966). However, natural populations inhabiting highly variable environments defy this expectation. If selection in these environments were to proceed without gene flow, all microhabitats would be fixed for the favoured allele(s) and might show less variation than homogenous habitats. Conversely, if gene flow is high, organisms will experience a dynamic fitness landscape in which fitness maximization becomes a problem (Bergland et al., 2014; Hedrick, 1976; Machado et al., 2019; Powell & Taylor, 1979). In these cases, populations will experience mosaics of heterogeneity in which alleles may be both advantageous and deleterious as a function of time and space. In diploids, this can result in marginal overdominance, a phenomenon where the heterozygous genotype displays higher harmonic (for spatial heterogeneity) or geometric (for temporal heterogeneity) mean fitness than either homozygote genotype with concomitant long-term maintenance of genetic polymorphisms via balancing selection (see Fijarczyk & Babik, 2015; Jamie & Meier, 2020). Over the past six decades, the role of adaptive maintenance of polymorphism on natural populations has been subject to extensive and contentious discussion. Much of this debate has been centred around three foci. First, many authors have argued for a transient role of polymorphism in adaptation (Asthana et al.,2005; Bubb et al., 2006; Kimura & Ohta, 1971). Second, classical models of population genetics have posited that, in the absence of loci interactions, widespread balancing selection is unplausible as it leads to unbearable amounts of genetic death (Kimura & Crow, 1964; King, 1967; Lewontin & Hubby, 1966; Milkman, 1967; Sved et al., 1967). Third, most models of maintenance of polymorphism have very restrictive assumptions. For example, the Levene (1953) model of spatially heterogeneous selection commonly allows for maintenance of variation when there is a large number of ecological niches available, when there is reduced migration among these niches, and when individuals can choose niches where their fitness is maximized (Hedrick et al., 1976). Empirical tests of these assumptions in natural populations present considerable challenges. Despite the long intellectual tradition associated with these questions, the evolutionary conditions leading to adaptations to

highly variable environments remain poorly understood in natural populations. Nevertheless, the recent expansion of high-throughput sequencing and the growing availability of reference genomes for non-model species has provided new avenues to revisit, and potentially resolve, this old and important conundrum (e.gBergland et al., 2014; Fijarczyk & Babik, 2015; Machado et al., 2019; Wagner et al., 2017; Westram et al., 2018).

Natural populations of the northern acorn barnacle (*Semibalanus balanoides*) provide an ideal system in which to investigate balancing selection in spatially heterogeneous environments. *S. balanoides* inhabits the rocky intertidal, a highly heterogeneous ecosystem in which multiple environmental stressors can be easily tracked (Brown et al., 2020; Holm & Bourget, 1994; Schmidt et al., 2000, 2008; Schmidt & Rand, 1999, 2001; Veliz et al., 2004). Individual barnacles mate yearly, followed by brooding in September–December, and subsequent hatching and release during the winter plankton blooms (Barnes & Barnes, 1959; Crisp, 1968). At this stage, larvae have high dispersal capabilities, swimming freely in the water column for up to 5 weeks from the point of spawning (Flowerdew, 1983). While young cyprids (i.e., juvenile free-swimming barnacles) have the capacity to survey potential habitats for ideal substrates (Bertness et al., 1992), once settlement has occurred, barnacles transition to a fully sessile life habit by growing a calcareous shell. *S. balanoides* recruits to a wide range of microhabitats extending from the upper to the lower intertidal zone. While spatially adjacent, each microhabitat displays striking differences in terms of abiotic heterogeneity (Figure 1a,b). For instance, during the summer months, individuals settling on the lower tidal microhabitat experience consistently lower thermal stress relative to upper tidal microhabitats (Nunez, Flight, et al., 2020; Schmidt & Rand, 1999). Differential exposure to the sun also generates microhabitat heterogeneity: in the northern hemisphere, the southwestern facing shores experience higher thermal stress relative to southeastern facing shores (Schmidt & Rand, 1999). These differential ecological pressures have drastic consequences on both life-history (Bertness et al., 1991) and morphological traits in barnacles (Figure 1c; Figure S1). Moreover, the stark contrast between strong selection at the adult stage and the free-swimming planktonic larvae results in the problem of ecological genetic load (Nunez, Rong, et al., 2020). That is, since larvae often settle in new habitats every generation, any

FIGURE 1 Ecological characterization of intertidal habitats in the North Pacific. (a) Model of abiotic stress due to intertidal zonation. (b) Model of abiotic stress due to sun aspect. (c) PCA of the size opercular plates of barnacles collected in upper (red) and lower (blue) intertidal microhabitats. Temperature profiles of the intertidal in the rocky shore in (d) Jamestown, RI. (e) Herdla, Norway. and (f) Tjärnö, Sweden. The x-axis shows hour of day from 10 am to 6 pm. The boxplots show all temperatures collected for that given hour during the period of time indicated in each graph. Microhabitats are indicated by colour: L*c* (blue), H*c* (light red; only shown for RI and SV), and H*h* (dark red)
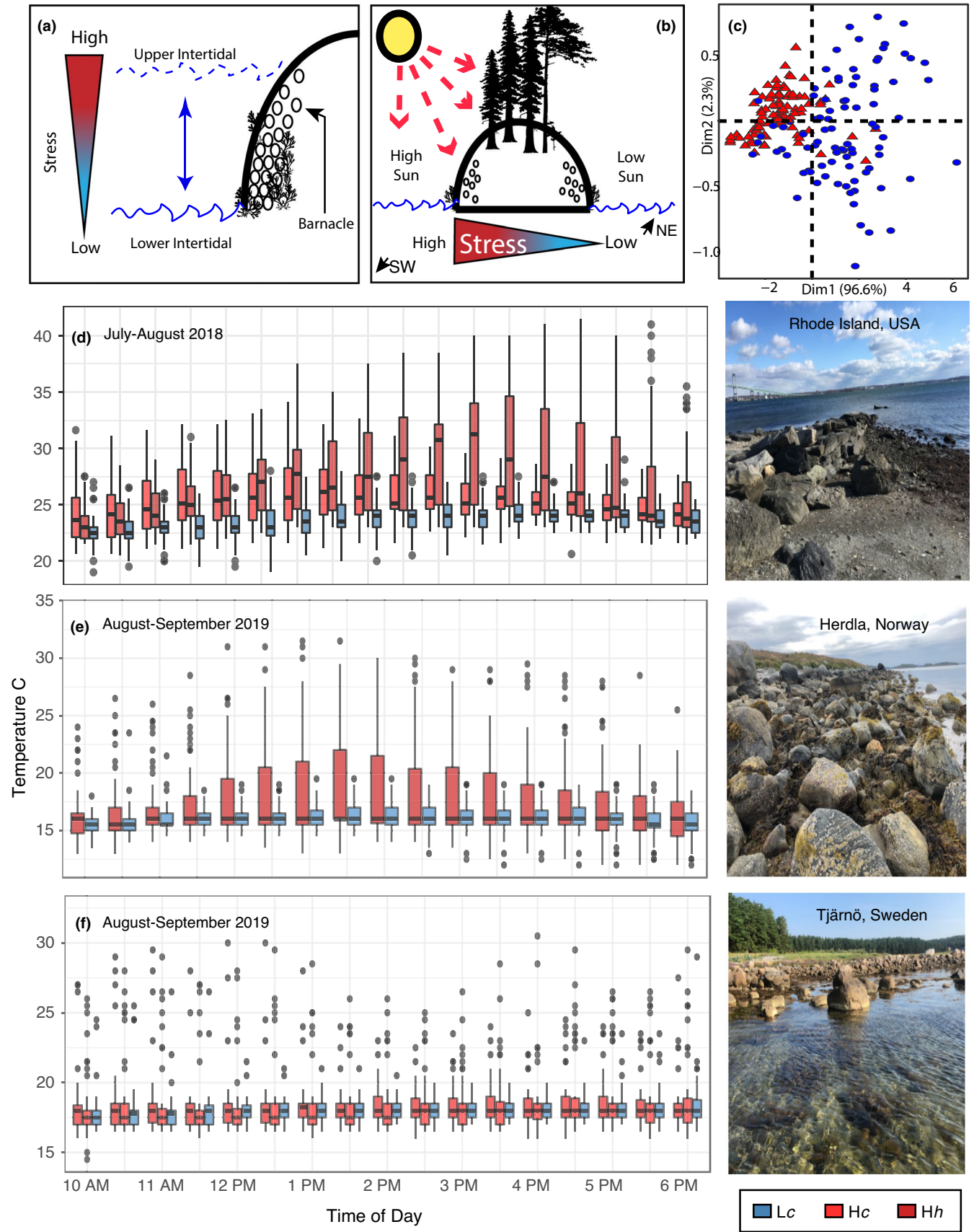
beneficial allele inherited from its parent may be neutral or deleterious, and, consequentially, fitness maximization within a microhabitat becomes a problem.

Previous work in *Semibalanus* has argued that the life history of this species is a good fit for the Levene (1953) model of spatially heterogeneous selection. For example, in their recent paper, Nunez,

Rong, et al., (2020) revealed the existence of hundreds of loci across the genome which harbour trans-species polymorphisms and show molecular signatures canonically associated with long-term balancing selection. As such, the study provides evidence supporting the hypothesis that spatially heterogeneous selection plays a key role in the maintenance of adaptation. Despite these discoveries, a fundamental question remains as to what particular ecological stressors, and spatial scales, produce the selection gradients leading to these genomic signatures. To date, only one gene displaying conspicuous signatures of balancing selection has been comprehensively characterized in the barnacle. This gene is the metabolic locus mannose-6 phosphate isomerase (*Mpi*). *Mpi* displays alleles whose frequencies vary spatially, conditional on their position in the rocky intertidal (i.e., the locus is "zonated"). Moreover, ecological and physiological experiments have shown that this allele zonation is associated with variations in fitness, indicating that zonation at this locus is caused by spatially heterogeneous selection (Nunez, Flight, et al., 2020; Schmidt et al., 2000; Schmidt & Rand, 1999, 2001). Notably, the zonation patterns of *Mpi* are not uniformly observed across the North Atlantic. For example, the gene shows zonation across intertidal microhabitats in the Gulf of Maine (Schmidt et al., 2000; Schmidt & Rand, 1999), at the level of estuaries in the gulf of St. Lawrence (Canada) (Dufresne et al., 2002; Veliz et al., 2004), and shows conflicting signs of zonation in the Narragansett Bay (Rhode Island; i.e., some findings suggests that the zonation pattern is reversed, relative to Maine; others suggest there is no zonation at all) (Nunez, Flight, et al., 2020; Rand et al., 2002).

