



Ecological Load and Balancing Selection in Circumboreal Barnacles

Joaquin C.B. Nunez ^{†,1} Stephen Rong,^{1,2} Alejandro Damian-Serrano,³ John T. Burley,^{1,4} Rebecca G. Elyanow,² David A. Ferranti,¹ Kimberly B. Neil,¹ Henrik Glenner,⁵ Magnus Alm Rosenblad,⁶ Anders Blomberg,⁶ Kerstin Johannesson,⁷ and David M. Rand ^{*,1,2}

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

²Center for Computational Molecular Biology, Brown University, Providence, RI

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

⁴Institute at Brown for Environment and Society, Brown University, Providence, RI

⁵Department of Biological Sciences, University of Bergen, Bergen, Norway

⁶Department of Chemistry and Molecular Biology, University of Gothenburg, Lundberg Laboratory, Göteborg, Sweden

⁷Department of Marine Sciences, University of Gothenburg, Tjärnö Marine Laboratory, Strömstad, Sweden

[†]Present address: Department of Biology, University of Virginia, Charlottesville, VA

*Corresponding authors: E-mails: joaquin_nunez@brown.edu; david_rand@brown.edu.

Associate editor Michael Rosenberg

Abstract

Acorn barnacle adults experience environmental heterogeneity at various spatial scales of their circumboreal habitat, raising the question of how adaptation to high environmental variability is maintained in the face of strong juvenile dispersal and mortality. Here, we show that 4% of genes in the barnacle genome experience balancing selection across the entire range of the species. Many of these genes harbor mutations maintained across 2 My of evolution between the Pacific and Atlantic oceans. These genes are involved in ion regulation, pain reception, and heat tolerance, functions which are essential in highly variable ecosystems. The data also reveal complex population structure within and between basins, driven by the trans-Arctic interchange and the last glaciation. Divergence between Atlantic and Pacific populations is high, foreshadowing the onset of allopatric speciation, and suggesting that balancing selection is strong enough to maintain functional variation for millions of years in the face of complex demography.

Key words: barnacles, *Semibalanus balanoides*, balancing selection, ecological genomics, ecological load.

Introduction

The relationship between genetic variation and adaptation to heterogeneous environments remains a central conundrum in evolutionary biology (Botero et al. 2015). Classical models of molecular evolution predict that populations should be locally adapted to maximize fitness (Williams 1966). However, species inhabiting highly heterogeneous environments violate this expectation: If gene flow is high in relation to the scale of environmental heterogeneity, species will harbor variation that is beneficial in one condition but deleterious in another (Gillespie 1973), and the resulting ecological load (i.e., the fitness difference between the best and the average genotype across the range of environments where offspring may settle) will prevent local adaptation. Conversely, if gene flow is low, favored alleles will become locally fixed and species should display low levels of genetic variation. Paradoxically, many natural populations living in variable environments possess high dispersal capabilities and harbor more variation than expected under classical models (Metz and Palumbi 1996;

Mackay et al. 2012; Messer and Petrov 2013; Bergland et al. 2014). This disconnect between data and theory has motivated the hypothesis that balancing selection, a process where selection favors multiple beneficial alleles at a given locus, is at play to maintain adaptations in these habitats (Levene 1953; Hedrick 2006).

The northern acorn barnacle (*Semibalanus balanoides*) is a model system to study adaptations to ecological variability. This barnacle is a self-incompatible, simultaneous hermaphrodite which outcrosses only with adjacent individuals. Adult barnacles are fully sessile and occupy broad swaths of intertidal shores in both the North Pacific and North Atlantic oceans. These habitats experience high levels of cyclical and stochastic ecological heterogeneity which impose strong selection at multiple spatial scales: microhabitats (intertidal shores), mesohabitats (bays and estuaries), and macrohabitats (continental seabords) (Schmidt et al. 2008; Nunez et al. 2020). Barnacle larvae, on the other hand, engage in extensive pelagic dispersal by ocean currents (70–100 km in 5–7 weeks) and may settle in habitats completely different from those of

