Like a rolling stone: colonization and migration dynamics of the grey reef shark (*Carcharhinus amblyrhynchos*)

Running title: Evolutionary history of the grey reef shark.

Pierre Lesturgie[1], Camrin D. Braun[2], Eric Clua[3,4], Jeremy Kiszka[5], Johann Mourier[3,6], Simon R. Thorrold[2], Thomas Vignaud[3], Serge Planes[3,4], Stefano Mona[1,3,*]

[1] Institut de Systématique, Evolution, Biodiversité, ISYEB (UMR 7205), Muséum National d'Histoire Naturelle, CNRS, Sorbonne Université, EPHE, Université des Antilles, Paris, France.

[2] Biology Department, Woods Hole Oceanographic Institution, Woods Hole, MA 02543, USA

[3] Laboratoire d'Excellence CORAIL, Papetoai, French Polynesia.

[4] PSL Research University: EPHE-UPVD-CNRS, USR 3278 CRIOBE, Université de Perpignan, 52 Avenue Paul Alduy, 66860, Perpignan, Cedex, France

[5] Institute of Environment, Department of Biological Sciences, Florida International University, North Miami, USA

[6] Université de Corse Pasquale Paoli, UMS 3514 Plateforme Marine Stella Mare, 20620 Biguglia, France

* Author for corresponding: Stefano Mona Institut de Systématique, Evolution, Biodiversité, ISYEB (UMR 7205), Muséum National d'Histoire Naturelle, CNRS, Sorbonne Université, EPHE, Université des Antilles, Paris, France, stefano.mona@mnhn.fr.

Abstract

Designing appropriate management plans requires knowledge of both the dispersal ability and what has shaped the current distribution of the species under consideration. Here we investigated the evolutionary history of the endangered grey reef shark (*Carcharhinus amblyrhynchos*) across its range by sequencing thousands of RAD-seq loci in 173 individuals in the Indo-Pacific (IP) . We first bring evidence of the occurrence of a range expansion (RE) originating close to the Indo-Australian Archipelago (IAA) where two stepping-stone waves (east and westward) colonized almost the entire IP. Coalescent modeling additionally highlighted a homogenous connectivity (*Nm*~10 per generation) throughout the range, and an isolation by distance model suggested the absence of barriers to dispersal despite the affinity of *C. amblyrhynchos* to coral reefs. This coincides with long-distance swims previously recorded, suggesting that the strong genetic structure at the IP scale (F_{ST} ~ 0.56 between its ends) is the consequence of its broad current distribution and organization in a large number of demes. Our results strongly suggest that management plans for the grey reef shark should be designed on a range-wide rather than a local scale due to its continuous genetic structure. We further contrasted these results with those obtained previously for the sympatric but strictly lagoon-associated *Carcharhinus melanopterus*, known for its restricted dispersal ability. *C. melanopterus* exhibits similar RE dynamic, but is characterized by stronger genetic structure and a non-homogeneous connectivity largely dependent on local coral reefs availability. This sheds new light on shark evolution, emphasizing the roles of IAA as source of biodiversity and of life history traits in shaping the extent of genetic structure and diversity.

Keywords: Meta-population, Rad-seq, demographic history, range expansion, *Carcharhinus amblyrhynchos*, *Carcharhinus melanopterus*.

Introduction

More than 37% of shark species are currently threatened with extinction (Dulvy et al., 2021) and less than 30% are on stable or increasing population trend according to the International Union for Conservation of Nature (IUCN) Red List of threatened species. As meso or apex predators, they hold important roles in their ecosystems (Bornatowski, Navia, Braga, Abilhoa, & Corrêa, 2014) and their decline has already shown negative cascading effects on food web structure (Friedlander & DeMartini, 2002; Myers, Baum, Shepherd, Powers, & Peterson, 2007). Although local-scale conservation programs have been established, their efficiency has been questioned for some species of sharks (Robbins, Hisano, Connolly, & Choat, 2006; Speed et al., 2016). For instance, local-scale management might not always be consistent with the home range size and the dispersal ability of sharks (but see Dwyer et al. (2020)). Genetics and ecological evidence have identified both species with very restricted home ranges (Mourier, Mills, & Planes, 2013; Whitney, Robbins, Schultz, Bowen, & Holland, 2012) and species capable of crossing large expanses of ocean (Bailleul et al., 2018; Corrigan et al., 2018; Pirog et al., 2019). Designing appropriate management actions is therefore a difficult task requiring the knowledge of both the dispersal ability of the species under investigation and the existence of barriers to gene flow, which are often hard to identify in the marine realm.

Population genomics is becoming increasingly important in this context, particularly because of the large amount of data provided by the emergence of next generation sequencing approaches (NGS). It is now possible to assess the genetic diversity of model or non-model species at an unprecedented level of accuracy (Benazzo et al., 2017; Steiner, Putnam, Hoeck, & Ryder, 2013). However, genetic diversity alone does not provide clues on the evolutionary trajectory of a species and a careful modelling is required to fully understand its demographic history as well as the

conservation challenges to be faced. Unfortunately, for computational reasons, many commonly used software implement, under different algorithms, *unstructured* models, i.e., models that consider the population under investigation as isolated or panmictic (Heled & Drummond, 2008; Heller, Chikhi, & Siegismund, 2013; Li & Durbin, 2011; Liu & Fu, 2015). Except for highly vagile species which are panmictic at a large scale (Corrigan et al., 2018; Lesturgie, Planes, & Mona, 2022; Pirog et al., 2019), broadly distributed sharks species are more likely organized in meta-population(s) throughout their range (Maisano Delser et al., 2019, 2016; Momigliano et al., 2017; Pazmiño et al., 2018). The application of *unstructured* models to species organised in meta-populations yield spurious signatures of effective populations size (N_e) changes through time (Chikhi, Sousa, Luisi, Goossens, & Beaumont, 2010; Maisano Delser et al., 2019; Mazet, Rodríguez, Grusea, Boitard, & Chikhi, 2016; Mazet, Rodríguez, & Chikhi, 2015), with potentially dangerous consequences in terms of conservation policies. However, recent studies have highlighted the usefulness of such models to characterize the gene genealogy of the sampled lineages which in turn reveals important features of the meta-population (Arredondo et al., 2021; Lesturgie et al., 2022; Rodríguez et al., 2018). This emphasizes the necessity to couple complex meta-population models and *unstructured* models when uncovering the demographic history of a species.

Here we investigated the evolutionary history of the grey reef shark *Carcharhinus amblyrhynchos*, a coral reef-associated shark inhabiting the tropical Indo-Pacific. While *C. amblyrhynchos* is considered one of the most abundant reef sharks in the Indo-Pacific, it is listed as Endangered on the IUCN red list of threatened species. With a mean size of ~190 cm (Compagno, 2001), *C. amblyrhynchos* inhabits either fringing or barrier reefs and displays patterns of reef fidelity (Barnett, Abrantes, Seymour, & Fitzpatrick, 2012; Espinoza, Heupel, Tobin, & Simpfendorfer,

2014) as well as philopatry (Field, Meekan, Speed, White, & Bradshaw, 2011). Tagging studies have indicated long range movement up to ~900 km (Barnett et al., 2012; Bonnin et al., 2019), which raise questions about the extent of residency patterns for this species. Previous molecular studies using both microsatellites and Rad-sequencing did not find signatures of genetic structure at a low geographic scale such as the Great Barrier Reef (Momigliano et al., 2017; Momigliano, Harcourt, Robbins, & Stow, 2015) and the Phoenix Islands archipelago (Boissin et al., 2019). Conversely, isolation by distance patterns have been found at larger scale and some evidence suggests that coastal abundance of reef can fuel genetic exchanges, while oceanic expanses are barriers to gene flow (Boissin et al., 2019; Momigliano et al., 2017).

To shed light on these contrasting findings, we sequenced DNA from 203 individuals of *C. amblyrhynchos* sampled at 18 sites covering most of its distribution range (Figure 1) following a double digest restriction site associated DNA protocol (dd-RADseq, Peterson et al. 2012). The large panel of assembled loci was used to: (*i*) detect the occurrence and origin location of a range expansion (RE); (*ii*) investigate its demographic history by implementing both meta-population and *unstructured* models; (*iii*) reassess the population structure of the grey reef shark in the Indo-Pacific. We finally compared the results here obtained with those previously found in the blacktip reef shark (*Carcharhinus melanopterus* (Maisano Delser et al., 2019, 2016)). The two species share a very similar distribution in the Indo-Pacific but are characterized by different habitat preferences and life-history traits, providing an excellent opportunity to improve our knowledge on the biology of sharks.

Material and Methods

Sampling and Rad sequencing

We collected 203 samples of *C. amblyrhynchos* that covered most of its longitudinal distribution range (Figure 1), with two sampling sites in the Mozambique Channel in the western Indian Ocean (IND – Juan de Nova and Zélée bank) and 16 in the Pacific Ocean (PAC). Among the PAC sampling sites, four were chosen in the Coral Sea (COR): two in the Chesterfield Islands (Bampton and Avond) and two in New Caledonia (Belep and Poindimie). The remaining samples came from the Central and Easter Pacific (CPA): six in the Phoenix Islands (Enderbury, Kanton, McKean, Niku, Orona and Birnie) one in Palmyra Island and five in French Polynesia (Fakarava, Moorea, Faaite, Raraga and Nengo) (Figure 1, Table 1). Total genomic DNA has been extracted and conserved in 96% ethanol using QIAGEN DNeasy Blood and Tissue purification kit (Qiagen, Hilden, Germany) according to the manufacturer's protocols. We followed the double digest restriction site associated DNA (dd-RADseq) protocol of (Peterson et al., 2012) to create a genomic library, using EcoRI and MSFI as restriction enzymes. We selected fragments of ~400 bp length and sequenced with Illumina HiSeq 2500 machine (single-end, 125 bp).

In the absence of a reference genome, we assembled loci *de novo* using *Stacks* v.2.5 (Rochette, Rivera-Colón, & Catchen, 2019). Briefly, we demultiplexed the reads through the *process_radtags.pl* script and assembled the loci using the *denovo_map.pl* pipeline with the parameters *m=3* (minimum read depth to create a stack), *M=3* (number of mismatches allowed between loci within individuals) and *n=3* (number of mismatches allowed between loci within catalogue). We found a mean depth of coverage (over individuals and loci) of ~10x (see Results). Previous work suggested that such low-coverage value may bias a correct genotype calling under the algorithm implemented in *Stacks* v.1, *Stacks* v.2 and *PyRAD* by skewing the site frequency

spectrum (SFS) towards an excess of low frequency variants (Mona et al. in prep; see supplementary materials for details). For this reason, we followed two different bioinformatics pipelines: the first to obtain a dataset to perform analyses based on the SFS (genetic diversity, range expansion and historical demographic inferences) and the second to investigate population structure, for which low frequency variants are not informative and are removed before the downstream analyses.

