

Life history traits and biogeographic features shaped the complex evolutionary history of an iconic apex predator (*Galeocerdo cuvier*).

Pierre Lesturgie

Institut de Systématique, Évolution, Biodiversité

Hugo Lainé

Institut de Systématique, Évolution, Biodiversité

Arnaud Suwalski

Institut de Systématique, Évolution, Biodiversité

Pascaline Chifflet-Belle

Institut de Systématique, Évolution, Biodiversité

Pierpaolo Maisano Delser

University of Cambridge

Eric Clua

Laboratoire d'Excellence CORAIL

Sébastien Jaquemet

University of Reunion Island

Hélène Magalon

University of Reunion Island

Stefano Mona (✉ stefano.mona@mnhn.fr)

Institut de Systématique, Évolution, Biodiversité

Research Article

Keywords: Agulhas leakage, coalescent modelling, demographic history, population genomics, Rad-seq, tiger shark.

Posted Date: May 20th, 2022

DOI: <https://doi.org/10.21203/rs.3.rs-1635778/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background. The tiger shark (*Galeocerdo cuvier*) is a large iconic marine predator inhabiting worldwide tropical and subtropical waters. So far, only mitochondrial markers and microsatellites studies have investigated its worldwide historical demography with inconclusive outcomes. Here, we assessed for the first time the genomic variability of tiger shark by Rad-sequencing 50 individuals from five sampling sites in the Indo-Pacific (IP) and one in the Atlantic Ocean (AO) to decipher the extent of the global connectivity and its demographic history.

Results. Clustering algorithm, F_{ST} and an approximate Bayesian computation framework revealed the presence of two clusters corresponding to the two oceanic basins. By modelling the two-dimensional site frequency spectrum, we tested alternative isolation/migration scenarios between these two populations. We found the highest support for a divergence time of $\sim 193,000$ years before present (B.P) and an ongoing but limited asymmetric migration ~ 176 times larger from the IP to the AO ($Nm \sim 3.9$) than *vice versa* ($Nm \sim 0.02$).

Conclusions. The two oceanic regions are isolated by a strong barrier to dispersal more permeable from the IP to the AO through the Agulhas leakage. We finally emphasized contrasting recent demographic histories for the two regions, with the IP characterized by a recent bottleneck around 2,000 years B.P. and the AO by an expansion starting 6,000 years B.P. The large differentiation between the two oceanic regions and the absence of population structure within highlight the existence of two large management units and call for future conservation programs at the oceanic rather than local scale, particularly in the Indo-Pacific where the population is declining.

Background

Predation plays a fundamental role in the top-down regulation of ecosystem dynamics, with apex predators being key actors in promoting species diversity (1). However, in marine ecosystems, many predatory species have declined across their range (2). As such, considerable effort should be devoted to develop conservation programs that correctly define units of managements for these species (3). Recent advances in sequencing technologies allow the characterization of thousands of independent loci giving the power to assess the genetic diversity of any target model or non-model species, which can inform management policies. However, genetic diversity should be complemented by the reconstruction of species connectivity and historical demography to better establish conservation priorities, since understanding how populations are spatially connected as well as the divergence time between lineages is essential to decipher at which geographic scale a species should be managed. Reconstructing the evolutionary history of a species is often a complex task that requires an educated choice of the most likely model of evolution, often selected among a reduced selection of biologically meaningful models. Unfortunately, selection of an inappropriate model can yield misleading estimates of critically important parameters as more data are collected. This has important implications for conservation genetic

applications that rely on accurate estimates of genetic diversity and changes in effective population size through time (4–7).

The tiger shark (*Galeocerdo cuvier*, Péron & Lesueur, 1822) is a large and iconic apex marine predator, that is considered “Near Threatened” by the International Union for Conservation of Nature (IUCN). Though the tiger shark is a predominantly coastal species, its distribution includes tropical and subtropical waters worldwide (8). The species is heavily impacted by fisheries (9) and shark control programs in the Indo-Pacific (10). Indirect estimates have suggested an annual number of tiger shark catches between 50,000 and 300,000 individuals (11), raising conservation concerns. Though not directly endangered by global climate changes, the species is likely to extend its habitat range poleward as a consequence of the increase in annual sea surface temperatures (12), which may increase the potential for greater trans-oceanic movements (13). Despite being found predominantly along the coastline, tiger sharks spend considerable time in pelagic waters and telemetry studies have shown that they can cross oceanic expanses (14–16), but no evidence of contemporary migration between the Indo-Pacific and the Atlantic Ocean has yet been found.

Knowledge of these ecological traits is important to devise meaningful evolutionary models, but it is not sufficient. Large marine predators with continuous distribution can be intuitively considered as capable of high dispersal due to the absence of clear physical barriers (17). Nevertheless, there are both examples of panmictic species, such as the blue shark *Prionace glauca* (18) or the mako shark *Isurus oxyrinchus* (19), and examples of species structured according to ocean basins such as *Carcharhinus galapagensis* and *Carcharhinus obscurus* (20). The degree of population structure, the extent of genetic diversity and the historical demography of the tiger shark remain largely mis-understood because of contrasting evidences mainly based on mitochondrial DNA (mtDNA) or nuclear microsatellite loci (21–27). Even though all studies recognize the existence of a clear separation between the Indo-Pacific (IP) and the Atlantic Ocean (AO), it remains controversial whether they represent two allopatric species as originally proposed by (21) or two divergent lineages as proposed by (22). In addition, an accurate characterization of divergence and migration is still lacking. Indeed, (22) provided a divergence time computed on mtDNA using a non-equilibrium model implemented in mdiv (28) but no confidence interval could be determined. At the same time, migration rates were estimated using the equilibrium model implemented in Migrate (29). Yet it remains unclear which model should be finally preferred to estimate which parameter. Population structure within the IP is also contentious, with (22) suggesting moderate population structure with a separation between Hawaii and the rest of IP, while (24) and (23) pointing to a panmictic population, which was recently supported by genomic data but few sampling locations (27). Furthermore, an in-depth analysis of the historical demography in both regions is still lacking, with only (24) supporting a recent decrease in effective population size in two sampling sites from the IP. Rad-seq provides a wealth of data that can help to resolve controversies originated by the analysis of mtDNA and of a small panel of microsatellites. To this end, we sequenced a total of 50 sharks from six sites in the IP and one in the AO (Fig. 1) to investigate the genetic diversity, the level of population structure, and the historical demography in all sampling sites. Finally, we tested alternative evolutionary scenarios to model the divergence and migration between IP and AO by fitting the observed two-dimensional site frequency

spectrum (2D-SFS) with extensive coalescent simulations. These analyses are necessary not only to reconstruct the evolutionary history of the tiger shark but also to better inform conservation strategies.

Results

Rad-seq sequencing

The average number of reads retained per individual after the quality filtering and demultiplexing step was 4,011,430 ($\pm 1,314,894$ s.d.). After a first round of *de novo* assembly and filtering using STACKS v.2, the depth of coverage was low with a mean of 12.65 (± 6.36 s.d.), which motivated the use of the genotype-free allele frequency estimation pipeline implemented in angsd (30) rather than the direct call. The final number of loci (variable and fixed) was highly variable between sampling locations (Table 1) ranging from 16,953 to 118,591 in Brazil (BRA) and New Caledonia (NCA) sampling sites respectively, with a number of SNPs with no missing data following a similar trend (from 5,868 to 26,075 for BRA and NCA, respectively).

Table 1

Sample size (n), mean pairwise difference (θ_{π}), Watterson theta (θ_w), Tajima's D (TD), and total number of loci (monomorphic included) (n_{loci}) and SNPs (n_{SNP}) for all sampling sites (ranged from west to east). AUS_E: East Coast of Australia; AUS_N: North Coast of Australia; BRA: Brazil; COR: Coral Sea; NCA: New Caledonia; RUN: Reunion Island.

	n	$\theta_{\pi} (10^{-3})$	$\theta_w (10^{-3})$	TD ^a	n_{loci}	n_{SNP}
BRA	7	0.97	1.16	0.14	16,953	5,868
RUN	15	0.59	0.73	-0.19	71,214	19,971
AUS _N	8	0.76	0.97	-0.19	38,420	11,627
COR	5	0.58	0.72	0.04	97,736	18,407
AUS _E	7	0.63	0.77	0.02	49,380	11,385
NCA	8	0.57	0.7	-0.03	118,591	26,075

^a Tajima's *D* values in bold are significantly different from 0 ($P < 0.001$).

