Coalescent times, life history traits and conservation concerns: an example from four shark species from the Indo-Pacific

Pierre Lesturgie[1], Serge Planes[2], and Stefano Mona[1]

[1]Museum National d'Histoire Naturelle
[2]Université de Perpignan

March 07, 2024

Abstract

Dispersal abilities play a crucial role in shaping the extent of population genetic structure, with more mobile species being panmictic over large geographic ranges and less mobile ones organized in meta-populations exchanging migrants to different degrees. In turn, population structure directly influences the coalescent pattern of the sampled lineages, but the consequences on the estimated variation of the effective population size (*Ne*) over time obtained by means of *unstructured* demographic models remain poorly understood. However, this knowledge is crucial for biologically interpreting the observed Ne trajectory and further devising conservation strategies in endangered species. Here we investigated the demographic history of four shark species (*Carharhinus melanopterus, Carharhinus limbatus, Carharhinus amblyrhynchos, Galeocerdo cuvier*) with different degrees of endangered status and life history traits related to dispersal distributed in the Indo-Pacific and sampled off New Caledonia. We compared several evolutionary scenarios representing both *structured* (meta-population) and *unstructured* models and then inferred the *Ne* variation through time. By performing extensive coalescent simulations, we provided a general framework relating the underlying population structure and the observed *Ne* dynamics. On this basis, we concluded that the recent decline observed in three out of the four considered species when assuming *unstructured* demographic models can be explained by the presence of population structure. Furthermore, we also demonstrated the limits of the inferences based on the sole site frequency spectrum and warn that statistics based on linkage disequilibrium will be needed to exclude recent demographic events affecting meta-populations.

Introduction

Reconstructing the evolutionary history of a species is a challenging exercise only partially eased by the growing size of genetic data available. However, larger amounts of data will indeed provide more precision but not more accuracy if the model(s) chosen to infer demographic parameters is distant from the true one. Species are dynamic entities whose geographic range has often changed in time trough range expansions, contractions and shifts (Arenas, Ray, Currat, & Excoffier, 2012; Excoffier, Foll, & Petit, 2009; Mona, Ray, Arenas, & Excoffier, 2014). This suggests that species are most likely organized in meta-populations (i.e. groups of demes or sub-populations exchanging migrants to some extent), even though the more vagile ones might be panmictic at a large scale (Corrigan et al., 2018; Karl et al., 2010). Neglecting the meta-population structure (i.e., performing demographic inferences under *unstructured* models) may lead to spurious inference of population size change (Chikhi, Sousa, Luisi, Goossens, & Beaumont, 2010; Maisano Delser et al., 2019, 2016; Olivier Mazet, Rodríguez, & Chikhi, 2015), which is particularly worrisome for species of conservation concern. Unfortunately, the link between the inferred temporal trajectory of *Ne* and the real demographic history of the meta-population remains largely under explored. However, the role of connectivity, particularly the number of migrants *Nm* exchanged each generation and the migration matrix, has been put forward as a key actor in shaping the gene genealogy of lineages sampled from a deme belonging to a meta-population (Chikhi et al., 2010; Mona et al., 2014; Ray, Currat, & Excoffier, 2003; Städler, Haubold, Merino, Stephan,

& Pfaffelhuber, 2009).

Understanding the relations between meta-population structure, the inferred *Ne* variation under *unstructured* models, and species dispersal abilities, is crucial to correctly interpret the pattern of genetic variability and to establish conservation priorities. To search for general rules describing such relations, we followed an inductive approach investigating species: i) with large distribution (which in principle should guarantee an organization in meta-populations); ii) with different life history traits (LHT) related to dispersal; iii) of conservation concerns. In this spirit, we selected for our genomic study four shark species (*Carcharhinus amblyrhynchos* , *Carcharhinus limbatus* , *Carcharhinus melanopterus,* and *Galeocerdo cuvier*) from New Caledonia. These species have a large and overlapping distribution in the Indo-Pacific (https://sharksrays.org/) and they differ for LHT features such as size (which is positively correlated with the capacity for long distance swimming and oceanic migration (Parsons, 1990)), residency pattern, and long-distance dispersal ability as measured by tagging data (Table S1). Moreover, the IUCN red list reported that *Carcharhinus limbatus*and *Galeocerdo cuvier* are Near Threatened (with a decreasing trend in the tiger shark), *Carcharhinus melanopterus* is vulnerable with decreasing trend, and *Carcharhinus amblyrhynchos*is Endangered with decreasing trend as well. We first compared several population genetics models by means of coalescent simulations coupled with an *approximate Bayesian computation* framework (Bertorelle, Benazzo, & Mona, 2010) to detect whether panmixia or a meta-population model best describe the genomic variation of each species. Then, we inferred the demographic parameters under the most likely model and applied the *stairwayplot* , which assumes a panmictic unstructured population (Liu & Fu, 2015), to detect the*Ne* variation through time in each species. We finally run extensive coalescent simulations under the tested meta-population models with parameters compatible to those observed in real data. The simulated datasets were in turn analysed with the *stairwayplot* to: i) help interpreting the observed data in the four shark species; ii) providing general coalescent arguments relating the demographic history of a meta-population and the reconstructed variation in *Ne* through time by means of *unstructured* models.

