

A tale of two reef sharks: evolutionary history of *Carcharhinus amblyrhynchos* and similarities with *Carcharhinus melanopterus* as revealed by population genomics data

Running title: Population genetics of the grey reef shark.

Pierre Lesturgie¹, Serge Planes^{2,3}, Stefano Mona^{1,2,4,*}

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

² Laboratoire d'Excellence CORAIL, Papetoai, French Polynesia.

³ PSL Research University : EPHE-UPVD-CNRS, USR 3278 CRIOBE, Université de Perpignan, 52 Avenue Paul Alduy, 66860, Perpignan, Cedex, France

⁴ EPHE, PSL Research University, Paris, France.

* Corresponding author. E-mail : stefano.mona@mnhn.fr

Abstract

We investigated the evolutionary history of the grey reef shark (*Carcharhinus amblyrhynchos*) in the Indo-Pacific (IP) by harnessing the power of thousands of RAD-seq loci sequenced in 175 individuals across its range. We bring strong evidences of the occurrence of a range expansion (RE) originating close to the Indo-Australian Archipelago (IAA) from which two stepping-stone waves (east and westward) started, colonizing almost the entire IP. Consequently, the demographic history of *C. amblyrhynchos* is best explained by a meta-population model, for which we estimated a homogenous connectivity throughout its range ($Nm \sim 10$ per generation) by means of coalescent modeling coupled with an Approximate Bayesian Computation framework. An isolation by distance model further highlights the absence of either genetic barriers or preferential migration corridors, despite the dependency of *C. amblyrhynchos* from coral reefs occurrence. This is consistent with the long-distance swims observed, suggesting that the strong genetic structure at the IP scale ($F_{ST} \sim 0.56$ between its ends) is rather the consequence of its large current distribution. We contrasted these results with those previously obtained for the sympatric but strictly lagoon-associated *Carcharhinus melanopterus*, a species well known for its restricted dispersal ability. While *C. melanopterus* exhibits a similar RE dynamic, it is characterized by stronger genetic structure and a non-homogeneous connectivity largely dependent on local coral reefs availability. The comparison between the two species sheds new light on shark evolution, emphasizing the role of IAA as source of biodiversity and of dispersal ability in shaping the extent of population structure and genetic diversity.

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

Introduction

More than 25% of shark species are currently threatened with extinction 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 their food web (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, low scale management might not always be consistent with the mobility range and the dispersal ability of sharks. Genetics and ecological evidences identified both species with very restricted home ranges (Mourier, Mills, & Planes, 2013; Whitney, Robbins, Schultz, Bowen, & Holland, 2012) and species capable of crossing large sea expanses (Bailleul et al., 2018; Corrigan et al., 2018; Pirog et al., 2019), which should be better managed at the oceanic or global scale. Designing conservation actions is therefore a difficult task requiring the knowledge both of 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. Demographic models need to be designed to be as consistent as possible with the biology of the species considered, still accounting for the trade-off between complexity and feasibility. Unfortunately, for computational reasons, many commonly used *software* implement, under different algorithms, mostly *unstructured* models, i.e., models that consider the population under investigation as isolated or panmictic (Heled & Drummond, 2008; Heller, Chikhi, & Siegmund, 2013; Li & Durbin, 2011; Liu & Fu, 2015). Except for highly vagile species which are panmictic at a large scale such as the tiger shark (Pirog et al., 2019) or the mako shark (Corrigan et al., 2018), 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 determines 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 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, Planes, & Mona, 2021; 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 living in the Indo-Pacific. Despite *C. amblyrhynchos* is believed to be one of the most abundant reef sharks in the Indo-Pacific, it is considered endangered with a decreasing trend by the IUCN red list of threatened species. Robbins et al.

(2006) projected a local decline in abundance of ~17% per year in Northern Australia. With a mean size of ~190 cm (Compagno, 2001), *C. amblyrhynchos* inhabits either the fringing or the barrier reef 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). However, tagging evidences 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 genetic studies using both microsatellites and RAD-sequencing did not find signature of genetic structure at a low geographic scale such the Great Barrier Reef (Momigliano et al., 2017; Momigliano, Harcourt, Robbins, & Stow, 2015) and the Phoenix archipelago (Boissin et al., 2019). Conversely, isolation by distance patterns have been found at larger scale and some evidences have been provided showing 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). These evidences cannot be easily reconciled with tagging data, and a deep reappraisal of grey reef shark connectivity is warranted. To shed light on these contrasting findings, we sampled 203 individuals of *C. amblyrhynchos* in 18 sites covering most of its distribution (Figure 1) and sequenced them all following a double digestion RAD-seq protocol (dd-RADseq; Peterson, Weber, Kay, Fisher, & Hoekstra, 2012). The large panel of assembled loci was then used to: (i) detect the occurrence and eventually locate the origin 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 back tip reef shark (*Carcharhinus melanopterus*). 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

Rad sequencing

We collected 203 samples of *C. amblyrhynchos* to cover most of its longitudinal distribution range (Figure 1), with two sampling sites in the Indian Ocean (IND – Juan and Zelee) 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 were chosen in the Central and Easter Pacific (CPA): six in the Phoenix Islands (Enderbury, Kanton, McKean, Niku, Orona and Birnie) one in the 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-digestion DNA sequencing* (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 (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 an average coverage (over individuals and loci) of ~10x (see Results).

Previous work suggested that this value may bias a correct genotype calling under commonly used algorithm by skewing the site frequency spectrum (SFS) towards an excess of low frequency variants (Mona et al. in prep). 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 can be removed before the downstream analyses.

Genetic diversity

To compute genetic diversity and perform demographic inferences, we followed the genotype free estimation of allele frequencies pipeline implemented in the software *ANGSD* v.0.923 (Korneliussen et al., 2014). This approach has been suggested to be more efficient for low to medium coverage *next-generation sequencing* (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) which we applied to each population 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; 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 *-minQ 20* 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 population (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 (Waterson's Theta, θ_w) and Tajima's D (*TD*) directly from the SFS. θ_π and θ_w were standardized per sites (i.e., taking into account both monomorphic and polymorphic sites) and significance of *TD* was evaluated under 1,000 coalescent simulations of a constant population model with size θ_π .

