Author for correspondence:
E. M. Salas
e-mail: salas.e@gmail.com

RADseq analyses reveal concordant Indian Ocean biogeographic and phylogeographic boundaries in the reef fish *Dascyllus trimaculatus*

E. M. Salas[1,2], G. Bernardi[2], M. L. Berumen[3], M. R. Gaither[4] and L. A. Rocha[1]

[1]Section of Ichthyology, California Academy of Sciences, 55 Music Concourse Drive, San Francisco, CA 94118, USA
[2]Department of Ecology and Evolutionary Biology, University of California Santa Cruz, 100 Shaffer Rd, Santa Cruz, CA 95060, USA
[3]Red Sea Research Center, Biological and Environmental Science and Engineering Division, King Abdullah University of Science and Technology, Thuwal, Saudi Arabia
[4]Department of Biology, Genomics and Bioinformatics Cluster, University of Central Florida, 4110 Libra Drive, Orlando, FL 32816, USA

EMS, 0000-0002-6941-0548; GB, 0000-0002-8249-4678; MLB, 0000-0003-2463-2742; MRG, 0000-0002-0371-5621

Population genetic analysis is an important tool for estimating the degree of evolutionary connectivity in marine organisms. Here, we investigate the population structure of the three-spot damselfish *Dascyllus trimaculatus* in the Red Sea, Arabian Sea and Western Indian Ocean, using 1174 single nucleotide polymorphisms (SNPs). Neutral loci revealed a signature of weak genetic differentiation between the Northwestern (Red Sea and Arabian Sea) and Western Indian Ocean biogeographic provinces. Loci potentially under selection (outlier loci) revealed a similar pattern but with a much stronger signal of genetic structure between regions. The Oman population appears to be genetically distinct from all other populations included in the analysis. While we could not clearly identify the mechanisms driving these patterns (isolation, adaptation or both), the datasets indicate that population-level divergences are largely concordant with biogeographic boundaries based on species composition. Our data can be used along with genetic connectivity of other species to identify the common genetic breaks that need to be considered for the conservation of biodiversity and evolutionary processes in the poorly studied Western Indian Ocean region.

1. Introduction

Coral reefs are home to a diversity of fish species that originated despite the presence of few geographical barriers to larval dispersal [1]. To understand the origins of marine biodiversity, evolutionary biologists have often turned to the highly diverse Indo-Pacific realm (as defined by Kulbicki *et al.* [2]), a large biogeographic region spanning from the Red Sea to the Central Pacific. This realm includes many wide-ranging species that, upon close study, often show structured populations [3], revealing the evolutionary and historical mechanisms that operate in the oceans. These studies can also be relevant to marine resource management and conservation planning. For example, in Hawaii, assessing the connectivity of multiple species enabled the identification of common genetic breaks that can be used to delineate management units [4].

Over the past two decades, there have been numerous phylogeographic studies of Indo-Pacific reef fishes, but comparatively few have examined patterns of population connectivity within the Western Indian Ocean [5,6]. At the largest scale, and based on species distributions, this biogeographic region was divided into the Northwestern Indian Ocean (NWIO) and the Western Indian Ocean (WIO) provinces ([2], figure 1). In general, biogeographic boundaries are created by one of two mechanisms [7]: either the boundary is where two separate faunas meet, or the boundary itself is the engine of faunal separation. In the latter case, sister species are expected to occur across the boundary, and in populations that have not yet reached the stage of speciation, the biogeographic boundary coincides with a population separation—a phylogeographic break. The goal of this study was to assess phylogeographic patterns in *Dascyllus trimaculatus* and determine whether they matched established biogeographic boundaries.

In general, studies that have sampled populations in the Indian Ocean find results ranging from no genetic structure [5,8–15], to little differentiation [16–18]. However, some of these studies have limited geographical coverage. In some cases, strong genetic breaks between the Red Sea (a hotspot of endemism) and the Indian Ocean were observed [19,20]. Biogeographic studies on the distribution and evolutionary origin of endemic species show changes in species composition between the Red Sea, Gulf of Aden and Arabian Sea [21,22]. Within the WIO province, there appears to be genetic breaks between the Seychelles and populations on the African coast in some species, such as the parrotfish *Scarus ghobban* [18] and the mangrove crab *Neosarmatium meinerti* [23]. Some datasets also show evidence of a population genetic separation between the Chagos Archipelago and the rest of the WIO, for example in the butterflyfish *Chaetodon trifasciatus* [24]. Detailed studies of the widespread fish species *Epinephelus merra* and *Myripristis berndti* also showed evidence of structure within the WIO [25,26].