Population structure

After filtering, the remaining number of SNP was 24,454 for the Principal Component Analysis (PCA) and the *non-negative matrix factorization* (*nmf*) inference, and ranged from 8,785 to 15,977 per population pair for the F_{ST} computation. Pairwise F_{ST} highlighted a moderate differentiation between Indo-Pacific (IP) and Atlantic Ocean (AO) sampling sites, with values ranging from 0.117 to 0.129 and systematically significant ($P \leq 0.001$, Fig. 2 and Table S1). Conversely, the average F_{ST} between IP sites was 0.023

(ranging from 0.018 to 0.029) and not significant for the majority of pairwise comparisons (Table S1). The Mantel test, computed between IP sampling sites only (given the evidence of a clear genetic discontinuity between AO and IP), showed no correlation between genetic and geographic distances ($r = 0.005$, $P = 0.62$, Figure S1). Clustering analyses were consistent with the observed pattern of differentiation. First, the PCA clearly segregated AO from IP individuals, with 38.71% of the total variance explained by the first axis (Fig. 3). Second, the *nmf* algorithm selected $K = 2$ ancestral populations corresponding to IP and AO (Fig. 2). Individuals from IP had a probability ancestry to cluster 1 ranging from 92.6–100% whereas individuals from AO showed a probability ancestry to cluster 2 ranging from 84.6 to 100%. Average ancestry of cluster 2 in IP individuals was only 0.7% while average ancestry of cluster 1 in AO individuals was 4.4%. Sharks from Reunion Island (RUN), the IP sampling site closest to AO, did not show higher cluster 2 contribution than other IP individuals nor lower F_{ST} or more proximity in the PCA plot to AO (Fig. 2 and Fig. 3). When computed on IP individuals only, the PCA identified a single cluster (Fig. 3) and the *nmf* did not show any meaningful geographic clusters with $K = 2$ (Figure S2). We further applied an Approximate Bayesian Computation (ABC) framework using a 500 trees random forest for all sampling sites after checking for the evolution of the out-of-bag error rate. The model selection between the *Stepping Stone* (SS), *Finite Island* (FIM) and *Non-Structured* (NS) models (Figure S3 and Table S2) highlighted NS as the most supported model with a posterior probability ranging from 0.48 to 0.89 in the IP sampling sites and of 0.62 for BRA (Table S3).

Genetic diversity and variation of N_e

Genetic diversity values were very similar among sampling sites, with BRA being slightly more variable than the IP counterpart (Table 1). Tajima's D (TD) was significantly positive for BRA ($TD = 0.137$; $P \leq 0.001$), while significantly negative for Northern Australia (AUS_N) and RUN ($P \leq 0.001$) and not significantly different from zero for the other populations (Table 1). Except for the AUS_N population, the reconstructions of the effective size (N_e) through time by the stairwayplot were very similar among IP locations: an ancestral expansion occurred between $\sim 100,000$ and $\sim 200,000$ years before present (BP), bringing the median N_e to $\sim 10,000$ followed by a very recent bottleneck $\sim 2,000$ to $\sim 4,000$ years BP (Fig. 4). The stairwayplot for AUS_N displayed a different signal, with an ancestral N_e median value similar to the one retrieved in the other sampling sites ($\sim 10,000$) followed by a strong and recent expansion that raised the modern N_e to $\sim 35,000$, contrasting with the recent decrease observed for the other IP sampling sites. The demographic history reconstructed for BRA was slightly more complex with the ancestral N_e of $\sim 12,000$ first decreasing to $\sim 9,000$ at $\sim 40,000$ years BP and then increasing (between $\sim 4,000$ and $6,000$ years BP) to a modern N_e of $\sim 20,000$ (Fig. 4).

Population divergence and migration rate estimation

In the light of the absence of population structure within the IP and the AO and the genetic distinctness between them, we tested five Isolation-Migration (IM) models to determine the divergence time and the

migration pattern between the two oceanic regions (Fig. 5). The likelihood distribution computed over 100 replicates was similar for models IM-bsc and IM-full, but the AIC values supported IM-bsc as the model with the highest probability since it was designed with less parameters (Figure S4). The distribution of the likelihood evaluated in each model under the Maximum Likelihood (ML) parameters proved that they can be distinguished based on the available data (Figure S4). The two oceanic regions appeared connected, though the migration rate is limited and strongly asymmetric, being ~ 176 times higher from IP to AO ($Nm \sim 3.9$) than *vice versa* ($Nm \sim 0.02$) (Table 2). Going backward in time the populations from the two regions merged $\sim 193,000$ years BP (90% CI: [77,000; 355,000]). The ancestral population size was almost half of those estimated in both IP and AO derived populations (Table 2), indicating an ancestral expansion, consistent with the observed stairwayplot dynamics for IP populations. An increase in effective size was estimated in the AO starting $\sim 45,000$ years BP (90% CI: [14,000; 53,000]), bringing the effective size from 1,406 (90% CI: [1,178; 4,781]) to 16,810 (90% CI: [12,885; 40,036]) which is consistent with the observed expansion in the stairwayplot of BRA. We observed a decrease in N_e in the IP from 49,000 to 17,000, but the timing was poorly estimated. Moreover, ancient and modern N_e in the IP showed largely overlapping confidence intervals (Table 2).

Table 2

Maximum Likelihood (ML), 90% confidence interval (5% lower bound and 95% upper bound) and search bounds for the parameters estimated by fastsimcoal under the IM-bsc model.

Parameter	ML	5% lower bound	95% upper bound	Parameter bounds
N_{anc}	9,274	1,310	80,372	U [¶] :{100;100,000}
$N_{mod_{IP}}$	17,591	16,643	31,915	U:{100;100,000}
$N_{mod_{AO}}$	16,810	12885	40,036	U:{100;100,000}
$T_{S_{IP}}^*$	153,090	27,460	769,750	U:{10;100,000}
$T_{S_{AO}}^*$	45,980	14,030	53,090	U:{10;100,000}
$N_{anc_{IP}}$	48,823	3,103	107,816	U:{100;100,000}
$N_{anc_{AO}}$	1,406	1178	4,781	U:{100;100,000}
T_{div}^*	193,850	77,110	355,910	U:{10;100,000}
$Nm_{AO IP}^{\S}$	0.022	0.008	0.124	U:{0;50}
$Nm_{IP AO}^{\S}$	3.873	3.171	8.919	U:{0;50}
*Times are expressed in years using a mutation rate of 1.93×10^{-8} per generation per site and a generation time of 10 years				
[§] Number of migrants per generation are expressed in forward.				
[¶] U: uniform distribution.				

Discussion

To shed light on the evolutionary history of the tiger shark, we sequenced thousands of loci in 50 individuals following a double digest Rad-seq protocol. We handled low coverage issues by applying an appropriate framework based on genotype free likelihood estimation of allele frequencies (30). The first and most impressive result is the presence of two highly divergent genetic clusters, corresponding to the Indo-Pacific (IP) region and the Atlantic Ocean (AO), and the signature of very weak population structure within each of them. Despite the large panel of SNPs that could potentially detect fine spatial structure compared to previous work based on microsatellites, we did not find any barrier to gene flow within the IP, but rather a signature consistent with a large panmictic population or a meta-population characterized by very large amount of gene flow (Fig. 2, Fig. 3, Figure S1 and Figure S2). This strongly confirms previous findings (23, 24, 27) and dismiss the conclusions of (22), who found a significant evidence for population structure. Using an unprecedented amount of genomic data, we provide conclusive evidences for the

presence of a single mating population in the IP since (1) the PCA and *nmf* analyses unambiguously detected one single cluster in the IP; (2) the F_{ST} values computed between sampling sites did not exceed 0.029 (as a comparison, an average value of ~ 0.124 was found between IP and AO) with no signature of isolation by distance (Figure S1); (3) all IP sharks showed a similar amount of genetic contribution from the AO cluster in the *nmf* analysis (Fig. 2). These results can only be explained considering that either tiger sharks randomly mate within IP or that the number of migrants exchanged each generation between sampling sites is so large to erase any signature of genetic structure. The absence of multiple sampling sites in AO prevented us to perform similar analyses in AO. To test the presence of a single panmictic population, we therefore followed an ABC strategy based on coalescence simulations to reconstruct the evolutionary history of BRA population and assess whether it is better described by unstructured or meta-population models. This approach has been successfully applied in both empirical and simulation-based works (7, 31, 32) and represents an alternative (or complementary, when it is possible) method to infer the presence of population structure. Similarly to IP, the NS (No Structure) model was by far the most supported one within the AO (Table S3). Even though more sampling sites and SNPs would help providing tighter estimates, this result also suggest the presence of a single mating population in the AO. Population structure have been detected before within the AO with few mitochondrial markers (25, 26). However, our results are consistent with the genome wide study of (27) who suggested low to no population structure in the AO. We stress that given the accuracy provided by the size of genomic datasets (thousands of SNPs) and the robustness of our demographic inferences, results based on few mitochondrial markers have to be carefully interpreted. Furthermore, it is not surprising to find a similar scenario (i.e., low to no population structure) in the two oceanic regions, as it is consistent with the fact that the tiger shark is highly migratory. However, it does not explain why a large predator that is able to cross ocean basins covering distances of several thousands of kilometers (13–15) could not erase the genetic differentiation between the two regions.

