scientific reports

Check for updates

OPEN

Population genomic structure of the sea urchin *Diadema africanum*, a relevant species in the rocky reef systems across the Macaronesian archipelagos

Marc Peralta-Serrano[1], José Carlos Hernández[2], Romain Guet[1,2], Sara González-Delgado[1], Laia Pérez-Sorribes[1,3], Evandro P. Lopes[4,5] & Rocio Pérez-Portela[1,6]✉

The sea urchin *Diadema africanum* is a macro-herbivore found in the rocky reef systems of the West African region and Macaronesian archipelagos. Over several decades, high densities of this species have generated marine barrens in certain areas at the Canary Islands. In contrast, more recently, during the last few years, the species has suffered mass mortality events that continue to the present day. In this study, we used 9,109 Single Nucleotide Polymorphisms (SNPs) and a fragment of a mitochondrial gene to evaluate the species' population structure, effects of mass mortalities on its diversity, and potential local adaptation across the Canary Islands and Cabo Verde. Our research provides compelling evidence of low genomic diversity and homogeneity across the studied area for neutral markers, along with recent demographic fluctuations. The high connectivity among distant areas likely allows a rapid recovering of the populations from local mortality events. Interestingly, we also observed genomic sub-structure from 405 SNPs identified as candidate loci under selection for seawater temperature. The lack of divergence among distant sites and the low diversity found can be attributed to the species' divergence from a small ancestral genomic pool, followed by a contemporary demographic expansion, and ongoinggene flow.

The species' genetic structure depends on the combination of different processes including genetic drift, gene flow, and natural selection[1]. The effect of genetic drift primarily depends on the populations' size[2], being stronger in smaller populations. In marine species with limited mobility during adulthood, gene flow is influenced by the larval and/or gamete dispersal potential[3], but this dispersal capability interplays with the oceanographic circulation and geographical barriers to generate connectivity patterns over space[4,5]. Hence, although the species' dispersal capacity has been traditionally considered as one of the major factors determining genetic divergence among populations, the most recent molecular studies exploring large areas of the species' genome have shown that natural selection can also influence populations' genetic structure[6–8]. This means that the intraspecific divergence patterns are influenced by both "neutral" processes, which equally act over the whole genome of the species and independently to natural selection, and selective or "non-neutral" processes, such as adaptation to different environmental conditions, which differently affect certain genes or areas of the genome[9–11]. The combination of both "neutral" and "non-neutral" processes shapes the genomic architecture of the species' populations. For this reason, exploring population genomic patterns can be used as an approximation to determine the potential effect of each of these forces, and to understand the relative influence of environmental variables on the divergence patterns found in the current populations. Additionally, it is important to keep

[1]Departament de Biologia Evolutiva, Ecologia i Ciències Ambientals, Facultat de Biologia, Universitat de Barcelona (UB), Av. Diagonal 643, 08028 Barcelona, Spain. [2]Marine Comunity Ecology and Conservation, Departamento de Biología Animal, Edafología y Geología, Faculty of Science (Biology), University of La Laguna, San Cristóbal de La Laguna, Tenerife, Canary Islands, Spain. [3]Department of Ecology and Evolution, Estación Biológica de Doñana (CSIC), Seville, Spain. [4]Instituto de Engenharias e Ciências do Mar, Universidade Técnica do Atlântico, C.P. 163, Mindelo, Republic of Cabo Verde. [5]CIBIO, Centro de Investigação em Biodiversidade e Recursos Genéticos, InBIO Laboratório Associado, BIOPOLIS Program in Genomics, Biodiversity and Land Planning, Campus de Vairão, Universidade do Porto, 4485-661 Vairão, Portugal. [6]Institut de Recerca de la Biodiversitat (IRBio), Universitat de Barcelona, Barcelona, Spain. ✉email: rocio_perez@ub.edu

in mind that the current patterns of genomic divergence have been shaped by contemporary and historical processes acting over generations[12–14] and, for this reason, currently genomic patterns can provide insight about past demographic events[15].

Considering the advantages that recent molecular techniques offer to investigate the processes shaping the genomic architecture, and the ecological relevance of sea urchins, numerous population genomic studies have been conducted on multiple species ranging from tropical to polar areas[8,16–20]. At the east temperate and subtropical Atlantic coast, some remarkable studies on population genetics of sea urchins included two of the most important species for shallow sublittoral ecosystems, the black sea urchin *Arbacia lixula* (Linnaeus, 1758)[8,21,22] and the purple sea urchin *Paracentrotus lividus* (Lamarck, 1816)[8,18]. These studies confirmed the large dispersal potential their larvae, with moderate patterns of divergence among major oceanographic areas[18,21,22], but sub-structure due to selection to temperature and salinity on some genes[7,8] in both species.

Our study here presented focuses on the populations' genomic structure of a sea urchin species of the genus *Diadema* (Echinoidea, Echinodermata). *Diadema* is one of the largest and ecologically most important tropical sea urchin genus found in shallow waters[23–25]. Herbivorous sea urchins can influence the structure of algal assemblages, even at relatively low population densities[26,27]. Species of the genus *Diadema* are also notorious for their large populations' density variations, so-called "outbreak" or "die-off" events[28–32] with the consequent cascading effects overall ecosystem[33–35]. One of the eight named species of this worldwide distributed genus is the recently described species, *Diadema africanum* (Rodríguez, Hernández, Clemente & Coppard, 2013). This species likely diverged from *D. antillarum* (Philippi, 1845) during an allopatric speciation process between western and eastern Atlantic pools separated by the mid-Atlantic barrier[24]. Despite its planktotrophic larva remaining for approximately 39 days in the water column[36], the distribution of *D. africanum* is restricted to the eastern Atlantic Ocean, from the Guinean Gulf[37] to the eastern Atlantic archipelagos[25], which comprises São Tomé and Príncipe Islands, Cabo Verde, Canary Islands, Salvage Islands, Madeira, and Azores[38]. The species *D. africanum*, like other species of the genus, plays an engineering role in the rocky reef ecosystems, due to the great grazing capacity[25]. The increase of *D. africanum* populations' densities in different sites of the Canary Islands and Madeira, associated with the overfishing of its potential predators and the effects of global warming favouring larvae survival[39,40], led to the formation of barren grounds since the late 80s[25,41,42], and an expansion of the species distribution to northerner latitudes[36]. However, the introduction and propagation of marine pathogens during the last decades[30,43–45], triggered mass mortality events for this species. The "Xynthia" and "Emma" meteorological events, in 2009 and 2018, respectively, negatively impacted on *D. africanum* populations across the Canary Islands, promoting the development of diseases[30,46,47] that decimated populations throughout the archipelagos[30,46]. Interestingly, similar cases of mass mortality were observed in the 80 s in the sibling species *D. antillarum* (Philippi, 1845) on the Caribbean coast, where mass mortality events drove the disappearance of 98% of individuals of the species[29]. These mass mortality events are important at the population level because when the reduction of the effective population size is extreme or persists over time it may result in a bottleneck and a higher susceptivity to the loss of genetic diversity due to genetic drift[48]. Despite the relevance of *D. africanum* across its distribution range no previous population genetic studies have been performed to date. Additionally, the effect of mass mortality events on intraspecific genetic diversity is still unknown.

For these reasons, we here conduct a genome-wide analysis in *D. africanum* with the general aim of evaluating the genomic structure and diversity of its populations, and the potential patterns of substructure due to local adaptation. The specific objectives of this study were: (a) to explore the genomic structure and connectivity patterns of *D. africanum* across the Canary Islands and Cabo Verde; (b) to identify whether demographic fluctuations had an impact on the current levels of genomic diversity in the species, and (c) to identify potential patterns of local adaptation due to changes in the environmental variables along the Canary Islands. Our initial hypothesis is that *D. africanum* holds low levels of divergence across the studied area due to its large dispersal capability but potential local adaptation, as those observed in other sea urchins. We also hypothesised that recent mass mortality events have reduced the genetic diversity of the affected populations. This information may help to predict the potential future vulnerability of the species in the Canary Islands and to predict future distribution shifts and abundance changes.

Results

Mitochondrial COI sequence diversity, structure, and demography

A 702 bp fragment of the mitochondrial cytochrome c oxidase I (COI) gene was obtained from 249 individuals of *Diadema africanum,* collected from 11 sampling sites at the Canary Islands and three from Cabo Verde (Fig. 1). All these sequences corresponded to 33 new haplotypes for the species (GenBank accession numbers: OP494354-OP494596). A total of 29 nucleotide variable sites were found, 28 of them being polymorphic sites (3.99%) and only seven representing non-synonymous substitutions. Cabo Verde retained only three haplotypes, two shared with the Canary Islands and a third one (H28) private from Cabo Verde islands. In contrast, 32 haplotypes were found in the Canary Islands. The number of haplotypes found per site comprised a minimum of two at the sites of C_C1 and C_C2 in Cabo Verde (see details in Fig. 1), and a maximum of eight haplotypes for the site PB in La Palma in the Canary Islands (Fig. 1). For the entiregeographical range, the calculated haplotype diversity (Hd) and nucleotide diversity (π) were 0.578 ($\pm$0.035 SD) and 0.00109 ($\pm$0.00009), respectively (Table 1).