The three-spot damselfish, *D. trimaculatus* is an abundant species found throughout the Indo-Pacific [27], from the Red Sea to French Polynesia, and has many characteristics that are 'typical' of a coral reef dwelling damselfish. It spawns 2–3 times a month for several successive months [28] producing demersal eggs that are guarded for 2–3 days [29], and has a pelagic larval stage that lasts 22–26 days [30]. Juveniles generally settle on anemones, which they abandon once they are large enough to avoid predation, to find shelter in reef crevices nearby [27,29]. *Dascyllus trimaculatus* belongs to a species complex that comprises four species, *D. trimaculatus*, *D. albisella*, *D. strasburgi* and *D. auripinnis* [31]. The recent divergence of these species is consistent with a parapatric speciation scenario [31], as *D. trimaculatus* occupies nearly the entire Indo-Pacific range except the Hawaiian Archipelago, the Marquesas Islands, and the Line Islands, whereas the closely related *D. albisella*, *D. strasburgi* and *D. auripinnis* are only present in those restricted peripheral ranges, respectively [31]. *Dascyllus trimaculatus* populations show patterns of allopatric divergence—as the Pacific Ocean populations are distinct from those of the Indian Ocean, these populations were probably isolated by restricted water flow in the Sunda Shelf during the Pleistocene sea-level changes [31]. In the Pacific Ocean, where all four species of the complex are present, genetic studies have found a lack of congruence between colour morphs and genetic groups. For example, *D. auripinnis*, which has a bright yellow ventral surface and fins, is not fully genetically differentiated from *D. trimaculatus*. By contrast, only *D. trimaculatus* is present in the Indian Ocean and mitochondrial DNA (d-loop) and microsatellite comparisons showed these populations belong to a single clade [31].

Here, we use single nucleotide polymorphisms (SNPs) developed using restriction-site associated DNA sequencing (RADseq) to assess the population structure in Indian Ocean *D. trimaculatus*. Since recent biogeographic studies have found that in some species genetic boundaries match biogeographic regions in the Indian Ocean [32], and because *D. trimaculatus* shows moderate amounts of genetic structure throughout its range with evidence of divergence in the species complex in peripheral habitats [31], we considered the possibility of finding genetic divergence in peripheral areas where biogeographic boundaries have been proposed (such as the Red Sea). Not only did sea-level changes repeatedly isolate the Red Sea from the Indian Ocean during the Pleistocene [19], but currently, the
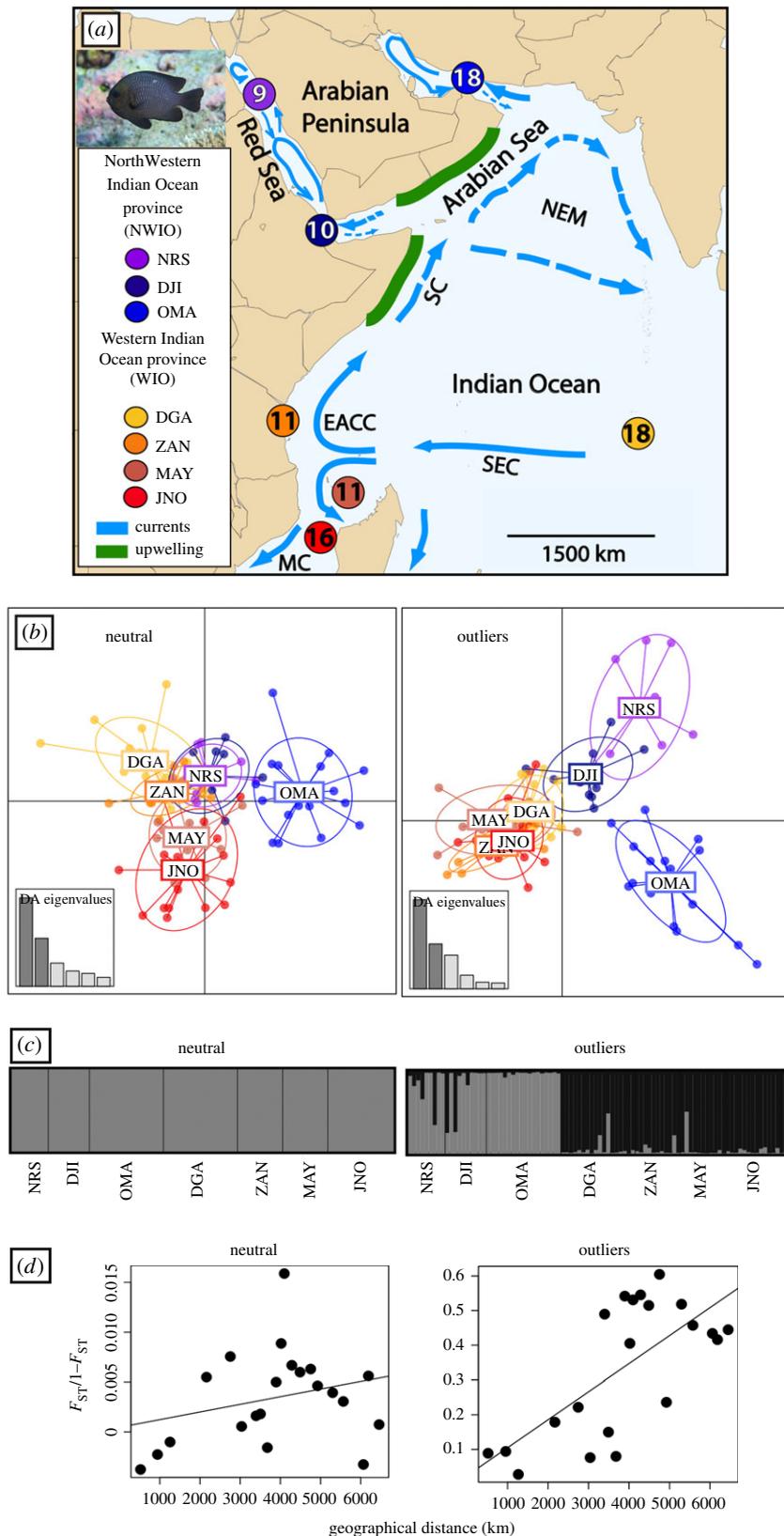
Figure 1. (a) Study region with populations and sample sizes (circles) indicated (NRS, Northern Red Sea; DJI, Djibouti; OMA, Oman; DGA, Diego de García, Chagos; ZAN, Zanzibar; MAY, Mayotte; JNO, Juan de Nova, Scattered Islands). Summer upwelling and currents are shown; dashed lines indicate winter reversals. Currents (C): NEM, North East Monsoon C; SC, Somali C; EACC, East African Coastal C; MC, Mozambique C; SEC, South Equatorial C. (b) DAPC for neutral (n = 1117) and outlier loci (n = 25). (c) STRUCTURE plot with most likely K for neutral (K = 1) and outlier loci (K = 2). (d) Isolation by distance (IBD) for neutral loci, mantel test p = 0.1990, r^2 = 0.0756, y = 8 × 10^{-7}x + 0.0005; and outlier loci, p = 0.016, r^2 = 0.4801, y = 8 × 10^{-5}x + 0.0239. *Dascyllus trimaculatus* picture by Tane Sinclair-Taylor.