One possible explanation was proposed by (21), who suggested the presence of two allopatric subspecies in IP and AO. This hypothesis was later refuted by (22), who still agreed on a long term genetic isolation between the two oceanic regions but proposed some genetic exchanges. By harnessing the power of Rad-seq genome wide data, our coalescent modelling could not only unequivocally disentangle the two hypotheses, but also quantitatively determine the tempo and mode of divergence. Comparing five Isolation/Migration (IM) scenarios, we found that the most supported model included a divergence around $\sim 193,000$ years BP (90% CI: [77,000; 356,000]) between the two regions (Table 2), which nevertheless remained in contact since then through a very limited (3.9 individuals per generation; 90% CI: [3.2; 8.9]) and asymmetric gene flow ~ 176 times higher from IP to AO than the opposite direction. The low number of migrants Nm exchanged each generation (Table 2) and the asymmetric exchange are consistent with the clustering results and the F_{ST} values (Fig. 2), which unambiguously differentiated the two regions and clearly highlighted a higher, but weak, genetic contribution from IP to AO than *vice versa* (Fig. 2). These results support the idea that populations from the AO and from the IP represent two lineages (22), rather than two allopatric species (21). A permeable barrier to gene flow between the two oceanic regions is therefore responsible for the observed pattern of divergence and asymmetric

migration. The presence of this barrier can be explained by the ecology of the tiger shark and by the environmental conditions governing the Indian-Atlantic water exchange, the so-called Agulhas leakage. As a tropical to sub-tropical species, tiger sharks prefer warm water and they show the peak of swimming activities at $\sim 22^{\circ}\text{C}$ (12) so that their movement from the Atlantic to the Indo-Pacific is hampered by the upwelling of cold water off South-Western Africa (the Benguela current). However, the AO receives warm water from the IP through the Agulhas leakage (33), which can account for the asymmetric migration reported, consistent with the pattern observed in other tropical sharks, bony fishes and turtles (34–37). The Agulhas leakage has not been constant through time, with an increasing intensity in the Holocene, preceded by a period of stasis and a strong peak in the late Pleistocene around 130,000 years BP (38, 39). This variation in Agulhas leakage intensity could have influenced the relation between the two basins. However, we could not distinguish pulses of migrations in our data since the model IM-bsc was preferred to those accounting for variation in migration rate through time. Model selection procedure robustly selected the IM-bsc model (Figure S4), which is neither the most nor the least parameters rich, supporting the idea that our results are not an artefact of wrong modelling (see also below). We are aware that in the next future there will still be space to improve our estimates: more data will help refining the confidence interval of each parameter and ameliorating the calibration of the molecular clock.

We found divergent demographic histories in the two oceanic regions (Fig. 4). First, we note that since both regions are most likely described by non-structured models and the migration rates between them are very low, it is possible to directly interpret the results of unstructured model such as the stairwayplot. It is important to stress this point, since otherwise the presence of population structure could determine spurious signature of population size changes that would be biologically misinterpreted (32, 40, 41). All the Indo-Pacific populations (except AUS_N) underwent a recent bottleneck between $\sim 2,000$ and 4,000 years BP. This is barely consistent with what was recently proposed by (24) based on 25 microsatellites combined with mitochondrial DNA. Here we refined their estimates and better characterized the intensity of the bottleneck since we applied a non-parametric approach (the stairwayplot) exploring a large parameter space to genome wide data. The demographic history in the AO contrasted with the IP. Indeed, we estimated a very high N_e ($\sim 20,000$) following a recent expansion that happened between $\sim 4,000$ and 6,000 years BP (Fig. 4), also consistent with the estimates retrieved by the IM-bsc model (Table 2). The strong signature of population expansion recovered implies that the tiger shark is profiting from recent environmental changes in AO, in contrast to the IP population. Consistently, (26, 42) found a recent demographic trend suggesting an expansion rather than a decrease in AO, while the recent demographic trends in the IP are rather conditioned by intense pressure from fisheries and shark-control programs (9–11). More investigations are needed to determine the origin of the difference between the two oceans. The applications of methods based on linkage disequilibrium applied to whole genome data will help detect more recent events (43), which will be important for planning conservation strategies. Given the very low genetic structure within IP, we would expect AUS_N to have the same demographic history than the other sampling sites in the IP. We cannot exclude that AUS_N represents an isolated population experiencing its own demographic history that separated from the rest of the IP too recently to

accumulate divergence. More data will be needed to shed light on this result, both by sampling more individuals and more loci at higher coverage.

Conclusions

Reconstructing the evolutionary history of a species relies on the application of a realistic demographic model, which is mostly unknown in empirical studies. Here, we carefully investigated population genetic structure of the tiger shark and found that its evolutionary history is characterized by an asymmetrical migration between the AO and the IP and a signature of random mating within each region. These findings let us modelling each oceanic region as a single unstructured population and advised us on the demographic scenarios to implement in order to investigate their divergence time and migration rates. The two regions are separated by the Benguela barrier, but our estimates strongly suggest that the Agulhas leakage allows an asymmetric migration between them, by far stronger from IP to AO than in the opposite direction. While we confirmed that the tiger shark is likely undergoing a reduction in N_e in the Indo-Pacific, we show that it probably underwent a strong expansion in the Atlantic Ocean. Even if a better calibration of the molecular clock and full genome analyses would still be needed to confirm our results, our findings support the existence of two management units. This implies that local conservation or shark control programs will have very limited impact on the dynamics of the species, which needs to be managed at the ocean basin level, demanding considerable communication efforts among different countries and coordination as suggested for other megafaunal organisms (45).

Material And Methods

Sampling, library preparation and sequencing

A total of 50 tiger shark individuals (*Galeocerdo cuvier*) from both the Indo-Pacific (IP) and the Atlantic Ocean (AO) were sampled off (from west to east) Brazil (BRA), Reunion Island (RUN), North Coast of Australia (AUS_N; North Territory), East Cost of Australia (AUS_E; Sunshine Coast), Coral Sea (COR) and New Caledonia (NCA). Sharks were grouped into six populations based on their sampling site (Fig. 1; Table 1). Total genomic DNA was extracted from muscle tissue or fin clips preserved in 96% ethanol using QIAGEN DNeasy Blood and Tissue kit (Qiagen, Hilden, Germany) according to the manufacturer's protocols. Double-digest restriction-associated DNA (ddRAD) libraries were prepared following (46) using EcoRI and MspI restriction enzymes, a 400-bp size selection, and a combination of two indexes and 24 barcodes to pool 48 individuals per lane. The genomic libraries obtained were sequenced with a HiSeq 2500 Illumina sequencer (single-end, 125 bp).

Rad-seq *de novo* assembly

Raw reads were first demultiplexed and quality filtered through the *process_radtags.pl* pipeline in stacks v.2.5 (47). In the absence of a reference genome of *G. cuvier* or of closely related species, RAD-seq loci (125 bp sequences) were *de novo* assembled under the *denovo_map.pl* pipeline in stacks. Preliminary

results based on 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) found an average depth of $\sim 10x$ (see Results). Despite the absence of a clear cut-off indicating an acceptable coverage value above which genotype calling may be considered reliable, simulation results suggest that $\sim 10x$ may produce inconsistent calling under different algorithm (48). To prevent this, we used on a genotype free estimation of allele frequencies implemented in the software *angsd* v.0.923 (Analysis of Next Generation Sequencing Data; (30)), which has been proven more efficient for low to medium coverage next-generation sequencing (NGS) data than SNPs calling algorithms (30). We describe below the bioinformatics steps required to apply *angsd* to Rad-seq data from a non-model organism and the downstream population genetic analyses applied to the filtered datasets.

