Footprints of natural selection at the mannose-6-phosphate isomerase locus in barnacles

Joaquin C. B. Nunez[a,1], Patrick A. Flight[a], Kimberly B. Neil[a], Stephen Rong[a,b], Leif A. Eriksson[c], David A. Ferranti[a], Magnus Alm Rosenblad[c], Anders Blomberg[c], and David M. Rand[a,1]

[a]Department of Ecology and Evolutionary Biology, Brown University, Providence, RI 02912; [b]Center for Computational Molecular Biology, Brown University, Providence, RI 02912; and [c]Department of Chemistry and Molecular Biology, Lundberg Laboratory, University of Gothenburg, 405 30 Göteborg, Sweden

The mannose-6-phosphate isomerase (*Mpi*) locus in *Semibalanus balanoides* has been studied as a candidate gene for balancing selection for more than two decades. Previous work has shown that *Mpi* allozyme genotypes (fast and slow) have different frequencies across Atlantic intertidal zones due to selection on postsettlement survival (i.e., allele zonation). We present the complete gene sequence of the *Mpi* locus and quantify nucleotide polymorphism in *S. balanoides*, as well as divergence to its sister taxon *Semibalanus cariosus*. We show that the slow allozyme contains a derived charge-altering amino acid polymorphism, and both allozyme classes correspond to two haplogroups with multiple internal haplotypes. The locus shows several footprints of balancing selection around the fast/slow site: an enrichment of positive Tajima's D for nonsynonymous mutations, an excess of polymorphism, and a spike in the levels of silent polymorphism relative to silent divergence, as well as a site frequency spectrum enriched for midfrequency mutations. We observe other departures from neutrality across the locus in both coding and noncoding regions. These include a nonsynonymous trans-species polymorphism and a recent mutation under selection within the fast haplogroup. The latter suggests ongoing allelic replacement of functionally relevant amino acid variants. Moreover, predicted models of *Mpi* protein structure provide insight into the functional significance of the putatively selected amino acid polymorphisms. While footprints of selection are widespread across the range of *S. balanoides*, our data show that intertidal zonation patterns are variable across both spatial and temporal scales. These data provide further evidence for heterogeneous selection on *Mpi*.

population genetics | ecological genomics | balancing selection | barnacles | *Semibalanus balanoides*

How natural populations evolve and adapt in highly heterogeneous environments is an old and important question that has motivated evolutionary biologist for decades. Enzyme polymorphisms have played a central role in studies addressing this question, for two reasons. First, allelic variation in enzymes (allozymes) was the first means of assessing levels of molecular polymorphisms in natural populations (1). Second, the existence of amino acid variation in proteins carrying out central metabolism raised the question of their functional significance and spurred the development of the neutral theory and the subsequent neutralist–selectionist debate. In the last 60 y, hundreds of allozyme polymorphisms have been identified in diverse species, and selection has been proposed repeatedly as the mechanism maintaining variation at many of these loci (2). While nucleotide variation and deviations from neutrality have been characterized for a large number of these loci, linking the evidence for selection to a particular biochemical or ecological process is rare (e.g., refs. 3 and 4).

Natural populations of the northern acorn barnacle (*Semibalanus balanoides*) provide an ideal system in which to investigate the biological significance of genetic polymorphisms in adaptation to heterogeneous environments. *S. balanoides* inhabits the rocky intertidal, a highly heterogeneous ecosystem in which multiple environmental stressors can be easily tracked (5, 6). This is due largely to its life history as a sessile, self-incompatible, simultaneous hermaphrodite that outcrosses with adjacent individuals in early-mid autumn. The fertilized eggs are brooded within the mantle cavity, and the release of the nauplii into the plankton occurs from January to March. Several nectarar instars precede the settling cypris larval stage, providing the opportunity for extensive pelagic larval dispersal by ocean currents (7). The cyprids settle out of the water column onto the rocky intertidal substrate from February to April, attach to the substrate, and metamorphose. Being sessile, genotype-specific selection based on mortality or growth

Significance

The rocky intertidal is a natural laboratory to study how natural selection acts on the genes and proteins responsible for organismal survival and reproduction. Alternative forms of enzymes that differ across the intertidal have been known for decades and have provided examples of selection, but the genetic basis of such enzyme variation is known in only a few cases. In this paper, we present molecular evidence of natural selection at the *Mpi* gene, a key enzyme in energy metabolism that alters survival of barnacles living across the stress gradient imposed by the intertidal. Our study demonstrates how natural selection can facilitate survival in highly heterogeneous environments through the maintenance of multiple molecular solutions to ecological stresses.

rates results from "decisions" to settle in distinct microhabitats (5, 8). But the wide dispersal of each brood virtually prevents local adaptation unless the "local" scale is on the order of 100s of kilometers. The strong selective gradients that exist within this dispersal range dictate that environmental heterogeneity is fine-grained relative to gene flow. *S. balanoides* recruits to a wide range of microhabitats across the intertidal zone (Fig. 1*A*). These microhabitats display high levels of heterogeneity due to many ecological stressors (9). For instance, individuals on the lower tidal microhabitat experience consistently lower thermal stress (Fig. 1*B*), relative to upper tidal microhabitats. Additionally, barnacles settling in the upper zone of sun-exposed shores (i.e., hot shores; facing southwest in the northern hemisphere) experience maximal levels of heat stress (5). Laboratory studies have shown that the maximum thermal range of *S. balanoides* lies between 35 °C and 37 °C with heat coma occurring above 35 °C, and mass mortality occurring above 37 °C (10). These temperatures repeatedly occur in sun-exposed upper tidal sites during the summer months (Fig. 1 *C* and *D*).

The gene for the glycolytic enzyme mannose-6-phosphate isomerase (*Mpi*) (EC 5.3.1.8) has been investigated in marine organisms as an example of the genetic basis for adaptation to heterogeneous environments (11–14). *Mpi* in *S. balanoides* displays two common allozyme alleles (fast/slow [F/S]; named after their electrophoretic mobility), which have been linked to fitness differentials across multiple spatial scales. For example, within rocky intertidal microhabitats in Maine (United States), the barnacles homozygous for the fast allele (FF genotypes) are favored in highly thermally stressed zones whereas barnacles homozygous for the slow allele (SS genotypes) are favored in the most thermally benign zones (i.e., the locus is "zonated") (5). These patterns have been validated in Maine with both transplant experiments (15) and laboratory experiments controlling for diet and heat exposure (8). Moreover, the allele frequencies at *Mpi* are panmictic at settlement, and zonation at the locus occurs only after full commitment to the rocky shore as a result of the differential survival of the FF, SS, and F/S genotypes and not habitat choice (6). Mannose derivatives are an important component of the barnacle's phytoplankton-based diet (16). Free mannose is phosphorylated by hexokinase(s), consuming ATP. If *Mpi*'s enzyme activity is reduced, an ATP deficit can result from low flux through glycolysis, leading to mortality (e.g., refs. 17 and 18). As such, Schmidt and Rand (5) proposed that heat-resistant fast alleles may protect barnacles from this ATP deficit in the face of extreme thermal stress.