presence of an upwelling along the Horn of Africa and the coast of Oman could further isolate Red Sea and Arabian Peninsula populations from those of the Indian Ocean [21]. In the Red Sea, the pelagic larval duration (PLD) of *D. trimaculatus*, *D. aruanus* and *D. marginatus* decreases as sea surface temperatures and food availability increase from north to south [33]. These results suggest that local environment variation affects the life history of *D. trimaculatus*, which in turn might influence its genetic structure. Currents near the African coast may also affect the genetic connectivity of *D. trimaculatus* in the Indian Ocean. The main oceanic current in the Indian Ocean is the South Equatorial Current (SEC, figure 1, [18]). This current splits near Tanzania into two continental currents, the East African Coastal Current (EACC) that flows north towards Somalia, and the Mozambique current that flows south and produces eddies in the Mozambique Channel [15,18], potentially isolating populations located in the Mozambique Channel from populations further north along the African coast.

In this study, we assessed differences between populations from the Red Sea, Arabian Peninsula, African coast, Mozambique Channel and Chagos Archipelago. We also tested for population divergence between the biogeographic NWIO and WIO provinces. To the best of our knowledge, this is one of the first studies based on extensive genomic data to compare population genetic structures in the Western Indian Ocean that simultaneously covers these two provinces.

2. Material and methods

2.1. Sample collection

A total of 93 individuals from seven populations were collected while diving or snorkelling between 1998 and 2013 (figure 1). Populations were grouped, following Kulbicki *et al.* [2], into Northwestern Indian Ocean Province, NWIO (Northern Red Sea: NRS, Djibouti: DJI and Oman: OMA); and Western Indian Ocean Province, WIO (Diego Garcia, Chagos Archipelago: DGA, Zanzibar, Tanzania: ZAN, Mayotte, Comoros Archipelago: MAY and Juan de Nova, Scattered Islands, Mozambique Channel: JNO, figure 1). The NRS population consisted of six individuals from Eilat, Israel and three from Jeddah, Saudi Arabia. Some of the samples were part of the earlier studies including individuals from Eilat, Oman, Mayotte and Zanzibar [31,34].

2.2. RADseq library preparation and sequencing

Genomic DNA was extracted using the Qiagen DNeasy animal blood and tissue kit (Qiagen, Valencia, USA). The library was prepared using the double-digest RADseq protocol [35], with modifications (see electronic supplementary material, methods) and sequenced on a single Illumina HiSeq 2000 lane, at the UCLA Neuroscience Genomics Core facility. Raw data were de-multiplexed, quality filtered and trimmed to 95 bp, using the 'process_rad_tags' script available in STACKS v. 1.09 [36]. Loci were assembled using the STACKS 'de novo_map.pl' pipeline, while the 'populations' script was used to filter loci and create output files (for raw data filtering and loci assembly see electronic supplementary material, methods). Loci were shared between the seven populations ($p = 7$), in at least 65% of individuals within a group ($r = 0.65$) and with a coverage of $8\times$ ($m = 8$). We used only the first SNP of each sequence and removed loci with minor allele frequencies lower than 1.5% (i.e. at least two individuals must have the unique allele). Our quality control and filtering resulted in a total of 1174 loci and a data matrix that was 84% complete. We used PGDSPIDER 2.0 [37] to convert the resulting STRUCTURE files into other formats.

2.3. Data analysis

First, we conducted population genetic analysis with all the loci ($n = 1174$). Then, to gain more perspective on the patterns of genetic structure and its potential mechanisms, we separated neutral from outlier loci, and repeated the population genetic analyses with the neutral subset and the subset of outliers that had higher than expected FST based on neutral expectations (see below).

To identify outlier loci, we used three methods. First, we used the modified FDIST approach [38] implemented in ARLEQUIN [39], which uses a hierarchical island model and simulates an F_{ST} null distribution across loci as a function of heterozygosity and determines outliers as being those outside of the distribution using a 99% confidence interval. We ran 50 000 simulations with 100 demes per group, with minimum and maximum expected heterozygosities of 0 and 0.5, respectively. To control for false positives, we adjusted probabilities by applying a false discovery rate of 0.01 [40] using the R

Table 1. F_{ST} values between populations, for neutral loci ($n = 1117$, below asterisks) and outlier loci ($n = 25$, above asterisks). Significant values ($p < 0.05$) are indicated in italics while significant values after sequential Bonferroni corrections are indicated in bolded.