Assembly pipeline and filtering

angsd requires a reference sequence to work, which we were lacking. To circumvent this issue, we followed the approach described in (49) by creating an artificial reference sequence from loci previously assembled by stacks under the parameter $m = 3$, $N = 3$, $M = 3$ (based on the results of Mona, Bertorelle, Benazzo, in preparation). To this end, we concatenated the consensus sequences of each locus spaced by a stretch of Ns and then map reads back from individual *fastq* files using the *bwa-mem* algorithm with default parameters (50). Using custom bash scripts coupled with *angsd*, we then discarded: (i) sites with coverage $< 3x$ (*-minIndDepth* = 3, corresponding to m in the first assembly performed by stacks) and/or of low quality (based on the per base alignment score, *-baq* = 1 flag); (ii) low quality bases and poorly aligned reads (*-minQ* and *-minMapQ* and *-C* flags with default values); (iii) SNPs present in the last 5 bp of each locus and SNPs genotyped as heterozygous in 80% or more of the individuals; (iv) loci with more than 5 SNPs that might be the result of paralog RAD loci alignment on the reference. Specific filters were further added according to the downstream analyses performed.

Population structure

A single reference sequence was created for all populations and we retained sites shared by at least 80% of the samples. The PCA was computed with PCAnsd v.0.97 based on genotype likelihood (51). Admixture was then investigated by running the *non-negative matrix factorization* algorithm (*nmf*) implemented in PCAnsd which is based on the same covariance matrix inferred for the PCA. The number of ancestral populations (K) was automatically chosen by PCAnsd to be $e + 1$, where e is the optimal number of significant principal components depicting population structure, resulting from the Velicier's minimum average partial test run on the covariance matrix. The sparseness regularization parameter α (used to reduce the noise in low depth NGS data) that best fitted the data was tested between 0 and 100 and it was chosen by comparing the resulting likelihood following (51). We generated pairwise *site allele frequency likelihood* files and then computed F_{ST} with the *realSFS* program in *angsd* (52) using SNPs with a minor allele frequency ≥ 0.05 (*-minMaf* flag). The significance of each pairwise F_{ST} comparison was evaluated with 1,000 permutations by randomly allocating individuals to one of the

two populations. We finally tested isolation by distance (IBD) using a Mantel test (53) and plotted the relationship between genetic vs. geographic distances.

We applied an *approximate Bayesian computation* (ABC) approach similar to previous studies (7, 31, 32) in all sampling sites to further investigate the presence of population structure. This approach is particularly helpful in the Atlantic Ocean (AO) where only one locality was sampled (Brazil; BRA population). Briefly, we designed three demographic models (Figure S3): (1) NS (No Structure) which represents a panmictic population where N_{mod} the modern effective size instantly changes to N_{anc} the ancestral effective size, at T_s (time shift) generations; (2) FIM (Finite Island Meta-population) which represents a finite island meta-population model composed of 100 demes exchanging symmetrically Nm migrants per generation with each other. All demes were instantaneously colonised, T_{col} generations ago, from an ancestral population of size N_{anc} . (3) SS (Stepping-Stone) which represents a stepping-stone model where the 100 demes are arranged in a two-dimensional grid and where migration is only allowed symmetrically in both directions between the four nearest neighbouring demes. We performed 50,000 coalescent simulations under each model using fastsimcoal v.2.6.0.3 (54) extracting parameters from the priors distributions displayed in Table S1. Model selection was evaluated by the random forest classification method implemented in the *abcRF* package in R (55). We used the SFS, θ_π and TD as summary statistics and further added the first two axes of the Linear Discriminant Analysis in the dataset as suggested by (55) to increase the classification method accuracy. The number of trees was chosen by checking the evolution of the out-of-bag error.

Genetic diversity and effective population size variation

We created one reference sequence per population in order to maximise the number of loci assembled. We filtered the sites with missing data by setting the *-minInd* flag in *angsd* to the total number of individuals in each population. The filtered dataset was then used to generate a *site allele frequency likelihood* (*saf*) file, where genotype likelihoods were computed using the SAMtools method (*-GL = 1* flag). The folded site frequency spectrum (SFS) was directly computed from the filtered *saf* datasets through the *realSFS* program (52). Nucleotide diversity (θ_π), Watterson's theta based on segregating sites (θ_w ; (56)) and Tajima's D (TD ; (57)) were computed with custom script from the SFS. Significance of TD was evaluated after 1,000 coalescent simulations of a constant population model with size θ_w . We reconstructed the variation in the effective population size (N_e) through time by running the *stairwayplot* v.0.2 software (58), where the composite likelihood is evaluated as the difference between the observed (folded) SFS and its expectation under a specific demographic history.

Population divergence and migration rate estimation

Based on the results of the previous analyses, we devised five alternative Isolation/Migration (IM) models of divergence between IP and AO regions using the composite likelihood method implemented in *fastsimcoal* (54). We presented in Fig. 3 the model richest in parameters, the remaining four representing simplified versions nested within it. Hereafter, a brief description of the five models going from the most

complex to the simplest: (a) **IM-full**: the two ocean regions with their respective modern effective population sizes, $N_{mod_{IP}}$ and $N_{mod_{AO}}$ diverged at T_{div} from an ancestral population of effective population size N_{anc} . Due to the stairwayplot results, we allowed the two modern effective population sizes $N_{mod_{IP}}$ and $N_{mod_{AO}}$ to change to $N_{anc_{IP}}$ and $N_{anc_{AO}}$ following an exponential dynamic in $T_{s_{IP}}$ and $T_{s_{AO}}$ years respectively. Migration is defined by two time periods: m_1 representing the migration rate occurring between time 0 until T_{mig} and m_2 between time T_{mig} until T_{div} . The migration matrix in each time period is asymmetric: for instance, $m_{1_{AO|IP}}$ represents the forward migration rate from AO to IP and $m_{1_{IP|AO}}$ from IP to AO. In summary, the model is defined by thirteen parameters: five effective population sizes, four migration rates and four historical events; (b) **IM-anc**: same as IM-full with ancestral migration only between T_{mig} and T_{div} . The model is defined by eleven parameters, the two m_1 migration rates being removed; (c) **IM-rec**: same as IM-anc, but with recent migration only occurring between time 0 until T_{mig} , keeping only the two m_1 migration rates; (d) **IM-bsc**: the classic model where migration is constant from time 0 until T_{div} (i.e. $m_1 = m_2 = m$ (28)). We modelled the variation in effective size of the two regions similarly to the other models, for a total of ten parameters; (e) **IM-div**: a pure divergence model with no migration. This is defined by eight parameters: the five effective population sizes and three historical events (T_{div} , $T_{s_{IP}}$, $T_{s_{AO}}$). The analyses are based on the folded 2D-SFS computed by *angsd* between six individuals from Brazil (representing the AO) and six from the Indo-Pacific (IP). This sample size was chosen to obtain a balanced design and to maximise the number of SNPs shared among the two ocean basins. Similarly, in each basin, we selected the individuals presenting the smaller proportion of missing data to further increase the number of joint SNPs. To maximize the observed 2D-SFS we applied the following options in *fastsimcoal*: -N 300,000 (number of coalescent simulations), -L 40 (number of expectation-maximization (EM) cycles), and -C 10 (minimum observed SFS entry count considered for parameter estimation). For all model parameters we used wide search ranges with uniform distributions (Table 2). We ran each model 100 times in order to determine the maximum likelihood parameters and to perform model selection using the Akaike's information criterion comparing the best run of each model (54). To check the robustness of the model selection procedure and to take into account the presence of linked sites in our dataset, we further examined the likelihood distribution obtained based on 100 expected 2D-SFS simulated under the parameters estimated in the best run of each model, each approximated with 10^6 coalescent simulations. This procedure is needed to take into account the variance in the likelihood estimation given our dataset: if the distributions obtained by the various models do overlap, the difference in the estimated likelihoods of our models is not significant (59). Finally, we determined the confidence interval of the parameter estimated under the best run of our best model by parametric bootstrapping. The 2D-SFS was bootstrapped 100 times using *fastsimcoal* and each of these datasets was analysed under the same conditions as the original data (one hundred independent runs for each dataset). Calibrating the molecular clock is crucial to obtain accurate estimates of demographic parameters and historical events, but it is challenging when fossil records and/or orthologous loci from an outgroup are lacking. Here, all demographic inferences were performed

using the Rad-seq mutation rate of $\mu = 1.93 \times 10^{-8}$ per site and per generation previously used for the tiger shark (7), and the generation time was set to 10 years (24, 60).

