Multi-model seascape genomics identifies distinct environmental drivers of selection among sympatric marine species

Erica S. Nielsen[1], Romina Henriques[1,3], Maria Beger[2], Robert J. Toonen[4], Sophie von der Heyden[1]

[1]Evolutionary Genomics Group, Department of Botany and Zoology, University of Stellenbosch, Private Bag X1, Matieland, 7602, South Africa
[2]School of Biology, Faculty of Biological Sciences, University of Leeds, LS2 9JT, UK
[3]Technical University of Denmark, National Institute of Aquatic Resources, Section for Marine Living Resources, Velsøvej 39, Silkeborg, 8600, Denmark
[4]Hawai‘i Institute of Marine Biology, University of Hawai‘i at Mānoa, Kāne‘ohe, HI, USA 96744

Corresponding author: Sophie von der Heyden
Email: svdh@sun.ac.za

Abstract

Background: As global change and anthropogenic pressures continue to increase, conservation and management increasingly needs to consider species' potential to adapt to novel environmental conditions. Therefore, it is imperative to characterise the main selective forces acting on ecosystems, and how these may influence the evolutionary potential of populations and species. Using a multi-model seascape genomics approach, we compare the dominant environmental drivers of selection in three sympatric southern African marine invertebrates with contrasting ecology and life histories: Cape urchin (*Parechinus angulosus*), Common shore crab (*Cyclograpsus punctatu*s), and Granular limpet (*Scutellastra granularis*).

Results: Using pooled (Pool-seq), restriction-site associated DNA sequencing (RAD-seq), and seven outlier detection methods, we characterise genomic variation between populations along a strong biogeographical gradient. Of the three species, only *S. granularis* showed significant isolation-by-distance, and isolation-by-environment driven by sea surface temperatures (SST). In contrast, sea surface salinity (SSS) and range in air temperature correlated more strongly with genomic variation in *C. punctatus* and *P. angulosus*. Differences were also found in genomic structuring between the three species, with outlier loci contributing to two clusters in the East and West Coasts for *S. granularis* and *P. angulosus*, but not for *C. punctatus*.

Conclusion: The findings illustrate distinct evolutionary potential across species, suggesting that species-specific habitat requirements and responses to environmental stresses better predict evolutionary patterns than the strong environmental gradients within the region. We also found large discrepancies between outlier detection methodologies, and thus offer a novel multi-model approach to identifying the principal environmental selection forces acting on species. Overall, this work highlights how adding a comparative approach to seascape genomics (both with multiple models and species) can elucidate the intricate evolutionary responses of ecosystems to global change.

Keywords: Pool-seq, RAD-seq, seascape genomics, environmental association, comparative phylogeography, marine invertebrates

Background

Anthropogenic climate change is altering the physical and chemical properties of coastal ecosystems at an unprecedented rate, ultimately threatening the persistence of biological communities 2020/06/10 10:55 AM. Nearshore environments are especially at risk from anthropogenic change as they are exposed to threats from both the terrestrial and marine realms [3]. Coastal systems experience strong environmental gradients, caused by complex interactions among features such as wind and wave action, ocean currents and upwelling cells, and exposure to sunlight and precipitation [4]. Environmental heterogeneity in coastal systems should therefore impose differential selection pressures, facilitating local adaptation and genetic differentiation [5,6]. While many marine species are thought to exhibit low genetic differentiation due to large population sizes and high dispersal potential, there is growing evidence suggesting that many coastal organisms display surprisingly fine-scale population structuring and local adaptation [7–10]. Along with oceanographic patterns and coastal topography, the support for climatic environmental gradients acting as barriers to gene flow is steadily increasing [11–14]. Uncovering patterns of genetic differentiation and possible local adaptation, and distinguishing which environmental conditions shape such patterns, is critical for effective conservation management in the face of global change [15–18].

Quantifying genomic differentiation and putative adaptive variation of marine species, and the resultant field of seascape genomics, relies on recent advances in Next Generation Sequencing (NGS; [19,20]). One of the main goals of seascape genomics is to use NGS to identify loci that differ significantly over environmental gradients, using gene-environment association analyses (GEAAs; [21,22]). GEAAs are powerful tools to detect putative adaptive loci (commonly termed ‘outlier loci’) by directly associating allele frequencies with environmental variables [5,23,24]. Sea surface temperature (SST) is the most common environmental structuring force identified among seascape studies to date [25], and has

been shown to strongly correlate with the genomic variation in abalone [14], mussels [26], oysters [12,27], sea cucumbers [28], and lobsters [11]. As SST is consistently identified as one of the prominent drivers of genomic variation in marine invertebrates, it has promise as a proxy for evolutionary processes, such as local selection, in conservation [29]. However, previous studies have solely investigated single-species GEAAs, which means that the effects of SST and other environmental variables on coastal species with similar distributions, but different micro-environmental niches, are still largely unexplored [1]. Furthermore, there are a multitude of GEA methods available, which differ in their statistical analyses and assumptions of demographic histories, often leading to diverse outputs [30,31]. Even though many studies use two or three outlier detection methods to account for false positives [23], there has yet to be a comprehensive comparison of various methods in their ability to identify the dominant selection forces acting on wild marine populations.

This study focusses on the environmental drivers of genomic differentiation in three rocky shore invertebrates: Cape urchin (*Parechinus angulosus*), Granular limpet (*Scutellastra granularis*), and Common shore crab (*Cyclograpsus punctatus*), that are widely distributed along the southern African coastline, which is known for its strong biogeographic gradients of temperature, productivity and other environmental variables (Fig. 1; [32]). Previous studies, consisting of mitochondrial DNA (mtDNA) data, have detected multiple lineages for each species, broadly differentiated into West and East Coast clades [33–35], with evidence of isolation-by-distance (IBD; e.g. [35]). However, a recent study using NGS data from the estuarine-restricted seagrass *Z. capensis* suggested that isolation-by-environment (IBE) plays a significant role in shaping the genomic differentiation [36], although the extent that IBD and IBE characterise the genomic variation of other marine species in the region currently remains unknown.

Broadly, the objectives of this study are to characterise phylogeographic patterns of three ecologically important rocky shore species, and to identify the dominant environmental drivers of putative adaptive variation within southern African rocky shore communities. A pooled (Pool-seq), restriction-site associated DNA sequencing (RAD-seq) approach was

used to characterise genomic variation across at least 13 sites per species, and describe population diversity within and differentiation amongst these species. Seven outlier detection methods were also used to distinguish the principal environmental drivers of selection in each species. We hypothesized that: 1) each species will be structured into West and East Coast populations in accordance with mtDNA population structure, 2) each species will show significant isolation-by-distance and isolation-by-environment, 3) SST will be the dominant driver of genomic variation for all three species.

Results

Sequencing and bioinformatic processing

DNA samples were collected from three species, including ~ 40 individuals from 13 or 14 sites each, per species (Fig. 1; Tables S1-S3, Additional File 1). A pooled ezRAD sequencing and *de novo* assembly approach was used, as this allowed for larger contig lengths (e.g. > 1000 base pairs; bp) compared to other RAD-seq approaches [37]. Various bioinformatic steps (see Methods section and Table S4, Additional File 1 for further details) were then used to obtain single nucleotide polymorphisms (SNPs), from which various comparative phylogeography analyses were performed.

To assure retained SNPs best reflect nuclear genome-wide variation, we first removed possible mtDNA reads, as well as compared the performance of three *de novo* assemblers. The average number of reads per sample that mapped onto the reference mitogenome was 12 363 for *C. punctatus*, 20 342 for *P. angulosus*, and 234 for *S. granularis* (Table S5, Additional File 1). These mitochondrial reads were subsequently removed from the raw reads during the mapping stage, as they reflect distinct evolutionary processes compared to nuclear loci [38]. As there are no reference genomes for these or closely-related species, *de novo* assemblies were compared between three programs, SPAdes,

AbySS, and MEGAHIT, for each species. SPAdes resulted in the more robust assembly, with the longest contig length, N50, and L50, as well as a higher number of BUSCO and NCBI hits on average for all three species (Table S6, Additional File 1), and thus was used for all downstream analyses.

The number of raw reads per species ranged from ~29 million for *C. punctatus* to ~47 million for *P. angulosus* (Table S7, Additional File 1). The average number of raw reads per pool was ~2.2 million for *C. punctatus*, ~2.5 million for *S. granularis*, and ~3.5 million for *P. angulosus* (Tables S8-10, Additional File 1). A total of 17 309, 3 946, and 10 416 SNPs were identified for each species, respectively (Table S7, Additional File 1). After pruning the SNP datasets to one SNP per 1 000bp (to account for linkage disequilibrium; LD), *C. punctatus*, *P. angulosus*, and *S. granularis* had 1 190, 822, and 1 658 SNPs, respectively (Table S7, Additional File 1).

Genomic structuring

To assess population structuring, all filtered and LD-pruned SNPs were used to calculate pairwise F_{ST} values and Nei's genetic distances. Further, Pool-seq specific scaled covariance (Ω) matrices were calculated, which can be interpreted as pairwise estimates of genomic differentiation and population structure. Isolation-by-distance (IBD) patterns were assessed by comparing genomic and geographic distance per species.

Pairwise F_{ST} values varied between species; *C. punctatus* F_{ST} = 0 - 0.021, *P. angulosus* F_{ST} = 0 - 0.127, and *S. granularis* F_{ST} = 0 - F_{ST} = 0.059 (Additional File 2). The PCoAs from Nei's genetic distance and the Ω matrices show no clear spatial clustering for *C. punctatus* and *P. angulosus*, but slight differentiation between West and East Coast sites for *S. granularis* (Fig. 2). Mantel tests suggest that of the three species, only *S. granularis* populations are characterised by IBD (r=0.48, p<0.01; Table 1).