Range Expansion

Decay of genetic diversity. Genetic diversity, here measured in each population as θ_π , is expected to decay as a function of the distance from the origin of the range expansion (Ramachandran et al., 2005). However, linear distances may poorly represent the capacity of individuals to move between two points, as migration can be influenced by many factors such as bathymetry, residency or habitat preference among others. To take all these features into account, we constructed a raster of 67894 cells using the R package *raster* (Hijmans, 2020), where each cell measures $\sim 60\text{km}^2$ and 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 *gdistance* R package under the *Least Cost* (LC) criterion algorithm (van Etten, 2017) and the genetic distances were measured by the F_{ST} (see below). After this step, we considered each marine cells of the raster to be a potential source of origin of the range expansion (RE) and computed its distance from the sampled populations under the LC criterion with the most likely permeability values previously estimated. We correlated these distances with the genetic diversity of each population to identify areas with more negative values, which are likely associated with the origin of the RE (Ramachandran et al., 2005).

Frequency of derived alleles. The second method used to detect RE is based on the comparison of the frequency of derived alleles shared between populations, which is expected to increase with the distance from the center of origin of the RE (Slatkin & Excoffier, 2012). This method is based on the directionality index ψ as defined in Peter and Slatkin (2013), where ψ represents the average difference in the frequency of derived alleles shared between two populations. Given the observed population structure in the Indo-Pacific (see results), we restrained our analysis to six sampling sites only: Juan, Bampton, Belep, Orona, Fakarava, and Palmyra. To maximise the number of loci with no missing data, we sampled one individual per sampling site only. We used one *Carcharhinus limbatus* individual to polarize each SNPs and followed the same bioinformatics pipeline previously adopted to estimate the unfolded two dimensional SFS with *ANGSD* (Korneliussen, Albrechtsen, & Nielsen, 2014), from which we computed the pairwise ψ . The significance of the pairwise matrix of ψ was tested by bootstrap resampling the two-dimensional SFS 1000 times. We then used the *time difference of arrival location estimation* (TDOA) algorithm

implemented in the library *RangeExpansion* in R (Peter & Slatkin, 2013) to locate the origin of the RE.

Historical demographic inferences

To account and test for meta-population structure, we performed model selection as well as parameters estimation using an Approximate Bayesian Computation (ABC) framework (Bertorelle, Benazzo, & Mona, 2010). We tested three demographic scenarios (Figure 2) for each population, namely NS, FIM, and SST. *Model NS (no structure)*: going backward in time, NS represents a panmictic population where the effective population size switches instantaneously at T_c generations from N_{mod} to N_{anc} . *Model FIM (Finite Island Model)*: FIM represents a meta-population composed of a two-dimensional array of 10x10 demes, each of size N which exchanges Nm migrants with any other deme each generation. Going backward in time all demes merge into a single population of size N_{anc} at T_{col} generations. *Model SST (Stepping Stone)*: SST is similar to FIM but demes exchange migrants only with their closest neighbours. We performed 50,000 simulations under each scenario and for each population 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 50,000 additional simulations under the most supported scenario in order to estimate the demographic parameters. To this end, we used 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. We

also performed a cross validation for both parameter inference and model selection (hereafter, the confusion matrix) procedures. The number of trees in the forest was chosen by checking the out-of-bag error rate (OOB). We finally modelled the variation of effective population size (N_e) through time in each population with the *stairwayplot* (Liu & Fu, 2015) which uses a composite likelihood estimator based on the comparison between the observed SFS and that predicted under specific demographic histories. The *stairwayplot* assumes that the sampled lineages come from an isolated 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 bears useful elements on the evolutionary history of the meta-population (Lesturgie et al., 2021).

Population structure

Population structure inferences were performed on the dataset obtained following the assembly pipeline implemented in *Stacks* as described above. After the *denovo* step, the *population* script was called to keep loci present in at least 80% of the individuals ($r = 0.8$) 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 $\sim 30x$ (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* (*sparse non-negative matrix factorization*), 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* (*discriminant analysis of principal components*), 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 populations with more than 5 individuals using the *PopGenome* library in R (Pfeifer, Wittelsb rger, Ramos-Onsins, & Lercher, 2014) and tested its significance with a 100 bootstrap 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).

Results

Genetic diversity

The number of loci (monomorphic included) and SNPs with no missing data ranged from 34,766 to 146,858 and from 36,380 to 103,258 respectively in the sampling sites considered (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 1:1.02:1.02 for open sea, coral reef habitat and seamounts respectively. We plotted the correlation map computed on the whole array of demes with and without IND samples in Figure 3. Both analyses pointed to a most likely origin of the RE closer to the COR region but with a remarkable difference in the distribution of correlation coefficients. When including all sampling sites, all PAC locations displayed negative correlation coefficients whereas the IND only positive values (Figure 3). Conversely, when including only PAC sampling sites, we observed negative values from the Indian Ocean to the Coral Sea and positive values eastward (Figure 3).

The directionality index (ψ) was significant ($p \leq 0.001$) for each pair of comparison after bootstrap resampling, suggesting that the species most likely underwent a RE (Table S1). We applied the TDOA algorithm to the pairwise ψ matrix both with and without the IND samples. We inferred the most likely position of the range expansion in both cases very close to the Belep sampling site (Figure 3).

Historical demographic inferences

We investigated the demographic history for all sampling site with $N_{ind} \geq 7$. We first used an ABC-RF framework in order to compare demographic scenarios (Figure 2). Model selection was performed using 500 trees in all sampling sites after monitoring the OOB error rate. SST was the most supported scenario in all locations, with posterior probabilities ranging from 0.48 to 0.73 and similar classification error rate among locations (Table 2 and S2). Parameter estimation was performed with 1,000 trees in the RF algorithm after monitoring the OOB error rate. The median number of migrants per generation (Nm) ranged from ~ 6 to ~ 14 (Table 2). Posterior distributions

were overlapping and clearly distinct from the prior distribution (Figure 4a), and both the squared mean error (SME) and the mean root squared error (MRSE) were small among locations, suggesting reliable estimates (Table S3). The posterior distributions of T_{col} overlapped among locations (Figure 4b), even though it was not clearly unimodal (and therefore less informative) in Juan. Juan additionally displayed a lower N_{anc} median value (~21k) than in PAC locations (ranging from ~33k to ~50k) although all credible intervals overlapped (Table 2). Surprisingly, the ABC estimates for the Mckean sampling site were inconsistent with any other PHO sampling locations (Figure 4 and Table 2). However, both SME and the MRSE for the colonization time (T_{col}) and the ancestral effective size (N_{anc}) were generally one order of magnitude larger than those estimated for N_m in all populations (Table S2).