Despite the patterns observed in Maine, *Mpi* has a complex biogeography, in many cases revealing very different intertidal zonation patterns. In Rhode Island, for instance, some habitats show no zonation while, in others, the FF genotype is more
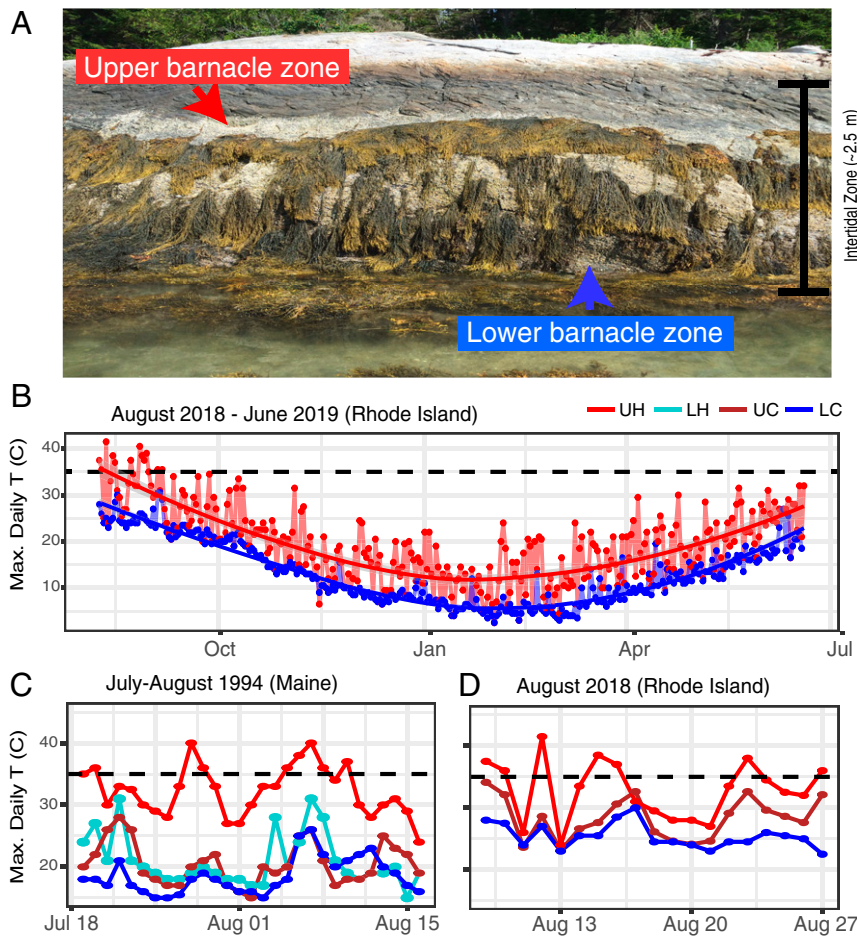
Fig. 1. (*A*) Diagram of the rocky shore in Maine, showing the intertidal range of *S. balanoides*. (*B*) Temperature plot showing the daily maximum temperature (Max. Daily T) in 2018 to 2019 Jamestown, RI, at two intertidal microhabitats: upper tidal zone in a hot shore (UH) and lower tidal zone in a cold shore (LC). "Hot shores" are shores which, because of their aspect, receive higher levels of sun exposure. A year-long trend line is shown for both. (*C*) Maximum daily temperature in Damariscotta in the summer of 1994 across four microhabitats: UH, upper tidal zone in a cold shore (UC), lower tidal zone in a hot shore (LH), and LC. Data obtained from Schmidt and Rand (5). (*D*) Maximum daily temperature in Rhode Island in the summer of 2018. A horizontal line marks heat coma temperatures in *B*–*D*.

common in the less stressed habitats (*SI Appendix*, Fig. S1*B*) (19). In Eastern Canada, selection in *Mpi* does not occur at the intertidal level but rather at estuary scales (20, 21). These studies provide evidence that the gene experiences heterogeneous selection across various spatial scales. Consequentially, the interaction of high planktonic gene flow, panmictic settlement (specifically for *Mpi*), and strong microhabitat selection presents an evolutionary problem for *Mpi* in barnacles: Beneficial alleles within one microhabitat may not become fixed because offspring are likely to settle into environments that drastically differ from their parents. As such, multiple authors have argued for a role of balancing selection as a multiple-niche polymorphism—sensu Levene (22)—as the mechanism for the maintenance of functional genetic variation at *Mpi* (8, 15, 23, 24). While ecological and physiological evidence lends support to the hypothesis, a full molecular description of the locus is lacking. In this paper, we present a detailed characterization of *Mpi* at the DNA level, with specific attention to footprints of natural selection in the gene. In particular, we ask if there is evidence of balancing selection working to maintain genetic variation in *Mpi*. A canonical signature of balancing selection is the presence of old, divergent alleles with excess variation segregating at intermediary frequencies (25). We also discuss the patterns of zonation in Maine and Rhode Island at the DNA level. We used multiple datasets [Pool-seq (26), RNA- and DNA-seq, cloning, and direct sequencing of amplicons] from populations across the North Atlantic, as well as intraspecific (North Pacific population) and interspecific (*Semibalanus cariosus*) outgroups (see *SI Appendix*, Table S1 for information on all datasets). To put our *Mpi* findings in a broader genomic context, we report patterns of genetic variation at mitochondrial DNAs (mtDNAs), and ~2,500 well-annotated barnacle genes (27), with special attention given to four classic allozyme loci: glucose-6-phosphate isomerase (*Gpi*), alcohol dehydrogenase (*Adh*), superoxide dismutase (*Sod*), and phosphoglucomutase (*Pgm*). Other than *Adh*, these loci were all used in an early study that first identified *Mpi* as an enzyme with unusual patterns of polymorphism in barnacles (28). Our analysis identifies the genetic mutations responsible for the fast and slow allozymes and their potential differences at the protein structure level, as well as evidence for a complex history of natural selection acting on both coding and noncoding parts of the gene. Additionally,

we provide evidence for the action of heterogeneous selection at the *Mpi* locus, both at spatial and temporal scales. Overall, these findings expand our understanding on the role of balancing selection in adaptations to highly variable environments.

Results

The Architecture of *Mpi* and of the Fast/Slow Allozymes. We identified the genomic structure of the *Mpi* locus using three approaches: two independent de novo DNA assemblies from whole genome data (*SI Appendix*, Text S1) and a dot blot "exon capture" technique followed by cloning, plus mapping of RNA-seq reads to the genomic scaffold (*SI Appendix*, Text S2). The organization of *Mpi*, which has six exons, is shown in *SI Appendix*, Fig. S2*C* (29). We used DNA- and RNA-seq from individual barnacles, and Pool-seq (*SI Appendix*, Table S1) to discover polymorphic sites across *Mpi* (Dataset S1) (30). Based on the sequence alignment of *Mpi* (Fig. 2*A*), we suspected that the charge-changing mutation Q390K (Table 1) is the molecular basis of the fast and slow allozymes. We validated this by genotyping 48 individuals (96 alleles) from Rhode Island using both allozyme electrophoresis and a restriction digest that distinguishes fast/slow variants by cutting the A392T site in the gene (which is linked to the Q390K site in the putative protein; see below). Ninety-one of 96 alleles were correctly diagnosed (i.e., the digest pattern matched the allozyme genotype; Fisher's exact test, $P < 0.001$). We sequenced amplicons for the five cases where the restriction site did not match the allozyme genotype (*Methods*). In all cases, the cut site was linked to the standard fast allele. This suggests that mischaracterized individuals may have occurred due to the presence of other charge-altering substitutions at low frequencies (e.g., C185R and G272E) (*SI Appendix*, Table S2).

Excess Variation in *Mpi*. We used Pool-seq from populations across the North Atlantic to discover segregating polymorphisms (Dataset S1), as well as survey levels of genetic variation (π) across the five allozyme loci (Table 2) (mtDNA shown in *SI Appendix*, Table S3). Our results show that the *Mpi* locus harbors the highest levels of nucleotide diversity (π = 2.8%) among the allozyme loci surveyed. Both neutral or adaptive processes could be invoked to explain these elevated levels of π in *Mpi*. To investigate this, we used a
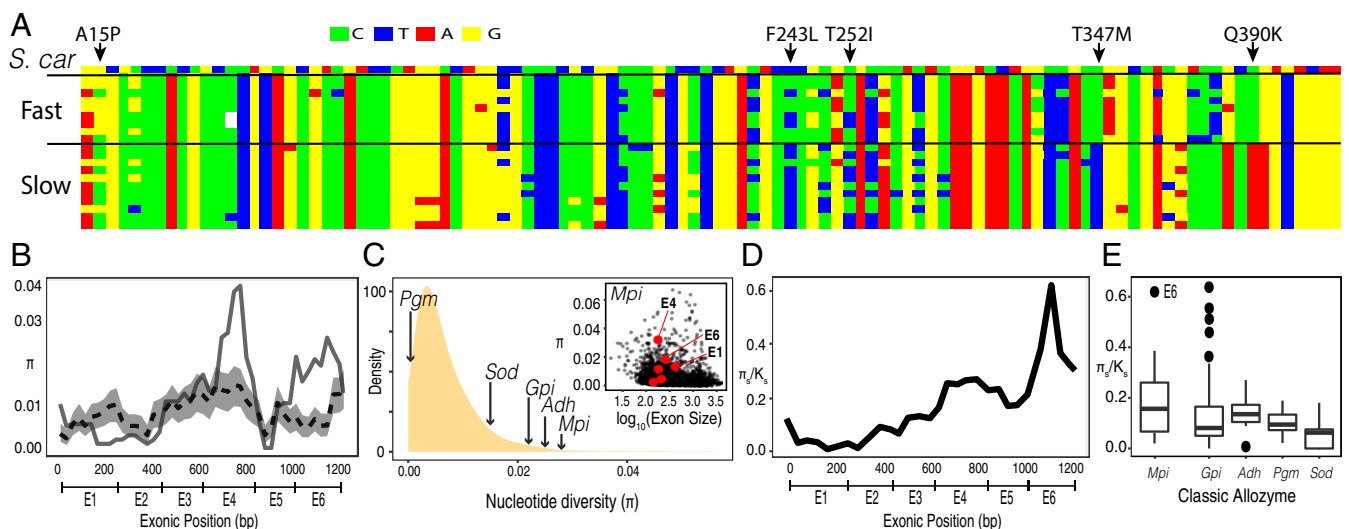
Fig. 2. (*A*) Multiple sequence alignment showing all polymorphic sites in *Mpi* across fast/slow haplogroups and the outgroup *S. cariosus*. The labeled mutations are discussed throughout the manuscript. (*B*) Window analysis of π across *Mpi* exons. The gray envelope is the 95% confidence interval of the neutral expectation. (*C*) Distribution of π for 2,555 genes across the barnacle genome (Sbal2.0). The π values for *Mpi* and other allozyme loci are shown. (*C, Inset*) Distribution of π values in the genome as a function of exon size (black dots). *Mpi* exons ("E") are shown in red. (*D*) Window analysis of π_s/K_s across *Mpi* exons. (*E*) Distributions of π_s/K_s across five allozyme loci. Values where estimated using a sliding window approach within each exon across genes.