Genetic diversity

We followed the genotype free estimation of allele frequencies pipeline implemented in the software *ANGSD* v.0.923 (Korneliussen, Albrechtsen, & Nielsen, 2014). This approach has been suggested to be more efficient for low to medium coverage NGS data than SNPs calling algorithms (Korneliussen et al., 2014). *ANGSD* requires a reference sequence to work. To this end, we followed the framework proposed by Khimoun et al. (2020) and Heller et al. (2021) which we applied to each sampling site separately to maximise the number of loci: i) we assembled Rad loci present in at least 80% of the sampled individuals using *Stacks* with the same parameters as above (i.e., *m=M=n=3*); ii) we concatenated the consensus sequences for each locus, to which we added a stretch of 120 "N" in order to facilitate mapping, to create an artificial reference sequence; iii) we mapped raw reads from individual *fastq* files using the *bwa-mem* algorithm with default parameters (Li & Durbin, 2009) against the artificial reference sequence. Using *ANGSD* filters, we discarded *(1)* sites with a coverage < 3 (using the flag *-minIndDepth 3*) *(2)* poor quality and mis-aligned reads (with default parameters and flags *-minQ20* and *-minMapQ 20*), *(3)* poor quality bases (with default parameters and flags *-baq 1* and *-C 50*). We further removed the last 5bp of each locus, SNPs heterozygous in at least 80% individuals, and loci with more than 5 SNPs. We

finally filtered all missing data by applying the *-minInd* filter equal to the total number of individual present in each sampling site (Table 1). We then created a *site allele frequency likelihood (saf)* file by using the SAMtools genotype likelihood computation method with the *-GL=1* flag (Li & Durbin, 2009) and finally computed the folded *site frequency spectrum* (SFS) from the *saf* files using the *RealSFS* program implemented in *ANGSD*. We computed the mean pairwise difference (θ_π), the number of segregating sites (Watterson's Theta, θ_w) and Tajima's D (*TD*) directly from the SFS. θ_π and θ_w were standardized per site (i.e., by taking into account both monomorphic and polymorphic loci) and significance of *TD* was evaluated under 1,000 coalescent simulations of a constant population model with size θ_π.

Range Expansion

Genetic diversity, here measured in each sampling site as θ_π, is expected to decay as a function of the distance from the origin of the range expansion (Ramachandran et al., 2005). Geographic distances were computed in order to take into account ecological features as it may better represent the capacity of individuals to move between two points than linear distances. To that end, we constructed a raster of 67894 cells using the R package *raster* (Hijmans, 2020) where each cell corresponds either to land, open sea, seamount or reef habitat. Permeability coefficients were fixed respectively to 0 and 1 for land and open sea, whereas coefficients for coral reefs and seamounts were varied between 1 and 100. We applied two constraints: coral reefs should always have the maximum relative permeability value (since they represent the only habitat for *C. amblyrhynchos*) and seamounts have permeability bounded within 1 and coral reefs' value. The most likely values were searched using a custom R script by maximising the correlation between the geographic and genetic distances between the sampled sites. Geographic distances were computed with the

183 *gdistance* R package under the *Least Cost* (LC) criterion algorithm (van Etten, 2017) and genetic
184 distances were measured by the F_{ST} (see below). After this step, we considered each marine cells
185 of the raster to be a potential source of origin of the range expansion (RE) and computed its
186 distance from the sampled sites under the LC criterion with the most likely permeability values
187 previously estimated. We correlated these distances with the genetic diversity of each sampling
188 site to identify areas with more negative values, which are likely associated with the origin of the
189 RE (Ramachandran et al., 2005). We limited these analyses to the PAC sites to avoid possible bias
190 due to the gap in our sampling distribution (i.e., the lack of samples between IND and the
191 westernmost PAC site). Nevertheless, we verified the robustness of our results to the inclusion of
192 IND sites.
193
194 ## Historical demographic inferences
195 To account and test for meta-population structure, we performed model selection as well as
196 parameters estimation using an Approximate Bayesian Computation (ABC) framework
197 (Bertorelle, Benazzo, & Mona, 2010). We tested three demographic scenarios (Figure 2) for each
198 sampling site, namely NS, FIM, and SST. *Model NS (no structure)*: going backward in time, NS
199 represents a panmictic population where the effective population size switches instantaneously at
200 T_c generations from N_{mod} to N_{anc}. *Model FIM (Finite Island Model)*: FIM represents a meta-
201 population composed of a two-dimensional array of 10x10 demes (D_i), each of the same size N
202 that exchanges Nm migrants with any other deme each generation. Going backward in time all
203 demes merge into a single population of size N_{anc} at T_{col} generations. *Model SST (Stepping STone)*:
204 SST is similar to FIM but demes exchange migrants only with their four closest neighbours. We
205 performed 50000 simulations under each scenario and for each sampling site independently using

fastsimcoal2 (Excoffier & Foll, 2011). We run the model selection with the Random Forest classification method implemented in the package *abcRF* (Pudlo et al., 2016) using the SFS, θ_π, θ_w and *TD* as summary statistics, to which we added the first two components of the Linear Discriminant Analysis performed on the previous summary statistics as suggested by Pudlo et al. (2016) to increase accuracy. We performed 50000 additional simulations under the most supported scenario in order to estimate the demographic parameters using the *abcRF* regression method (Raynal et al., 2019) with the same summary statistics as for the model selection. For all analyses, we performed the estimation twice to check for the consistency of the inferences. The number of trees was chosen by checking the out-of-bag error rate (OOB), and cross validation was performed for both parameter inference and model selection (hereafter, the confusion matrix) procedures. We finally modelled the variation of effective population size (N_e) through time in each sampling site with the *stairwayplot* (Liu & Fu, 2015). The *stairwayplot* assumes that the sampled lineages come from an isolated (panmictic) population (i.e., *unstructured*), which is not true in our case (see results). However, this method allows a powerful investigation of the underlying gene genealogy which provides useful elements for interpreting the evolutionary history of a meta-population (Lesturgie et al., 2022). All demographic inferences were performed using a generation time of 10 years and a mutation rate of 1.93e-8 per generation and per site following Lesturgie et al. (2022).

Population structure

Population structure inferences were performed on the dataset obtained following the assembly pipeline implemented in *Stacks 2.5* as described above. After the *de novo* assembly step, the *population* script was called to keep loci present in at least 80% of the individuals per sampling site (r = 0.8) and with a *minor allele frequency* of 0.05, hence removing low frequency variants.

We finally retained one random SNP per locus. Using a custom R script, we further filtered: *(i)* SNPs heterozygotes in more than 80% of the sample; *(ii)* loci with coverage higher than ~30x (which corresponds to the mean coverage plus twice the standard deviation); *(iii)* SNPs in the last 5bp of the assembled locus; and *(iv)* loci containing more than five SNPs, after visual inspection of the distribution of segregating sites per locus. The resulting dataset was used for the following analyses. 1) *sNMF* implemented in the R package *LEA* (Frichot & François, 2015): we investigated the number of ancestral clusters *K* by running the algorithm 10 times, with values of *K* ranging from 1 to 8. We chose the most likely *K* using the cross-entropy criterion and displayed the admixture coefficients under the best run. 2) *DAPC* implemented in the R package *Adegenet* (Jombart, Devillard, & Balloux, 2010): we varied *K* from 1 to 8 and chose the best values based on the BIC criterion. Linear discriminant functions were used to test whether individuals were correctly reassigned to the inferred clusters. 3) F_{ST}: we computed overall and pairwise F_{ST} between sampling sites with more than 5 individuals using the *PopGenome* (flag *nucleotide.F_ST*) library in R (Pfeifer, Wittelsbürger, Ramos-Onsins, & Lercher, 2014) and tested its significance with 1000 permutations using a custom R script. Isolation by distance (IBD) was computed with a Mantel Test (Mantel, 1967) between the genetic ($F_{ST}/(1-F_{ST})$) and the geographic or LC distance matrices and tested by 1000 permutations with the *ade4* R package (Thioulouse & Dray, 2007). The Mantel test, similarly as before, was limited to PAC sites. To check for IBD in the Indian Ocean, we fit a linear model to the pairwise F_{ST} values computed between the PAC and IND sites and their respective geographic distances.

Results

Genetic diversity

We discarded 30 individuals based on an excess of missing data after an initial *de novo* assembly. We found a mean depth of coverage of 10.77x (s.d. = 2.32) for the whole dataset. Summary statistics for all sampling sites are displayed in Table 1. The number of loci (monomorphic included) and SNPs with no missing data ranged from 35594 to 146858 and from 36982 to 103258 respectively across sampling sites (Table 1). Genetic diversity (θ_π and θ_w) was lower in IND sampling sites than in PAC (Table 1). Tajima's *D* values were positive in IND sampling sites and in Fakarava, suggesting an excess of high frequency variants when compared to the standard neutral model. Conversely, we found negative and significant Tajima's *D* values in all other PAC locations (except for Moorea and Mckean), suggesting an excess of low frequency variants compared to the standard neutral model (Table 1).

Range Expansion

The permeability coefficients maximising the correlation between genetic and the LC distances were very similar between the three habitat types. Indeed, we estimated the values of 1:1.02:1.02 for open sea, coral reef habitat and seamounts respectively. These values were retained for the following RE and IBD analyses. We plotted the correlation map computed using PAC sites only in Figure 3. The most negative correlation coefficients are concentrated close to the COR sampling sites, suggesting that the most likely origin of the RE is slightly east to the IAA region (Figure 3). We found consistent results when adding IND sites to the analysis (Figure S1), despite the geographic unbalanced distribution of our samples.

Historical demographic inferences

We investigated the demographic history for all sampling sites with n ≥ 7. We first used an ABC-RF framework to compare demographic scenarios (Figure 2). SST was the most supported scenario in all locations, with posterior probabilities ranging from 0.48 to 0.78 and similar classification error rate among locations (Table 2 and S1). The median *Nm* ranged from ~6 to ~14 (Table 2). Posterior distributions of *Nm* were overlapping and clearly distinct from the prior distribution (Figure S2), and both the squared mean error (SME) and the mean root squared error (MRSE) were small among locations, suggesting reliable estimates (Table S2). Posterior distributions of T_{col} overlapped among locations (Figure S2). Juan de Nova displayed a lower N_{anc} median value (~21k) than PAC sampling sites (ranging from ~34k to ~50k) although all credible intervals overlapped (Figure S2 and Table 2). Surprisingly, the ABC estimates of T_{col} and N_{anc} for the Mckean sampling site were inconsistent with any other PHO sampling sites (Figure S2 and Table 2). However, both SME and the MRSE for these two parameters were generally one order of magnitude larger than those estimated for *Nm* in all sampling sites (Table S2), suggesting less accurate estimates for T_{col} and N_{anc}.