Material & Methods

Sampling

Eight specimens of tiger shark (*G. cuvier*), 13 black tip shark (*C. limbatus*), and 12 grey reef shark (*C. amblyrhynchos*) were collected off New Caledonia. Total genomic DNA was extracted from muscle tissue or fin clips, and preserved in 96% ethanol using QIAGEN DNeasy Blood and Tissue purification kit (Qiagen, Hilden, Germany) according to the manufacturer's protocols. Double-digest restriction-associated DNA (ddRAD) libraries were prepared following Peterson et al. (2012) using EcoRI and MspI restriction enzymes and a 400-bp size selection. The genomic libraries obtained were sequenced with a HiSeq 2500 Illumina sequencer (single-end, 125 bp). Exon capture data of eight black tip reef shark (*Carcharhnius melanopterus*) from New Caledonia (Maisano Delser et al. 2019) were included in this study for comparative purposes.

De novo assembly and data filtering (dd-RADseq samples)

Raw reads were first demultiplexed and quality filtered through the*process_radtags.pl* pipeline in stacks v.2.5 (Rochette, Rivera-Colón, & Catchen, 2019). In the absence of a reference genome for any of the three species, RAD-seq loci were *de novo* assembled independently in each species under the *denovo_map.pl* pipeline in stacks. We used the following assembly parameters:m =3 (minimum read depth to create a stack), M =4 (number of mismatches allowed between loci within individuals), and n=*4* (number of mismatches allowed between loci within catalogue). We found an average coverage per species of ˜10x (see results). A consensus on the threshold below which SNP calling may be considered unreliable is still lacking. However, genotype free estimation of allele frequency is generally recommended with low to medium coverage (Korneliussen, Albrechtsen, & Nielsen, 2014). This approach, implemented in the software angsd v.0.923 (Korneliussen et al., 2014), has been rarely applied to Rad-seq data (however, see Warmuth and Ellegren (2019) for an exception) and, to our knowledge, never to Rad-seq data from non-model organisms, probably due to need of a reference sequence for the software to work. Here, we followed the approach of Khimoun et al., (2020) by creating an artificial reference sequence. First, we used the *population* script in stacks to assembly loci present in at least 80% of the individuals (using the flag *r=0.8*); then, we concatenated the consensus sequences of the retrieved loci spaced by a stretch of 120 N (unknown) characters (the same length of the Rad-loci) to facilitate the subsequent mapping. Raw reads were then mapped back to the novel reference sequence by means of the *bwa-mem* algorithm with default parameters (Li & Durbin, 2009). Using custom bash scripts coupled with angsd, we applied a number of filters to the aligned data and eliminated: *i)* sites with coverage <3 (-*minIndDepth =3* flag),*ii)* bad quality bases and poorly aligning reads (-*minQ* and -*minMapQ* and -*C* flags with default values); *iii)*poor quality sites based on the per base alignment quality (-*baq=1* flag); : *iv)* SNPs in the last 5 bp of each locus;*v)* SNPs heterozygote in at least 80% of individuals; *vi)*loci with more than 5 SNPs that could potentially be paralogous;*vii*) sites with missing data by setting the -*minInd* flag to the total number of individuals retained in each species. The filtered dataset was then used to generate a site allele frequency likelihood file, with the genotype likelihoods computed with the SAMtools method (-*GL=1* flag), further optimised to compute a folded *site frequency spectrum* (SFS) with no missing data for downstream analyses. An alternative (and simpler) approach would have been to augment m to achieve an higher coverage (Paris, Stevens, & Catchen, 2017). However, beside the considerable loss in the number of assembled loci (and hence of retrieved SNPs), we found by extensive simulation of *in silico* Rad experiments that selecting high coverage loci biases the SFS towards low frequency variants (Mona, Bertorelle, Benazzo, in preparation). The SFS for *C. melanopterus*was estimated directly from the high coverage exon-capture dataset of Maisano Delser et al. (2019).