Nunez et al.

Table 1. Nonsynonymous mutations in *Mpi*

Exon pos.	AA pos.	Ancestral/Derived	Notation (Type)
43	15	Ala^{gcc}/Pro^{ccc}	A15P (C)
553	185	Cys^{tgt}/Arg^{cgt}	C185R (E)
727	243	Phe^{ttc}/Leu^{ctc}	F243L (C)
755	252	Thr^{acc}/Iso^{atc}	T252I (R)
815	272	Gly^{gga}/Glu^{gaa}	G272E (E)
956	319	Ser^{agc}/Thr^{acc}	S319T (C)
1040/1041	347	Thr$^{acg/aca}$/Met^{atg}	T347M (R)
1168	390	Glu^{cag}/Lys^{aag}	Q390K* (E)
1174	392	Ala^{gcg}/Thr^{acg}	A392T (R)

Exon and protein positions are listed, as well as their products (ancestral and derived states) and notations. The fast/slow mutation is indicated with an asterisk. C, conservative; E, charge changing; R, radical.

McDonald–Kreitman test (MKT) to characterize the potential excess of amino acid polymorphism in our allozyme loci. MKT was used on 10 phased haplotypes (five individuals) from Maine and two phased haplotypes (one individual) from *S. cariosus* (Table 2). After phylogenetic correction, MKT values were not significant for all loci except for *Pgm* ($P < 0.05$; this significant MKT signature in *Pgm* will be investigated in future studies). With the exception of *Sod*, all allozyme loci show hints of directional selection in their neutrality indexes (NIs) (31). Notably, *Mpi* is the only gene displaying NI > 1, suggesting an excess of amino acid polymorphism. In addition, we estimated π in windows across *Mpi*'s exons, using all sequences from Maine and Rhode Island. With all sequences pooled, π shows three regions with excess variation (only exon regions were used for these analyses). Two major π peaks, one in exon 6 (where Q390K resides) and another in exon 4, as well as an elevated area of π in exon 1. All peaks are noticeably higher than surrounding areas, and the two major peaks are well above neutral expectation using divergence from the sister species (*S. cariosus*) as a reference (Fig. 2*B*; dashed line and 95% confidence interval of simulated data, which assumes that genetic variation has segregated neutrally since the divergence of the two species) (*Methods*). To understand how *Mpi* compares to the rest of the genome, we estimated π across 2,555 annotated genes across the barnacle genome (Sbal2.0) (27, 32, 33). Values for π were calculated using a genome-wide panel of genetic variation from Maine published by Flight and Rand ($\pi_{mean} = 0.65\%$, $\pi_{sd} = 0.59\%$) (33, 34). This analysis reveals that *Mpi* is among the most polymorphic genes in the genome (Fig. 2*C*). We also estimated the ratio of silent intraspecies diversity (π_s) to silent interspecies divergence (K_s) using a window approach within all exons of *Mpi*. Under a scenario of balancing selection, mutations around the balanced polymorphism are expected to show much older ages relative to neutrality and will produce a spike of π_s/K_s. In *Mpi*, we observed a spike of π_s/K_s in exon 6 (Fig. 2*D*) centered at T347M (a mutation linked to Q390K; see below) (*SI Appendix*, Fig. S3). We also observed elevated π_s/K_s at exon 1 around A15P, a nonsynonymous trans-species polymorphism (*SI Appendix*, Fig. S4). These spikes are not observed in *Adh*, *Pgm*, or *Sod* (Fig. 2*E*). Notably, we observed π_s/K_s outliers in *Gpi*, a locus that has also been hypothesized to be under balancing selection (19, 21).

Population Level Patterns. To investigate how genetic variation in *Mpi* is distributed in space, we constructed a site frequency spectrum (SFS) using samples from Maine and Rhode Island. The SFS analysis shows an excess of midfrequency mutations in both Maine and Rhode Island relative to both the neutral expectation and other allozyme loci (*SI Appendix*, Fig. S5). In addition, we conducted linkage disequilibrium (LD; r^2) analyses using both our DNA sequences as well as our Pool-seq data. Notably, estimates of LD in *Mpi* (Fig. 3*A* and Dataset S2) (35) are different between

populations. In the DNA sequence data, half decay distance occurs in Maine after 101 base pairs (bp), and after 291 bp in Rhode Island (Fig. 3*A*, solid lines). LD patterns estimated using our Pool-seq data (*SI Appendix*, Fig. S6) (36) recapitulate our findings with DNA sequences ($\rho = 0.66$, $P < 0.001$). In these data, half LD decay occurs after 168 bp in Maine, and after 416 bp in Rhode Island (Fig. 3*A*, dashed lines). These analyses showed that the fast/slow mutation (Q390K) has high levels of LD with two nonsynonymous mutations, T347M and A392T. In our DNA data, these mutations always show $r^2 = 1$. LD estimates from our Pool-seq data suggest that the association between these three mutations, while high, is population-specific. In Maine, the maximum likelihood estimator of r^2 in the pool (r^2_{MLE}) is 0.99 between Q390K and A392T. In Rhode Island, r^2_{MLE} between Q390K and A392T is 0.84. For mutations Q390K and T347M, r^2_{MLE} is ~0.75 in both Maine and Rhode Island. These high levels of LD among nonsynonymous mutations around the fast/slow site are additional footprints of balancing selection (25, 37).

We also estimated the Tajima's D statistic (D) across all loci. The D statistic leverages deviation from the SFS expected under neutrality and can be interpreted as signatures of selection or nonequilibrium demography. A case of balancing selection is expected to produce an inflation of positive D values in polymorphisms closely linked to the balanced site. We first estimated Tajima's D across all 2,555 Sbal2.0 genes to generate an empirical distribution for the D statistic. Mean genome-wide D values for our Maine reference panel are mildly skewed toward negative values ($D_{mean} = -0.29$, $D_{sd} = 0.55$) (Fig. 3*B*). This is consistent with the role of purifying selection on protein coding loci. The classic allozyme loci (as whole genes) are not outliers relative to genome-wide patterns, *Pgm* being the only exception (this is consistent with the MKT results). Remarkably, *Mpi*'s exon 6 shows significantly positive levels of Tajima's D in Maine (Fig. 3*C* and *SI Appendix*, Fig. S7*A*). This signal is driven by the nonsynonymous mutations Q390K and T347M. These two mutations consistently display positive values of D in populations across the western, middle, and eastern North Atlantic Ocean (Fig. 3*D* and *SI Appendix*, Fig. S7*C*). We validated these findings using custom primers (*SI Appendix*, Table S4) and Sanger sequencing around Q390K in Maine, Rhode Island, and United Kingdom populations (*SI Appendix*, Tables S5 and S6). Notably, these patterns of variation in *Mpi*'s exon 6 are absent in the North Pacific population (Fig. 3*D*).

