MOLECULAR ECOLOGY

Molecular Ecology (2013) **22**, 3580–3597 doi: 10.1111/mec.12337

Signals of selection in outlier loci in a widely dispersing species across an environmental mosaic

MELISSA H. PESPENI† and STEPHEN R. PALUMBI

Department of Biology, Hopkins Marine Station, Stanford University, Oceanview Blvd, Pacific Grove, CA 93950, USA

Abstract

Local adaptation reflects a balance between natural selection and gene flow and is classically thought to require the retention of locally adapted alleles. However, organisms with high dispersal potential across a spatially or temporally heterogeneous landscape pose an interesting challenge to this view requiring local selection every generation or when environmental conditions change to generate adaptation in adults. Here, we test for geographical and sequence-based signals of selection in five putatively adaptive and two putatively neutral genes identified in a previous genome scan of the highly dispersing purple sea urchin, *Strongylocentrotus purpuratus*. Comparing six populations spanning the species' wide latitudinal range from Canada to Baja California, Mexico, we find positive tests for selection in the putative adaptive genes and not in the putative neutral genes. Specifically, we find an excess of low-frequency and nonsynonymous polymorphisms in two transcription factors and a transporter protein, and an excess of common amino acid polymorphisms in the two transcription factors, suggestive of spatially balancing selection. We test for a genetic correlation with temperature, a dominant environmental variable in this coastal ecosystem. We find mild clines and a stronger association of genetic variation with temperature than latitude in four of the five putative adaptive loci and a signal of local adaptation in the Southern California Bight. Overall, patterns of genetic variation match predictions based on spatially or temporally balancing selection in a heterogeneous landscape and illustrate the value of geographical and coalescent tests on candidate loci identified in a genome-wide scan for selection.

Keywords: balancing selection, candidate gene, gene flow, genome scan, natural selection, purple sea urchin, *Strongylocentrotus purpuratus*, transcription factor evolution

Received 19 September 2011; revision received 28 February 2013; accepted 5 March 2013

Introduction

Adaptive evolution reflects a balance between the strength of natural selection and the rate of gene flow (Endler 1977). In this light, organisms with high dispersal potential that live across a spatially heterogeneous landscape, such as many marine invertebrates, fishes or plants, pose an interesting challenge for local adaptation (Grosberg & Cunningham 2001; Taylor & Hellberg 2003; Petit & Hampe 2006; Nielsen *et al.* 2009; Hart & Marko 2010; Sanford & Kelly 2011). Such organisms may be sessile or have small home ranges as adults, yet may broadcast eggs, pollen or larvae that disperse to habitats distinct from their parents. When this mix of propagules settles from near or far locations, survival to adulthood may vary depending on the fitness of an individual's alleles in that particular environment (Marshall *et al.* 2010). After this selective filter, the progeny of successful adults may then disperse to near or far locations where they experience a different set of environmental conditions (Levene 1953; Hedrick 2006). In this scenario, there is a mismatch between larval genotypes and local environment that is reduced by local selection in each generation (Marshall *et al.* 2010; Sanford & Kelly 2011; Fraser *et al.* 2011).

Correspondence: Melissa H. Pespeni, Fax: 812-855-6705; E-mail: mpespeni@indiana.edu

†Present address: Department of Biology, Indiana University, Bloomington, IN,47405, USA

However, even in species with high dispersal potential, actual dispersal can be limited by larval behaviour, oceanography or other mechanisms (Weersing & Toonen 2009) allowing the build-up of locally adapted alleles over time (Wares 2002; Sotka *et al.* 2004; Carvalho 2005; Dionne *et al.* 2007; Burford 2009). The combination of local and distant sources for recruits and a heterogeneous landscape of natural selection is expected to lead to a wide variety of different gene frequency patterns at selected and neutral loci.

Identifying the genetic signatures of local adaptation has been facilitated in recent years through the growing accessibility of genomic approaches to ecological model organisms (De Wit *et al.* 2012). Outlier tests for selection have been developed to test whether some loci show genetic differentiation, usually measured as F_{ST} from SNP or AFLP markers, that is significantly higher than expected based on a presumably neutral distribution (Beaumont & Balding 2004; Antao *et al.* 2008; Foll & Gaggiotti 2008). The availability of data for 10 000s of expressed loci across the transcriptome provides the opportunity to test functional groups of loci for larger-than-expected differentiation. In this case, loci might not be significant outliers individually but groups of genes in the same metabolic or functional pathway may show high differentiation (Pespeni *et al.* 2012).

Outlier tests identify candidate loci that have higher than expected genetic differentiation, but they do not by themselves explore patterns and causes of selection (Bierne 2010; Bierne *et al.* 2011). There is a strong need to test candidate loci that are revealed in outlier tests for patterns of selection. But these loci are chosen for their high F_{ST} levels, so high F_{ST} in follow-up studies, by itself, does not constitute a strong test. Luckily, nucleotide and amino acid sequence data can be tested to a much greater extent for signals of natural selection using coalescence and ratios of amino acid-changing polymorphisms (Nei & Gojobori 1986; Tajima 1989; Fu & Li 1993; Fay & Wu 2000; Fay *et al.* 2001, 2002). These sequenced-based signals of selection can be used to complement data on frequency-based tests (Pogson 2001; Bierne 2010; Kemppainen *et al.* 2010).

Here, we tested for geographical and sequence-based signals of selection in five putatively adaptive genes identified as outliers in a previous genome scan of 10 purple sea urchins from each of two distant, wild populations at 12 431 polymorphisms (Boiler Bay, OR and San Diego, CA; Pespeni *et al.* 2010). In the present study, we more than doubled sample sizes from each of these populations and sampled an additional four populations to span most of the species range, from Bamfield, Canada to Baja California, Mexico (Table 1 and Fig. 1). The west coast of North America experiences a latitudinally associated range of environmental changes including

Table 1 Sampling locations, abbreviations and sample sizes. Boiler Bay, OR and San Diego, CA, were the sites compared in the previous genome scan. The 24 individuals sampled from Boiler Bay, OR and San Diego, CA, for the present study did not include the 10 previously studied individuals for each location

Location	Location code	Lat/Long	No. individuals sampled
Bamfield, Canada	BM	48.8N/125.2W	24
Boiler Bay, Oregon	BB	44.8N/124.1W	24
Patrick's Point, California	PP	41.1N/124.2W	27
Monterey, California	MY	36.6N/121.9W	25
San Diego, California	SD	32.8N/117.3W	24
Baja California, Mexico	Baja		(41)
La Bufadora	BJ	31.7N/116.7W	14
Punta Baja	PB	29.9N/115.8W	17
Punta San Carlos	SC	29.6N/115.5W	10
Total			165

Fig. 1 Map of sampling locations. Bamfield, Canada (BM), Boiler Bay, Oregon (BB), Patrick's Point, California (PP), Monterey, California (MY), San Diego, California (SD) and three sites pooled for samples from Baja California, Mexico (Baja). See Table 1 for latitude/longitude details for each location. BB and SD were the sites compared in the previous genome scan (Pespeni *et al.* 2010).

day length, spawning season, temperature, primary productivity, settlement and pH due to upwelling (Helmuth *et al.* 2006). Populations near San Diego, however, are affected by an unusually broad continental shelf that reduces the environmental influence of cold, deep water and breaks the latitudinal association with environment. As a result, there is a local reversal of temperature patterns with southern populations in Baja being cooler than San Diego (see Materials and Methods for more details). Because temperature has been shown to have a broad influence on patterns of allele frequency and fitness in marine invertebrates and fish (Place & Powers 1979; Schmidt & Rand 2001), we use mean sea surface temperature to test for allele differentiation associated with latitude vs. environment.

We compare five putatively selected outlier loci with two loci expected to show neutral patterns. Our goals were to (i) test outlier loci for signals of natural selection and relationships between genetic variation and environment, (ii) investigate broad-scale biogeographic patterns of alleles across latitude and (iii) determine whether patterns of genetic variation along the species range are consistent with clines or with balancing selection in response to spatial or temporal heterogeneity.

Materials and methods

Study system

The purple sea urchin species presents an ideal system to test for patterns of clinal vs. mosaic signatures of selection. Purple sea urchins live in rocky inter- and subtidal habitats from Alaska to Baja California, Mexico (Schultz 2005). They are broadcast spawners, releasing eggs and sperm into the water column where fertilization occurs. Developing larvae may then spend several weeks to months (Strathmann 1978) in the water column potentially travelling 10s to 100s of kilometres before recruiting to a suitable habitat and metamorphosing into juvenile urchins. Accordingly, previous genetic studies using putatively neutral markers have found no population structure at large geographical scales (Palumbi & Wilson 1990; Edmands *et al.* 1996; Olivares-Banuelos *et al.* 2008). All alleles among 12 431 polymorphisms studied by Pespeni *et al.* (2010) were shared across a 1700-km expanse of the species' range (Pespeni *et al.* 2010). In addition, purple sea urchins are highly fecund (Strathmann 1978) and have dramatically large population sizes (Ebert & Russell 1988). These natural history characteristics make the species a particularly robust system in which to study natural selection. Large population sizes and high fecundity maximize the effects of selection and minimize the effects of random genetic drift (Hartl & Clark 1997).

Despite the species distribution across a broad gradient in sea surface temperature, there exists a considerable amount of heterogeneity at smaller spatial scales along the species range (Helmuth *et al.* 2002, 2006; Lester *et al.* 2007; Ebert 2010; Sanford & Kelly 2011). For example, Helmuth *et al.* (2006) characterized the coastline as a geographical mosaic of thermal 'hot spots' and 'cold spots' where intertidal temperatures are hotter or colder than expected based on latitude. These patterns were revealed by the estimated body temperatures of mussels, a midintertidal sessile species. 'Hot spots' in the north were largely driven by summer low tides occurring during the daytime, while 'cold spots' in the south experienced low tides at night (Helmuth *et al.* 2006). In addition, there is not a perfect correlation between temperature and latitude: populations in the Southern California Bight region experience consistently warmer temperatures than those that live in the Baja upwelling system further south (Walsh *et al.* 1974; Huyer 1983; Zaytsev *et al.* 2003).

Located in the Southern California Bight, San Diego is furthest from the along-shore current transport systems, the California and Davidson currents (Mann & Lazier 2006). The distance from the current systems, the geography of the coastline and the relatively shallow bathymetry of the region result in warmer waters entrained in the Bight, high-temperature anomalies and unique environmental conditions, such as distinct primary productivity seasons (Bakun & Parrish 1982; Smith & Eppley 1982; Mann & Lazier 2006). Particle tracking data have shown water movements in the Southern California Bight are highly complex with numerous localized eddies resulting in a high degree of connectivity among sites in the region (Mitarai *et al.* 2009).