Declarations

Ethics approval and consent to participate

This research was conducted in accordance with Permit No. 6024- 4916/DENV/SMer (New Caledonia), Permit No. QS2010 GS065 and Permit No. 143005 (Australia), and Permit No. TREL2118715S/541 (Reunion). Samples from Brazil were provided by Jennifer Schults from Hawaii, in accordance with local regulations. All methods are reported in accordance with ARRIVE guidelines (<https://arriveguidelines.org>).

Consent for publication

Not applicable.

Availability of data and materials

Fastq sequence files are available from the GenBank at the National Center for Biotechnology Information short-read archive database (accession number: forthcoming).

Competing interests

The authors declare no conflict of interest.

Funding

This work was supported by two ATM grants (2016 and 2017) from the Muséum National d'Histoire Naturelle to Stefano Mona.

Authors' contribution

S.M. and H.M. conceived the project. S.M., P.L and H.L. analysed the data with input from P.M.D. E.C., H.M. and S.J. provided reagents and samples. A.S. and P.C.B performed the molecular lab work. P.L., S.M. and H.M. wrote the manuscript with input from all the others.

Acknowledgements

We thank Serge Planes for providing DNA samples of all specimens but Reunion Island, Gavin Naylor for critical reading of the manuscript, and Thibaut Caley for discussion above the Agulhas current. We thank Andrea Benazzo for help with angsd. We are grateful to the Genotoul bioinformatics platform Toulouse Midi-Pyrenees (Bioinfo Genotoul; <http://bioinfo.genotoul.fr/>) for providing computing resources. All data and analyses from Reunion Island were produced with the help of A. Pirog and financially supported by DEAL Réunion under the project ECoReCo-Run.

References

1. Terborgh JW. Toward a trophic theory of species diversity. *Proc Natl Acad Sci U S A*. 2015;112(37):11415–22.
2. Myers RA, Worm B. Rapid worldwide depletion of predatory fish communities. *Nature* [Internet]. 2003 May;423(6937):280–3. Available from: <http://www.nature.com/articles/nature01610>
3. Palsbøll PJ, Bérubé M, Allendorf FW. Identification of management units using population genetic data. *Trends Ecol Evol* [Internet]. 2007 Jan;22(1):11–6. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S0169534706002965>
4. Chikhi L, Sousa VC, Luisi P, Goossens B, Beaumont MA. The Confounding Effects of Population Structure, Genetic Diversity and the Sampling Scheme on the Detection and Quantification of Population Size Changes. *Genetics* [Internet]. 2010 Nov 1;186(3):983–95. Available from: <https://academic.oup.com/genetics/article/186/3/983-995/6063626>
5. Mona S, Ray N, Arenas M, Excoffier L. Genetic consequences of habitat fragmentation during a range expansion. *Heredity (Edinb)* [Internet]. 2014 Mar 23;112(3):291–9. Available from: <http://www.nature.com/articles/hdy2013105>
6. Mazet O, Rodríguez W, Chikhi L. Demographic inference using genetic data from a single individual: Separating population size variation from population structure. *Theor Popul Biol* [Internet]. 2015 Sep;104:46–58. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S0040580915000581>
7. Lesturgie P, Planes S, Mona S. Coalescence times, life history traits and conservation concerns: An example from four coastal shark species from the Indo-Pacific. *Mol Ecol Resour* [Internet]. 2022 Feb 28;22(2):554–66. Available from: <https://onlinelibrary.wiley.com/doi/10.1111/1755-0998.13487>
8. Compagno LJV. FAO species catalogue, Vol. 4: Sharks of the world: An annotated and illustrated catalogue of shark species known to date. Part 2 - Carcharhiniformes. In: FAO Fisheries Synopsis. Rome, Italy; 1984. p. 503–12.
9. Temple AJ, Kiszka JJ, Stead SM, Wambiji N, Brito A, Poonian CNS, et al. Marine megafauna interactions with small-scale fisheries in the southwestern Indian Ocean: a review of status and challenges for research and management. *Rev Fish Biol Fish* [Internet]. 2018 Mar 18;28(1):89–115. Available from: <http://link.springer.com/10.1007/s11160-017-9494-x>
10. Sumpton W, Taylor S, Gribble N, McPherson G, Ham T. Gear selectivity of large-mesh nets and drumlines used to catch sharks in the Queensland Shark Control Program. *African J Mar Sci*

- [Internet]. 2011 Apr;33(1):37–43. Available from:
<http://www.tandfonline.com/doi/abs/10.2989/1814232X.2011.572335>
11. Clarke SC, McAllister MK, Milner-Gulland EJ, Kirkwood GP, Michielsens CGJ, Agnew DJ, et al. Global estimates of shark catches using trade records from commercial markets. *Ecol Lett* [Internet]. 2006 Oct;9(10):1115–26. Available from: <http://doi.wiley.com/10.1111/j.1461-0248.2006.00968.x>
 12. Payne NL, Meyer CG, Smith JA, Houghton JDR, Barnett A, Holmes BJ, et al. Combining abundance and performance data reveals how temperature regulates coastal occurrences and activity of a roaming apex predator. *Glob Chang Biol* [Internet]. 2018 May;24(5):1884–93. Available from: <http://doi.wiley.com/10.1111/gcb.14088>
 13. Holland KN, Anderson JM, Coffey DM, Holmes BJ, Meyer CG, Royer MA. A Perspective on Future Tiger Shark Research. *Front Mar Sci* [Internet]. 2019 Feb 14;6(FEB):1–7. Available from: <https://www.frontiersin.org/article/10.3389/fmars.2019.00037/full>
 14. Werry JM, Planes S, Berumen ML, Lee KA, Braun CD, Clua E. Reef-Fidelity and Migration of Tiger Sharks, *Galeocerdo cuvier*, across the Coral Sea. Klimley AP, editor. *PLoS One* [Internet]. 2014 Jan 8;9(1):e83249. Available from: <https://dx.plos.org/10.1371/journal.pone.0083249>
 15. Lea JSE, Wetherbee BM, Queiroz N, Burnie N, Aming C, Sousa LL, et al. Repeated, long-distance migrations by a philopatric predator targeting highly contrasting ecosystems. *Sci Rep* [Internet]. 2015 Sep 9;5(1):11202. Available from: <http://www.nature.com/articles/srep11202>
 16. Meyer CG, Anderson JM, Coffey DM, Hutchinson MR, Royer MA, Holland KN. Habitat geography around Hawaii's oceanic islands influences tiger shark (*Galeocerdo cuvier*) spatial behaviour and shark bite risk at ocean recreation sites. *Sci Rep* [Internet]. 2018 Dec 21;8(1):4945. Available from: <http://www.nature.com/articles/s41598-018-23006-0>
 17. Palumbi SR. Genetic divergence, reproductive isolation, and marine speciation [Internet]. Vol. 25, *Annual Review of Ecology and Systematics*. 1994. p. 547–72. Available from: <http://www.annualreviews.org/doi/10.1146/annurev.es.25.110194.002555>
 18. Bailleul D, Mackenzie A, Sacchi O, Poisson F, Bierne N, Arnaud-Haond S. 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? *Evol Appl*. 2018;11(5):614–30.
 19. Corrigan S, Lowther AD, Beheregaray LB, Bruce BD, Cliff G, Duffy CA, et al. Population Connectivity of the Highly Migratory Shortfin Mako (*Isurus oxyrinchus Rafinesque 1810*) and Implications for Management in the Southern Hemisphere. *Front Ecol Evol* [Internet]. 2018 Nov 20;6(NOV):1–15. Available from: <https://www.frontiersin.org/article/10.3389/fevo.2018.00187/full>
 20. Corrigan S, Maisano Delsler P, Eddy C, Duffy C, Yang L, Li C, et al. Historical introgression drives pervasive mitochondrial admixture between two species of pelagic sharks. *Mol Phylogenet Evol* [Internet]. 2017 May;110(December):122–6. Available from: <http://dx.doi.org/10.1016/j.ympev.2017.03.011>
 21. Naylor GJP, Caira JN, Jensen K, Rosana KAM, White WT, Last PR. A DNA sequencebased approach to the identification of shark and ray species and its implications for global elasmobranch diversity and