Genetic Variation in Exon 6. We conducted a phylogenetic analysis using the haplotype sequences of *Mpi*'s exon 6 (*SI Appendix*, Fig. S8*A*), as well as an exon-wide haplotype network (*SI Appendix*, Fig. S8*B*). These analyses show the existence of two haplogroups corresponding to the fast and slow allozymes and suggest that the slow mutation is the derived state. The Pool-seq data confirm that the derived states of Q390K and T347M are only observed in the North Atlantic. The derived state of A392T exists at low

Table 2. Population genetic metrics of classic allozyme loci in Maine

Locus	π	$F_{ns(JC)}$	$F_{s(JC)}$	P_{ns}	P_s	MKT	NI
Mpi	0.028	15.16	41.25	20	30	0.153	1.82
Adh	0.025	33.53	14.15	6	24	0.343	0.59
Gpi	0.022	20.22	44.34	5	26	0.114	0.42
Pgm	0.008	13.18	26.30	1	14	**0.004**	0.14
Sod	0.017	9.14	12.86	10	14	0.994	1.00

Nucleotide diversity (π); number of fixed nonsynonymous sites ($F_{ns(JC)}$) and number of fixed synonymous sites ($F_{s(JC)}$), both corrected with Jukes–Cantor; and number of polymorphic nonsynonymous sites (P_{ns}) and number of polymorphic synonymous sites (P_s). MKT, P values of the McDonald–Kreitman test; NI, neutrality index. P values: Bold font represents statistical significance.
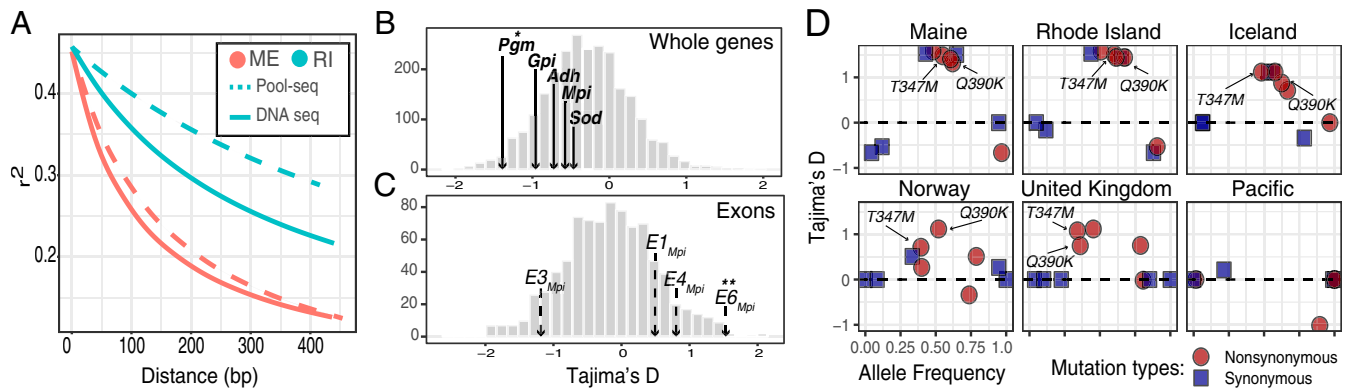
Fig. 3. (*A*) Linkage disequilibrium decay at *Mpi* in Maine (ME) and Rhode Island (RI) estimated using our Pool-seq and DNA sequences datasets. (*B*) Distribution of Tajima's D for 2,555 genes across the barnacle genome (Sbal2.0). Whole gene Tajima's D values are shown for the classic allozymes (data from Maine). The vertical solid lines indicate the values of Tajima's D for the classic allozyme loci. (*C*) Distribution of Tajima's D for the constituent exons of the 2,555 Sbal2.0 genes (exons with less than five segregating sites were excluded). The values of *Mpi* exon 1, 3, 4, and 6 are shown (data from Maine). The vertical dashed lines indicate the values of Tajima's D for exons of *Mpi*. In both *B* and *C*, the * symbol represents a significant deviation from the genome-wide distribution (*$P < 0.05$, **$P < 0.01$). (*D*) Tajima's D for synonymous and nonsynonymous variants in *Mpi*'s exon 6 across North Atlantic and North Pacific populations. The horizontal dashed lines represent a Tajima's D value of zero (a proxy for neutrality).

frequency in the North Pacific and thus is shared across ocean basins ([Dataset S1](#)). Genetic diversity within each haplogroup is similar across the locus although haplogroups differ on an exon-by-exon basis (*SI Appendix*, Fig. S9). To gain insights of allele ages, we estimated extended haplotype homozygosity (EHH) (38) and a bifurcation analysis focusing on Q390K. EHH is the probability that two randomly chosen haplotypes are identical by descent. Backgrounds showing high levels of EHH, and few bifurcations in the bifurcation analysis, over long stretches of DNA correspond to young haplotypes (39). For our data, both the fast and slow haplogroups appear to be old, with EHH rapidly decaying around the locus (*SI Appendix*, Fig. S10). This is reflected in the bifurcation analysis (Fig. 4 *A* and *B*). Notably, the fast haplogroup

shows marginally higher levels of EHH relative to the slow haplogroup. This is not expected since the fast haplogroup is the ancestral state.

Genetic Variation in Exon 4. Our window-based analysis in Fig. 2*B* shows an excess of variation also occurring in other exons. To determine if these signatures were the product of nonneutral processes, we conducted two scans for selection (d_N/d_S [the ratio of the number of nonsynonymous substitutions per nonsynonymous site to the number of synonymous substitutions per synonymous site] and $\beta^{(2)}$) (40, 41) on the locus (*SI Appendix*, Text S3). Our analyses suggest that the signal in exon 4 is driven by two nonsynonymous mutations: F243L and T252I. Both these
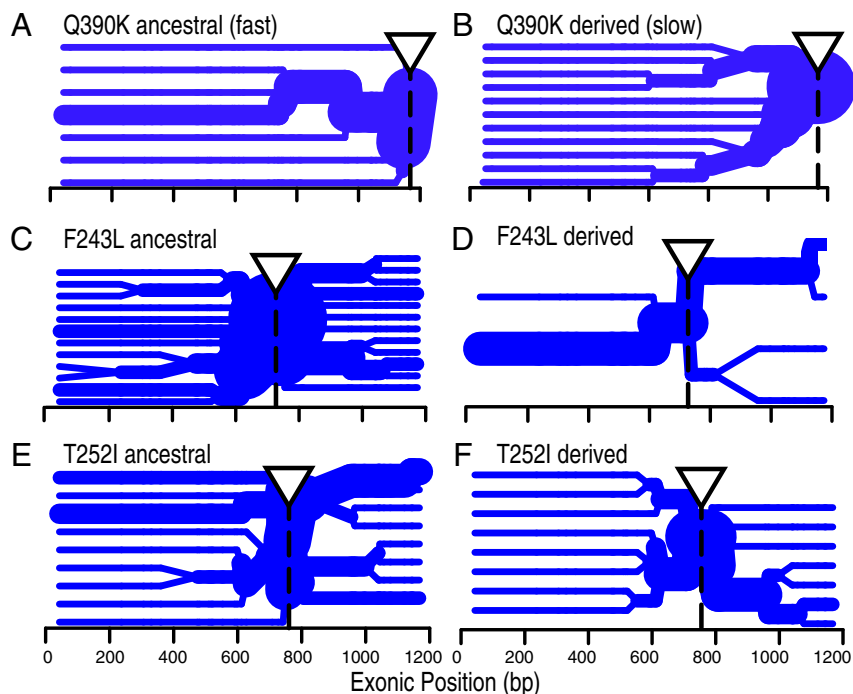
Fig. 4. Bifurcation analysis for Q390K ancestral (*A*) and derived (*B*), F243L ancestral (*C*) and derived (*D*), and T252I ancestral (*E*) and derived (*F*). The dashed lines and arrow indicate the position of the focal mutation.

mutations are polymorphic only in the North Atlantic, but their LD relative to Q390K is population-dependent (*SI Appendix*, Table S7A and Dataset S2). F243L's ancestral state harbors high levels of variation and low EHH (Fig. 4C). Its derived state, however, appears young due to its high levels of EHH (Fig. 4D), has low π (*SI Appendix*, Text S4), and occurs primarily within the fast haplogroup (*SI Appendix*, Table S7). On the other hand, mutation T252I shows low levels of EHH for both states (Fig. 4 *E* and *F*).

Genetic Variation in Exon 1 and the 5′ UTR. Another mutation showing signatures of selection in the d_N/d_S scan is the nonsynonymous trans-species polymorphism A15P (exon 1). This variant is polymorphic across the entire North Atlantic, North Pacific, and the outgroup *S. cariosus* (and it is seen across all of our data sets) (*SI Appendix*, Fig. S4 *A and B*). Notably, the region surrounding A15P displays elevated levels of π (*SI Appendix*, Fig. S4 *C and D*) and positive Tajima's D (*SI Appendix*, Fig. S4E). LD between A15P and mutations in exons 4 and 6 is habitat-specific (*SI Appendix*, Table S7 and Dataset S2). This analysis also reveals an association with a noncoding multiallelic mutation in the 5′ UTR region of *Mpi*, 51 bp away from the start codon (−51T > C;A, position 4936 in the scaffold). In Maine, the r^2_{MLE} between A15P and −51T > C;A is 0.37. In Rhode Island r^2_{MLE} is 0.82. This 5′ UTR mutation shows very elevated levels of positive Tajima's D, and it is only polymorphic in the North Atlantic. These footprints suggest that the regulatory region of *Mpi* may be the target of linked balancing selection.