Based on these data, we expected that San Diego urchins would show the greatest differences in allele frequency due to reduced gene flow, local adaptation or both. By contrast, we expected that populations of urchins along the Baja Peninsula would be more genetically similar to Central and Northern California urchins due to the close nearshore proximity of the California Current along the Baja Peninsula and the resulting prevalence of strong upwelling (Zaytsev *et al.* 2003), similar to the upwelling conditions off Northern California. These conditions could connect populations through larval transport and provide similar selective habitats with respect to temperature and nutrients despite a large geographical distance.

Sample collection and DNA extraction

A total of 165 purple sea urchins were sampled (either spine base muscle or tube foot tissue) from six regions

spanning the distribution of purple sea urchins along the west coast of North America from Bamfield, Canada to Baja California, Mexico (Table 1, Fig. 1). We attempted to collect purple sea urchins from Sitka, Alaska, to have both range edges represented in this study; however, with the assistance of several offices of Alaska's Department of Fish and Game, we could not find enough urchins for a sufficient sample size. Localities with historic abundances were depauperate of purple sea urchins, possibly due to the successful resurgence of sea otters, a dominant predator of the purple sea urchin, in southeast Alaska. Sampling locations were each separated by a geographical distance of 400–600 km, except for the three locations in Baja California that were an average of 100 km apart. Spine samples were collected in the field (Patrick's Point, Monterey, Punta Baja and Punta San Carlos) or collected from live animals in aquaria that had been recently collected from a nearby location in the wild (Bamfield) and preserved in 70% ethanol for later DNA extraction. For Boiler Bay and San Diego samples, live urchins were collected from the wild and shipped to the Hopkins Marine Station where they were maintained in outdoor aquaria (see also Pespeni *et al.* 2013a). For these individuals, fresh tube foot tissue was collected from live individuals for direct DNA extraction. For all samples, total genomic DNA was extracted from ~10 mg tissue using the NucleoSpin extraction kit following manufacturer's protocol (Macherey-Nagel, Bethlehem, PA, USA).

Locus selection, primer design and sequencing

We selected three loci as putatively adaptive candidates from the extreme tail (top 1%, $F_{ST} > 0.18$) of the F_{ST} distribution of polymorphisms in coding regions from the previous study that compared 10 individuals from each of two sites, Boiler Bay, OR and San Diego, CA (Pespeni *et al.* 2010). These loci were identified as outliers based on results from two outlier detection programs, LOSITAN (Antao *et al.* 2008) and BayeScan (Foll & Gaggiotti 2008; Pespeni *et al.* 2010). We focused on coding polymorphisms because all previously identified outliers were in coding regions (Pespeni *et al.* 2010) and for ease of sequencing in this highly polymorphic species. These outlier loci were transcription factor 25, cubilin and a transporter protein (Table 2). To include potential false negative outliers, we randomly selected an additional three loci from among the top 1% highest F_{ST} coding polymorphisms that had gene annotations and comparable heterozygosity levels as the outlier loci. Thirty-eight loci met these criteria and could have been interesting candidates for this study. The randomly selected additional three candidate loci included another transcription factor, a gaba-b receptor and a WD repeat domain-containing protein (Table 2). Two nonoutlier loci with comparable heterozygosity were selected, the metabolic enzyme pyruvate kinase and the structural protein collagen (Table 2). The pyruvate kinase locus was selected because it had been arbitrarily chosen in the previous study to validate the RSTA methodology (Pespeni *et al.* 2010). For the purposes of the present study, the F_{ST} value of this locus comparing Boiler Bay, OR and San Diego, CA, was just below the 5% tail of the genome-wide F_{ST} distribution ($F_{ST} = 0.09$), making it a high F_{ST} locus relative to the genome-wide mean but outside the 5% tail of the distribution and significantly outside the 1% tail of the distribution. This position would serve as a good putatively neutral, negative control. The additional locus, 2-alpha fibrillar collagen, was randomly selected among a set of 50 loci with similar F_{ST} and heterozygosity levels as the pyruvate kinase locus.

Primers were designed to amplify the largest fragment possible around the previously identified SNP within exon boundaries using whole gene and exon alignments generated from the sea urchin genome

Table 2 Six putatively adaptive and two putatively neutral genes selected for direct sequencing along the purple sea urchin species range. Genes were selected based on F_{ST}s from the previous RSTA genome scan comparing 10 individuals from each of two populations, Boiler Bay, Oregon and San Diego, California (Pespeni *et al.* 2010). False discovery rate–corrected *P*-values were generated using two outlier detection programs, LOSITAN (Antao *et al.* 2008) and BayeScan (Foll & Gaggiotti 2008; Pespeni *et al.* 2010)

Annotation	Abbreviation	Gene number	Why chosen	Original RSTA F_{ST}	*P* value
Transcription factor 25	TF25	SPU_015723	Statistical outlier	0.42	<< 0.0001
Cubilin	CU	SPU_012783	Statistical outlier	0.33	<< 0.0001
Transporter protein	TP	SPU_000860	Statistical outlier	0.21	<< 0.0001
WD repeat domain 49	WD	SPU_011144	Random from high F_{ST}, high H_T list	0.25	> 0.05
Gaba-b receptor	GR	SPU_004567	Random from high F_{ST}, high H_T list	0.21	> 0.05
Transcription factor	TF	SPU_002852	Random from high F_{ST}, high H_T list	0.18	> 0.05
Pyruvate kinase	PK	SPU_001817	Random, putatively neutral	0.09	>> 0.05
2 alpha fibrillar collagen	CO	SPU_026008	Random, putatively neutral	0.09	>> 0.05

sequencing project (Sea Urchin Genome Sequencing Consortium *et al.* 2006; Spur v.2.5). Amplicon size ranged from 115 to 501 bp with an average of 240 bp per exon and a total of 1914 bp of sequence length across eight exons (Table 3). Primers were designed using Primer3 (Rozen & Skaletsky 2000) and designed to terminate in the second base position of a codon to maximize amplification success (Palumbi 1991). All primer pairs were amplified using a touchdown PCR amplification protocol from 62 to 48 °C for 40 cycles. For each exon, PCR product was confirmed to amplify a single band of predicted length by agarose gel electrophoresis. Amplified products were cleaned and sequenced in both directions using forward and reverse primers (Sequetech, Mountain View, CA, USA). Sequence data were aligned and cleaned using Sequencher software (Gene Codes Corporation). Heterozygous SNPs in an individual were called as such when confirmed with both forward and reverse sequences. Ambiguous sequences were resequenced, often from newly extracted genomic DNA. For each exon, sequences from all individuals were trimmed to a common length ranging from 50 to 447 bp.

Data analyses

Genetic diversity. We used DnaSP (Rozas *et al.* 2003) to phase diploid sequence data and calculate polymorphism statistics: haplotype diversity (Hd), mean number of pairwise differences (k), pi and theta. We estimated recombination rate statistics for each exon and linkage disequilibrium among all nucleotide pairs in DnaSP using only common polymorphisms (allele frequency >12.5%). The cut-off of 12.5% was chosen to be consistent throughout the present study and with a previous study (Fay *et al.* 2002). Cut-offs of 10% and 15% yield similar results. Significance for linkage disequilibrium was calculated with Fisher's exact test and Bonferroni corrected.

Population and biogeographic statistics. We used GENALEX (Peakall & Smouse 2006) to calculate allele frequencies, Hardy–Weinberg equilibrium (HWE) for each SNP, and global and pairwise F_{ST}s using AMOVA for each exon and for each common SNP. We performed all population-level analyses using the common SNPs for each locus rather than haplotypes due to the high haplotypic diversity and haplotype numbers that derive from rare polymorphisms and high recombination rates (Tables 4 and 5). We were unable to unambiguously identify a nonrecombining region to use in defining haplotypes for all loci (Table 5). The number of haplotypes and haplotype diversity was only slightly reduced when reconstructing haplotypes using only common polymorphisms (Table 5, reduced from an average 0.84 to 0.83). We tested for HWE for each common SNP for each locus. Four of the eight exons were in HWE. Cubilin and the transcription factor 25 had a fraction of individual nucleotides out of HWE (15–20%, 2 and 3 nucleotides, respectively); however because these two loci showed some linkage disequilibrium among nucleotide pairs, we were able to test the multi-SNP haplotypes for HWE and found that the haplotypes were in HWE. At locus seven, the WD repeat domain, 49 of the 59

Table 3 Primers used for sequencing eight nuclear exons in the purple sea urchin genome. WD repeat domain 49 was excluded from further analyses as these primers likely amplify multiple paralogous regions of the genome (see Materials and Methods and Results for details)

Gene annotation	Gene no.	Primer sequences	Fragment length	Reference
Transcription factor 25	SPU_015723	5′-CGTGCCATGAGAAAACTGTG-3′ 5′-CAACTCGAATGGGTTCACAA-3′	180	This paper
Cubilin	SPU_012783	5′-GTTCAGCATCTCCCATGGTT-3′ 5′-GTGAAGCCAGAGCTGGTGAC-3′	132	This paper
Transporter protein	SPU_000860	5′-CAGCAAACACCACCATGAACT-3′ 5′-ATGCCGAAGAACACAGGGTA-3′	192	This paper
WD repeat domain 49	SPU_011144	5′-AGCTCAGAGAACAGCCGAAA-3′ 5′-GTCCTCCGATTGGACAGACT-3′	501	This paper
Gaba-b receptor	SPU_004567	5′-GGATGATGAAACGACCGACT-3′ 5′-CATTTTTGAGATGGTTTCGTT-3′	115	This paper
Transcription factor	SPU_002852	5′-GAGCTCCTTGCTGAGATTGG-3′ 5′-GGGTTCTTGGTTCGTTGTGA-3′	270	This paper
Pyruvate kinase	SPU_001817	5′-GAAGCAGAGGCTGCTGTGTT-3′ 5′-CTTCCAGTCTTGGTGAGTACAAT-3′	158	This paper
2 alpha fibrillar collagen	SPU_026008	5′-GCTGTAAGTGCCATCCCTGT-3′ 5′-GGAGTTGGGAACAGGAACAA-3′	366	This paper