Potential environmental drivers of genomic structuring

To assess possible environmental drivers of genomic structuring, we ran isolation-by-environment (IBE) tests, which compare genomic and environmental distance, accounting for geographic distance. To identify environmental variables for the IBE and GEA analyses, a total of 20 environmental variables were originally included, and subsequently filtered based on Spearman's correlation coefficients < 0.65 and variance inflation factors <10. There were multiple correlations between the 20 environmental predictor variables (Additional File 3). After filtering for collinearity, five final environmental predictor variables remained: mean sea surface salinity (SSSmean), sea surface salinity range (SSSrange), mean sea surface temperature (SSTmean), sea surface temperature range (SSTrange) and air temperature range (Trange; Additional File 3).

Partial Mantel tests showed significant IBE by SSTmean for *C. punctatus* ($r=0.43$, $p<0.05$), but this did not remain significant after correcting for multiple testing ($q=0.19$; Table 1). SSTmean was also found to significantly correlate with genomic differentiation in *S. granularis* ($r=0.40$, $p<0.01$), which remained significant after multiple testing correction ($q=0.001$; Table 1). The partial Mantel tests did not find a significant correlation between any of the three environmental predictor variables and genomic differentiation in *P. angulosus* (Table 1).

Characterising selective forces via outlier loci identification

As GEAA methods have been shown to vary in the type and number of outliers detected [23,30], seven different outlier-detection methods were compared, including six GEAAs to investigate possible associations between SNPs and environmental variables. The analyses included BayPass Bayesian hierarchical models (both core and auxiliary models), Latent factor mixed models (LFMM), Moran spectral outlier detection (MSOD) and Moran spectral

randomization outlier detection (MSR), and Redundancy analyses (RDA) and Distance-based redundancy analyses (dbRDA).

Overall, there was a large range in the number of outliers detected, with little overlap between models (Table 2). LFMM detected the most outliers and had the highest number of unique outliers, followed by MSOD (Table 2). Generally, *S. granularis* had the highest number of outliers detected for each model, with the exception of LFMM (Table 2). The model type with the lowest number of outliers selected was dbRDA (Table 2). For the dbRDA analyses, a forward selection process retained zero dbMEMs for *C. punctatus* and *P. angulosus,* and one dbMEM for *S. granularis*. The dbRDA for *S. granularis* had an adjusted R^2 value of 0.02 (p=0.33), with one outlier locus selected. The standard RDAs had adjusted R^2 values of 0.021 (p= 0.31), 0.021 (p=0.65), and 0.084 (p=0.01) for *C. punctatus*, *P. angulosus*, and *S. granularis*, respectively. The single population-differentiation based outlier detection method, BayPass core model (BPC), identified nine outliers for *C. punctatus*, five outliers in *P. angulosus*, and 19 in *S. granularis*, with two, two, and eight outliers unique to that method, respectively (Table 2).

The environmental variable that most strongly correlates with genomic variation differed between outlier detection methods and across species. The majority of methods for *C. punctatus* identified the most outlier loci in association with SSSmean, with the exception of LFMM that identified the most outliers with Trange (Fig. 3). Trange and SSTmean were the two variables that identified outliers in at least three models for *P. angulosus* (Fig. 3). SSTmean identified the most outlier loci in all methods except LFMM for *S. granularis* (Fig. 3).

Spatial structure of neutral vs. outlier SNPs

Loci that were selected by two or more outlier detection methods (2X outliers) were used to create an ‘outlier dataset’, and these loci were removed from the total SNP dataset to create

a putatively 'neutral dataset'. We compared the genomic structuring between putative neutral and outlier SNPs via Principal Components Analysis (PCA) ordinations of allele frequencies from each dataset.

The number of SNPs used to create 'outlier' datasets was 13, 12, and 26 for *C. punctatus*, *P. angulosus*, and *S. granularis*, respectively. The PCAs of allele frequencies differed between the neutral and outlier SNP datasets for all three species (Fig. 4). For *C. punctatus*, the putatively neutral SNPs show most of the sites within one main cluster, with the YZ and MG sample sites each forming individual clusters. In contrast, the outliers show more differentiation between sites, with MG and YZ as most divergent. The neutral SNPs of *P. angulosus* do not separate sites following any geographical pattern, however the outlier SNPs clearly distinguish between the East and West Coast sites (Fig 5). In *S. granularis,* the neutral dataset separates East Coasts and West Coast sites, a pattern even more pronounced when examining the outlier dataset, where sampling sites are clearly differentiated according to geography (Fig 5).

Potential functionality of outlier SNPs

We investigated the potential functional roles of the outlier SNPs selected by three or more detection methods (3X outliers), by BLASTing them to the NCBI non-redundant protein sequence database, and assessing gene ontology (GO) with Blast2GO.

Of the 3X outliers, none of the contigs resulted in quality BLAST hits for *C. punctatus*, but three contigs from *P. angulosus* and four contigs from *S. granularis* resulted in BLAST results passing quality filters (Additional File 4). All three of the contigs from *P. angulosus*, and one contig from *S. granularis* matched to histone complexes, with GO terms relating to DNA-binding, protein heterodimerization activity, and regulation of DNA recombination and chromatin silencing (Additional File 4). Of the remaining three contigs for *S. granularis*, two BLASTed to hypothetical proteins, and thus had no GO terms, and one BLASTed to a

protein kinase-like protein, with GO terms relating to regulation of transcription and sequence-specific DNA binding (Additional File 4).

Discussion

This study utilised high-throughput genomic data to elucidate phylogeographic patterns of three southern African coastal marine invertebrates. We hypothesized that the study species would each demonstrate West and East Coast genomic structuring as well as isolation-by-distance (IBD) and isolation-by-environment (IBE). However, these hypotheses were only confirmed in the limpet, *S. granularis*. We also used a multispecies approach to explore environmental drivers of genomic variation within this unique marine biogeographical gradient. Here we hypothesized that sea-surface temperature (SST) would be the dominant driver of genomic variation, and yet again this hypothesis was rejected in all species except for *S. granularis*. Overall, the results reveal species-specific evolutionary patterns, highlighting the complexity of interacting factors shaping natural genomic variation, which is discussed in detail in the following sections.

Genomic markers elucidate distinct patterns of population structuring

Our first hypothesis was that each species would reflect previously described mtDNA patterns, with two clusters separated into West and East Coast individuals, reflecting the biogeographic breakpoint around the Southwestern Cape. However, only the limpet *S. granularis* follows this pattern, with *C. punctatus* showing high connectivity between populations, and *P. angulosus* showing no clear population structuring (Fig. 2; Additional File 2). The discordance in genomic differentiation found between mtDNA datasets in previous studies and the SNPs datasets here could be owing to the differences between the two marker types, as mtDNA markers are comprised of a single maternally inherited locus, while

SNP markers represent a broad range of loci across the nuclear genome [38]. Additionally, mtDNA markers are expected to reflect relatively historical evolutionary events compared to the more contemporary processes captured by microsatellite and SNP markers [39].

Scutellastra granularis was also the only species which supported our hypothesis of IBD and IBE influencing genomic structure. This pattern of IBD and IBE in *S. granularis* could also result from repeated founder effects and allele surfing, caused by colonization generating an allele frequency gradient which co-varies with the environmental gradient [40,41]. However, *S. granularis* and *P. angulosus* were shown to have similar evolutionary histories [33,34], and thus it seems more likely that contemporary environmental, rather than historical demographic, processes are leading to the distinct patterns found in *S. granularis*.

Even though *S. granularis* is the only species to show distinct West/East Coast structuring in all SNPs, both *P. angulosus* and *S. granularis* show strong West and East Coast clustering when using only outlier SNPs (Fig. 4). The West and East Coast bioregions exhibit profound differences in not only temperature, but other environmental variables such as primary productivity [42], which can potentially lead to local selection despite high levels of connectivity [43]. This finding builds on multiple other studies which have found outlier SNPs showing fine-scale genomic structuring in populations characterised by high genetic connectivity, yet situated within strong environmental gradients [14,44,45], and suggests that environmental variation along coastal South Africa plays an important role in the evolutionary dynamics of species in the region.

In contrast to the other two species, the crab *C. punctatus*, did not show a strong separation between the West and East Coast sites. Instead, two range-edge sampling locations (YZ and MG) are highly differentiated in both the neutral and outlier SNP datasets (Fig. 4). It could be that this species is less affected by large-scale environmental gradients of the coastline, but rather that an edge effect driven by historical demographic processes such as range contractions and expansions are underlying this pattern [46]. The distinct distribution of genomic variation of *C. punctatus* could also result from it being the most generalist of the three species, inhabiting both estuarine and marine environments [47,48]. A

previous study by [49] found that among 10 rocky intertidal invertebrates, the ability to utilize sheltered habitat was the strongest predictor of genetic structure. Of the three species here, *S. granularis* is the most restricted in its habitat, compared to *P. angulosus* and *C. punctatus* which are able to shelter under rocks and macroaglae, rather than remain exposed on rocky surfaces [48]. Broadly, the results show that while each species exhibits high levels of gene flow, there are fine-scale patterns of genomic differentiation which appear to vary based on the ecology of the species.

Identifying drivers of selection using a multi-model approach

The number of outlier loci, as well as the environmental variable most strongly associated with outlier loci, varied across GEAA methodologies (Table 2; Fig. 3), which mirrors previous studies describing differences in outlier detection methods (e.g., [22,30,31]). Each GEAA accounts for past demographic histories differently, leading to differences in outliers detected [31,50], and thus it has been suggested to use multiple models in combination when the principal environmental variables are unknown [23]. At present, most studies use one or two outlier detection methods, and identify the most important environmental drivers of selection based on which variable identifies the most outlier SNPs [23]. However, our results, as well as those from [30], indicate that the number of SNPs identified as outliers varies greatly among detection methods. We argue that the importance of environmental variables should not be measured by the total number of outliers it detects, but rather by the number of GEAA approaches in which the variable identifies outliers. For example, even though Trange identified the most outliers for *C. punctatus*, we argue that the most important environmental driver of genetic differentiation for *C. punctatus* is SSS, because it identified outliers by three detection methods rather than only for Trange (Fig. 3). Given that outlier detection methods are highly variable and subject to false positives [31], we believe that using multi-model approaches will increase the robustness of GEAAs, especially in studies

identifying potential drivers of selection across species with varying evolutionary histories. Hence, in the following section, we discuss the dominant environmental drivers identified for each species based on the number of GEAA models in which outliers were identified.