Intertidal Zonation at *Mpi*. We used Pool-seq to investigate allele frequency zonation at *Mpi* (Dataset S1). We collected 37 adult barnacles from two replicate microhabitats experiencing high and low stress in Maine and Rhode Island intertidal shores (Fig. 1A and *SI Appendix*, Fig. S11). Estimates of genetic differentiation are shown in *SI Appendix*, Fig. S12. We tested for zonation using a Cochran–Mantel–Haenszel (CMH) test. CMH is an association test used in population genetics to identify constant and consistent changes in allele frequencies across independent biological replicates (42–44). For this analysis, we tested for independence between "stress treatment" and allelic ratio "outcome." We conducted the test separately for Maine and Rhode Island. To account for the statistical associations due to LD, we aggregated individual mutation-level *P* values into gene-level *P* values (45). Only the *Mpi* gene in Maine shows significant gene-level zonation ($P < 0.001$) (*SI Appendix*, Table S8). Surprisingly, the zonation in Maine at electromorph mutations Q390K, and associated variants A392T, and T347M, does not occur in the same direction as the allozyme results reported by Schmidt and Rand in 1999 (*SI Appendix*, Fig. S1 *C and D*) (5). This is because the derived states for these mutations are more abundant in the most stressful microhabitats. We also observed no signatures of zonation in Rhode Island, which is consistent with earlier allozyme results (19).

Structural Analyses. In order to explore the potential functional consequences of the aforementioned mutations, we conducted a series of protein-folding analyses, which are fully discussed in *SI Appendix*, Text S5. Overall, the predicted protein structures for the fast and slow haplogroups are highly similar, but our analyses predict different correlative patterns of movement among residues between the fast and slow proteins, especially relative to T347M. Correlation among atomic motions is essential in determining key functions, such as substrate–enzyme binding affinity (46). This is consistent with the findings of Rand et al. (19), which show different *Mpi* allozyme activities across varying temperatures and substrate concentrations. In addition, our analyses also reveal that the fast and slow haplogroups produce loops of different orientations around amino acids 320–325, which, in the fast

haplogroup, is bent down, away from the active site region, as compared to the slow (Fig. 5 *A* and *B*). These different physical arrangements result in different residue interactions of T347M between haplogroups (Fig. 5B). Our results also show that the allozyme mutations (Q390K, T347M, and A392T) lie on the surface of the protein (Movie S1 and *SI Appendix*, Text S5). As such, they may impact *Mpi*'s capacity to engage in physical interactions with other proteins. Notably, we observe that the trans-species polymorphism, A15P, is located near the vicinity of the active site. However, it does not appear to induce any major effects on the structure of the model. Lastly, mutation F243L is predicted to alter solvent accessibility (47), likely affecting the folding thermodynamics and hydrophobicity of the protein (46). We note that, because our protein models are based on homology, our mutation effect estimates are limited, and the model is unable to accurately estimate additive or epistatic interactions which may be occurring in the wild (48). More research is required to determine potential impacts to the catalytic function of the protein.

Discussion

Footprints of Selection at *Mpi*. Enzyme polymorphisms of central metabolism that are differentiated between ecological habitats provide compelling systems to study natural selection in the wild. The *Mpi* locus in the acorn barnacle, and in many other marine organisms, shows this pattern, with alternative genotypes associated with distinct intertidal habitats (5, 21, 23). These studies have provided support for balancing selection maintaining alternative alleles. However, the strength and even direction of selection on *Mpi* and other allozymes varies geographically. While the ecological aspects of selection on *Mpi* are variable in space, the patterns of molecular evidence seen at *Mpi* appear consistent with the action of balancing selection originally proposed for the locus (6, 8, 15). The locus is enriched for positive nonsynonymous Tajima's D values across the entire North Atlantic habitat of the species (Fig. 3D; a pattern not observed in the other genes surveyed; *SI Appendix*, Fig. S7), the site frequency spectrum (SFS) shows an excess of midfrequency polymorphism, and our neutral divergence analyses show an excess of polymorphism at exons 1, 4, and 6. While some of these tests have, by themselves, reduced power to detect selection in the face of complex demography (49), the combined evidence suggests that selection on standing genetic variation operates at broad spatial scales to maintain genetic variation in *Mpi*.

Selection in the Fast/Slow Allozymes. Our analyses show that multiple mutations throughout the gene are candidates of selection (*SI Appendix*, Text S3). Among those targets, Q390K and linked nonsynonymous substitutions A392T and T347M appear to be the targets of balancing selection characterized by Schmidt and Rand (6) in the Maine intertidal. Linkage between these variants is high and shared between two populations separated by Cape Cod (a known phylogeographic break for the species) (7). Moreover, these
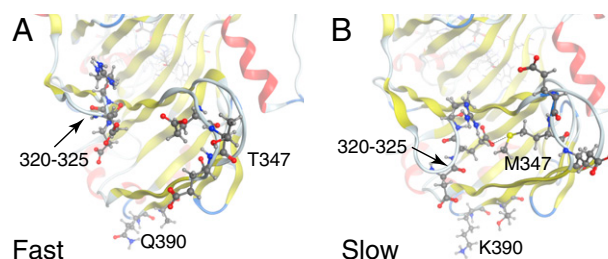
Fig. 5. Structural analyses based on homology models for the predicted proteins encoded by the fast (*A*) and slow (*B*) haplogroups. For both models, residues in the 320–325 loop, 347, and 390 are shown in the ball-and-stick representation.

concordant patterns of LD show that the fast/slow electromorphs are defined by Q390K. Notably, T347M shows an excess of intra- to interspecific silent divergence (*SI Appendix*, Fig. S3) and has the largest impact in the three-dimensional (3D) structure of *Mpi*. One hypothesis is that T347M is the balanced polymorphism, and all linked cosegregating mutations are either neutral variants, have interactive effects, or represent a case of sheltered load (37): a phenomenon where balancing selection shelters deleterious mutations surrounding the balanced mutation, leading to the accumulation of nonsynonymous divergence between alleles.

The Time Scales of the Fast/Slow Variants. In this paper, we have shown that *Mpi*'s Q390K polymorphism is neither trans-species nor shared in the ancestral North Pacific population. This poses the question: Could the derived state for Q390K be a putatively neutral polymorphism segregating privately in the North Atlantic? As a species, *S. balanoides* colonized the North Atlantic during the transarctic interchange 1 to 3 million years ago (mya) (50). Assuming that the derived mutation arose sometime after the interchange, whether or not enough time has passed for neutral coalescence depends on estimates of effective population size (N_e) ($\theta = 4N_e\mu$). Depending on mutation rate (μ), N_e in barnacles may vary between 250,000 and 1,500,000 (*SI Appendix*, Text S6). Lower estimates of N_e would allow for mutations arising after the interchange to reach expected time to coalescence ($4N_e$ generations) before the present day. Upper estimates of N_e, on the other hand, do not allow enough time. Despite this uncertainty, the functional evidence available for Q390K supports rejecting neutrality.

Complex Landscape of Selection at *Mpi*. While Q390K and associated variants show footprints of balancing selection, there are three pieces of evidence that suggest the action of other selection events occurring in the gene. First, the presence of a nonsynonymous trans-species polymorphism (A15P) is not expected under neutrality and may constitute a case of ancient balancing selection (25). *S. cariosus* does not occur in the North Atlantic, and there is no evidence for modern gene flow between the species across ocean basins (51). As such, an alternative hypothesis of recent introgression is unlikely. This would suggest that trans-species polymorphisms have been preserved in populations since the species split ~8 mya (52). A15P shows patterns of linkage disequilibrium with −51T > C;A (5′ UTR multiallelic mutation), and the region containing both these mutations harbors elevated levels of Tajima's D (*SI Appendix*, Fig. S4E). This signal indicates the action of linked selection on noncoding regulatory DNAs in *Mpi* (e.g., 5′ UTR in CCR5) (53). Also, the putative maintenance of multiallelic genotypes at −51T > C;A is a canonical footprint of balancing selection (e.g., MHC) (54, 55). Overall, the relative low levels of LD between exons 1 and 6 suggest that Q390K and A15P are independent cases of balancing selection. Further testing of this hypothesis will be the target of future research. Second, the pattern of variation at T252I indicates that the variant may be experiencing its own kind of selection. However, it is currently unknown if this is related to intertidal stress or a different ecological stressor. Third, under long-term balancing selection, both fast and slow haplogroups should display high levels of genetic variation, but the fast (ancestral) haplogroup should harbor marginally more variation. Our analysis, however, shows that genetic variation within *Mpi*'s fast haplogroup has been reduced by the presence of a young haplotype associated with a derived mutation in exon 4 (F243L). Notably, variation at this site shows footprints similar to those of a selective sweep (e.g., low genetic variation and high EHH) (Fig. 4 *C* and *D*). Moreover, our protein structure analysis suggests that F243L alters solvent accessibility (*SI Appendix*, Text S5), potentially affecting folding thermodynamics and hydrophobic properties of the resulting protein. The fact that such a young

mutation is seen throughout the North Atlantic provides a biogeographic frame of reference. For *S. balanoides*, the North Atlantic current has been proposed as a barrier for modern east-to-west gene flow across the basin (27). This suggests that F243L may have originated prior to the last glacial maximum (∼20,000 ya) when these populations shared a most recent common ancestor (7). Based on this, we hypothesize that F243L may be driving an allelic replacement event within the fast haplogroup of *Mpi* (25). This hypothesis is similar to that described for the superoxide dismutase locus (*Sod*) in *Drosophila* (56), but restricted primarily to the fast haplogroup of *Mpi* in barnacles. In general, allele replacement events may serve as mechanisms to "fine tune" balanced polymorphisms. We note that this assumes that the derived state of F243L is a beneficial mutation. Now that this amino acid variant has been characterized, ecological zonation and transplant experiments could be used to test this hypothesis. Such studies would need to control for haplotype combinations that occur in nature.