Table 4 High diversity and an excess of low-frequency and nonsynonymous polymorphisms in sequenced exons considering all polymorphisms (rare and common). We used DnaSP (Rozas *et al.* 2003) to tabulate the number of polymorphisms and haplotypes, to calculate haplotype and nucleotide diversity statistics and to perform the tests for selection (D_{Taj}, D_{FL}, H_{FW} and D_N/D_S). Bold values in the final four columns indicate statistical significance ($P < 0.05$)

Annotation	No. individuals	Sequence length	Polymorphic sites	No. of amino acid changes	No. of haplotypes estimated	Hd	Mean no. of pairwise differences (k)	pi	Theta	D_{Taj}	D_{FL}	H_{FW}	D_N/D_S
Transcription factor 25	97	50	17	10	21	0.75	2.80	0.06	0.08	−0.70	**−3.02**	**−3.14**	**3.90**
Cubilin	123	112	13	2	36	0.87	2.40	0.02	0.02	0.30	0.79	−0.76	−1.64
Transporter protein	96	159	24	6	32	0.85	2.51	0.02	0.03	−1.34	**−5.37**	**−3.75**	**4.77**
Gaba-b receptor	87	113	9	2	25	0.69	1.63	0.03	0.03	−0.37	−0.16	**−5.11**	−0.11
Transcription factor	111	228	78	31	126	0.97	10.75	0.05	0.07	−0.99	**−2.72**	—	**3.44**
Pyruvate kinase	135	114	15	0	36	0.91	3.10	0.03	0.02	0.31	1.01	—	na
2 alpha fibrillar collagen	125	297	27	3	44	0.82	2.80	0.01	0.02	−1.13	**−3.97**	−1.69	1.05

Table 5 High diversity and recombination rates even when considering only common polymorphisms (>12.5% frequency). We used DnaSP (Rozas *et al.* 2003) to tabulate the number of polymorphisms and haplotypes and to calculate haplotype diversity and recombination statistics. We calculated the fraction of strongly linked nucleotide pairs as the proportion of nucleotide pairs in significant linkage disequilibrium for all pairs and for pairs including the previously identified SNP (RSTA). Global F_{ST} was calculated using all common polymorphisms in each exon using GENALEX (Peakall & Smouse 2006); bold values indicate statistical significance ($P < 0.05$)

Annotation	No. individuals	No. common SNPs	No. common AA SNPs	No. haplotypes estimated	Hd	Average R across nt pairs	Fraction strongly linked nt pairs	Fraction strongly linked RSTA pairs	R per gene	R between adjacent SNPs	No. pairs w/four gametic types	Rm, min. recomb events	Global F_{ST}
Transcription factor 25	145	5	4	12	0.70	0.92	1.00	1.00	3.4	0.07	0/3	0	0.013
Cubilin	132	5	0	19	0.81	0.34	0.70	1.00	33.3	0.30	10/10	4	**0.059**
Transporter protein	145	6	1	18	0.82	0.31	0.67	0.67	23.1	0.15	5/6	3	0.013
Gaba-b receptor	118	6	1	23	0.70	0.45	0.70	0.75	2.4	0.04	10/10	4	0.007
Transcription factor	111	15	7	48	0.91	0.46	0.83	0.75	4.7	0.03	76/78	10	0.000
Pyruvate kinase	135	8	0	27	0.88	0.31	0.73	0.80	25.7	0.23	8/15	3	0.000
2 alpha fibrillar collagen	125	8	1	26	0.78	0.22	0.43	1.00	18.8	0.06	16/21	5	0.000

common SNPs were out of HWE so we excluded this locus from further analyses. Likewise, at an eighth locus (pyruvate kinase), there was a strong heterozygote deficit across all SNPs suggesting uneven amplification of single alleles from some heterozygous individuals. Close inspection of the sequence data revealed for this locus a high number of unlikely homozygous individuals considering the low levels of linkage across nucleotide pairs ($P < 10^{-4}$ for an individual to be homozygous at each of the eight common SNP sites at this locus). To correct for selective amplification, we uniformly excluded all unlikely homozygotes, reducing sample sizes to approximately half across all populations for this locus. Allele frequency shifts among populations were not affected by this procedure. The remaining individuals were included in further analyses.

To explore biogeographic patterns, we used the polymorphism identified in our previous genome scan (Pespeni *et al.* 2010). This SNP tended to be significantly linked to other nearby nucleotides ($P < 0.01$ after Bonferroni correction, Table 5). To explore shifts in allele frequency between neighbouring populations, we calculated the absolute value of the difference in allele frequency between each closest sampled population pair (BM–BB, BB–PP, PP–MY, MY–SD, SD–Baja) for the RSTA SNP in each of the five putatively adaptive loci. For each population pair, we then averaged the difference in allele frequency between closest sampled populations across the five loci. We used ANOVA to test for differences across the five population pairs and a one-tailed *t*-test to compare allele frequency differences in northern vs. southern population pairs to test our a priori hypothesis of limited gene flow into or out of the Southern California Bight.

Correlations between genetic and environmental data. We utilize sea surface temperature data because they are readily and publically available for the entire North American coastline since 1990. However, it should be noted that satellite sea surface temperature data do not capture the variability due to upwelling and microhabitat variation characteristic of the nearshore environment (Helmuth *et al.* 2006). Ocean pH is another important variable to consider when studying local adaptation in this upwelling ecosystem; however, these data are only recently being generated for specific localities along the US west coast (Evans *et al.* 2013). Here, we calculated average annual and monthly sea surface temperatures from NOAA Pathfinder data (http://data.nodc.noaa.gov/pathfinder/Version5.2/). For each of the eight sites (Table 1), we extracted data using the xtracto_3D script with the third dimension as time (http://coastwatch.pfel.noaa.gov/xtracto/R/code/xtracto_3D_bdap.R) in the R programming environment (R Development Core Team 2009). We used latitude and longitude points approximately 5–10 km offshore from each site because satellite imaging did not consistently produce data for the very nearshore environment likely due to cloud cover or other sources of interference. We calculated average monthly and annual temperatures spanning a 20-year period from 1 January 1990 to 30 December 2009.

We used partial Mantel tests implemented in R with the 'vegan' and 'ade4' packages (as in Goldberg & Waits 2010) to test for a relationship between allele frequency and temperature at each site, for each locus, while controlling for geographical distance. We standardized the variance for all variables by dividing by the mean and subtracting 1 standard deviation (as in Selkoe *et al.* 2010).

Allele-based statistics. To test for signals of neutrality vs. selection using the distribution of allele frequency variants, we calculated Tajima's *D*, Fu and Li's *D*, and Fay and Wu's *H* (Tajima 1989; Fu & Li 1993; Fay & Wu 2000) using DnaSP (Rozas *et al.* 2003) with all polymorphisms (common and rare). These statistics are sensitive to the relative frequency of rare vs. common variants. An excess of rare variants results in a negative statistic, and values much less than −1 are suggestive of positive selection, a selective sweep or a population bottleneck. To calculate Fay and Wu's *H*, we used whole-transcriptome sequence data from *Allocentrotus fragilis*, a sister species to *Strongylocentrotus purpuratus*, available for five genes to provide a reference for identifying derived allele states (Oliver *et al.* 2010).

A second test for selection compares the rate of amino acid-changing mutations (d_N) to the rate of synonymous mutations (d_S) in a gene (Nei & Gojobori 1986). A ratio of d_N to d_S much >1 indicates an excess of nonsynonymous mutations over neutral expectations and is suggestive of positive selection.

In addition, the relative amino acid polymorphism (A)–to–synonymous polymorphism (S) ratios among rare vs. common polymorphisms can also reveal signals of selection (Fay *et al.* 2001, 2002). Based on the neutral theory and studies in humans and *Drosophila* and because most amino acid changes are expected to be deleterious, it is expected that the amino acid-to-synonymous ratio should be greater among rare polymorphisms than among common polymorphisms (Kimura 1983; Fay *et al.* 2001, 2002). Negative or purifying selection should prevent deleterious mutations from reaching common frequencies, while positive selection will lead to the quick fixation of beneficial mutations. Common amino acid polymorphisms may therefore be the result of (i) relaxed purifying or negative selection (Fay *et al.* 2001) or (ii) an environment-dependent beneficial mutation maintained

by spatial or temporal balancing selection and gene flow across a heterogeneous landscape (Levene 1953; Hedrick *et al.* 1976; Schmidt *et al.* 2000). To compare our SNP data to the expectation of an excess of rare amino acid SNPs, we counted the number of amino acid and synonymous variants among common and rare polymorphisms (>12.5% and $\leq$12.5%, respectively) for the five exons with at least one common amino acid polymorphism. We calculated the ratio of A/S for common and for rare SNPs and divided the A/S ratios to reveal any excess of common amino acid SNPs. Because the A/S ratio is expected to be greater among rare polymorphisms, there is the convention of calculating the excess number of rare amino acid polymorphisms. We calculated the excess number of rare amino acid SNPs as the observed minus expected number of rare amino acid SNPs, where expected is the product of S (for rare SNPs) and A/S (in common SNPs; as in Fay *et al.* 2001). Negative values represent a deficit in the number of rare amino acid SNPs and show an excess of common amino acid polymorphisms.

Results

Genetic diversity in purple sea urchin exons

We found high levels of diversity in each of the eight nuclear exons sequenced (Table 4) with an average nucleotide diversity per site between two sequences (π) of 0.03 across seven exons. This estimate did not include the WD repeat protein 49 because 52% of the nucleotide sites in this exon were polymorphic. Considering just the common polymorphisms (frequency >12.5%), the WD repeat domain 49 showed almost no linkage across nucleotide pairs (Table 5). The high polymorphism and recombination levels in the WD repeat domain 49 may have been due to spurious amplification of another genomic region of similar size and sequence homology and so this exon was excluded from further analyses. Among the remaining seven loci, considering all polymorphisms, 18% of nucleotide sites on average were polymorphic. The average θ_W was 0.04, reflecting the large population size and high mutation rate known for purple sea urchins (Britten *et al.* 1978; Ebert & Russell 1988; Sea Urchin Genome Sequencing Consortium *et al.* 2006).

Considering just the common polymorphisms, we found high levels of recombination, with an average R of 0.4 across the seven exons (Table 5). However, rates differed across loci (see individual nucleotide frequencies in three genes in Fig. 2). The transcription factor 25 was the only exon that showed near complete linkage across the five common polymorphisms considered (Fig. 2A).