We further investigated the variation of N_e through time using the *stairwayplot* algorithm (Figure 5). We detected a broadly similar N_e 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 5). However, we found three main differences between IND and PAC sampling sites: i) the expansion time was around twice more 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) N_e during the constant period reached a value of ~40,000 in PAC sampling sites and of only ~20,000 in IND, consistent with the computed θ (Table 1). The PAC sampling sites showed a remarkable homogeneous *stairwayplot*, with only the more peripheral population of Fakarava having a slightly more recent ancestral expansion (Figure 5).

Population structure

We discarded 30 individuals based on an excess of missing data after an initial *denovo* assembly. We found an average coverage of 10.77x (s.d. = 2.32) in the total sample. After filtering, we retained 88,276 variable loci 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 S1 and S2), perfectly matching the two oceanic regions, namely the Indian and the Pacific Ocean (Figure 6a and S2). 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 6a). This highlights that IND samples are slightly more admixed than those from PAC. 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 S2). We further investigated $K=3$ under both algorithms and found three main results: i) the ancestral population or clusters clearly identify three geographic areas corresponding to IND, COR, and CPA regions (Figure 6a and S2); ii) the ancestry proportion of *cluster 3* follows a clinal distribution, steadily increasing in frequency from West (Indian Ocean) to East (French Polynesia) (Figure 6a); 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 S2). The PCA 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 6b). 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 variance. The second axis also suggested a clinal differentiation between the two clusters, with New Caledonia and Phoenix samples being more closely related (Figure 6b).

Population based analyses were performed on a reduced dataset excluding sampling sites with less than $N_{ind}=5$ individuals. We therefore retained 14 sampling sites, $N_{ind}=168$ individuals, and 88,824 variable loci and obtained an overall $F_{ST} = 0.25$ ($p\text{-value} < 0.01$). However, the pairwise F_{ST} highlighted a strong differentiation between Indian and Pacific sampling sites with values ranging from 0.53 to 0.56 (and always significant, $p\text{-value} < 0.01$, Table S4). In contrast, comparisons within oceanic regions never exceed 0.03 (Figure 7a) with only few values statistically significant. Consistently with clustering results, a heatmap displaying pairwise F_{ST} values visually suggest the existence of the three clusters previously identified (Figure 7a). However, the average differentiation between COR and CPA is only slightly higher than within group comparisons (Figure 7a). Moreover, we found a strong signature of isolation by distance within the Pacific Ocean, since the correlation between the F_{ST} and geographic or LC distance matrices was large and significant (Mantel test: $r = 0.93$; $p\text{-value} < 0.001$ in both cases) (Figure 7b). We additionally computed the correlation between genetic and geographic distances by considering only IND vs. PAC pairwise distances ($r = 0.77$, Figure S4).

Discussion

Range expansion

Range expansions occur by series of founder effects leading to the fixation of novel (derived) alleles and the decay in genetic diversity as colonization progresses (Excoffier, Foll, & Petit, 2009). They also leave specific signatures in the gene genealogy of lineages sampled from a deme of the meta-population (Maisano Delser et al., 2016; Ray, Currat, & Excoffier, 2003) and in the extent of population structure (Mona, 2017; Mona, Ray, Arenas, & Excoffier, 2014). Testing for the occurrence of a RE is therefore a fundamental step to decipher the evolutionary history of a species using genomic data. To this end, we first computed the directionality index ψ (Peter &

362 Slatkin, 2013) between *C. amblyrhynchos* sampling sites. All pairwise comparisons were
363 significantly different from 0 after bootstrap resampling ($p < 0.001$) (Table S1), suggesting that
364 the distribution of derived allele frequencies is not consistent with an equilibrium isolation by
365 distance model (Peter & Slatkin, 2013). We identified the most likely origin of the RE close to
366 New Caledonia using the TDOA algorithm (Figure 3). This result was robust to the inclusion of
367 the Juan sample from the Indian Ocean (Figure 3), despite the unbalanced sampling scheme. We
368 further investigated the RE by constructing a grid of 67894 points in the Indo-Pacific
369 corresponding to three different types of habitats (excluding land): coral reef, sea-mounts, open
370 sea. For each point of the grid, we computed the correlation coefficient between genetic diversity
371 and least cost distances to the point (representing the putative origin) for various combination of
372 permeability values of the three habitats. The presence of zones of more negative correlations is
373 suggestive of the occurrence of a RE and such zones are candidate for its origin (Ramachandran
374 et al., 2005). The inferred location of the origin of the RE covered a large area, which might be
375 attributable to the low differences in θ_π between Pacific populations (Table 1). Nevertheless,
376 including or not IND samples, we identified the most likely area west of the Indo-Australian
377 Archipelago (IAA), consistent with the finding based on the directionality index (Figure 3). The
378 scenario of a RE was corroborated by other evidences: first, the large and significant correlation
379 coefficient between genetic and geographic distances in the Pacific Ocean ($r=0.93$; Mantel p -value
380 < 0.001 , Figure 7b, see also Figure S4). This result alone would not be conclusive, since a similar
381 pattern is also expected under an equilibrium isolation by distance, but it strengthens our previous
382 findings. Second, the historical demography inferences performed in each sampled deme showed
383 that the pattern of genetic variability is most likely the outcome of a non-equilibrium meta-
384 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 4) and the expansion times retrieved by the *stairwayplot* (Figure 5) harbor the signature of the RE process (Lesturgie et al., 2021): the oldest times are expected to be close to the center 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 to detect any clear pattern through the Indo-Pacific, the expansion times highlighted by the *stairwayplot* are consistent with the RE scenario. All populations display a simultaneous expansion time around ~400 ky B.P. (Figure 5) except for Fakarava and Juan, which are the sampling sites respectively further east and west to the inferred origin of the RE. We note that the *stairwayplot* slightly overestimates the true colonization time in SST models (Lesturgie et al., 2021) but this bias does not affect the relative ranking among sampling sites. In summary, all the evidences presented thus far point to an origin of *C. amblyrhynchos* somewhere west to New Caledonia, from which two migration waves took place, one to the East Pacific and the other to the Indian Ocean, with Juan being probably one of the last spots to be colonized. This scenario, characterized by a center 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 in the IAA for *C. melanopterus*, a well known centre of origin for many teleost fishes (Cowman & Bellwood, 2013), and a current biodiversity hotspot (Allen, 2008). The difference observed between *C. amblyrhynchos* and *C. melanopterus* could simply results 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. Moreover, 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 many marine species in order 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 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 (Figure 4). *C. amblyrhynchos*, similarly to *C. melanopterus*, is strictly dependent on the reef habitat, which distribution is not homogenous through the Indo-Pacific (Figure S3). We would have expected the connectivity in each sampled deme to be highly correlated to the distribution of coral reef in its neighbourhood, as it was previously observed in *C. melanopterus* (Maisano Delser et al., 2019). However, the two sharks differ in their dispersal ability: while the grey reef shark is capable of long-distance movement of up to ~900 km (Barnett et al., 2012; Bonnín et al., 2019; Heupel, Simpfendorfer, & Fitzpatrick, 2010; Speed et al., 2016; White et al., 2017) the black tip reef shark shows a very restricted home range not exceeding ~50 km (Mourier & Planes, 2013). Our results reinforce the idea that the neighbouring size in the two species is very different, with *C. amblyrhynchos* being able to cross larger sea expanses becoming less sensitive to corals' concentration than *C. melanopterus*.