Genetic diversity and demographic inferences

Nucleotide diversity ( ),$_w$ (Watterson's theta, based on segregating sites (Watterson, 1975)) and Tajima's *D (TD,* (Tajima, 1989)) were computed from the SFS for each species with custom scripts. Significance of *TD* was evaluated after 1,000 coalescent simulations of a constant population model with scaled size $_\pi$. To test whether sampled demes are isolated or belong to a structured meta-population and to eventually estimate connectivity, we devised three alternative evolutionary models for each species (Figure 1) within an *approximate bayesian computation* framework (ABC) framework. Model NS (non-structured) defined an isolated population characterized by a modern effective population size ($Nmod$) switching instantaneously into an ancestral population size ($Nanc$) atTc generations before present. Model FIM specifies a non-equilibrium finite island model defined by 100 demes exchangingNm migrants each generation under a symmetric migration matrix. The array of demes is instantaneously colonized $Tcol$ generations before present from a population with an ancestral size ($Nanc$). Model SST is similar to FIM but demes exchanging migrants only with their four neighbours (or less, if they are at the border of the array), in a steppingstone fashion. We performed 50,000 coalescent simulations from prior distributions using fastsimcoal v.2.6.0.3 (Excoffier, Dupanloup, Huerta-Sánchez, Sousa, & Foll, 2013)*,*reproducing the exact number of individuals and loci for each species (Table 1). We first performed model selection through the random forest (RF) classification method implemented in the abcRF R package (Pudlo et al., 2016). We then performed 50,000 additional simulations under the most supported model in order to estimate demographic parameters with the abcRF regression method (Raynal et al., 2019). Both model selection and parameter estimation were computed with the following set of summary statistics: the SFS, ,$_w$ and *TD.* The first two axes of a Linear Discriminate Analysis performed on the previous statistics were also included for model selection in order to increase the accuracy of the estimates (Pudlo et al., 2016). Even though $_\pi$,$_w$ and *TD* are function of the SFS, they convey additional information by the non-linear feature of the functions. Information redundancy among the considered summary statistics is accounted for by the RF algorithm. Model selection and parameter estimation were performed twice to check the consistency of the analyses, and cross validation (or confusion matrix for the model selection) was performed on the first of the two runs. The number of trees in each RF algorithm was chosen by monitoring the evolution of the out-of-bag error (Pudlo et al., 2016).

We investigated the variation in the effective population size (Ne) through time by running the composite likelihood approach implemented in the stairwayplot v.0.2 software (Liu & Fu, 2015). We set the generation time to seven years for *C. melanopterus* (Maisano Delser et al., 2016) and to 10 years for the other species

(Cortés, 2002; Pirog et al., 2019) for all demographic inferences. We applied a mutation rate per generation per site of 8.4×10^{-9} to the exon capture data of *C. melanopterus* (Maisano Delser et al. 2016) and of 1.93×10^{-8} to the RADseq data for the remaining three species. This mutation rate was determined by scaling genetic diversity between ddRAD (obtained under the same protocol of this study) and Exon Capture data from 12 *C. melanopterus* individuals from Moorea, French Polynesia (Supplementary Material).

Simulation study

We ran coalescent simulations under FIM, SST and their modified version FIM-CH and SST-CH, where the Nm parameter is changed at Tch generations B.P. (Figure 1), to first inspect the shape of the SFS and to further uncover the variation of Ne over time assuming a panmictic population by means of the *stairwayplot* . We investigated in total 288 demographic scenarios under the four meta-population models (Tables 2, S4, S7, S8, S9 and S10). Similarly, to the analyses performed on the real data, all scenarios were represented by 100 demes exchanging migrants. We sampled 10 diploid individuals either from a randomly selected deme in the case of FIM/FIM-CH (since all demes have the same coalescent history) or from the central deme of the array in the case of SST/SST-CH (to avoid border effects). Deme size was fixed to N=5000 with m varying accordingly to obtain a long-term Nm of 1, 5, 10, and 15 in order to encompass the range of the estimated values (see results). $Tcol$ was fixed to 5,000, 15,000 and 50,000 generations B.P. or to [?] (i.e., equilibrium model). Change of connectivity occurred at Tch =10 or 50 generations B.P., to mimic human induced effects due to overfishing and/or habitat modifications (i.e., climate changes). Looking forward in time, we modelled the change in connectivity by instantaneously decreasing m or N by a factor 10 or 100 with respect to the long-term Nm (Tables S4, S7, S8, S9 and S10). For each combination of parameters, we performed 100 coalescent simulations of 50,000 Rad-like loci of 115 bp. Mutation rate per site per generations was set to $1.93\text{x}10^{-8}$ and the generation time to 10 years. We computed for each scenario (averaged over the 100 replicates): a) summary statistics ($,_w$, and TD); b) the normalised SFS as in (Lapierre, Lambert, & Achaz, 2017); c) the *stairwayplot* , to reconstruct the apparent variation of Ne through time.