We further investigated the variation of *Ne* through time using the *stairwayplot* algorithm (Figure 4). We detected a broadly similar *Ne* dynamic across sampling sites that we summarized for simplicity in three time periods: looking forward in time we observed an ancestral expansion followed by a constant phase and a final systematic strong decrease in recent times (Figure 4). However, we found three main differences between IND and PAC sampling sites: i) the expansion time was around twice as recent in IND than in PAC (~180ky B.P. vs. ~400ky B.P); ii) the strength of the expansion is much stronger in PAC sampling sites; iii) *Ne* during the constant period reached a value of ~40000 in PAC sampling sites and of only ~20000 in IND, consistent with the computed

θ (Table 1). The PAC sampling sites showed a remarkably homogeneous *stairwayplot*, with only the peripheral sites (Fakarava and Palmyra) having a slightly more recent ancestral expansion (Figure 4).

Population structure

After filtering, 88276 variable loci were retained to perform individual based structure analyses. Both *sNMF* and the *DAPC* clustering algorithms found K=2 as the most likely number of ancestral populations or clusters (Figure S3 and S4a). The ancestral populations inferred by *sNMF* perfectly matched the two oceanic regions, namely the Indian and the Pacific Ocean: the ancestry proportion of *cluster 1* in IND samples ranged from 70% to 100% while the ancestry proportion of *cluster 2* in PAC samples ranged from 87% to 100% (Figure 5a). This highlights slightly more admixture in IND than in PAC samples. We retained one LD function in the *DAPC* which correctly re-assigned all individuals from IND and PAC to *cluster 1* and *cluster 2* respectively (Figure S4b). We further investigated K=3 under both algorithms and found three main results: i) the ancestral populations or clusters clearly identify three geographic areas corresponding to IND, COR, and CPA regions (Figure 5a and S5); ii) the ancestry proportion of *cluster 3* follows a clinal distribution, steadily increasing in frequency from West (Indian Ocean) to East (French Polynesia) (Figure 5a); iii) all individuals belonging to the three areas are correctly re-assigned to the three clusters by the *DAPC* computed with two LD functions (Figure S4b). We then computed a PCA which showed similar results, with the first principal component explaining ~14.5% of the total variance and clearly separating individuals coming from the two oceans (Figure 5b). The second axis segregated CPA from COR samples. In agreement with the cluster analyses, CPA and COR are only slightly differentiated as the second principal component explains only ~1% of the total

318 variance. The second axis also suggested a clinal differentiation between the two clusters (Figure
319 5b).
320 Population based analyses were performed on a reduced dataset excluding sampling sites with less
321 than n=5 individuals. We therefore retained 14 sampling sites, n=168 individuals, and 88824
322 variable loci and obtained an overall F_{ST} = 0.25 (*p-value*< 0.001). The pairwise F_{ST} highlighted a
323 strong differentiation between Indian and Pacific sampling sites with values ranging from 0.53 to
324 0.56 (and always significant, *p-value* ≤ 0.001, Table S3). In contrast, comparisons within oceanic
325 regions never exceed 0.03 (Figure 6a) with values not always statistically significant. Consistently
326 with clustering results, a heatmap displaying pairwise F_{ST} values visually suggest the existence of
327 the three clusters previously identified (Figure 6a). However, the average differentiation between
328 COR and CPA is only slightly higher than within group comparisons (Figure 6a). Moreover, we
329 found a strong signature of isolation by distance (IBD) within the Pacific Ocean (using PAC sites
330 only), since the correlation between the F_{ST} and geographic or LC distance matrices was high and
331 significant (Mantel test: r = 0.93; *p-value* < 0.001 in both cases, Figure 6b). The correlation
332 between genetic and geographic distances by considering only IND *vs*. PAC pairwise distances
333 was also considerable although lower than in PAC region only (r = 0.77, Figure S5).

334 # Discussion

335 ## Range expansion

336 Range expansions (RE) occur by a series of founder effects leading to the fixation of novel alleles
337 and the decay in genetic diversity as colonization progresses (Excoffier, Foll, & Petit, 2009). They
338 also leave specific signatures in the gene genealogy of lineages sampled from a deme of the meta-
339 population (Maisano Delser et al., 2016; Ray, Currat, & Excoffier, 2003) and in the extent of
340 population structure (Mona, 2017; Mona, Ray, Arenas, & Excoffier, 2014). Testing for the

occurrence of a RE is therefore fundamental to understanding the evolutionary history of a species. Here, the spatial distribution of genetic diversity suggested the occurrence of a RE most likely starting east of the Indo-Australian Archipelago (IAA). The inferred origin area was large (Figure 3), likely due to low differences in θ_π between Pacific sampling sites (Table 1), but robust to the inclusion of samples from the Indian Ocean (Figure S1). The scenario of a RE was corroborated by other evidence. First, the strong and significant correlation coefficient between genetic and geographic distances in the Pacific Ocean (r=0.93; Mantel *p*-value < 0.001, Figure 6b and S5). This result alone would not be conclusive, since a similar pattern is also expected under an equilibrium isolation by distance, but it strengthens our previous findings. Second, the historical demography inferences performed in each sampled deme showed that the pattern of genetic variability was most likely the outcome of a non-equilibrium meta-population structured according to a stepping stone migration matrix (Table 2). In this context, both the colonization times of the meta-population estimated by the ABC (Figure S2) and the expansion times retrieved by the *stairwayplot* (Figure 4) harbour the signature of the RE process (Lesturgie et al., 2022): the oldest times are expected to be close to the centre of origin of the RE, while the more recent ones are likely associated to the edge of the colonization wave(s). While the large variance in T_{col} estimated by ABC does not allow for an accurate interpretation of the temporal dynamics of colonisation through the Indo-Pacific, the expansion times highlighted by the *stairwayplot* are consistent with the RE scenario. Indeed, all sampling sites display a simultaneous expansion time around ~400 ky B.P. (Figure 4) except for Palmyra, Fakarava and Juan de Nova, which are the sites respectively further east (Palmyra and Fakarava) and west (Juan de Nova) to the inferred origin of the RE. In summary, all the evidence presented thus far point to an origin of *C. amblyrhynchos* east of IAA (particularly, east of New Caledonia), from which two migration waves took place, one to the East

Pacific and the other to the Indian Ocean, with the Mozambique Channel being probably one of the last areas to have been colonized.

Our hypothesis is in line with the recent results of Walsh et al. (2022), but they detected the origin of the RE within rather than eastward the IAA, using a similar genetic diversity decay approach on low-coverage RAD-seq data. However, Walsh et al. (2022) assembled loci with *PyRAD* (Eaton, 2014), whose calling algorithm requires high coverage data to correctly identify genotypes (Rochette et al., 2019). When coverage is low, the direct call of genotypes is likely to skew the SFS towards an excess of singletons. This leads to biases in both genetic diversity estimates and downstream demographic inferences: genotype-free approaches such as those implemented in *ANGSD* (Mona et al. in prep; Heller et al. 2021) should be therefore preferred, as we did in this work. To demonstrate this bias, we carefully compared our results (obtained with *ANGSD*) to those obtained by three assembly and calling pipelines (namely, *PyRAD* (Eaton, 2014), *Stacks* v.1.48 (Catchen, Hohenlohe, Bassham, Amores, & Cresko, 2013) and *Stacks* v.2.5 (Rochette et al., 2019), see Supplementary Methods) using the Bampton sampling site as a test case. Bampton belongs to the Chesterfield islands, one of the sites considered by Walsh et al. (2022), and it is therefore expected to share their same demographic history. All three SFS displayed an excess of singletons in comparison to the one inferred by *ANGSD* (Figure S6b), clearly determining not only a stronger ancestral expansion but also the absence of the recent bottleneck when fed to the *stairwayplot* algorithm (Figure S6a). We note that by using the generation time and mutation rate of Walsh et al. (2022) (a mutation rate only relevant to exon capture data (Maisano Delser et al., 2016) thus lower than the one estimated for RAD-seq data (Lesturgie et al., 2022)), the *Ne* variation reconstructed by the *stairwayplot* applied to our *PyRAD* assembly was highly similar to the results they obtained in their Chesterfield sampling sites (compare our Figure 6b with their Figure 3).

Consequently, we highlight that the SFS reported by (Walsh et al., 2022) is biased toward an excess of low frequency variants, skewing their inferred genetic diversity and the subsequent demographic modelling. In turn, this implies that their results (and conclusions) might need to be revised using a pipeline that explicitly takes into account low-coverage data.

The RE scenario, characterized by a centre of origin and two independent colonization waves, is similar to the one inferred for *C. melanopterus* by (Maisano Delser et al., 2019), a species whose range distribution overlaps with that of the grey reef shark. However, the most likely origin of the RE was located within the IAA for *C. melanopterus*, a well-known centre of origin for many teleost fishes (Cowman & Bellwood, 2013), and a biodiversity hotspot (Allen, 2008). The difference observed between *C. amblyrhynchos* and *C. melanopterus* could result from the more balanced sampling scheme of (Maisano Delser et al., 2019), who could cover more homogeneously the Indo-Pacific. More samples from the IAA will be needed to refine our estimates. More generally, it will be interesting in the next future to explicitly investigate the role of the IAA for coral reef biodiversity fauna and to reconstruct the colonisations routes in the Indo-Pacific, using population genetics modelling applied to genomics data on multiple marine species to extract more general patterns (see for example Delrieu-Trottin et al. (2020)).