© The Author(s) 2020. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

Open Access

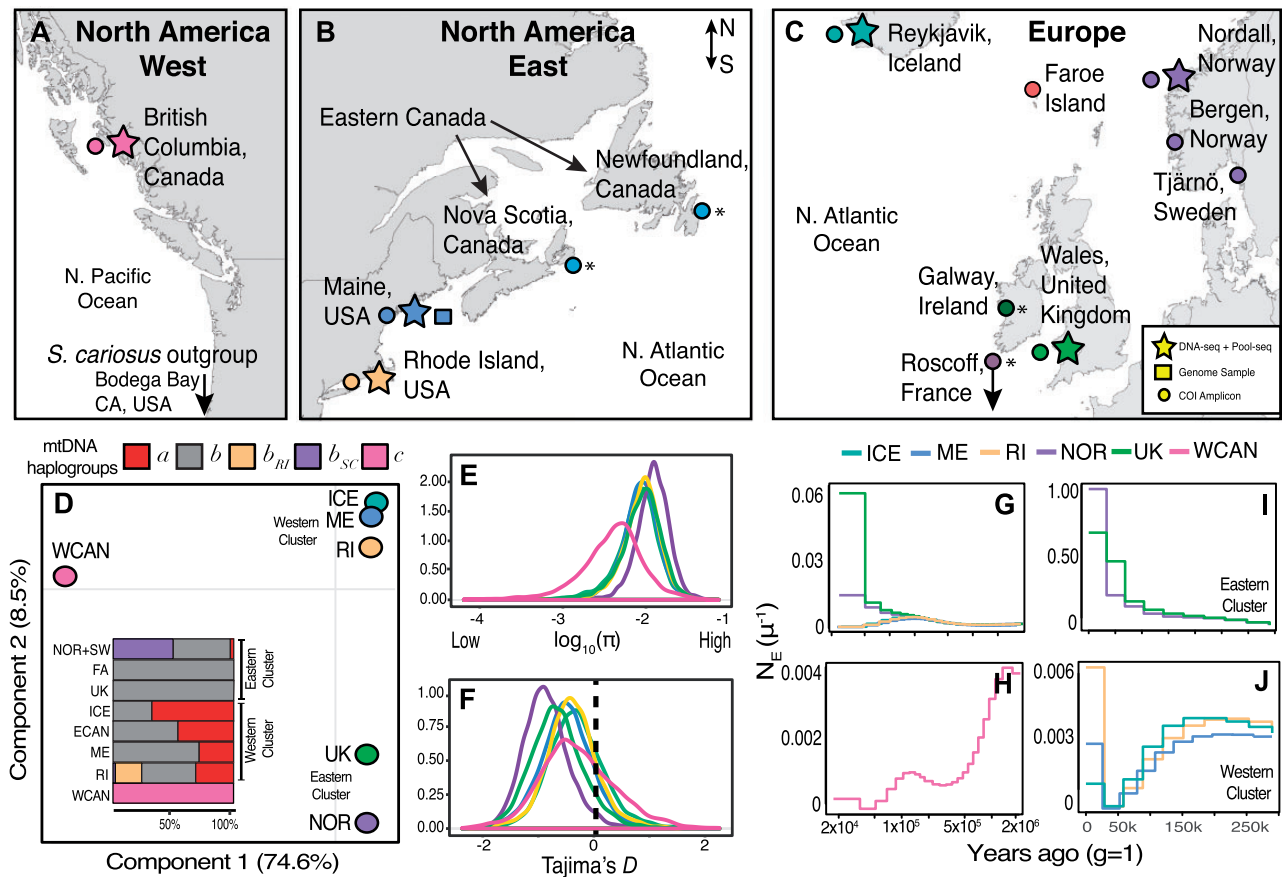


Fig. 1. Genetic variation and phylogeography. (A) Map of the North Pacific coast of North America with collection sites indicated. (B) Collections in the Atlantic Eastern coast of North America. (C) Collections in the Atlantic European coast. For (A), (B), and (C), stars indicate sites where a single individual and a pool of multiple individuals were collected, the square indicates the site from which the reference genome was constructed, and the circles indicate sites where *COX I* data were collected. The asterisks indicate cases where *COX I* data were downloaded. (D) PCA with pool-seq data from all populations. The colors represent populations: Pacific Canada (WCAN; pink), Maine (ME; blue), Rhode Island (RI; yellow), Iceland (ICE; dark green), Norway (NOR; purple), United Kingdom (UK; light green). (D-inset) Distribution of mitochondrial haplotypes across all populations. The names *a*, *b* (including *b_{RI}* and *b_{SC}*), and *c* represent common mtDNA haplotypes observed in populations. (E) Nucleotide diversity ($\log_{10} \pi$) for all nuclear genes across all populations. (F) Tajima's *D* for all nuclear genes across all populations. The dashed vertical line marks 0, the expected value under a neutral model. The y -axis in (E) and (F) shows the density of observations. (G) Demographic reconstruction for North Atlantic individuals showing demographic changes from 2 Ma to 200 ka. (H) North Pacific individual showing demographic changes from 2 Ma to 200 ka. (I) Demographic changes in British and Norwegian individuals. (J) Plot of recent (today–250 ka) demographic changes in the North American and Icelandic individuals.

their parents (Flowerdew 1983). This contrast between strong adult selection and high juvenile dispersal prevents local adaptation. In addition, *S. balanoides* has a complex demography. It originated in the Pacific, and colonized the Atlantic during the many waves of the trans-Arctic interchange (1–3 Ma) (Vermeij 1991). Like most circumboreal species, it was subjected to drastic range shifts due to the Pleistocene glacial cycles (Wares and Cunningham 2001; Flight et al. 2012), and more recently due to anthropogenic climate change (Jones et al. 2012). As such, *S. balanoides* is a premier system to study how adaptive genetic variation is maintained over broad spatial and evolutionary scales, in the face of ecological load.

Three decades of work have shown that balancing selection, via marginal overdominance (a case where the harmonic mean fitness of heterozygous genotypes must be larger than that of either homozygote) (Levene 1953), maintains adaptive variation at the metabolic gene Mannose-6-phosphate

isomerase (*Mpi*) in barnacles across the entire North Atlantic basin (Schmidt and Rand 1999; Dufresne et al. 2002; Rand et al. 2002; Veliz et al. 2004; Nunez et al. 2020). These findings motivate two questions which are addressed in this study. First, how pervasive are balanced polymorphisms in the barnacle genome? And, second, what genes are targets of balancing selection? To investigate functional polymorphism in *S. balanoides*, we quantified genomic variation in North Pacific and North Atlantic populations (fig. 1A–C). In the Pacific, we analyzed samples from British Columbia, Canada (WCAN) as well as a sample of the sister taxon *S. cariosus*. In the Atlantic, we analyzed samples from Maine (ME), Rhode Island (RI), Iceland (ICE), Norway (NOR), and the United Kingdom (UK). For all populations, we sequenced multiple libraries including: a single individual barnacle genome to $\sim 50\times$ coverage, pools of 20–38 individuals per population (i.e., pool-seq; Schlotterer et al. 2014), as well

as ~600-bp amplicons from the mitochondrial (mtDNA) *COX I* gene (including previously published *COX I* data (Wares and Cunningham 2001)). We mapped these data sets to our newly assembled *S. balanoides* genome (supplementary appendix 1, Supplementary Material online) and characterized genetic diversity across all populations (supplementary appendix 2, Supplementary Material online). We first present our findings in the context of the barnacle's phylogeography and demographic history. This is pivotal to understand the historical conditions which can contribute to ecological load. Then, we characterize the pervasiveness of balancing selection across the genome, as well as the age of balanced polymorphisms and their putative functional significance in highly heterogeneous environments.

Results

Standing Variation across Oceans

Our pool-seq panels discovered ~3M high-quality single nucleotide polymorphisms (SNPs) across populations at common allele frequencies (>5%). When linkage is removed at 500 bp, the SNP panel thins to ~690,000. Principal component analysis (PCA), on the Linkage disequilibrium (LD) thinned SNPs, shows that variation is strongly subdivided by ocean basins (fig. 1D). PC1 captures 74% of the variation and partitions populations across basins. PC2 (8.5% var.) partitions Atlantic populations into two discrete east–west clusters. The western cluster contains ME, RI, and ICE, and the eastern cluster contains UK and NOR. These clusters are supported by the abundance of mtDNA haplotypes within and between ocean basins (fig. 1D inset; supplementary table S1, Supplementary Material online) (Wares and Cunningham 2001; Flight et al. 2012; Nunez et al. 2018). The large divergence between oceans is also captured in levels of nucleotide diversity (π ; a metric of standing genetic variation). Surprisingly, North Atlantic populations harbor more genetic variation ($\pi = 1.05\%$) than their Pacific, ancestral, conspecifics ($\pi = 0.55\%$; fig. 1E; supplementary fig. S1, Supplementary Material online). We also estimated the Tajima's *D* statistic (*D*), a neutrality test based on the comparison of two measures of nucleotide polymorphism: π and Watterson theta (θ). The value of the *D* statistic is a good proxy for the excess ($D < 0$), or deficit ($D > 0$), of rare alleles in populations. These data indicate that all North Atlantic populations, especially NOR, have negatively skewed genome-wide values of *D* (fig. 1F, supplementary fig. S2, Supplementary Material online).

Historical Phylogeography and Structure

We reconstructed changes of historical effective population sizes (N_e) with the multiple sequentially Markovian coalescent model (MSMC) using individual whole genomes (Schiffels and Durbin 2014). Our results provide evidence for different phylogeographic trajectories in response to the events of the glaciations (fig. 1G and H). For instance, the Eastern Cluster and the Western Cluster populations shared a common demography throughout the Pleistocene (fig. 1G) but diverged in recent geological time. Namely, Eastern

populations (especially NOR) experienced striking increases in N_e in the recent past (fig. 1I), likely following the asynchronous deglaciation of the Fennoscandian ice sheet (Ruddiman and McIntyre 1981; Patton et al. 2017). Western populations, on the other hand, experienced a demographic contraction which started during the last glacial period and ended during the last glacial maxima (~20 ka; fig. 1J) (Brochmann et al. 2003; Maggs et al. 2008; Flight et al. 2012).

We estimated gene flow by computing f_3 statistics (Reich et al. 2009) for all possible combinations of target, source 1, and source 2 populations, using individual whole genomes (supplementary fig. S3 and table S2, Supplementary Material online). Our analysis finds no evidence of recent gene flow across oceans. This result is supported by two additional lines of evidence. First, a mtDNA molecular clock analysis (Drummond et al. 2002) suggests that Pacific and Atlantic populations have not exchanged migrants in nearly 2 My (supplementary appendix 3, Supplementary Material online); and second, estimates of genetic differentiation (F_{ST}) reveal large amounts of genome-wide divergence (supplementary fig. S4, Supplementary Material online) and foreshadow the onset of allopatric speciation across oceans. Within the North Atlantic, F_{ST} is low (likely due to shared demography until the glacial maximum) and the f_3 analysis suggests that admixture is pervasive (supplementary fig. S3 and table S2, Supplementary Material online). These findings are supported by additional ABBA–BABA tests for gene tree heterogeneity (Green et al. 2010) (see supplementary appendix 4, Supplementary Material online). Overall, these findings present three important points: First, they exemplify the complex demography that underlies standing variation in natural populations; second, they confirm that barnacles harbor high levels of genetic variation genome-wide; and third, they reveal the pervasiveness of gene flow and shared variation within ocean basins, where environmental heterogeneity is extensive across “micro” (1–3 m) and “meso” (1–10 km) scales. These conditions provide the environmental context for ecological load at the genomic scale.