- parasitology [Internet]. Vol. 367, Bulletin of the American Museum of Natural History. 2012. p. 1–262. Available from: <http://www.bioone.org/doi/abs/10.1206/754.1>
22. Bernard AM, Feldheim KA, Heithaus MR, Wintner SP, Wetherbee BM, Shivji MS. Global population genetic dynamics of a highly migratory, apex predator shark. *Mol Ecol* [Internet]. 2016 Nov;25(21):5312–29. Available from: <http://doi.wiley.com/10.1111/mec.13845>
 23. Holmes BJ, Williams SM, Otway NM, Nielsen EE, Maher SL, Bennett MB, et al. Population structure and connectivity of tiger sharks (*Galeocerdo cuvier*) across the Indo-Pacific Ocean basin. *R Soc Open Sci* [Internet]. 2017 Jul 5;4(7):170309. Available from: <https://royalsocietypublishing.org/doi/10.1098/rsos.170309>
 24. Pirog A, Jaquemet S, Ravigné V, Cliff G, Clua E, Holmes BJ, et al. Genetic population structure and demography of an apex predator, the tiger shark *Galeocerdo cuvier*. *Ecol Evol* [Internet]. 2019 May 4;9(10):5551–71. Available from: <https://onlinelibrary.wiley.com/doi/abs/10.1002/ece3.5111>
 25. Carmo CB, Ferrette BLS, Camargo SM, Roxo FF, Coelho R, Garla RC, et al. A new map of the tiger shark (< scp > *Galeocerdo cuvier* </ scp >) genetic population structure in the western Atlantic Ocean: Hypothesis of an equatorial convergence centre. *Aquat Conserv Mar Freshw Ecosyst* [Internet]. 2019 May 2;29(5):760–72. Available from: <https://onlinelibrary.wiley.com/doi/abs/10.1002/aqc.3029>
 26. Andrade FRS, Afonso AS, Hazin FHV, Mendonça FF, Torres RA. Population genetics reveals global and regional history of the apex predator *Galeocerdo cuvier* (carcharhiniformes) with comments on mitigating shark attacks in north-eastern brazil. *Mar Ecol* [Internet]. 2021 Apr 21;42(2):1–16. Available from: <https://onlinelibrary.wiley.com/doi/10.1111/maec.12640>
 27. Bernard AM, Finnegan KA, Pavinski Bitar P, Stanhope MJ, Shivji MS. Genomic Assessment of Global Population Structure in a Highly Migratory and Habitat Versatile Apex Predator, the Tiger Shark (*Galeocerdo cuvier*). *J Hered*. 2021;112(6):497–507.
 28. Nielsen R, Wakeley J. Distinguishing migration from isolation: A Markov chain Monte Carlo approach. *Genetics*. 2001;158(2):885–96.
 29. Beerli P, Felsenstein J. Maximum likelihood estimation of a migration matrix and effective population sizes in n subpopulations by using a coalescent approach. *Proc Natl Acad Sci U S A*. 2001;98(8):4563–8.
 30. Korneliussen TS, Albrechtsen A, Nielsen R. ANGSD: Analysis of Next Generation Sequencing Data. *BMC Bioinformatics* [Internet]. 2014 Dec 25;15(1):1–13. Available from: <https://bmcbioinformatics.biomedcentral.com/articles/10.1186/s12859-014-0356-4>
 31. Peter BM, Wegmann D, Excoffier L. Distinguishing between population bottleneck and population subdivision by a Bayesian model choice procedure. *Mol Ecol* [Internet]. 2010 Nov 23;19(21):4648–60. Available from: <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1365-294X.2010.04783.x>
 32. Maisano Delser P, Corrigan S, Duckett D, Suwalski A, Veuille M, Planes S, et al. Demographic inferences after a range expansion can be biased: the test case of the blacktip reef shark (*Carcharhinus melanopterus*). *Heredity (Edinb)* [Internet]. 2019 Jun 21;122(6):759–69. Available from: <http://dx.doi.org/10.1038/s41437-018-0164-0>

33. Beal LM, Ruijter WPM De, Biastoch A, Zahn R, Wcrp S, Working I. On the role of the Agulhas system in ocean circulation and climate. *Nature*. 2011;472(7344):429–36.
34. Gaither MR, Bowen BW, Rocha LA, Briggs JC. Fishes that rule the world: circumtropical distributions revisited. *Fish Fish* [Internet]. 2016 Sep 30;17(3):664–79. Available from: <http://doi.wiley.com/10.1111/faf.12136>
35. Maduna SN, Rossouw C, da Silva C, Soekoe M, Bester-van der Merwe AE. Species identification and comparative population genetics of four coastal houndsharks based on novel NGS-mined microsatellites. *Ecol Evol* [Internet]. 2017 Mar;7(5):1462–86. Available from: <http://doi.wiley.com/10.1002/ece3.2770>
36. Reid BN, Naro-Maciel E, Hahn AT, FitzSimmons NN, Gehara M. Geography best explains global patterns of genetic diversity and postglacial co-expansion in marine turtles. *Mol Ecol*. 2019;28(14):3358–70.
37. van der Zee JP, Christianen MJA, Bérubé M, Nava M, Schut K, Humber F, et al. The population genomic structure of green turtles (*Chelonia mydas*) suggests a warm-water corridor for tropical marine fauna between the Atlantic and Indian oceans during the last interglacial. *Heredity (Edinb)*. 2021;127(6):510–21.
38. Caley T, Giraudeau J, Malaizé B, Rossignol L, Pierre C. Agulhas leakage as a key process in the modes of Quaternary climate changes. *Proc Natl Acad Sci U S A* [Internet]. 2012 May 1;109(18):6835–9. Available from: <http://www.pnas.org/cgi/doi/10.1073/pnas.1115545109>
39. Caley T, Peeters FJC, Biastoch A, Rossignol L, van Sebille E, Durgadoo J, et al. Quantitative estimate of the paleo-Agulhas leakage. *Geophys Res Lett* [Internet]. 2014 Feb 28;41(4):1238–46. Available from: <http://doi.wiley.com/10.1002/2014GL059278>
40. Chikhi L, Rodríguez W, Grusea S, Santos P, Boitard S, Mazet O. The IICR (inverse instantaneous coalescence rate) as a summary of genomic diversity: Insights into demographic inference and model choice. *Heredity (Edinb)* [Internet]. 2018 Jan 8;120(1):13–24. Available from: <http://dx.doi.org/10.1038/s41437-017-0005-6>
41. Mazet O, Rodríguez W, Grusea S, Boitard S, Chikhi L. On the importance of being structured: instantaneous coalescence rates and human evolution—lessons for ancestral population size inference? *Heredity (Edinb)* [Internet]. 2016 Apr 9;116(4):362–71. Available from: <http://www.nature.com/articles/hdy2015104>
42. Peterson CD, Belcher CN, Bethea DM, Driggers WB, Frazier BS, Latour RJ. Preliminary recovery of coastal sharks in the south-east United States. *Fish Fish* [Internet]. 2017 Sep;18(5):845–59. Available from: <http://doi.wiley.com/10.1111/faf.12210>
43. Kerdoncuff E, Lambert A, Achaz G. Testing for population decline using maximal linkage disequilibrium blocks. *Theor Popul Biol* [Internet]. 2020 Aug;134:171–81. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S0040580920300289>
44. Lapierre M, Lambert A, Achaz G. Accuracy of Demographic Inferences from the Site Frequency Spectrum: The Case of the Yoruba Population. *Genetics* [Internet]. 2017 May 1;206(1):439–49.