This study seeks to provide us with a solid understanding of what loci, other than *Mpi*, may experience spatially heterogeneous selection across the habitat of *Semibalanus*. Particularly, we tackle the following questions: how many genes show repeatable patterns of zonation across the North Atlantic habitat of the barnacle? Is zonation localized to particular areas in the genome? Is the genetic basis of ecological zonation fueled by young, short-lived, alleles, or by old, balanced polymorphisms (e.g., those discovered in Nunez, Rong, et al., 2020)? While *S. balanoides* experiences ecological stressors at multiple spatial scales, we have chosen to focus our study at the spatial scale of the rocky intertidal since it has been exhaustively characterized in the species. For our study, we used pool-seq (Schlotterer et al., 2014a, 2014b) to characterize genetic variation across intertidal extremes of the rocky shore. We conducted all analyses in populations collected from Maine (ME), Rhode Island (RI), and Norway (Norge; NO). We also included samples from Sweden (Sverige; SV), collected at a site which does not seem to experience drastic thermal zonation across intertidal microhabitats due to small and irregular sea-level variation. We discovered over 383 regions across the genome that exhibit repeatable intertidal zonation at these sites, spanning the North Atlantic. Moreover, we show that the genetic footprints around these loci suggest zonation is driven by spatially heterogeneous selection. Finally, we discuss the ecological consequences of this zonated variation to understand how species adapt to highly variable and changing environments.

2 | MATERIALS AND METHODS

2.1 | Barnacle sampling

Barnacle samples were collected from Damariscotta (ME, United States), Jamestown (RI, United States), Herdla (Hordaland, Norway), and Tjärnö (Strömstad, Sweden). Samples were collected from microhabitats at the edges of the intertidal range of *Semibalanus*. We collected samples across a full stress gradient of two tidal extremes (high tidal edge [H] and lower tidal edge [L]) and 2 levels of sun exposure ("hot shores" [*h*] and "cold shores" [*c*]), for a total of four possible microhabitats: H*h*, L*h*, H*c*, L*c* (see Figure 1a,b). We defined H and L microhabitats based on the cycles of neap and spring tides. H microhabitats represent the 5% of the upper intertidal zone where barnacles settle and are always exposed to air during low tides. L microhabitats, on the other hand, are intertidal spaces only exposed to air for long periods of time during the lowest low tides of the lunar cycle. In Maine, we sampled all four microhabitats (H*h*, L*h*, H*c*, L*c*) across two biological replicates (i.e., different pools of barnacles collected in adjacent rocks experiencing the same types of microhabitat stressors, in the same geographical region; see Table S2). These replicates consist of two parallel islands (i.e., two islands located in the same latitude, separated by 150 m; ME1 [Hodgson Island] and ME2 [Farmer's Island]) in the Damariscotta river, in Damariscotta Maine. In Rhode Island, and Norway, we only sampled the two most extreme microhabitats (H*h*, L*c*) at two replicates. In Rhode Island, replicates were taken from a Jetty (RI1) and a boulder (RI2) in Jamestown, RI. In Norway, replicates were taken from beaches in the southwest side (NO1; *Herdla golfbana*), and north side (NO2; *Herdla naturreservat*) of Herdla Island. In Sweden, we sampled one site near Tjärnö (*Yttre Vattenholmen*). We only sampled barnacles who were at least 1 year or older, such that they had survived at least one full summer season. All individuals were removed from the rocky shore using screwdrivers or putty knifes and transported back to the laboratory in ethanol.

2.2 | Ecological sampling

At each of our sites, we obtained summer temperature data using high resolution sensors or from previously published works. Sites in RI as well as NO, and SV sites were surveyed using high resolution Thermochron sensors (part no. DS1921G; Maxim Integrated) deployed inside waterproof cases (part no. DS9107+; Maxim Integrated) to prevent water damage. Cases were glued to the rock using marine epoxy (1919324; Loctite) and fastened with screws installed with a hammer drill. Data from Maine was obtained from Schmidt and Rand (1999). For RI, additional temperature data was obtained from (Nunez, Rong, et al., 2020). The RI data which we collected includes the period from 11 July to 9 August 2018. The downloaded RI data covers the rest of the year. Our NO data spans 7 August to 4 September 2019. Lastly, our SV data documents 9

August to 8 September 2019. Note that temperature sensors were placed in the same patches from which barnacles were collected. We also characterized phenotypic differences between 183 barnacles from Maine and Rhode Island in H*h* and L*c* microhabitats using opercular plate dimensions as measures of phenotype. Accordingly, 91 opercular plates were from barnacles collected in the upper intertidal and 92 opercular plates were from barnacles collected in the lower intertidal. A total of 118 plates were from ME and 65 from RI. Dissections and measurements were done on barnacles of age >1. Opercular plates were separated from the rest of the barnacle shell and stored in a bleach solution for 24–48 h. Afterwards, the opercular plates were rinsed with water and photographed. The two-dimensional area of each component (left and right scutum, left and right tergum) of the opercular plates was calculated in ImageJ by drawing a multisided polygon around the component and finding the area within the polygon.

2.3 | DNA work

DNA was extracted from individual barnacles using the Qiagen DNeasy blood and tissue extraction kits (including an RNAse step). DNA was quantified using Qbit 3.0 and nanodrop 1000. For all samples, species identification was confirmed using Sanger sequencing of the mtDNA *COX I* region (Bucklin et al., 2011). The forward primer used was "TCAACAAATCACAAAGATATTGG" and the reverse primer used was "TAAACTTCAGGGTGTCCAAAG". PCR reaction was done using the EconoTaq PLUS GREEN 2X master mix (Cat no. 30033–2). Using a standard PCR protocol: 2 min at 94°C for initial denaturation. A total of 40 cycles were of 30 s at 94°C for cycle denaturation, 30 s at 52°C for cycle annealing, and 60 s at 72°C for cycle extension. Finally, we applied a 5-min final extension at 72°C. After species identity was confirmed by COXI sequence, pools of genomic DNA were prepared from equimolar contributions from 37, 38, or 50 individual barnacles for ME, RI, and NO respectively. Additional quality control, library preparation and sequencing were done by GENEWIZ LLC using 2 × 150 paired-end configuration. Each pool was sequenced and mapped to the reference genome to a read depth of ~57x. In total, our sequencing effort consisted of 12 pools sequenced in 12 lanes of Illumina.

2.4 | Genome mapping

Quality of DNA reads were assessed using FastQC (Babraham Bioinformatics, 2019), quality trimmed with BBmap's bbduk.sh module (Bushnell, 2016), and mapped to barnacle genome (Sbal3.1; NCBI GenBank: VOPJ00000000.1) (Nunez et al., 2019) using bwa mem (Li, 2013). Quality trimming option used were: qtrim =rl, trimq =20, maq=20, minlen =80, ftl =5, and ftr =145. Alignment post-processing and calling of single nucleotide polymorphism (SNPs) was done with a standard samtools/bcftools v1.9 (Li et al., 2009) pipeline. We used SAM flags -f 0x0002 -F 0x0004 -F 0x0008. This keeps reads mapped in proper pair, removes unmapped reads, and removes reads with mates unmapped. We also used picard tools (https://broadinstitute.github.io/picard/) to sort the output file and to remove duplicate reads. We made VCF files from each pool enforcing a minimum quality score of 35 and the --m option in the samtools/bcftools pipeline (using a μ "--prior" of 2.8e^{-9} mutations/per site/per generation). For all pools, we generated standard SYNC files, for pooled-sequencing analysis, using Popoolation-2's mpileup2sync.jar (Kofler, Orozco-terWengel, et al., 2011; Kofler et al., 2011). We filtered variants using a three-step approach. First, we removed indel regions and all SNPs flanking indels by five base pairs. Next, we removed singleton and doubleton SNPs (i.e., SNPs for which there are only one or two reads as supporting evidence). We accomplished this using a sed command on the SYNC file. Lastly, we estimated allele counts using the Perl script snp-frequency-diff.pl, a module part of Popoolation2 (Kofler, Orozco-terWengel, et al., 2011; Kofler, Pandey, et al., 2011), enforcing the following parameters --min-count 8, --min-coverage 10, and --max-coverage 300. These parameters were chosen based on previous work, see (Nunez, Rong, et al., 2020) and (Nunez, Flight, et al., 2020).

2.5 | Genomic analyses: demography

While the primary goal of our analysis was to understand the patterns of allele zonation across the North Atlantic rocky shores, our population panel was composed of samples with a complex biogeographic history. As such, prior to assessing zonation, it was pivotal to explore the genomic signatures of demography and assess how these patterns may influence our zonation analyses. Accordingly, we first conducted a principal component analysis (PCA) using a genome-wide sample of SNPs and including all populations regardless of microhabitat of origin. For these analyses, we filtered our SNPs, trimming linkage disequilibrium (LD) at 200 bp (LD threshold chosen based on Nunez, Rong, et al., 2020) and applying a minor allele frequency (mAF) filter to 5%. This resulted in ~300,000 LD-thinned SNPs. Principal component analysis on the genome-wide data was done using the R packages FactoMiner v2.4 and factoextra v1.0.7 (Lê et al., 2008). In addition, we conducted PCA analyses along windows of the barnacle genome using the method described by Li and Ralph (2019). We applied the local PCA on scaffolds larger than 1,000 bp and across windows of ~100 SNPs, thus partitioning the genome in 27,275 windows. These local PCA analyses were important to gain insight about potential "clusters of zonation signal" across the genome. We estimated levels of nucleotide diversity (π) and Tajima's D using the Popoolation1 suite (Kofler, Orozco-terWengel, et al., 2011; Kofler, Pandey, et al., 2011). Estimations of allele frequencies were done using the Popoolation2 suite. We also estimated genome-wide fixation index (F_{ST}; a proxy for genetic differentiation between populations or subpopulations) among and between all samples for populations in ME, RI, and NO. These populations were chosen based on our temperature data which suggested that only those habitats experience temperature fluctuations at the microhabitat level (see Results). These analyses were done using

a sliding window approach using the standard method implemented in the Popoolation2 suite (window size =1,000 bp; step size =500 bp).

2.6 | Genomic analyses: zonation and relative ages of alleles

To formally assess targets of zonation in our pools, we implemented Cochran–Mantel–Haenszel (CMH) and Fisher's exact (FET) tests. The CMH test is an association test which identifies constant and consistent changes in allele frequencies across independent biological replicates (Cochran, 1954; Mantel & Haenszel, 1959; Orozco-terWengel et al., 2012; Schlötterer et al., 2014; Wiberg et al., 2017). In this study, we applied CMH to the sampling hierarchy of intertidal extremes (Hh vs. Lc) of each biological replicate across the basin. Analyses were implemented in R (R Core Team, 2020). We used an false discovery rate (FDR) corrected (Benjamini & Hochberg, 1995) p-value threshold of .1 to classify SNP as zonated. We also estimated theta statistics (π and D) for all zonated SNPs (and neighbouring areas) on SAM files made by merged four individual microhabitat files into a single "superpool" to >200x read depth. Accordingly, we combined all pools from ME, RI, and from NO into their individual merged SAM files. This is appropriate given that all populations belong to the same demographic units. Whenever appropriate, we included pools from other populations obtained from Nunez, Rong, et al., (2020). These pools are Iceland (ICE), Wales (UKW), and British Columbia (WCAN; an ancestral population from the Pacific). In addition, our broad geographical sampling allowed us to leverage phylogeographic principles to better understand properties of the zonated loci. Accordingly, we constructed two allele trees using a neighbour-joining algorithm using ape v5.4.1 (Paradis & Schliep, 2019) and dendextend v1.1.4 (Galili, 2015): one with all SNPs, and another with only zonated SNPs discovered using the CMH test. We included all microhabitat populations as well as additional samples from across the North Atlantic and the North Pacific.