Balancing Selection in Barnacles

Balancing selection is expected to produce molecular and phylogenetic footprints not consistent with neutrality (Fijarczyk and Babik 2015). Molecular footprints include: enrichment of old alleles (e.g., trans-species polymorphisms; TSPs), elevated genetic variation (high π), deficit of rare alleles ($D > 0$), excess SNPs at medium allele frequencies, reduced divergence around the balanced locus (low F_{ST}), as well as the accumulation of nonsynonymous variation in the vicinity of balanced polymorphisms, a phenomenon known as sheltered load (Uyenoyama 2005). Likewise, balancing selection will produce a phylogenetic signal composed of diverged clades, corresponding to the balanced haplotypes. Deeply diverged clades will occur when balancing selection has maintained variation over long evolutionary times (i.e., ancestral balancing selection; Fijarczyk and Babik 2015). Notably, these signatures may become highly localized in the genome as the action of recombination over long periods of time will erode long-distance haplotype signatures.

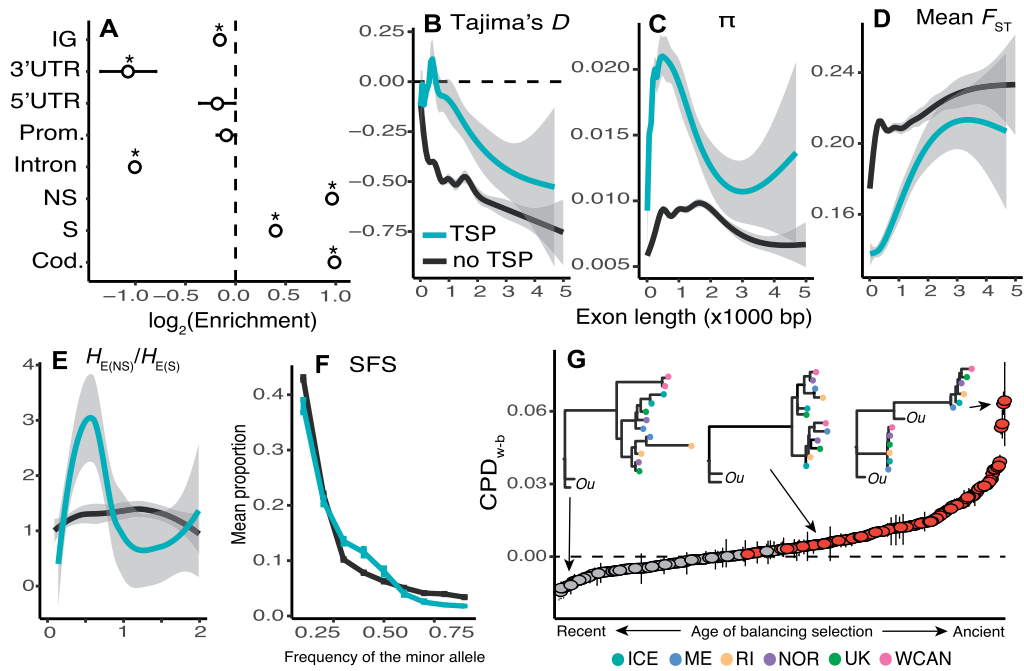


FIG. 2. Evidence for balancing selection across the genome. (A) Enrichment analysis of TSPs across the genome of *Semibalanus balanoides* based on all populations studied. The asterisk symbols represent statistical significance. Prom., promoters; NS, nonsynonymous loci; S, synonymous loci; Cod., coding loci. (B) Plot of Tajima's D (as a function of length) of exons bearing TSPs versus all other exons not bearing TSPs. (C) Same as (B) but for nucleotide diversity (π). (D) Same as (B) but for mean F_{ST} . (E) Same as (B) but for the ratio of nonsynonymous heterozygosity to synonymous heterozygosity. (F) SFS for whole genes with TSPs versus other genes. Vertical bars are 95% confidence intervals. (G) Candidate genes under balancing selection ranked according to their CPD_{w-b} values (interquartile ranges shown as error bars). Red values indicate statistical significance. Horizontal dashed line indicates $CPD_{w-b} = 0$. In the x-axis, the label "ancient" refers to allele trees whose topology violates the genome-wide phylogeographic expectation (e.g., fig. 1D). "Recent" denotes the opposite case. Three example allele tree topologies are shown. The sister taxon, *Semibalanus cariosus*, is shown as "Ou" (for outgroup). The x-axis for (B), (C), (D), and (E) is exon length ($\times 1,000$ bp).

A joint analysis of our Pacific, Atlantic, and outgroup (*S. cariosus*) data sets reveal 11,917 cosmopolitan SNPs (i.e., SNPs that segregate in all populations across both oceans) which are also TSPs (supplementary data set S1, Supplementary Material online). TSPs, genome-wide, occur in 0.14% coding regions, 0.21% in introns, 0.02% in promoters, 0.01% in 5'-UTRs, and $<0.01\%$ in 3'-UTRs. The remainder of TSPs occurs in 0.09% of intergenic regions. An enrichment analysis which compares the abundance of TSPs, of each genomic class, relative to all discovered SNPs, reveals that TSPs are significantly overenriched in coding loci (fig. 2A), and 4,415 segregate at high frequencies in all populations (TSPs with heterozygosity $[H_E] > 0.30$; supplementary fig. S5, Supplementary Material online). These patterns of variation could be the result of neutral processes such as recurrent mutation (homoplasy) across all populations of either species. However, the enrichment of cosmopolitan, nonsynonymous, TSPs at common frequencies is not consistent with neutrality. Under a model of strict neutrality, segregating mutations are eventually lost in populations after speciation (Clark 1997). Moreover, coding regions are subjected to purifying selection which removes deleterious and mildly deleterious nonsynonymous variants (Hartl and Clark 1997).

We compared patterns of genetic variation in exons bearing TSPs and other exons. When accounting for exon length, we observe consistently elevated values of D and π for TSP-

bearing exons relative to other exons (fig. 2B and C; supplementary fig. S6, Supplementary Material online). Except for the ME versus RI comparison (supplementary fig. S7, Supplementary Material online), TSP-bearing exons have consistently low F_{ST} values (fig. 2D). To quantify sheltered load, we compared the ratio of H_E values at nonsynonymous and synonymous mutations in TSP-bearing and other exons. Our results show that medium sized TSP-bearing exons (~ 500 bp) harbor an excess of nonsynonymous H_E (fig. 2E). Notably, we observed that differences between TSP-bearing and other exons become less apparent as exons get longer. The relationship between exon size and the intensity of the balancing selection signatures depends on local recombination rates. Although exact recombination rates are not yet available for *Semibalanus*, empirical data suggest that LD decays at distances <1 kb (supplementary fig. S8, Supplementary Material online). As such, the signals of deviation from neutrality are more apparent on shorter exons, relative to longer ones. We observe 1,107 TSPs that cause nonsynonymous changes and occur in 312 genes with high-confidence annotations (4%; supplementary data set S2, Supplementary Material online). Consistent with our expectation of balancing selection, site frequency spectrum (SFS) analyses show that these 312 candidate genes harbor an excess of SNPs at medium allele frequencies relative to other annotated genes (fig. 2F).