Historical demography

The ABC framework not only provided another evidence in favour of a non-equilibrium meta-population scenario through the model selection analysis, but also allowed us to further refine our understanding of the evolutionary history of the grey reef shark. By analysing each deme separately, we found an overlapping posterior distribution of *Nm* with an average mode of ~10 (Table 2 and Figure S2). *C. amblyrhynchos*, similarly to *C. melanopterus*, is strongly dependent

410 on reefs, whose distribution is not homogenous in the Indo-Pacific (Figure S7). We would have
411 expected the connectivity in each sampled deme to be highly correlated to the distribution of coral
412 reef in its neighbourhood, as it was previously observed in *C. melanopterus* (Maisano Delser et
413 al., 2019). However, the two species differ in their dispersal behaviours: while grey reef sharks
414 perform long-distance movements of at least ~900 km (Barnett et al., 2012; Bonnin et al., 2019;
415 White et al., 2017), the blacktip reef shark exhibits a range of movement not exceeding ~50 km
416 (Johann Mourier & Planes, 2013). Our results reinforce the idea that the neighbourhood size in the
417 two species is very different, with *C. amblyrhynchos* being able to cross expanses of open ocean
418 and therefore being less sensitive to coral reef concentration than *C. melanopterus*.
419 The homogeneity in the signature of genetic variation in each deme was confirmed by the
420 *stairwayplot* analyses (Figure 4), contrasting with the heterogeneity previously described for *C.*
421 *melanopterus* (Maisano Delser et al., 2019). All demes showed an ancestral expansion followed
422 by a period of stasis and a strong bottleneck in recent times. We recently showed that these three
423 time periods are the typical signature of the variation in the coalescence rate through time due to
424 the meta-population structure, with the slight differences observed between sites being only due
425 to their specific colonization time (Lesturgie et al., 2022). This result confirms the similarity of
426 dispersal pattern throughout the Indo-Pacific. Similarly, the signature of bottleneck observed in
427 recent times for all demes (Figure 4) is also the expected consequence of population structure
428 (Chikhi et al., 2018; Lesturgie et al., 2022; Mazet et al., 2015; Rodríguez et al., 2018) and cannot
429 be interpreted as a demographic decline. Unfortunately, population structure and demographic
430 decline affect the SFS in a similar fashion making impossible to quantitatively disentangle the
431 contribution of both to the observed bottleneck estimated using RAD-seq data (Lesturgie et al.,
432 2022). We stress that investigating local recent changes in connectivity or demographic events will

clearly requires whole genome sequencing coupled with inferential methods based on the IICR (Arredondo et al., 2021) and/or linkage disequilibrium (Boitard, Rodríguez, Jay, Mona, & Austerlitz, 2016). More generally, the next challenge will be to perform a full modelling of species structured in many demes as the grey reef shark. Here we took a simplified approach by considering each sampling site separately and by modelling the unsampled demes to estimate local migration rates. We are aware that in the future more data will be needed to explore complex demographic scenarios integrating RE that include both all sampled demes and the unsampled ones. In this light, the biological interpretation of parameters estimated under simplified models discarding unsampled demes and ignoring RE (as in Walsh et al. (2022)) remains unclear.

Population structure

The results presented so far suggest that dispersal abilities of *C. amblyrhynchos* are similar throughout the Indo-Pacific and independent of the availability of coral reefs. However, this cannot exclude the presence of barriers to gene flow which may have shaped the connectivity between demes. For widely distributed marine species, detecting such barriers may help to delineate management units and to take proper conservation measures in relation to fisheries (Dudgeon et al., 2012). Several evidence point to an absence of barriers to gene flow in the grey reef shark. First of all, we found a strong IBD pattern with a significant correlation between genetic and geographic distances of > 0.9 when considering only PAC samples (Figure 6b) and a linear relation of smaller intensity between IND and PAC samples (Figure S5). Remarkably, these values are not affected by computing geographic distances between sampling sites under an LC approach. Indeed, the permeability values maximizing the correlation are (almost) the same for the different type of habitats. This suggest that different geographic features do not affect the direction of grey

456 reef shark migrations, indicating, albeit indirectly, the absence of barriers to dispersal, consistently
457 with the occasional long-distance swims detected across the open ocean (Barnett et al., 2012;
458 Bonnin et al., 2019; White et al., 2017). When strong IBD is present, it is difficult to attribute a
459 biological meaning to groups identified by clustering algorithms (Meirmans, 2012). Both the
460 *sNMF* and PCA analyses suggested a strong separation between IND and PAC samples (Figure
461 5), with the latter subdivided into two weakly divergent clusters (Figure 5 and S8). The IND
462 ancestral components diminished remarkably continuously eastward, once again supporting an
463 IBD structure (Figure 5a) rather than the presence of barriers to gene flow. This is consistent with
464 the pairwise F_{ST} matrix, where intra Pacific comparisons did not exceed ~0.03 while the inter-
465 oceanic comparisons have an average F_{ST} of ~0.54 (Figure 6a). Defining management units within
466 the PAC seems therefore inappropriate in the case of the grey reef shark, as genetic variations are
467 rather continuous. This contrasts with what was previously suggested by Boissin et al. (2019) at
468 the Pacific scale: however, their results were based on a small number of microsatellites and they
469 did not model IBD between the sampling points.
470 The pitfall of our study is to extrapolate the dynamic of the grey reef shark at the scale of its whole
471 range by focusing mostly on the Pacific Ocean. Indeed, even if the species seems to follow an IBD
472 pattern also from Chagos to Eastern Australia (Momigliano et al., 2017), the level of population
473 differentiation appears to be higher than what we found in the Pacific for similar geographic
474 distances. However, while the distribution of coral reef in the Pacific Ocean is scattered due to the
475 presence of many archipelagos, coral reefs in the Indian Ocean are more concentrated on the
476 coastal edge of the Asian and African continents (Figure S7). The effective distance between
477 sampling sites within the Indian Ocean would therefore be larger than in the Pacific Ocean, where
478 coral reefs would act as stepping stones to facilitate the colonization process and further

migrations. This could also account for the different linear relationship estimated in the Pacific *vs.* the one estimated between Pacific and Indian sampling sites (Figure S5).

Conclusion

We explored the evolutionary history of the grey reef shark throughout most of its range in the Indo-Pacific and contrasted the results with those previously obtained for the blacktip reef shark (Maisano Delser et al., 2019). The two species are among the most abundant reef sharks (MacNeil et al., 2020), share an almost overlapping distribution in the Indo-Pacific and are both strictly coral reef-dependent species. Despite similarities in the RE dynamic, patterns of genetic diversity and population structure are very different between the two species. First, *C. melanopterus* is significantly more structured than *C. amblyrhynchos* at similar spatial distances (for comparison, F_{ST} values are ~30 times higher when comparing French Polynesia vs New Caledonia, see Table S5 of Maisano Delser et al. (2019) and our Table S3). Second, *C. amblyrhynchos* shows homogeneous migration rates and demographic signals throughout its whole distribution whereas *C. melanopterus* is more sensitive to the spatial distribution of coral reef with a connectivity largely dependent on the short scale reef-availability (Maisano Delser et al., 2019). Indeed, migration rates estimated in areas with extensive coral reefs coverage (e.g., the Great Barrier Reef) are much higher compared to those estimated in isolated islands/atolls in the Indo-Pacific (Maisano Delser et al., 2019), something that we did not observe for *C. amblyrhynchos*. All these differences can be explained by the life history traits related to dispersal abilities of the two species, with *C. amblyrhynchos* moving more freely in open sea expanses compared to *C. melanopterus*, lowering the impact of coral density on the observed genetic diversity. However, it will be important in the next future to precisely characterize the extent of the neighbourhood size for both species. To this

end, ecological and genomic data need to be coupled: this will help to carefully decipher how many management units are necessary for species conservation and at which scale they should be established.

Acknowledgement

We are grateful to the Genotoul bioinformatics platform Toulouse Midi-Pyrenees (Bioinfo Genotoul; http://bioinfo.genotoul.fr/) for providing computing resources. We are thankful to Valeriano Parravicini for his input and for providing resources on coral reef distribution in the Indo-Pacific and Romuald Laso-Jadart for critical reading. We thank Jenn Caselle and Darcy Bradley for providing samples from the Phoenix archipelago. This work was supported by two ATM grants (2016 and 2017) from the Muséum National d'Histoire Naturelle to S.M.

Authors contribution

S.M coordinated the study. P.L carried out data analysis with inputs from S.M. P.L. and S.M conceived the study and wrote the paper. C.B., E.C., J.K., J.M., S.T., T.V. and S.P. collected samples, provided reagents and revised the manuscript.

Data archiving

VCF files, SFS and scripts are available from the Dryad Digital Repository: doi:10.5061/dryad.547d7wm9b. Fastq sequence files are available from the GenBank at the National Center for Biotechnology Information short-read archive database (accession number: forthcoming).

References

Allen, G. R. (2008). Conservation hotspots of biodiversity and endemism for Indo-Pacific coral reef fishes. *Aquatic Conservation: Marine and Freshwater Ecosystems*, *18*(5), 541–556. https://doi.org/10.1002/aqc.880

Arredondo, A., Mourato, B., Nguyen, K., Boitard, S., Rodríguez, W., Noûs, C., … Chikhi, L. (2021). Inferring number of populations and changes in connectivity under the n-island model. *Heredity*, *126*(6), 896–912. https://doi.org/10.1038/s41437-021-00426-9

Bailleul, D., Mackenzie, A., Sacchi, O., Poisson, F., Bierne, N., & Arnaud-Haond, S. (2018). Large-scale genetic panmixia in the blue shark (Prionace glauca): A single worldwide population, or a genetic lag-time effect of the "grey zone" of differentiation? *Evolutionary Applications*, *11*(5), 614–630. https://doi.org/10.1111/eva.12591

Barnett, A., Abrantes, K. G., Seymour, J., & Fitzpatrick, R. (2012). Residency and spatial use by reef sharks of an isolated seamount and its implications for conservation. *PLoS ONE*, *7*(5), 1–12. https://doi.org/10.1371/journal.pone.0036574

Benazzo, A., Trucchi, E., Cahill, J. A., Delser, P. M., Mona, S., Fumagalli, M., … Bertorelle, G. (2017). Survival and divergence in a small group: The extraordinary genomic history of the endangered Apennine brown bear stragglers. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(45), E9589–E9597. https://doi.org/10.1073/pnas.1707279114

Bertorelle, G., Benazzo, A., & Mona, S. (2010). ABC as a flexible framework to estimate demography over space and time: Some cons, many pros. *Molecular Ecology*, *19*(13), 2609–2625. https://doi.org/10.1111/j.1365-294X.2010.04690.x

Boissin, E., Thorrold, S. R., Braun, C. D., Zhou, Y., Clua, E. E., & Planes, S. (2019). Contrasting

global, regional and local patterns of genetic structure in gray reef shark populations from the Indo-Pacific region. *Scientific Reports*, *9*(1), 1–9. https://doi.org/10.1038/s41598-019-52221-6

Boitard, S., Rodríguez, W., Jay, F., Mona, S., & Austerlitz, F. (2016). Inferring Population Size History from Large Samples of Genome-Wide Molecular Data - An Approximate Bayesian Computation Approach. *PLOS Genetics*, *12*(3), e1005877. https://doi.org/10.1371/journal.pgen.1005877

Bonnin, L., Robbins, W. D., Boussarie, G., Kiszka, J. J., Dagorn, L., Mouillot, D., & Vigliola, L. (2019). Repeated long-range migrations of adult males in a common Indo-Pacific reef shark. *Coral Reefs*, *38*(6), 1121–1132. https://doi.org/10.1007/s00338-019-01858-w

Bornatowski, H., Navia, A. F., Braga, R. R., Abilhoa, V., & Corrêa, M. F. M. (2014). Ecological importance of sharks and rays in a structural foodweb analysis in southern Brazil. *ICES Journal of Marine Science*, *71*(7), 1586–1592. https://doi.org/10.1093/icesjms/fsu025

Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A., & Cresko, W. A. (2013). Stacks: an analysis tool set for population genomics. *Molecular Ecology*, *22*(11), 3124–3140. https://doi.org/10.1111/mec.12354

Chikhi, L, Rodríguez, W., Grusea, S., Santos, P., Boitard, S., & Mazet, O. (2018). The IICR (inverse instantaneous coalescence rate) as a summary of genomic diversity: Insights into demographic inference and model choice. *Heredity*, *120*(1), 13–24. https://doi.org/10.1038/s41437-017-0005-6

Chikhi, L, Sousa, V. C., Luisi, P., Goossens, B., & Beaumont, M. A. (2010). The Confounding Effects of Population Structure, Genetic Diversity and the Sampling Scheme on the Detection and Quantification of Population Size Changes. *Genetics*, *186*(3), 983–995.

https://doi.org/10.1534/genetics.110.118661

Compagno, L. J. . (2001). Sharks of the world. An annotated and illustrated catalogue of shark species known to date. Volume 2. Bullhead, mackerel and carpet sharks (Heterodontiformes, Lamniformes and Orectolobiformes). *FAO Species Catalogue for Fishery Purposes*, *2*(1), 108–125.