The homogeneity in the signature of genetic variation in each deme was confirmed by the *stairwayplot* analyses (Figure 5), contrasting with the heterogeneity previously described for *C. melanopterus* (Maisano Delser et al., 2019). All demes showed an ancestral expansion followed by a period of stasis and a strong bottleneck in recent times. We recently showed that these three

time periods are the typical signature of the variation in the coalescence rate through time due to the meta-population structure (Lesturgie et al., 2021). The differences observed are only related to the colonization time of the deme under investigation for the estimated range of Nm values (Lesturgie et al., 2021), confirming the similarity of dispersal pattern throughout the Indo-Pacific. In the same spirit, the signature of bottleneck observed in recent times for all demes (Figure 5) is the expected consequence of population structure (Chikhi et al., 2018; Lesturgie et al., 2021; Mazet et al., 2015; Rodríguez et al., 2018) and cannot be interpreted as a demographic decline. Unfortunately, population structure and demographic decline affect the SFS in a similar fashion making impossible to quantitatively disentangle the contribution of both to the observed bottleneck estimated using RAD-seq data (Lesturgie et al., 2021). Investigating recent changes in connectivity or demographic events 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).

Population structure

The results presented so far suggest that dispersal abilities of *C. amblyrhynchos* populations 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 populations. 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 evidences 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 7b) and a linear

454 relation of smaller intensity between IND and PAC samples (Figure S4). Remarkably, these values
455 are not affected by computing geographic distances between population under an LC approach.
456 Indeed, the permeability values maximizing the correlation are (almost) the same for the different
457 type of cells. This suggest that different geographic features do not affect the direction of
458 migrations of the grey reef sharks, which indirectly reinforces the absence of obstacles to shark
459 dispersal, consistently with the occasional long-distance swims detected across the open ocean
460 (Barnett et al., 2012; Bonnin et al., 2019; Heupel et al., 2010; Speed et al., 2016; White et al.,
461 2017). When strong IBD is present, it is difficult to attribute a biological meaning to groups
462 identified by clustering algorithms (Meirmans, 2012). Here the *sNMF* pointed out to K=1 as the
463 most likely number of clusters in PAC samples (Figure S5) and to K=2 when considering all
464 samples (Figure S1), with the latter result clearly driven by the distance between IND and PAC
465 populations. We further tested K=3 and noticed that the IND ancestral components diminish
466 continuously eastward, once again supporting an IBD structure (Figure 6a) rather than the presence
467 of barriers to gene flow. Similar considerations can be applied to the PCA, with the first axis,
468 explaining most of the variance, highlighting the difference between IND and PAC populations
469 and the second axis, explaining only ~1% of the total variance, rather identifying a cline in PAC
470 sampling sites (Figure 6b). This is consistent with the pairwise F_{ST} matrix, where intra Pacific
471 comparisons do not exceed ~0.03 and are mostly non-significant while the inter Ocean comparison
472 have an average F_{ST} of ~0.54 (Figure 7a). Defining management units seems therefore not
473 appropriate in the case of the grey reef shark, as the genetic variations appears to be rather
474 continuous. This contrasts with what previously suggested by Boissin et al. (2019) at the Pacific
475 scale: however, not only their results are based on few microsatellites only but they did not

consider IBD between the sampling points. This have likely artificially inflated the number of inferred clusters (Meirmans, 2012).

The pitfall of our study is to extrapolate the dynamic of the grey reef shark at the scale of its whole range by focusing mostly on the Pacific Ocean. Indeed, even if the species seems to follow an IBD pattern also from Chagos to Eastern Australia (Momigliano et al., 2017), the level of population differentiation appears to be higher than what we found in the Pacific for similar geographic distances. However, while the distribution of coral reef in the Pacific Ocean is scattered due to the presence of many archipelagos, coral reefs in the Indian Ocean are more concentrated on the coastal edge of the Asian and African continents (Figure S3). The effective distance between populations within the Indian Ocean would therefore be larger than in the Pacific Ocean, where 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 S4). An extensive sampling in the Indian Ocean will be mandatory to confirm this hypothesis and to shed more light on the dispersal ability of *C. amblyrhynchos*, in particular to understand the magnitude of what we previously defined as its neighbouring size.

Conclusions

Here 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 by Maisano Delser et al. (2019). The two species are among the most abundant reef sharks (MacNeil et al., 2020) and they share an almost overlapping distribution in the whole Indo-Pacific, being both strictly dependent on the availability of coral reef habitat. We showed that despite the

similarities in the RE dynamic, the pattern of genetic diversity and population structure are very different between the two species. First, *C. melanopterus* is significantly more structured than *C. amblyrhynchos* for 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 Figure 7). Second, *C. amblyrhynchos* shows a homogeneity in migration rates and demographic signals throughout its whole distribution, while *C. melanopterus* is more sensitive to the spatial distribution of coral reef, and the connectivity between populations is largely dependent on the short scale availability of this habitat (Maisano Delser et al., 2019). Indeed, migration rates estimated in areas with extensive coral reefs coverage (such as 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 in the light of the life history traits related to dispersal abilities of the two species, with *C. amblyrhynchos* being able to move more freely in open sea expanses rather than *C. melanopterus*. However, it will be important in the next future to precisely characterize the extent of the neighboring 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 the coral reefs distribution in

521 the Indo-Pacific. This work was supported by two ATM grants (2016 and 2017) from the Muséum
522 National d'Histoire Naturelle to S.M.