Different environmental drivers of selection across species

Of the three species, only *S. granularis* supported our hypothesis of SST being most important environmental predictor of genomic variation (Fig. 3). Previous seascape genomic studies in temperate regions have frequently identified some measure of SST as the best predictor of genomic variation of marine invertebrates [9,11,12,25,28,51], which is most likely due to SST affecting both cellular processes, and life-history events such as spawning and larval development [52]. However, for *P. angulosus,* Trange and SSTmean best explained genomic variation, whereas SSSmean best explained the structure of *C. punctatus*. Salinity emerging as a major selective force on *C. punctatus* is understandable, as this species is an osmoconformer that inhabits estuarine environments [53], and because the larval development of decapods is influenced by changes in salinity [54]. The lack of clear correlations with any environmental variables is unexpected for the urchin *P. angulosus*, given that previous studies have shown genomic variation corresponding to SST gradients in other echinoderms [28,55]. Additionally, the paucity of annotated genomes for marine invertebrates makes it difficult to identify the functionality of outlier loci, which likely led to the limited number of BLAST hits for the 3X outliers in each species (Additional File 4). Despite this limitation, outliers from both *S. granularis* and *P. angulosus* BLASTed to histone variants, which could indicate responses to environmental pressures [56].

Previous terrestrial comparative GEAA studies have found distinct results in co-occurring species, which the authors attribute to either differences in ecological niche ranges [57] or phenotypic plasticity [41]. A multitude of factors could be driving the interspecific differences observed here, as the study species not only inhabit different ecological niches,

but also exhibit differential behaviours to remain within their physiological niches [48]. It is also likely that the study species exhibit phenotypic plasticity in response to environmental pressures, as plasticity and epigenetic effects have been noted in response to temperature and salinity at multiple life stages in marine invertebrates [58]. Additionally, the rocky shore is a highly variable environment, and it is likely that species within different zonations are under differential selection pressures at fine spatial scales [59,60], which might interact with large-scale environmental gradients to create complex patterns of genomic variation.

Conclusions

The results of our comparative genomic study suggest that environmental drivers, and the impacts from their future change, may be highly species-specific, even among co-occurring species living within regions of strong environmental gradients. Further, the results contrast many single-species marine GEAA studies by showing that SST does not consistently emerge as an important environmental force structuring the distribution of genomic variation in marine organisms. This finding brings into question the use of SST clines as simple surrogates for selection in marine conservation spatial planning with regards to global change. Yet the results here provide exciting opportunities to investigate the relationships between ecological or behavioural traits and environmental drivers of selection across species.

This is one of the first comparative seascape genomic studies to date, and it is imperative that future seascape genomic studies aim to understand how climatic change will impact not just individual species, but communities [61]. Multispecies GEAA studies remain a challenge due to costs associated with high-throughput sequencing and the lack of annotated genomes in non-model species, particularly marine invertebrates [62,63]. Here we used a pooled RAD-seq approach, which allowed us to conduct a multispecies comparative GEAA study with relatively low costs, albeit with some limitations such as reduced genomic

representation and the identification of individuals and polygenic scores being unavailable [64,65]. However, for our research question, a pooled RAD-seq approach is beneficial as it allowed us to maximize the number of individuals per location to obtain accurate population allele frequency estimates [66,67], as well as maximize the number of sample sites, both of which are essential for GEAAs [50,68], without the full cost of sequencing every individual.

Finally, we also provide a novel approach to identify drivers of selection across a diverse array of species, by using multiple GEAA methods and inferring the importance of each variable across methods. Ultimately, we argue that future seascape genomics studies can benefit from widening their scope with species and model comparisons, to more robustly identify environmental drivers of selection.

Methods

Study region and species

The study domain lies along the South African coastline, which is one of the most biodiverse marine systems in the world [69]. This region has also been identified as hotspot for ocean warming as it is experiencing environmental change at a faster rate than predicted [70]. In South Africa, the coastline is characterised by SST increasing with longitude, from the cool-temperate Benguela region on the West Coast to the sup-tropical Dalgoa region on the East Coast (Fig. 1).

The study species were selected as their distributions span several bioregions and the natural environmental gradients of southern Africa (Fig. 1), and can represent the high- (*C. punctatus*), mid- (*S. granularis*), and low- (*P. angulosus*) rocky shore ecotypes [48]. They also differ in life history traits with *C. punctatus* being a brooder, and *S. granularis* and *P. angulosus* being broadcast spawners, with pelagic larval durations varying from ~ 5-15 days (*S. granularis* and *C. punctatus*) to potentially up to 50 days (*P. angulosus*; [34,35,71]).

These species are each ecologically important; either as dominant grazers or scavengers, as substrates for other species to either live on, or as shelter for juvenile abalone [48].

A total of 14 sites, spanning ~ 2 200 km of the South African coastline, were sampled for *S. granularis* and *P. angulosus*, and 13 sites spanning ~1 800 km were sampled for *C. punctatus* (Fig. 1). These sites incorporate the natural environmental (e.g. SST, salinity, air temperature) gradients in the region, as well as the distributional range per study species [47].

Laboratory protocols and bioinformatics

Genomic data consisted of pooled ezRAD-seq samples, as it is a cost-effective approach to obtain precise allele frequency data [67]. Dorant et al. (2019; [66]) found that Pool-seq inflated F_{ST} values relative to individual-based sequencing approaches, but still gave highly similar allele frequency outputs and patterns of population structure. Thus, while the absolute magnitude of F_{ST} values may be upwardly biased relative to sequencing individuals, for a fraction of the cost Pool-seq data still allow us to infer relative patterns of population structure with confidence [72].

Genomic RAD-seq data was previously obtained for the study species from 11 of the 20 sample sites ([73]; L. Mertens pers. comm.). Additional sampling was conducted at the remaining sites during July 2018, with 30-40 individuals collected from each site (Tables S1-S3, Additional File 1). Individuals were preserved in 100% ethanol, from which < 25mg tissue (gonad from *P. angulosus*, foot from *S. granularis*, and muscle from *C. punctatus*) was taken for DNA extractions. Extractions were performed with the Qiagen DNeasy Blood & Tissue kit following the manufacturer's protocols. The quality of the DNA extractions was assessed on 1% agarose gels and quantity was determined using the Qubit Quant iT dsDNA HS Assay system at the Central Analytical Facility at Stellenbosch University (CAF-SU). All extractions passing quality and quantity checks were stored at -20^{0}C. For each species,

equimolar amounts of DNA from each individual were pooled per sample site, flash frozen and sent to the Hawaii Institute of Marine Biology (HIMB) for library preparation following [74] (further outlined in [73]). Equimolar pooled ezRAD libraries [37] were sequenced (V3, 2x300PE) on the Illumina Mi-Seq platform at University of California, Riverside.

The quality of raw FASTA reads were viewed with FastQC [75], and then uploaded onto the CAF-SU high performance cluster (HPC) for further analyses (see Table S4, Additional File 1 for outline of analyses). Bases with low quality scores (Q<20), overrepresented sequences and adapter sequences were removed using TrimGalore! [76].

As mitochondrial DNA (mtDNA) markers have different evolutionary characteristics than nuclear markers [39,77], we chose to filter mtDNA-mapped reads from the datasets [45]. In order to separate mtDNA from nuclear sequences, the quality-trimmed reads were first mapped onto mitogenome references of closely related species, using BWA-MEM ([78]; Table S5, Additional File 1). The mapped reads were converted to BAM files, sorted and filtered using SAMtools v.1.3 [79], and then merged using BAMtools [80]. The merged BAM files were converted back to SAM and used to filter the quality-trimmed reads, removing putative mtDNA markers before mapping, using the 'filterbyname' command in BBMap [81].

Given that there are no reference genomes for these or closely-related species, *de novo* assemblies were created, using quality-trimmed reads that were normalized to a coverage of 100X with BBMap 'bbnorm', and using k-mer value ranges identified with K-mer Genie [82]. The reads were assembled with three different programs: ABySS [83], MEGAHIT [84] and SPAdes [85]. Because SPAdes can only handle nine input samples at a time, we assembled half of each species' samples at a time, and then merged the two SPAdes assemblies using GARM [86]. The outputs of the three assemblers were compared using QUAST v.4.1.1 [87], BUSCO v3.0 [88], and the NCBI BLASTN v.2.4.0+ algorithm [89]. Metrics such as N50 and L50 values, and number of BLAST and BUSCO hits, were used to select a *de novo* assembly for further analysis.

The mtDNA-free, but not normalized, reads were mapped onto the *de novo* assemblies with BWA-MEM. The subsequent SAM files were converted into BAM files,

sorted, indexed and filtered with SAMtools. To control for sequencing biases, we down-sampled SAM files to the median number of reads across all pools with SAMtools. A synchronized multiple pileup file was created for each species with SAMtools 'mpileup', followed by the Popoolation2 'mpileup2sync.jar' commands [90]. Final SNP calling was performed with the 'popsync2pooldata' function of the *poolfstat* v.0.0.1 R package [91], with a minimum coverage of > 20, and a minor allele frequency (MAF) of > 0.01 in each pool [66,92]. To account for the possibility of loci being physically linked (linkage disequilibrium: LD) we further used custom R scripts to randomly select one SNP per 1 000bp per contig.

Assessing gene flow and potential drivers of population structuring

Characterising genomic differentiation

To assess genomic population structuring, pairwise F_{ST} values from the LD-pruned SNP dataset were calculated using the 'computeFST' function of *poolfstat*, the confidence interval (CI) values of which were computed with a custom bash script from [66] using 1 000 bootstrap iterations. Nei's genetic distances matrices were generated with the 'stamppNeisD' function of the R package *StAMPP*, and visualized in Principal Coordinates Analyses (PCoAs) generated with the 'pco' function in the *ecodist* R package [93].