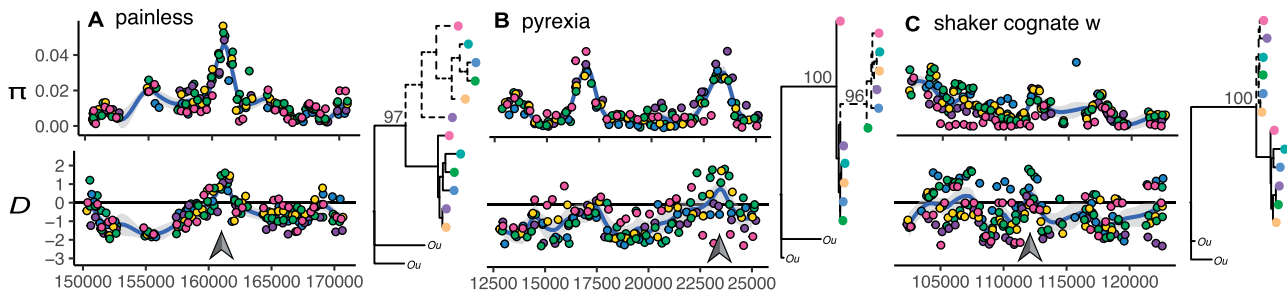


Fig. 3. Balancing selection on ecologically important genes. We present patterns of genetic variation (π and D estimated from pool-seq data, and allele tree topologies estimated from single individuals) for three example genes: (A) *painless* (*Pain*), (B) *pyrexia* (*Pyx*), and (C) *shaker cognate w* (*Shaw*). Gray arrows show regions that contain TSPs. In Tajima's D panels, the horizontal line marks the $D = 0$ point. For all trees, the sister taxon, *Semibalanus cariosus*, is shown as “Ou.” The colors represent populations: WCAN (pink), ME (blue), RI (yellow), ICE (dark green), NOR (purple), UK (light green). The x-axis shows base pair position within scaffolds.

Age of Balanced Polymorphisms

To determine the age of the putatively balanced polymorphisms, we ran topological tests on the allele trees for each TSP region across the 312 candidate genes. We built trees using phased haplotypes for each TSP-bearing region for all single-individual genomes. We used these allele trees to compute the cophenetic distance (CPD) between tips. We classified allele trees as having or lacking highly diverged alleles based on the relative mean CPD between haplotypes from the same population versus from different populations (CPD_{w-b} ; see [supplementary methods](#), [Supplementary Material](#) online). The analysis reveals that of the 312 allele trees, 150 carry a significant signature of ancestral balancing selection ($CPD_{w-b} > 0$, Bonferroni $P < 1 \times 10^{-9}$; [fig. 2G](#); [supplementary data](#) set S2, [Supplementary Material](#) online). This suggests maintenance of diverged haplotypes for more than 2 My, with extreme cases in which haplotypes are shared across species (8–10 My) ([Perez-Losada et al. 2008](#); [Herrera et al. 2015](#)). The remaining genes with $CPD_{w-b} < 0$ may represent either cases where the balanced alleles are younger or oversampling of homozygous individuals for any given marker.

Targets of Selection

We partitioned our data set among candidate genes with positive and negative CPD_{w-d} allele trees and conducted gene ontology (GO) enrichment analyses. The 150 genes with positive CPD_{w-d} trees show enrichment for terms related to “ion channel regulation,” including genes involved in environmental sensing and circadian rhythm regulation ([supplementary table S3](#), [Supplementary Material](#) online). We show examples for three candidate genes under ancestral balancing selection involved in environmental sensing: 1) the *painless* gene (*Pain*; g1606; [fig. 3A](#)), which is involved in nociception (i.e., pain reception), as well as detection of heat and mechanical stimuli ([Tracey et al. 2003](#); [Xu et al. 2006](#)); 2) the *Pyrexia* gene (*Pyx*; g3472; [fig. 3B](#)), which is involved in negative geotaxis, and responses to heat ([Lee et al. 2005](#)); and 3) the *shaker cognate w* gene (*Shaw*; g3310; [fig. 3C](#)), which is involved in regulation of circadian rhythm ([Hodge and Stanewsky 2008](#); [Buhl et al. 2016](#)). These three examples showcase canonical footprints of balancing selection around the TSP, concomitant with a bimodal allele tree.

Among genes with negative CPD_{w-d} we observe enriched functions for “anatomical structure formation” including genes coding for motor proteins and muscle genes ([supplementary table S4](#), [Supplementary Material](#) online). In all cases, we used RNA-seq data from ME individuals to confirm that these loci are expressed in adult barnacles.

Discussion

In intertidal barnacles, the dichotomy of strong adult selection and high offspring dispersal means that any allele that is beneficial to parental fitness in one generation may be neutral or deleterious in the next ([Gillespie 1973](#)). This leads to a fundamental question in evolutionary biology: How are adaptations maintained in the face of extreme ecological variability? In this article, we provide evidence suggesting that balancing selection is widespread across the barnacle genome, with 4% of annotated genes harboring putatively balanced polymorphisms. Notably, these polymorphisms occur in genes with functions that may be important for life in variable environments, and many have been maintained for at least 2 My despite a complex phylogeographic history ([Wares and Cunningham 2001](#); [Flight and Rand 2012](#)). Naturally, the heterogeneous nature of the rocky intertidal imposes a segregation “cost” for these balanced polymorphisms, as they occur in individuals that, due to high dispersal, recruit in suboptimal habitats for any given genetic makeup. This ecological load, defined as $L_e = (W_{max} - \bar{W}) / W_{max}$ (where \bar{W} is mean fitness, and W_{max} is optimal fitness, across all habitats), will be substantial, as demonstrated by comprehensive recruitment studies in natural habitats ([Bertness 1989](#); [Bertness et al. 1992](#); [Pineda et al. 2006](#)). For example, at initial settlement, barnacle density can be as high as 76 individuals per cm^2 , but at maturity, it can be as low as 0.15 individuals per cm^2 (0.2% survival) ([Pineda et al. 2006](#)). This mass mortality is habitat- and genotype dependent ([Schmidt and Rand 2001](#)). This is the type of “fitness cost” envisioned in the Levene model of balancing selection ([Levene 1953](#)). As such, our data suggest that the problem of ecological load is a defining condition of the barnacle life cycle. And, more generally, it argues that balancing selection, via marginal overdominance, may be the fundamental process underlying maintenance of adaptation in variable environments.

Is Pervasive Balancing Selection Plausible in Nature?

Under classical models of population genetics, when loci are considered to be independent of each other, the additive effects of widespread balanced polymorphism result in unbearable amounts of fitness variance and genetic death (Kimura and Crow 1964; Lewontin and Hubby 1966). However, if balanced loci have interactive effects (e.g., epistasis), multiple polymorphisms could be maintained with minimum effects on the distribution of fitness variance (King 1967; Milkman 1967; Sved et al. 1967; Wittmann et al. 2017). Based on this theoretical framework, multiple models have been developed to describe the conditions that favor the long-term maintenance of functional variation in spatially varying environments (Gillespie 1973; Hedrick et al. 1976). Moreover, polymorphisms will be less likely to be lost if there is a large number of ecological niches available, if there is migration among niches, and if individuals are proactive in choosing niches where their fitness is maximized (Hedrick et al. 1976). We argue that barnacles satisfy these conditions to some degree.