2.7 | Genomic analyses: molecular footprints of selection

We explored the genomic footprint of zonated loci by assessing the degree to which the F_{ST} summary statistic behaves near the zonated locus (i.e., focal site) relative to neighbouring sites (i.e., proximal sites; $\Delta F_{ST\ [focal-prox]}$). In order to define numerical ranges for focal and proximate sites, we tested the behaviour of F_{ST} at a range of window sizes for the focal region ranging from +/- 10 bp immediately after the zonated SNP to 1,500 bp. Accordingly, we surveyed the window sizes ("k") at which the F_{ST} distribution of the zonated focal region is indistinguishable from the F_{ST} calculated at across a random background. For each k, we generated a random background F_{ST} distribution by randomly sampling the same number of genomic loci found at the focal and proximal window

TABLE 1 Temperatures in the intertidal for the summer months

Pop	Rep.	Hab.	All (°C)	All$_{sd}$	Day (°C)	Day$_{sd}$
RI	1	Hh	24.0	4.19	27.3	4.59
RI	1	Hc	23.4	2.67	25.4	2.44
RI	1	Lc	22.6	1.68	23.7	1.48
RI	2	Hh	24.2	4.10	27.3	4.76
RI	2	Lc	22.7	1.28	23.4	1.08
NO	1	Hh	15.7	2.71	17.3	3.45
NO	1	Lc	15.4	1.30	16.0	1.15
NO	2	Hh	15.9	3.20	17.9	3.97
NO	2	Lc	15.5	1.41	16.1	1.30
SV	1	Hh	17.8	1.85	18.5	2.48
SV	1	Hc	17.9	1.82	18.6	2.42
SV	1	Lc	17.8	1.15	18.1	1.45

Note: Rep: shows the biological replicate id as well as the name of site. Hab: indicates the microhabitat label. All and All$_{sd}$ show the pooled mean and sd temperatures collected by the sensors. Day and Day$_{sd}$ show the mean and sd of temperatures collected during the day.

for the actual comparisons (e.g., ~4 SNPs at k = 10, ~20 SNPs at k = 100, ~100 SNPs at k = 500, etc.). Another fundamental question in our analysis is whether or not SNPs responding to ecological zonation are "old" or "young" alleles. In this context, we refer to age in relative phylogenetic terms. For example, we considered alleles to be old if they are shared between ocean basins (i.e., time to most common ancestor [T_{MRCA}] >2 My; [Nunez, Rong, et al., 2020]); conversely, young alleles are those shared across the entire North Atlantic (T_{MRCA} >20 kya; [Flight et al., 2012]). Accordingly, we classify loci by asking whether zonated alleles are shared with the ancestral population from Pacific Canada.

2.8 | Genomic analyses: functional annotations

Functional analyses were done based on the Sbal3.1 annotation. These were done by pairwise blast against the *Drosophila melanogaster* genome (Dmel6; NCBI GenBank: GCA_000001215.4). GO enrichment analysis was done using GOrilla (Eden et al., 2009) and GO terms inferred from our *Drosophila* annotation. The enrichment was assessed by comparing two gene lists, the first composed of the genes of interest (i.e., the gene targets), the second one by all the genes annotated in Sbal3.1 (i.e., the gene universe).

3 | RESULTS

3.1 | Ecological variation in the intertidal

We first characterized the extreme differences in thermal variation between the upper and lower intertidal. Our temperature data showed thermal differences among microhabitats (Table 1).

In Rhode Island, these differences are most apparent during the afternoon hours in H*h*, when the low tides coincide with the afternoon sun (Figure 1d and Figure S1). Of all temperature measurements, differences among standard deviation of H*h* and L*c* sites were most conspicuous (at least three-fold more variance in H*h* sites). This was observed in both RI (Figure 1d) and NO (Figure 1e) sites. Thermal differences were seen in SV but are generally weaker (Figure 1f). Next, we characterized the morphological differences of upper and lower intertidal barnacles. Particularly, we measured the area of complete opercular plates from 183 barnacles. We conducted this analysis with the intention of documenting that intertidal ecological variation has an impact on barnacle phenotypes. We measured the area of the barnacle operculum and summarized the data using a PCA. The first component of the PCA explains 99.6% of the variation and separates individuals in the upper intertidal from individuals in the lower intertidal (Figure 1c; Table S1). The data highlights that barnacles, of age >1, settling in H*h* microhabitats are smaller in size and have a lower size variance that those settling in L*c* sites across habitats (see Figure S2), probably due to phenotypic plasticity during post-settlement growth.

Overall, our temperature and morphological surveys provide us with confidence that our samples are capturing the biological complexities of intertidal zonation.

3.2 | Genome-wide genetic diversity

Using pool-seq, we characterized genetic diversity in the upper and lower microhabitats in the North Atlantic. Overall, our sequencing effort produced an average output of ~213 M high quality reads (which passed filter). The mean read depth was ~52x, mean mapping quality was high (~56), and general error rates were low (0.029). The total number of single nucleotide polymorphism (SNPs) discovered across populations was 23,142,252 (see summary in Table S2). We estimated the levels of nucleotide diversity (π) across the genome for all microhabitats. The π summary statistic is a proxy for the amount of standing genetic variation segregating in populations (Table S3). Overall, all microhabitats have similar levels of π (mean π = 1.02%; Figure 2a). We conducted a similar analysis for the Tajima's D estimator (Table S3). This estimator assesses the excess of rare alleles
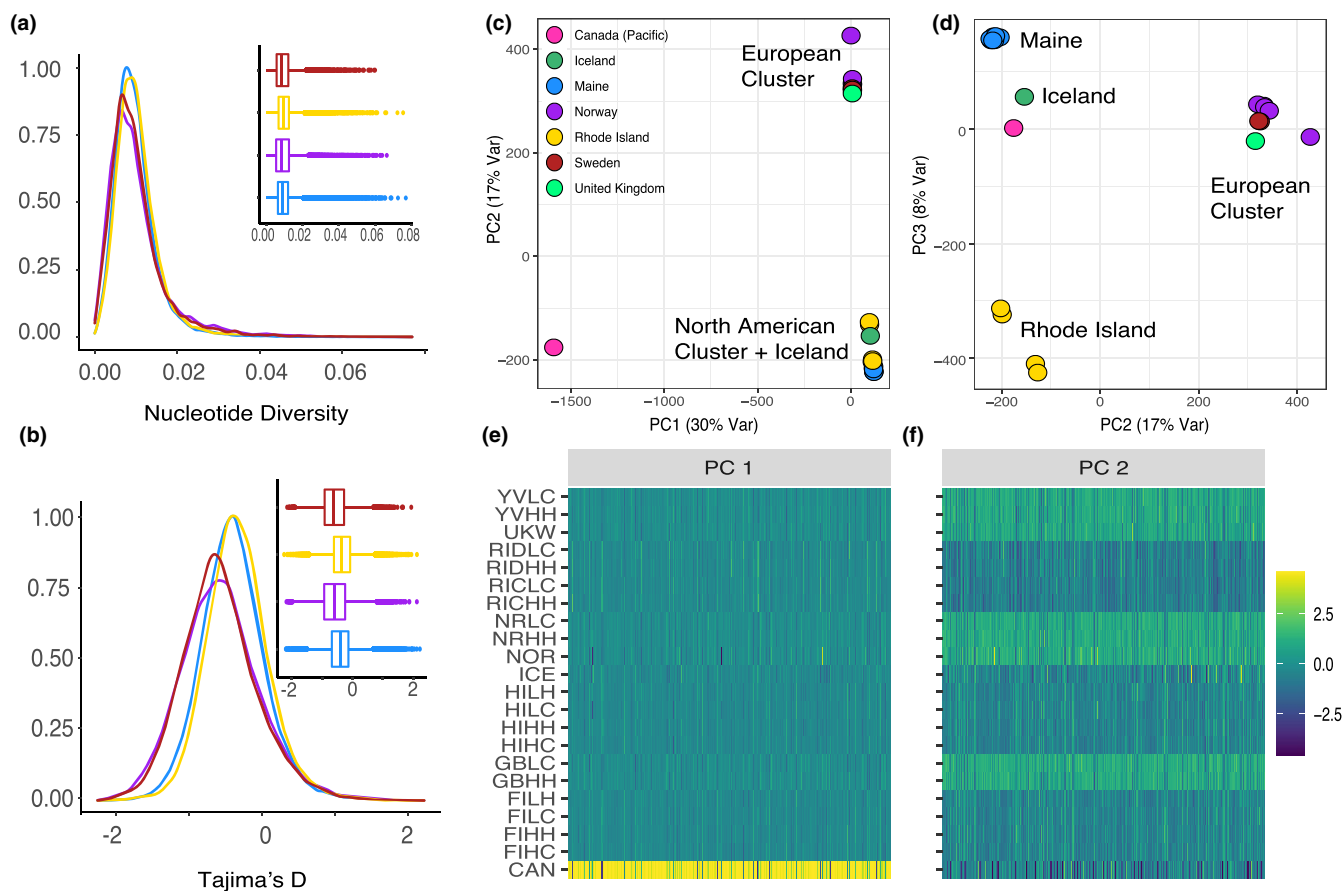

FIGURE 2 Population structure and genetic variation among samples. (a) Nucleotide diversity (π) plot across all samples. (b) Same as A, but for Tajima's *D*. The inset plots for a and b are the corresponding distributions visualized as box plots. The colours are consistent with the legend in panel C. (c) and (d) PCA of all populations analysed plus phylogeographic reference populations. (e) Local PCA projections across the genome. The x-axis corresponds to windows along the barnacle genome. The colour gradient (yellow to dark green) corresponds to the projections of each sample (per genomic window) onto the first PC of the local PCAs. The samples in the y-axis are sorted alphabetically from bottom to top. (f) Same as e, but the projections are done relative to the second PC

in populations and has multiple interpretations. We observed similar levels of negatively skewed Tajima's *D* across samples (mean *D* = −0.43; Figure 2b). However, European samples have a noticeably larger negative skew of D. This is consistent with a different post-glacial expansion history in European barnacles relative to North American barnacles (discussed in Nunez, Rong, et al., 2020).