Additionally, the allele frequencies of all SNPs per species were input into the core model of BayPass v2.1 [94] to estimate scaled covariance (Ω) matrices. BayPass is specifically designed to handle Pool-seq data, and uses allele-frequencies to create an Ω matrix, which can be interpreted as pairwise estimates of differentiation and population structure. BayPass was run under default conditions to create the Ω matrices, which were then converted into a correlation matrices using the 'cov2cor' function in R *stats* package, and visualized as similarity matrix heatmaps.

Seascape features

The various seascape genomic analyses included a standard set of environmental features as predictor variables. A total of 20 environmental features were considered (Additional File 3), including air temperature and precipitation of the coldest month, warmest month, the range between coldest and warmest months, as well as annual mean between 1970 and 2000, which were downloaded from the WorldClim database at a resolution of ~1 km [95]. Annual mean, coldest ice-free month, and warmest ice-free month, and the range in SST between 2002 and 2010 and annual mean, monthly minimum and maximum, and range in sea surface salinity between 1955 and 2006 were downloaded from the MARSPEC database at a resolution of ~1 km [96]. Mean surface dissolved oxygen, diffuse attenuation coefficient, pH, and chlorophyll concentration between 2000 and 2014 were downloaded from the BIO-ORACLE database at a resolution of ~9.2 km [97]. Environmental features were downloaded for each sample site with the 'load_layers' function of the *sdmpredictors* R package [98]. We tested collinearity between predictor variables using pairwise Spearman's correlation coefficients and Benjamini-Hochberg (BH) corrected p-values ($p < 0.05$; [99]). We removed variables that were significantly correlated ($r>0.65$), and those with a variance inflation factor (VIF) >10.

Isolation-by-distance (IBD) versus isolation-by-environment (IBE)

Isolation-by-distance (IBD) and isolation-by-environment (IBE) were tested using Mantel tests. Mantel tests are widely used in landscape genetics to test which spatial features are significant drivers of genetic differentiation [100]. IBD was assessed with a standard Mantel test, which evaluates the relationship between two matrices (i.e. geographic versus genetic distances) and IBE was tested with Partial Mantel tests, which compare the relationship

between two matrices while taking into account the effect of a third (i.e. temperature versus genetic distance, accounting for geographic distance; [100]).

IBD analyses consisted of Slatkin's linearized pairwise F_{ST} (F_{ST} = [F_{ST} /(1-F_{ST})]; [101]), and log-transformed geographic distances along the coastline calculated with the roadmap tool in QGIS [102], starting from the western-most site for each species. IBE analyses additionally included pairwise Euclidean climatic distances. Partial Mantel tests were performed for each climatic variable separately, with geographic distance as a conditioning variable. Individual Mantel test significance was assessed in *ecodist*, using 1,000 permutations. To account for multiple tests, *p*- were converted to *q-values* and significance was assessed using a False Discovery Rate of 0.05 (FDR) based on BH criteria with the *qvalue* R package [103].

A multi-model approach to identifying environmental associations with SNPs

To investigate possible associations between SNPs and environmental variables, we used seven different outlier detection methods, using the same seascape features as stated above as predictor variables. As GEAA methods have been shown to vary in the type and number of outliers detected [23,30], the multi-model approach used here allows for more robust inferences. The protocol pertaining to each outlier detection method are outlined below.

BayPass Bayesian hierarchical models

For an F_{ST}-like outlier detection approach, the core model of BayPass was run, which uses a hierarchical Bayesian model to create per-locus *XTX* values, which can be interpreted as an F_{ST} values corrected for the scaled covariance (Ω) of population allele frequencies [94]. BayPass v.2.1 was run under default conditions to create *XTX* values. As described in [94],

a pseudo-observed dataset (POD) was created to estimate the posterior predictive distribution of *XTX* values, and candidate SNPs were selected if they fell within the 99.9% quantile of the POD *XTX* distribution.

For a GEAA-like approach, the auxiliary model in BayPass was run to identify candidate SNPs due to associations with environmental variables. The auxiliary covariate model includes a binary auxiliary variable to classify the association and compute a Bayes factor (BF) for each locus while accounting for multiple testing [94]. After running the model under default conditions, we followed the general rule derived from [104], which identifies outliers as those having a log10 Bayes factor (db) >20 [94].

Latent factor mixed models (LFMM)

Latent factor mixed models (LFMM) use mixed linear models to test for correlation between allele frequencies and an environmental predictor variable while correcting for population structure with latent factors [105]. As such, these models require *a priori* knowledge of the number of genetic clusters (K). K was inferred from previous mtDNA clustering analyses (K=2 for each species; [33–35]), as it is recommended to estimate K from independent genetic datasets [105]. LFMMs were run separately for each environmental variable using the R package *LEA* [106] with 10 000 cycles of the Gibbs sampling algorithm, 5 000 burn-in cycles, and 10 replicate runs. For all runs per predictor variable, z-scores were combined, genomic inflation factor was calculated, and candidate loci were selected following using R scripts available from: http://membres-timc.imag.fr/Olivier.Francois/LEA.

Moran spectral outlier detection (MSOD) & Moran spectral randomization (MSR)

Moran spectral outlier detection (MSOD) uses Moran's eigenvector maps (MEMs) to create power spectrums for each individual SNP, by taking the squared correlation coefficient of

allele frequencies with MEM eigenvectors [107]. Candidate SNPs are then identified as having power spectra outside of the average spectrum across all SNPs. Moran spectral randomization (MSR), is then used to identify candidate SNPs that show a strong correlation to environmental variables by building the observed spatial structure into the null model, while accounting for spatial autocorrelation [107].

MEM axes were first created from geographic coordinates using the *spdep* R package [108], then power spectra corresponding MAFs and MEMs at each site were calculated. Z-scores were calculated for each locus based on the deviation from the average power spectrum following R code from: https://popgen.nescent.org/2016-12-13_MEM_outlier.html. The outlier loci identified by MSOD were then subjected to the MSR randomization approach, which tests the correlation between outlier MAFs and environmental variables, given the power spectra of each SNP. Using the *adespatial* R package, the MSR was run individually for each environmental variable, with 1 000 permutations. We followed the suggested cut-offs of [107] of 0.01 and 0.05 for MSOD and MSR candidates, respectively.

Redundancy analysis (RDA)

Redundancy analyses (RDAs) are an extension of linear regressions that compare a matrix of dependent variables with multiple independent predictor variables. Linear regressions are calculated between allele frequencies and the climate variables at each site, while the fitted values are simultaneously constrained using a PCA. Environmental variables were centred and scaled, and allele frequencies were Hellinger transformed [109]. All RDAs were performed with the 'rda' function of the *vegan* R package [110]. Significance was assessed from the adjusted R^2 value and with an ANOVA following 1 000 permutations. Candidate loci were those that had loading scores ±3 Standard Deviations (SD) of the mean loading for each of the first two constrained axes [28,30].

Distance-based RDAs (dbRDAs) were also run to account for autocorrelation between environmental and geographic distance. Distance-based Moran's eigenvector maps (dbMEMS), which decompose Euclidean distances into a set of spatial variables [111], were created with the R package *adespatial* [112]. Significant dbMEMs were selected by first running an RDA solely using the dbMEMs as predictor variables, then using the adjusted R^2 value from that RDA as the threshold for the forward selection procedure with the 'forward.sel' function in the *packfor* R package [113].

Outlier variation and functional annotation

Loci that were selected by two or more detection methods (2X outliers) were used to create a statistical 'outlier dataset', and these loci were removed from the total SNP dataset to create a 'neutral dataset'. Intraspecific outlier and neutral variation was compared by running PCA ordinations on the MAFs of each dataset with the *vegan* package, and plotting the ordinations with the *ggplot2* package in Rstudio [114].

Furthermore, we investigated the potential functional roles of outlier SNPs selected by three or more detection methods (3X outliers). The contigs containing the 3X outliers were BLASTed against NCBI non-redundant protein sequence database for arthropods (for *C. punctatus*), molluscs (for *S. granularis*), and sea urchins (for *P. angulosus*) using Blast2GO [115]. Search results were filtered to only include those which had an E-value less than 10^{-4}, and a minimal alignment length of 20 bp. Gene Ontology (GO) mapping and annotation was conducted on BLAST searches passing quality filters, using default parameters in Blast2GO.

List of abbreviations

2X outliers: outliers selected by two or more outlier-detection methods

3X outliers: outliers selected by three or more outlier-detection methods

BF: Bayes factor

BH: Benjamini-Hochberg

BPA: BayPass auxiliary model

BPC: BayPass core model

CAF-SU: Central Analytical Facility at Stellenbosch University

CI: confidence interval

dbRDA: distance-based redundancy analysis

GEAA: gene-environment association analysis

GO: gene ontology

HPC: high performance cluster

IBD: isolation-by-distance

IBE: isolation-by-environment

LD: linkage disequilibrium

LFMM: latent factor mixed models

MAF: minor allele frequency

MEM: Moran's eigenvector map

MSOD: Moran spectral outlier detection

mtDNA: mitochondrial DNA

NGS: next generation sequencing

PCA: principal components analysis

PCoA: principal coordinates analysis

POD: pseudo-observed dataset

Pool-seq: pooled DNA sequencing

RAD-seq: restriction site associated DNA sequencing

RDA: redundancy analysis

SNP: single nucleotide polymorphism

SSS: sea surface salinity

SSSmean: mean sea surface salinity

SSSrange: range in sea surface salinity

SST: sea surface temperature

SSTmean: mean sea surface temperature

SSTrange: range in sea surface temperature

Trange: range in air temperature

VIF: variance inflation factor

Declarations

Ethics approval and consent to participate: Not applicable.

Consent for publication: Not applicable.

Availability of data and materials: Raw reads accessible via Genbank accessions: As our campus is currently closed due to COVID-19, we will only be able to upload the raw sequences once we will have access to our lab again. This should be before the end of June. Allele frequencies and R scripts accessible via GitHub: https://github.com/vonderHeydenLab/Nielsen_et_al_2020_BMC_Evol_Biol

Competing interests: The authors declare no competing interests.