First, although it is useful to summarize intertidal heterogeneity in the form of discrete microhabitats (Schmidt et al. 2000), individual barnacles experience the rocky shore as a complex tapestry of interactive stressors at three spatial levels. At microhabitats scales, the upper and lower tidal zones pose diametrically different ecological challenges in terms of food availability, competition, predation, and risk of desiccation (Bertness et al. 1991; Schmidt and Rand 1999, 2001). At mesohabitat scales, open coasts versus sheltered estuaries vary in their exposure to wave action, upwelling dynamics, and biotic interactions (Sanford and Menge 2001; Dufresne et al. 2002; Veliz et al. 2004). These, in turn, modify microlevel stressors. Lastly, at macrohabitat scales, topological differences across shores and latitudinal variations in tidal range produce a mosaic of thermal stress along continents (Helmuth et al. 2002). Consequentially, what selection pressures are more important for any given barnacle will emerge from the interactions among these stress gradients. This complex landscape of selection has been captured in studies of the barnacle *Mpi* gene. Accordingly, the locus is under selection at microlevels in the Gulf of Maine (Schmidt and Rand 1999; Schmidt et al. 2000), at mesolevels in the gulf of St Lawrence (Canada) (Dufresne et al. 2002; Veliz et al. 2004), yet it shows tepid signs of selection in the Narragansett Bay (Rhode Island) (Rand et al. 2002; Nunez et al. 2020). Similar complexity has also been captured in temperate populations of *Drosophila*. In these, idiosyncratic weather effects can alter the dynamics of seasonal adaptation (Bergland et al. 2014; Machado et al. 2019). Second, the high dispersal capacity of the larval stage ensures constant migration between these niches across generations. Finally, barnacles also have the ability to choose preferred substrates during settlement. This occurs during the spring when barnacle larvae extensively survey microhabitats for biological, chemical, and physical cues produced by previous settlers before making final commitments of where to settle (Bertness et al. 1992). Unfortunately for the barnacle, this capacity for substrate choice does not mitigate mass mortality during late summer, which leads to strong selection for

particular genotypes (Schmidt and Rand 2001). Currently, there is limited evidence for genotype-specific substrate selection or nonrandom settlement (Veliz et al. 2006). A cohort-tracking and sequencing experiment could be utilized to address this question (these experiments are underway). If true, these behaviors may constitute a form of adaptive plasticity, helping barnacles choose habitats where their fitness may be marginally improved. Overall, this suggests that the barnacle's life history is conducive to the maintenance of balanced polymorphisms.

Pool-seq in Ecological Genomics

Our analysis was conducted using pool-seq data. As such, it is important to highlight known caveats associated with this genotyping technology (Anderson et al. 2014; Anand et al. 2016; Nunez et al. 2018). Although incredibly cost-effective, the accuracy of pool-seq is highly dependent on the number of individuals pooled, sequencing coverage, as well as sequencing technology. These caveats can become pronounced when working on nonmodel systems where enforcing uniform sample sizes across populations may be logistically challenging. Nevertheless, pool-seq experiments that deviate from the recommended design (Gautier et al. 2013) result in inaccurate estimates of allele frequency, including undersampling of rare alleles and oversampling of fixed sites (Anderson et al. 2014). These systemic errors have notable impacts when estimating demographic parameters. We ameliorated these shortcomings using a two-pronged approach. First, for each population sampled, we sequenced both a single individual and a pool. The single individual allowed us to estimate demographic parameters. The pool, on the other hand, allowed us to survey common variation across populations. And thus, although each individual approach has unique shortcomings, their combination provides a robust data set to address the questions presented in this study. In addition, because one must filter out sequencing errors, most implementations of high-throughput sequencing in ecological genomics produce skewed SFS distributions by undersampling low-frequency mutations (Achaz 2008, 2009). This problem is exacerbated for pool-seq experiments and can produce biased estimates of common statistics such as θ and, consequently, Tajima's D . However, because we are interested in understanding patterns of genetic variation at common variants, our analyses are less susceptible to this drawback.

What Variation Is under Selection?

Our analyses suggest that 4% (312) of all annotated genes are candidates of balancing selection across the entire range of the species. Although follow-up experiments are needed to determine the replicability and functional importance of these variants, our evidence for balancing selection is consistent with patterns reported for other species. For example, the number of candidate genes in *Semibalanus* is like that observed in *Arabidopsis thaliana* and its close relative *Capsella rubella* (433 genes) (Wu et al. 2017). Similar to *Semibalanus*, these plants diverged ~ 8 Ma, and their natural populations experience high levels of ecological heterogeneity (Bakker et al. 2006). We must acknowledge that our number may

be an underestimation driven by the nascent state of the genomic tools in *Semibalanus*. Future genome assemblies, combined with improved annotations, will undoubtedly yield a more complete picture of functional variation in the species. In addition, it will allow for a more comprehensive characterization of selection in structural variants and regulatory loci, which have been shown to be fundamental in the evolution of complex phenotypes (Wray 2007; Faria et al. 2019). Despite these limitations, our analysis recovered many candidate genes involved in functions which may be key for life in variable environments. Without more functional validation, the connections between these genes and barnacle ecology are merely speculative. However, many of these candidates have been studied in other systems in the context of stress tolerance. Consequentially, they are fertile grounds for hypothesis generation and follow-up experiments. For instance, the general enrichment for ion channel genes suggests selection related to osmotic regulation (Sundell et al. 2019). This hypothesis is highly plausible given that intertidal ecosystems experience strong salinity fluctuations, repeatedly exposing barnacles to osmotic challenges at all spatial scales. In addition, we observe targets of selection involved in environmental sensing loci (e.g., *pain*, *pyx*, and *shaw*; fig. 3). Similar to osmotic regulation, selection on these genes is entirely plausible given the inherent variability of intertidal habitats. An important hypothesis from the allozyme era is the idea that balancing selection would target genes at the node of metabolic fluxes (Eanes 1999; Watt and Dean 2000). In such cases, balanced variation would provide biochemical flexibility to cope with environmental heterogeneity. In the same vein, we hypothesize that balancing selection may act more often on “sensor genes” which control plastic responses to ecological variation. Testing this hypothesis is beyond the scope of this study and would require the use of allele-specific differential expression experiments in barnacles. We also note that evidence of balancing selection and TSPs at the *Mpi* gene are discussed in Nunez et al. (2020).

Complex Demography and Speciation

Our demographic analyses provide clues about how historical events affected genetic variation in barnacle populations. In the Atlantic, our evidence suggests a shared demography throughout the Pleistocene, and that the modern Eastern and Western clusters formed in response to recent events of the last glacial cycle. These findings highlight that the low F_{ST} values observed within the basins arise due to shared ancestry. Moreover, they also suggest that population structure persists in the presence of gene flow. As such, although larvae have the capacity to disperse for hundreds of kilometers, ocean currents (Nunez et al. 2018) and different estuarine flushing times (Brown et al. 2001) allow regions to retain some level of geographical structuring (Johannesson et al. 2018; Nunez et al. 2018). Comparisons between oceans reveal a stark pattern of genome-wide divergence. This pattern is driven by the separation of Pacific and Atlantic populations following the events of the trans-Arctic interchange (Vermeij 1991). Accordingly, the negative levels of D in the North Atlantic may reflect the effect of bottlenecks during

the trans-Arctic interchange. Notably, the high levels of π in the Atlantic are not concordant with predictions of common colonization models in which variation of the younger population is a subset of the ancestral population (Maggs et al. 2008). We hypothesize that this could be the result of ancient admixture due to repeated trans-Arctic invasions from the Pacific (Väinölä 2003). We recognize that ancestral admixture could generate artificial signatures of balancing selection via the mixing of highly differentiated haplotypes. However, such an occurrence would affect most genes in the genome. Our evidence shows that the signatures of balancing selection are highly localized in TSP regions. For example, although D is elevated in TSP regions, it is negatively skewed genome-wide. Our data do not support recent gene flow between ocean basins. As such, after 2 My of separation, neutral divergence appears to be driving Atlantic and Pacific populations to speciate in allopatry. A closer look to this hypothesis will require crossing individuals from both basins, and surveying offspring fitness and viability. More salient, however, is the observation of shared haplotypes between oceans in our candidate genes for balancing selection. In light of such strong background divergence, this provides evidence that balancing selection on most of these genes is strong and that polymorphisms have been maintained for long periods of time.