Over the entire dataset (Table 1), the measures of haplotype and nucleotide diversity were relatively consistent across all sites within Cabo Verde and the Canary Islands. Considering the haplotype richness (Hk) and diversity (Hd) in the Canary Islands, the site of T_C2 in Tenerife retained the highest values (Hk = 3.288, and Hd = 0.725 $\pm$ 0.086), while the lowest values were found in the site of L_C5 in the easternmost region of the Canary Islands (Hk = 0.890, and Hd = 0.264 $\pm$ 0.136). In comparison, haplotype richness ranged between 0.750 and 1 in Cabo Verde. For nucleotide diversity, a similar pattern was observed. Indeed, diversity values were quite low, ranging from 0.00038 $\pm$ 0.00019 in L_C5 (Lanzarote) to 0.00149 $\pm$ 0.00034 in P_C2 (La Palma). The

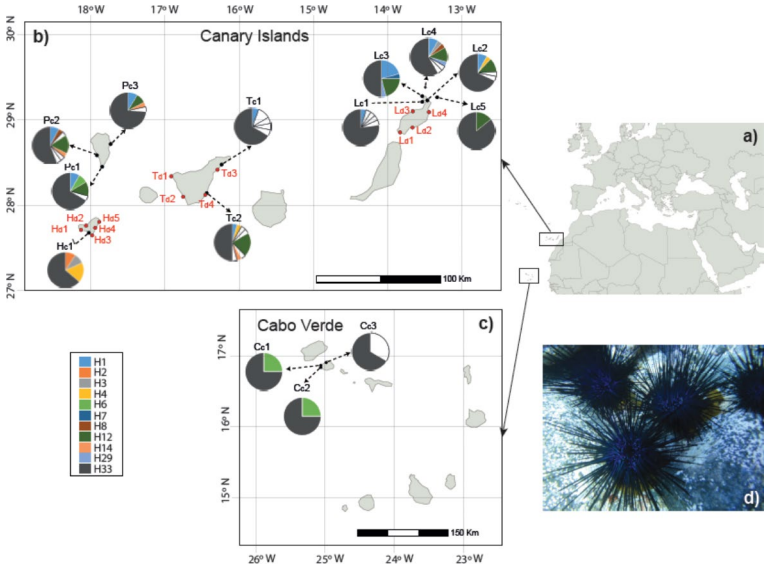

Fig. 1. Sampling sites of *Diadema africanum* for the COI analyses (black dots) and the SNPs analyses (red dots). The pie charts represent the frequency of COI haplotypes in which each colour partitions correspond to a different haplotype, except for the white colour that represents private haplotypes. (**a**) General map of the area, (**b**) the Canary Islands, (**c**) Cabo Verde, and (**d**) image of individuals of *D. africanum* from Tenerife (Canary Islands). (HAP = haplotypes from COI). The maps were created using QGIS v.3.16.16 and edited in Adobe Illustrator. The *D. africanum* picture belongs to J.C. Hernández.

Locality	Code	Island	N	Lat	Long	Nh	Hk	Hd	π
Ponta Cabrea	Cc1	São Vicente	4	16.8538	−25.0722	2	0.750	0.5 ± 0.265	0.00071 ± 0.00038
Santo André	Cc2	São Vicente	4	16.8283	25.0717	2	0.750	0.5 ± 0.265	0.00071 ± 0.00039
Laginha	Cc3	São Vicente	4	16.8946	−24.9931	2	1.000	0.833 ± 0.222	0.00142 ± 0.00048
Cabo Verde	ALL Cabo Verde		12			3	–	0.561 ± 0.154	0.00091 ± 0.00041
Tacorón	Hc1	El Hierro	11	27.6699	−18.0265	3	2.000	0.6 ± 0.024	0.00124 ± 0.00037
Fuentecaliente	Pc1	La Palma	25	28.4557	−17.8466	5	2.389	0.547 ± 0.112	0.00096 ± 0.00022
La Bombilla	Pc2	La Palma	25	28.5911	−17.9181	8	3.256	0.673 ± 0.096	0.00149 ± 0.00034
Puerto del Trigo	Pc3	La Palma	23	28.7242	−17.7328	5	2.063	0.455 ± 0.123	0.00072 ± 0.00022
Santa Cruz de Tenerife (Escuela de Náutica)	Tc1	Tenerife	18	28.4812	−16.2399	6	2.765	0.562 ± 0.134	0.00094 ± 0.00028
Abades	Tc2	Tenerife	24	28.1427	−16.4390	8	3.288	0.671 ± 0.094	0.00124 ± 0.00026
Montaña Amarilla	Lc1	La Graciosa Island, Lanzarote	24	29.2205	−13.5385	6	2.045	0.504 ± 0.123	0.00094 ± 0.00029
Punta Fariones	Lc2	Lanzarote	25	29.2386	−13.4705	5	2.145	0.493 ± 0.117	0.00087 ± 0.00023
Cuevas Coloradas	Lc3	Montana Clara Islet, Lanzarote	24	29.2884	−13.5393	5	2.609	0.645 ± 0.081	0.0011 ± 0.0002
Punta de La Mareta	Lc4	Alergransa Islet, Lanzarote	24	29.3924	−13.4970	8	3.270	0.656 ± 0.0105	0.00132 ± 0.00034
Roque del Este (Sotavento)	Lc5	Lanzarote	14	29.2744	−13.3386	2	0.890	0.264 ± 0.136	0.00038 ± 0.00019
Canary Islands	ALL Canary Islands		237			32	–	0.579 ± 0.036	0.00110 ± 0.00660
Canary Islands & Cabo Verde	ALL		249			33	–	0.578 ± 0.035	0.00109 ± 0.00009

Table 1. Diversity values for the mitochondrial COI fragment in *Diadema africanum*. Sampling locality, code of the sampling locality, island of collection, number of individuals analysed (N), geographical coordinates of latitude (Lat) and longitude (Long), number of haplotypes (Nh), haplotype richness (Hk), haplotype diversity (Hd), and nucleotide diversity (π) including their standard deviations. Data not available (−).

haplotype richness was low for almost all the sampled populations ranging from 0.890 for the L_C5 site to 3.288 for the T_C2 site in Tenerife. Nevertheless, comparing both haplotype and nucleotide diversity between regions, the results showed no significant differences in diversity between them ($s = 0.6546$; p-value > 0.05). Additionally, the Canary Islands showed the highest diversity of haplotypes.

A computed haplotype network for all haplotypes obtained showed a star-like shape (Fig. 2), characterised by a remarkable number of unique haplotypes, mostly related to a central and most abundant haplotype (H33). This pattern was detected in all the studied sites and most of them were separated by only one or a few mutational

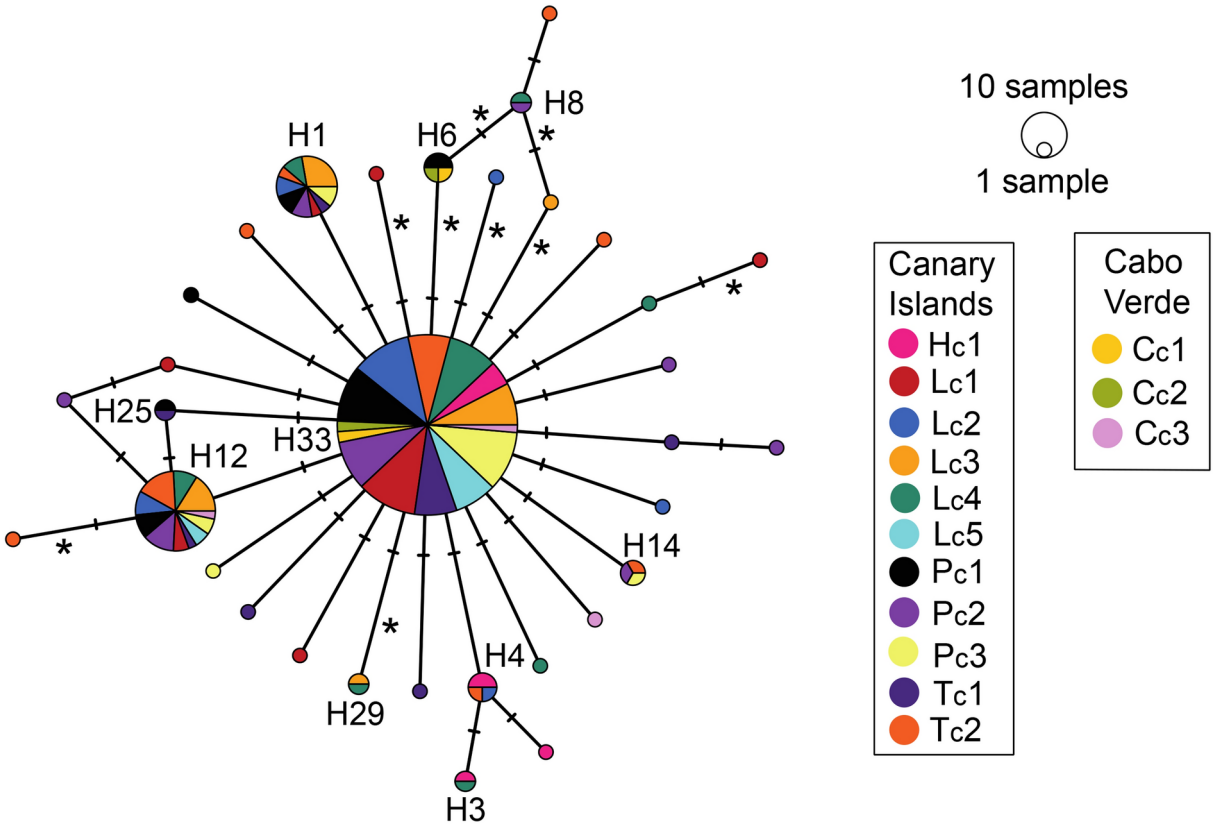

Fig. 2. Haplotype network from the COI dataset. Each haplotype is depicted by a circle, and colour partitions indicate the sampling sites where it was sampled. The size of each circle is proportional to the haplotype frequency. Connecting lines among haplotypes represents a mutational step, and asterisks illustrate the non-synonymous substitution. Only the most common haplotypes are presented by their numbers (H1, H3, H4, H6, H8, H12, H14, H25, H29, and H33). This image was created using Popart v.1.7 and edited in Adobe Illustrator.

steps from the main haplotype. In contrast, eight other haplotypes had more than one mutation and, in nine cases, with a non-synonymous substitution (Fig. 1).