The two transcription factors had the highest fraction of amino acid-changing polymorphisms across all SNPs (Table 4, TF25: 60% and TF: 41% of codons sequenced) and considering only common SNPs (Table 5, TF25: 24% and TF: 9% of codons sequenced) and also showed the strongest levels of linkage disequilibrium (Table 5, TF25: $R = 0.92$, TF: $R = 0.46$). In contrast, cubilin and pyruvate kinase had no common amino acid polymorphisms, while the transporter protein, the gaba-b receptor and the collagen protein each had one (Table 5).

Population and biogeographic patterns

Considering all common SNPs together, one locus, the cubilin receptor, had a significant global F_{ST} (0.06, $P = 0.02$) and showed a clinal pattern of variation along the species range (Table 5, Fig. 3B). One amino acid polymorphism, glycine to serine in the gaba-b receptor, had a significant global F_{ST} (0.05, $P = 0.02$) and showed a clinal pattern along the species range (Fig. 3F). 95% of alleles in the northernmost populations coded for glycine shifting to only 67% of alleles in Baja California with a glycine residue and 33% with a serine. This nearly 30% shift in allele frequency could represent a strong functional cline considering the high levels of gene flow along the species range. This polymorphism showed a near significant correlation to latitude (Mantel test: $r = 0.41$, $P = 0.09$), but not to sea surface temperature (partial Mantel test controlling for latitude: $r = -0.70$, $P = 1$).

Four of the seven loci showed clinal shifts in allele frequencies across at least four closest sampled populations: the two transporter proteins including cubilin, one of the transcription factors and the gaba-b receptor (Fig. 3, panels B, C, D and E). In each of these cases, however, there were one or two populations that did not follow this clinal pattern. For the two putatively neutral loci, pyruvate kinase and the collagen protein, allele frequency patterns shift up and down between the closest sampled populations (Fig. 3, panels G and H). These patterns in the putative adaptive loci suggest local adaptation to spatially heterogeneous environmental conditions in specific localities and dispersion of adaptive alleles in space. Patterns in the putative neutral loci show no such signature of selection.

The transcription factor 25 shows highly linked SNPs across the exon with many amino acid polymorphisms and had large shifts in allele frequencies between populations (Fig. 2A). Two major haplotypes differed at three amino acid positions (across only 17 codons sequenced), including a nonconservative asparagine to aspartic acid change. The allele frequency pattern showed a 26% decrease in the frequency of one haplo-

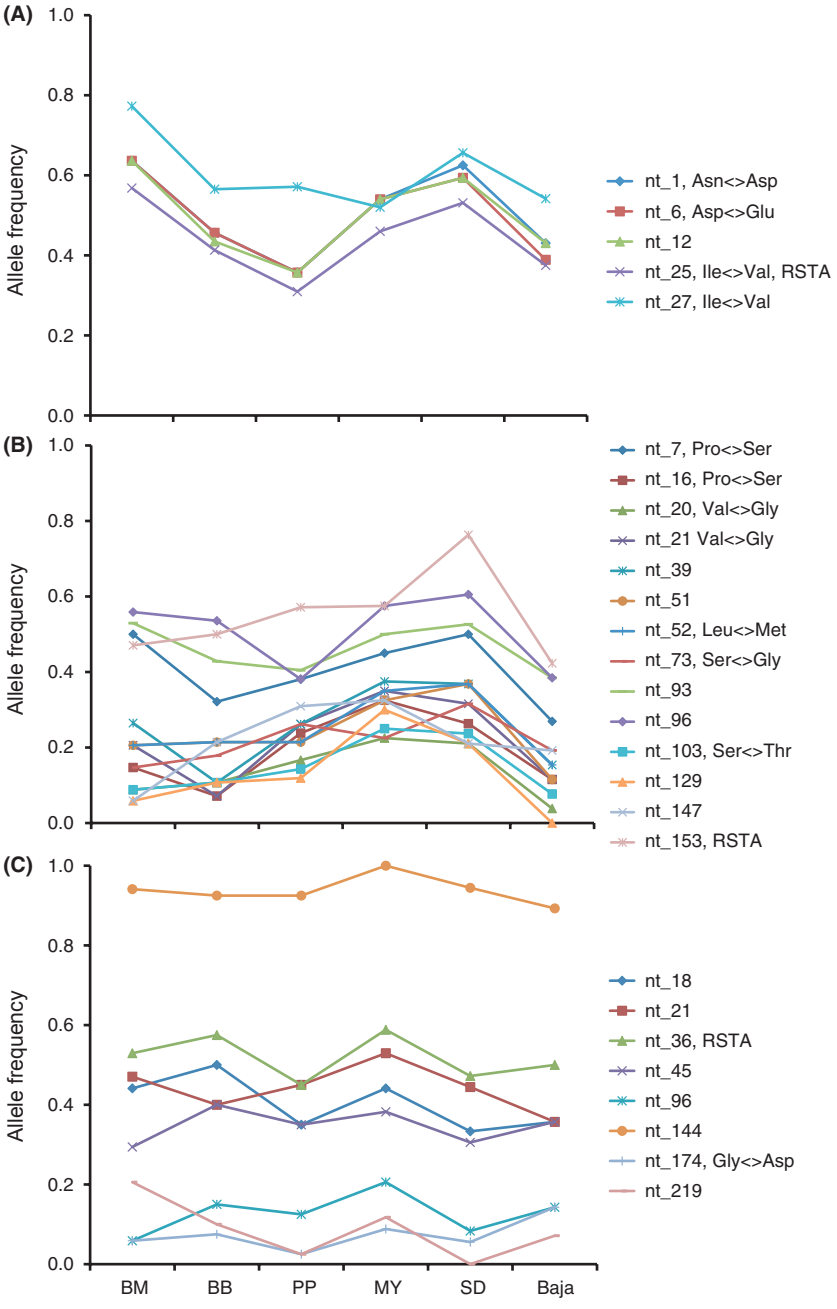

Fig. 2 High genetic diversity, heterozygosity and variation in recombination rate across three exons. Each line represents the allele frequency of a single common polymorphism along the geographical range of the species from north to south. Amino acid polymorphisms and the RSTA polymorphism are noted in the legend for each panel. (A) Transcription factor 25 (SPU_015723)—low recombination and high diversity, (B) transcription factor (SPU_002852)—high recombination and high diversity and (C) 2 alpha fibrillar collagen (SPU_026008)—high recombination and lower diversity.

type from Bamfield, Canada to Patrick's Point in Northern California followed by a 23% increase to San Diego, California, then a 15% decrease to Baja California, Mexico. This locus had the second highest global F_{ST} but was not significant, possibly because of the odd geographical pattern of allele frequencies. Bamfield and San Diego were similar in haplotype and SNP frequencies compared with the other six localities.

To measure the shift in allele frequency between population pairs, we compared the average shift in frequency for a single sequenced polymorphism representing each locus (the SNP previously identified in the genome scan) for each of the five closest sampled population pairs (BM–BB, BB–PP, PP–MY, MY–SD and SD–Baja). The average shifts in allele frequency were greatest between Monterey and San Diego (0.12) and San Diego and Baja (0.14) compared with other population pairs (BM–BB = 0.08, BB–PP = 0.09, PP–MY = 0.08) suggesting either restricted gene flow in and out of San Diego compared with other populations or a signal of local adaptation to unique environmental conditions in the Southern California Bight (Figs 3 and 4; northern

population pairs vs. southern population pairs including San Diego: one-tailed *t*-test *P = 0.0071*).

We found that allele frequencies in two loci, cubilin (CU, Fig. 3B) and the transcription factor (TF, Fig. 3D), were correlated with average annual temperature (Fig. 3I, partial Mantel: CU: $r = 0.52$, $P = 0.04$, TF: $r = 0.47$, $P = 0.03$). The transporter protein also showed a noteworthy relationship to average annual temperature (Fig. 3C, partial Mantel: TP; $r = 0.43$, $P = 0.09$). Genetic distances were also correlated with geographical distances (Mantel: $r = 0.53$, $P = 0.04$) for cubilin. The two other putatively adaptive loci, the transcription factor 25 and the gaba-b receptor, and the two putatively neutral loci showed no relationship with temperature or distance (Fig. 3C,E,G,H).

Tests for selection

Considering all polymorphisms for the following frequency-based statistics, we found that four of the five putative adaptive loci showed an excess of low-frequency polymorphisms using one or both of Fu and Li's *D* and Fay and Wu's *H* tests for selection (Table 4). There were no deviations from neutrality using the more conservative Tajima's D test (Table 4). Among the putative neutral loci, the collagen protein also showed an