Corrigan, S., Lowther, A. D., Beheregaray, L. B., Bruce, B. D., Cliff, G., Duffy, C. A., … Rogers, P. J. (2018). Population Connectivity of the Highly Migratory Shortfin Mako (Isurus oxyrinchus Rafinesque 1810) and Implications for Management in the Southern Hemisphere. *Frontiers in Ecology and Evolution*, *6*(NOV), 1–15. https://doi.org/10.3389/fevo.2018.00187

Cowman, P. F., & Bellwood, D. R. (2013). The historical biogeography of coral reef fishes: Global patterns of origination and dispersal. *Journal of Biogeography*, *40*(2), 209–224. https://doi.org/10.1111/jbi.12003

Delrieu-Trottin, E., Hubert, N., Giles, E. C., Chifflet-Belle, P., Suwalski, A., Neglia, V., … Saenz-Agudelo, P. (2020). Coping with Pleistocene climatic fluctuations: Demographic responses in remote endemic reef fishes. *Molecular Ecology*, *29*(12), 2218–2233. https://doi.org/10.1111/mec.15478

Dudgeon, C. L., Blower, D. C., Broderick, D., Giles, J. L., Holmes, B. J., Kashiwagi, T., … Ovenden, J. R. (2012). A review of the application of molecular genetics for fisheries management and conservation of sharks and rays. *Journal of Fish Biology*, *80*(5), 1789–1843. https://doi.org/10.1111/j.1095-8649.2012.03265.x

Dulvy, N. K., Pacoureau, N., Rigby, C. L., Pollom, R. A., Jabado, R. W., Ebert, D. A., … Simpfendorfer, C. A. (2021). Overfishing drives over one-third of all sharks and rays

toward a global extinction crisis. *Current Biology*, 1–15. https://doi.org/10.1016/j.cub.2021.08.062

Dwyer, R. G., Krueck, N. C., Udyawer, V., Heupel, M. R., Chapman, D., Pratt, H. L., … Simpfendorfer, C. A. (2020). Individual and Population Benefits of Marine Reserves for Reef Sharks. *Current Biology*, *30*(3), 480-489.e5. https://doi.org/10.1016/j.cub.2019.12.005

Eaton, D. A. R. (2014). PyRAD: Assembly of de novo RADseq loci for phylogenetic analyses. *Bioinformatics*, *30*(13), 1844–1849. https://doi.org/10.1093/bioinformatics/btu121

Espinoza, M., Heupel, M. R., Tobin, A. J., & Simpfendorfer, C. A. (2014). Residency patterns and movements of grey reef sharks (Carcharhinus amblyrhynchos) in semi-isolated coral reef habitats. *Marine Biology*, *162*(2), 343–358. https://doi.org/10.1007/s00227-014-2572-x

Excoffier, L., & Foll, M. (2011). fastsimcoal: A continuous-time coalescent simulator of genomic diversity under arbitrarily complex evolutionary scenarios. *Bioinformatics*, *27*(9), 1332–1334. https://doi.org/10.1093/bioinformatics/btr124

Excoffier, L., Foll, M., & Petit, R. J. (2009). Genetic Consequences of Range Expansions. *Annual Review of Ecology, Evolution, and Systematics*, *40*(1), 481–501. https://doi.org/10.1146/annurev.ecolsys.39.110707.173414

Field, I. C., Meekan, M. G., Speed, C. W., White, W., & Bradshaw, C. J. A. (2011). Quantifying movement patterns for shark conservation at remote coral atolls in the Indian Ocean. *Coral Reefs*, *30*(1), 61–71. https://doi.org/10.1007/s00338-010-0699-x

Frichot, E., & François, O. (2015). LEA: An R package for landscape and ecological association studies. *Methods in Ecology and Evolution*, *6*(8), 925–929. https://doi.org/10.1111/2041-210X.12382

Friedlander, A. M., & DeMartini, E. E. (2002). Contrasts in density, size, and biomass of reef

615 fishes between the northwestern and the main Hawaiian islands: The effects of fishing down
616 apex predators. *Marine Ecology Progress Series*, *230*, 253–264.
617 https://doi.org/10.3354/meps230253
618 Heled, J., & Drummond, A. J. (2008). Bayesian inference of population size history from
619 multiple loci. *BMC Evolutionary Biology*, *8*(1), 289. https://doi.org/10.1186/1471-2148-8-
620 289
621 Heller, R., Chikhi, L., & Siegismund, H. R. (2013). The Confounding Effect of Population
622 Structure on Bayesian Skyline Plot Inferences of Demographic History. *PLoS ONE*, *8*(5).
623 https://doi.org/10.1371/journal.pone.0062992
624 Heller, R., Nursyifa, C., Garcia-Erill, G., Salmona, J., Chikhi, L., Meisner, J., … Albrechtsen, A.
625 (2021). A reference-free approach to analyse RADseq data using standard next generation
626 sequencing toolkits. *Molecular Ecology Resources*, *21*(4), 1085–1097.
627 https://doi.org/10.1111/1755-0998.13324
628 Hijmans, R. J. (2020). *Raster: Geographic Data Analysis and Modeling.*
629 Jombart, T., Devillard, S., & Balloux, F. (2010). Discriminant analysis of principal components:
630 A new method for the analysis of genetically structured populations. *BMC Genetics*, *11*(1),
631 94. https://doi.org/10.1186/1471-2156-11-94
632 Khimoun, A., Doums, C., Molet, M., Kaufmann, B., Peronnet, R., Eyer, P. A., & Mona, S.
633 (2020). Urbanization without isolation: The absence of genetic structure among cities and
634 forests in the tiny acorn ant Temnothorax nylanderi. *Biology Letters*, *16*(1).
635 https://doi.org/10.1098/rsbl.2019.0741
636 Korneliussen, T. S., Albrechtsen, A., & Nielsen, R. (2014). ANGSD: Analysis of Next
637 Generation Sequencing Data. *BMC Bioinformatics*, *15*(1), 1–13.

https://doi.org/10.1186/s12859-014-0356-4

Lesturgie, P., Planes, S., & Mona, S. (2022). Coalescence times, life history traits and conservation concerns: An example from four coastal shark species from the Indo-Pacific. *Molecular Ecology Resources*, *22*(2), 554–566. https://doi.org/10.1111/1755-0998.13487

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, *25*(14), 1754–1760. https://doi.org/10.1093/bioinformatics/btp324

Li, H., & Durbin, R. (2011). Inference of human population history from individual whole-genome sequences. *Nature*, *475*(7357), 493–496. https://doi.org/10.1038/nature10231

Liu, X., & Fu, Y.-X. (2015). Exploring population size changes using SNP frequency spectra. *Nature Genetics*, *47*(5), 555–559. https://doi.org/10.1038/ng.3254

MacNeil, M. A., Chapman, D. D., Heupel, M., Simpfendorfer, C. A., Heithaus, M., Meekan, M., … Cinner, J. E. (2020). Global status and conservation potential of reef sharks. *Nature*, *583*(7818), 801–806. https://doi.org/10.1038/s41586-020-2519-y

Maisano Delser, P., Corrigan, S., Duckett, D., Suwalski, A., Veuille, M., Planes, S., … Mona, S. (2019). Demographic inferences after a range expansion can be biased: the test case of the blacktip reef shark (Carcharhinus melanopterus). *Heredity*, *122*(6), 759–769. https://doi.org/10.1038/s41437-018-0164-0

Maisano Delser, P., Corrigan, S., Hale, M., Li, C., Veuille, M., Planes, S., … Mona, S. (2016). Population genomics of C. melanopterus using target gene capture data: Demographic inferences and conservation perspectives. *Scientific Reports*, *6*(April), 1–12. https://doi.org/10.1038/srep33753

Mantel, N. (1967). The Detection of Disease Clustering and a Generalized Regression Approach.

Cancer Research, *27*(2), 209–220.