- Available from: <https://academic.oup.com/genetics/article/206/1/439-449/6064209>
45. Barkley AN, Gollock M, Samoilyis M, Llewellyn F, Shivji M, Wetherbee B, et al. Complex transboundary movements of marine megafauna in the Western Indian Ocean. *Anim Conserv* [Internet]. 2019 Oct 10;22(5):420–31. Available from: <https://onlinelibrary.wiley.com/doi/abs/10.1111/acv.12493>
 46. Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE. Double Digest RADseq: An Inexpensive Method for De Novo SNP Discovery and Genotyping in Model and Non-Model Species. Orlando L, editor. *PLoS One* [Internet]. 2012 May 31;7(5):e37135. Available from: <https://dx.plos.org/10.1371/journal.pone.0037135>
 47. Rochette NC, Rivera-Colón AG, Catchen JM. Stacks 2: Analytical methods for paired-end sequencing improve RADseq-based population genomics. *Mol Ecol*. 2019;28(21):4737–54.
 48. Fountain ED, Pauli JN, Reid BN, Palsbøll PJ, Peery MZ. Finding the right coverage: the impact of coverage and sequence quality on single nucleotide polymorphism genotyping error rates. *Mol Ecol Resour* [Internet]. 2016 Jul 25;16(4):966–78. Available from: <https://onlinelibrary.wiley.com/doi/10.1111/1755-0998.12519>
 49. Khimoun A, Doums C, Molet M, Kaufmann B, Peronnet R, Eyer PA, et al. Urbanization without isolation: The absence of genetic structure among cities and forests in the tiny acorn ant *Temnothorax nylanderi*. *Biol Lett*. 2020;16(1).
 50. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* [Internet]. 2009 Jul 15;25(14):1754–60. Available from: <https://academic.oup.com/bioinformatics/article-lookup/doi/10.1093/bioinformatics/btp324>
 51. Meisner J, Albrechtsen A. Inferring Population Structure and Admixture Proportions in Low-Depth NGS Data. *Genetics* [Internet]. 2018 Oct 1;210(2):719–31. Available from: <https://academic.oup.com/genetics/article/210/2/719-731/5931101>
 52. Nielsen R, Korneliussen T, Albrechtsen A, Li Y, Wang J. SNP Calling, Genotype Calling, and Sample Allele Frequency Estimation from New-Generation Sequencing Data. Awadalla P, editor. *PLoS One* [Internet]. 2012 Jul 24;7(7):e37558. Available from: <https://dx.plos.org/10.1371/journal.pone.0037558>
 53. Mantel N. The Detection of Disease Clustering and a Generalized Regression Approach. *Cancer Res*. 1967;27(2):209–20.
 54. Excoffier L, Dupanloup I, Huerta-Sánchez E, Sousa VC, Foll M. Robust Demographic Inference from Genomic and SNP Data. Akey JM, editor. *PLoS Genet* [Internet]. 2013 Oct 24;9(10):e1003905. Available from: <https://dx.plos.org/10.1371/journal.pgen.1003905>
 55. Pudlo P, Marin J-MM, Estoup A, Cornuet J-MM, Gautier M, Robert CP. Reliable ABC model choice via random forests. *Bioinformatics* [Internet]. 2016 Mar 15;32(6):859–66. Available from: <https://academic.oup.com/bioinformatics/article/32/6/859/1744513>
 56. Watterson GAA. On the number of segregating sites in genetical models without recombination. *Theor Popul Biol* [Internet]. 1975 Apr;7(2):256–76. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/1145509>

57. Tajima F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* [Internet]. 1989 Nov 1;123(3):585–95. Available from: <https://academic.oup.com/genetics/article/123/3/585-595/5998755>
58. Liu X, Fu YX. Stairway Plot 2: demographic history inference with folded SNP frequency spectra. *Genome Biol.* 2020;21(1):1–9.
59. Meier JI, Sousa VC, Marques DA, Selz OM, Wagner CE, Excoffier L, et al. Demographic modelling with whole-genome data reveals parallel origin of similar *Pundamilia* cichlid species after hybridization. *Mol Ecol* [Internet]. 2017 Jan;26(1):123–41. Available from: <http://doi.wiley.com/10.1111/mec.13838>
60. Cortés E. Incorporating uncertainty into demographic modeling: Application to shark populations and their conservation. *Conserv Biol.* 2002;16(4):1048–62.

Figures

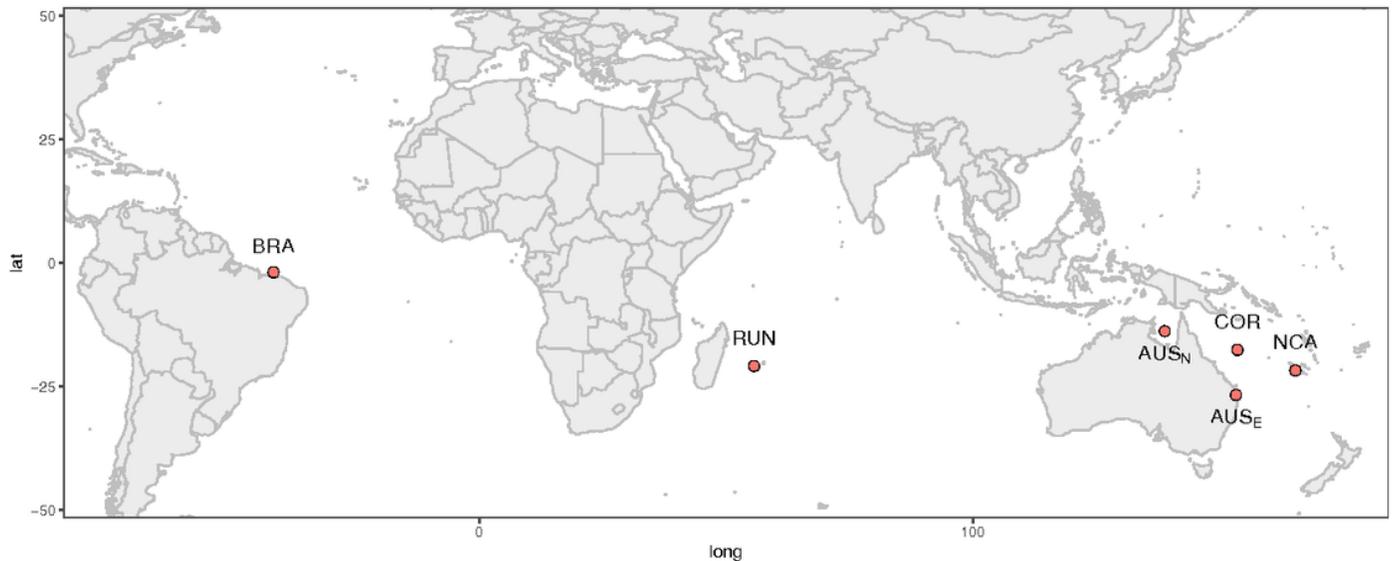


Figure 1

Map of the sampling sites. From west to east: Brazil (BRA; n = 7), Reunion Island (RUN; n = 15), North Coast of Australia (AUS_N; n = 8), Coral Sea (COR; n = 5), East Coast of Australia (AUS_E; n = 7) and New Caledonia (NCA; n = 8).