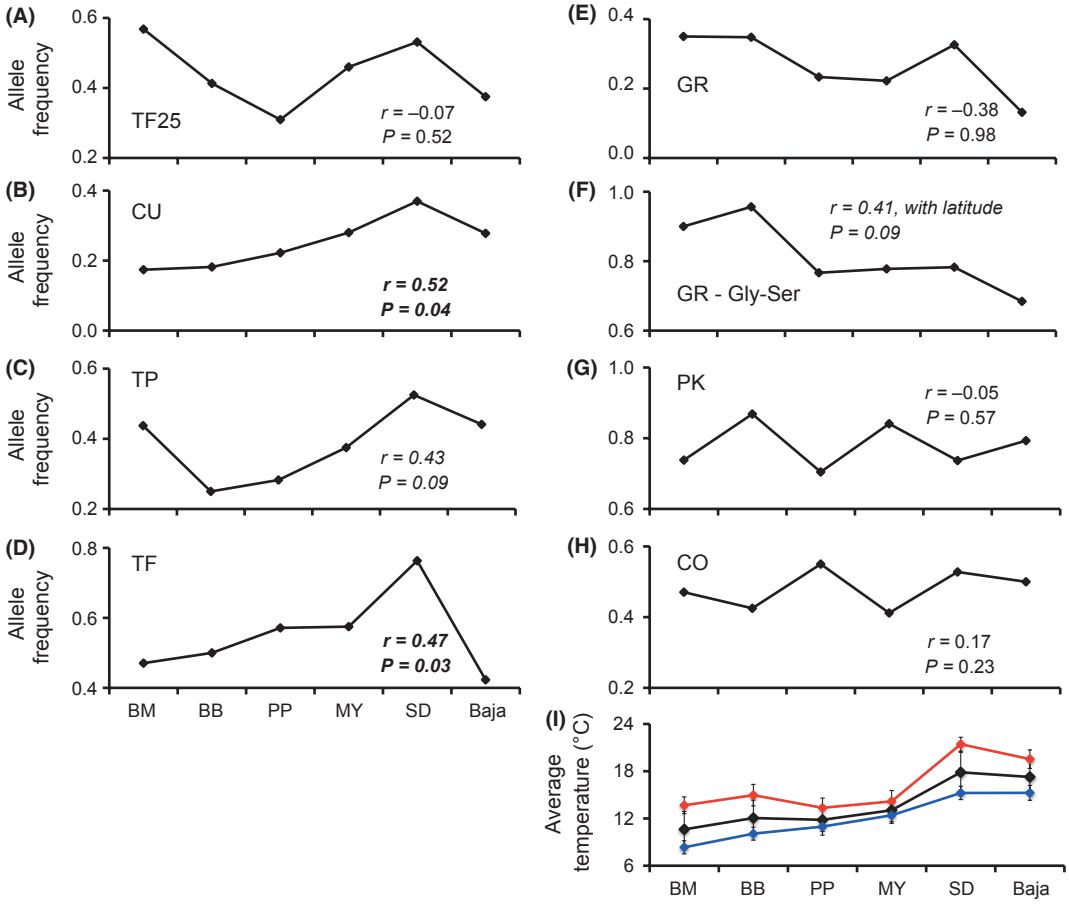

Fig. 3 Allele frequency patterns from north to south along the species range for five putative adaptive (A–E) and two putative neutral exons (G and H) represented by sequence data from the present study of the polymorphism identified in the previous genome scan. Geographical patterns in allele frequency correlate with average annual temperature (I) for two loci (cubilin and the transcription factor, panels B and D) and show a positive trend for a third locus (transporter protein, panel C). Correlation coefficients and *P*-values are the results of partial Mantel and Mantel tests (in the case of cubilin); bold indicates $P < 0.05$, italics $P < 0.10$. The glycine-to-serine amino acid polymorphism that shows significant structure along the species range is also shown (F). This locus correlates with latitude rather than temperature. Error bars in the temperature data (I) represent standard deviation across 20 years of satellite images. The red line shows temperatures in the warmest months, July and August, while the blue line shows temperatures in the coldest months, February and March. Note that offshore sea surface temperature data do not capture the variability due to upwelling and microhabitats that characterize the nearshore environment but rather represent a course approximation of site differences in temperature (see Materials and Methods for more details). See Table 2 for locus abbreviations.

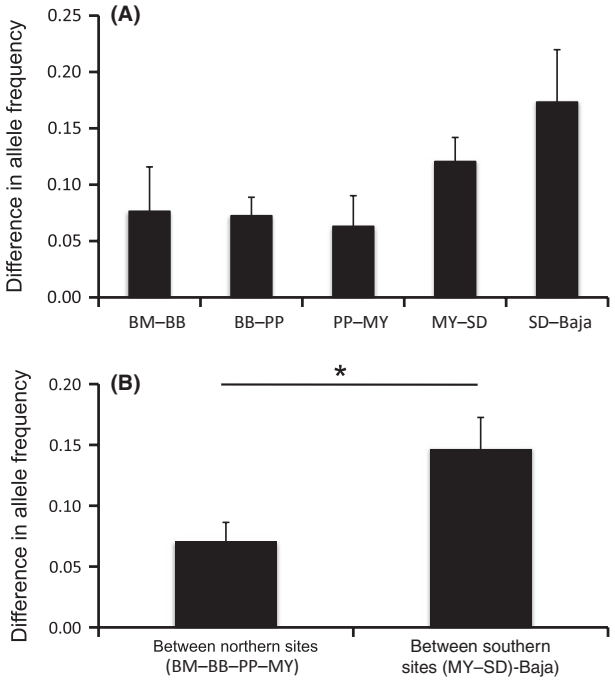

Fig. 4 Shifts in allele frequency between closest sampled populations suggest local adaptation or local retention of alleles in San Diego urchins or both, considering the five putative adaptive loci. (A) Differences in allele frequency for each closest sampled population pair (ANOVA $P = 0.14$) and (B) considering changes in allele frequency between population pairs including San Diego (MY–SD and SD–Baja) vs. changes in allele frequency between the four northern populations (BM–BB, BB–PP and PP–MY) (one-tailed *t*-test $P = 0.0071$). Error bars represent standard error across the five putative adaptive loci. Note that changes in allele frequency are not significant when putative neutral loci are included, suggesting that adaptive alleles may be retained in the Southern California Bight by selection.

excess of low-frequency polymorphisms using Fu and Li's *D* (Table 4). This result is the only significant deviation from neutrality for a putatively neutral locus out of all frequency and biogeographic tests for selection.

We calculated d_N/d_S for each locus and found a significant excess of amino acid-changing polymorphisms in three of the seven exons, the transporter protein and the two transcription factors, the same genes that showed an excess of low-frequency polymorphisms (Table 4). Amino acid substitutions among common polymorphisms (frequency >12.5%) were generally moderate in physiochemical difference (Table 6, 10 of 14 moderate, 71%) based on Grantham's D index of differences between amino acid pairs (Grantham 1974), and none were radical amino acid substitutions (0 of 14). By contrast, among the rare amino acid polymorphisms (frequency $\leq$12.5%), there were many radical changes (10 of 63, 16%). Taken together, these results suggest the action of purifying selection among common amino acid polymorphisms limiting the frequency of radical amino acid variants in populations.

In the context of neutral evolution, the ratio of amino acid to synonymous variants is expected to be greater among rare polymorphisms than among common polymorphisms due to purifying selection removing deleterious mutations (Kimura 1983). In contrast, we found that the A/S ratio was greater among common than rare variants in three of five comparisons, in the two transcription factors and in the collagen protein (Table 7). The average ratio of common to rare A/S across the five loci was 1.12 showing an overall excess of common amino acid polymorphisms (value >1). The observed A/S ratio in common SNPs can be used to calculate the expected number of rare amino acid SNPs for a given number of rare synonymous SNPs (see Materials and Methods). We observed an overall deficit rather than excess of rare amino acid SNPs revealing an excess of common amino acid SNPs. This pattern was driven by the high number of common amino acid polymorphisms in the two transcription factors (TF25: 4 of 5, TF: 7 of 15, Table 7).

Discussion

We present evidence of geographical and sequence-based signals of selection in the genomic regions of loci previously identified as F_{ST} outliers. We find that biogeographic patterns of genetic variation correlate with latitude and temperature in the putative adaptive loci. Among allele genealogies, we find non-neutral patterns at five of seven loci and high d_N/d_S at three. At the nucleotide level, we find an excess of common amino acid polymorphisms in the two transcription factors suggestive of balancing selection. We consider models of spatial and temporal balancing selection, soft sweeps from standing genetic variation and the ascertainment bias inherent in a follow-up study to explain these combined results in this highly dispersing, broadly distributed species.

Polymorphism and recombination

Compared with most other animals whose genetic diversity has been explored in detail, we found very high levels of polymorphism and recombination. However, our estimates are similar to what has been observed previously in the purple sea urchin (Britten *et al.* 1978; Flowers *et al.* 2002; Sea Urchin Genome Sequencing Consortium *et al.* 2006) and in *Drosophila* autosomes (Begun *et al.* 2007). As expected, diversity is two orders of magnitude greater than what has been observed in humans [$\theta = 0.04$ in this study vs. 0.0004 in

Table 6 All common amino acid polymorphisms (frequency >12.5%). D is an index of physiochemical differences between amino acid pairs that ranges from 5 (Ile-Leu) to 215 (Cys-Trp; Grantham 1974), where <50 is conservative, 50–100 is moderately conservative, 100–150 is moderately radical and >150 is radical (Li *et al*. 1985). F_{ST} was calculated for each polymorphism using GENALEX (Peakall & Smouse 2006); bold values indicate statistical significance ($P < 0.05$)

Annotation	SNP site	Amino acids	Groups	D	F_{ST}
Transporter protein	nt_79	Glu< >Lys	Polar acidic–Polar basic	56	0.01
Transcription factor	nt_7	Pro< >Ser	NonpolarHp–Polar uncharged	74	0.00
	nt_16	Pro< >Ser	NonpolarHp–Polar uncharged	74	0.01
	nt_20	Val< >Gly	NonpolarHp–Polar uncharged	109	0.02
	nt_21	Val< >Gly	NonpolarHp–Polar uncharged	109	0.01
	nt_52	Leu< >Met	NonpolarHp	15	0.00
	nt_73	Ser< >Gly	Polar uncharged	56	0.00
	nt_103	Ser< >Thr	Polar uncharged	58	0.00
Gaba-b receptor	nt_37	Ser< >Gly	Polar uncharged	56	**0.05**
Transcription factor 25	nt_1	Asn< >Asp	Polar acidic–Polar uncharged	23	0.02
	nt_6	Asp< >Glu	Polar acidic	45	0.01
	nt_25	Ile< >Val	NonpolarHp	29	0.01
	nt_27	Ile< >Val	NonpolarHp	29	0.00
2 alpha fibrillar collagen	nt_174	Gly< >Asp	Polar acidic	94	0.00

Table 7 Relative numbers of rare ($\leq$12.5%) and common (>12.5%) amino acid and synonymous polymorphisms for each of the five exons that have at least one common amino acid polymorphism. The ratio of common over rare A/S ratios shows for values >1 that there are more amino acid polymorphisms among common polymorphisms than among rare polymorphisms, and the reverse true for values <1. The excess number of rare amino acid polymorphisms was calculated as the observed number of rare amino acid SNPs minus the expected number of rare amino acid SNPs, where the expected number of rare amino acid SNPs is calculated as the product of S (for rare SNPs) and A/S (in common SNPs, as in Fay *et al*. (2001)). Negative values represent a deficit in the number of rare amino acid SNPs and show an access of common amino acid SNPs

Annotation	Class	Amino acid polymorphism, A	Synonymous polymorphism, S	A/S	$A/S_{Common}/A/S_{Rare}$	Excess Rare A
Transcription factor 25	Rare	6	5	1.20	3.33	−14
	Common	4	1	4.00		
Transporter protein	Rare	5	13	0.38	0.52	2
	Common	1	5	0.20		
Gaba-b receptor	Rare	1	2	0.50	0.40	1
	Common	1	5	0.20		
Transcription factor	Rare	24	39	0.62	1.42	−10
	Common	7	8	0.88		
2 alpha fibrillar collagen	Rare	2	17	0.12	1.21	0
	Common	1	7	0.14		
Total	Rare	38	76	0.50	1.08	−3
	Common	14	26	0.54		
Total among putatively adaptive loci	Rare	36	59	0.61	1.12	−4
	Common	13	19	0.68		

humans (Sachidanandam *et al*. 2001; Wang *et al*. 1998)]. Recombination rates among these sea urchin loci differed greatly from locus to locus, as has been seen in humans and *Drosophila* (Begun & Aquadro 1992; Nachman 2001). For some loci, SNPs were highly linked across 100s of bp, whereas for other loci, significant linkage did not occur between nearby nucleotides. The apparent lack of linkage between nearby nucleotides in some loci in this study is likely the result of several potential causes: (i) we sampled individuals from wild, out-crossed populations that have large effective population sizes, and as a result, sampled individuals are likely unrelated and drawn from a huge reservoir of genetic variation and (ii) the high mutation and recombination rates in purple sea urchins likely result in homoplasy and other mutational events that break linkage patterns. Both explanations are likely in this study system.