Regarding to the genetic structure obtained from the COI across the two archipelagos, the AMOVA analysis showed no significant differences between Cabo Verde and the Canary Islands, with 0% of genetic variation between them ($\Phi_{ST} = -0.00087$; $p = 0.419$), and most genetic diversity variation retained within sampling sites (> 99%). A more detailed analysis of genetic differentiation, using F_{ST}, showed no significant differences among sampling sites in the Canary Islands and Cabo Verde (see Supplementary material, Table S1).

The three different neutrality tests applied to the COI, the Tajima's *D*, the Fu's Fs and the Ramos-Onsins & Rozas' R^2, and the mismatch distribution analysis calculated for the entire COI dataset showed coherent results. The first three tests displayed significant values (see Supplementary material, Table S2). Indeed, negative values of the Tajima's D and the Fu's Fs, as well as small values of R^2, have been interpreted as features of population expansion. Additionally, the R^2 statistic, considered a better descriptor for demographic expansions, showed a significant value for the overall dataset. (R^2 statistic 0.0806; *p*-value = 0.02). Therefore, information recovered from these statistics points to a recent demographic expansion of the species in the studied area. The mismatch distribution of the studied data also showed a unimodal curve close to the y-axis (see Supplementary material, Fig. S1), suggesting a very recent expansion (*p*-value > 0.05). The hypothesis that the studied data have a fairly good fit to the expansion model was tested through the sum of squared deviations (SSD) and the raggedness index. In this case, non-significant values for SSD suggested that the data fit the model of expansion. Additionally, the non-significant value of raggedness is also a parameter of population expansion. The estimated time of the demographic expansion in *D. africanum*, using the known mutation rate for *Diadema*[49], was estimated between 28,800 and 45,500 years ago.

Genomic diversity and divergence at the Canary Islands using single nucleotide polymorphisms (SNPs)

The total dataset of *D. africanum* from ddRADseq loci included 9,109 SNPs from 179 individuals, from thirteen sampling sites in the Canary Islands (five from El Hierro, four from Tenerife, and four from Lanzarote) (see Fig. 1 and Table 2), after removing 4 SNPs that were not in Hardy–Weinberg equilibrium in all populations. These sampling scheme included five sites affected by the last 2018 mass mortality event and eight sampling sites

Locality	Code	Island	N (n)	Lat	Long	Biosamples	Neutral				Candidate		
							A_R	Ho	He	F_{IS}	Ho	He	F_{IS}
Embarcadero de Orchilla*	Hd1	El Hierro	19 (13)	27.7044	−18.1329	SAMN39456681-SAMN39456699	1.514	0.218	0.227	−0.204	0.195	0.199	−0.202
Roque de la Hoya	Hd2	El Hierro	4 (2)	27.7555	−18.0642	SAMN39456723-SAMN39456726	–	0.207	0.234	0.049	0.172	0.177	−0.082
Muelle de La Restinga*	Hd3	El Hierro	20 (19)	27.6394	−17.9820	SAMN39456727-SAMN39456746	1.529	0.203	0.230	−0.010	0.170	0.187	−0.128
Roque de la Bonanza	Hd4	El Hierro	19 (17)	27.7292	−17.9408	SAMN39456747-SAMN39456765	1.535	0.205	0.235	0.009	0.165	0.196	0.046
La Caleta	Hd5	El Hierro	19 (15)	27.8002	−17.8873	SAMN39456662-SAMN39456680	1.518	0.205	0.227	−0.053	0.174	0.196	−0.022
Playa de Teno	Td1	Tenerife	21 (12)	28.3421	−16.9179	SAMN39456766-SAMN39456786	1.498	0.194	0.223	−0.160	0.173	0.190	−0.203
Caleta de Adeje*	Td2	Tenerife	19 (12)	28.0991	−16.7565	SAMN39456643-SAMN39456661	1.495	0.204	0.220	−0.320	0.171	0.189	−0.264
La Jaca*	Td3	Tenerife	23 (15)	28.1207	−16.4605	SAMN39456709-SAMN39456722	1.514	0.193	0.227	−0.087	0.153	0.175	−0.096
Muelle de Añaza	Td4	Tenerife	21 (15)	28.4198	−16.2952	SAMN39456622-SAMN39456642	1.520	0.194	0.230	−0.137	0.168	0.203	−0.118
Playa Flamingo	Ld1	Lanzarote	20 (17)	28.8565	−13.8423	SAMN39456787-SAMN39456806	1.525	0.194	0.230	−0.015	0.190	0.221	−0.029
Punta Tiñoza*	Ld2	Lanzarote	20 (16)	28.9183	−13.6702	SAMN39456845-SAMN39456864	1.519	0.194	0.227	−0.008	0.178	0.207	−0.013
La Santa	Ld3	Lanzarote	16 (10)	29.1108	−13.6655	SAMN39456829-SAMN39456844	1.502	0.193	0.227	−0.154	0.189	0.218	−0.156
Mala	Ld4	Lanzarote	22 (16)	29.1007	−13.4482	SAMN39456807-SAMN39456828	1.533	0.200	0.234	−0.017	0.189	0.216	−0.030

Table 2. Diversity values for the SNPs datasets of *Diadema africanum*. Sampling locality, code of the sampling site, island, number of individuals sampled (N) and number of individuals retained after filtering (n), geographical coordinates of latitude (Lat) and longitude (Long), biosample number, and main genetic features of the neutral SNPs (Neutral) and candidate SNPs under selection of temperature (candidate), including allelic richness (A_R (− = no data), observed heterozygosity (Ho), expected heterozygosity (He) and F_{IS}. *Sampling sites affected by the 2018 mass mortality event.

not affected by it. The total number of reads obtained was 1,269,736,321, with an average coverage of 7,093,499 (ranging from 727,179 to 43,584,550) reads per sample and 19.13% of missing data in the whole dataset. Among them, a generated neutral dataset, excluding all SNPs under potential selection, included 8,704 SNPs, and the candidate SNPs under selection for temperature were 405 SNPs, identified using the Redundancy analyses (RDA) (see the Methods section for more details about how these datasets were created), while no SNPs under selection were identified with BayeScan.

Genetic diversity, measured as observed heterozygosity (Ho) and expected heterozygosity (He), ranged from 0.193 in T_d3 and L_d3 to 0.218 in H_d1, with an average of 0.200 (±0.018) for the Ho, and from 0.220 in T_d2 to 0.235 in H_d4, with an average of 0.229 (±0.009) for He. Observed heterozygosity was lower than expected heterozygosity in all sampling sites. The F_{IS} statistic mostly displayed low negative values (Table 2), with non-significant *p*-values (*p*-value > 0.05) after the FDR correction. Allele richness (A_R) per locus and sampling site were very similar between them ranging from 1.495 in T_d2 to 1.535 in H_d4.

The populations affected and unaffected by the 2018 mass mortality event did not show significant differences in diversity for either allelic richness or observed heterozygosity (*p*-value > 0.05).

Regarding the genomic structure in the Canary Islands, all analyses applied showed a lack of genomic divergence among sampling sites and across the three islands analysed, El Hierro, Tenerife and Lanzarote, from the neutral SNPs. The AMOVA analysis from neutral SNPs did not show significant differences between islands within the Canary Archipelago either, with 0% of variation among islands, only 12.6% of the genetic variation between populations, and most genetic diversity variation, 87.5%, retained within individuals ($\Phi_{ST} = 0.125, -0.001$ and 0.124, respectively, and $p > 0.05$ for all of them). Again, all pairwise F_{ST} distances among sampling sites displayed low and/or no significant values (after the FDR correction of the *p*-values), suggesting genomic homogeneity (see the heatmap of the F_{ST} values in Fig. 3 and the complete table of F_{ST} values in Supplementary material, Table S3). The migration network generated from the G_{ST} values indicated high connectivity between all the sampling sites and islands. We can see that the sites T_d2 and T_d2 from Tenerife, and L_d3 from Lanzarote, are mainly a source of migrants to other localities (see Supplementary material, Fig. S2). The discriminant analyses of principal components (DAPC) also evidenced genomic homogeneity, with only one genomic cluster grouping all the thirteen sampling sites. The sampling sites H_d4 and H_d2 are slightly separated from the centre, being respectively in the east part of El Hierro, between H_d5 and H_d3, and the north-western part of El Hierro (see Supplementary material, Fig. S3). The difference observed in the H_d2 site was likely caused by the small sampling size. The Bayesian clustering analysed performed in STRUCTURE (Fig. 4) attributed the most likely K value to three genetic clusters (see Supplementary material, Table S5), but the thirteen sampling sites, ordered from west to east across the Canary Islands in the figure, had mixed individuals from the three genetic clusters with no clear geographical structure.

Local adaptation across the Canary Islands from candidate SNPs

From the 405 candidate SNPs under selection to temperature detected from RDA analyses (see the Methods section), the observed heterozygosity (Ho) ranged from 0.165 in DRB to 0.195 in H_d1, with an average of 0.176 (±0.019), while the expected heterozygosity (He) ranged from 0.175 in T_d3 to 0.218 in L_d3, with an average of 0.198 (±0.023). The F_{IS} estimated from heterozygosity showed non-significant results (*p*-values > 0.05) after the FDR correction (Table 2).

In contrast to the neutral SNPs dataset, we detected genomic divergence among islands associated with seawater temperature. Pairwise F_{ST} distances showed in general low values, ranging from 0 to 0.047, but the sites of Lanzarote (L_d1—4), the coldest island among the analysed here, were significantly different to other islands (see Fig. 3 and Supplementary material, Table S4). Similarly, the DAPC revealed two genomic clusters: one grouping the sites of El Hierro and Tenerife, and the other grouping the Lanzarote sites. The STRUCTURE result was consistent with the previous findings, suggesting the existence of two main genomic groups (see Fig. 4 and Supplementary material, Table S5). The sampling sites from the south-western and warmer islands, El Hierro and Tenerife, were mainly represented by the first genomic cluster, and the sites from Lanzarote by the second one (see Fig. 4).