Funding: The running costs for field, laboratory and analytical work were funded by grants to S.V.D.H. (National Research Foundation Grant numbers 92788, 105949).

Author contributions: E.S.N., S.V.D.H., M.B., and R.H. conceptualized the study. R.T. assisted with data generation. E.S.N. collected and analysed data. E.S.N. led the writing, to which all other authors contributed. All authors read and approved the manuscript.

Acknowledgements: We would like to thank the following for assisting in data collection: A. Steele, M. Czachur, C. Waspe, A. Shurtey, N. and D. Phair. A special thank you to Lisa Mertens for collecting specimens and extracting DNA from the following locations: Cape Agulhas (CA), Knysna (KY), Cape St. Francis (CF), and Haga Haga (HH). We thank A. Vorster for Qubit assistance, and C. Van Der Vyver for freeze-drying assistance. We would like to especially thank Ingrid Knapp for her assistance with library preparation. Further thanks to G. Van Wageningen for his assistance with the HPC, and A. Le Moan for his code to LD-prune the SNP datasets.

References

1. Harvey BP, Gwynn-Jones D, Moore PJ. Meta-analysis reveals complex marine biological responses to the interactive effects of ocean acidification and warming. Ecol Evol. 2013;3(4):1016–30.

2. Pecl GT, Ward TM, Doubleday ZA, Clarke S, Day J, Dixon C, et al. Rapid assessment of fisheries species sensitivity to climate change. Clim Change. 2014;127(3–4):505–20.

3. Glavovic B, Limburg K, Liu K-K, Emeis K-C, Thomas H, Kremer H, et al. Living on the Margin in the Anthropocene: engagement arenas for sustainability research and action at the ocean–land interface. Curr Opin Environ Sust. 2015;14:232–8.

4. Helmuth B, Mieszkowska N, Moore P, Hawkins SJ. Living on the Edge of Two Changing Worlds: Forecasting the Responses of Rocky Intertidal Ecosystems to Climate Change. Annu Rev Ecol Evol Syst. 2006;37(1):373–404.

5. Riginos C, Crandall ED, Liggins L, Bongaerts P, Treml EA. Navigating the currents of seascape genomics: how spatial analyses can augment population genomic studies. Curr Zool. 2016;62(6):581–601.

6. Ravinet M, Westram A, Johannesson K, Butlin R, André C, Panova M. Shared and nonshared genomic divergence in parallel ecotypes of *Littorina saxatilis* at a local scale. Mol Ecol. 2016;25(1):287–305.

7. Tisthammer KH, Forsman ZH, Toonen RJ, Richmond RH. Genetic structure is stronger across human-impacted habitats than among islands in the coral *Porites lobata*. PeerJ. 2020;8:e8550.

8. Selkoe KA, Gaggiotti OE, Bowen BW, Toonen RJ. Emergent patterns of population genetic structure for a coral reef community. Mol Ecol. 2014;23(12):3064–79.

9. Stanley RRE, DiBacco C, Lowen B, Beiko RG, Jeffery NW, Wyngaarden MV, et al. A climate-associated multispecies cryptic cline in the northwest Atlantic. Sci Adv. 2018;4(3):eaaq0929.

10. De Wit P, Jonsson PR, Pereyra RT, Panova M, André C, Johannesson K. Spatial genetic structure in a crustacean herbivore highlights the need for local considerations in Baltic Sea biodiversity management. Evol Appl. 2019.

11. Benestan L, Quinn BK, Maaroufi H, Laporte M, Clark FK, Greenwood SJ, et al. Seascape genomics provides evidence for thermal adaptation and current-mediated population structure in American lobster (*Homarus americanus*). Mol Ecol. 2016;25(20):5073–92.

12. Bernatchez S, Xuereb A, Laporte M, Benestan L, Steeves R, Laflamme M, et al. Seascape genomics of eastern oyster (*Crassostrea virginica*) along the Atlantic coast of Canada. Evol Appl. 2019;12(3):587–609.

13. Saenz-Agudelo P, Dibattista JD, Piatek MJ, Gaither MR, Harrison HB, Nanninga GB, et al. Seascape genetics along environmental gradients in the Arabian Peninsula: insights from ddRAD sequencing of anemonefishes. Mol Ecol. 2015;24(24):6241–55.

14. Sandoval□Castillo J, Robinson NA, Hart AM, Strain LW, Beheregaray LB. Seascape genomics reveals adaptive divergence in a connected and commercially important mollusc, the greenlip abalone (*Haliotis laevigata*), along a longitudinal environmental gradient. Mol Ecol. 2018;27(7):1603–20.

15. Barshis DJ, Birkeland C, Toonen RJ, Gates RD, Stillman JH. High-frequency temperature variability mirrors fixed differences in thermal limits of the massive coral *Porites lobata*. J Exp Biol. 2018;221(24):jeb188581.

16. Narum SR, Buerkle CA, Davey JW, Miller MR, Hohenlohe PA. Genotyping-by-sequencing in ecological and conservation genomics. Mol Ecol. 2013;22(11):2841–7.

17. Razgour O, Taggart JB, Manel S, Juste J, Ibáñez C, Rebelo H, et al. An integrated framework to identify wildlife populations under threat from climate change. Mol Ecol Res. 2018;18(1):18–31.

18. von der Heyden S. Making evolutionary history count: biodiversity planning for coral reef fishes and the conservation of evolutionary processes. Coral Reefs. 2017;36(1):183–94.

19. Aldrich DP, Meyer MA. Social Capital and Community Resilience. Am Behav Sci. 2015;59(2):254–69.

20. Stapley J, Reger J, Feulner PGD, Smadja C, Galindo J, Ekblom R, et al. Adaptation genomics: the next generation. Trends Ecol Evol. 2010;25(12):705–12.

21. Balkenhol N, Dudaniec RY, Krutovsky KV, Johnson JS, Cairns DM, Segelbacher G, et al. Landscape Genomics: Understanding Relationships Between Environmental Heterogeneity and Genomic Characteristics of Populations. Population Genomics: Concepts, Approaches and Applications. Springer International Publishing; 2019 p. 261–322. https://doi.org/10.1007/13836_2017_2

22. Joost S, Bonin A, Bruford MW, Després L, Conord C, Erhardt G, et al. A spatial analysis method (SAM) to detect candidate loci for selection: towards a landscape genomics approach to adaptation. Mol Ecol. 2007;16(18):3955–69.

23. Forester BR, Lasky JR, Wagner HH, Urban DL. Comparing methods for detecting multilocus adaptation with multivariate genotype-environment associations. bioRxiv. 2017. http://biorxiv.org/lookup/doi/10.1101/129460

24. Rellstab C, Gugerli F, Eckert AJ, Hancock AM, Holderegger R. A practical guide to environmental association analysis in landscape genomics. Mol Ecol. 2015;24(17):4348–70.

25. Selkoe K, D'Aloia C, Crandall E, Iacchei M, Liggins L, Puritz J, et al. A decade of seascape genetics: contributions to basic and applied marine connectivity. Mar Ecol Prog Ser. 2016;554:1–19.

26. Wei K, Wood A, Gardner J. Seascape genetics of the New Zealand greenshell mussel: sea surface temperature explains macrogeographic scale genetic variation. Mar Ecol Prog Ser. 2013;477:107–21.

27. Takeuchi T, Masaoka T, Aoki H, Koyanagi R, Fujie M, Satoh N. Divergent northern and southern populations and demographic history of the pearl oyster in the western Pacific revealed with genomic SNPs. Evol Appl. 2020;13:837-853.

28. Xuereb A, Kimber CM, Curtis JMR, Bernatchez L, Fortin M. Putatively adaptive genetic variation in the giant California sea cucumber (*Parastichopus californicus*) as revealed by environmental association analysis of restriction□site associated DNA sequencing data. Mol Ecol. 2018;mec.14942.

29. Hanson JO, Rhodes JR, Riginos C, Fuller RA. Environmental and geographic variables are effective surrogates for genetic variation in conservation planning. Proc Natl Acad Sci. 2017;114(48):12755–60.

30. Dalongeville A, Benestan L, Mouillot D, Lobreaux S, Manel S. Combining six genome scan methods to detect candidate genes to salinity in the Mediterranean striped red mullet (*Mullus surmuletus*). BMC Genom. 2018;19(1):217.

31. Lotterhos KE, Whitlock MC. The relative power of genome scans to detect local adaptation depends on sampling design and statistical method. Mol Ecol. 2015;24(5):1031–46.

32. Smit AJ, Roberts M, Anderson RJ, Dufois F, Dudley SFJ, Bornman TG, et al. A Coastal Seawater Temperature Dataset for Biogeographical Studies: Large Biases between In Situ and Remotely-Sensed Data Sets around the Coast of South Africa. PLOS ONE. 2013;8(12):e81944.

33. Mmonwa K, Teske P, McQuaid C, Barker N. Historical demography of southern African patellid limpets: congruence of population expansions, but not phylogeography. Afr J Mar Sci. 2015;37(1):11–20.

34. Muller C, von der Heyden S, Bowie R, Matthee C. Oceanic circulation, local upwelling and palaeoclimatic changes linked to the phylogeography of the Cape sea urchin *Parechinus angulosus*. Mar Ecol Prog Ser. 2012;468:203–15.

35. Wright D, Bishop JM, Matthee CA, von der Heyden S. Genetic isolation by distance reveals restricted dispersal across a range of life histories: implications for biodiversity conservation planning across highly variable marine environments. Divers Distrib 2015;21(6):698–710.

36. Phair NL, Toonen RJ, Knapp I, Heyden S von der. Shared genomic outliers across two divergent population clusters of a highly threatened seagrass. PeerJ. 2019;7:e6806.

37. Toonen RJ, Puritz JB, Forsman ZH, Whitney JL, Fernandez-Silva I, Andrews KR, et al. ezRAD: a simplified method for genomic genotyping in non-model organisms. PeerJ. 2013;1:e203.

38. Morin PA, Luikart G, Wayne RK, the SNP workshop group. SNPs in ecology, evolution and conservation. Trends Ecol Evol. 2004;19(4):208–16.