Mazet, O., Rodríguez, W., Grusea, S., Boitard, S., & Chikhi, L. (2016). On the importance of being structured: instantaneous coalescence rates and human evolution—lessons for ancestral population size inference? *Heredity*, *116*(4), 362–371. https://doi.org/10.1038/hdy.2015.104

Mazet, O, Rodríguez, W., & Chikhi, L. (2015). Demographic inference using genetic data from a single individual: Separating population size variation from population structure. *Theoretical Population Biology*, *104*, 46–58. https://doi.org/10.1016/j.tpb.2015.06.003

Meirmans, P. G. (2012). The trouble with isolation by distance. *Molecular Ecology*, *21*(12), 2839–2846. https://doi.org/10.1111/j.1365-294X.2012.05578.x

Momigliano, P., Harcourt, R., Robbins, W. D., Jaiteh, V., Mahardika, G. N., Sembiring, A., & Stow, A. (2017). Genetic structure and signatures of selection in grey reef sharks (Carcharhinus amblyrhynchos). *Heredity*, *119*(3), 142–153. https://doi.org/10.1038/hdy.2017.21

Momigliano, P., Harcourt, R., Robbins, W. D., & Stow, A. (2015). Connectivity in grey reef sharks (Carcharhinus amblyrhynchos) determined using empirical and simulated genetic data. *Scientific Reports*, *5*(August), 1–9. https://doi.org/10.1038/srep13229

Mona, S. (2017). On the role played by the carrying capacity and the ancestral population size during a range expansion. *Heredity*, *118*(2), 143–153. https://doi.org/10.1038/hdy.2016.73

Mona, S., Ray, N., Arenas, M., & Excoffier, L. (2014). Genetic consequences of habitat fragmentation during a range expansion. *Heredity*, *112*(3), 291–299. https://doi.org/10.1038/hdy.2013.105

Mourier, J., Mills, S. C., & Planes, S. (2013). Population structure, spatial distribution and life-

history traits of blacktip reef sharks Carcharhinus melanopterus. *Journal of Fish Biology*, *82*(3), 979–993. https://doi.org/10.1111/jfb.12039

Mourier, Johann, & Planes, S. (2013). Direct genetic evidence for reproductive philopatry and associated fine-scale migrations in female blacktip reef sharks (Carcharhinus melanopterus) in French Polynesia. *Molecular Ecology*, *22*(1), 201–214. https://doi.org/10.1111/mec.12103

Myers, R. A., Baum, J. K., Shepherd, T. D., Powers, S. P., & Peterson, C. H. (2007). Cascading effects of the loss of apex predatory sharks from a coastal ocean. *Science*, *315*(5820), 1846–1850. https://doi.org/10.1126/science.1138657

Pazmiño, D. A., Maes, G. E., Green, M. E., Simpfendorfer, C. A., Hoyos-Padilla, E. M., Duffy, C. J. A., … Van Herwerden, L. (2018). Strong trans-Pacific break and local conservation units in the Galapagos shark (Carcharhinus galapagensis) revealed by genome-wide cytonuclear markers. *Heredity*, *120*(5), 407–421. https://doi.org/10.1038/s41437-017-0025-2

Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S., & Hoekstra, H. E. (2012). Double Digest RADseq: An Inexpensive Method for De Novo SNP Discovery and Genotyping in Model and Non-Model Species. *PLoS ONE*, *7*(5), e37135. https://doi.org/10.1371/journal.pone.0037135

Pfeifer, B., Wittelsbürger, U., Ramos-Onsins, S. E., & Lercher, M. J. (2014). PopGenome: An efficient swiss army knife for population genomic analyses in R. *Molecular Biology and Evolution*, *31*(7), 1929–1936. https://doi.org/10.1093/molbev/msu136

Pirog, A., Jaquemet, S., Ravigné, V., Cliff, G., Clua, E., Holmes, B. J., … Magalon, H. (2019). Genetic population structure and demography of an apex predator, the tiger shark

Galeocerdo cuvier. *Ecology and Evolution*, *9*(10), 5551–5571. https://doi.org/10.1002/ece3.5111

Pudlo, P., Marin, J.-M. M., Estoup, A., Cornuet, J.-M. M., Gautier, M., & Robert, C. P. (2016). Reliable ABC model choice via random forests. *Bioinformatics*, *32*(6), 859–866. https://doi.org/10.1093/bioinformatics/btv684

Ramachandran, S., Deshpande, O., Roseman, C. C., Rosenberg, N. A., Feldman, M. W., & Cavalli-Sforza, L. L. (2005). Support from the relationship of genetic and geographic distance in human populations for a serial founder effect originating in Africa. *Proceedings of the National Academy of Sciences*, *102*(44), 15942–15947. https://doi.org/10.1073/pnas.0507611102

Ray, N., Currat, M., & Excoffier, L. (2003). Intra-deme molecular diversity in spatially expanding populations. *Molecular Biology and Evolution*, *20*(1), 76–86. https://doi.org/10.1093/molbev/msg009

Raynal, L., Marin, J. M., Pudlo, P., Ribatet, M., Robert, C. P., & Estoup, A. (2019). ABC random forests for Bayesian parameter inference. *Bioinformatics*, *35*(10), 1720–1728. https://doi.org/10.1093/bioinformatics/bty867

Robbins, W. D., Hisano, M., Connolly, S. R., & Choat, J. H. (2006). Ongoing Collapse of Coral-Reef Shark Populations. *Current Biology*, *16*(23), 2314–2319. https://doi.org/10.1016/j.cub.2006.09.044

Rochette, N. C., Rivera-Colón, A. G., & Catchen, J. M. (2019). Stacks 2: Analytical methods for paired-end sequencing improve RADseq-based population genomics. *Molecular Ecology*, *28*(21), 4737–4754. https://doi.org/10.1111/mec.15253

Rodríguez, W., Mazet, O., Grusea, S., Arredondo, A., Corujo, J. M., Boitard, S., & Chikhi, L.

(2018). The IICR and the non-stationary structured coalescent: towards demographic inference with arbitrary changes in population structure. *Heredity*, *121*(6), 663–678. https://doi.org/10.1038/s41437-018-0148-0

Speed, C. W., Meekan, M. G., Field, I. C., McMahon, C. R., Harcourt, R. G., Stevens, J. D., … Bradshaw, C. J. A. (2016). Reef shark movements relative to a coastal marine protected area. *Regional Studies in Marine Science*, *3*, 58–66. https://doi.org/10.1016/j.rsma.2015.05.002

Steiner, C. C., Putnam, A. S., Hoeck, P. E. A., & Ryder, O. A. (2013). Conservation Genomics of Threatened Animal Species. *Annual Review of Animal Biosciences*, *1*(1), 261–281. https://doi.org/10.1146/annurev-animal-031412-103636

Thioulouse, J., & Dray, S. (2007). Interactive multivariate data analysis in R with the ade4 and ade4TkGUI packages. *Journal of Statistical Software*, *22*(5), 1–14. https://doi.org/10.18637/jss.v022.i05

van Etten, J. (2017). R package gdistance: Distances and routes on geographical grids. *Journal of Statistical Software*, *76*(1). https://doi.org/10.18637/jss.v076.i13

Walsh, C. A. J., Momigliano, P., Boussarie, G., Robbins, W. D., Bonnin, L., Fauvelot, C., … Manel, S. (2022). *Genomic insights into the historical and contemporary demographics of the grey reef shark*. (February). https://doi.org/10.1038/s41437-022-00514-4

White, T. D., Carlisle, A. B., Kroodsma, D. A., Block, B. A., Casagrandi, R., De Leo, G. A., … McCauley, D. J. (2017). Assessing the effectiveness of a large marine protected area for reef shark conservation. *Biological Conservation*, *207*, 64–71. https://doi.org/10.1016/j.biocon.2017.01.009

Whitney, N. M., Robbins, W. D., Schultz, J. K., Bowen, B. W., & Holland, K. N. (2012).

753 Oceanic dispersal in a sedentary reef shark (Triaenodon obesus): Genetic evidence for
754 extensive connectivity without a pelagic larval stage. *Journal of Biogeography*, *39*(6),
755 1144–1156. https://doi.org/10.1111/j.1365-2699.2011.02660.x
756
757
758

Tables

Table 1. Summary Statistics. Sample size (n), total number of loci (monomorphic included) (n_{loci}) and SNPs (n_{SNP}), mean pairwise difference (θ_π), Watterson theta (θ_w), Tajima's *D* (TD) for all sampling sites (ranged from west to east).

Region	Group	Sampling site	n	n_{loci}	n_{SNP}	θ_π[†]	θ_w[†]	TD[‡]
IND	IND	Juan	13	95027	45635	1.18	1.09	**0.32**
		Zelee	6	146858	62674	1.30	1.23	**0.26**
COR[§]	CHE	Bampton	10	89958	82869	2.14	2.26	**-0.22**
		Avond	5	125710	87817	2.10	2.15	**-0.12**
	NCA	Belep	7	120038	103258	2.30	2.35	**-0.11**
		Poindimie	5	107464	72995	2.07	2.09	**-0.05**
CPA[§]	PHO	Niku	21	49922	53349	2.02	2.16	**-0.25**
		McKean	7	112711	88258	2.13	2.14	-0.01
		Orona	11	81725	75423	2.15	2.20	**-0.09**
		Kanton	10	99720	87202	2.12	2.14	**-0.05**
		Birnie[¶]	2	-	-	-	-	-
		Enderbury	13	76314	72221	2.09	2.16	**-0.12**
	PAL	Palmyra	38	35594	36982	1.66	1.84	**-0.35**
	POL	Moorea	5	104050	68380	2.03	2.02	0.02
		Fakarava	17	71715	66559	2.01	1.97	**0.08**
		Faaite[¶]	1	-	-	-	-	-
		Raraka[¶]	1	-	-	-	-	-
		Nengo[¶]	1	-	-	-	-	-

[†] Mean pairwise difference and Watterson theta are expressed per site and are multiplied by a 10^3 factor.

[‡] Tajima's D values in bold are significant (P <0.001).

[§] COR and CPA regions are from the Pacific Ocean (PAC).

[¶] Summary statistics were not computed in sampling sites with n < 5.

Table 2. ABC estimation. Posterior probability (PP) of the Stepping Stone model (SST) and its parameters (median value and 95% credible interval in parentheses).

Region	Group	Sampling site	PP	*Nm*	T_{col}	N_{anc}
IND	IND	Juan	0.67	5.7 (1.77 - 17.72)	257800 (8086 - 658471)	21086 (399 - 52652)
COR§	CHE	Bampton	0.73	11.41 (3.97 - 19.03)	188782 (127761 - 577503)	45965 (27556 - 49856)
	NCA	Belep	0.51	7.8 (2.84 - 20.82)	241218 (112840 - 843171)	49239 (7346 - 56316)
CPA§	PHO	Enderbury	0.65	8.36 (2.9 - 20.9)	197070 (95260 - 678828)	43602 (14665 - 51030)
		Kanton	0.7	8.16 (2.84 - 16.55)	257718 (118094 - 789320)	41236 (2534 - 52613)
		McKean	0.6	7.09 (2.98 - 15.25)	621535 (158650 - 836223)	18881 (4968 - 51387)
		Niku	0.59	14.1 (3 - 30.55)	152035 (66928 - 598129)	43495 (9184 - 48625)
		Orona	0.48	7.7 (2.93 - 15.31)	269621 (137304 - 799518)	41680 (4575 - 51152)
	PAL	Palmyra	0.73	13.39 (4.16 - 27.22)	142756 (62402 - 445380)	32542 (9502 - 37524)
	POL	Fakarava	0.72	10.2 (2.68 - 15.34)	256744 (110875 - 780150)	40502 (3091 - 49533)
			Priors	*U [0.0001 ; 100]	U [100 ; 1500000]	U [100 ; 100000]

***** The prior distribution of *Nm* is the product of two uniforms (one for *N* and one for *m*)
§ COR and CPA regions are from the Pacific Ocean (PAC).