Materials and Methods

Barnacle Collections

Barnacle samples were collected from Damariscotta (Maine, United States; ME), Jamestown (Rhode Island, United States; RI), Calvert Island (British Columbia, Canada; WCAN), Reykjavik (Iceland; ICE), Porthcawl (Wales, United Kingdom; UK), and Norddal (Norway; NOR). Additional samples were collected in Bergen (Norway), Tórshavn (Faroe Island), and Tjärnö (Sweden). For all samples, species identities were confirmed using Sanger sequencing of the mtDNA COX I region (Bucklin et al. 2011). For the WCAN, RI, ME, ICE, UK, and NOR populations, we collected a single individual for DNA-seq, and a group of 20–40 individuals for pool-seq (supplementary appendix 2, Supplementary Material online). RNA-seq was done on four individuals from Maine. DNA-seq was done on a single individual from the sister taxa *S. cariosus*. DNA/RNA was extracted using Qiagen DNeasy/RNeasy kits. All pools and single individuals were sequenced in their own lanes of an Illumina machine by GENEWIZ LLC using 2 × 150 paired-end configuration.

Mapping Data Sets to the Genome

Samples were mapped to a genome assembled de novo for the species (Sba13.1; NCBI GenBank accession: VOPJ000000000; BioProject: PRJNA557548; BioSample: SAMN12406453; supplementary appendix 1, Supplementary Material online). The genome was assembled using a hybrid approach which combines PacBio reads and Illumina reads using DBG2OLC (Ye et al. 2016) and Redundans (Pryszcz and Gabaldon 2016). Gene models were constructed using an ab initio method, AUGUSTUS (Stanke and Waack 2003), informed by evidence from the RNA-seq. A gene feature file (GFF) is available as

supplementary data set S4, [Supplementary Material](#) online. The model used for gene prediction was trained in *Drosophila melanogaster*. Genes were annotated by pairwise blast against the *D. melanogaster* genome (Dmel6; NCBI GenBank: GCA_000001215.4). All annotations are available as [supplementary data](#) set S5, [Supplementary Material](#) online. DNA reads from all populations were mapped to Sbal3.1 using bwa mem (Li 2013). RNA reads were mapped using HiSat2 (Kim et al. 2015). SNPs were called using the samtools pipeline (Li et al. 2009). Short-read phasing was done using HAPCUT2 (Edge et al. 2017). LD in pools was estimated using LDx (Feder et al. 2012).

Genome Analyses

Estimates of π and D were done using the popoolation-1 suite (Kofler, Orozco-terWengel, et al. 2011). Estimations of allele frequencies and F_{ST} were done using the popoolation-2 suite (Kofler, Pandey, et al. 2011). Demographic reconstructions were done using MSMC (Schiffels and Durbin 2014). The f_3 statistics were estimated using treemix (Pickrell and Pritchard 2012). Bayesian molecular clock analyses were done in BEAST2 (Bouckaert et al. 2014). ABBA/BABA statistics were calculated in *Dsuite* (Malinsky et al. 2020). Phylogenetic inferences were done in iQtree (Chernomor et al. 2016). 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 genes list. 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). A detailed description of our analyses can be found in the [supplementary methods](#) section, [Supplementary Material](#) online, as well as in GitHub: <https://github.com/Jcbnunez/BarnacleEcoGenomics>.

Supplementary Material

[Supplementary data](#) are available at *Molecular Biology and Evolution* online.

Acknowledgments

All the authors would like to acknowledge the Centre for Marine Evolutionary Biology (CeMEB) at the University of Gothenburg which organized the Marine Evolution 2018 meeting in which most of the authors met and started this collaborative work. In addition, they acknowledge S. Ramachandran, E. Huerta-Sanchez, D. Sax, D.R. Gaddes, and R.E.F. Gordon for their support and helpful insights, E. Sanford for providing the sample of *Semibalanus cariosus*, M.D. Rand and family for collecting barnacles in Nordall, Norway, and C. Harley for providing the samples from British Columbia. We thank the Natural Environment Research Council, the European Research Council, the Swedish Research Councils VR and Formas (Linnaeus Grant to CeMEB), and *SciLife* Laboratory. This research was conducted using computational resources and services at the Center for Computation and Visualization, Brown University. J.C.B.N., K.B.N., and S.R. were supported by Graduate Research Fellowships (GRFP) from the National Science Foundation

(NSF). J.C.B.N. received additional support from the NSF-Graduate Research Opportunities Worldwide (GROW) fellowship, as well as a fellowship from the Swedish Royal Academy of Sciences (*Kungliga Vetenskapsakademien*; KVA). D.A.F. was supported by a Brown University Karen T. Romer Undergraduate Teaching and Research Award (UTRA). A.D.S. was supported by the NSF (Grant No. OCE-1829835) and a Fulbright Spain Graduate Studies Scholarship. This work was supported by the NSF Integrative Graduate Education and Research Traineeship Grant (IGERT: DGE-0966060) and the National Institutes of Health (NIH: 2R01GM067862) to D.M.R., a Carl Trygger Foundation (CTS 11:14) to M.A.R., the Swedish Research Council (*Vetenskapsrådet*; Grant Nos. 2017-04559 to A.B., 2017-03798 to K.J.), the Meltzer Research Fund to H.G., and the Bushnell Graduate Research and Education Fund (EEB Doctoral Dissertation Enhancement Grant) to J.C.B.N.

Data Deposition

Data used in this study are available in the National Center for Biotechnology Information (NCBI), <https://www.ncbi.nlm.nih.gov> (last accessed September 13, 2020). Raw reads were deposited under submission ID: SUB6188969. SRAs are as follows: DNA-seq data sets: SRR10011798, SRR10011802, SRR10011804, SRR10011805, SRR10011807–SRR10011810, SRR10011812–SRR10011814, SRR10011819, and SRR10011825; PacBio data set: SRR10011818; and RNA-seq data sets: SRR10011820–SRR10011823. MtDNA sequences for the *COX I* genes can be accessed from the following GenBank accession numbers MG925538–MG925662, MG928281–MG928323, and MT329074–MT329592. Whole mtDNAs were deposited under accession numbers MG010647, MG010648, MG010649, MT528636, and MT528637. The barnacle genome (Sbal3.1) is available at NCBI (accession number VOPJ00000000). A GitHub repository with code as well as with the [supplementary data](#) sets S1–S5 can be found at <https://github.com/Jcbnunez/BarnacleEcoGenomics> (last accessed September 13, 2020).

References

- Achaz G. 2008. Testing for neutrality in samples with sequencing errors. *Genetics* 179(3):1409–1424.
- Achaz G. 2009. Frequency spectrum neutrality tests: one for all and all for one. *Genetics* 183(1):249–258.
- Anand S, Mangano E, Barizzone N, Bordoni R, Sorosina M, Clarelli F, Corrado L, Martinelli Boneschi F, D'Alfonso S, De Bellis G. 2016. Next generation sequencing of pooled samples: guideline for variants'. *Sci Rep.* 6(1):33735.
- Anderson EC, Skaug HJ, Barshis DJ. 2014. Next-generation sequencing for molecular ecology: a caveat regarding pooled samples. *Mol Ecol.* 23(3):502–512.
- Bakker EC, Stahl EA, Toomajian C, Nordborg M, Kreitman M, Bergelson J. 2006. Distribution of genetic variation within and among local populations of *Arabidopsis thaliana* over its species range. *Mol Ecol.* 15(5):1405–1418.
- Bergland AO, Behrman EL, O'Brien KR, Schmidt PS, Petrov DA. 2014. Genomic evidence of rapid and stable adaptive oscillations over seasonal time scales in *Drosophila*. *PLoS Genet.* 10(11):e1004775.
- Bertness MD. 1989. Intraspecific competition and facilitation in a northern acorn barnacle population. *Ecology* 70(1):257–268.