39. Epps CW, Keyghobadi N. Landscape genetics in a changing world: disentangling historical and contemporary influences and inferring change. Mol Ecol. 2015;24:6021-6040.

40. de Lafontaine G, Ducousso A, Lefèvre S, Magnanou E, Petit RJ. Stronger spatial genetic structure in recolonized areas than in refugia in the European beech. Mol Ecol. 2013;22(17):4397–412.

41. Nadeau S, Meirmans PG, Aitken SN, Ritland K, Isabel N. The challenge of separating signatures of local adaptation from those of isolation by distance and colonization history: The case of two white pines. Ecol Evol. 2016;6(24):8649–64.

42. Sink K, van der Bank MG, Majiedt PA, Harris LR, Atkinson LP, Kirkman SP, et al. South African National Biodiversity Assessment 2018 Technical Report Volume 4: Marine Realm. South African National Biodiversity Institute, Pretoria. South Africa; 2018.

43. Teske PR, Sandoval-Castillo J, Golla TR, Emami-Khoyi A, Tine M, Heyden S von der, et al. Thermal selection as a driver of marine ecological speciation. Proc R Soc B. 2019; 286(1896): 20182023.

44. Diopere E, Vandamme SG, Hablützel PI, Cariani A, Van Houdt J, Rijnsdorp A, et al. Seascape genetics of a flatfish reveals local selection under high levels of gene flow. ICES J Mar Sci. 2018;75(2):675–89.

45. Pujolar JM, Jacobsen MW, Als TD, Frydenberg J, Munch K, Jónsson B, et al. Genome-wide single-generation signatures of local selection in the panmictic European eel. Mol Ecol. 2014;23(10):2514–28.

46. Eckert CG, Samis KE, Lougheed SC. Genetic variation across species' geographical ranges: the central–marginal hypothesis and beyond. Mol Ecol. 2008;17(5):1170–88.

47. Branch G. Two Oceans: A guide to the marine life of southern Africa. Penguin Random House South Africa; 2017.

48. Branch G, Branch M. The Living Shores of Southern Africa. Cape Town; 2018.

49. Ayre DJ, Minchinton TE, Perrin C. Does life history predict past and current connectivity for rocky intertidal invertebrates across a marine biogeographic barrier? Mol Ecol. 2009;18(9):1887–903.

50. Storfer A, Patton A, Fraik AK. Navigating the Interface Between Landscape Genetics and Landscape Genomics. Front Genet. 2018;9:68.

51. Coscia I, Wilmes SB, Ironside JE, Goward-Brown A, O'Dea E, Malham SK, et al. Fine-scale seascape genomics of an exploited marine species, the common cockle *Cerastoderma edule*, using a multi-modelling approach. bioRxiv. 2019;724062.

52. O'Connor MI, Bruno JF, Gaines SD, Halpern BS, Lester SE, Kinlan BP, et al. Temperature control of larval dispersal and the implications for marine ecology, evolution, and conservation. Proc Natl Acad Sci. 2007;104(4):1266–71.

53. Winch JJWW, Hodgson AN. The effect of temperature and salinity on oxygen consumption in the brachyuran crab *Cyclograpsus punctatus* (Crustacea: Decapoda: Grapsidae). Afr Zoo. 2007;42(1):118–23.

54. Anger K. Salinity as a key parameter in the larval biology of decapod crustaceans. Invert Rep Develop. 2003;43(1):29–45.

55. Pespeni MH, Palumbi SR. Signals of selection in outlier loci in a widely dispersing species across an environmental mosaic. Mol Ecol. 2013;22(13):3580–97.

56. Suarez-Ulloa V, Gonzalez-Romero R, Eirin-Lopez JM. Environmental epigenetics: A promising venue for developing next-generation pollution biomonitoring tools in marine invertebrates. Mar Poll Bull. 2015;98(1–2):5–13.

57. Prates I, Penna A, Rodrigues MT, Carnaval AC. Local adaptation in mainland anole lizards: Integrating population history and genome–environment associations. Ecol Evol. 2018;8(23):11932–44.

58. Foo SA, Byrne M. Chapter Two - Acclimatization and Adaptive Capacity of Marine Species in a Changing Ocean. Advances in Marine Biology. Academic Press; 2016. p. 69–116.

59. Osores SJA, Ruz GA, Opitz T, Lardies MA. Discovering divergence in the thermal physiology of intertidal crabs along latitudinal gradients using an integrated approach with machine learning. J Therm Biol. 2018;78:140–50.

60. Andrade SCS, Solferini VN. Fine-scale genetic structure overrides macro-scale structure in a marine snail: nonrandom recruitment, demographic events or selection? Biol J Linn Soc. 2007;91(1):23–36.

61. Gamboa M, Watanabe K. Genome-wide signatures of local adaptation among seven stoneflies species along a nationwide latitudinal gradient in Japan. BMC Genom. 2019;20(1):84.

62. The Global Invertebrate Genomics Alliance (GIGA): Developing Community Resources to Study Diverse Invertebrate Genomes. J Hered. 2014;105(1):1–18.

63. Goodwin S, McPherson JD, McCombie WR. Coming of age: ten years of next-generation sequencing technologies. Nat Rev Genet. 2016;17(6):333–51.

64. Lowry DB, Hoban S, Kelley JL, Lotterhos KE, Reed LK, Antolin MF, et al. Responsible RAD: Striving for best practices in population genomic studies of adaptation. Mol Ecol Resour. 2017;17(3):366–9.

65. Tiffin P, Ross-Ibarra J. Advances and limits of using population genetics to understand local adaptation. Trends Ecol Evol. 2014;12:673-680.

66. Dorant Y, Benestan L, Rougemont Q, Normandeau E, Boyle B, Rochette R, et al. Comparing Pool-seq, Rapture, and GBS genotyping for inferring weak population structure: The American lobster (*Homarus americanus*) as a case study. Ecol Evol. 2019;9(11):6606–23.

67. Schlötterer C, Tobler R, Kofler R, Nolte V. Sequencing pools of individuals — mining genome-wide polymorphism data without big funding. Nat Rev Genet. 2014;15(11):749–63.

68. Fischer MC, Rellstab C, Tedder A, Zoller S, Gugerli F, Shimizu KK, et al. Population genomic footprints of selection and associations with climate in natural populations of *Arabidopsis halleri* from the Alps. Mol Ecol. 2013;22(22):5594–607.

69. Griffiths CL, Robinson TB, Lange L, Mead A. Marine Biodiversity in South Africa: An Evaluation of Current States of Knowledge. PLOS ONE. 2010;5(8):e12008.

70. Hobday AJ, Pecl GT. Identification of global marine hotspots: sentinels for change and vanguards for adaptation action. Rev Fish Biol Fisheries. 2014;24(2):415–25.

71. Mertens LEA, Treml EA, von der Heyden S. Genetic and Biophysical Models Help Define Marine Conservation Focus Areas. Front Mar Sci. 2018;5:268.

72. Kurland S, Wheat CW, Mancera M de la PC, Kutschera VE, Hill J, Andersson A, et al. Exploring a Pool-seq-only approach for gaining population genomic insights in nonmodel species. Ecol Evol. 2019;9(19):11448–63.

73. Nielsen ES, Henriques R, Toonen RJ, Guo B, von der Heyden S. Complex signatures of genomic variation of two non-model marine species in a homogeneous environment. BMC Genom. 2018;19:347.

74. Knapp I, Puritz J, Bird C, Whitney M, Sudek M, Forsman Z, et al. ezRAD—an accessible next-generation RAD sequencing protocol suitable for non-model organisms_v3. Protocols. io Life Sciences Protocol Repository; 2016.

75. Andrews S. FastQC: a quality control tool for high throughput sequence data. 2010. http://www.bioinformatics.babraham.ac.uk/ projects/fastqc

76. Krueger F. Trim Galore!: a wrapper tool around Cutadapt and FastQC to consistently apply quality and adapter trimming to FastQ files. Babraham Bioinformatics, Cambridge, United Kingdom; 2015.

77. Teske PR, Golla TR, Sandoval-Castillo J, Emami-Khoyi A, Lingen CD van der, Heyden S von der, et al. Mitochondrial DNA is unsuitable to test for isolation by distance. Sci Rep. 2018;8(1):1–9.

78. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv:13033997. 2013; http://arxiv.org/abs/1303.3997

79. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009;25(16):2078–9.

80. Barnett DW, Garrison EK, Quinlan AR, Stromberg MP, Marth GT. BamTools: a C++ API and toolkit for analyzing and managing BAM files. Bioinformatics. 2011;27(12):1691–2.

81. Bushnell B. BBMap short-read aligner, and other bioinformatics tools. University of California, Berkeley, CA; 2015.

82. Chikhi R, Medvedev P. Informed and automated k-mer size selection for genome assembly. Bioinformatics. 2014;30(1):31–7.

83. Simpson JT, Wong K, Jackman SD, Schein JE, Jones SJM, Birol I. ABySS: A parallel assembler for short read sequence data. Genom Res. 2009;19(6):1117–23.

84. Li D, Liu C-M, Luo R, Sadakane K, Lam T-W. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. Bioinformatics. 2015;31(10):1674–6.

85. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. J Comput Biol. 2012;19(5):455–77.

86. Mayela Soto-Jimenez L, Estrada K, Sanchez-Flores A. GARM: Genome Assembly, Reconciliation and Merging Pipeline. Curr Top Medl Chem. 2014;14(3):418-424.

87. Gurevich A, Saveliev V, Vyahhi N, Tesler G. QUAST: quality assessment tool for genome assemblies. Bioinformatics. 2013;29(8):1072–5.

88. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics. 2015;31(19):3210–2.

89. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST+: architecture and applications. BMC Bioinformatics. 2009;10(1):421.

90. Kofler R, Pandey RV, Schlotterer C. PoPoolation2: identifying differentiation between populations using sequencing of pooled DNA samples (Pool-Seq). Bioinformatics. 2011 Dec 15;27(24):3435–6.

91. Hivert V, Leblois R, Petit EJ, Gautier M, Vitalis R. Measuring genetic differentiation from Pool-seq data. Genetics. 2018;210:315-330.