	NRS	DJI	OMA	DGA	ZAN	MAY	JNO
NRS	***	**0.1522**[a]	**0.1919**	**0.2944**	**0.3139**	**0.3030**	**0.3080**
DJI	*0.0055*	***	**0.1825**	**0.2889**	**0.3292**	**0.3516**	**0.3533**
OMA	*0.0046*	**0.0075**	***	**0.3469**	**0.3399**	**0.3768**	**0.3420**
DGA	**0.0056**	**0.0088**	**0.0157**	***	*0.0738*	**0.0714**[b]	**0.1309**[c]
ZAN	*0.0030*	*0.0016*	**0.0060**	−0.0016	***	**0.0870**[b]	**0.0273**
MAY	−0.0033	**0.0050**	**0.0063**	*0.0006*	−0.0022	***	**0.0821**[a]
JNO	*0.0008*	**0.0067**	**0.0039**	*0.0018*	−0.0010	−0.0038	***

[a]Comparisons not significant when using the datasets with 9 and 7 outlier loci.
[b]Comparison not significant when using the dataset with 9 outliers.
[c]Comparison not significant when using the dataset with 7 outliers.

function p.adjust. Second, we detected outliers using the program LOSITAN [41] that also uses the FDIST method, but without a hierarchical approach. We ran 50 000 simulations, the false discovery rate was set at 0.05, and used an infinite allele mutation model. Finally, we used a Bayesian approach to estimate the probability that each locus is subject to selection, using BAYESCAN 2.1 [42]. The analysis was run with 5000 iterations and prior odds of eight, with a false discovery rate of 0.01. Based on these results, we classified each locus into one of three categories: (i) loci with F_{ST} values significantly higher than expected under neutrality, (ii) loci with F_{ST} values significantly lower than expected, and (iii) neutral loci with F_{ST} values within the expected range. The possible adaptive value of the outlier loci was explored by blasting their sequences in the NCBI nucleotide database and looking for match with genes and their functions. The search on the database nucleotide collection (nr/nt) was optimized for somewhat similar sequences (Blastn), with the default algorithm options.

To test for genetic structure, we conducted hierarchical AMOVAs and calculated pairwise F_{ST} [43] using ARLEQUIN; for the latter, sequential Bonferroni corrections were applied [44]. Discriminant analyses of principal components (DAPC) [45] were executed using ADEGENET [46] for R (R Development Core Team 2015). In addition, we ran the Bayesian clustering method implemented in STRUCTURE [47]. To test for isolation by distance (IBD), we compared matrixes of $F_{ST}/(1 - F_{ST})$ and minimum ocean distance with Mantel tests performed using GENEPOP [48]. For details, see electronic supplementary material, methods.

3. Results

A total of 1174 loci were obtained for 93 individuals. The three outlier loci methods combined identified 26 outlier loci with high F_{ST}. The ARLEQUIN method identified 25 loci before the false discovery rate corrections and seven after the corrections. Nine loci, and one locus, were identified as outliers by LOSITAN and BAYESCAN, respectively. Of the nine loci identified by LOSITAN, eight were also identified by ARLEQUIN. The outlier locus identified by BAYESCAN was also found with LOSITAN and ARLEQUIN. All 26 outlier loci identified were compared with GenBank entries (BLAST search) to identify potential gene functions, but no significant alignments were found (electronic supplementary material, table S1).

For the population genetic analysis with outliers, we performed analyses with 7, 9 and 25 loci (the latter being the outlier loci identified by ARLEQUIN before further corrections) and all population genetic results remained unchanged, except for some F_{ST} pairwise comparisons (table 1).

The FDIST method from ARLEQUIN without false discovery rate correction that identified 25 outlier loci with high F_{ST} (potentially under directional selection), also identified 32 outliers potentially under balancing selection (lower than expected F_{ST}). Based on these results, we generated three datasets for the population genetic analysis: (i) all loci ($n = 1174$), (ii) neutral loci ($n = 1117$), and (iii) outlier loci ($n = 25$). Note that the 32 outliers thought to be under balancing selection were excluded from the neutral dataset.

Analyses of all loci ($n = 1174$) showed evidence of a population structure within our study region. We found a global F_{ST} of 0.0127 ($p < 0.0001$). Pairwise F_{ST} comparisons (electronic supplementary material, table S2) indicate that DJI and OMA are significantly different from all other populations, while the NRS

is different from all except MAY and JNO. An AMOVA showed a very low but significant divergence between the NWIO (Red Sea, Djibouti and Oman) and the WIO (Chagos, Zanzibar, Mayotte, Juan de Nova) ($F_{CT} = 0.0099$, $p = 0.0225$). However, STRUCTURE analysis did not resolve any population clusters, while DAPC results suggest that Oman is distinct from other populations (electronic supplementary material, figure S1). Finally, there was statistically significant IBD (electronic supplementary material, figure S2), but the slope is low and there is not a strong model fit.