3.3 | Population structure across oceans

Our global PCA (Figure 2c,d) revealed patterns of general population structure concordant with previous findings Nunez, Rong, et al., (2020). Namely, a stark differentiation between Atlantic and Pacific samples (Figure 2c; PC1). Within the Atlantic, samples subdivide into
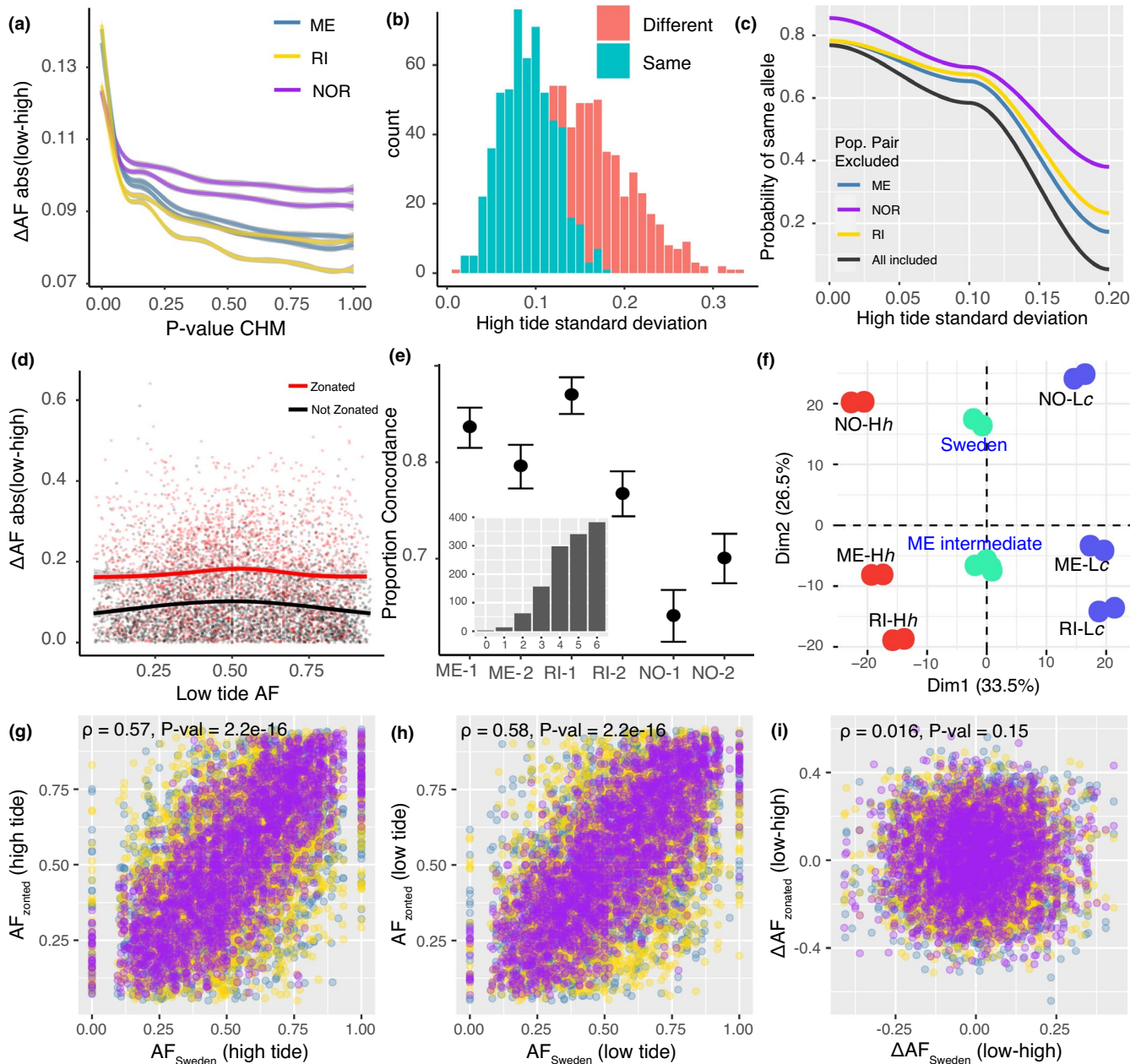

FIGURE 3 Zonation across ocean basins. (a) Distribution of allele frequency differences (high vs. low tide) as a function of CHM *p*-values. (b) Number of zonated loci which have the same allele call (in high tide populations) as a function of the standard deviation of the high tide allele frequency. (c) Probability of populations having the same allele as a function of allele frequency standard deviation. Each colour represents the probability after a given population pair (ME, NO, or RI) has been excluded. (d) Difference in allele frequency of zonated and nonzonated variants as a function of the frequency of low tide variants. (e) Concordance between CMH and FET analysis in each population. Inset: Proportion of concordant test across all populations. (f) PCA visualization of zonated mutations. In the plot, all populations samples from H*h* habitats are shown in red, and all samples from L*c* habitats are shown in blue. The plot includes samples from Sweden and intermediary ME habitats as supplementary points. (g) Correlation between zonated AF in high tide habitats and AF in Swedish high tide habitats. (h) Correlation between zonated AF in low tide habitats and AF in Swedish low tide habitats. (i) Correlation between zonated ΔAF in zonated populations and ΔAF in Sweden

two discrete clusters regardless of tidal level: a western cluster (RI, ME, ICE) and an eastern cluster (UK, NO, SV; Figure 2d; PC2). We also observed a further subdivision of the western cluster of populations north and south of Cape Cod, a known phylogeographic break for the species (Figure 2d; PC3) (Flight et al., 2012). In general, we did not observed any clustering patterns among microhabitat treatments (tide level, sun exposure, or sun *x* tide effects) among the firsts 20 PCs which explain >98% of the total variation in the data (Table S4). The only exception occurred in PC 17 (variance explained =1.2%), where we observed a significant association between PCA projections and the sun *x* tide model (*p*-value =.046; R^2 = .301; this association is not significant after false discovery rate adjustment). Our F_{ST} results were also highly concordant with the phylogeographic expectation from the PCA (Figure S3). Similar to the π analysis, the genome-wide distributions of F_{ST} did not reveal any conspicuous differentiation across the intertidal for any population. Overall, these results indicated that while microhabitat treatment is not a primary source of genome-wide variation, a signal of microhabitat zonation existed in the data.

3.4 | Population structure across the genome

Our PCA using the genome-wide sample of SNPs provided a bird's eye view of variation across the barnacle genome. However, there was a possibility for zonation signals to exist as regional clusters across the genome. To further understand how these types of demographic signal differs across the genome, we conducted local PCA in windows of 100 SNPs across the scaffolds of Sbal3.1. Accordingly, we plotted the projections of each sample onto PC space for dimensions 1 and 2. Our results indicated that, for PC1, the demographic signal is homogeneous across the genome with no appreciable regions or clusters deviating from the global PCA signal (Figure 2e). Notably, the local differentiation of North Atlantic samples relative to the ancestral sample from western Canada is always high across the windows. For PC2, the signal was similarly homogeneous across the genome (Figure 2f).

3.5 | Identifying zonated loci

We formally tested whether zonation is experienced at the level of individual SNPs using the Cochran–Mantel–Haenszel (CMH) test among all extreme microhabitat comparisons (H*h* vs. L*c*) in ME, RI, and NO. The CMH test is a statistical framework used in population genetics to assess constant and consistent changes in allele frequencies across independent replicates (Orozco-terWengel et al., 2012; Schlötterer et al., 2014; Wiberg et al., 2017). In our particular case, we used the microhabitat stress as the treatment variable, and the allele ratio of the pools as the outcome. Accordingly, our analysis found 1,257 zonated SNPs across the North Atlantic (Figure 3a). These zonated targets occured at 112 coding loci, seven 5'UTRs, 182 intergenic loci, 632 intron loci, 43 promoters, three 3'UTR. The remaining loci occured at unknown regions.

3.6 | Properties of zonated loci

Overall, zonated loci can be divided into two groups: 561 loci which have the same major alleles across all H*h* microhabitats, and 696 loci where major alleles differ in at least one population (Figure 3b). Accordingly, loci with the same major allele have similar allele frequencies within microhabitats. This was measured as the allele frequency standard deviation (AF_{sd}) within microhabitats. On average, the mean AF_{sd} of sites with the same allele is 0.091 +/- 0.03, and those with different allele states is 0.149 +/- 0.0544 (Figure 3b). The probability of zonated loci having the same allele state decreased towards zero after AF_{sd} ~0.2 (Figure 3c). Notably, when the analysis is repeated removing the Norwegian populations, the probability of having the same major allele increases drastically (Figure 3c; purple line). This suggests that phylogeographic clusters and demography is an important determinant on shared allele identity across microhabitats. Regardless of allelic state similarity, the allele frequency differences (ΔAF) between high and low tides sites was statistically different between zonated and nonzonated loci across the entire data set (Figure 3d; t = −38.667, df = 2,715.3, *p*-value <2.2 × 10^{-16}). As such, the mean $\Delta AF_{high-low}$ for zonated loci was 0.172 (ΔAF_{sd} = 0.107), and the $\Delta AF_{high-low}$ for nonzonated was 0.0928 (ΔAF_{sd} = 0.0748).

The individual *p*-values of the CMH test are derived from biological replicates from multiple populations across the basin. As such, there is an inherent risk of loci failing to replicate at any given site. Accordingly, any SNP with a significant *p*-value across the ocean may not guarantee that the same SNP has a significant *p*-value at any given pairwise comparison. To survey this, we conducted a leave-one-out (LOO) concordance analysis on the 1,257 zonated SNPs targets discovered above. Accordingly, we conducted an additional CMH analysis by removing one H*h* and L*c* pair. In parallel, we applied a Fisher's exact test (FET) to the removed H*h*-L*c* pair. We documented the proportion of loci which were significant in both the LOO-CMH and the FET. Our results show that 60%–80% of all loci have concordant results among the tests within any given population (Figure 3e). Proportion of concordance varies by population, with NO samples having the lowest proportion of concordance (62%–65%). When the data is considered together, we observed that 30% of all loci always replicate, 27% fail to replicate at least once, 24% fails to replicate at least twice, and the rest of the loci fail to replicate ≥3 times (Figure 3e inset). We kept loci which replicated in all 12 populations as zonated loci for follow-up characterizations. This resulted in a conservative panel of 382 zonated SNPs. We used PCA to visualize the patterns among zonated loci (Figure 3f). This visualization showcased a stark contrast with the global PCA (see Figure 2c). Accordingly, while phylogeography remained a strong signal in the data (PC2), the first PC captured consistent patterns of ecological zonation across ME, RI and NO. Notably, when adding the intermediate-zone samples from ME, and those from SV, they were intuitively projected in the middle of PC space. This finding suggests that, while the zonated variants exist in Sweden, they do not experience zonation. This can be appreciated by calculating correlations

between both the AF of Swedish populations at high and low tide and those from the rest of the data set and comparing that value with the $\Delta AF_{low\text{-}high}$ for the same comparison. Our analysis shows that 97% of all zonated variants are polymorphic in Sweden (Figure 3g,h); however, the $\Delta AF_{high\text{-}low}$ between zonated and nonzonated populations showed no correlation (Figure 3i).