Results

The number of assembled loci with no missing data ranged from 60,812 to 117,916 (Table 1). Mean coverage (and standard deviation) per sample was 9.02 (± 2.62), 7.93 (± 0.48), 8.39 (± 0.81) for *G. cuvier, C. limbatus* and *C. amblyrhynchos* respectively. Genetic diversity was very heterogeneous, with varying from 0.0004 (*C. melanopterus*) to 0.0021 (*C. amblyrhynchos*) per site (Table 1). Tajima's D was significantly positive for *C. melanopterus* (TD = 0.691, p<0.001) and *C. limbatus* (TD = 0.43, p<0.001), significantly negative for *C. amblyrhynchos* (TD = -0.23, p<0.001) and not significant for *G. cuvier* (TD = -0.03, p=0.059) (Table 1).

We compared the models NS, FIM, and SST (Figure 1) in the four species by means of an ABC-RF algorithm and estimated demographic parameters for the most supported model. After checking for the evolution of the out-of-bag error of the RF, model selection and parameter estimation were computed using respectively 500 and 1,000 trees in each species. We found that NS had the higher posterior probability (p =0.84) for *G. cuvier* , suggesting that an *unstructured* model fit the data better than structured models for this species (Table 1). In contrast, demographic histories of the three other species were best described by SST, with a posterior probability ranging from 0.53 to 0.88 (Table 1 and S2). The estimated median number of migrants per generation Nm was 1.8 (95% CI: 0.7-3.0) for *C. melanopterus* , 6.6 (95% CI: 1.5-15.4) for *C. limbatus,* and 11.5 (95% CI: 3.0-22.0) for *C. amblyrhynchos* (Figure 2, Table 1). The posterior distribution of Nm strongly differed from the prior distribution and showed a clear unimodal peak with small credible intervals, and low mean square error (SME) and mean root square error (SMRE) in all three species (Figure 2, Table S3), suggesting that these estimates are highly reliable. Conversely, both $Tcol$ and $Nanc$ had larger SME and SMRE errors in all species (Table S3), but it was only in *C. melanopterus* where posterior and prior distribution could not be distinguished (Figure 2). $Tcol$ has a clear unimodal distribution in *C. amblyrhynchos* but a more disperse one (and with wider credible intervals) in *C. limbatus* (Figure 2, Table 1).

The *stairwayplots* showed a nearly similar dynamic for *C. amblyrhynchos* and *C. limbatus* , characterized by a strong ancestral expansion (Figure 3). During a short time interval, the *Ne* increased rapidly about ~2 times, reaching a value between 35,000 and 42,000 individuals. When approaching T=0, both species underwent a bottleneck but of distinct strength: *Ne* dropped down to ~12,500 for *C. amblyrhynchos* but almost to 0 for *C. limbatus* . This is consistent with the shape of the normalized SFS, which clearly shows a stronger deficit in low frequency variants for *C. limbatus* compared to *C. amblyrhynchos* (Figure 3). Similarly to *C. limbatus* , *C. melanopterus* experienced a recent 10-fold population collapse around 20,000 years B.P. starting from a long term constant *Ne* . However, *C. melanopterus* showed no signature of ancestral expansion, consistent to the results obtained by Maisano Delser et al. (2019) using *abc-skyline* method. Finally, *G. cuvier* displayed an ancestral expansion around 100,000 years B.P. with *Ne* reaching ~12,000 before dropping to ~3000 at T ~1,600 years B.P. Remarkably, the ancestral expansion retrieved by the *stairwayplot* (Figure 3) for both *C. amblyrhynchos* and *C. limbatus* is overlapped with the posterior distribution of *Tcol* estimated by the SST model (Table 1). This analogy holds too for *C. melanopterus,* where *Tcol* could not be properly estimated under the structured model (we obtained a flat posterior distribution, Figure 2) and there was no signature of ancestral expansion in the *stairwayplot* (Figure 3).