Why Is Zonation at *Mpi* Discordant across Space and Time? Data from intertidal zonation analyses in Maine, Rhode Island, and Canada reveal significant differences in the patterns of zonation for the *Mpi* fast and slow alleles in these locations (5, 19, 21, 57). These contradictory results for *Mpi* zonation patterns preclude a simple ecological model for the gene. Our sequence data shed light on the variable ecological and geographic zonation at *Mpi* from two lines of evidence. First is the distinction between "fast" or "slow" alleles at the level of allozyme electromorph vs. DNA sequence. The DNA sequences show that variation at *Mpi* is anything but binary, and the fast and slow electromorph mutations segregate within haplogroups harboring high levels of genetic variation and complex haplotype structures capable of affecting the properties of resulting proteins (Fig. 5 and *SI Appendix*, Text S5). For instance, the presence of mutations, other than just the electromorph mutation (Q390K), showing deviations from neutrality (e.g., polymorphisms in exons 1, 4, and the 5′ UTR) showcases the potential for a complex landscape of selection where additive or epistatic effects may produce phenotypes with different fitness values across space or time. It is clear that further studies are required to connect the various haplotypes segregating at *Mpi* to the ecology of *Semibalanus*.

Second are the temporal and spatial scales at which selection operates. In this paper, we have shown various footprints of selection acting on the *Mpi* locus that appear consistent across the entire North Atlantic range of the species. However, zonation of the fast/slow haplogroups is highly variable across space. It is known that thermal stress in intertidal habitats is not explained by a simple north-to-south latitude gradient, but rather by local variation in tidal range and timing of low tides. Consequentially, intertidal shores displaying large variation in tidal range may experience more thermal stress than shores located at lower latitudes (58). This "mosaic" model of thermal stress provides a potential explanation for why *Mpi* zonation varies from site to site. In addition, our evidence suggests that zonation is a habitat-specific process emerging from the interaction of many ecological variables, as exemplified by the controlled experiments of Flight et al. (24), which showed that mortality from thermal stress in barnacles is affected by osmotic stress.

Our data also show that patterns of allele zonation in Maine are inconsistent relative to earlier allozyme studies (5, 6, 19). Such variability in zonation patterns suggests that *Mpi* also experiences temporal variation. While puzzling, the simplest explanation is that zonation is variable in time and in space as the allozyme data were collected 20 y ago. Moreover, zonation per se is only one part of the selection acting on *Mpi* and may not represent the most important aspect, given the sampling of intertidal extremes. Notably, barnacles that survive the extreme of the high intertidal typically have relatively little reproductive

contribution to the larval pool of the following generation (9). This begs the question: What processes could be driving temporal changes in *Mpi*? Recent studies have shown that adaptations to seasonality are a general feature of *Drosophila* populations (59, 60). In this case, mutations that are advantageous in winter may be disadvantageous in summer, and vice versa. In barnacles, however, the temporal dimensions of *Mpi* selection are less intuitive. Recruitment in *S. balanoides* is a well-studied process consisting of tightly regulated yearly windows (61) occurring in synchrony with ecological cues (e.g., winter temperatures and plankton bloom dynamics) during the later winter/early spring. Notably, these ecological cues show high degrees of year-to-year variation and can have drastic effects on the survival and recruitment success of barnacle cohorts (62). The data reported here, and the implications from the structural modeling, point to possible explanations for the temporally variable zonation patterns. We consider three scenarios. First, the physiological properties of *Mpi*'s fast/slow haplogroups could be consistent over time, but year-to-year ecological variation during recruitment season may produce environments in which the fast/slow haplogroups have different fitnesses. A second possibility is that year-to-year variation in settlement ecology doesn't affect the overall physiological demands of intertidal microhabitats, but additive or epistatic interactions between *Mpi* mutations in exons 1, 4, and 6 and the 5′ UTR produce phenotypes with different fitness. As these mutations have LD patterns which are both low and population-specific, variation in allele frequency over time could result in the observed discordance among studies. Finally, a third option is the potential interaction between genetic variation and environmental variation. While *Mpi* is an important metabolic gene, its protein product is known to be pleiotropic, with functions in cell signaling pathways other than glycolysis (e.g., ref. 63). As such, if selection is imposed on a side function of *Mpi* during early life stages (e.g., during recruitment), this could result in fitness trade-offs leading to antagonistic pleiotropy (64) between life stages. Cell signaling is a plausible target for antagonistic pleiotropy since we have already shown that the fast/slow mutation results in amino acid changes on the surface of *Mpi*'s protein (Fig. 5) and may affect its physical interactions with other proteins. Moreover, thermal selection in *Mpi* only manifests during summer months (Fig. 1) (6). Thus, only the genotypes surviving settlement will contribute to the standing genetic variation participating in spatially varying selection. As such, the magnitude and direction of zonation at *Mpi* may emerge from the antagonism (or synergism) of these temporally and spatially varying selective pressures across different life stages. Overall, there is little doubt that *Mpi* is experiencing a complex landscape of selection, but a model of selection focused on a single site (e.g., Q390K) and one single stress gradient (e.g., temperature) is too simplistic to account for the molecular evolution of the locus.

What About Demography? The evidence presented in this paper highlights complex footprints of selection at *Mpi* and captures demographic differences between Maine and Rhode Island populations. For instance, the high levels of LD in Rhode Island may indicate unequal rates of gene conversion or different histories of local admixture relative to Maine (65). In light of this complexity, we must consider whether the signatures of nucleotide variation in *Mpi* can result from demographic processes operating in the North Atlantic basin. As previously highlighted, *S. balanoides* has a complex demographic history defined by the transarctic interchange and the last glaciation cycle, which have resulted in multiple habitat contractions, followed by expansions (7). As such, some patterns that we have attributed to balancing selection (e.g., the positive values of Tajima's D) may be due to postglacial nonequilibrium demography. However, demographic processes such as these are expected to generate genome-wide signatures.

In our datasets, however, we do not observe any systematic inflation of positive D across the genome (Fig. 3B).

In Conclusion. This paper presents a detailed characterization of nucleotide variation at *Mpi*, as well as various lines of evidence for the action of nonneutral processes acting on the locus. We show that, while the footprints of selection at *Mpi* are seen throughout the North Atlantic, zonation, per se, varies across time and space. This spatial–temporal variation in *Mpi* zonation is the primary evidence that the gene experiences heterogeneous selection. As such, we propose that genetic variation at the locus is maintained by a complex landscape of natural selection in which long term balancing selection cooccurs with recent instances of allelic replacement. Although more work is necessary to fully resolve *Mpi*'s connection to fitness, our results provide a robust example regarding the roles and spatial scales of adaptive responses in a natural population evolving in a highly heterogeneous environment.

Methods

Temperature Measures in the Intertidal. Temperatures were measured across intertidal microhabitats using high resolution Thermochron sensors (part no. DS1921G; Maxim Integrated). During the summer of 2018, sensors were installed in the upper tidal zone-hot shore (UH), upper tidal zone-cold shore (UC), and lower tidal zone-cold shore (LC) microhabitats across two sites in Jamestown, RI. Thermochron sensors were 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 (datasets and additional details are available as supplementary material) (Datasets S3 and S4).