Analyses of the neutral dataset ($n = 1117$) indicated high genetic connectivity and weak differentiation between the two provinces (NWIO and WIO). Global F_{ST} was low but significant (0.0057, $p < 0.0001$), Pairwise F_{ST} comparisons showed that the population in Oman (OMA) stood out and it was significantly different from all other populations (table 1), while Djibouti (DJI) was significantly different from all except Zanzibar (ZAN). An AMOVA demonstrated low but significant divergence between the NWIO and the WIO provinces ($F_{CT} = 0.0041$, $p = 0.0283$, electronic supplementary material, table S3), while the DAPC analysis shows a close relationship among all populations except Oman (figure 1). STRUCTURE analysis indicated $K = 1$ (figure 1). Finally, there was evidence of weak and not significant IBD (figure 1).

Analyses of the outlier loci ($n = 25$) showed strong genetic differentiation between provinces. Global F_{ST} for the outliers was 0.3271 ($p < 0.0001$), and all the pairwise comparisons were significant except between Zanzibar and Diego Garcia (Chagos Archipelago), and between Zanzibar and Juan de Nova (table 1). An AMOVA supports the distinction between the NWIO and WIO provinces ($F_{CT} = 0.2349$, $p = 0.0342$, electronic supplementary material, table S3). The DAPC analysis also identified separation of the NWIO and WIO provinces (figure 1), while the results from STRUCTURE suggest the presence of two clusters ($K = 2$) that closely match the DAPC results (figure 1). Outliers revealed significant IBD (figure 1), but the slope is low and there is not a strong model fit.

4. Discussion

The present study revealed significant genetic structure between Northwestern and Western Indian Ocean populations of *D. trimaculatus*, demonstrating concordance between intraspecific phylogeographic boundaries and biogeographic boundaries [49]. In addition, Oman appears as a distinct population from all the others, and Djibouti is distinct from most populations. At least some of the divergence appears to be driven by the outlier loci. These showed a clear difference between provinces that was consistent across analyses (STRUCTURE, DAPC, AMOVA), and the DAPC results for this dataset indicate that the populations of Oman, Djibouti and the Northern Red Sea are distinct from each other. These results suggest that gene flow is variable across the genome and it may be affected by different processes and/or operate at different scales. However, it is important to acknowledge that variation in recombination rate could also explain patterns of divergence across the genome [50].

Habitat discontinuities, deep-water upwellings and the direction of prevailing currents (figure 1) could be responsible for contemporary isolation between the Northwestern and Western Indian Ocean provinces. Seasonal upwelling brings cold and nutrient-rich waters to southern Oman and the Somali coast, creating large areas unsuitable for the development of coral reef habitat. In addition, currents and complex topography may divert larvae and prevent dispersal between these provinces [21,51]. If the divergences revealed in the outlier dataset are due to isolation and not adaptation, then these loci should be subject to the effects of drift and show similar patterns to the neutral loci. Our outlier dataset does, in fact, show similar—but stronger—signals compared to the neutral loci. AMOVAs based on the neutral and outlier loci demonstrate weak but significant structures between provinces (see also electronic supplementary material, table S2).

Sea-level fluctuations may also contribute to the observed pattern, as the Red Sea was subject to periods of extreme isolation when sea level dropped as much as 130 m below current levels during the Pleistocene [21]. In some cases, this repeated isolation led to speciation, while in others it only led to population differentiation, as seen here in *D. trimaculatus*. After the last glacial maximum 26.5 to 19 kya, populations of many species began to expand into the Red Sea and Persian Gulf as habitat opened up [52]. When a subset of individuals at the leading edge of a population expansion moves into a new territory, their particular alleles increase in frequency, a phenomenon called 'allele surfing' [53]. Unlike most other demographic effects, allele surfing generally does not affect all loci, so it can impact neutral allele frequencies in ways that mimic the patterns of directional selection [53,54] and could be responsible for the results that are more evident in outlier loci.

In contrast with the more stable WIO, the NWIO is one of the most variable and environmentally extreme regions in the tropical oceans [21]. Such differences could be selecting for different traits across provinces in

D. trimaculatus and other species. During the summer months, the waters between the Arabian Peninsula and the Red Sea become the world's hottest sea, while in the winter they become one of the coldest environments for coral reef growth [55]. The Red Sea experiences large spatio-temporal fluctuations in physical conditions and a unique north–south environmental gradient in salinity, temperature and primary productivity [21] (see electronic supplementary material, figure S3). Reefs in both the Red Sea and Gulf of Oman are known to have high variability in environmental factors such as temperature and salinity [22,56,57]. Adaptation to these highly variable environments might drive the high rates of endemism in the region [21] and may affect the survival of recruits from non-native populations. There is a possibility that the outlier loci are under selection and reflect adaptive divergence; however, this hypothesis is less likely. Isolation by distance is a neutral pattern, yet a weak trend was detected in the outlier dataset. In addition, we could not clearly identify genes involved in specific adaptations (electronic supplementary material, table S3), nor exclude the possibility of false positives.

While it is difficult to distinguish between divergence driven by selection and drift, it is important to note that these processes are not mutually exclusive and could be acting in concert on populations found around the region, given its complex geologic history and heterogeneous environment. It is possible that physical barriers between the provinces are semipermeable, allowing for restricted dispersal, and environmental contrasts between provinces reinforce those barriers through selection. Because population sizes fluctuate with sea level, the founder effect can also influence our results: population expansions after isolation can promote adaptation if colonizing individuals carry beneficial mutations [58]. In our view, multiple processes are probably at play in the study region, and carefully designed experiments are needed to disentangle their particular roles.