Our broad geographical sampling allowed us to leverage phylogeographic principles to better understand properties of the zonated loci. As such, we estimated tree structures based on the allele frequencies of two data sets: One, the set of the 300,000 LD-thinned SNPs used to build the global PCA, and the other one, the set of zonated SNPs (Figure 4a). Our results shows that the tree with all SNPs displays a topology fully concordant with the PCA results shown in Figure 2c and in previous work (Nunez, Rong, et al., 2020). In this tree, populations clustered according to their phylogeographic provenance and no microhabitat clustering was observed. The zonated tree, on the other hand, showed clear clustering based on zonation patterns. Figure 4a shows the difference between tree topologies. Our analysis of relative age of zonated variants showed that 87% of all zonated variants were fixed in Pacific Canada: 57% were fixed for the ancestral allele and 29% were fixed for the derived allele (Figure 4b). The remaining 13% of variants were polymorphic across both ocean basins (Figure 4c). These analyses suggest that the bulk of zonated variants are "young" and private to the North Atlantic (Fisher's exact test p-value $<7 \times 10^{-13}$).

3.7 | Are zonated SNPs driven by spatially heterogeneous selection?

We estimated various summary statistics on and around (flanks 2,000 bp away at both 3' and 5' directions) each zonated SNP to look for footprints of natural selection. First, we calculated differences in allele frequency between upper and lower intertidal sites, $\Delta AF_{high\text{-}low}$. Notably, $\Delta AF_{high\text{-}low}$ was elevated around the zonated loci, however the signal decreased back to background levels at less than ~500 bp (Figure 5a). The $\Delta AF_{high\text{-}low}$ between zonated loci (including variants +/- 50 bp; the average size of the sequenced read) and other variants +/- 2,000 bp (i.e., proximal variants) was statistically significant (Figure 5b; $t = 11.051$, df = 32.229, p-value $=1.69 \times 10^{-12}$). Next, we estimated pairwise F_{ST} in and around the zonated loci for two types of comparisons: microhabitats with the same level of stress, and microhabitats with different levels of stress (Figure 5c). We only calculated these pairwise comparisons among pools collected from the same phylogeographic cluster (i.e., West Atlantic, or East Atlantic). This decision was done in order to avoid the confounding effects of demography on the F_{ST} estimator. Similar to the $\Delta AF_{high\text{-}low}$ analysis, F_{ST} shows that the zonation signal is highly localized to the vicinity of the discovered variant. Notably, the comparison among extreme microhabitats (Hh vs. Lc; within the same geographic region) showed very elevated levels of F_{ST}, relative to neighbouring variants (Figure 5d, p-values <<.01; red dashed lines). Notably, F_{ST} across similar microhabitats (e.g., Lc vs. Lc; within the same region) was significantly lower
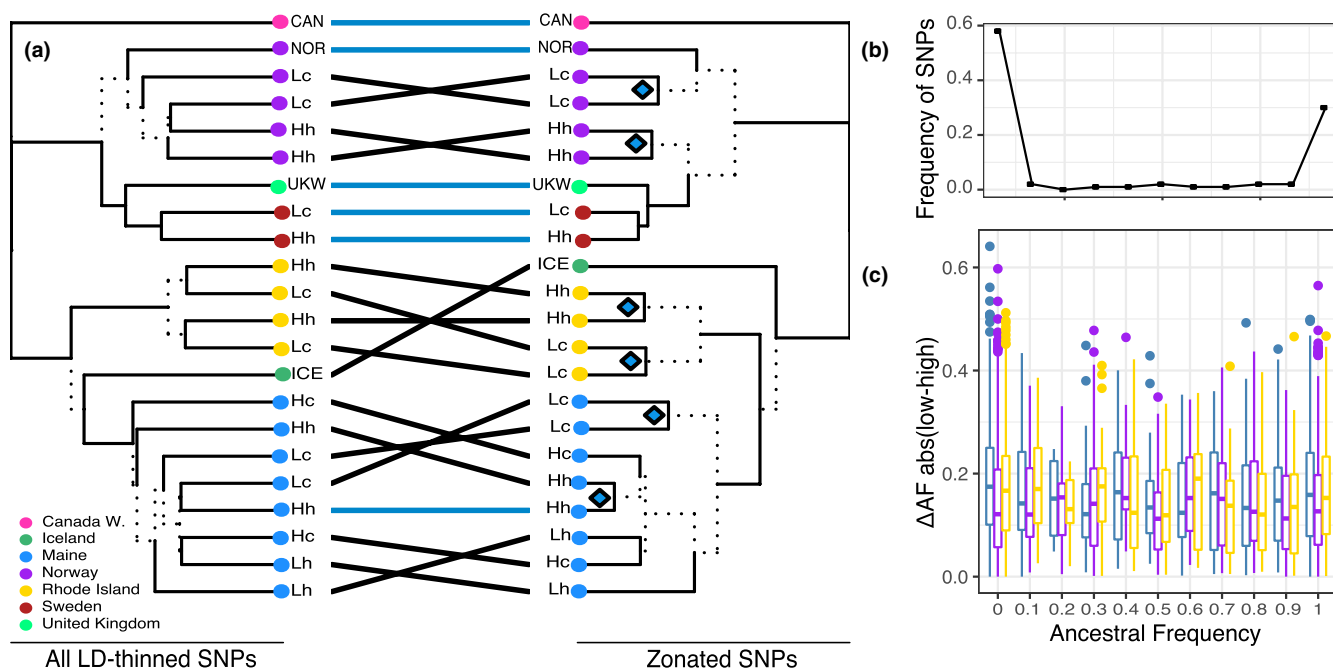

FIGURE 4 Phylogeographic context of zonated variants. (a) Tanglegram comparing two allele frequency dendrograms: one estimated using all SNPs (left) and one estimated using only zonated variants (right). Population whose tips are different across topologies are indicated with black lines. Populations with stable tips are shown with blue lines. Discordances in topologies are shown as dotted branches. Whenever a zonated pair builds a monophyletic clade are indicated with a blue rhombus. (b) Frequency of zonated SNPs as a function of ancestral frequency (in Pacific Canada). (c) Allele frequency differences (high vs. low tide) as a function of ancestral frequency. Blue boxplots represent Maine populations, purple is Norway, and yellow is Rhode Island
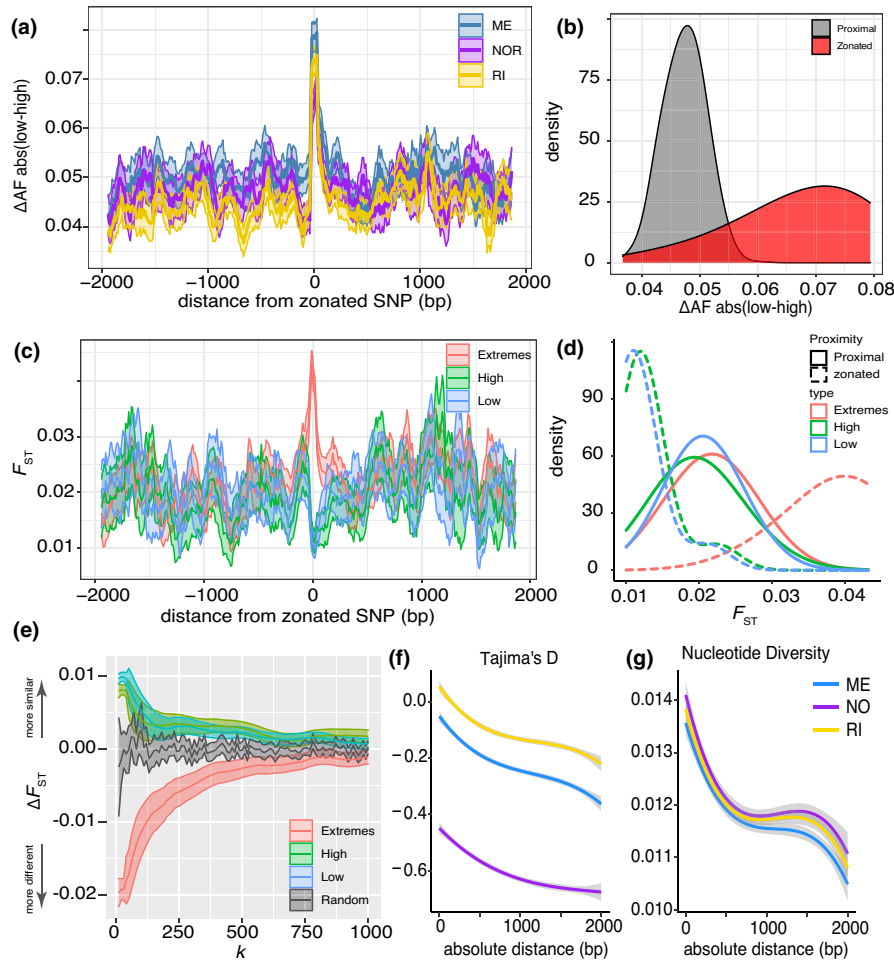
FIGURE 5 Properties of zonated SNPs. (a) Allele frequency differences (high vs. low tide) in and around zonated variants (+/- 2000 bp). The envelopes represent the 95% interval. (b) Distribution of allele frequency differences for two types of variants: zonated (all SNPs near +/- 50 bp of the zonated SNP; red) and proximal (all SNPs around +/- 2,000 bp of the zonated SNP; grey). (c) Fixation index (F_{ST}) of three types of comparisons: high vs. low microhabitats (red; labelled "extremes"), all samples from high microhabitats (green, "high"), and all samples from low microhabitats (blue, "low"). (d) Distributions of F_{ST} for zonated vs. proximal variants. (e) Mean differentiation and 95% confidence intervals of F_{ST} distribution between zonated regions and proximal regions for all comparisons. The x-axis indicates the window size around the zonated SNP consider part of the "focal zonated" zone (i.e., "k"). The grey envelop represents random sampling of the data. (f) Conditional mean (i.e., rolling mean) of Tajimas's D as a function of distance from the zonated SNPs (in the x-axis, 0, represents the location of the zonated SNP). (g) Same as F but for nucleotide diversity

than the region's average (Figure 5d, *p*-values <<.01; blue and green dashed lines). As such, all L*c* (or H*h*) sites were more similar to other L*c* (or H*h*) sites—at zonated loci—relative to the local genome average. Overall, these findings indicate that the genomic footprint of zonation is confined to the close proximity of the zonated SNPs discovered by the CMH test. However, these discovered loci may not be the drivers of the signal. Accordingly, all loci segregating immediately next to the zonated candidates may also have a role—or be a driver of—the zonation signal.