Obtaining Allozyme Sequences. DNA sequences of *Mpi* and other allozymes (*Adh*, *Sod*, *Gpi*, and *Pgm*) were extracted bioinformatically from the draft assembly of the *S. balanoides* genome (version 3.1) (*SI Appendix*, Text S1). Extraction of the genes was done using a protein-to-genome alignment using references from UniProt (*SI Appendix*, Table S9). We validated each sequence by mapping RNA-seq reads and obtaining intron–exon boundary evidence. Mapping was conducted on hard-masked sequences. Genomic features were predicted with Augustus (66) using a *Drosophila melanogaster* model and RNA evidence. The assembly of the *Mpi* sequence required additional bioinformatic steps outlined in *SI Appendix*, Text S2. We further validated the *Mpi* sequence using a dot blot "exon capture" technique and cloning.

***Mpi* Barnacle Datasets.** Five types of datasets were used for the present data analysis: 1) pooled sequencing, 2) whole DNA-sequencing, 3) whole RNA-sequencing, 4) haploid sequencing via cloning, and 5) amplicon sequencing. These datasets are thoroughly described in *SI Appendix*, Table S1. Pool-seq samples came from Damariscotta (Maine, United States), Jamestown (Rhode Island, United States), British Columbia (Canada), Reykjavik (Iceland), Wales (United Kingdom), and Norddal (Norway). We sequenced single individuals from the same sites in Maine, Rhode Island, and British Columbia. RNA-seq was done on four individuals from Maine. In addition, we sequenced DNA from a single *S. cariosus* individual. All pools, single individuals, and the four RNA samples were sequenced in their respective single lanes of an Illumina machine by GENEWIZ LLC. Pool-seq, whole DNA-sequences, and whole RNA-sequences contain information for *Mpi*, all other classical allozyme loci, and mtDNA. The haploid sequences of *Mpi* came from Rhode Island and were sequenced via cloning (Topo-TA cloning kit; Invitrogen) and Sanger. DNA/RNA was extracted using Qiagen DNeasy/RNeasy kits.

Population Genetic Analyses. Genotype validation of the fast/slow mutations was done using allozyme electrophoresis and restriction digest using the AhdI restriction enzyme (New England Biolabs). Validation of the mischaracterized individuals in the restriction digest was done using Sanger amplification and the exon 6 primers shown in *SI Appendix*, Table S4. For all datasets, we filtered singleton loci and enforced a 5% minimum allele frequency. Various tools were used for data analysis; their versions and corresponding citations are available in *SI Appendix*, Table S10. An extended discussion of our analysis pipeline is available in *SI Appendix*, Text S7. Population genetic simulations were done in msms (67). We conducted simulations using a window approach (size = 100, step = 25, sample size = 20, 1,000 iterations)

sensu Kreitman and Hudson (68). Gene-level *P* values were obtained using Pegasus (45). EHH was estimated in R using rehh (38). Protein analyses are described in *SI Appendix*, Text S5 and Table S10. Mutations were polarized using the sequence of *S. cariosus*. Whole genome analyses were done with 2,555 complete and partially complete genes extracted from the currently published *S. balanoides* genome (Sbal2.0; National Center for Biotechnology Information [NCBI] ID PHFM00000000). Genes used in the analyses are shown in Dataset S5. The mtDNA reference for *S. balanoides* is also in NCBI (27) (NCBI RefSeq ID NC_039849.1). The data used to estimate genome-wide Tajima's D and π are available from Flight and Rand (34) (SRR10011826). Gene features for Sbal2.0 are shown in Dataset S6. Gene features for allozyme loci are shown in Dataset S7.

1. R. C. Lewontin, J. L. Hubby, 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 (1966).
2. W. F. Eanes, Analysis of selection on enzyme polymorphisms. *Annu. Rev. Ecol. Syst.* **30**, 301–326 (1999).
3. C. W. Wheat, W. B. Watt, D. D. Pollock, P. M. Schulte, From DNA to fitness differences: Sequences and structures of adaptive variants of Colias phosphoglucose isomerase (PGI). *Mol. Biol. Evol.* **23**, 499–512 (2006).
4. R. S. Burton, M. W. Feldman, Physiological effects of an allozyme polymorphism: Glutamate-pyruvate transaminase and response to hyperosmotic stress in the copepod Tigriopus californicus. *Biochem. Genet.* **21**, 239–251 (1983).
5. P. S. Schmidt, D. M. Rand, Intertidal microhabitat and selection at Mpi: Interlocus contrasts in the northern acorn barnacle, Semibalanus balanoides. *Evolution* **53**, 135–146 (1999).
6. P. S. Schmidt, D. M. Rand, Adaptive maintenance of genetic polymorphism in an intertidal barnacle: Habitat- and life-stage-specific survivorship of Mpi genotypes. *Evolution* **55**, 1336–1344 (2001).
7. P. A. Flight, M. A. O'Brien, P. S. Schmidt, D. M. Rand, Genetic structure and the North American postglacial expansion of the barnacle, Semibalanus balanoides. *J. Hered.* **103**, 153–165 (2012).
8. P. S. Schmidt, The effects of diet and physiological stress on the evolutionary dynamics of an enzyme polymorphism. *Proc. Biol. Sci.* **268**, 9–14 (2001).
9. M. D. Bertness, S. D. Gaines, D. Bermudez, E. Sanford, Extreme spatial variation in the growth and reproductive output of the acorn barnacle semibalanus balanoides. *Mar. Ecol. Prog. Ser.* **75**, 91–100 (1991).
10. A. J. Southward, Note on the temperature tolerances of some intertidal animals in relation to environmental temperatures and geographical distribution. *J. Mar. Biol. Assoc. U. K.* **37**, 49–66 (1958).
11. D. Hedgecock, Is gene flow from pelagic larval dispersal important in the adaptation and evolution of marine invertebrates? *Bull. Mar. Sci.* **39**, 550–564 (1986).
12. J. H. McDonald, Repeated geographic variation at three enzyme loci in the amphipod platorchestia platensis. *Evolution* **41**, 438–441 (1987).
13. J. H. McDonald, Contrasting amounts of geographical variation as evidence for direct selection: The Mpi and Pgm loci in eight crustacean species. *Heredity* **67**, 215–219 (1991).
14. H. R. Siegismund, Genetic studies of Gammarus. IV. Selection component analysis of the Gpi and the Mpi loci in Gammarus oceanicus. *Hereditas* **102**, 241–250 (1985).
15. P. S. Schmidt, M. D. Bertness, D. M. Rand, Environmental heterogeneity and balancing selection in the acorn barnacle Semibalanus balanoides. *Proc. Biol. Sci.* **267**, 379–384 (2000).
16. D. R. Kreger, "Cell walls" in *Physiology and Biochemistry of Algae*, R. A. Lewin, Ed. (Academic Press, New York, 1962), pp. 315–332.
17. H. Arnold, U. Seitz, G. W. Löhr, Die Hexokinase und die Mannosetoxizität der Biene. *Hoppe Seylers Z. Physiol. Chem.* **355**, 266–272 (1974).
18. D. Hernández, M. De la Fuente, Mannose toxicity in Ehrlich ascites tumor cells. *Biochem. Cell Biol.* **67**, 311–314 (1989).
19. D. M. Rand, P. S. Spaeth, T. B. Sackton, P. S. Schmidt, Ecological genetics of Mpi and Gpi polymorphisms in the acorn barnacle and the spatial scale of neutral and nonneutral variation. *Integr. Comp. Biol.* **42**, 825–836 (2002).
20. D. Veliz, P. Duchesne, E. Bourget, L. Bernatchez, Genetic evidence for kin aggregation in the intertidal acorn barnacle (Semibalanus balanoides). *Mol. Ecol.* **15**, 4193–4202 (2006).
21. D. Véliz, E. Bourget, L. Bernatchez, Regional variation in the spatial scale of selection at MPI* and GPI* in the acorn barnacle Semibalanus balanoides (Crustacea). *J. Evol. Biol.* **17**, 953–966 (2004).
22. H. Levene, Genetic equilibrium when more than one ecological niche is available. *Am. Nat.* **87**, 331–333 (1953).
23. E. R. Holm, E. Bourget, Selection and population genetic-structure of the barnacle semibalanus-balanoides in the Northwest Atlantic and Gulf of St-Lawrence. *Mar. Ecol. Prog. Ser.* **113**, 247–256 (1994).
24. P. A. Flight, S. D. Schoepfer, D. M. Rand, Physiological stress and the fitness effects of Mpi genotypes in the acorn barnacle Semibalanus balanoides. *Mar. Ecol. Prog. Ser.* **404**, 139–149 (2010).
25. A. Fijarczyk, W. Babik, Detecting balancing selection in genomes: Limits and prospects. *Mol. Ecol.* **24**, 3529–3545 (2015).
26. C. Schlötterer, R. Tobler, R. Kofler, V. Nolte, Sequencing pools of individuals - mining genome-wide polymorphism data without big funding. *Nat. Rev. Genet.* **15**, 749–763 (2014).
27. J. C. B. Nunez, R. G. Elyanow, D. A. Ferranti, D. M. Rand, "Population genomics and biogeography of the northern acorn barnacle (semibalanus Balanoides) using pooled sequencing approaches" in *Population Genomics: Marine Organisms, Population Genomics*, M. F. Oleksiak, O. P. Rajora, Eds. (Springer, Cham, 2018), pp. 139–168.
28. M. W. Flowerdew, Electrophoretic investigation of populations of the Cirripede balanus-balanoides (L) around the North-Atlantic seaboard. *Crustaceana* **45**, 260–278 (1983).
29. J. C. B. Nunez et al., Scaffold containing the sequence of Mpi (MK953005). NCBI GenBank. https://www.ncbi.nlm.nih.gov/nuccore?term=MK953005. Deposited 20 May 2019.
30. J. C. B. Nunez et al., Data for Footprints of natural selection at the mannose-6-phosphate isomerase locus in barnacles. NCBI Sequence Read Archive. https://trace.ncbi.nlm.nih.gov/Traces/sra/?study=SRP199303. Deposited 23 May 2019.
31. D. M. Rand, L. M. Kann, Excess amino acid polymorphism in mitochondrial DNA: Contrasts among genes from Drosophila, mice, and humans. *Mol. Biol. Evol.* **13**, 735–748 (1996).
32. J. C. B. Nunez, R. G. Elyanow, D. A. Ferranti, D. M. Rand, Semibalanus balanoides isolate SBAL2_MERISW, whole genome shotgun sequencing project (PHFM00000000.1). NCBI GenBank. https://www.ncbi.nlm.nih.gov/nuccore/PHFM00000000.1. Deposited 2 November 2018.
33. P. A. Flight, D. M. Rand, Genome-wide pooled sequencing data from Maine (SRR10011826). NCBI Sequence Read Archive. https://trace.ncbi.nlm.nih.gov/Traces/sra/?run=SRR10011826. Deposited 21 August 2019.
34. P. A. Flight, D. M. Rand, Genetic variation in the acorn barnacle from allozymes to population genomics. *Integr. Comp. Biol.* **52**, 418–429 (2012).
35. W. G. Hill, B. S. Weir, Variances and covariances of squared linkage disequilibria in finite populations. *Theor. Popul. Biol.* **33**, 54–78 (1988).
36. A. F. Feder, D. A. Petrov, A. O. Bergland, LDx: Estimation of linkage disequilibrium from high-throughput pooled resequencing data. *PLoS One* **7**, e48588 (2012).
37. M. K. Uyenoyama, Evolution under tight linkage to mating type. *New Phytol.* **165**, 63–70 (2005).
38. P. C. Sabeti et al., Detecting recent positive selection in the human genome from haplotype structure. *Nature* **419**, 832–837 (2002).
39. B. F. Voight, S. Kudaravalli, X. Wen, J. K. Pritchard, A map of recent positive selection in the human genome. *PLoS Biol.* **4**, e72 (2006).
40. K. M. Siewert, B. F. Voight, Detecting long-term balancing selection using allele frequency correlation. *Mol. Biol. Evol.* **34**, 2996–3005 (2017).
41. K. M. Siewert, B. F. Voight, BetaScan2: Standardized statistics to detect balancing selection utilizing substitution data. *Biology and Evolution*, evaa013, https://doi.org/10.1093/gbe/evaa01 (2020).
42. R. S. Brennan, A. D. Garrett, K. E. Huber, H. Hargarten, M. H. Pespeni, Rare genetic variation and balanced polymorphisms are important for survival in global change conditions. *Proc. Biol. Sci.* **286**, 20190943 (2019).
43. C. Schlötterer, R. Kofler, E. Versace, R. Tobler, S. U. Franssen, Combining experimental evolution with next-generation sequencing: A powerful tool to study adaptation from standing genetic variation. *Heredity* **114**, 431–440 (2015).
44. P. Orozco-terWengel et al., Adaptation of Drosophila to a novel laboratory environment reveals temporally heterogeneous trajectories of selected alleles. *Mol. Ecol.* **21**, 4931–4941 (2012).
45. P. Nakka, B. J. Raphael, S. Ramachandran, Gene and network analysis of common variants reveals novel associations in multiple complex diseases. *Genetics* **204**, 783–798 (2016).
46. A. M. Lesk, C. Chothia, Solvent accessibility, protein surfaces, and protein folding. *Biophys. J.* **32**, 35–47 (1980).
47. S. Sunyaev, V. Ramensky, P. Bork, Towards a structural basis of human nonsynonymous single nucleotide polymorphisms. *Trends Genet.* **16**, 198–200 (2000).