In order to study the shape of the SFS under the considered models and to understand how *Ne* changes over time assuming a panmictic population, we run sets of coalescent simulations. The first set of coalescent simulations was run under FIM and SST only (Table 2 and S4) to check if simulated data could reproduce the pattern of genetic variability (both ϑ estimators and *TD*) observed for *C. melanoptetus* , *C.limbatus* , and *C. amblyrynchos* . The simulated ϑ values (excluding the equilibrium model) ranged between 0.001 and 0.003 per site, in line with the observed values (Table 1, 2 and S4). *TD* follows a U-shaped distribution for each *Nm* value as a function of *Tcol* , being more positive at recent *Tcol* and at equilibrium and less positive (or negative for higher *Nm*) at intermediate values. Therefore, species demography with *Nm* ~10 (and higher) and *Tcol* within 15k and 50k generations B.P. will have negative *TD* values. In contrast, species with lower *Nm* and very recent or very ancient *Tcol* will have positive *TD* . This matches strikingly the *TD* observed for the three shark species and their estimated demographic parameters under SST (Table 1). We plot the normalized SFS and the *stairwayplot* for all scenarios presented in Table 2 (Figure 4, 5, 6, S1 and S2). First, we note that none of our scenarios, even those at equilibrium and with no variation in *Nm* through time, showed a normalized SFS compatible with a constant size population (Figure 4, 5, 6, S1 and S2). The normalized SFS and the reconstructed *stairwayplot* depend generally on the interaction between *Nm* and *Tcol* with a dynamic strikingly similar to *TD* (which is indeed a summary of the SFS). For *Nm=1* we observed the signature of a recent decrease in *Ne* for all scenarios and independently of *Tcol* (Figure 4). The normalized SFS showed consistently a strong deficit of low frequency variants, typical of a demographic bottleneck and in agreement with the positive *TD* (Figure 4 and Table 1). Furthermore, the *stairwayplot* could never detect the ancestral expansion for any *Tcol* . These results are consistent with the normalized SFS and the *stairwayplot* observed in *C. melanopterus* , which had a median *Nm* of ~1 and for which it was not possible to estimate *Tcol* (Figure 2, 3 and Table 1). For growing *Nm,* the interplay with *Tcol* becomes more complex. A general result is that, once again, all scenarios were characterized by a recent decrease of *Ne* when looking at the *stairwayplot* and a deficit of singletons compared to the other low frequency classes when looking at the normalized SFS (Figure 5, 6 and S1). However, a strong signature of ancestral expansion appeared for *Nm* >10 and *Tcol* between 15k and 50k generations B.P., mirroring the results of *TD* for which most of these scenarios displayed a negative value. Remarkably, the *stairwayplot* retrieved the ancestral expansion only slightly overestimating the simulated *Tcol* (Figure 5, 6, S1). The signature of the expansion becomes weaker for growing *Tcol* (up to the equilibrium) even for larger *Nm* and the magnitude of the deficit of low frequency variants increases (Figure 5, 6, S1 and S2). These results are consistent with the *stairwayplot* and the normalized SFS observed in *C. amblyrynchos* and *C. limbatus* , with the former showing larger *Nm* and more recent *Tcol* than the latter. This can explain why *C. limbatus* showed a stronger decrease in *Ne* in recent times compared to *C. amblyrhynchos* (Figure 2, Table 1). Similar results were obtained for FIM (Figure S3).

We further tested whether a change in connectivity occurred in any of the three species best described by SST. To this end, we compared SST vs SST-CH model (Figure 1) by means of the same ABC-RF model selection framework previously adopted. The two models cannot be clearly distinguished in any of the three species since: *i*) they showed similar posterior probability (˜0.50); *ii*) the prior error rates are large ˜0.40 (Table S5);*iii*) posterior distributions of *Nm* before and after *Tch* are wide and largely overlapping (Table S6); *iv*) the normalized SFS closest to the observed data retrieved under the two models are very similar (Figure S4). These results can be explained either because SST and SST-CH produce a similar pattern in genetic diversity, or because we lacked enough data given our framework. To explore this issue, we ran a second set of coalescent simulations focusing on the consequences of a recent change in connectivity on the observed SFS and the reconstructed *stairwayplot* (Table S7 and S8). The decrease in connectivity was simulated by reducing either *m* (the migration rate per generation) or *N* (the effective population size of each deme) (Figure 7 and 8). As expected, we found a signature of recent population decline in all simulated scenarios, with its intensity only slightly affected by the change in *Nm* (Figure 7, 8, S5 and S6). However, the drop in *N* (Figure 8) had larger effect compared to the drop in *m* (Figure 7) on both the normalized SFS and the expansion time estimated by the *stairwayplot* . In scenarios with 100x reduction in *N* , the *stairwayplot* could not retrieve the ancestral expansion even for large long-term *Nm* (Figure 8). FIM-CH models displayed a behaviour similar to SST-CH models but more pronounced (Figure S7, S8, S9 and S10, Table S9 and S10). While at *Tch* =10 a decrease in *Nm* slightly affected the SFS and the reconstructed *stairwayplot* , the consequence of the change in connectivity are more substantial at *Tch* =50, with a stronger deficit in singletons and a more pronounced recent decline in *Ne* particularly in scenarios with a 100-fold reduction of *N* (Figure S8, S10).