92. Leblois R, Gautier M, Rohfritsch A, Foucaud J, Burban C, Galan M, et al. Deciphering the demographic history of allochronic differentiation in the pine processionary moth *Thaumetopoea pityocampa*. Mol Ecol. 2018;27(1):264–78.

93. Goslee SC, Urban DL. The ecodist Package for Dissimilarity-based Analysis of Ecological Data. J Stat Soft. 2007;22(7).

94. Gautier M. Genome-Wide Scan for Adaptive Divergence and Association with Population-Specific Covariates. Genetics. 2015;4:1555–79.

95. Fick SE, Hijmans RJ. WorldClim 2: new 1□km spatial resolution climate surfaces for global land areas. Int J Climatol. 2017;37(12):4302–15.

96. Sbrocco EJ, Barber PH. MARSPEC: ocean climate layers for marine spatial ecology. Ecology. 2013;94(4):979–979.

97. Tyberghein L, Verbruggen H, Pauly K, Troupin C, Mineur F, Clerck OD. Paper. Glob Ecol Biogeogr. 2012;21(2):272–81.

98. Bosch S, Tyberghein L, De Clerck O. sdmpredictors: an R package for species distribution modelling predictor datasets. Marine Species Distributions: From data to predictive models; 2017.

99. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. J R Stat Soc Ser B. 1995;57(1):289–300.

100. Diniz-Filho JAF, Soares TN, Lima JS, Dobrovolski R, Landeiro VL, Telles MP de C, et al. Mantel test in population genetics. Gen Mol Biol. 2013;36(4):475–85.

101. Slatkin M. A Measure of Population Subdivision Based on Microsatellite Allele Frequencies. Genetics. 1995;139:457–62.

102. Team QD. QGIS geographic information system. Open Source Geospatial Foundation; 2009.

103. Dabney A, Storey JD, Warnes GR. qvalue: Q-value estimation for false discovery rate control. R package version. 1(0); 2010.

104. Jefferys H. Theory of probability (3rd edition). New York: Oxford university press; 1961.

105. Frichot E, Schoville SD, Bouchard G, François O. Testing for Associations between Loci and Environmental Gradients Using Latent Factor Mixed Models. Mol Biol Evol. 2013;30(7):1687–99.

106. Frichot E, François O. LEA: An R package for landscape and ecological association studies. Methods Ecol Evol. 2015;6(8):925–9.

107. Wagner HH, Chávez-Pesqueira M, Forester BR. Spatial detection of outlier loci with Moran eigenvector maps. Mol Ecol Resour. 2017;17(6):1122–35.

108. Bivand R, Anselin L, Berke O, Bernat A, Carvalho M, Chun Y, et al. spdep: Spatial dependence: weighting schemes, statistics and models. 2011.

109. Legendre P, Gallagher ED. Ecologically meaningful transformations for ordination of species data. Oecologia. 2001;129(2):271–80.

110. Oksanen J, Blanchet FG, Kindt R, Legendre P, Minchin PR, O'hara RB, et al. Package 'vegan.' Community ecology package, version 2. 2013;9:1-295.

111. Dray S, Legendre P, Peres-Neto PR. Spatial modelling: a comprehensive framework for principal coordinate analysis of neighbour matrices (PCNM). Ecoll Model. 2006;196(3–4):483–93.

112. Dray S, Blanchet G, Borcard D, Guenard G, Jombart T, Larocque G, et al. Adespatial: multivariate multiscale spatial analysis. R package version 0.0-9. 2017.

113. Blanchet FG, Legendre P, Borcard D. Forward Selection of Explanatory Variables. Ecology. 2008;89(9):2623–32.

114. Wickham H. ggplot2: elegant graphics for data analysis. J Stat Soft. 2016;77.

115. Conesa A, Götz S, García-Gómez JM, Terol J, Talón M, Robles M. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics. 2005;21(18):3674–6.

Supplementary information

Additional file 1: Sampling information, bioinformatic pipeline parameters, results of mitogenome mapping and de novo assembly comparisons, and single nucleotide polymorphism (SNP) results are shown per species.

Additional file 2: Per species pairwise F_{ST} values.

Additional file 3: The seascape features considered in the gene-environment association analyses, shown as values per sample site, as well as the Spearman's R coefficients and p-values between variables.

Additional file 4: BLAST results against NCBI non-redundant protein sequences, including the species, contig ID, protein match, E-value, and percent identical. The Gene Ontology (GO) terms from Blast2GO are also shown.

Tables

Table 1. Mantel and partial Mantel test results for *C. punctatus*, *P. angulosus*, and *S. granularis*.

	C. punctatus			*P. angulosus*			*S. granularis*		
Test	r	p	q	r	p	q	r	p	q
FST~D	0.18	0.20	0.39	0.13	0.18	0.54	0.48	**0.00**	**0.01**
FST~SSSmean|D	0.18	0.45	0.67	-0.02	0.87	0.87	-0.23	0.27	0.34
FST~SSSrange|D	-0.50	0.06	0.19	-0.02	0.77	0.87	-0.23	**0.03**	0.06
FST~SSTmean|D	0.43	**0.03**	0.19	-0.20	0.14	0.54	0.40	**0.00**	**0.01**
FST~SSTrange|D	0.04	0.79	0.79	-0.08	0.46	0.87	-0.14	0.39	0.39
FST~Trange|D	0.08	0.61	0.73	-0.06	0.71	0.87	-0.19	0.28	0.34

Correlation coefficients (r), *p-values* (p), and *q-values* (q) are given for tests between genetic distance (FST) and geographic distance (D), and distance matrices between each of the five environmental variables: mean sea surface salinity (SSSmean), sea surface salinity range (SSSrange), mean sea surface temperature (SSTmean), sea surface temperature range (SSTrange), and surface air temperature range (Trange). Significant values are denoted in bold.

Table 2. **Comparisons in number of outlier SNPs detected between seven outlier detection methods.**

Method (Abbreviation)	Model type	Correction for spatial or population structure	*C. punctatus* # outliers (# unique)	*P. angulosus* # outliers (# unique)	*S. granularis* # outliers (# unique)
BayPass core model (BPC)	Bayesian	Yes, population	9 (2)	5 (0)	19 (5)
BayPass auxiliary model (BPA)	Bayesian	Yes, population	0	0	4 (0)

Latent factor mixed model (LFMM)	Mixed model	Yes, population	134 (121)	72 (60)	125 (101)
Moran spectral outlier detection (MSOD)	Multivariate model	Yes, spatial	15 (14)	9 (7)	20 (18)
Moran spectral randomization outlier detection (MSR)	Multivariate model	Yes, spatial	3 (NA)	3 (NA)	8 (NA)
Redundancy analysis (RDA)	Multivariate model	No	9 (3)	9 (1)	16 (2)
Distance-based redundancy analysis (dbRDA)	Multivariate model	Yes, spatial	0	0	1 (1)

Descriptions of outlier detection methods, and the number of total and unique outliers (restricted to that method) detected by each method for each species. Note that MSR could not have unique outliers as it uses those identified by MSOD.

Figure legends

Figure 1. The distributions of the study species (*Cyclograpsus punctatus*, *Parechinus angulosus*, *Scutellastra granularis*), imposed over average sea surface temperatures within the study region (a), and the 20 sample sites included in the study (b).

Figure 2. Population differentiation is shown by PCoAs of Nei's genetic distance from all quality-filtered SNPs (a, c, e) and covariance (Ω) matrices represented as heatmaps (b, d, f), shown for *C. punctatus* (a, b), *P. angulosus* (c, d), and *S. granularis* (e, f). Letters in the PCoAs (a, c, e) correspond to the sample sites shown in Figure 1, with darker shaded letters corresponding to western sites, and lighter shaded letters corresponding to eastern sites.

Figure 3. The number of outlier SNPs detected per method for *C. punctatus* (a), *P. angulosus* (b), and *S. granularis* (c). See Table 2 for method abbreviations.

Figure 4. Genomic differentiation as shown by PCAs of allele frequencies in either the putatively neutral (a-c) or outlier (d-f) datasets for *C. punctatus* (a, d), *P. angulosus* (b, e), and *S. granularis* (c, f). Letters correspond to the sample sites shown in Figure 1, with darker shaded letters corresponding to western sites, and lighter shaded letters corresponding to eastern sites.

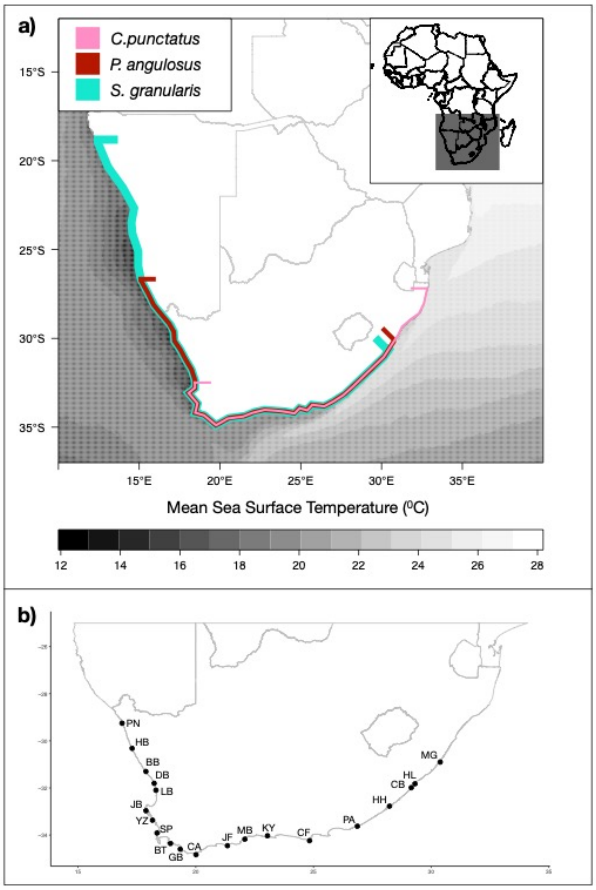

Figure 1. The distributions of the study species (*Cyclograpsus punctatus*, *Parechinus angulosus*, *Scutellastra granularis*), imposed over average sea surface temperatures within the study region (a), and the 20 sample sites included in the study (b).