Despite the lack of a clear causal mechanism, our data can be used along with data on genetic connectivity of other species to identify the common genetic breaks that need to be considered for the conservation of biodiversity and evolutionary processes in the poorly studied Western Indian Ocean region. Our results suggest that the Red Sea and Arabian populations should be managed separately from the greater Western Indian Ocean population, and the role of adaptive versus neutral variation must be examined further.

References

1. Rocha L, Bowen B. 2008 Speciation in coral reef fishes. *J. Fish Biol.* **72**, 1101–1121. (doi:10.1111/j.1095-8649.2007.01770.x)

2. Kulbicki M *et al.* 2013 Global biogeography of reef fishes: a hierarchical quantitative delineation of regions. *PLoS ONE* **8**, e81847. (doi:10.1371/journal.pone.0081847.g001)

3. Gaither M, Rocha LA. 2013 Origins of species richness in the Indo-Malay-Phillipine biodiversity hotspot: evidence for the centre of overlap hypothesis. *J. Biogeogr.* **40**, 1638–1648. (doi:10.1111/jbi.12126)

4. Toonen RJ *et al.* 2011 Defining boundaries for ecosystem-based management: a multispecies case study of marine connectivity across the Hawaiian Archipelago. *J. Marine Biol.* **2011**, 1–13. (doi:10.1155/2011/460173)

5. Ridgway T, Sampayo EM. 2005 Population genetic status of the Western Indian Ocean: what do we know? *West. Indian Ocean J. Mar. Sci.* **4**, 1–10. (doi:10.4314/wiojms.v4i1.28468)

6. Wilson N, Kirkendale L. 2016 Putting the 'Indo' back into the Indo-Pacific: resolving marine phylogeographic gaps. *Invertebr. Syst.* **30**, 86–94. (doi:10.1071/IS15032)

7. Burton RS. 1998 Intraspecific phylogeography across the Point Conception biogeographic boundary. *Evolution* **52**, 734–745. (doi:10.1111/j.1558-5646.1998.tb03698.x)

8. Dorenbosch M, Pollux B, Pustjens A, Rajagopal S, Nagelkerken I, van der Velde G, Moon-van der Staay S. 2006 Population structure of the Dory snapper, *Lutjanus fulviflamma*, in the western Indian Ocean revealed by means of AFLP fingerprinting. *Hydrobiologia* **568**, 43–53. (doi:10.1007/s10750-006-0020-8)

9. Horne JB, van Herwerden L, Choat JH, Robertson D. 2008 High population connectivity across the Indo-Pacific: congruent lack of phylogeographic structure in three reef fish congeners. *Mol. Phylogen. Evol.* **49**, 629–638. (doi:10.1016/j.ympev.2008.08.023)

10. Craig MT, Eble JA, Bowen BW, Robertson D. 2007 High genetic connectivity across the Indian and Pacific Oceans in the reef fish *Myripristis berndti* (Holocentridae). *Mar. Ecol. Prog. Ser.* **334**, 245–254. (doi:10.3354/meps334245)

11. Gaither MR, Toonen RJ, Robertson DR, Planes S, Bowen BW. 2010 Genetic evaluation of marine biogeographical barriers: perspectives from two widespread Indo-Pacific snappers (*Lutjanus kasmira* and *Lutjanus fulvus*). *J. Biogeogr.* **37**, 133–147. (doi:10.1111/j.1365-2699.2009.02188.x)

12. Ridgway T, Hoegh-Guldberg O, Ayre D. 2001 Panmixia in *Pocillopora verrucosa* from South Africa. *Mar. Biol.* **139**, 175–181. (doi:10.1007/s002270100573)

13. Silva IC, Mesquita N, Paula J. 2010 Lack of population structure in the fiddler crab *Uca annulipes* along an East African latitudinal gradient: genetic and morphometric evidence. *Mar. Biol.* **157**, 1113–1126. (doi:10.1007/s00227-010-1393-9)

14. Muths D, Gouws G, Mwale M, Tessier E, Bourjea J. 2012 Genetic connectivity of the reef fish *Lutjanus kasmira* at the scale of the western Indian Ocean. *Can. J. Fish. Aquat. Sci.* **69** 842–853. (doi:10.1139/f2012-012)

15. Otwoma LM, Reuter H, Timm J, Meyer A. 2018 Genetic connectivity in a herbivorous coral reef fish (*Acanthurus leucosternon* Bennet, 1833) in the Eastern African region. *Hydrobiologia* **806**, 237–250. (doi:10.1007/s10750-017-3363-4)

16. Duda T, Palumbi S. 1999 Population structure of the black tiger prawn, *Penaeus monodon*, among western Indian Ocean and western Pacific populations. *Mar. Biol.* **134**, 705–710. (doi:10.1007/s002270050586)

17. Fratini S, Vannini M. 2002 Genetic differentiation in the mud crab *Scylla serrata* (Decapoda: Portunidae) within the Indian Ocean. *J. Exp. Mar. Biol. Ecol.* **272**, 103–116. (doi:10.1016/S0022-0981(02)00052-7)