Signals of selective sweeps

Although we detect signals of selection at the nucleotide and amino acid level, we do not observe a reduction in genetic diversity as is commonly predicted with the hitchhiking of neutral, close or 'linked' loci to a favourable mutation in a selective sweep (Maynard-Smith and Haigh 1974). The spatial extent of a selective sweep is small when recombination rate is high (Begun & Aquadro 1992; Nachman 2001). As a result, we may not have observed reduced genetic diversity around putative selected sites because recombination rates are very high in four of the five putatively adaptive loci.

If an adaptive mutation is common because it is maintained by balancing selection, hitchhiking in areas of high recombination will lead to increased nucleotide diversity (Hudson and Kaplan 1988, Kaplan *et al.* 1989). This model could explain the signals of selection without reduced genetic diversity we observe. The lack of reduced genetic diversity around potentially beneficial mutations could be the sign of a 'soft sweep', a beneficial mutation that increases in frequency from standing genetic variation or recurrent beneficial mutations (Hermisson & Pennings 2005; Pennings & Hermisson 2006). Adaptation from standing genetic variation or recurrent mutations would not result in a strong reduction in linked neutral variation (Hermisson & Pennings 2005; Pennings & Hermisson 2006). Under these models, the lack of reduced diversity around a beneficial polymorphism reduces the power of frequency-based neutrality tests (Pennings & Hermisson 2006). This fact coupled with the high recombination rates and high heterozygosity in this species may explain the lack of significant results using the more conservative Tajima's *D* test for neutrality. In species with large effective population sizes and high levels of standing genetic variation, adaptation may be more likely to occur from a soft sweep from standing variation than a hard sweep in a mutation-limited scenario. Such soft sweeps have been observed in other species with large effective population sizes similar to the purple sea urchin: for example, they have led to pesticide resistance in *Drosophila* (Karasov *et al.* 2010) and see (Barrett & Schluter 2008) for a number of other examples.

Population and biogeographic patterns

In four of the five putatively adaptive loci, we found associations between genetic variation and latitude or temperature (Table 8 and Fig. 3). As predicted based on differences in environmental conditions, we measured the greatest shifts in allele frequency between San Diego and populations sampled to the north and south of the Southern California Bight (Fig. 4). These results suggest that San Diego urchins may harbour or require alleles with higher fitness at warmer temperatures or at higher pH (Hofmann *et al.* 2011). These biogeographic patterns showing San Diego as an outlier could be an artefact of the choice of outlier loci based on a comparison of Boiler Bay to San Diego in the previous study (Pespeni *et al.* 2010). However, if these results were due strictly to such geographical ascertainment bias, it would be equally likely to observe Boiler Bay as an outlier as it would be to observe San Diego as an outlier. We find no signals of Boiler Bay as a geographical outlier, which suggests that results cannot be due strictly to ascertainment bias and that San Diego is likely an outlier for environmental reasons.

Two loci showed clinal patterns of variation and significant global F_{ST} measurements: the cubilin exon and a glycine-to-serine polymorphism in the gaba-b receptor. These clines showed significant shifts in allele frequency of about 20 and 30% along the species range, respectively. Such shifts could be the result of moderate selection balanced against high gene flow (Sotka & Palumbi 2006). Several previous studies have shown that physiochemically moderate amino acid changes, such as glycine to serine, can have major effects on protein flexibility and thermal stability (reviewed in Hochachka & Somero 2002). Interestingly, these two proteins, cubilin and the gaba-b receptor, of the seven in this study have been shown to play important roles in vitamin and nutrient uptake during development in mammal models (Kozyraki 2001) and provide cues for larval settlement in abalone (Morse *et al.* 1979), respectively. This suggests that larvae may experience a significant latitudinal gradient in some selective pressure, perhaps related to temperature.

Interestingly, only two of the five loci previously identified as F_{ST} outliers (Pespeni *et al.* 2010) show significant global F_{ST} in the present study. There are several potential explanations for this finding. First, allele frequency estimates from the RSTA study were based on fewer individuals sampled, 10, from fewer locations, 2, and therefore could have been less accurate. However, a comparison of two genome-wide data sets that estimated allele frequencies in purple sea urchins, from the RSTA study (Pespeni *et al.* 2010) and another study using RNA sequencing from 20 individuals from seven populations (Pespeni *et al.* 2013b), showed a strong correlation in allele frequency estimates (slope = 1, $R^2 = 0.75$, $P < 0.0001$). This concordance suggests that both approaches yield accurate allele frequency estimates. Second, global F_{ST} in the present study is calculated based on all SNPs in the sequenced exon, while the F_{ST} outlier is just one SNP. High recombination and mutation rates, as is observed in this species, will obscure significant global F_{ST} estimates in these exon sequence data. Third, the initial identification of an outlier locus might be spurious because of

unknown spatial structure or variation in population size and demography through time (Excoffier *et al.* 2009; Bierne 2010; Ralph & Coop 2010) (see (Bierne *et al.* 2011) for a review of alternative explanations for outlier loci). Choosing 'outlier' loci could select for regions of the genome that have similar genealogical histories as those expected under directional selection, yielding false positive results in follow-up studies (Thornton & Jensen 2007). However, this artefact would not explain the amino acid polymorphism results and so is unlikely to be responsible for the patterns we report here.

Signals of selection

All five candidate loci show multiple signals of selection at the nucleotide, amino acid or biogeographic level (see summary Table 8 and Tables 4–7). A particularly interesting result is that the common SNPs have a 12% higher A/S ratio (amino acid/synonymous polymorphism) than rare SNPs. Typically, by contrast, 31 autosomal loci in *Drosophila melanogaster* showed a 41% drop in A/S for rare SNPs (Fay *et al.* 2002). The two transcription factors show the highest levels of excess of common amino acid polymorphisms. An excess of common amino acid polymorphisms supports a model of the maintenance of alternatively adaptive alleles by balancing selection in an environment where the selective pressures change in space or time (Levene 1953; Hedrick *et al.* 1976). Under this model, recruits to a population arrive from a common larval pool or a distant population with different habitat conditions and are selected by the local environment. Their progeny then are mixed together in the dispersal pool for the next generation (Fig. 5). If selection pressures vary significantly from place to place or time to time, then the dispersal pool will contain more common polymorphisms than expected under a neutral model. This spatial or temporal balancing selection acts to increase functional diversity.

An excess of common amino acid polymorphisms could also be due to relaxed purifying selection; however, in our data, we see a signal of purifying selection. Common amino acid polymorphisms were moderate and none were radical in physiochemical differences (Table 6). By contrast, several of the low-frequency amino acid polymorphisms were radical in physiochemical difference (16%). These results suggest that purifying selection is in fact acting to limit radical amino acid polymorphisms from becoming common. This supports the idea that common amino acid polymorphisms may be maintained by balancing selection, rather than permitted by relaxed purifying selection.

Another alternative explanation for excess nonsynonymous and common SNPs could be the hitchhiking of weakly deleterious mutations with a beneficial mutation when the deleterious effects do not outweigh the advantageous effects (Fay 2011). As a consequence, the hitchhiking of weakly deleterious mutations could inflate estimates of nonsynonymous and common SNPs when there is sufficient linkage. This could explain patterns in the transcription factor 25, the one locus that shows strong linkage and has the highest excess of common amino acid polymorphisms. The beneficial mutation being selected could be in the transcription factor or in a physically close, linked locus.

The high-frequency amino acid polymorphisms could be (i) maintained by balancing selection or (ii) the result of a long period of isolation and subsequent unification of alleles through secondary introgression. Isolation and subsequent introgression is less parsimonious given the degree of genetic connectivity previously measured among populations of this species showing all alleles shared at multiple spatial scales (Palumbi & Wilson 1990; Edmands *et al.* 1996; Olivares-Banuelos *et al.* 2008; Pespeni *et al.* 2010). Further, previous studies have found signals of balancing selection in the purple sea

Table 8 Summary of signals of selection across five putatively adaptive exons (first five in table) and two putatively neutral exons (last two). Check marks indicate positive tests for selection ($P < 0.05$), the dot indicates a noteworthy trend ($P < 0.10$), and dashes indicate the test could not be performed either because of lack of an outgroup sequence for H_{FW} or because of an absence of common amino acid polymorphisms. Blank spaces indicate negative results for tests for selection

Annotation	D_{Taj}	D_{FL}	H_{FW}	D_N/D_S	Excess Common AA SNPs	Global F_{ST}	Correlation with temperature
Transcription factor 25		✓	✓	✓	✓		
Cubilin					—	✓	✓
Transporter protein		✓	✓	✓			•
Gaba-b receptor			✓			✓	
Transcription factor		✓	—	✓	✓		✓
Pyruvate kinase			—		—		
2 alpha fibrillar collagen		✓					

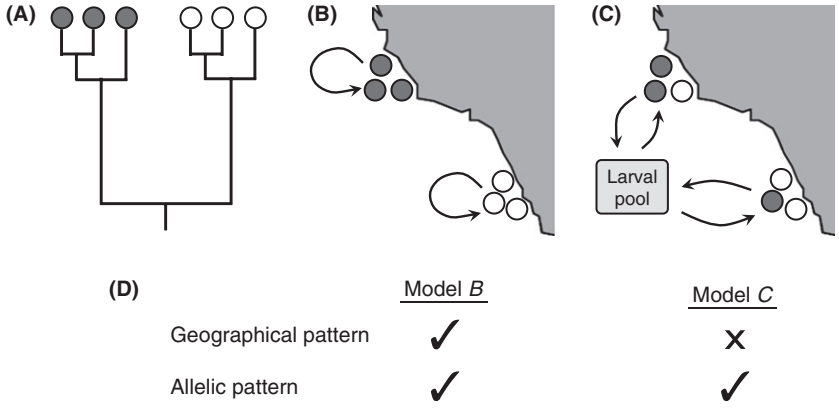

Fig. 5 Geographical and allelic patterns of selection. (A) Genealogical differentiation of alleles with functional differences acted on by selection (shaded vs. open clades). (B) Hypothetical coastal populations without larval dispersal among environmentally different habitats selecting for different allele clades. (C) Hypothetical populations with larval dispersal among environments and selection (C). Model *B* represents the classic view of the mechanism of local adaptation via retention of locally adapted alleles, while model *c* represents recruitment from a mixed larval pool and repeated selection every generation. (D) Model *B* will show both geographical patterns between habitats and genealogically based allelic patterns of selection between clades, while model *C* will show allelic patterns of selection and mild geographical patterns depending on migration/selection balance.

urchin, particularly excess genetic diversity in immune-related genes (Hibino *et al.* 2006; Terwilliger *et al.* 2006; Pespeni *et al.* 2012).