48. M. A. DePristo, D. M. Weinreich, D. L. Hartl, Missense meanderings in sequence space: A biophysical view of protein evolution. *Nat. Rev. Genet.* **6**, 678–687 (2005).

49. A. Ramírez-Soriano, S. E. Ramos-Onsins, J. Rozas, F. Calafell, A. Navarro, Statistical power analysis of neutrality tests under demographic expansions, contractions and bottlenecks with recombination. *Genetics* **179**, 555–567 (2008).

50. G. J. Vermeij, Anatomy of an invasion: The trans-Arctic interchange. *Paleobiology* **17**, 281–307 (1991).

51. A. F. Brown, L. M. Kann, D. M. Rand, Gene flow versus local adaptation in the northern acorn barnacle, Semibalanus balanoides: Insights from mitochondrial DNA variation. *Evolution* **55**, 1972–1979 (2001).

52. M. Pérez-Losada et al., The tempo and mode of barnacle evolution. *Mol. Phylogenet. Evol.* **46**, 328–346 (2008).

53. M. J. Bamshad et al., A strong signature of balancing selection in the 5′ cis-regulatory region of CCR5. *Proc. Natl. Acad. Sci. U.S.A.* **99**, 10539–10544 (2002).

54. P. W. Hedrick, Pathogen resistance and genetic variation at MHC loci. *Evolution* **56**, 1902–1908 (2002).

55. C. Wills, Maintenance of multiallelic polymorphism at the MHC region. *Immunol. Rev.* **124**, 165–220 (1991).

56. R. R. Hudson, A. G. Sáez, F. J. Ayala, DNA variation at the Sod locus of Drosophila melanogaster: An unfolding story of natural selection. *Proc. Natl. Acad. Sci. U.S.A.* **94**, 7725–7729 (1997).

57. F. Dufresne, E. Bourget, L. Bernatchez, 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**, 113–123 (2002).

58. B. Helmuth et al., Climate change and latitudinal patterns of intertidal thermal stress. *Science* **298**, 1015–1017 (2002).

59. A. O. Bergland, R. Tobler, J. González, P. Schmidt, D. Petrov, Secondary contact and local adaptation contribute to genome-wide patterns of clinal variation in Drosophila melanogaster. *Mol. Ecol.* **25**, 1157–1174 (2016).

60. H. Machado et al., Broad geographic sampling reveals predictable and pervasive seasonal adaptation in Drosophila. bioRxiv:10.1101/337543 (3 June 2018).

61. J. Pineda, V. Starczak, T. A. Stueckle, Timing of successful settlement: Demonstration of a recruitment window in the barnacle Semibalanus balanoides. *Mar. Ecol. Prog. Ser.* **320**, 233–237 (2006).

62. D. S. Wethey et al., Response of intertidal populations to climate: Effects of extreme events versus long term change. *J. Exp. Mar. Biol. Ecol.* **400**, 132–144 (2011).

63. A. Cazet, J. Charest, D. C. Bennett, C. L. Sambrooks, J. N. Contessa, Mannose phosphate isomerase regulates fibroblast growth factor receptor family signaling and glioma radiosensitivity. *PLoS One* **9**, e110345 (2014).

64. M. R. Rose, Antagonistic pleiotropy, dominance, and genetic variation. *Heredity* **48**, 63–78 (1982).

65. L. Frisse et al., Gene conversion and different population histories may explain the contrast between polymorphism and linkage disequilibrium levels. *Am. J. Hum. Genet.* **69**, 831–843 (2001).

66. M. Stanke, S. Waack, Gene prediction with a hidden Markov model and a new intron submodel. *Bioinformatics* **19** (suppl. 2), ii215–ii225 (2003).

67. G. Ewing, J. Hermisson, MSMS: A coalescent simulation program including recombination, demographic structure and selection at a single locus. *Bioinformatics* **26**, 2064–2065 (2010).

68. M. Kreitman, R. R. Hudson, Inferring the evolutionary histories of the Adh and Adh-dup loci in Drosophila melanogaster from patterns of polymorphism and divergence. *Genetics* **127**, 565–582 (1991).

EVOLUTION