18. Visram S, Yang M-C, Pillay RM, Said S, Henriksson O, Grahn M, Chen CA. 2010 Genetic connectivity and historical demography of the blue barred parrotfish (*Scarus ghobban*) in the western Indian Ocean. *Mar. Biol.* **157**, 1475–1487. (doi:10.1007/s00227-010-1422-8)

19. DiBattista J, Berumen M, Gaither M, Rocha L, Eble J, Choat J, MT Craig, Skillings D, Bowen B. 2013 After continents divide: comparative phylogeography of reef fishes from the Red Sea

and Indian Ocean. *J. Biogeogr.* **40**, 1170–1181. (doi:10.1111/jbi.12068)

20. DiBattista JD, Waldrop E, Rocha LA, Craig MT, Berumen ML, Bowen BW. 2015 Blinded by the bright: a lack of congruence between colour morphs, phylogeography and taxonomy for a cosmopolitan Indo-Pacific butterflyfish, *Chaetodon auriga*. *J. Biogeogr.* **42**, 1919–1929. (doi:10.1111/jbi.12572)

21. DiBattista JD et al. 2016 On the origin of endemic species in the Red Sea. *J. Biogeogr.* **43**, 13–30. (doi:10.1111/jbi.12631)

22. DiBattista JD et al. 2016 A review of contemporary patterns of endemism for shallow water reef fauna in the Red Sea. *J. Biogeogr.* **43**, 423–439. (doi:10.1111/jbi.12649)

23. Ragionieri L, Cannicci S, Schubart CD, Fratini S. 2010 Gene flow and demographic history of the mangrove crab *Neosarmatium meinerti*: a case study from the western Indian Ocean. *Estuar. Coast. Shelf Sci.* **86**, 179–188. (doi:10.1016/j.ecss.2009.11.002)

24. Waldrop E, Hobbs JPA, Randall JE, DiBattista JD, Rocha LA, Kosaki RK, Berumen ML, Bowen BW. 2016 Phylogeography, population structure and evolution of coral-eating butterflyfishes (Family Chaetodontidae, genus *Chaetodon*, subgenus *Corallochaetodon*). *J. Biogeogr.* **43**, 1116–1129. (doi:10.1111/jbi.12680)

25. Muths D, Tessier E, Gouws G, Craig M, Mwale M, Mwaluma J, Mwandya A, Bourjea J. 2011 Restricted dispersal of the reef fish *Myripristis berndti* at the scale of the SW Indian Ocean. *Mar. Ecol. Prog. Ser.* **443**, 167–180. (doi:10.3354/meps09394)

26. Muths D, Tessier E, Bourjea J. 2015 Genetic structure of the reef grouper *Epinephelus merra* in the West Indian Ocean appears congruent with biogeographic and oceanographic boundaries. *Mar. Ecol.* **36**, 447–461. (doi:10.1111/maec.12153)

27. Bernardi G, Holbrook SJ, Schmitt RJ. 2001 Gene flow at three spatial scales in a coral reef fish, the three-spot damselfish, *Dascyllus trimaculatus*. *Mar. Biol.* **138**, 457–465. (doi:10.1007/s002270000484)

28. Garnaud J. 1957 Ethologie de *Dascyllus trimaculatus* (Rüppell). *Bull. Inst. Oceangr. Monaco*. **54**, 1–10.

29. Bernardi G, Beldade R, Holbrook SJ, Schmitt RJ. 2012 Full-sibs in cohorts of newly settled coral reef fishes. *PLoS ONE* **7**, e44953. (doi:10.1371/journal.pone.0044953)

30. Wellington G, Victor B. 1989 Planktonic larval duration of one hundred species of Pacific and Atlantic damselfishes (Pomacentridae). *Mar. Biol.* **101**, 557–567. (doi:10.1007/BF00541659)

31. Leray M, Beldade R, Holbrook S, Schmitt R, Planes S, Bernardi G. 2010 Allopatric divergence and speciation in a coral reef fish: the three-spot dascyllus, *Dascyllus trimaculatus* species complex. *Evolution* **64**, 1218–1230. (doi:10.1111/j.1558-5646.2009.00917.x)

32. Berumen ML, DiBattista JD, Rocha LA. 2017 Introduction to virtual issue on Red Sea and Western Indian Ocean biogeography. *J. Biogeogr.* **44**, 1923–1926. (doi:10.1111/jbi.13036)

33. Robitzch V, Lozano-Cortes D, Kandler N, Salas E, Berumen ML. 2016 Productivity and sea surface

temperature are correlated with the pelagic larval duration of damselfishes in the Red Sea. *Mar. Pollut. Bull.* **105**, 566–574. (doi:10.1016/j.marpolbul.2015.11.045)

34. Bernardi G, Holbrook SJ, Schmitt RJ, Crane NL, DeMartini E. 2002 Species boundaries, populations and colour morphs in the coral reef three–spot damselfish (*Dascyllus trimaculatus*) species complex. *Proc. R. Soc. Lond. B* **269**, 599–605. (doi:10.1098/rspb.2001.1922)

35. Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE. 2012 Double digest RADseq: an inexpensive method for *de novo* SNP discovery and genotyping in model and non-model species. *PLoS ONE* **7**, e37135. (doi:10.1371/journal.pone.0037135)