We attempted to formally estimate the size of the zonation footprint by assessing the degree to which the F_{ST} distribution of the focal zonated region varies from the background of the proximal variants (i.e., $\Delta F_{ST\ [focal-prox]}$). Our findings indicated two general behaviours: one for comparisons among extreme microhabitats (Figure 5e; H*h* vs. L*c*; red line), and one for comparisons among the

same microhabitats (Figure 5e; H*h* or L*c* vs. themselves; blue and green line). For extreme comparisons, the signal (difference between high and low) decayed to random background at ~490–500 bp. For same habitat comparisons, the signal (similarity among sites of same stress) decayed at ~90 bp (this is consistent with the graphical pattern shown in Figure 5a,c). These results suggest that the genomic footprint of zonation exists at a range of ~500 bp from the focal SNP. Based on these findings, we estimated the levels of genetic variation (D and π) and observe that, similar to other metrics, their values rapidly decay away from the vicinity of the focal SNP (Figure 5f,g). Table 2 shows the mean values of D and π for three ranges around the zonated loci (10, 500, and 2,000 bp). Values were estimated in a sliding window approach (window size =50 bp, step size =10 bp). With the exception of Norway, mean values of D in the immediate proximity of the zonated site are positive.

TABLE 2 Diversity metrics of zonated loci

Pop	Range	D	π
ME	1–2,000	−0.242	0.012
ME	1–500	−0.144	0.012
ME	1–10	0.058	0.015
RI	1–2,000	−0.131	0.012
RI	1–500	−0.034	0.013
RI	1–10	0.143	0.016
NOR	1–2,000	−0.613	0.012
NOR	1–500	−0.524	0.013
NOR	1–10	−0.324	0.016

Note: Estimates of mean Tajima's *D* (D) and mean nucleotide diversity (π) for zonated loci as a function of distance from zonated loci: +/- 10 bp, +/- 500 bp, +/- 2,000 bp.

Following our finding that the range of the zonated signal exists at ~500 bp from the discovered SNP, we expanded our zonated SNP panel to include both the loci discovered by the LOO-CMH/FET tests, as well as those in the 500 bp flanks. This expanded our data set of candidate SNPs to contain 7,536 SNPs (see Figure S4; 382 focal +7,154 neighbouring; see File S1). These loci are portioned into annotated regions: 1186 (10.2%) coding, 80 (0.69%) 5'UTR, 1603 (13.2%) intergenic, 4275 (37.0%) introns, 366 (3.17%) promoters, and 26 (0.22%) 3'UTR. These results showed that zonated-adjacent loci are over-enriched for coding loci compared to distant loci (odds ratio [OR] =1.29, *p*-value =.0058) while under-enriched for intronic, and 3'UTR loci (OR =0.844, *p*-values: 2.20×10^{-16}; and OR = 0.494, 4.79×10^{-4}, respectively). Moreover, zonated regions contain 162 protein coding genes. Given that our relative age analysis (see above) suggested that 17% of zonated targets are old alleles, we compared how many of the zonated genes overlaped with the genes reported in Nunez, Rong, et al., (2020) which contain trans-species polymorphisms (TSPs) and are candidates of ancestral balancing selection. We only observed 13 loci which are both zonated in all populations and harbour TSPs. This proportion of shared loci (~3%) does not suggest enrichment of zonated targets among genes experiencing ancestral balancing selection.

3.8 | A note on zonated loci functions

Lastly, we characterized the putative functions of our zonated genes based on annotations inferred from *D. melanogaster* homology (including gene ontology [GO] terms). GO enrichment analysis was performed on the 162 zonated genes across all populations. Analyses suggested an enrichment for terms including ion channel complex (GO:0034702; *p*-value 1.97×10^{-4}), transmembrane signalling receptor activity (GO:0004888, *p*-value 2.19×10^{-5}) and similar terms. Among the 13 genes with zonated SNPs which also occur in genes with TSPs (Nunez, Rong, et al., 2020), we observed a notable candidate: the gene painless (*pain*) which encodes a nonselective cation

channel required for the detection of noxious heat and mechanical stimuli (Tracey et al., 2003; Xu et al., 2006).

4 | DISCUSSION

In this study, we have shown evidence that zonation of alleles is a repeatable phenomenon across the North Atlantic basin, with at least ~300 genomic regions and ~7,000 SNPs associated with the ecological gradient of the rocky shore. This zonation signal appears to be localized at scales of ~500 bp, consistent with the presence of small linkage blocks observed in *Semibalanus* (Nunez, Flight, et al., 2020). Moreover, the majority of zonated alleles appear to be young and restricted to North Atlantic populations. This observation suggests that the age of mutations falls between the estimated dates of the transarctic T_{MRCA} of 2 million years, when the species first entered the Atlantic basin, and the North Atlantic T_{MRCA} of 20 kya (the last glacial maxima), when North American and European barnacles shared a common ancestor (Nunez, Rong, et al., 2020). These findings tap into a question with larger implications in ecological genomics: are consistently zonated loci under spatially heterogeneous selection such as the one illustrated by the *Mpi* locus (see Nunez, Flight, et al., 2020)? In this sense, all the zonated regions discovered in this paper can be seen as candidate loci in need of functional validation in order to understand what types of selection are at play.

4.1 | Is repeatable zonation evidence for the Levene model?

One hypothesis to explain these patterns of zonation is that spatially heterogeneous selection is at play following the Levene model (Levene, 1953). An important assumption of this model is the existence of mutations which are advantageous in each microhabitat (i.e., H*h* and L*c*). Three lines of evidence suggest that this is a plausible hypothesis: first, the footprints of molecular evolution at our zonated loci are concordant with balancing selection (e.g., elevated F_{ST}, π, and Tajima's D, at the focal site; Vitti et al., 2013; Fijarczyk & Babik, 2015). Second, while the loci discovered by the CMH test are present across all North Atlantic habitats (e.g., Figure 3g,h), zonation only occurs at habitats that experience drastic intertidal temperature variation (Figure 3f). This is most salient in the case of Norway and Sweden. Both populations are very similar at the genome-wide level; however, zonated loci in Norway segregate in the same patterns as in North American populations. Finally, the ecology of intertidal ecosystems is highly conducive to reciprocal selection coefficients in H*h* and L*c* habitats. For example, while spatially adjacent, barnacles settling in different intertidal microhabitats experience diametrically different ecological challenges. These challenges have well-studied phenotypic and physiological consequences for individual barnacles (Schmidt, 2001). To this point, our data shows that barnacles that settled in

upper intertidal sites (e.g., H*h*) and are >1 year old, have a smaller variance in the size of their opercular plates (Figure 1c), and it is known that they make reproductive investments later in life (Bertness et al., 1991). While there is no doubt that environmental factors, age, and phenotypic plasticity play a major role in this process, the combination of these lines of evidence provide support for a selective hypothesis in which zonation is the product of spatially heterogeneous selection.

4.2 | Ecological load at the habitat margins?

An alternative to the Levene model is a "load at the margins" model. This hypothesis emerges because the Levene model requires beneficial alleles to exists in both microhabitats. Conversely, it is possible that many of the zonated mutations have a fitness value close to the population mean in most microhabitats, but they are deleterious at the most extreme edges of the habitat. This hypothesis is plausible for two reasons. First, most zonated alleles are young variants unique to the north Atlantic, which indicates a lack of long-term maintenance. Second, there is empirical evidence to suggest that individuals that survive in the fringes of the intertidal have reduced contributions to the gene pool compared to other regions of the rocky shore (Bertness et al., 1991). This alternative hypothesis has interesting implications for the evolutionary potential of intertidal organisms. On one hand, abundance of deleterious, or mildly deleterious, mutations at the margins will doom barnacles to experience high levels of ecological load across its habitat. Yet, it would also suggest that, similar to *Drosophila*, barnacles are not mutation limited (Karasov et al., 2010). Meaning that the species has abundant evolutionary fuel to elicit rapid adaptive responses. This suggests that barnacles—and species with similar genomic landscapes—may be better prepared for future evolutionary responses to novel stressors within their realized niche. For example, as anthropogenic climate change continues to disturb ecological systems, the habitat distribution and stressors normally experienced by intertidal species is bound to undergo rapid and drastic changes (Bertness et al., 1999; Helmuth et al., 2006). For example, it has been documented that *S. balanoides* have experienced a northward habitat contraction of 350 km due to climate change (Jones et al., 2012). This provides a future opportunity to characterize how these latitudinal habitat changes have modified genetic variation at the intertidal level.

4.3 | The potential role of demography in zonation

Another alternative hypothesis to both selective models is the role of demography. Specifically, that our zonated loci are caused by processes such as cohort behavior, differential habitat availability, or timing of settlement of the barnacle larvae (Gaines & Bertness, 1992; Hills & Thomason, 2003). While this hypothesis could be very plausible for individual loci, we do not believe that all zonated loci

are derived from demographic processes. We argue this is the case for three reasons: first, drastic cohort recruitment is likely to affect genome-wide levels of genetic variation, and our global and local PCA analyses do not detect such signals (Figure 2e,f). Second, the evidence presented by Schmidt and Rand (2001), using allozyme and mitochondrial data, suggests that barnacles recruit out of the water column and into the rocky shore as a well-mixed population (but see Veliz et al., 2006). Lastly, our evidence for zonation is contingent on successful replication across the North Atlantic basin. As such, allele frequencies must replicate across populations living thousands of miles apart, collected across different years, and belonging to two different phylogeographic clusters. On top of this, when one considers that the dynamics and timing of recruitment is different between European and North American shores (Crisp, 1964, 1968), a global demography hypothesis—for all zonated loci—becomes less likely. We recognize that resolving all of these hypotheses is a daunting task. Nevertheless, we could gain novel insights by genotyping barnacles across their life cycle from the water column (January–February) to reproductive stage (August–September). Such experiments would reveal whether allele frequency changes occur due to demography, random chance, or selection.

4.4 | The missing pieces: are zonated loci clustered in the genome?

The analysis presented thus far leverages the power of the nascent genomic tools for *Semibalanus*. However, due to the current state of the draft genome, our analyses are mostly effective when describing patterns at SNP loci. Consequentially, while our local PCA analysis (Figure 2e,f) does not highlight notable patterns across the genome, the fragmentation of the reference leaves a lot of room for future studies on structural variation. For example, it is still unclear the degree to which zonated variants may be clustered in distinct genomic regions (e.g., particular chromosome arms). Overall, studies of structural variation are highly promising, given the recent discoveries on the role of inversions in the maintenance of ecologically important polymorphisms (Faria et al., 2019). For example, in the intertidal snail *Littorina saxatilis*, the genomic basis of ecotypes differentially distributed across "high wave action" and "high predation" microhabitats has been associated with SNPs located in large inversions (Westram et al., 2018). It is currently unknown if any of the putative zonation regions in barnacles may occur—or be driven by—inversions. Tackling these questions require improvements to the barnacle genome as well as generation of additional datasets, other than pool-seq, which capture individual genotype and haplotype information. In addition, a panel of individual haplotypes would allow association studies between genetic variation and organismal traits. For example, while we have morphological measurements for individual barnacles, the pooled nature of our data prevents whole genome association studies (GWAS). Future studies will seek to generate these data sets and conduct these analyses.