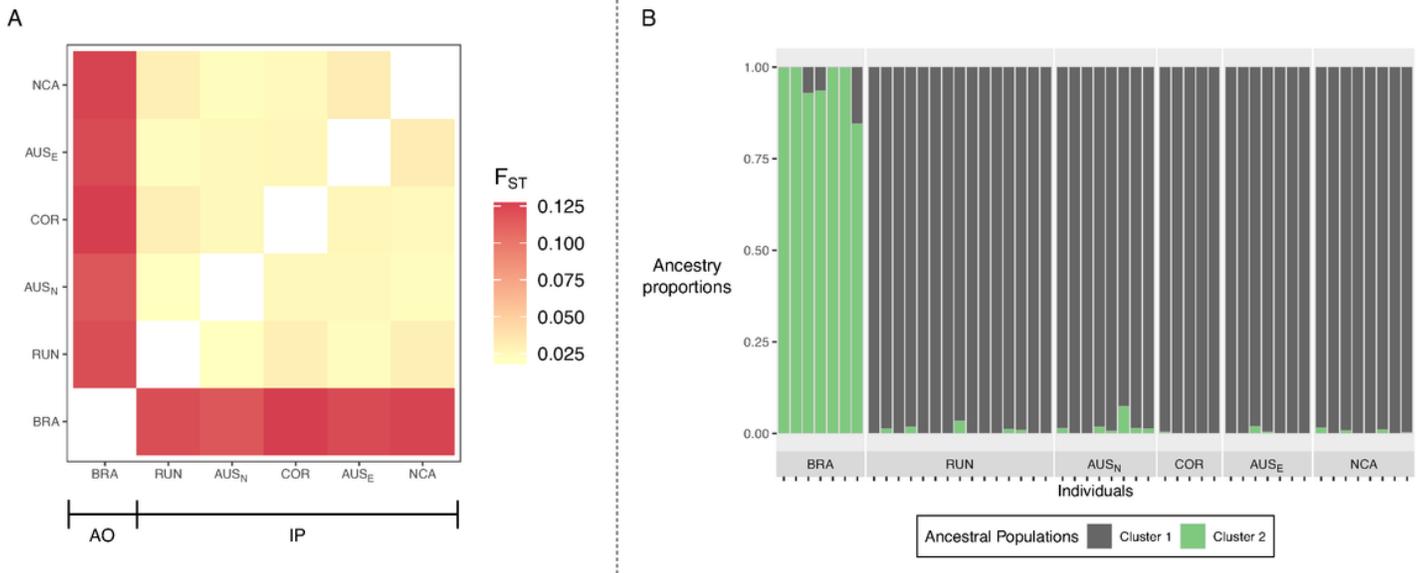


Figure 2

Ancestry proportions retrieved using the nmf algorithm with $K=2$ ancestral populations (A) and heat map representing the pairwise Reynold's F_{ST} values between sampling sites (B). Both analyses were performed with PCAngsd.

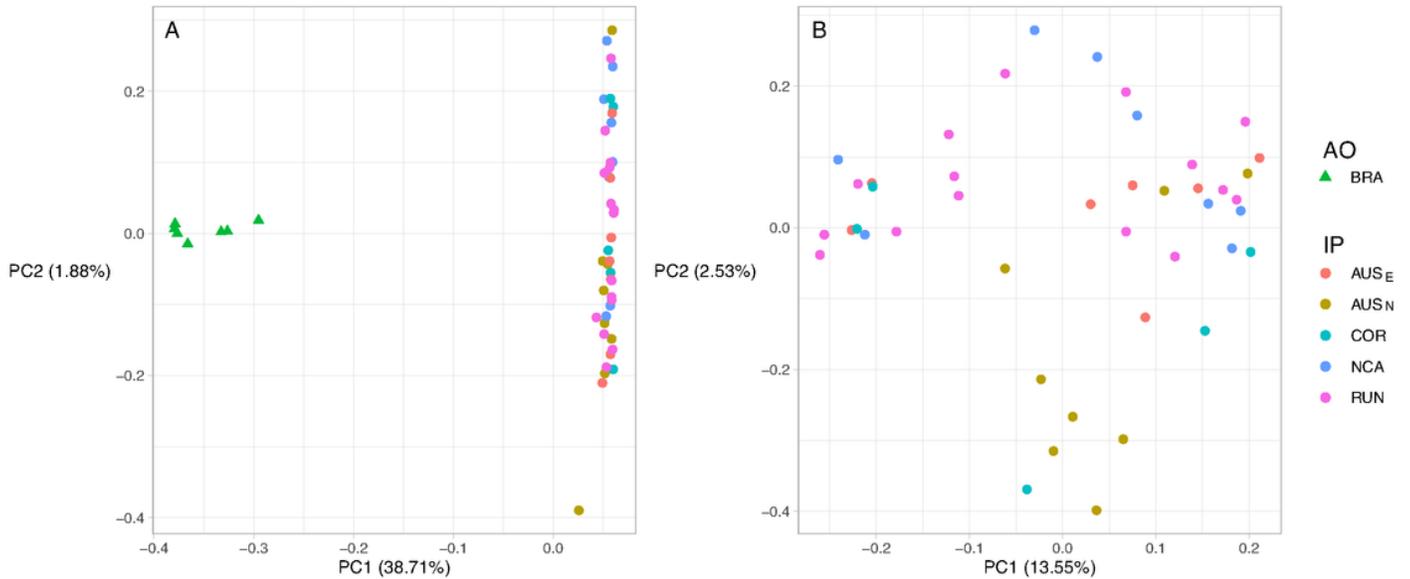


Figure 3

Principal Component Analysis (PCA) computed with: (A) all individuals ($n = 50$) and (B) Indo-Pacific individuals only ($n = 43$).

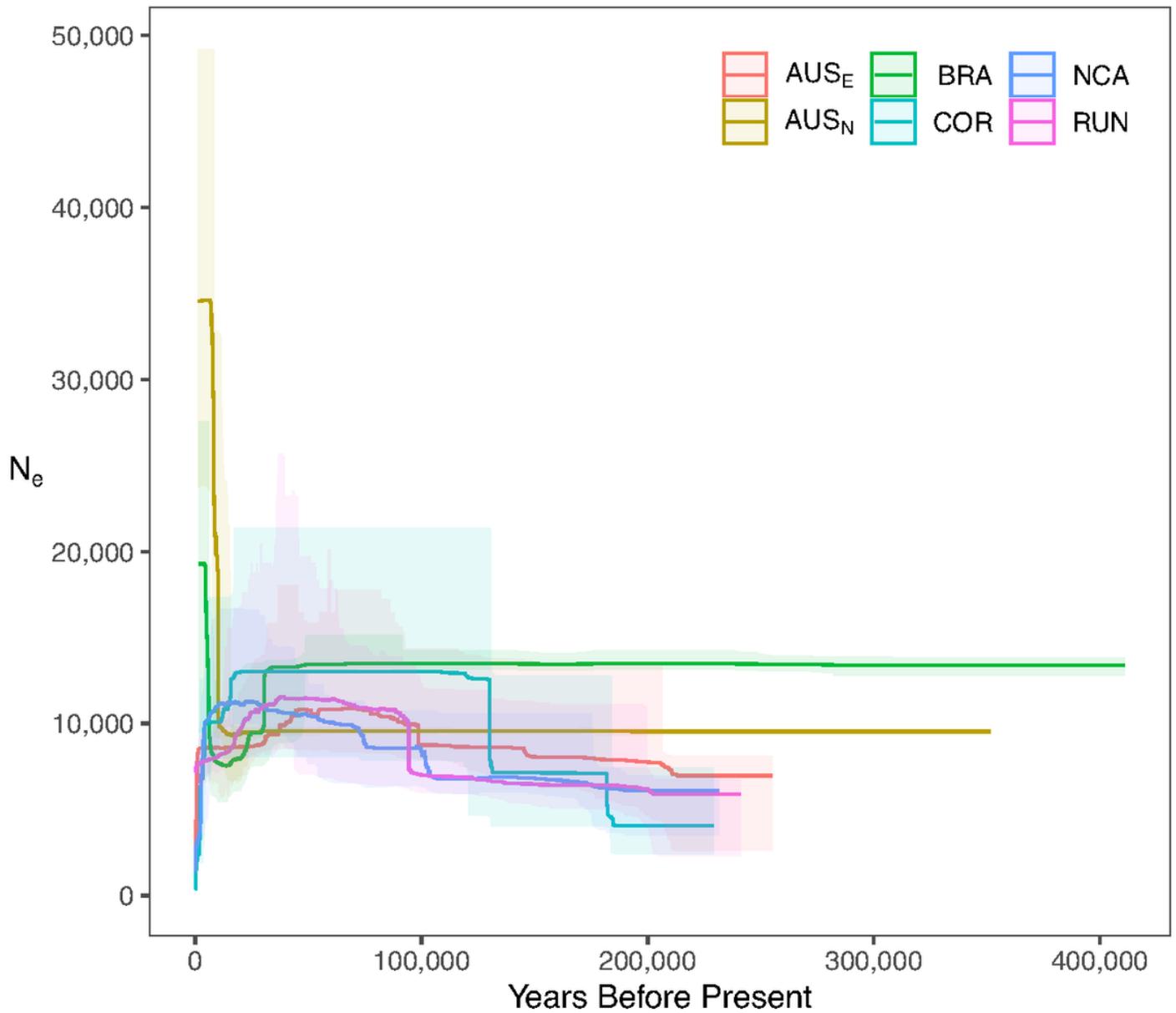


Figure 4

Variation of the effective population size (N_e) through time and its 75% confidence interval estimated by the stairwayplot for all sampling sites. AUS_E: East Coast of Australia; AUS_N: North Coast of Australia; BRA: Brazil; COR: Coral Sea; NCA: New Caledonia; RUN: Reunion Island.

Figure 5

See image above for figure legend

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [suppinfo.docx](#)