Past demography from neutral SNPs

The Stairway Plot 2 results (Fig. 5) revealed the existence of a demographic expansion more than 100,000 years ago, reaching the population size a plateau approximately for ~ 90,000 years. According to the data, approximately 10,000–12,000 years ago, the effective population size of *D. africanum* has drastically decreased till today.

Discussion

We here address the first population genomic analysis of *Diadema africanum*, a recently described sea urchin species (Rodriguez et al. 2013), and the most important key herbivore of the shallow-waters rocky reef ecosystems in the West African region and the Macaronesian archipelagos[36]. Our study undertakes sampling throughout a large part of the species' distribution range, including populations that have undergone recent mass mortality events. Using 9,109 SNPs and a fragment of the mitochondrial COI gene, we investigate levels of diversity, divergence patterns and historical demographic changes in this species.

In general, *D. africanum* showed genomic homogeneity over short, medium, and long- distances across the area studied, for both neutral SNPs and COI, and for all the analyses applied (e.g. STRUCTURE, F_{ST}, DAPC, etc.). The weak genomic structure found a priori supports that *D. africanum* form a unique population across the Canary Islands and Cabo Verde. This homogeneity observed from the neutral markers is explained by high levels of gene flow. This species has a planktotrophic larva with an extended development time and the ability to survive under conditions of low food supply[36]. Larvae that remain several weeks in the water column may promote dispersal over large geographical distances[36], as observed from molecular data in other echinoderms[17,21,50–53].

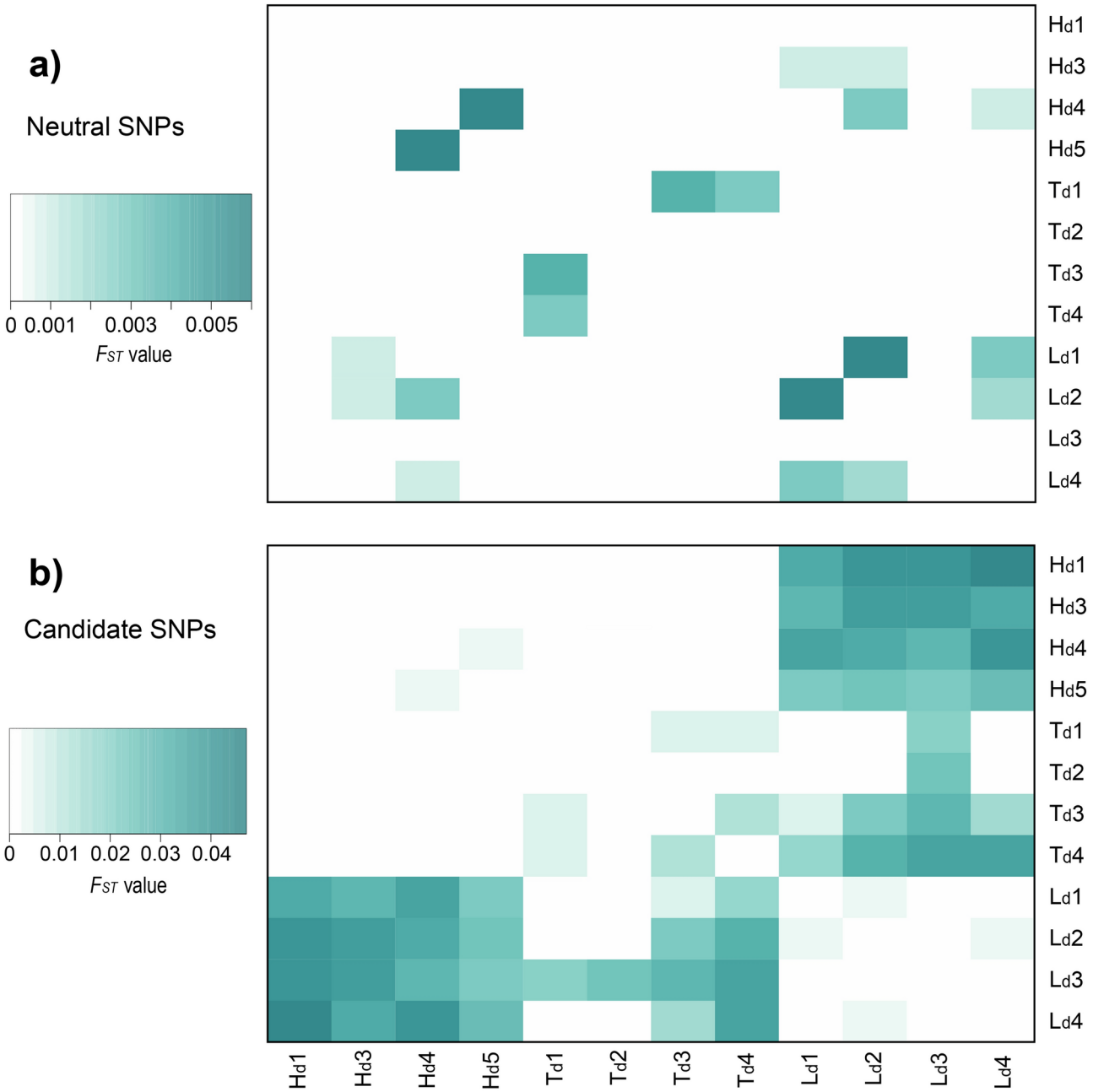

Fig. 3. Heatmaps obtained from the F_{ST} values using SNPs. (**a**) Results from the F_{ST} obtained from the neutral SNPs dataset, and the candidate SNPs dataset (**b**). Values and significance of the F_{ST} statistics are included as Supplementary material (Tables S3 and S4). Codes are described in Table 1.

For instance, *D. antillarum*, neither displayed evidence of genetic divergence along the coast of Florida over distances of hundreds of kilometres[19]. Additionally, the lack of potential large predators in the eastern Atlantic archipelagos on this species boosts the survival of larvae and high recruitment success[36,54]. The high connectivity, and larvae exchange between sites and islands, together with the absence of predators, can explain the rapid recovery of the populations affected by the first mass mortality events[47] through the recruitment of larvae from close and/or distant sites that could act as genomic sources.

Nevertheless, the underlying sub-structure found in *D. africanum*, with the significant differentiation of all the Lanzarote sampling sites (corresponding to the coldest sites studied), but undetectable from the neutral SNPs and the COI, was revealed from 405 SNPs. These loci were potentially located on genes under selection, prompting local adaptation to temperature under high levels of gene flow. Previous studies with sea urchins, as those with *Arbacia lixula* and *Paracentrotus lividus*, also unmasked significant genomic sub-structure related to local adaptation to seawater temperature and salinity[7,8], even when neutral markers did not show genomic divergence over large geographical distances across the Atlantic-Mediterranean area[18,21,22,55].

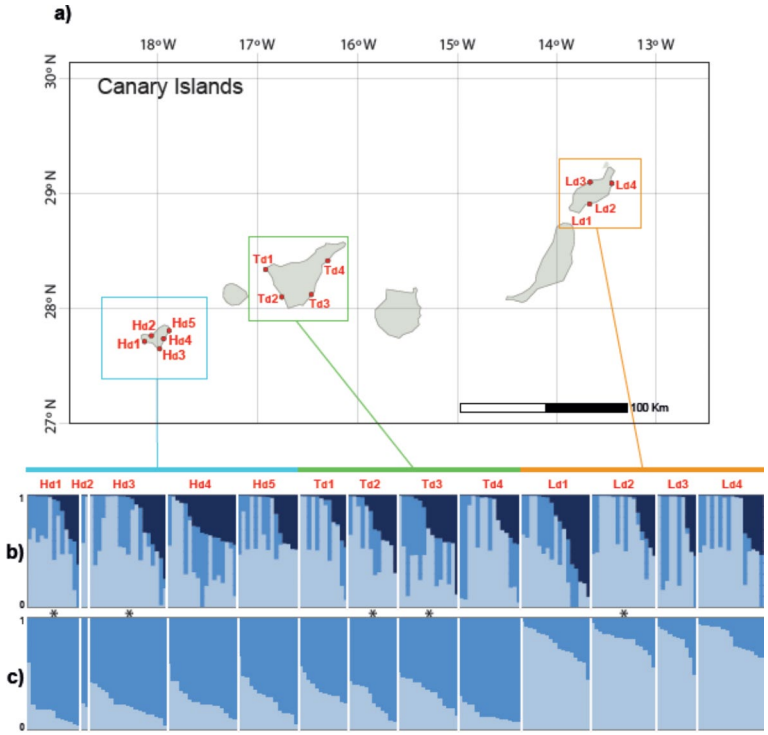

Fig. 4. Structure results. Sampling sites map (**a**), STRUCTURE barplots from the neutral SNPs (**b**), and candidate SNPs (**c**). The barplots are represented with the most probable number of genetic clusters (K = 3 and 2 for (**b**) and (**c**), respectively). *The asterisk indicates the sites affected by the 2018 mass mortality event. This map was created in QGIS v.3.16.16 and edited with Adobe Illustrator.