We find several signals of selection acting on the two transcription factors included in this study. Differences in gene regulation have been shown to be fundamental to adaptive and morphological evolution (Davidson 2001; Wray *et al.* 2003; Carroll *et al.* 2005; Gompel *et al.* 2005). Several genome-wide surveys of genetic diversity have identified transcription factor proteins as targets of selection in humans (Kayser *et al.* 2003; Bustamante *et al.* 2005; Gilad *et al.* 2006). Among Strongylocentrotid sea urchins including the purple sea urchin, Oliver *et al.* found enhanced evolutionary rates in several transcription factor genes (Oliver *et al.* 2010). There are several other lines of evidence that suggest an important role for regulatory evolution along the purple sea urchin species range: (i) an equal proportion of genetic variation in upstream putative regulatory regions as in coding regions suggests evolutionary constraint of functionally important motifs rather than neutral evolution in these non-coding regions (Pespeni *et al.* 2010), (ii) enrichment of high F_{ST} polymorphisms (nonrandom distribution with respect to protein function) among only regulatory polymorphisms (Pespeni *et al.* 2012) and (iii) lasting differences in gene expression, particularly in genes related to growth and biomineralization, after 3 years of common garden acclimation in adult purple sea urchins (Pespeni *et al.* 2013a). These previous results combined with the data on transcription factors in the present study suggest that polymorphisms of regulatory alleles in purple sea urchins may be maintained by natural selection.

Conclusions

We have used sequence data from seven nuclear exons to demonstrate signals of selection that match predictions from initial outlier tests and the biogeography of populations of the purple sea urchin. The putatively adaptive loci exhibit signals of soft sweeps and balancing selection. Our results illustrate the value of follow-up studies on candidates of selection identified in genome scans with direct sequencing to characterize the potential modes of natural selection acting in a species. Biogeographic patterns also illustrate that environmental gradients and hot spots may shape genetic variation in key locations across the genome and across the species range.

Acknowledgements

The authors would like to thank K. Bartlett, E. Metz, A. Haupt, S. Lester, T. Tydingco, R. Kelly, V. Vacquier and A. Sivasundar for assistance with sample collection and A. Beichman, particularly, and K. Yeung, G. Ruiz-Jones and M. Oster for their assistance in generating and cleaning sequence data. We also thank the anonymous reviewers whose comments helped significantly improve the manuscript. This work was supported by NSF SGER-0714997 (S.R.P.), NSF Graduate Research Fellowship (M.H.P.) and the Partnership for Interdisciplinary Studies of Coastal Oceans (PISCO). This is contribution 425 from PISCO, funded primarily by the Gordon and Betty Moore Foundation and the David and Lucille Packard Foundation.

References

Antao T, Lopes A, Lopes RJ, Beja-Pereira A, Luikart G (2008) LOSITAN: a workbench to detect molecular adaptation based on a F_{ST}-outlier method. *BMC Bioinformatics*, **9**, 323.

Bakun A, Parrish RH (1982) Turbulence, transport, and pelagic fish in the California and Peru Current systems. *California Cooperative Oceanic Fisheries Investigations Report*, **23**, 99–112.

Barrett RD, Schluter D (2008) Adaptation from standing genetic variation. *Trends in Ecology & Evolution*, **23**, 38–44.

Beaumont MA, Balding DJ (2004) Identifying adaptive genetic divergence among populations from genome scans. *Molecular Ecology*, **13**, 969–980.

Begun DJ, Aquadro CF (1992) Levels of naturally occurring DNA polymorphism correlate with recombination rates in *Drosophila melanogaster*. *Nature*, **356**, 519–520.

Begun DJ, Holloway AK, Stevens K *et al.* (2007) Population genomics: whole-genome analysis of polymorphism and divergence in *Drosophila simulans*. *PLoS Biology*, **5**, e310.

Bierne N (2010) The distinctive footprints of local hitchhiking in a varied environment and global hitchhiking in a subdivided population. *Evolution*, **64**, 3254–3272.

Bierne N, Welch J, Loire E, Bonhomme F, David P (2011) The coupling hypothesis: why genome scans may fail to map local adaptation genes. *Molecular Ecology*, **20**, 2044–2072.

Britten RJ, Cetta A, Davidson EH (1978) The single-copy DNA sequence polymorphism of the sea urchin *Strongylocentrotus purpuratus*. *Cell*, **15**, 1175–1186.

Burford M (2009) Demographic history, geographical distribution and reproductive isolation of distinct lineages of blue rockfish (*Sebastes mystinus*), a marine fish with a high dispersal potential. *Journal of Evolutionary Biology*, **22**, 1471–1486.

Bustamante CD, Fledel-Alon A, Williamson S *et al.* (2005) Natural selection on protein-coding genes in the human genome. *Nature*, **437**, 1153–1157.

Carroll SB, Grenier JK, Weatherbee SD (2005) *From DNA to Diversity: Molecular Genetics and the Evolution of Animal Design*. Blackwell Publishing, Malden, MA.

Carvalho G (2005) Evolutionary aspects of fish distribution: genetic variability and adaptation. *Journal of Fish Biology*, **43**, 53–73.

Davidson EH (2001) Genomic regulatory systems: development and evolution. Academic Press, San Diego, CA.

De Wit P, Pespeni MH, Ladner JT *et al.* (2012) The Simple Fool's Guide to population genomics via RNA-Seq: an introduction to high-throughput sequencing data analysis. *Molecular Ecology Resources*, **12**, 1058–1067.

Dionne M, Miller KM, Dodson JJ, Caron F, Bernatchez L (2007) Clinal variation in MHC diversity with temperature: evidence for the role of host–pathogen interaction on local adaptation in Atlantic salmon. *Evolution*, **61**, 2154–2164.

Ebert T (2010) Demographic patterns of the purple sea urchin *Strongylocentrotus purpuratus* along a latitudinal gradient, 1985–1987. *Marine Ecology Progress Series*, **406**, 105–120.

Ebert TA, Russell MP (1988) Latitudinal variation in size structure of the West Coast Purple Sea Urchin: a correlation with headlands. *Limnology and Oceanography*, **33**, 286–294.

Edmands S, Moberg PE, Burton RS (1996) Allozyme and mitochondrial DNA evidence of population subdivision in the purple sea urchin *Strongylocentrotus purpuratus*. *Marine Biology*, **126**, 443–450.

Endler JA (1977) *Geographic Variation, Speciation and Clines*. Princeton University Press, Princeton, New Jersey.

Evans TG, Chan F, Menge BA, Hofmann GE (2013) Transcriptomic responses to ocean acidification in larval sea urchins from a naturally variable pH environment. *Molecular Ecology*, **22**, 1609–1626.

Excoffier L, Hofer T, Foll M (2009) Detecting loci under selection in a hierarchically structured population. *Heredity*, **103**, 285–298.

Fay JC (2011) Weighing the evidence for adaptation at the molecular level. *Trends in Genetics*, **27**, 343–349.

Fay J, Wu C (2000) Hitchhiking under positive Darwinian selection. *Genetics*, **155**, 1405.

Fay JC, Wyckoff GJ, Wu CI (2001) Positive and negative selection on the human genome. *Genetics*, **158**, 1227–1234.

Fay JC, Wyckoff GJ, Wu CI (2002) Testing the neutral theory of molecular evolution with genomic data from Drosophila. *Nature*, **415**, 1024–1026.

Flowers JM, Schroeter SC, Burton RS (2002) The recruitment sweepstakes has many winners: genetic evidence from the sea urchin *Strongylocentrotus purpuratus*. *Evolution*, **56**, 1445–1453.

Foll M, Gaggiotti O (2008) A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. *Genetics*, **180**, 977–993.

Fraser DJ, Weir LK, Bernatchez L, Hansen MM, Taylor EB (2011) Extent and scale of local adaptation in salmonid fishes: review and meta-analysis. *Heredity*, **106**, 404–420.

Fu Y, Li W (1993) Statistical tests of neutrality of mutations. *Genetics*, **133**, 693.

Gilad Y, Oshlack A, Smyth G, Speed T, White K (2006) Expression profiling in primates reveals a rapid evolution of human transcription factors. *Nature*, **440**, 242–245.

Goldberg C, Waits L (2010) Comparative landscape genetics of two pond-breeding amphibian species in a highly modified agricultural landscape. *Molecular Ecology*, **19**, 3650–3663.

Gompel N, Prud'homme B, Wittkopp PJ, Kassner VA, Carroll SB (2005) Chance caught on the wing: cis-regulatory evolution and the origin of pigment patterns in Drosophila. *Nature*, **433**, 481–487.

Grantham R (1974) Amino acid difference formula to help explain protein evolution. *Science*, **185**, 862–864.

Grosberg R, Cunningham C (2001) Genetic structure in the sea. In: *Marine Community Ecology*(eds Bertness MD, Gaines S & Hay ME), pp. 61–84. Sinauer Associates, Sunderland, MA.

Hart MW, Marko PB (2010) It's about time: divergence, demography, and the evolution of developmental modes in marine invertebrates. *Integrative and Comparative Biology*, **50**, 643–661.

Hartl D, Clark A (1997) *Principles of Population Genetics*. Sinauer Associates, Sunderland, MA.

Hedrick P (2006) Genetic polymorphism in heterogeneous environments: the age of genomics. *Annual Review of Ecology, Evolution, and Systematics*, **37**, 67–93.

Hedrick PW, Ginevan ME, Ewing EP (1976) Genetic polymorphism in heterogeneous environments. *Annual Review of Ecology and Systematics*, **7**, 1–32.