Discussion

Discriminating whether the most appropriate model to reconstruct the demographic history of a species is *structured* or *unstructured* should be the first step in empirical population genetics investigations, particularly when targeting species of conservation concerns. Even when an extensive spatial sampling is lacking, an ABC model selection approach can actually distinguish whether the sampled deme belongs or not to a meta-population (similarly to previous studies (Maisano Delser et al., 2019; Peter, Wegmann, & Excoffier, 2010)). This should inform on how to interpret the variation in *Ne* through time traced using *unstructured* models. Among the four species considered here, the tiger shark is the only panmictic. This means that the reconstructed *stairwayplot* has a direct biological interpretation, with the shark experiencing a mild ancestral expansion and a recent ˜4-fold bottleneck around 2,000 years B.P. (consistent with the results of (Pirog et al., 2019)) (Figure 3). How to interpret the *stairwayplot* in the remaining three species? Coalescence simulations provided helpful and general insights into the understanding of the relation between the inferences performed under *unstructured* and *structured* models. We first focus on scenarios simulated under the SST, with parameters close to those estimated in real data. The first and most striking result is that we systematically observed a recent bottleneck under all simulated scenarios (Table 2, Figure 4, 5, 6, S1, and S2). This result could seem at a first glance surprising and due to an artefact. However, this is not the case, as: i) the signal does not depend on the inferential algorithm chosen to analyse the data (i.e., the *stairwayplot*), since the normalized spectra showed a deficit in singletons compared to the other low frequency classes (Figure 4, 5, 6 and S1), which is typical of a recent population decline; ii) it is consistent with the distribution of the Inverse Instantaneous Coalescent Rate (IICR) computed in one diploid individual, which shows a signature of decline under similar meta-population models (Chikhi et al., 2018; O. Mazet, Rodríguez, Grusea, Boitard, & Chikhi, 2016; Rodríguez et al., 2018). The results of our simulations are consistent with the recent bottleneck observed in the three shark species (Figure 3), with its intensity inversely correlated to the estimated *Nm* (i.e., stronger for *C. melanopterus* and *C. limbatus* than for *C. amblyrhynchos*). In our SST model there is an instantaneous colonization of the array of demes at *Tcol* , which corresponds also to a demographic expansion (i.e., the total number of individuals in the array of deme is larger than those in the ancestral deme). However, this signature is detected only for *Nm* [?]5 when *Tcol* is neither too recent nor too old (at equilibrium) (Figure 5, 6, S1 and S2). In these scenarios, the beginning of the expansion retrieved by the *stairwayplot* broadly corresponds to the simulated *Tcol* . This again corroborates the results obtained for the three shark species, since the two species with higher *Nm* displayed indeed an ancestral expansion in the *stairwayplot* with a timing consistent with the estimated *Tcol* (Table 1, Figure 2 and 3). Similarly, it explains why we could not retrieve the ancestral expansion for *C. melanopterus* nor estimate $Tcol$ under the SST model: this appears to be a property of the coalescent pattern and it is not related to the amount of data available (see below).

It is now straightforward to frame all these findings under the coalescent perspective. The coalescent history of the lineages sampled from a single deme in an SST (or FIM) model can be separated for simplicity into three phases: the *scattering* , the*collecting* and the *ancestral* phase. Going backward in time, lineages will coalesce in the sampled deme with a rate according to both Nm and N until all lineages either have coalesced or migrated to another deme. This is the *scattering* phase described in the seminal works of (Wakeley, 1998, 1999). The*scattering* phase was considered instantaneous for mathematical tractability, with its outcome dependent on Nm only, but later works could disentangle the effect of N and m on the shape of the gene genealogy (Mona, 2017). The *collecting* phase starts when the lineages which did not coalesce have migrated to other demes of the array: they will then coalesce according to a Kingman process with a rate scaled by Nm and the number of demes of the array (Wakeley, 1999). Finally, all surviving lineages (in non-equilibrium model) will reach the ancestral deme at $Tcol$, where they will coalesce at a rate depending only on the $Nanc$ parameter. In species with lowNm , the rate of coalescent during the *scattering* phase is very fast since lineages have low probability of emigrating from the sampled deme and high probability of coalescence due to the small$N$. Once all the lineages are dispersed in the array of demes, there will be two possible outcomes: i) in equilibrium model, we shift to the *collecting* phase, where the rate of coalescent drops since lineages will hardly fall in the same deme again; ii) in non-equilibrium model, with the parameters we have simulated here, there will be very few (if any) coalescent events during the collecting phase and the transition will be directly from the *scattering* to the*ancestral* phase. Both the *collecting* and the*ancestral* phases have a rate of coalescent lower than the*scattering* phase, which determines the observed recent drop inNe for all simulated scenarios. Remarkably, the decline inNe is much stronger in equilibrium model, since the rate of coalescent is much lower in the *collecting* than in the*ancestral* phase (Figure 4, 5, 6, S1, and S2). Low Nmspecies will therefore have only two coalescent phases, the*scattering* and either the *collecting* (in equilibrium model) or the *ancestral* (in non-equilibrium model) which is why the signature of the ancestral expansion is lost. For growing Nm , in equilibrium model there will be again only two coalescent phases, namely the *scattering* and *collecting* , with the latter having lower rate of coalescent than the former independently of the simulated parameters. This is why we observed always a strong bottleneck consistent with the distribution of the IICR statistics in any equilibrium model (Chikhi et al., 2018; Olivier Mazet et al., 2015; Rodriguez et al., 2018). In non-equilibrium model, there will be two different situations: a) $Tcol$ (in generations) is of the same order of the deme size N . In this setting, going backward in time few lineages would have escaped the sampled demes before $Tcol$. This corresponds to a shift in the coalescent rate directly from the*scattering* to the *ancestral* phase, resulting in a bottleneck of lower intensity compared to an equilibrium model (Figure 4, 5, 6 and S1), for the same reasons as above; b) $Tcol$ (in generations) is larger than N . In this setting, some coalescent events may occur during the *collecting* phase, at a rate much slower than the two other phases. This determines the hump observed in the *stairwayplot* (Figure 4, 5, 6) and explains why in this window of parameters it is also possible to correctly estimate $Tcol$using our ABC framework. Further simulations under the FIM model confirmed those patterns even though the ancestral expansion could be detected for lower long-term Nm than the corresponding SST scenario (Figure S3). This is probably due to a higher apparent connectivity underlined the by FIM, where lineages can move more freely during the collecting phase in comparison to SST where migrants only come from the closest neighbours. If many coalescent events occur during the *collecting* phase, the change in coalescent rate will affect the resulting gene genealogy and it will be detected by the*stairwayplot* (or any other *unstructured* method based on coalescent theory).