36. Catchen J, Hohenlohe PA, Bassham S, Amores A, Cresko WA. 2013 Stacks: an analysis tool set for population genomics. *Mol. Ecol.* **22**, 3124–3140. (doi:10.1111/mec.12354)

37. Lischer H, Excoffier L. 2012 PGDSpider: an automated data conversion tool for connecting population genetics and genomics programs. *Bioinformatics* **28**, 298–299. (doi:10.1093/bioinformatics/btr642)

38. Excoffier L, Hofer T, Foll M. 2009 Detecting loci under selection in a hierarchically structured population. *Heredity* **103**, 285–298. (doi:10.1038/hdy.2009.74)

39. Excoffier L, Lischer HE. 2010 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. (doi:10.1111/j.1755-0998.2010.02847.x)

40. Benjamini Y, Hochberg Y. 1995 Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. B (Methodological)*. **57**, 289–300. (doi:10.1111/j.2517-6161.1995.tb02031.x)

41. Antao T, Lopes A, Lopes RJ, Beja-Pereira A, Luikart G. 2008 LOSITAN: a workbench to detect molecular adaptation based on a F_{st}-outlier method. *BMC Bioinf.* **9**, 323. (doi:10.1186/1471-2105-9-323)

42. Foll M, Gaggiotti OE. 2008 A genome scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. *Genetics* **180**, 977–993. (doi:10.1534/genetics.108.092221)

43. Weir BS, Cockerham CC. 1984 Estimating F-statistics for the analysis of population structure. *Evolution* **38**, 1358–1370. (doi:10.2307/2408641)

44. Rice WR. 1989 Analyzing tables of statistical tests. *Evolution* **43**, 223–225. (doi:10.2307/2409177)

45. Jombart T, Devillard S, Balloux F. 2010 Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genet.* **11**, 94. (doi:10.1186/1471-2156-11-94)

46. Jombart T. 2008 Adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics* **24**, 1403–1405. (doi:10.1093/bioinformatics/btn129)

47. Pritchard JK, Stephens M, Donnelly P. 2000 Inference of population structure using multilocus genotype data. *Genetics* **155**, 945–959.

48. Raymond M., Rousset F. 1995 GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. *J. Hered.* **86**, 248–249. (doi:10.1093/oxfordjournals.jhered.a111573)

49. Bowen BW, Gaither MR, Dibattista JD, Iacchei M, Andrews KR, Grant WS, Toonen RJ, Briggs JC. 2016 Comparative phylogeography of the ocean planet. *Proc. Natl Acad. Sci. USA* **113**, 7962–7969 (doi:10.1073/pnas.1602404113)

50. Burri R. 2017 Interpreting differentiation landscapes in the light of long-term linked selection. *Evol. Lett.* **1**, 118–131. (doi:10.1002/evl3.14)

51. Kemp J. 1998 Zoogeography of the coral reef fishes of the Socotra Archipelago. *J. Biogeogr.* **25**, 919–933. (doi:10.1046/j.1365-2699.1998.00249.x)

52. Ludt WB, Rocha LA. 2015 Shifting seas: the impacts of Pleistocene sea-level fluctuations on the evolution of tropical marine taxa. *J. Biogeogr.* **42**, 25–38. (doi:10.1111/jbi.12416)

53. Excoffier L, Ray N. 2008 Surfing during population expansions promotes genetic revolutions and structuration. *Trends Ecol. Evol.* **23**, 347–351. (doi:10.1016/j.tree.2008.04.004)

54. Kirk H, Freeland JR. 2011 Applications and implications of neutral versus non-neutral markers in molecular ecology. *Int. J. Mol. Sci.* **12**, 3966–3988. (doi:10.3390/ijms12063966)

55. Riegl BM, Purkis SJ. 2012 Coral reefs of the Gulf: adaptation to climatic extremes in the world's hottest sea. In *Coral reefs of the Gulf* (eds BM Riegl, SJ Purkis), pp. 1–4. Dordrecht, The Netherlands: Springer.

56. Thoppil PG, Hogan PJ. 2009 On the mechanisms of episodic salinity outflow events in the Strait of Hormuz. *J. Phys. Oceanogr.* **39**, 1340–1360. (doi: 10.1175/2008jpo3941.1)

57. Cavalcante GH, Feary DA, Burt JA. 2016 The influence of extreme winds on coastal oceanography and its implications for coral population connectivity in the southern Arabian Gulf. *Mar. Pollut. Bull.* **105**, 489–497. (doi:10.1016/j.marpolbul.2015.10.031)

58. Gralka M, Stiewe F, Farrell F, Moebius W, Waclaw B, Hallatschek O. 2016 Allele surfing promotes microbial adaptation from standing variation. *bioRxiv*, 049353. (doi:10.1101/049353)

59. Salas EM, Bernardi G, Berumen ML, Gaither M, Rocha LA. 2019 Data from: RADseq analyses reveal concordant Indian Ocean biogeographic and phylogeographic boundaries in the reef fish *Dascyllus trimaculatus*. Dryad Digital Repository. (doi:10.5061/dryad.bn457rr)

60. Deck J, Gaither MR, Ewing R, Bird CE, Davies N, Meyer C, Riginos C, Toonen RJ, Crandall ED. 2017 The genomic observatories metadatabase (GeOMe): a new repository for field and sampling event metadata associated with genetic samples. *PLoS Biol.* **15**, e2002925. (doi:10.1371/journal.pbio.2002925)