523

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. doi: 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. doi: 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. doi: 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. doi: 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. doi: 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. doi: 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. doi: 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. doi: 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. doi: 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. doi: 10.1093/icesjms/fsu025
- Chikhi, Lounès, 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. doi: 10.1038/s41437-017-0005-6
- Chikhi, Lounès, 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. doi: 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. doi: 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. doi: 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. doi: 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. doi: 10.1111/j.1095-8649.2012.03265.x
- 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. doi: 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. doi: 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. doi: 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. doi: 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. doi: 10.1111/2041-210X.12382
- Friedlander, A. M., & DeMartini, E. E. (2002). Contrasts in density, size, and biomass of reef fishes between the northwestern and the main Hawaiian islands: The effects of fishing down apex predators. *Marine Ecology Progress Series*, 230, 253–264. doi: 10.3354/meps230253
- Heled, J., & Drummond, A. J. (2008). Bayesian inference of population size history from multiple loci. *BMC Evolutionary Biology*, 8(1), 289. doi: 10.1186/1471-2148-8-289
- Heller, R., Chikhi, L., & Siegmund, H. R. (2013). The Confounding Effect of Population Structure on Bayesian Skyline Plot Inferences of Demographic History. *PLoS ONE*, 8(5). doi: 10.1371/journal.pone.0062992
- Heupel, M. R., Papastamatiou, Y. P., Espinoza, M., Green, M. E., & Simpfendorfer, C. A. (2019). Reef shark science - Key questions and future directions. *Frontiers in Marine Science*, 6(JAN), 1–14. doi: 10.3389/fmars.2019.00012
- Heupel, M. R., Simpfendorfer, C. A., & Fitzpatrick, R. (2010). Large-scale movement and reef fidelity of grey reef sharks. *PLoS ONE*, 5(3), 1–5. doi: 10.1371/journal.pone.0009650
- Hijmans, R. J. (2020). *Raster: Geographic Data Analysis and Modeling*.
- Jombart, T., Devillard, S., & Balloux, F. (2010). Discriminant analysis of principal components: A new method for the analysis of genetically structured populations. *BMC Genetics*, 11(1),

94. doi: 10.1186/1471-2156-11-94

- Khimoun, A., Doums, C., Molet, M., Kaufmann, B., Peronnet, R., Eyer, P. A., & Mona, S. (2020). Urbanization without isolation: The absence of genetic structure among cities and forests in the tiny acorn ant *Temnothorax nylanderi*. *Biology Letters*, 16(1). doi: 10.1098/rsbl.2019.0741
- Korneliussen, T. S., Albrechtsen, A., & Nielsen, R. (2014). ANGSD: Analysis of Next Generation Sequencing Data. *BMC Bioinformatics*, 15(1), 1–13. doi: 10.1186/s12859-014-0356-4
- Lesturgie, P., Planes, S., & Mona, S. (2021). Coalescence times, life history traits and conservation concerns: an example from four coastal shark species from the Indo-Pacific. *Molecular Ecology Resources*, 1755-0998.13487. doi: 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. doi: 10.1093/bioinformatics/btp324
- Li, H., & Durbin, R. (2011). Inference of human population history from individual whole-genome sequences. *Nature*, 475(7357), 493–496. doi: 10.1038/nature10231
- Liu, X., & Fu, Y.-X. (2015). Exploring population size changes using SNP frequency spectra. *Nature Genetics*, 47(5), 555–559. doi: 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. doi: 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. doi: 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. doi: 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. doi: 10.1038/hdy.2015.104
- Mazet, Olivier, 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. doi: 10.1016/j.tpb.2015.06.003
- Meirmans, P. G. (2012). The trouble with isolation by distance. *Molecular Ecology*, 21(12), 2839–2846. doi: 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. doi: 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. doi: 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. doi: 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. doi: 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. doi: 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. doi: 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. doi: 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. doi: 10.1038/s41437-017-0025-2
- Peter, B. M., & Slatkin, M. (2013). Detecting range expansions from genetic data. *Evolution*, 67(11), 3274–3289. doi: 10.1111/evo.12202
- 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. doi: 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. doi: 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. doi: 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. doi: 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 in human populations for a serial founder effect originating in Africa. *Proceedings of the National Academy of Sciences of the United States of America*, 102(44), 15942–15947. doi: 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. doi: 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. doi: 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. doi: 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. doi: 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. doi: 10.1038/s41437-018-0148-0
- Slatkin, M., & Excoffier, L. (2012). Serial Founder Effects During Range Expansion: A Spatial Analog of Genetic Drift. *Genetics*, 191(1), 171–181. doi: 10.1534/genetics.112.139022
- 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. doi: 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. doi: 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. doi: 10.18637/jss.v022.i05
- van Etten, J. (2017). R package gdistance: Distances and routes on geographical grids. *Journal of Statistical Software*, 76(1). doi: 10.18637/jss.v076.i13
- 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. doi: 10.1016/j.biocon.2017.01.009
- Whitney, N. M., Robbins, W. D., Schultz, J. K., Bowen, B. W., & Holland, K. N. (2012). Oceanic dispersal in a sedentary reef shark (*Triaenodon obesus*): Genetic evidence for extensive connectivity without a pelagic larval stage. *Journal of Biogeography*, 39(6), 1144–1156. doi: 10.1111/j.1365-2699.2011.02660.x

Data availability statement

Fastq sequence files are available from the GenBank at the National Center for Biotechnology

Information short-read archive database (accession number: forthcoming).

Authors contribution

S.M. and P.L. conceived the project. S.P. provided reagents and samples. S.M. and P.L. analysed the data and wrote the manuscript with input from S.P.

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}^{\dagger}	θ_w^{\dagger}	TD [‡]
IND	IND	Juan	13	95027	45635	1.18	1.48	0.32
		Zelee	6	146858	62674	1.30	1.31	0.26
COR [§]	CHE	Bampton	10	89958	82869	2.14	2.83	-0.22
		Avond	5	125710	87817	2.10	2.15	-0.12
	NCA	Belep	7	120038	103258	2.30	2.64	-0.11
		Poindimie	5	107464	72995	2.07	2.09	-0.05
CPA [§]	PHO	Niku	21	49922	53349	2.02	3.28	-0.25
		McKean	7	112711	88258	2.13	2.41	-0.01
		Orona	11	81725	75423	2.15	2.84	-0.09
		Kanton	10	99720	87202	2.12	2.69	-0.05
		Birnie [¶]	2	-	-	-	-	-
		Enderbury	13	76314	72221	2.09	2.91	-0.12
	PAL	Palmyra	40	34766	36380	1.64	3.22	-0.37
	POL	Moorea	5	104050	68380	2.03	2.02	0.02
		Fakarava	17	71715	66559	2.01	2.85	0.08
		Faaite [¶]	1	-	-	-	-	-
		Raraka [¶]	1	-	-	-	-	-
		Nengo [¶]	1	-	-	-	-	-

[†] Mean pairwise difference and Watterson theta 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).