4.5 | Strengths and weaknesses of pool-seq

In this study, we used pool-seq to survey genetic variation in our samples. This method is a powerful and cost-effective tool to quantify genetic variation (Schlotterer et al., 2014a, 2014b). Aside from its cost-effectiveness, pool-seq has the added advantage of producing genome-wide data, thus allowing it to outperform other cost-effective methods, such as RAD-seq or GBS (Baird et al., 2008; Elshire et al., 2011), in terms of raw number of SNPs discovered. Nevertheless, the method is not without drawbacks. For example, while pool-seq is very effective at sampling large numbers of SNPs across the genome, the LD information among SNPs is limited, relative to individual sequencing. Moreover, the method is very dependent on parameters such as number of individuals pooled, sequencing coverage, as well as sequencing technology. As such, the method can present additional challenges for non-model systems, where uniform sample sizes, or high DNA quality across populations may be logistically difficult. As a consequence, experimental designs that deviate from the recommended settings (e.g., Gautier et al., 2013), may produce inaccurate estimates of allele frequency (Anderson et al., 2014). We ameliorated these shortcomings by leveraging biological replicates as well as the phylogeography of the species.

4.6 | Zonated functional variation

While the annotation of the barnacle genome is still in its early stages, homology comparisons with the model system *D. melanogaster* provide an avenue of hypothesis generation. A notable finding of our homology analysis is the putative enrichment of GO terms related to signalling processes, and ion channel processes. How are these genes linked to survival in highly heterogeneous intertidal habitats? One explanation is that the signal is purely artificial. The annotations in the barnacle genome are reliant on their functional homology to *Drosophila*, and thus all results should be interpreted with caution. Nevertheless, the enrichment of ion channel genes, for example the *Painless* gene, may indicate a potential role of environmental sensing phenotypes. As discussed by Nunez, Rong, et al., (2020), these phenotypes may be pivotal for survival in ecologically heterogeneous environments. For example, selection on the mechanisms responsible for sensing ecological heterogeneity may allow barnacles to make physiological decisions which maximize individual survival (see Jenkins, 2005). As genomic tools for the barnacle continue to mature (e.g., Jonsson et al., 2018; Nunez et al., 2018), the kinds of experiments needed to test these hypotheses and connect genotype and phenotype will become increasingly feasible and will allow us to capitalize on the century-old body of ecological knowledge for barnacles.

AUTHOR CONTRIBUTIONS

JCBN: Conceptualization, Data curation, Formal Analysis, Funding acquisition, Investigation, Methodology, Project Administration, Resources, Software, Supervision, Visualization, Writing - original draft, Writing - review & editing. SR: Formal Analysis, Methodology, Software, Visualization, Writing - original draft, Writing - review & editing. DAF: Formal Analysis, Methodology, Software, Visualization, Writing - original draft, Writing - review & editing. ADS: Resources, Methodology, Writing - review & editing. KBN: Formal Analysis, Writing - review & editing. HG: Resources, Writing - review & editing. RGE: Resources, Formal Analysis, Writing - review & editing. BRPB: Resources, Writing - review & editing. MAR: Conceptualization, Funding acquisition, Project Administration, Resources, Supervision, Writing - original draft, Writing - review & editing. AB: Conceptualization, Funding acquisition, Project Administration, Resources, Supervision, Writing - original draft, Writing - review & editing. KJ: Conceptualization, Funding acquisition, Investigation, Methodology, Project Administration, Resources, Writing - original draft, Writing - review & editing. DMR: Conceptualization, Data curation, Formal Analysis, Funding acquisition, Investigation, Methodology, Project Administration, Resources, Software, Supervision, Visualization, Writing - original draft, Writing - review & editing.

CONFLICTS OF INTEREST

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

Data used in this paper has been made available in the National Center for Biotechnology Information (NCBI), https://www.ncbi.nlm.nih.gov. Raw reads have been deposited under Bioproject ID PRJNA560680. SRA Ids: SRR10011795-SRR10011826, and SRR12037155-SRR12037160. Please note that in the NCBI SRA repository some of the H*h* samples were recorded as "HS" (HS means High South facing which is equivalent to H*h*). Likewise, some of the L*c* samples were recorded as "LN" (LN means Low North facing which is equivalent to L*c*). Please consult Table S2 for detailed sample-to-SRR correspondence. The barnacle genome (Sbal3.1) is available at NCBI (accession no. VOPJ00000000.1). A GitHub repository can be found at https://github.com/Jcbnunez/BarnacleZonation. Data for this paper can be found in Dryad: https://doi.org/10.5061/dryad.02v6wwq2f.

ORCID

Joaquin C. B. Nunez https://orcid.org/0000-0002-3171-8918

Stephen Rong https://orcid.org/0000-0002-6584-1391

Alejandro Damian-Serrano https://orcid.org/0000-0002-2544-0489

David M. Rand https://orcid.org/0000-0001-6817-3459

REFERENCES

Anderson, E. C., Skaug, H. J., & Barshis, D. J. (2014). Next-generation sequencing for molecular ecology: A caveat regarding pooled samples. *Molecular Ecology*, 23(3), 502–512.

Asthana, S., Schmidt, S., & Sunyaev, S. (2005). A limited role for balancing selection. *Trends in Genetics*, 21(1), 30–32.

Babraham Bioinformatics, B. I. (2019). *FastQC: A quality control tool for high throughput sequence data*. Retrieved from https://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Baird, N. A., Etter, P. D., Atwood, T. S., Currey, M. C., Shiver, A. L., Lewis, Z. A., Selker, E. U., Cresko, W. A., & Johnson, E. A. (2008). Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS One*, 3(10), e3376.

Barnes, H., & Barnes, M. (1959). A comparison of the annual growth patterns of *Balanus balanoides* (L.) with particular reference to the effect of food and temperature. *Oikos*, 10(1), 1.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, 57(1), 289–300.

Bergland, A. O., Tobler, R., Gonzalez, J., Schmidt, P., & Petrov, D. (2014). Secondary contact and local adaptation contribute to genome-wide patterns of clinal variation in *Drosophila melanogaster*. *Molecular Ecology*, 25, 1157–1174. https://doi.org/10.1111/mec.13455

Bertness, M. D., Gaines, S. D., Bermudez, D., & Sanford, E. (1991). Extreme spatial variation in the growth and reproductive output of the acorn barnacle *Semibalanus-balanoides*. *Marine Ecology Progress Series*, 75(1), 91–100.

Bertness, M. D., Gaines, S. D., Stephens, E. G., & Yund, P. O. (1992). Components of recruitment in populations of the acorn barnacle *Semibalanus balanoides* (Linnaeus). *Journal of Experimental Marine Biology and Ecology*, 156(2), 199–215.

Bertness, M. D., Leonard, G. H., Levine, J. M., & Bruno, J. F. (1999). Climate-driven interactions among rocky intertidal organisms caught between a rock and a hot place. *Oecologia*, 120(3), 446–450.

Brown, B. R. P., Nunez, J. C. B., & Rand, D. M. (2020). Characterizing the cirri and gut microbiomes of the intertidal barnacle *Semibalanus balanoides*. *Animal Microbiome*, 2(1), 41.

Bubb, K. L., Bovee, D., Buckley, D., Haugen, E., Kibukawa, M., Paddock, M., Palmieri, A., Subramanian, S., Zhou, Y., Kaul, R., Green, P., & Olson, M. V. (2006). Scan of human genome reveals no new Loci under ancient balancing selection. *Genetics*, 173(4), 2165–2177.

Bucklin, A., Steinke, D., & Blanco-Bercial, L. (2011). DNA barcoding of marine metazoa. *Annual Review of Marine Science*, 3, 471–508.

Bushnell, B. (2016). *BBMap short read aligner*. University of California. Retrieved from http://sourceforge.net/projects/bbmap

Cochran, W. G. (1954). Some methods for strengthening the common χ2 tests. *Biometrics*, 10(4), 417–451.

Crisp, D. J. (1964). Racial differences between North American and European forms of *Balanus balanoides*. *Journal of the Marine Biological Association of the United Kingdom*, 44(01), 33.

Crisp, D. J. (1968). Differences between North American and European populations of *Balanus balanoides* revealed by transplantation. *Journal of the Fisheries Research Board of Canada*, 25(12), 2633–2641.

Dufresne, F., Bourget, E., & Bernatchez, L. (2002). Differential patterns of spatial divergence in microsatellite and allozyme alleles:

Further evidence for locus-specific selection in the acorn barnacle, Semibalanus balanoides? *Molecular Ecology*, 11(1), 113–123.

Eden, E., Navon, R., Steinfeld, I., Lipson, D., & Yakhini, Z. (2009). GOrilla: A tool for discovery and visualization of enriched GO terms in ranked gene lists. *BMC Bioinformatics*, 10, 48.

Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S., & Mitchell, S. E. (2011). A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS One*, 6(5), e19379.

Faria, R., Johannesson, K., Butlin, R. K., & Westram, A. M. (2019). Evolving inversions. *Trends in Ecology and Evolution*, 34(3), 239–248.

Fijarczyk, A., & Babik, W. (2015). Detecting balancing selection in genomes: Limits and prospects. *Molecular Ecology*, 24(14), 3529–3545.

Flight, P. A., O'Brien, M. A., Schmidt, P. S., & Rand, D. M. (2012). Genetic Structure and the North American postglacial expansion of the barnacle, *Semibalanus balanoides*. *Journal of Heredity*, 103(2), 153–165.

Flowerdew, M. W. (1983). Electrophoretic investigation of populations of the Cirripede Balanus-Balanoides (L) around the North-Atlantic Seaboard. *Crustaceana*, 45, 260–278.

Gaines, S. D., & Bertness, M. D. (1992). Dispersal of juveniles and variable recruitment in sessile marine species. *Nature*, 360(6404), 579–580.

Galili, T. (2015). dendextend: An R package for visualizing, adjusting and comparing trees of hierarchical clustering. *Bioinformatics*, 31(22), 3718–3720.

Gautier, M., Foucaud, J., Gharbi, K., Cezard, T., Galan, M., Loiseau, A., Thomson, M., Pudlo, P., Kerdelhue, C., & Estoup, A. (2013). Estimation of population allele frequencies from next-generation sequencing data: Pool-versus individual-based genotyping. *Molecular Ecology*, 22(14), 3766–3779.

Hedrick, P. W. (1976). Genetic variation in a heterogeneous environment. II. Temporal heterogeneity and directional selection. *Genetics*, 84(1), 145–157.

Hedrick, P. W., Ginevan, M. E., & Ewing, E. P. (1976). Genetic polymorphism in heterogeneous environments. *Annual Review of Ecology and Systematics*, 7(1), 1–32.

Helmuth, B., Mieszkowska, N., Moore, P., & Hawkins, S. J. (2006). Living on the edge of two changing worlds: Forecasting the responses of rocky intertidal ecosystems to climate change. *Annual Review of Ecology, Evolution, and Systematics*, 37(1), 373–404.

Hills, J. M., & Thomason, J. C. (2003). The 'ghost of settlement past' determines mortality and fecundity in the barnacle, *Semibalanus balanoides*. *Oikos*, 101(3), 529–538.