Together with the low level of divergence found in *D. africanum*, we also found lower levels of genomic diversity, for the COI (e.g., haplotype diversity = 0.578 ± 0.035) and SNPs (SNPs mean Ho = 0.200 ± 0.018), than in other co-occurring sea urchins as *A. lixula* (haplotype diversity = 0.912 ± 0.007 for COI, and mean Ho = 0.379 ± 0.002 for SNPs)[7,22] and *P. lividus* (haplotype diversity = 0.961 ± 0.009 for COI, and mean Ho = 0.370 ± 0.005 for SNPs)[8,18]. Although comparing different numbers and types of genomic markers affects the genetic diversity measures, we still detected lower diversity values in *D. africanum* than in other species for the same fragment of the COI. In the closely related species, *D. antillarum,* genetic studies of populations were mostly based on a reduced number of microsatellite loci, making the comparison of diversity values with our SNPs inaccurate[19]. In general, and despite the limitations of comparing genetic diversity levels among species, it seems that the low diversity found in *D. africanum* is not a specific feature of sea urchins, or the genus *Diadema,* but a singularity of *D. africanum*. The lack of significant differences in diversity between sampled sites that passed through mass mortality events and those that were not affected by mortalities points out that low diversity is not caused by the recent events of mass mortality[30,46,47]. Contrary, our results suggest that its cause is related to evolutionary origin of the species and a historical trend. Previous studies in the Eastern Atlantic indicated that *D. africanum* populations were isolated from a *Diadema* ancestor as the result of sea level fluctuations and change in the direction of currents during glacial cycles in the Pleistocene or Late Pliocene, and the formation of the mid-Atlantic barrier that isolated several species at both sides of the Atlantic Ocean[24,56]. This isolation promoted the allopatric speciation of *D. africanum* at the east Atlantic, a speciation process like others occurred in the genus[24]. This hypothesis proposes that *D. africanum* has recently diverged from an ancestor, likely from a small genomic pool that passed through a strong founder effect[57]. Although our demographic analyses indicate that *D. africanum* effective population size grew up until ~ 100,000 years ago, probably since its origin, the strong founder effect fingerprint persist until its populations reach the equilibrium[58,59]. This expansion inferred from molecular data is coherent with the empirical observations of the geographical expansion of the species in norther areas in the Atlantic archipelagos[38]. What is not clear from our demographic inference is whether the drastic reduction of the species' effective population size during the last 10,000 years is related to recurrent mortality events suffered by *D. africanum,* which, as observed, is reaching minimum values of effective population size[30]. Considering the lack of an accurate molecular clock in *Diadema* for nuclear DNA, demographic results should be taken with caution and only considered as a broad approximation, but what they clearly reflect is the decline of *D. africanum* populations.

The general pattern of low diversity, weak genomic structure and recent demographic fluctuations resembles, in some ways, to that in the Atlantic-Mediterranean sea star *Echinaster sepositus* (Retzius, 1783)[60], suggesting that successful marine invertebrate species displaying low levels of diversity are more common than initially expected. To conclude, our study demonstrates potential local adaptation to temperature, despite the existence of high levels of connectivity. This finding reinforces the idea that local adaptation can happen under high

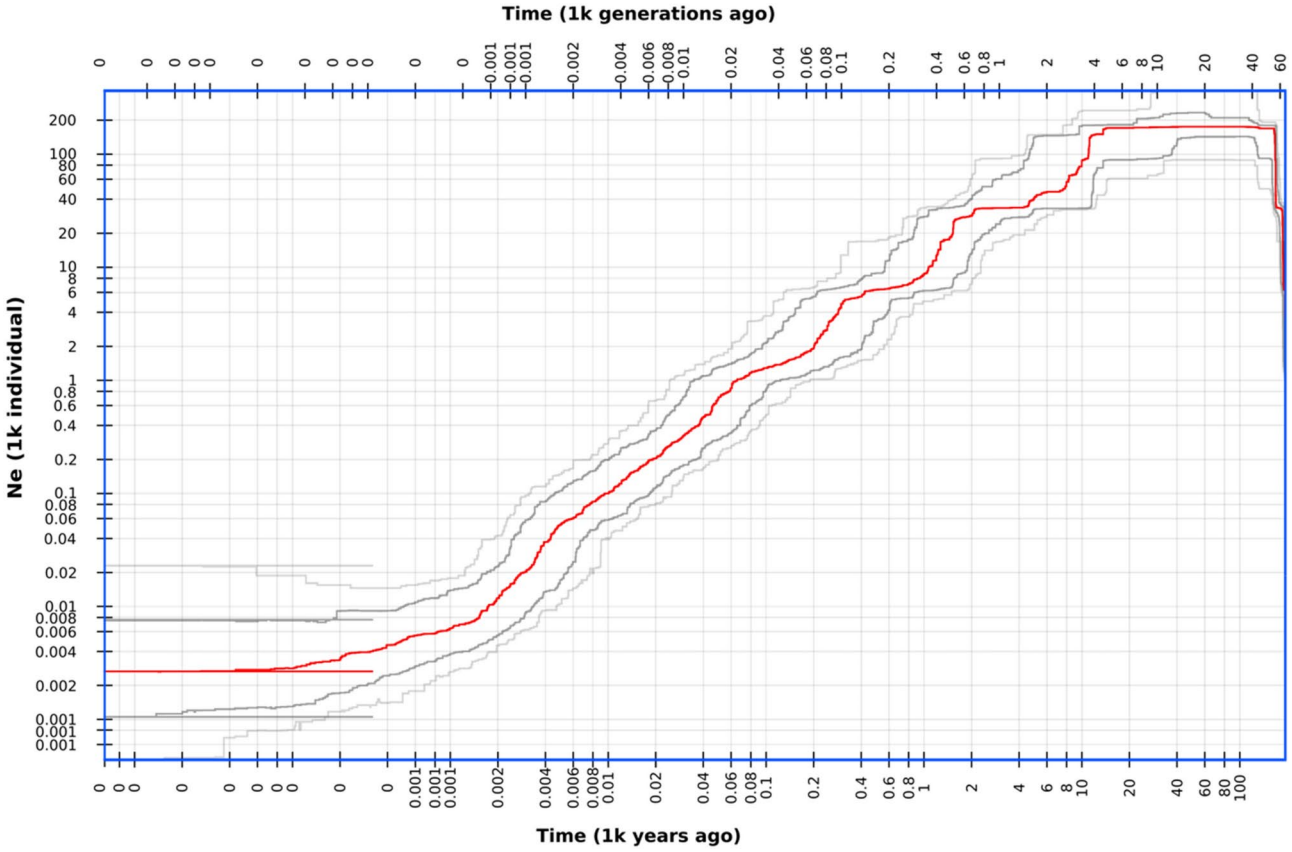

Fig. 5. Demographic history of *Diadema africanum* inferred by Stairway Plot 2. The red line represents the effective population size (N_e) estimated over time, and the grey lines the 75% and 95% intervals of confidence. Axes represent time (x) and effective population size (y), and both are logarithmically scaled.

levels of gene flow by differential selective pressures across the genome. The local adaptation patterns can turn some populations, such as those from Lanzarote, into unique evolutionary units. The existence of this genomic mosaic would claim the need of applying specific management strategies according to the characteristics of each geographical area. Nevertheless, further identification of the involved genes in this adaptation would be needed to delve deeper into the biological, phenotypical, and/or physiological impacts of these genomic particularities. We should also keep in mind that only a small representation of the whole genome of *D. africanum* has been analysed, and therefore, other genome regions related to adaptation to environmental variables could be relevant to understand the genomic architecture of this species. Finally, we cannot anticipate how adaptation to local conditions of temperature could limit the recovery decimated populations, an interesting question that should be addressed in further studies.

Methods

Mitochondrial DNA gene sequencing and analyses

A total of 249 individuals of *Diadema africanum*, from 11 sites in the Canary Islands and 3 sites in Cabo Verde, were collected between 2016 and 2017 (Fig. 1 and Table 1), before the 2018 mass mortality event. Small pieces of gonad were extracted, immersed in RNAlater Stabilization Solution (Thermo-Fisher Scientific, USA), and preserved at − 20 °C in the laboratory until DNA extraction. Total DNA was extracted from each gonad sample using the "Speedtools tissue DNA extraction kit" from Biotools (www.biotools.eu), following the manufacturer's protocol. A fragment of the mitochondrial gene cytochrome c oxidase I (COI) was amplified using specific primers designed from the complete mitochondrial genome of *Diadema setosum* (Acc. number KX385835.1 available from GenBank). Primers were designed with the software Geneious v 8.0.4 as follows: forward primer Di_311F- TAGTGCCCCCTTCCTTTATT, and reverse primer Di_1205R- CCGGAGAATAAAGGAAACCA. PCR amplification products of the COI fragment were purified and sequenced by Macrogen, Inc. (Seoul, Korea) using the same primers. COI sequences obtained were edited, aligned, and translated into amino acids to exclude pseudogenes, with Geneious v.2022.1 (www.geneious.com).

Ethics statement

No endangered or protected species were involved in this study. Authors possessed the required permits to collect echinoderm samples for research.

General genetics descriptors and divergence among populations from the COI

From the COI sequences, genetic diversity descriptors such as number of haplotypes (Nh), haplotype diversity (Hd), nucleotide diversity (π) and number of polymorphic sites (S) were computed using DnaSP v.6.12.03[61]. An ANOVA[62] statistical test was performed in Rstudio, to compare the haplotype and nucleotide diversity between Cabo Verde and the Canary Islands. The software CONTRIB 1.02[63] was used to calculate the haplotype richness (Hk) after rarefaction (adjusted to the minimum sample size) per population to characterise diversity levels within and between sites and to compare populations with different sampling sizes. The same software, DnaSP, was also used to identify non-synonymous substitutions among haplotypes. To define the relationships among the haplotypes, a network was constructed with Popart v.1.7 (Population Analysis with Reticulate Tree)[64], implementing a median-joining method[65].

To examine genetic divergence among sites, an analysis of the molecular variance (AMOVA)[66] was performed using Arlequin v.3.5.2.2[67], with a total of 10,000 permutations. The test was performed at different levels: grouping sampling sites in two archipelagos (Canary Islands and Cabo Verde) to test potential genetic subdivision between the two archipelagos, among sampling sites within archipelagos, among individuals within sampling sites. Additionally, pairwise distances (F_{ST}) between sampling sites were calculated to estimate genetic differentiation and using 10,000 random permutations to obtain their associated p-values in Arlequin. Significance thresholds were set at p-value < 0.05 after applying a Benjamini-Yekutieli FDR corrections[68], according to the number of comparisons between sampling sites.