Using coalescent arguments, we clarified why simple meta-population models with constant connectivity generate a gene genealogy harbouring a signature of a recent decline for any parameters' combination. The signature of bottleneck detected by the *stairwayplot* in the three shark species best described by SST can be therefore interpreted as a consequence of the underlying structure. However, connectivity likely changes through time. For instance, human activities have likely impacted the evolutionary history of a large number of species either by decreasing their effective population size and/or by fragmenting their habitat (i.e., reducing migration rates between demes). This intuitively should exacerbate the signature of population decline in the resulting gene genealogy. However, it remains to be shown whether this signature is qualitatively and quantitatively distinguishable from models with constant connectivity. This is a question of fundamental importance to understand whether it is possible to detect recent bottleneck in structured populations. To this end, we further investigated by coalescent simulations the expected gene genealogy in SST-CH (and FIM-CH) models with a change in connectivity 10 or 50 generations B.P., which matches the beginning of extensive anthropogenic influence on biodiversity considering our species' generation time (Ceballos et al., 2015). The resulting gene genealogies were poorly affected by the recent drop in connectivity, with both the normalized SFS and the inferred Ne dynamic following the same trajectory of the corresponding scenario with the same long-term Nm and $Tcol$ (Figures 7, 8, S5, S6, S7, S8, S9 and S10). We noticed the drop in N (Figure 8, S6, S9 and S10) had stronger influence than the drop in m(Figure 7, S5, S7 and S8), consistent with previous finding showing that the distribution of coalescent events depends not only by the Nmcompound parameter but also by their individuals values (Mona, 2017). These results imply that the simulated change in connectivity is too recent to significantly alter the pattern of coalescent events during the *scattering* phase and that a recent drop can be hardly detected on the basis of the SFS only. Our empirical data are consistent with these findings: when we compared SST vs. SST-CH models in the three shark species using the ABC framework, we failed to clearly distinguish the two models (Table S5, Table S6, Figure S4). This seems a paradox: we observed a recent bottleneck in species of conservation concern using*unstructured* model, but we cannot exclude that this is just the consequence of population structure.

This study highlight once more the importance to explicitly test for meta-population structure before interpreting the demographic signals detected by *unstructured* models, similarly to what advocated previously by (Maisano Delser et al., 2019; Rodriguez et al., 2018). If the meta-population structure hypothesis is rejected, the variation ofNe through time can be directly interpreted as the demographic history of the population under investigation, such as the case of tiger shark. Otherwise, this variation is still related to demographic events, but it has to be explained in the light of population structure and its consequence on the rate of coalescent events. We showed by coalescent simulations how to interpret such variation: the recent bottleneck detected by the *stairwayplot* in demes belonging to a meta-population is a consequence of the coalescent process. In other words, any inferential method implementing an *unstructured* model will detect such decline (if enough data is available) since it is a property of the gene genealogy. Importantly, the gene genealogy is only slightly affected by recent changes in connectivity if the time of this change in generations is of the same order of the size of the deme.