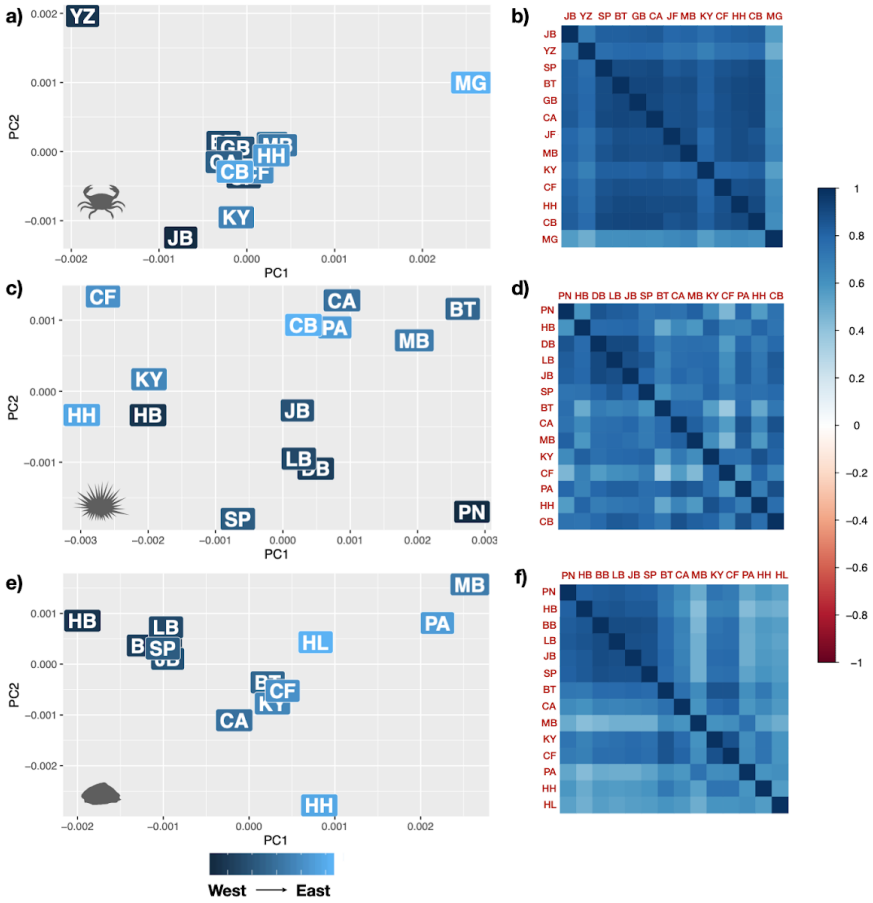

Figure 2. Population differentiation is shown by PCoAs of Nei's genetic distance from all quality-filtered SNPs (a, c, e) and covariance (Ω) matrices represented as heatmaps (b, d, f), shown for *C. punctatus* (a, b), *P. angulosus* (c, d), and *S. granularis* (e, f). Letters in the PCoAs (a, c, e) correspond to the sample sites shown in Figure 1, with darker shaded letters corresponding to western sites, and lighter shaded letters corresponding to eastern sites.

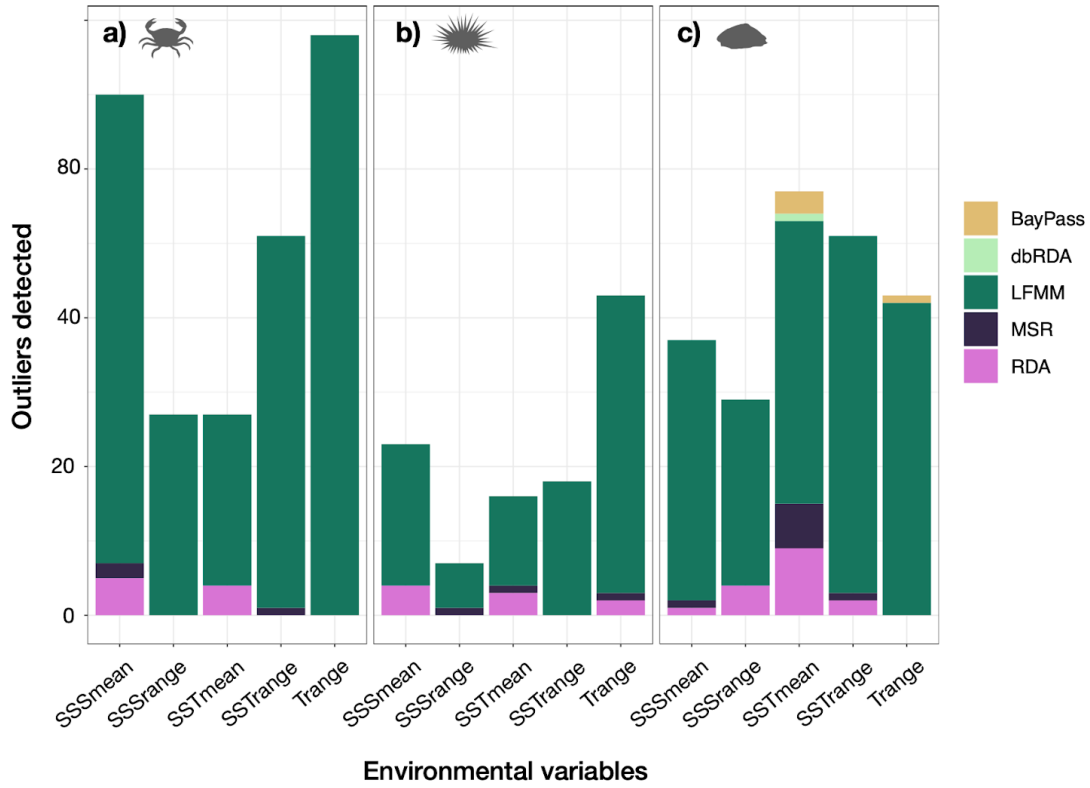

Figure 3. The number of outlier SNPs detected per method for *C. punctatus* (a), *P. angulosus* (b), and *S. granularis* (c). See Table 2 for method abbreviations.

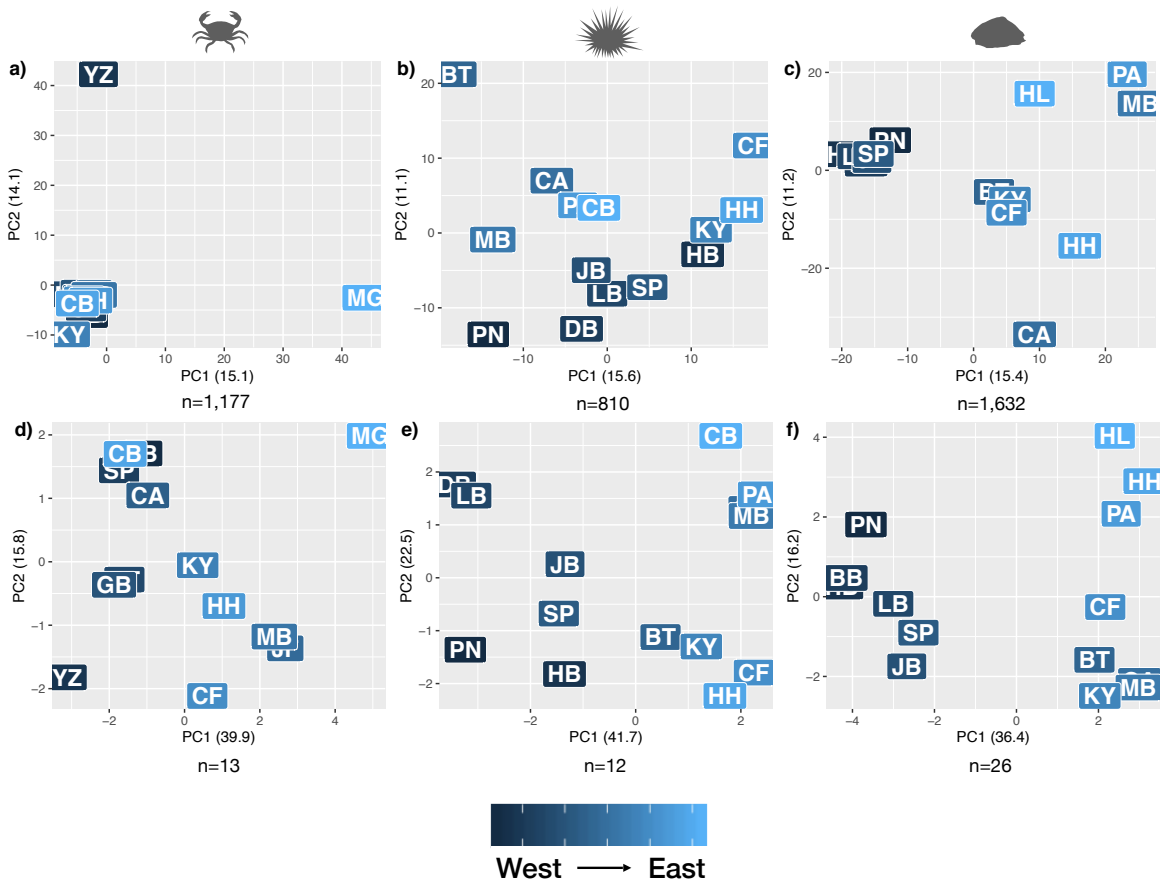

Figure 4. Genomic differentiation as shown by PCAs of allele frequencies in either the putatively neutral (a-c) or outlier (d-f) datasets for *C. punctatus* (a, d), *P. angulosus* (b, e), and *S. granularis* (c, f). Letters correspond to the sample sites shown in Figure 1, with darker shaded letters corresponding to western sites, and lighter shaded letters corresponding to eastern sites.