Past demography from COI

To detect recent demographic changes from the COI, we calculated different neutrality tests, including the Tajima's D[69], Fu's Fs[70] and the Ramos-Onsins and Rozas[71] to infer potential departures from population equilibrium in DnaSP. To better understand the demographic history of the studied regions, a mismatch distribution analysis was performed. The demographic changes were examined by calculating Harpending's raggedness index[72] and the sum of squared deviations (SSD) between the observed and expected mismatch for each one of the sites. The spatial expansion hypothesis (both raggedness index and SSD) was tested using a parametric bootstrap approach (1000 replicates). These last tests were calculated using the methods of Schneider and Excoffier[73] in Arlequin. This measure quantifies the smoothness of the observed mismatch distribution, and a non-significant result indicates an expanding population[72].

To determine the estimated time of the possible demographic expansion (t), the formula $\tau = 2\mu kt$[74] was applied, with τ as the mode of mismatch distribution, μ as the mutation rate per nucleotide, k as the number of nucleotides assayed and t the time since population expansion[22]. The mutation rates used were the ones used on *Diadema* species from[49] with values of either 1.2×10^{-5} or 1.9×10^{-5} substitutions per site per millennium.

ddRADseq libraries construction, isolation of markers and analyses

DNA quality, ddRADseq library construction and single nucleotide polymorphisms (SNPs) isolation

Individuals used for the COI sequencing, collected in 2016 and 2017, could not be further used for ddRADseq library construction due to their low quality of the nuclear DNA. For this reason, 243 new individuals of *D. africanum* were collected from 13 sampling sites in the Canary Islands in 2021, after the 2018 mass mortality event. Due to the low resolution of the COI for population genetics in this species (see the Results sections), no sequencing of the COI was performed for the newly recollected samples. For tissue collection, a small portion of the muscular tissue from Aristotle's lantern was preserved in absolute ethanol at -20ºC.

Genomic DNA was extracted with a CTAB protocol (Hexadecyl Trimethyl Ammonium Bromide)[75]. DNA quality, purity and quantity were controlled in agarose electrophoresis gels, Nanodrop (Thermo Fisher, www.thermofisher.com), and using a Qubit DNA BR assay (Life Technologies, www.thermofisher.com), respectively. From the 243 high-quality extractions, we made double digest restriction site-associated DNA (ddRADseq) libraries from an initial amount of 500 ng of total DNA, with the restriction enzymes *EcoRI* and *MseI* (New England Biolabs, www.neb.com) following a described protocol[76]. Briefly, after double enzymatic digestion and DNA clearing, individuals were barcoded and pooled in 10 libraries that were later amplified, and sequenced in a NovaSeq 6000 (Illumina Inc., San Diego, CA) at the Novogene (https://www.novogene.com) generating 150 bp paired-end reads. See more details of the ddRADseq library construction by Leiva and coauthors[6].

The Stacks software v. 2.59[77] was used to process the sequence data with de novo analysis without a reference genome. The module process_radtags was used to demultiplex and trim raw reads to remove low-quality and unidentifiable reads. The ustacks program was used to assemble the matching reads per individual and call SNPs at each locus, with a maximum distance allowed of 4 (-M) and a minimum of 3 matching reads to create a stack (-m). The cstacks program was used to build a catalogue of consensus loci with 4 mismatches allowed between sample loci (-n). Putative loci were searched against the catalogue using sstacks program and paired-end reads were associated with each single-end locus. The module populations was conducted retaining SNPs present in at least 75% of the individuals, with a minimum allele frequency of 5%. From this dataset, SNPs that presented either significantly greater observed than expected heterozygosity (Hardy Weinberg Equilibrium-HWE, p-value < 0.01), which can be interpreted as a sign of sequencing errors, or significant linkage disequilibrium (p-value < 0.01) in all sampling sites, were removed for further analyses.

Identifying candidate ddRADseq-derived SNPs under selection

The total SNPs dataset was analysed using two methods to detect candidate loci under selection: (I) a Bayesian analysis implemented in BayeScan v. 2.1[78] with 100,000 iterations and 50% burn-in without including any environmental variable; and (II) a Redundancy Analysis (RDA)[79] using the sea surface temperature (average temperature over the last 28 years) as an environmental variable (Supplementary material, Fig. S4). Other marine environmental variables such as salinity and pH did not show differences across the study area. The

"RDA" function was performed in the R package "vegan" v. 2.5-6[80]. The seawater temperatures were obtained from the E.U. Copernicus Marine Service information (product ID: MULTIOBS_GLO_PHY_TSUV_3D_MYNRT_015_012, https://doi.org/https://doi.org/10.48670/moi-00052). This approach allows us to detect potential SNPs under selection to unknown variables (with BayeScan), and specifically under selection to temperature (with RDA).

Using the information of potential SNPs under selection from both methods, we created two datasets: a dataset of SNPs excluding all candidate SNPs under selection, and therefore, considered as a neutral dataset (neutral SNPs), and a dataset including candidate SNPs under selection to temperature, obtained from RDA (candidate SNPs). Hence, most analyses explained below were computed for both datasets separately to highlight differences in structure between datasets as a sign of local adaptation.

Genomic diversity and divergence for the ddRADseq-derived SNPs

Genomic diversity from neutral and candidate SNPs including observed and expected heterozygosity (Ho and He, respectively), was calculated using GenoDive v. 3.0[81], and allelic richness per locus per sampling site using the R package "PopGenReport" 3.0.7[82] (Table 2). The F_{IS} statistic was calculated in Arlequin, with 10,000 permutations to obtain their associated *p*-values. The relation between the population size and the number of alleles per sampling site was plotted with the package "ggplot2" in RStudio to check if the population size of our sampling was enough to represent the genomic diversity within the sites (data not shown).

To explore whether the 2018 mass mortality event affected the genomic diversity of *D. africanum* in the Canary Islands, we applied a two-tailed Student's t-test[83] in RStudio comparing values of diversity (allelic richness and observed heterozygosity from the neutral SNPs dataset) between sampling sites affected and unaffected by this mortality event. We used the neutral SNPs dataset considering that mass mortality events equally influence the whole genome of the species.

To investigate genomic divergence among sampling sites of *D. africanum* in the Canary Islands, we applied different methods such as genomic distances, principal component analyses and Bayesian clustering. These analyses were separately applied to the neutral and candidate SNPs datasets, except for an AMOVA performed only with the neutral dataset (for comparison with the COI data). For the AMOVA, we grouped sampling sites in islands, and tested differences among islands, among sampling sites within islands and among individuals within sampling sites. The analysis was performed in Arlequin with 10,000 permutations. Genomic distances among sampling sites were calculated using the pairwise F_{ST} statistics in the same software, with also 10,000 permutations to obtain their associated *p*-values. Significance thresholds were set at p-value < 0.05 after applying a B-Y FDR correction. The results were plotted with the r function "heatmap" in the "ggplot2" package for a better visualisation of them. Relative migration between sites was estimated for the neutral SNPs dataset using the divMigrate function of the diversity R package[84], based on the Nei's G_{ST} distance-method, and setting the filter threshold to values > 0.7 to interpret levels of connectivity among sampling sites. Discriminant analyses of principal components (DAPCs) were also carried out using the "adegenet" R package[85] using sites *as prior* grouping, to detect clusters of genetically related individuals. The function "find.clusters" was used to infer genetic clusters and the function "dapc" to determine the number of PCs and the diversity between sites. Bayesian clustering analyses were performed with the software Structure v. 2.3.4[86] to investigate the most likely number and distribution of genomic clusters (K). An initial run was performed with K from 2 to 14 with two replicates, 150,000 Markov chain Monte Carlo (MCMC) per replicate and a 50,000 burn-in period. Once, the parameters were stabilised and confirmed, final runs were performed with K from 2 to 14 with five replicates, 200,000 MCMC per replicate and a 50,000 burn-in period. The most likely K was calculated by evaluating the rate of change in the likelihood of K, using the ad hoc statistic ΔK with the Evanno method[87] in Structure Harvester[88].

Past demographic events from SNPs data

Past demographic events were explored using the neutral SNPs dataset. Potential fluctuations in the past effective population size were inferred with the Stairway Plot v. 2.1.1[89], which detects historical changes using a flexible multi-epoch demographic model. The folded site frequency spectrum (SFS) for each sampling site was built with the script easySFS.py (https://github.com/isaacovercast/easySFS) from the VCF file obtained in the population module of Stacks. The parameter settings were as follows: 67% of sites for training, 3 years per generation (estimated for this species), and random breakpoints were set as 67, 134, 201, 268. Based on the output from Stacks, the total number of observed nucleic sites was set as 3,307,887. Since there are no estimations of the nuclear mutation rate for the *Diadema* nuclear genome or any closely related species, we used the nuclear mutation rate of the sea star *Acanthaster* spp. (8.8×10^{-9}, 9.4×10^{-9}, and 9.9×10^{-9})[90]. We performed the analyses applying separately the three mutation rates, and as no large differences were observed, we finally show in this study the ones associated with the 9.4×10^{-9}. Each estimation was generated using 200 bootstrapped folded SFS.

Data availability

The COI haplotypes are available on NCBI GeneBank (accession numbers: OP494354-OP494596) and Illumina raw reads on the SRA database (BioProject ID PRJNA1065271; Biosamples SAMN39456622-SAMN39456864).

Received: 11 March 2024; Accepted: 17 September 2024
Published online: 28 September 2024

References

1. Santangelo, J. S., Johnson, M. T. J. & Ness, R. W. Modern spandrels: The roles of genetic drift, gene flow and natural selection in the evolution of parallel clines. *Philos. Trans. R. Soc. B: Biol. Sci.***285**, 20180230 (2018).