Figure Legends

Figure 1. Map of the sampling sites. From west to east, Indian Ocean (IND): Juan (n = 13) and Zelee (n = 6); Chesterfield islands (CHE): Bampton (n = 10) and Avond (n = 5), New Caledonia (NCA): Belep (n = 7) and Poindimie (n = 5); Phoenix islands (PHO): Niku (n = 21), Mckean (n = 7), Orona (n = 11), Kanton (n = 10), Birnie (n = 2) and Enderbury (n = 13); Palmyra (PAL, n = 38); French Polynesia (POL): Moorea (n = 5), Fakarava (n = 17), Faaite (n = 1), Raraka (n = 1), and Nengo (n = 1). Colours represent the region of origin of the sampling sites: Indian Ocean (IND, yellow), Coral Sea (COR, red) and Central Pacific Ocean (CPA, blue).

Figure 2. Demographic scenarios investigated in all populations with Nind ≥ 7 through an Approximate Bayesian Computation (ABC) framework. N_{anc}: ancestral effective population size; T_c: time of effective population size change (NS only); N_{mod}: modern effective population size (NS only); T_{col}: colonization time of the array of demes (FIM and SST); $D_{1\text{-}100}$: demes (FIM and SST). Arrows represent the migrants exchanged (Nm) between demes. Details on each scenario are presented in the main text.

Figure 3. Correlation map between genetic diversity (θ_π) and Least Cost (LC) distances when considering Pacific Ocean sampling sites only. Each cell is coloured according to the correlation coefficient value computed between θ_π and the LC distance from the putative origin of the range expansion (RE). Black dots represent the sampling sites considered.

Figure 4. Variation of the effective population size (*Ne*) through time and its 75% confidence interval estimated by the *stairwayplot* for sampling sites of n ≥ 7 in IND (a), COR (b) and CPA (c) regions.

Figure 5. Individual-based population structure analyses. Ancestry proportions retrieved using the *sNMF* algorithm with K=2 and K=3 ancestral populations (a) and Principal Component Analysis (b).

Figure 6. Population-based population structure analyses computed with populations of n ≥ 5. Heat map representing the pairwise F_{ST} values between sampling sites (a) and Isolation by distance (IBD) plot considering Pacific sampling sites only (b).

Figures

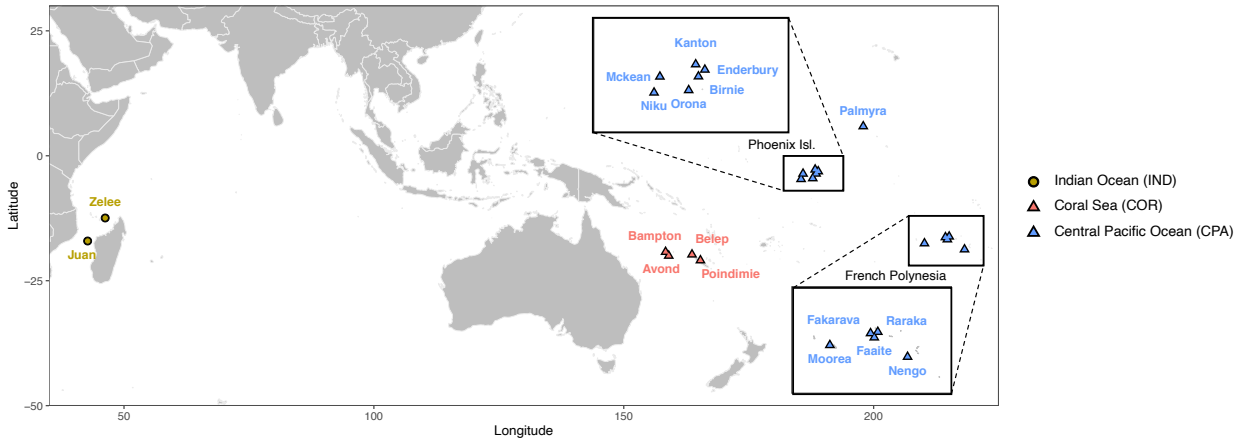


Figure 1. Map of the sampling sites. From west to east, Indian Ocean (IND): Juan (n = 13) and Zelee (n = 6); Chesterfield islands (CHE): Bampton (n = 10) and Avond (n = 5), New Caledonia (NCA): Belep (n = 7) and Poindimie (n = 5); Phoenix islands (PHO): Niku (n = 21), Mckean (n = 7), Orona (n = 11), Kanton (n = 10), Birnie (n = 2) and Enderbury (n = 13); Palmyra (PAL, n = 38); French Polynesia (POL): Moorea (n = 5), Fakarava (n = 17), Faaite (n = 1), Raraka (n = 1), and Nengo (n = 1). Colours represent the region of origin of the sampling sites: Indian Ocean (IND, yellow), Coral Sea (COR, red) and Central Pacific Ocean (CPA, blue).

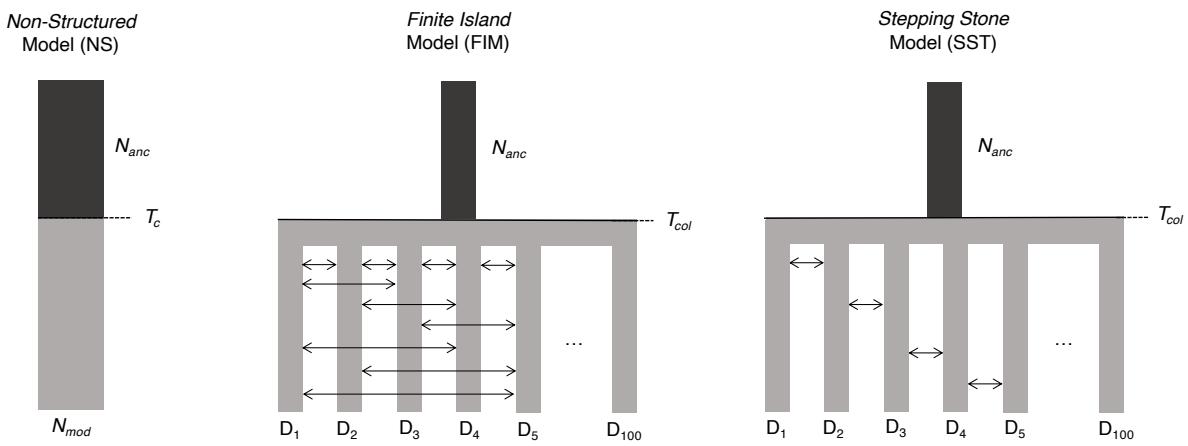


Figure 2. Demographic scenarios investigated in all populations with Nind ≥ 7 through an Approximate Bayesian Computation (ABC) framework. N_{anc}: ancestral effective population size; T_c: time of effective population size change (NS only); N_{mod}: modern effective population size (NS only); T_{col}: colonization time of the array of demes (FIM and SST); $D_{1\text{-}100}$: demes (FIM and SST). Arrows represent the migrants exchanged (Nm) between demes. Details on each scenario are presented in the main text.

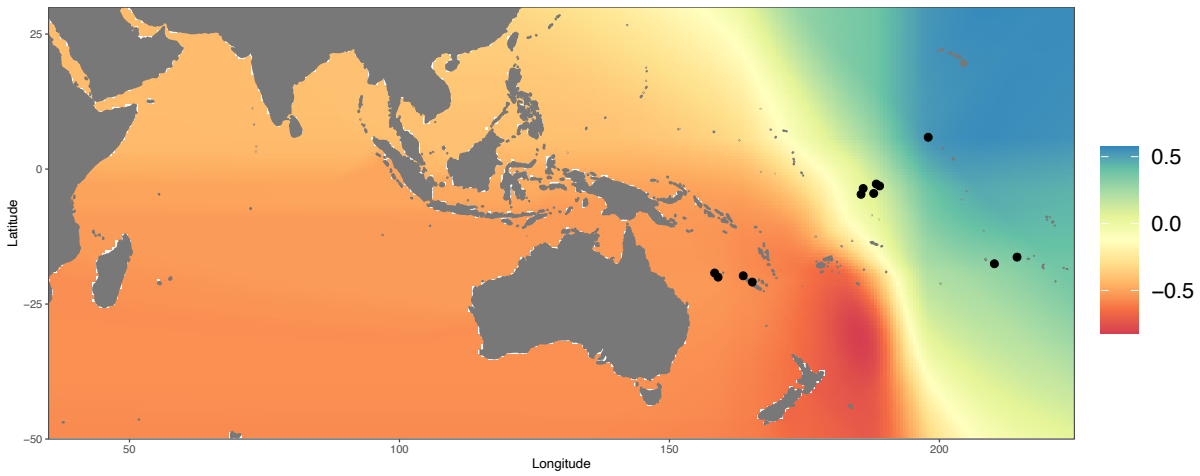
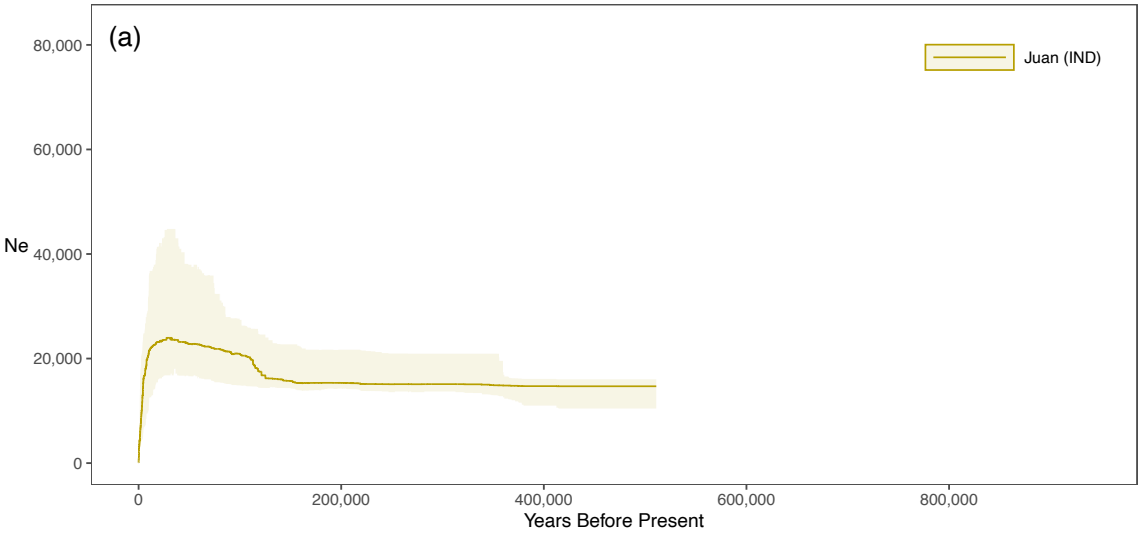
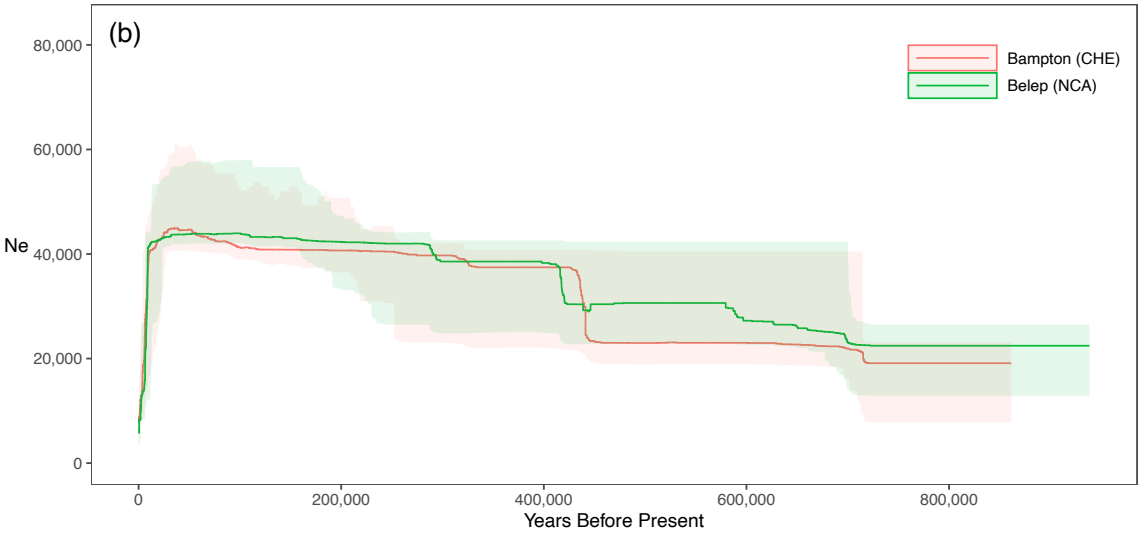