	PP	Nm	T_{col}	N_{anc}
Juan	0.67	5.7 (1.77 - 17.72)	257800 (8086 - 658471)	21086 (399 - 52652)
Bampton	0.73	11.41 (3.97 - 19.03)	188782 (127761 - 577503)	45965 (27556 - 49856)
Belep	0.51	7.8 (2.84 - 20.82)	241218 (112840 - 843171)	49239 (7346 - 56316)
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)
Palmyra	0.73	13.39 (4.16 - 27.22)	142756 (62402 - 445380)	32542 (9502 - 37524)
Fakarava	0.72	10.2 (2.68 - 15.34)	256744 (110875 - 780150)	40502 (3091 - 49533)
Priors		${}^{\dagger}U[0.0001 ; 100]$	$U[100 ; 1500000]$	$U[100 ; 100000]$

† The prior distribution of Nm is the product of two uniforms (one for N and one for m)

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, $n = 40$); 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 (CAP, blue).

Figure 2. Demographic scenarios investigated in all populations with $N_{ind} \geq 7$ through an Approximate Bayesian Computation (ABC) framework. 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 all sampling sites (upper panel) or only Pacific Ocean sampling sites (lower panel). 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). The white area represents the most negative correlation values, i.e., the most likely origin of RE. The black cross represents the most likely origin of RE inferred using the directionality index. Black dots represent the sampling sites considered in the analysis.

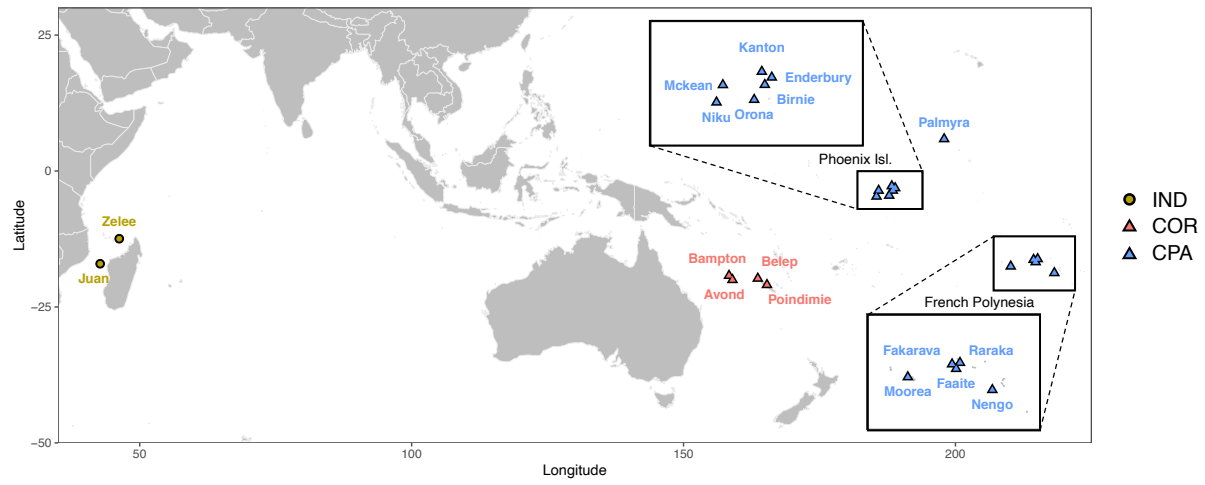
Figure 4. Posterior distribution of the number of migrants per generation Nm (a) and of the colonisation time of the array of deme T_{col} (b) estimated under the stepping stone model (SST) for all sampling sites with $N_{ind} \geq 7$. Colours represent the origin of the populations: Indian Ocean (yellow), Chesterfield islands (red), New Caledonia (green), Phoenix islands (blue), Palmyra (cyan) and Polynesia (purple). Line types represent the different populations from the Phoenix islands: Enderbury (solid), Kanton (dashes), Mckean (dots), Niku (dot-dashes) and Orona (long-dashes). The prior distribution is coloured in grey.

Figure 5. Variation of the effective population size (N_e) through time and its 75% confidence interval estimated by the *stairwayplot* for sampling sites of $N_{ind} \geq 7$ in IND (a), COR (b) and CPA (c) regions. Colours represent the origin of the populations: Indian Ocean (yellow), Chesterfield islands (red), New Caledonia (green), Phoenix islands (blue), Palmyra (cyan) and Polynesia (purple). Line types represent the different populations from the Phoenix islands: Enderbury (solid), Kanton (dashes), Mckean (dots), Niku (dot-dashes) and Orona (long-dashes).

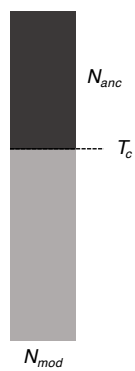
Figure 6. Individual-based population structure analyses. Ancestry proportions retrieved using the *sNMF* algorithm with $K=2$ and $K=3$ ancestral populations (a) and Principal Component Analysis (b).

Figure 7. Population-based population structure analyses computed with populations of $N_{ind} \geq 5$. Heat map representing the pairwise Reynold's F_{ST} values between sampling sites (a) and Isolation by distance (IBD) plot, with pairwise genetic distances ($F_{ST}/(1 - F_{ST})$) plotted against geographic distances between Pacific sampling sites only (b).

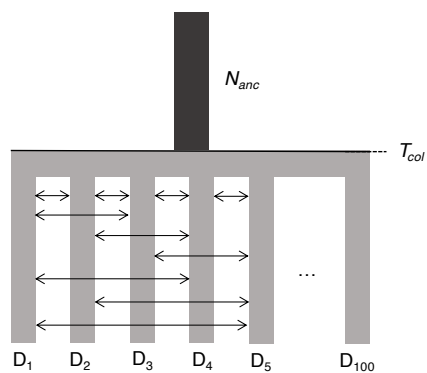
Figures



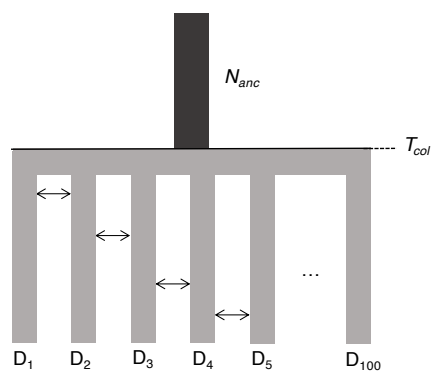
Non-Structured
Model (NS)