2. Star, B. & Spencer, H. G. Effects of genetic drift and gene flow on the selective maintenance of genetic variation. *Genetics***194**, 235–244 (2013).
3. Cowen, R. K. & Sponaugle, S. Larval dispersal and marine population connectivity. *Annu. Rev. Mar. Sci.***1**, 443–466 (2009).
4. Thompson, D. M. *et al.* Variability in oceanographic barriers to coral larval dispersal: Do currents shape biodiversity?. *Prog. Oceanogr.***165**, 110–122 (2018).
5. Peluso, L. *et al.* Contemporary and historical oceanographic processes explain genetic connectivity in a Southwestern Atlantic coral. *Sci. Rep.***8**, 2684 (2018).
6. Leiva, C., Pérez-Portela, R. & Lemer, S. Genomic signatures suggesting adaptation to ocean acidification in a coral holobiont from volcanic CO2 seeps. *Commun. Biol.***6**, 769 (2023).
7. Carreras, C. *et al.* The two sides of the Mediterranean: Population genomics of the black sea urchin *Arbacia lixula* (Linnaeus, 1758) in a warming sea. *Front. Mar. Sci*.**8**, 739008 (2021).
8. Carreras, C. *et al.* East is East and West is West: Population genomics and hierarchical analyses reveal genetic structure and adaptation footprints in the keystone species *Paracentrotus lividus* (Echinoidea). *Divers. Distrib.***26**, 382–398 (2020).
9. He, Q. *et al.* Networks of genetic similarity reveal non-neutral processes shape strain structure in Plasmodium falciparum. *Nat. Commun.***9**, 1817 (2018).
10. Baselga, A., Gómez-Rodríguez, C. & Vogler, A. P. Multi-hierarchical macroecology at species and genetic levels to discern neutral and non-neutral processes. *Glob. Ecol. Biogeogr.***24**, 873–882 (2015).
11. Haddrill, P. R. & Charlesworth, B. Non-neutral processes drive the nucleotide composition of non-coding sequences in *Drosophila*. *Biol. Lett.***4**, 438–441 (2008).
12. Ferchaud, A. L. & Hansen, M. M. The impact of selection, gene flow and demographic history on heterogeneous genomic divergence: Three-spine sticklebacks in divergent environments. *Mol. Ecol.***25**, 238–259 (2016).
13. Schreiber, D. & Pfenninger, M. Genomic divergence landscape in recurrently hybridizing chironomus sister taxa suggests stable steady state between mutual gene flow and isolation. *Evol. Lett.***5**, 86–100 (2021).
14. Knowles, L. L. & Richards, C. L. Importance of genetic drift during Pleistocene divergence as revealed by analyses of genomic variation. *Mol. Ecol.***14**, 4023–4032 (2005).
15. Frankham, R., Ballou, J. D., Briscoe, D. A. & McInnes, K. H. *Introduction to Conservation Genetics* (Cambridge University Press, 2002).
16. Díaz, A. *et al.* Genetic structure and demographic inference of the regular sea urchin *Sterechinus neumayeri* (Meissner, 1900) in the Southern Ocean: The role of the last glaciation. *PLoS One***13**, e0197611 (2018).
17. Casilagan, I. L. N., Juinio-Meñez, M. A. & Crandall, E. D. Genetic diversity, population structure, and demographic history of exploited sea urchin populations (*Tripneustes gratilla*) in the Philippines. *J. Exp. Mar. Biol. Ecol.***449**, 284–293 (2013).
18. Duran, S., Palacín, C., Becerro, M. A., Turon, X. & Giribet, G. Genetic diversity and population structure of the commercially harvested sea urchin *Paracentrotus lividus* (Echinodermata, Echinoidea). *Mol. Ecol.***13**, 3317–3328 (2004).
19. Chandler, L. M., Walters, L. J., Sharp, W. C. & Hoffman, E. A. Genetic structure of natural & broodstock populations of the long-spined sea urchin, *Diadema antillarum*, throughout the Florida Keys. *Bull. Mar. Sci.***93**, 881–889 (2017).
20. Banks, S. C. *et al.* Oceanic variability and coastal topography shape genetic structure in a long-dispersing sea urchin. *Ecology***88**, 3055–3064 (2007).
21. Pérez-Portela, R. *et al.* Spatio-temporal patterns of genetic variation in Arbacia lixula, a thermophilous sea urchin in expansion in the Mediterranean. *Heredity***122**, 244–259 (2019).
22. Wangensteen, O. S., Turon, X., Pérez-Portela, R. & Palacín, C. Natural or naturalized? Phylogeography suggests that the abundant sea urchin *Arbacia lixula* is a recent colonizer of the Mediterranean. *PLoS One***7**, e45067 (2012).
23. Coppard, S. E. & Campbell, A. C. Distribution and abundance of regular sea urchins on two coral reefs in Fiji. *Micronesica***37**, 249–269 (2005).
24. Lessios, H. A., Kessing, B. D. & Pearse, J. S. Population structure and speciation in tropical seas: Global phylogeography of the sea urchin *Diadema*. *Evolution***55**, 955–975 (2001).
25. Hernández, J. C., Clemente, S., Sangil, C. & Brito, A. The key role of the sea urchin *Diadema* aff. *antillarum* in controlling macroalgae assemblages throughout the Canary Islands (eastern subtropical Atlantic): An spatio-temporal approach. *Mar. Environ. Res.***66**, 259–270 (2008).
26. Prado, P. *et al.* Habitat and scale shape the demographic fate of the keystone sea urchin *Paracentrotus lividus* in Mediterranean macrophyte communities. *PLoS One***7**, e35170 (2012).
27. Ling, S. D. *et al.* Global regime shift dynamics of catastrophic sea urchin overgrazing. *Philos. Trans. R. Soc. B: Biol. Sci.***370**, 1–10 (2015).
28. Uthicke, S., Schaffelke, B. & Byrne, M. A boom-bust phylum? Ecological and evolutionary consequences of density variations in echinoderms. *Ecol. Monogr.***79**, 3–24 (2009).
29. Lessios, H. A. Mass mortality of *Diadema antillarum* in the Caribbean: What have we learned?. *Annu. Rev. Ecol. Syst.***19**, 371–393 (1988).
30. Hernández, J. C., Sangil, C. & Lorenzo-Morales, J. Uncommon southwest swells trigger sea urchin disease outbreaks in Eastern Atlantic archipelagos. *Ecol. Evol.***10**, 7963–7970 (2020).
31. Hylkema, A. *et al.* The 2022 *Diadema antillarum* die-off event: Comparisons with the 1983–1984 mass mortality. *Front. Mar. Sci.***9**, 1067449 (2023).
32. Zirler, R. *et al.* Mass mortality of the invasive alien echinoid *Diadema setosum* (Echinoidea: Diadematidae) in the Mediterranean Sea. *R. Soc. Open. Sci.***10**, 230251 (2023).
33. Wallner-Hahn, S. *et al.* Cascade effects and sea-urchin overgrazing: An analysis of drivers behind the exploitation of sea urchin predators for management improvement. *Ocean Coast Manag.***107**, 16–27 (2015).
34. Kriegisch, N., Reeves, S. E., Johnson, C. R. & Ling, S. D. Top-down sea urchin overgrazing overwhelms bottom-up stimulation of kelp beds despite sediment enhancement. *J. Exp. Mar. Biol. Ecol.***514–515**, 48–58 (2019).
35. Sangil, C. & Hernández, J. C. Recurrent large-scale sea urchin mass mortality and the establishment of a long-lasting alternative macroalgae-dominated community state. *Limnol. Oceanogr.***67**, S430–S443 (2022).
36. Hernández, J. C., Clemente, S., García, E. & McAlister, J. S. Planktonic stages of the ecologically important sea urchin, *Diadema africanum*: larval performance under near future ocean conditions. *J. Plankton Res.***42**, 286–304 (2020).
37. John, D. M., Lieberman, D. & Lieberman, M. A Quantitative study of the structure and dynamics of benthic subtidal algal vegetation in Ghana (Tropical West Africa). *J. Ecol.***65**, 497–521 (1977).
38. Minderlein, R. & Wirtz, P. A sea urchin (*Diadema africanum* Rodriguez et al., 2013) and a pipe fish (*Syngnathus phlegon* Risso, 1827): Two new records for the Azores. *Arquipélago,***31.** (2014).
39. Clemente, S., Hernández, J. C., Rodríguez, A. & Brito, A. Identifying keystone predators and the importance of preserving functional diversity in sublittoral rocky-bottom areas. *Mar. Ecol. Prog. Ser.***413**, 55–67 (2010).
40. Hernández, J. C., Clemente, S., Girard, D., Pérez-Ruzafa, Á. & Brito, A. Effect of temperature on settlement and postsettlement survival in a barrens-forming sea urchin. *Mar. Ecol. Prog. Ser.***413**, 69–80 (2010).
41. Sangil, C. *et al.* Spatial variability, structure and composition of crustose algal communities in *Diadema africanum* barrens. *Helgol. Mar. Res.***68**, 451–464 (2014).
42. Carrillo, M. & Cruz, T. *Estudio de las Comunidades vegetales Marinas y Poblaciones Faunísticas del Litoral del Parque Nacional de Timfaya*. 153–223 (Servicio de Publicaciones de la Caja General de Ahorros de Canarias, 1992).