Figure 3. Correlation map between genetic diversity (θ_π) and Least Cost (LC) distances when considering Pacific Ocean sampling sites only. Each cell is coloured according to the correlation coefficient value computed between θ_π and the LC distance from the putative origin of the range expansion (RE). Black dots represent the sampling sites considered.


(a)
Juan (IND)
Ne
80,000
60,000
40,000
20,000
0
0
200,000
400,000
600,000
800,000
Years Before Present


(b)
Bampton (CHE)
Belep (NCA)
Ne
80,000
60,000
40,000
20,000
0
0
200,000
400,000
600,000
800,000
Years Before Present

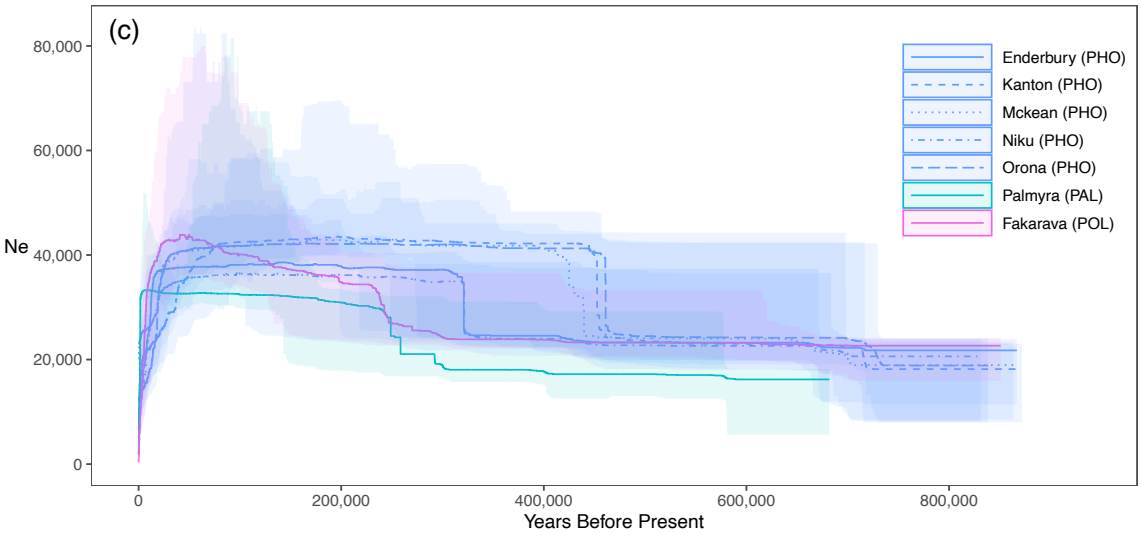

(c)
Enderbury (PHO)
Kanton (PHO)
Mckean (PHO)
Niku (PHO)
Orona (PHO)
Palmyra (PAL)
Fakarava (POL)
Ne
80,000
60,000
40,000
20,000
0
0
200,000
400,000
600,000
800,000
Years Before Present

Figure 4. Variation of the effective population size (*Ne*) through time and its 75% confidence interval estimated by the *stairwayplot* for sampling sites of n ≥ 7 in IND (a), COR (b) and CPA (c) regions.

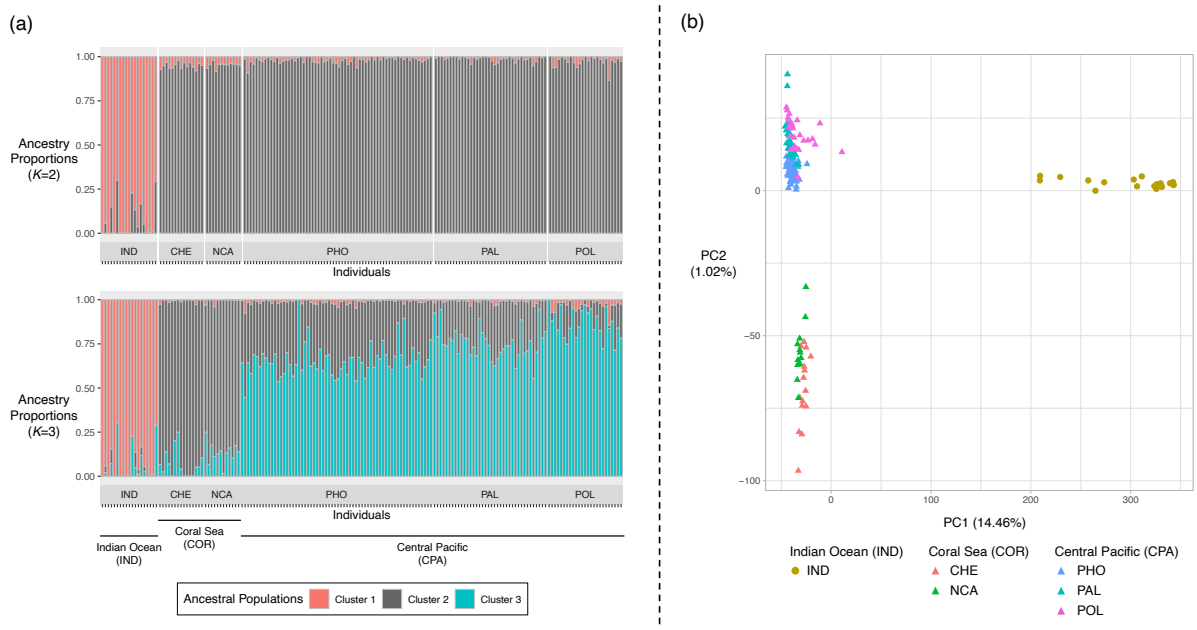


Figure 5. Individual-based population structure analyses. Ancestry proportions retrieved using the *sNMF* algorithm with K=2 and K=3 ancestral populations (a) and Principal Component Analysis (b).

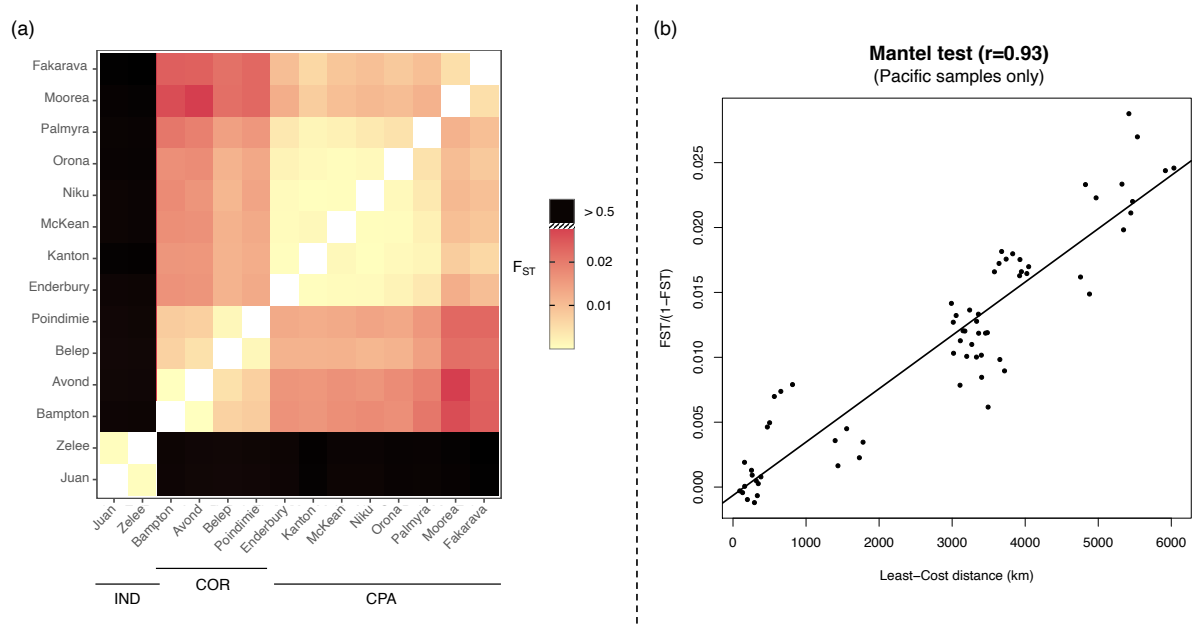


Figure 6. Population-based population structure analyses computed with populations of n ≥ 5. Heat map representing the pairwise F_{ST} values between sampling sites (a) and Isolation by distance (IBD) plot considering Pacific sampling sites only (b).