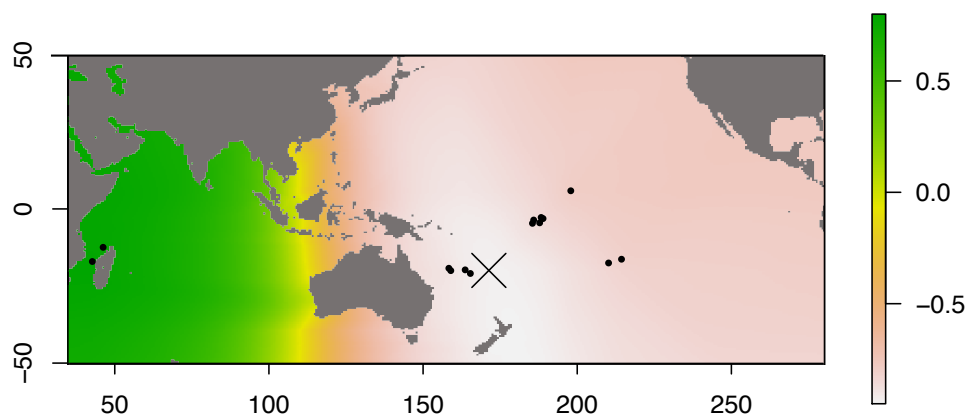
Finite Island
Model (FIM)



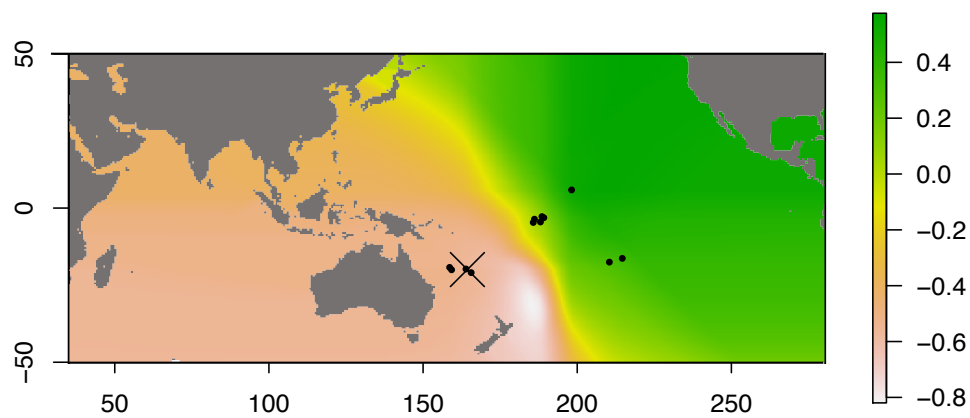
Stepping Stone
Model (SST)

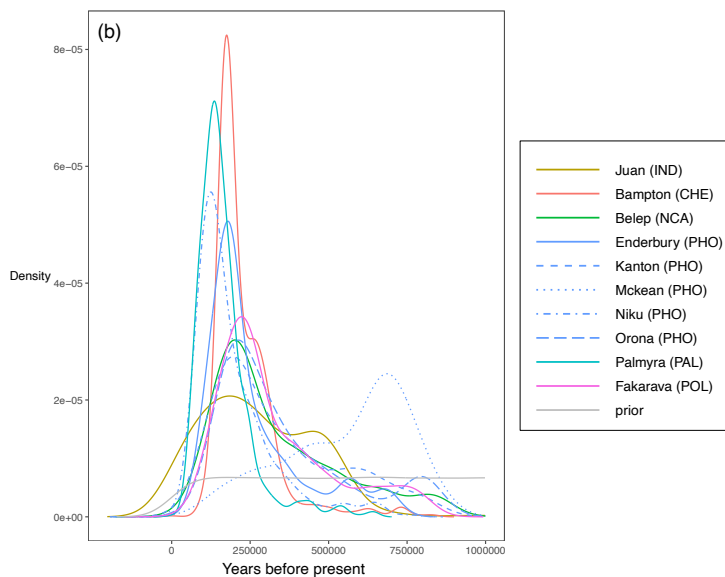
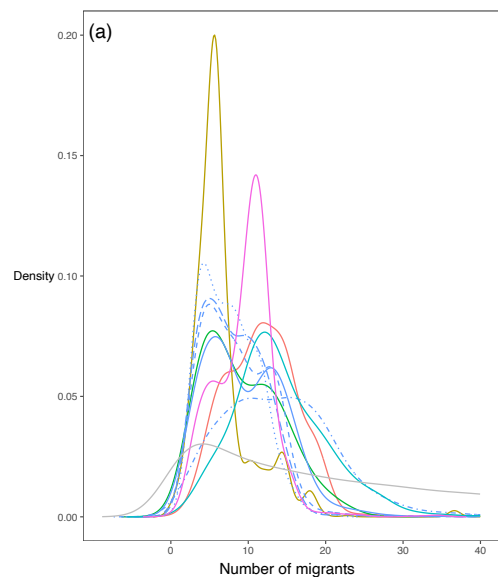