- Bertness MD, Gaines SD, Bermudez D, Sanford E. 1991. Extreme spatial variation in the growth and reproductive output of the acorn barnacle *Semibalanus balanoides*. *Mar Ecol Prog Ser* . 75:91–100.
- Bertness MD, Gaines SD, Stephens EG, Yund PO. 1992. Components of recruitment in populations of the acorn barnacle *Semibalanus balanoides* (Linnaeus). *J Exp Mar Biol Ecol*. 156(2):199–215.
- Botero CA, Weissing FJ, Wright J, Rubenstein DR. 2015. Evolutionary tipping points in the capacity to adapt to environmental change. *Proc Natl Acad Sci U S A*. 112(1):184–189.
- Bouckaert R, Heled J, Kuhner T, Vaughan T, Wu CH, Xie D, Suchard MA, Rambaut A, Drummond AJ. 2014. BEAST 2: a software platform for Bayesian evolutionary analysis. *PLoS Comput Biol*. 10(4):e1003537.
- Brochmann C, Gabrielsen TM, Nordal I, Landvik JY, Elven R. 2003. Glacial survival or tabula rasa? The history of North Atlantic biota revisited. *Taxon* 52(3):417.
- Brown AF, Kann LM, Rand DM. 2001. Gene flow versus local adaptation in the northern acorn barnacle, *Semibalanus balanoides*: insights from mitochondrial DNA variation. *Evolution* 55(10):1972–1979.
- Bucklin A, Steinke D, Blanco-Bercial L. 2011. DNA barcoding of marine metazoa. *Annu Rev Mar Sci*. 3(1):471–508.
- Buhl E, Bradlaugh A, Ogueta M, Chen KF, Stanewsky R, Hodge JJ. 2016. Quasimodo mediates daily and acute light effects on *Drosophila* clock neuron excitability. *Proc Natl Acad Sci U S A*. 113(47):13486–13491.
- Chernomor O, von Haeseler A, Minh BQ. 2016. Terrace aware data structure for phylogenomic inference from supermatrices. *Syst Biol*. 65(6):997–1008.
- Clark AG. 1997. Neutral behavior of shared polymorphism. *Proc Natl Acad Sci U S A*. 94(15):7730–7734.
- Drummond AJ, Nicholls GK, Rodrigo AG, Solomon W. 2002. Estimating mutation parameters, population history and genealogy simultaneously from temporally spaced sequence data. *Genetics* 161(3):1307–1320.
- 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*? *Mol Ecol*. 11(1):113–123.
- Eanes WF. 1999. Analysis of selection on enzyme polymorphisms. *Annu Rev Ecol Syst*. 30(1):301–326.
- 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(1):48.
- Edge P, Bafna V, Bansal V. 2017. HapCUT2: robust and accurate haplotype assembly for diverse sequencing technologies. *Genome Res*. 27(5):801–812.
- Faria R, Johannesson K, Butlin RK, Westram AM. 2019. Evolving inversions. *Trends Ecol Evol*. 34(3):239–248.
- Feder AF, Petrov DA, Bergland AO. 2012. LDx: estimation of linkage disequilibrium from high-throughput pooled resequencing data. *PLoS One* 7(11):e48588.
- Fijarczyk A, Babik W. 2015. Detecting balancing selection in genomes: limits and prospects. *Mol Ecol*. 24(14):3529–3545.
- Flight PA, O'Brien MA, Schmidt PS, Rand DM. 2012. Genetic structure and the North American postglacial expansion of the barnacle, *Semibalanus balanoides*. *J Hered*. 103(2):153–165.
- Flight PA, Rand DM. 2012. Genetic variation in the acorn barnacle from allozymes to population genomics. *Integr Comp Biol*. 52(3):418–429.
- Flowerdew MW. 1983. Electrophoretic investigation of populations of the cirripede *Balanus balanoides* (L) around the North Atlantic seaboard. *Crustaceana* 45(3):260–278.
- 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. *Mol Ecol*. 22(14):3766–3779.
- Gillespie J. 1973. Polymorphism in random environments. *Theor Popul Biol*. 4(2):193–195.
- Green RE, Krause J, Briggs AW, Maricic T, Stenzel U, Kircher M, Patterson N, Li H, Zhai W, Fritz MHY, et al. 2010. A draft sequence of the Neandertal genome. *Science* 328(5979):710–722.
- Hartl DL, Clark AG. 1997. Principles of population genetics. Sunderland (MA): Sinauer Associates.
- Hedrick PW. 2006. Genetic polymorphism in heterogeneous environments: the age of genomics. *Annu Rev Ecol Syst*. 37(1):67–93.
- Hedrick PW, Ginevan ME, Ewing EP. 1976. Genetic polymorphism in heterogeneous environments. *Annu Rev Ecol Syst*. 7(1):1–32.
- Helmuth B, Harley CD, Halpin PM, O'Donnell M, Hofmann GE, Blanchette CA. 2002. Climate change and latitudinal patterns of intertidal thermal stress. *Science* 298(5595):1015–1017.
- Herrera S, Watanabe H, Shank TM. 2015. Evolutionary and biogeographical patterns of barnacles from deep-sea hydrothermal vents. *Mol Ecol*. 24(3):673–689.
- Hodge JJ, Stanewsky R. 2008. Function of the Shaw potassium channel within the *Drosophila* circadian clock. *PLoS One*. 3(5):e2274.
- Johannesson K, Ring AK, Johannesson KB, Renborg E, Jonsson PR, Havenhand JN. 2018. Oceanographic barriers to gene flow promote genetic subdivision of the tunicate *Ciona intestinalis* in a North Sea archipelago. *Mar Biol*. 165(8):126.
- Jones SJ, Southward AJ, Wethey DS. 2012. Climate change and historical biogeography of the barnacle *Semibalanus balanoides*. *Glob Ecol Biogeogr*. 21(7):716–724.
- Kim D, Langmead B, Salzberg SL. 2015. HISAT: a fast spliced aligner with low memory requirements. *Nat Methods*. 12(4):357–360.
- Kimura M, Crow J. 1964. The number of alleles that can be maintained in a finite population. *Genetics* 49:725–738.
- King JL. 1967. Continuously distributed factors affecting fitness. *Genetics* 55(3):483–492.
- Kofler R, Orozco-terWengel P, De Maio N, Pandey RV, 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 RV, Schlotterer C. 2011. PoPoolation2: identifying differentiation between populations using sequencing of pooled DNA samples (Pool-Seq). *Bioinformatics* 27(24):3435–3436.
- Lee Y, Lee Y, Lee J, Bang S, Hyun S, Kang J, Hong ST, Bae E, Kaang BK, Kim J. 2005. Pyrexia is a new thermal transient receptor potential channel endowing tolerance to high temperatures in *Drosophila melanogaster*. *Nat Genet*. 37(3):305–310.
- Levene H. 1953. Genetic equilibrium when more than one ecological niche is available. *Am Nat*. 87(836):331–333.
- Lewontin RC, Hubby JL. 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* 54:595–609.
- 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, 1000 Genome Project Data Processing Subgroup. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25(16):2078–2079.
- Machado HE, Bergland AO, Taylor R, Tilk S, Behrman E, Dyer K, Fabian DK, Flatt T, González J, Karasov TL. 2019. Broad geographic sampling reveals predictable and pervasive seasonal adaptation in *Drosophila*. bioRxiv (unpublished data). Available from: <https://www.biorxiv.org/content/biorxiv/early/2019/10/11/337543.full.pdf> (last accessed December 12, 2019).
- Mackay TF, Richards S, Stone EA, Barbadilla A, Ayroles JF, Zhu D, Casillas S, Han Y, Magwire MM, Cridland JM, et al. 2012. The *Drosophila melanogaster* Genetic Reference Panel. *Nature* 482(7384):173–178.
- Maggis CA, Castilho R, Foltz D, Henzler C, Jolly MT, Kelly J, Olsen J, Perez KE, Stam W, Väinölä R, et al. 2008. Evaluating signatures of glacial refugia for North Atlantic benthic marine taxa. *Ecology* 89(Suppl 11):S108–S122.
- Malinsky M, Matschiner M, Svoldal H. 2020. Dsuite - fast D-statistics and related admixture evidence from VCF files. bioRxiv (unpublished data). Available from: <https://www.biorxiv.org/content/10.1101/634477v2> (last accessed May 1, 2020).
- Messer PW, Petrov DA. 2013. Population genomics of rapid adaptation by soft selective sweeps. *Trends Ecol Evol*. 28(11):659–669.