Holm, E. R., & Bourget, E. (1994). Selection and population genetic-structure of the barnacle *Semibalanus-balanoides* in the Northwest Atlantic and Gulf of St-Lawrence. *Marine Ecology Progress Series*, 113(3), 247–256.

Jamie, G. A., & Meier, J. I. (2020). The persistence of polymorphisms across species radiations. *Trends in Ecology and Evolution*, 35(9), 795–808.

Jenkins, S. R. (2005). Larval habitat selection, not larval supply, determines settlement patterns and adult distribution in two chthamalid barnacles. *Journal of Animal Ecology*, 74(5), 893–904.

Jones, S. J., Southward, A. J., & Wethey, D. S. (2012). Climate change and historical biogeography of the barnacle Semibalanus balanoides. *Global Ecology and Biogeography*, 21(7), 716–724.

Jonsson, P. R., Wrange, A. L., Lind, U., Abramova, A., Ogemark, M., & Blomberg, A. (2018). The *Barnacle balanus* improvisus as a marine model - culturing and gene expression. *Journal of Visualized Experiments*, (138), e57825. https://doi.org/10.3791/57825

Karasov, T., Messer, P. W., & Petrov, D. A. (2010). Evidence that adaptation in Drosophila is not limited by mutation at single sites. *PLoS Genetics*, 6(6), e1000924.

Kimura, M., & Crow, J. (1964). The number of alleles that can be maintained in a finite population. *Genetics*, 49, 725–738.

Kimura, M., & Ohta, Y. (1971). Protein polymorphism as a phase of molecular evolution. *Nature*, 229, 467–469.

King, J. L. (1967). Continuously distributed factors affecting fitness. *Genetics*, *55*(3), 483–492.

Kofler, R., Orozco-terWengel, P., De Maio, N., Pandey, R. V., Nolte, V., Futschik, A., Kosiol, C., & Schlotterer, C. (2011). PoPoolation: A toolbox for population genetic analysis of next generation sequencing data from pooled individuals. *PLoS One*, *6*(1), e15925.

Kofler, R., Pandey, R. V., & Schlotterer, C. (2011). PoPoolation2: Identifying differentiation between populations using sequencing of pooled DNA samples (Pool-Seq). *Bioinformatics*, *27*(24), 3435–3436.

Lê, S., Josse, J., & Husson, F. (2008). FactoMineR: An R package for multivariate analysis. *Journal of Statistical Software*, *25*(1), 1–18.

Levene, H. (1953). Genetic equilibrium when more than one ecological niche is available. *American Naturalist*, *87*(836), 331–333.

Lewontin, R. C., & Hubby, J. L. (1966). A molecular approach to the study of genic heterozygosity in natural populations. II. Amount of variation and degree of heterozygosity in natural populations of Drosophila pseudoobscura. *Genetics*, *54*(2), 595–609.

Li, H. (2013). *Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM*. arXiv:1303.3997.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., & S. Genome Project Data Processing (2009). The sequence alignment/Map format and SAMtools. *Bioinformatics*, *25*(16), 2078–2079.

Li, H., & Ralph, P. (2019). Local PCA shows how the effect of population structure differs along the genome. *Genetics*, *211*(1), 289–304.

Machado, H. E., Bergland, A. O., Taylor, R., Tilk, S., Behrman, E., Dyer, K., Fabian, D. K., Flatt, T., González, J., Karasov, T. L., Kozeretska, I., Lazzaro, B. P., Merritt, T. J. S., Pool, J. E., O'Brien, K., Rajpurohit, S., Roy, P. R., Schaeffer, S. W., Serga, S., & Petrov, D. A. (2019). *Broad geographic sampling reveals predictable and pervasive seasonal adaptation in Drosophila*. bioRxiv (unpublished data). Retrieved from https://www.biorxiv.org/content/biorxiv/early/2019/10/11/337543.full.pdf

Mantel, N., & Haenszel, W. (1959). Statistical aspects of the analysis of data from retrospective studies of disease. *Journal of the National Cancer Institute*, *22*(4), 719-748.

Milkman, R. D. (1967). Heterosis as a major cause of heterozygosity in nature. *Genetics*, *55*(3), 493.

Nunez, J. C. B., Elyanow, R. G., Brown, B. R. P., & Rand, D. M. (2019). *The genome of* Semibalanus balanoides, *version 3.1*. NCBI GenBank. Retrieved from https://www.ncbi.nlm.nih.gov/nuccore/VOPJ00000000.1. Deposited Aug 7, 2019.

Nunez, J. C. B., Elyanow, R. G., Ferranti, D. A., & Rand, D. M. (2018). Population genomics and biogeography of the Northern Acorn Barnacle (*Semibalanus balanoides*) using pooled sequencing approaches. In M. F. Oleksiak, & O. P. Rajora (Eds.), *Population genomics: Marine organisms* (pp. 139–168). Springer.

Nunez, J. C. B., Flight, P. A., Neil, K. B., Rong, S., Eriksson, L. A., Ferranti, D. A., Rosenblad, M. A., Blomberg, A., & Rand, D. M. (2020). Footprints of natural selection at the mannose-6-phosphate isomerase locus in barnacles. *Proceedings of the National Academy of Sciences*, *117*(10), 5376–5385.

Nunez, J. C. B., Rong, S., Damian-Serrano, A., Burley, J. T., Elyanow, R. G., Ferranti, D. A., Neil, K. B., Glenner, H., Rosenblad, M. A., Blomberg, A., Johannesson, K., & Rand, D. M. (2020). Ecological load and balancing selection in circumboreal barnacles. *Molecular Biology and Evolution*, *38*(2), 676–685.

Orozco-terWengel, P., Kapun, M., Nolte, V., Kofler, R., Flatt, T., & Schlotterer, C. (2012). Adaptation of Drosophila to a novel laboratory environment reveals temporally heterogeneous trajectories of selected alleles. *Molecular Ecology*, *21*(20), 4931–4941.

Paradis, E., & Schliep, K. (2019). ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics*, *35*(3), 526–528.

Powell, J. R., & Taylor, C. E. (1979). Genetic variation in ecologically diverse environments: Environmental diversity and intraspecific habitat choice, rather than heterosis, may be the answer to the question "Why so many genotypes?". *American Scientist*, *67*(5), 590–596.

R Core Team (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Rand, D. M., Spaeth, P. S., Sackton, T. B., & Schmidt, P. S. (2002). Ecological genetics of Mpi and Gpi polymorphisms in the acorn barnacle and the spatial scale of neutral and non-neutral variation. *Integrative and Comparative Biology*, *42*(4), 825–836.

Schlötterer, C., Kofler, R., Versace, E., Tobler, R., & Franssen, S. U. (2014). Combining experimental evolution with next-generation sequencing: A powerful tool to study adaptation from standing genetic variation. *Heredity*, *114*(5), 431–440.

Schlotterer, C., Tobler, R., Kofler, R., & Nolte, V. (2014a). Sequencing pools of individuals - mining genome-wide polymorphism data without big funding. *Nature Reviews Genetics*, *15*(11), 749–763.

Schlotterer, C., Tobler, R., Kofler, R., & Nolte, V. (2014b). Sequencing pools of individuals-mining genome-wide polymorphism data without big funding. *Nature Reviews Genetics*, *15*(11), 749–763.

Schmidt, P. S. (2001). The effects of diet and physiological stress on the evolutionary dynamics of an enzyme polymorphism. *Proceedings of the Royal Society B: Biological Sciences*, *268*(1462), 9–14.

Schmidt, P. S., Bertness, M. D., & Rand, D. M. (2000). Environmental heterogeneity and balancing selection in the acorn barnacle *Semibalanus balanoides*. *Proceedings of the Royal Society B: Biological Sciences*, *267*(1441), 379–384.

Schmidt, P. S., & Rand, D. M. (1999). Intertidal microhabitat and selection at Mpi: Interlocus contrasts in the northern acorn barnacle, *Semibalanus balanoides*. *Evolution*, *53*(1), 135–146.

Schmidt, P. S., & Rand, D. M. (2001). Adaptive maintenance of genetic polymorphism in an intertidal barnacle: Habitat- and life-stage-specific survivorship of Mpi genotypes. *Evolution*, *55*(7), 1336–1344.

Schmidt, P. S., Serrao, E. A., Pearson, G. A., Riginos, C., Rawson, P. D., Hilbish, T. J., Brawley, S. H., Trussell, G. C., Carrington, E., Wethey, D. S., Grahame, J. W., Bonhomme, F., & Rand, D. M. (2008). Ecological genetics in the North Atlantic: Environmental gradients and adaptation at specific loci. *Ecology*, *89*(11), S91–S107.

Sved, J. A., Reed, T. E., & Bodmer, W. F. (1967). The number of balanced polymorphisms that can be maintained in a natural population. *Genetics*, *55*(3), 469–481.

Tracey, W. D., Wilson, R. I., Laurent, G., & Benzer, S. (2003). Painless, a Drosophila gene essential for nociception. *Cell*, *113*(2), 261–273.

Veliz, D., Bourget, E., & Bernatchez, L. (2004). Regional variation in the spatial scale of selection at MPI* and GPI* in the acorn barnacle *Semibalanus balanoides* (Crustacea). *Journal of Evolutionary Biology*, *17*(5), 953–966.

Veliz, D., Duchesne, P., Bourget, E., & Bernatchez, L. (2006). Genetic evidence for kin aggregation in the intertidal acorn barnacle (*Semibalanus balanoides*). *Molecular Ecology*, *15*(13), 4193–4202.

Vitti, J. J., Grossman, S. R., & Sabeti, P. C. (2013). Detecting natural selection in genomic data. *Annual Review of Genetics*, *47*, 97–120.

Wagner, D. N., Baris, T. Z., Dayan, D. I., Du, X., Oleksiak, M. F., & Crawford, D. L. (2017). Fine-scale genetic structure due to adaptive divergence among microhabitats. *Heredity (Edinb)*, *118*(6), 594–604.

Westram, A. M., Rafajlović, M., Chaube, P., Faria, R., Larsson, T., Panova, M., Ravinet, M., Blomberg, A., Mehlig, B., Johannesson, K., & Butlin, R. (2018). Clines on the seashore: The genomic architecture underlying rapid divergence in the face of gene flow. *Evolution Letters*, *2*(4), 297–309.

Wiberg, R. A. W., Gaggiotti, O. E., Morrissey, M. B., Ritchie, M. G., & Johnson, L. (2017). Identifying consistent allele frequency differences in studies of stratified populations. *Methods in Ecology and Evolution*, *8*(12), 1899–1909.

Williams, G. (1966). *Adaptation and natural selection: A critique of some current evolutionary thought*. Princeton University Press.

Xu, S. Y., Cang, C. L., Liu, X. F., Peng, Y. Q., Ye, Y. Z., Zhao, Z. Q., & Guo, A. K. (2006). Thermal nociception in adult Drosophila: Behavioral characterization and the role of the painless gene. *Genes, Brain and Behavior, 5*(8), 602–613.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.