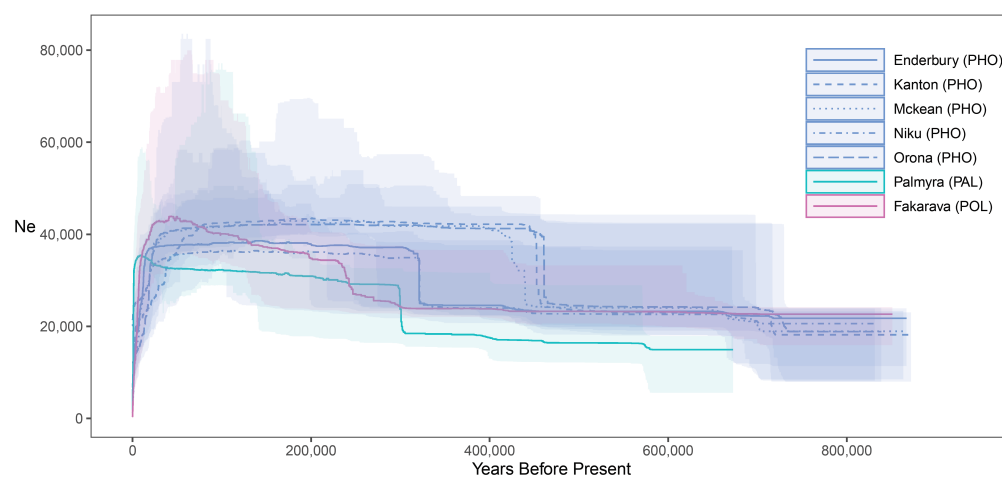
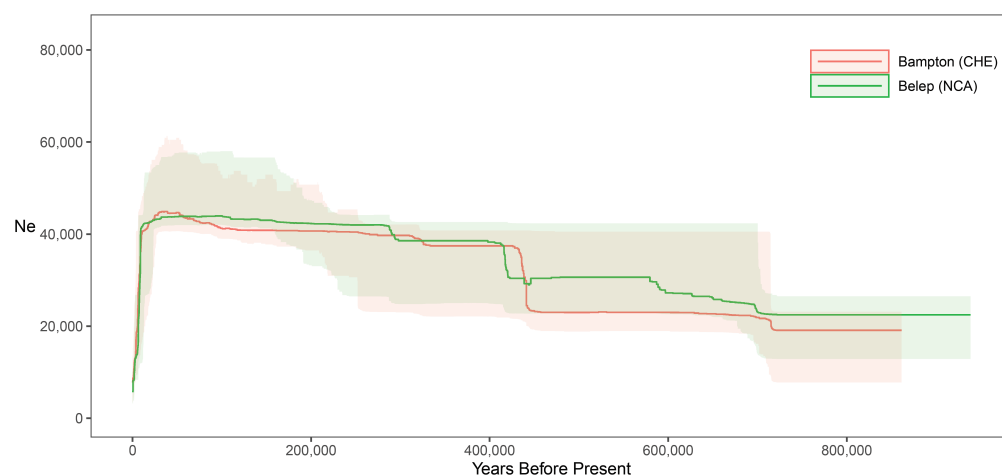
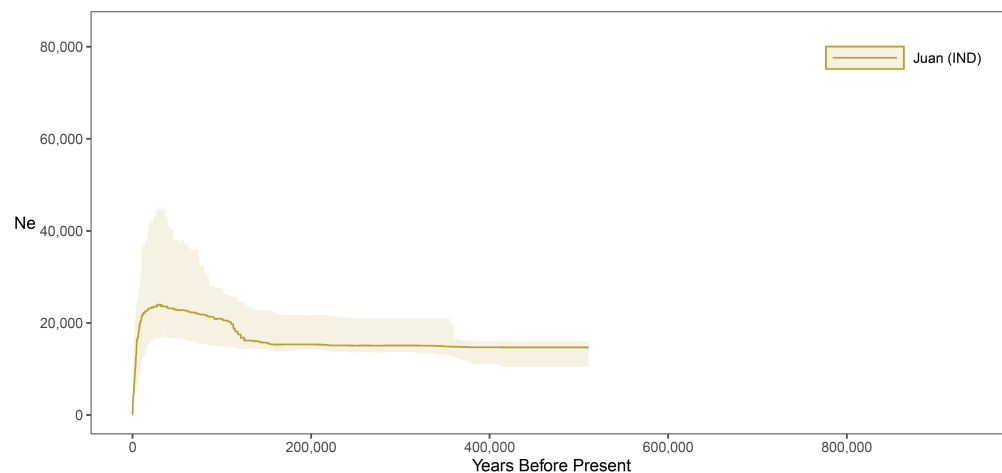
Indo-Pacific
samples



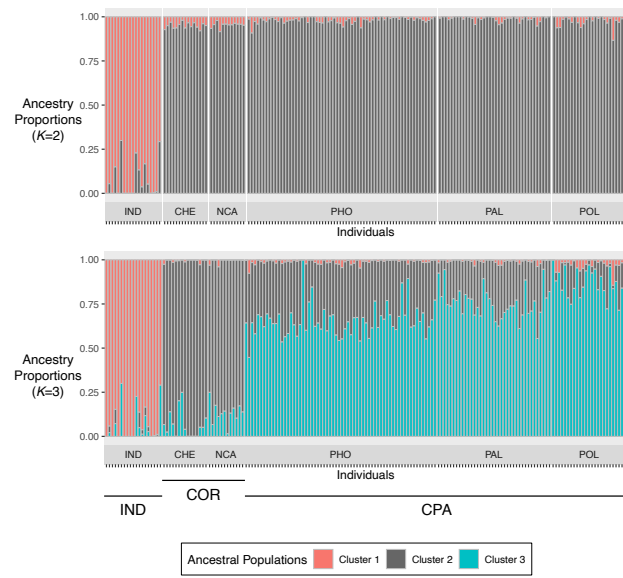
Pacific
samples



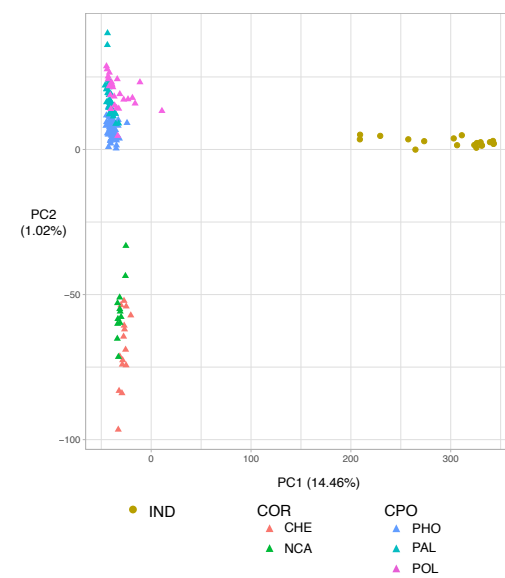




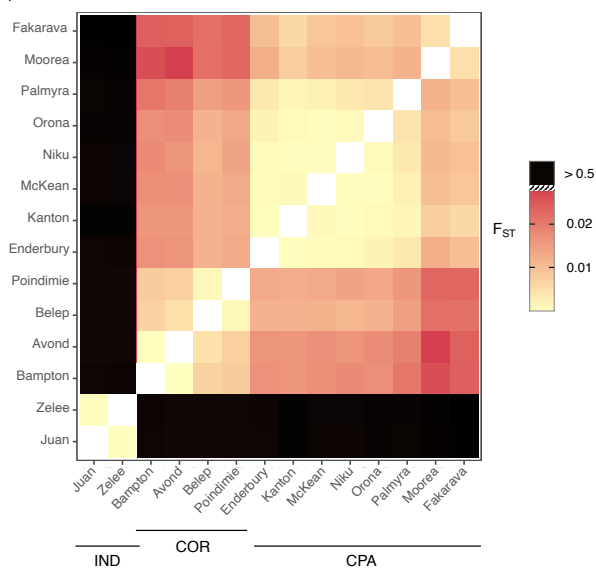
(a)



(b)



(a)



(b)