- Metz EC, Palumbi SR. 1996. Positive selection and sequence rearrangements generate extensive polymorphism in the gamete recognition protein bindin. *Mol Biol Evol.* 13(2):397–406.
- Milkman RD. 1967. Heterosis as a major cause of heterozygosity in nature. *Genetics* 55(3):493–495.
- Nunez JCB, Elyanow RG, Ferranti DA, Rand DM. 2018. Population genomics and biogeography of the northern acorn barnacle (*Semibalanus balanoides*) using pooled sequencing approaches. In: Oleksiak MF, Rajora OP, editors. Population genomics: marine organisms. Cham (Switzerland): Springer. p. 139–168.
- Nunez JCB, Flight PA, Neil KB, Rong S, Eriksson LA, Ferranti DA, Rosenblad MA, Blomberg A, Rand DM. 2020. Footprints of natural selection at the mannose-6-phosphate isomerase locus in barnacles. *Proc Natl Acad Sci USA.* 117(10):5376–5385.
- Patton H, Hubbard A, Andreassen K, Auriac A, Whitehouse PL, Stroeve AP, Shackleton C, Winsborrow M, Heyman J, Hall AM. 2017. Deglaciation of the Eurasian ice sheet complex. *Quat Sci Rev.* 169:148–172.
- Perez-Losada M, Harp M, Hoeg JT, Aчитuv Y, Jones D, Watanabe H, Crandall KA. 2008. The tempo and mode of barnacle evolution. *Mol Phylogenet Evol.* 46(1):328–346.
- Pickrell JK, Pritchard JK. 2012. Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genet.* 8(11):e1002967.
- Pineda J, Starczak V, Stueckle TA. 2006. Timing of successful settlement: demonstration of a recruitment window in the barnacle *Semibalanus balanoides*. *Mar Ecol Prog Ser.* 320:233–237.
- Pryszcz LP, Gabaldon T. 2016. Redundans: an assembly pipeline for highly heterozygous genomes. *Nucleic Acids Res.* 44(12):e113.
- Rand DM, Spaeth PS, Sackton TB, Schmidt PS. 2002. Ecological genetics of Mpi and Gpi polymorphisms in the acorn barnacle and the spatial scale of neutral and non-neutral variation. *Integr Comp Biol.* 42(4):825–836.
- Reich D, Thangaraj K, Patterson N, Price AL, Singh L. 2009. Reconstructing Indian population history. *Nature* 461(7263):489–494.
- Ruddiman WF, McIntyre A. 1981. The North Atlantic Ocean during the last deglaciation. *Palaeogeogr Palaeoclimatol Palaeoecol.* 35:145–214.
- Sanford E, Menge BA. 2001. Spatial and temporal variation in barnacle growth in a coastal upwelling system. *Mar Ecol Prog Ser.* 209:143–157.
- Schiffels S, Durbin R. 2014. Inferring human population size and separation history from multiple genome sequences. *Nat Genet.* 46(8):919–925.
- Schlotterer C, Tobler R, Kofler R, Nolte V. 2014. Sequencing pools of individuals-mining genome-wide polymorphism data without big funding. *Nat Rev Genet.* 15(11):749–763.
- Schmidt PS, Bertness MD, Rand DM. 2000. Environmental heterogeneity and balancing selection in the acorn barnacle *Semibalanus balanoides*. *Proc R Soc Lond B.* 267(1441):379–384.
- Schmidt PS, Rand DM. 1999. Intertidal microhabitat and selection at Mpi: interlocus contrasts in the northern acorn barnacle, *Semibalanus balanoides*. *Evolution* 53(1):135–146.
- Schmidt PS, Rand DM. 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 PS, Serrão EA, Pearson GA, Riginos C, Rawson PD, Hilbish TJ, Brawley SH, Trussell GC, Carrington E, Wetthey DS, et al. 2008. Ecological genetics in the North Atlantic: environmental gradients and adaptation at specific loci. *Ecology* 89(Suppl 11):S91–S107.
- Stanke M, Waack S. 2003. Gene prediction with a hidden Markov model and a new intron submodel. *Bioinformatics* 19(Suppl 2):ii215–ii225.
- Sundell K, Wrangé AL, Jonsson PR, Blomberg A. 2019. Osmoregulation in barnacles: an evolutionary perspective of potential mechanisms and future research directions. *Front Physiol.* 10:877.
- Sved JA, Reed TE, Bodmer WF. 1967. The number of balanced polymorphisms that can be maintained in a natural population. *Genetics* 55:469–481.
- Tracey WD, Wilson RI, Laurent G, Benzer S. 2003. painless, a *Drosophila* gene essential for nociception. *Cell* 113(2):261–273.
- Uyenoyama MK. 2005. Evolution under tight linkage to mating type. *New Phytol* 165(1):63–70.
- Väinölä R. 2003. Repeated trans-Arctic invasions in littoral bivalves: molecular zoogeography of the *Macoma balthica* complex. *Mar Biol.* 143:935–946.
- 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). *J Evol Biol.* 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*). *Mol Ecol.* 15(13):4193–4202.
- Vermeij GJ. 1991. Anatomy of an invasion: the trans-Arctic interchange. *Paleobiology* 17(3):281–307.
- Wares JP, Cunningham CW. 2001. Phylogeography and historical ecology of the North Atlantic intertidal. *Evolution* 55(12):2455–2469.
- Watt WB, Dean AM. 2000. Molecular-functional studies of adaptive genetic variation in prokaryotes and eukaryotes. *Annu Rev Genet.* 34(1):593–622.
- Williams G. 1966. Adaptation and natural selection: a critique of some current evolutionary thought. Princeton (NJ): Princeton University Press.
- Wittmann MJ, Bergland AO, Feldman MW, Schmidt PS, Petrov DA. 2017. Seasonally fluctuating selection can maintain polymorphism at many loci via segregation lift. *Proc Natl Acad Sci U S A.* 114(46):E9932–E9941.
- Wray GA. 2007. The evolutionary significance of cis-regulatory mutations. *Nat Rev Genet.* 8(3):206–216.
- Wu Q, Han TS, Chen X, Chen JF, Zou YP, Li ZW, Xu YC, Guo YL. 2017. Long-term balancing selection contributes to adaptation in *Arabidopsis* and its relatives. *Genome Biol.* 18(1):217.
- Xu SY, Cang CL, Liu XF, Peng YQ, Ye YZ, Zhao ZQ, Guo AK. 2006. Thermal nociception in adult *Drosophila*: behavioral characterization and the role of the painless gene. *Genes Brain Behav.* 5(8):602–613.
- Ye C, Hill CM, Wu S, Ruan J, Ma ZS. 2016. DBG2OLC: efficient assembly of large genomes using long erroneous reads of the third generation sequencing technologies. *Sci Rep.* 6(1):31900.