Helmuth B, Harley C, Halpin P *et al.* (2002) Climate change and latitudinal patterns of intertidal thermal stress. *Science*, **298**, 1015.

Helmuth B, Broitman B, Blanchette C *et al.* (2006) Mosaic patterns of thermal stress in the rocky intertidal zone: implications for climate change. *Ecological Monographs*, **76**, 461–479.

Hermisson J, Pennings P (2005) Soft sweeps: molecular population genetics of adaptation from standing genetic variation. *Genetics*, **169**, 2335.

Hibino T, Loza-Coll M, Messier C *et al.* (2006) The immune gene repertoire encoded in the purple sea urchin genome. *Developmental Biology*, **300**, 349–365.

Hochachka PW, Somero GN (2002) *Biochemical Adaptation: Mechanism and Process in Physiological Evolution*. Oxford University Press, New York.

Hofmann GE, Smith JE, Johnson KS *et al.* (2011) High-frequency dynamics of ocean pH: a multi-ecosystem comparison. *PLoS One*, **6**, e28983.

Hudson R, Kaplan N (1988) The coalescent process in models with selection and recombination. *Genetics*, **120**, 831.

Huyer A (1983) Coastal upwelling in the California Current system. *Progress in Oceanography*, **12**, 259–284.

Kaplan N, Hudson R, Langley C (1989) The "hitchhiking effect" revisited. *Genetics*, **123**, 887–899.

Karasov T, Messer PW, Petrov DA (2010) Evidence that adaptation in Drosophila is not limited by mutation at single sites. *PLoS Genetics*, **6**, e1000924.

Kayser M, Brauer S, Stoneking M (2003) A genome scan to detect candidate regions influenced by local natural selection in human populations. *Molecular Biology and Evolution*, **20**, 893–900.

Kemppainen P, Lindskog T, Butlin R, Johannesson K (2010) Intron sequences of arginine kinase in an intertidal snail suggest an ecotype-specific selective sweep and a gene duplication. *Heredity*, **106**, 808–816.

Kimura M (1983) *The Neutral Theory of Molecular Evolution*. Cambridge University Press, Cambridge, UK.

Kozyraki R (2001) Cubilin, a multifunctional epithelial receptor: an overview. *Journal of Molecular Medicine*, **79**, 161–167.

Lester S, Gaines S, Kinlan B (2007) Reproduction on the edge: large-scale patterns of individual performance in a marine invertebrate. *Ecology*, **88**, 2229–2239.

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

Li WH, Wu CI, Luo CC (1985) A new method for estimating synonymous and nonsynonymous rates of nucleotide substitution considering the relative likelihood of nucleotide and codon changes. *Molecular Biology and Evolution*, **2**, 150.

Mann KH, Lazier JRN (2006) *Dynamics of Marine Ecosystems: Biological-Physical Interactions in the Oceans*, 3rd edn. Wiley-Blackwell, Nova Scotia, Canada.

Marshall D, Monro K, Bode M, Keough M, Swearer S (2010) Phenotype-environment mismatches reduce connectivity in the sea. *Ecology Letters*, **13**, 128–140.

Maynard Smith J, Haigh J (1974) The hitch-hiking effect of a favourable gene. *Genetical Research*, **23**, 23–35.

Mitarai S, Siegel D, Watson J, Dong C, McWilliams J (2009) Quantifying connectivity in the coastal ocean with application to the Southern California Bight. *Journal of Geophysical Research*, **114**, C10026.

Morse DE, Hooker N, Duncan H, Jensen L (1979) Gamma-aminobutyric acid, a neurotransmitter, induces planktonic abalone larvae to settle and begin metamorphosis. *Science*, **204**, 407–410.

Nachman M (2001) Single nucleotide polymorphisms and recombination rate in humans. *Trends in Genetics*, **17**, 481–485.

Nei M, Gojobori T (1986) Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Molecular Biology and Evolution*, **3**, 418.

Nielsen EE, Hemmer-Hansen J, Larsen PF, Bekkevold D (2009) Population genomics of marine fishes: identifying adaptive variation in space and time. *Molecular Ecology*, **18**, 3128–3150.

Olivares-Banuelos NC, Enriquez-Paredes LM, Ladah LB, De La Rosa-Velez J (2008) Population structure of purple sea urchin *Strongylocentrotus purpuratus* along the Baja California peninsula. *Fisheries Science*, **74**, 804–812.

Oliver T, Garfield D, Manier M *et al.* (2010) Whole-genome positive selection and habitat-driven evolution in a shallow and a deep-sea urchin. *Genome Biology and Evolution*, **2**, 800–814.

Palumbi S (1991) *The Simple Fool's Guide to PCR*. Department of Zoology and Kewalo Marine Laboratory, Honolulu, HI.

Palumbi SR, Wilson AC (1990) Mitochondrial DNA diversity in the Sea Urchins *Strongylocentrotus purpuratus* and *S. droebachiensis*.. *Evolution*, **44**, 403–415.

Peakall R, Smouse P (2006) GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes*, **6**, 288–295.

Pennings PS, Hermisson J (2006) Soft sweeps III: the signature of positive selection from recurrent mutation. *PLoS Genetics*, **2**, e186.

Pespeni MH, Oliver TA, Manier MK, Palumbi SR (2010) Restriction Site Tiling Analysis: accurate discovery and quantitative genotyping of genome-wide polymorphisms using nucleotide arrays. *Genome Biology*, **11**, R44.

Pespeni MH, Garfield DA, Manier MK, Palumbi SR (2012) Genome-wide polymorphisms show unexpected targets of natural selection. *Proceedings of the Royal Society B: Biological Sciences*, **279**, 1412–1420.

Pespeni MH, Barney BT, Palumbi SR (2013a) Differences in the regulation of growth and biomineralization genes revealed through long-term common garden acclimation and experimental genomics in the purple sea urchin. *Evolution*, doi: 10.1111/evo.12036.

Pespeni MH, Sanford E, Gaylord B *et al.* (2013b) Evolutionary change during experimental ocean acidification. *Proceedings of the National Academy of Sciences of the United States of America*, **110**, 6397–6942.

Petit RJ, Hampe A (2006) Some evolutionary consequences of being a tree. *Annual Review of Ecology, Evolution, and Systematics*, **37**, 187–214.

Place AR, Powers DA (1979) Genetic variation and relative catalytic efficiencies: LDH-B allozymes of Fundulus heteroclitus. *Proceedings of the National Academy of Sciences of the United States of America*, **76**, 2354–2358.

Pogson GH (2001) Nucleotide polymorphism and natural selection at the pantophysin (Pan I) locus in the Atlantic cod, Gadus morhua (L.). *Genetics*, **157**, 317–330.

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

R Development Core Team (2009) *R: A Language and Environment for Statistical Computing*. Foundation for Statistical Computing, Vienna, Austria.

Rozas J, Sanchez-DelBarrio J, Messeguer X, Rozas R (2003) DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* **19**, 2496.

Rozen S, Skaletsky H (2000) Primer3 on the WWW for general users and for biologist programmers. *Methods in Molecular Biology*, **132**, 365–386.

Sachidanandam R, Weissman D, Schmidt S *et al.* (2001) A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature*, **409**, 928–933.

Sanford E, Kelly M (2011) Local adaptation in marine invertebrates. *Annual Review of Marine Science*, **3**, 509–535.

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**, 1336–1344.

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

Schultz H (2005) *Sea-urchins, a Guide to Worldwide Shallow Water Species*. Heinke & Peter Schultz Partner Scientific Publications, Germany.

Sea Urchin Genome Sequencing Consortium, Sodergren E, Weinstock GM *et al.* (2006) The genome of the sea urchin *Strongylocentrotus purpuratus*. *Science* **314**, 941–952.

Selkoe K, Watson J, White C *et al.* (2010) Taking the chaos out of genetic patchiness: seascape genetics reveals ecological and oceanographic drivers of genetic patterns in three temperate reef species. *Molecular Ecology*, **19**, 3708.

Smith PE, Eppley RW (1982) Primary production and the anchovy population in the Southern California Bight: comparison of time series. *Limnology and Oceanography*, **27**, 1–17.

Sotka EE, Palumbi SR (2006) The use of genetic clines to estimate dispersal distances of marine larvae. *Ecology*, **87**, 1094–1103.

Sotka EE, Wares JP, Barth JA, Grosberg RK, Palumbi SR (2004) Strong genetic clines and geographical variation in gene flow in the rocky intertidal barnacle *Balanus glandula*. *Molecular Ecology*, **13**, 2143–2156.

Strathmann R (1978) Length of pelagic period in echinoderms with feeding larvae from the Northeast Pacific. *Journal of Experimental Marine Biology and Ecology*, **34**, 23–28.

Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, **123**, 585–596.

Taylor MS, Hellberg ME (2003) Genetic evidence for local retention of pelagic larvae in a Caribbean Reef fish. *Science*, **299**, 107–109.

Terwilliger DP, Buckley KM, Mehta D, Moorjani PG, Smith LC (2006) Unexpected diversity displayed in cDNAs expressed by the immune cells of the purple sea urchin, *Strongylocentrotus purpuratus*. *Physiological Genomics*, **26**, 134.

Thornton KR, Jensen JD (2007) Controlling the false-positive rate in multilocus genome scans for selection. *Genetics*, **175**, 737–750.

Walsh J, Kelley J, Whitledge T, MacIsaac J, Huntsman S (1974) Spin-up of the Baja California upwelling ecosystem. *Limnology and Oceanography*, **19**, 553–572.

Wang DG, Fan JB, Siao CJ *et al.* (1998) Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. *Science*, **280**, 1077–1082.

Wares J (2002) Community genetics in the Northwestern Atlantic intertidal. *Molecular Ecology*, **11**, 1131–1144.

Weersing K, Toonen RJ (2009) Population genetics, larval dispersal, and connectivity in marine systems. *Marine Ecology Progress Series*, **393**, 12.

Wray GA, Hahn MW, Abouheif E *et al.* (2003) The evolution of transcriptional regulation in eukaryotes. *Molecular Biology and Evolution*, **20**, 1377.

Zaytsev O, Cervantes-Duarte R, Montante O, Gallegos-Garcia A (2003) Coastal upwelling activity on the Pacific shelf of the Baja California peninsula. *Journal of Oceanography*, **59**, 489–502.

M.H.P. and S.R.P. designed research, analyzed data, and wrote the paper. M.H.P. performed research.

Data accessibility

The data reported in this study have been deposited in the Dryad Repository, http://dx.doi.org/10.5061/dryad.h32j9.