43. Harvell, C. D. *et al.* Climate warming and disease risks for terrestrial and marine biota. *Science***296**, 2158–2162 (2002).
44. Scheibling, R. E. & Lauzon-Guay, J. S. Killer storms: North Atlantic hurricanes and disease outbreaks in sea urchins. *Limnol. Oceanogr.***55**, 2331–2338 (2010).
45. Salazar-Forero, C. E., Reyes-Batlle, M., González-Delgado, S., Lorenzo-Morales, J. & Hernández, J. C. Influence of winter storms on the sea urchin pathogen assemblages. *Front. Mar. Sci.***9**, 812931 (2022).
46. Gizzi, F. *et al.* Before and after a disease outbreak: Tracking a keystone species recovery from a mass mortality event. *Mar. Environ. Res.***156**, 104905 (2020).
47. Clemente, S. *et al.* Sea urchin *Diadema africanum* mass mortality in the subtropical eastern Atlantic: Role of waterborne bacteria in a warming ocean. *Mar. Ecol. Prog. Ser.***506**, 1–14 (2014).
48. Leberg, P. L. Effects of population bottlenecks on genetic diversity as measured by allozyme electrophoresis. *Evolution***46**, 477–494 (1992).
49. Lessios, H. A., Garrido, M. J. & Kessing, B. D. Demographic history of *Diadema antillarum*, a keystone herbivore on Caribbean reefs. *Proc. R. Soc. B Biol. Sci.***268**, 2347–2353 (2001).
50. Borrero-Pérez, G. H., González-Wangüemert, M., Marcos, C. & Pérez-Ruzafa, A. Phylogeography of the Atlanto-Mediterranean sea cucumber *Holothuria (Holothuria) mammata*: The combined effects of historical processes and current oceanographical pattern. *Mol. Ecol.***20**, 1964–1975 (2011).
51. Pérez-Portela, R., Rius, M. & Villamor, A. Lineage splitting, secondary contacts and genetic admixture of a widely distributed marine invertebrate. *J. Biogeogr.***44**, 446–460 (2017).
52. Taboada, S. & Pérez-Portela, R. Contrasted phylogeographic patterns on mitochondrial DNA of shallow and deep brittle stars across the Atlantic-Mediterranean area. *Sci. Rep.***6**, 32425 (2016).
53. Leiva, C. *et al.* Exceptional population genomic homogeneity in the black brittle star *Ophiocomina nigra* (Ophiuroidea, Echinodermata) along the Atlantic-Mediterranean coast. *Sci. Rep.***13**, 12349 (2023).
54. Landeira, J. M., Lozano-Soldevilla, F., Hernández-León, S. & Desmond Barton, E. Horizontal distribution of invertebrate larvae around the oceanic island of Gran Canaria: The effect of mesoscale variability. *Sci. Mar.***73**, 761–771 (2009).
55. Calderón, I., Pita, L., Brusciotti, S., Palacín, C. & Turon, X. Time and space: Genetic structure of the cohorts of the common sea urchin *Paracentrotus lividus* in Western Mediterranean. *Mar. Biol.***159**, 187–197 (2012).
56. Rodríguez, A., Hernández, J. C., Clemente, S. & Coppard, S. E. A new species of *Diadema* (Echinodermata: Echinoidea: Diadematidae) from the eastern Atlantic Ocean and a neotype designation of *Diadema antillarum* (Philippi, 1845). *Zootaxa***3636**, 144–170 (2013).
57. Willi, Y., Fracassetti, M., Bachmann, O. & van Buskirk, J. Demographic processes linked to genetic diversity and positive selection across a species' range. *Plant. Commun.***1**, 100111 (2020).
58. Ben-Shlomo, R., Paz, G. & Rinkevich, B. Postglacial-period and recent invasions shape the population genetics of botryllid ascidians along European Atlantic coasts. *Ecosystems***9**, 1118–1127 (2006).
59. Maggs, C. A. *et al.* Evaluating signatures of glacial refugia for North Atlantic benthic marine taxa. *Ecology***89**, S108–S122 (2008).
60. Garcia-Cisneros, A., Palacín, C., Ben Khadra, Y. & Pérez-Portela, R. Low genetic diversity and recent demographic expansion in the red starfish *Echinaster sepositus* (Retzius 1816). *Sci. Rep.***6**, 33269 (2016).
61. Rozas, J. *et al.* DnaSP 6: DNA sequence polymorphism analysis of large data sets. *Mol. Biol. Evol.***34**, 3299–3302 (2017).
62. Stahle, L. & Wold, S. Analysis of variance (ANOVA). *Chemom. Intell. Lab. Syst.***6**, 259–272 (1989).
63. Petit, R. J., El Mousadik, A. & Pons, O. Identifying populations for conservation on the basis of genetic markers. *Conserv. Biol.***12**, 844–855 (1998).
64. Leigh, J. W. & Bryant, D. POPART: Full-feature software for haplotype network construction. *Methods Ecol. Evol.***6**, 1110–1116 (2015).
65. Bandelt, H.-J., Forster, P. & Röhl, A. Median-joining networks for inferring intraspecific phylogenies. *Mol. Biol. Evol.***16**, 37–48 (1999).
66. Excoffier, L., Smouse, P. E. & Quattro, J. M. Analysis of molecular variance inferred from metric distances among DNA haplotypes: Application to human mitochondrial DNA restriction data. *Genetics***131**, 479–491 (1992).
67. Excoffier, L. & Lischer, H. E. L. Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Mol. Ecol. Resour.***10**, 564–567 (2010).
68. Benjamini, Y. & Yekutieli, D. The control of the false Discovery rate in multiple testing under dependency. *Ann. Stat.***29**, 1165–1188 (2001).
69. Tajima, F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics***123**, 585–595 (1989).
70. Fu, Y. X. & Li, W. H. Statistical tests of neutrality of mutations. *Genetics***133**, 693–709 (1993).
71. Ramos-Onsins, S. E. & Rozas, J. Statistical properties of new neutrality tests against population growth. *Mol. Biol. Evol***19**, 2092–2100 (2002).
72. Harpending, H. Signature of ancient population-growth in a low-resolution mitochondrial DNA mismatch distribution. *Hum. Biol.***66**, 591–600 (1994).
73. Schneider, S. & Excoffier, L. Estimation of past demographic parameters from the distribution of pairwise differences when the mutation rates vary among sites: Application to human mitochondrial DNA. *Genetics***152**, 1079–1089 (1999).
74. Rogers, A. R. & Harpending, H. Population growth makes waves in the distribution of pairwise genetic differences. *Mol. Biol. Evol***9**, 552–569 (1992).
75. Doyle, J. J. & Doyle, J. L. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem. Bull.***19**, 11–15 (1987).
76. Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S. & Hoekstra, H. E. Double digest RADseq: An inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PLoS One***7**, e37135 (2012).
77. Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A. & Cresko, W. A. Stacks: An analysis tool set for population genomics. *Mol. Ecol.***22**, 3124–3140 (2013).
78. Foll, M. & Gaggiotti, O. A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: A Bayesian perspective. *Genetics***180**, 977–993 (2008).
79. Forester, B. R., Lasky, J. R., Wagner, H. H. & Urban, D. L. Comparing methods for detecting multilocus adaptation with multivariate genotype–environment associations. *Mol. Ecol.***27**, 2215–2233 (2018).
80. Oksanen, J. *et al.* Vegan community ecology package version 2.5–6. (2019).
81. Meirmans, P. G. genodive version 3.0: Easy-to-use software for the analysis of genetic data of diploids and polyploids. *Mol. Ecol. Resour.***20**, 1126–1131 (2020).
82. Adamack, A. T. & Gruber, B. PopGenReport: Simplifying basic population genetic analyses in R. *Methods Ecol. Evol.***5**, 384–387 (2014).
83. Student. The probable error of a mean. *Biometrika* 1–25 (1908).
84. Keenan, K., Mcginnity, P., Cross, T. F., Crozier, W. W. & Prodöhl, P. A. DiveRsity: An R package for the estimation and exploration of population genetics parameters and their associated errors. *Methods Ecol. Evol.***4**, 782–788 (2013).
85. Jombart, T. & Ahmed, I. adegenet 1.3–1: New tools for the analysis of genome-wide SNP data. *Bioinformatics***27**, 3070–3071 (2011).
86. Raj, A., Stephens, M. & Pritchard, J. K. fastSTRUCTURE: variational inference of population structure in large SNP data sets. *Genetics***197**, 573–589 (2014).

87. Evanno, G., Regnaut, S. & Goudet, J. Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. *Mol. Ecol.***14**, 2611–2620 (2005).
88. Earl, D. A. & VonHoldt, B. M. STRUCTURE HARVESTER: A website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conserv. Genet. Resour.***4**, 359–361 (2012).
89. Liu, X. & Fu, Y. X. Stairway Plot 2: demographic history inference with folded SNP frequency spectra. *Genome Biol.***21**, 280 (2020).
90. Yuasa, H. *et al.* Elucidation of the speciation history of three sister species of crown-of-thorns starfish (*Acanthaster* spp.) based on genomic analysis. *DNA Res.***28**, 012 (2021).

Acknowledgements

The authors thank Beatriz Alonzo and Carlos Sangil for helping during the sampling collection, Jorge Rodriguez López-Rey for helping during the DNA extractions, and Carlos Leiva for helping with the demographic analyses. G.R. thanks the Marine Environment Program (MER+). This research was funded by different projects from the Spanish Government: ADAPTIVE- PGC2018-100735-B-I00 (MCIU/AEI/FEDER/ UE), ENVIOME (PID2021-128094NB-I00/MCIN/AEI/https://doi.org/10.13039/501100011033/ and FEDER una manera de hacer Europa), and a "Ramón y Cajal" contract to R.P-P. (RYC2018-025070-I), and the project "DIVERGEN- Ayudas Fundación BBVA a Proyectos Investigación Científica 2021". This paper is a contribution of the Consolidated Research Team: 2021 SGR 01271 Marine Biodiversity and Evolution (MBE).

Author contributions

J.C.H., S.G-D, and R.P-P designed this study. J.C.H., S.G-D, and E.P.L collected all the samples. M.P-S, L.P-S., R.G, and R.P-P performed the analyses. M.P-S, R.G., and R.P-P wrote the first draft of the manuscript, and all the other authors improved this first draft. All authors agreed with the final version of this manuscript.

Funding

Spanish Government projects: ADAPTIVE- PGC2018-100735-B-I00 (MCIU/AEI/FEDER. UE), ENVIOME (PID2021-128094NB-I00/MCIN/AEI/https://doi.org/10.13039/501100011033/ and FEDER una manera de hacer Europa), and a "Ramón y Cajal" gran to R.P-P. (RYC2018-025070-I), and the project "DIVERGEN- *Ayudas Fundación BBVA a Proyectos Investigación Científica 2021*".

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-024-73354-3.

Correspondence and requests for materials should be addressed to R.P.-P.