Our study highlights a key issue in conservation genetics as a recent decline inferred by an *unstructured* model can be mis-interpreted as a consequence of recent anthropic pressures (Ceballos et al., 2015) when it actually results from meta-population structure. This is all the more alarming since the majority of species is likely organised in meta-populations across their range rather than panmictic at a large scale. We therefore stress the necessity for an educated choice of tools to correctly uncover the recent trend of a species and design proper conservation programs. For instance, detecting a recent bottleneck in meta-populations will require summary statistics measuring the linkage disequilibrium (Boitard, Rodriguez, Jay, Mona, & Austerlitz, 2016; Kerdoncuff, Lambert, & Achaz, 2020) and/or the inferential framework based on the IICR (Chikhi et al., 2018; Rodriguez et al., 2018) coupled with whole genome data. On a positive note, we showed that the colonization time of the array of demes can be estimated to some extent (and under some combinations of parameters) by *unstructured*models. We believe that this is particularly important because it has been shown that the simple instantaneous colonization process we used here behaves similarly to a spatial explicit range expansion (Hamilton, Stoneking, & Excoffier, 2005; Mona, 2017), which is certainly a more realistic model but more difficult to investigate. We are aware that the meta-population models here tested are simple and the parameters chosen are specific of the three shark species we focused on. Nevertheless, the time-scale separation of the coalescent process is general, and it allows explaining intuitively any structured models. The four shark species here used as an example has the merit to cover a large spectrum of LHT and consequently a large spectrum of demographic scenarios, going from a highly structured to a panmictic population: this has strong implications on the distribution of coalescent times and therefore on the interpretation of the observed data.

In this study we found that population structure, independently from the degree of connectivity between demes and the migration matrix relating them, intrinsically determines a variation in the rate of coalescent events through time. We showed that the intensity and the direction(s) of such variation related to the demographic parameters of the meta-population in a predictable way. Our results highlight the importance of detecting population structure (which depends on LHT among other factors) before performing any demographic inferences but, at the same time, they reveal the utility of *unstructured* models to describe the shape of the gene genealogy, which is the final product of the evolutionary history of a species. A combination of structured and*unstructured* models is therefore the key to best characterize the evolutionary history of a species. We call for a change in perspective when investigating the demographic history of a species: the focus should be put in the reconstruction of the variation of both $N$and m through time, which requires certainly new methodological development and probably more data.

Acknowledgement

We are grateful to the Genotoul bioinformatics platform Toulouse Midi-Pyrenees (Bioinfo Genotoul; http://bioinfo.genotoul.fr/) for providing computing resources. We are indebted to Oscar Lao for fruitful discussions and careful reading of the manuscript. This work was supported by two ATM grants (2016 and 2017) from the Museum National d'Histoire Naturelle to S.M.

References

Arenas, M., Ray, N., Currat, M., & Excoffier, L. (2012). Consequences of range contractions and range shifts on molecular diversity.*Molecular Biology and Evolution* , *29* (1), 207–218. doi: 10.1093/molbev/msr187

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

Boitard, S., Rodriguez, 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

Ceballos, G., Ehrlich, P. R., Barnosky, A. D., Garcia, A., Pringle, R. M., & Palmer, T. M. (2015). Accelerated modern human-induced species losses: Entering the sixth mass extinction. *Science Advances* ,*1* (5), 9–13. doi: 10.1126/sciadv.1400253

Chikhi, L., Rodriguez, 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, L., Sousa, V. C., Luisi, P., Goossens, B., & Beaumont, M. A. (2010). The Confounding Effects of Population Structure, Genetic Diversity and the Sampling Scheme on the Detection and Quantification of Population Size Changes. *Genetics* , *186* (3), 983–995. doi: 10.1534/genetics.110.118661

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

Cortes, E. (2002). Incorporating uncertainty into demographic modeling: Application to shark populations and their conservation.*Conservation Biology* , *16* (4), 1048–1062. doi: 10.1046/j.1523-1739.2002.00423.x

Excoffier, L., Dupanloup, I., Huerta-Sanchez, E., Sousa, V. C., & Foll, M. (2013). Robust Demographic Inference from Genomic and SNP Data.*PLoS Genetics* , *9* (10), e1003905. doi: 10.1371/journal.pgen.1003905

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

Hamilton, G., Stoneking, M., & Excoffier, L. (2005). Molecular analysis reveals tighter social regulation of immigration in patrilocal populations than in matrilocal populations. *Proceedings of the National Academy of Sciences of the United States of America* ,*102* (21), 7476–7480. doi: 10.1073/pnas.0409253102

Karl, S. A., Motta, P. J., Stewart, B. S., Wilson, S. G., Bowen, B. W., Castro, A. L. F., … Hueter, R. E. (2010). Population genetic structure of Earth's largest fish, the whale shark (Rhincodon typus).*Molecular Ecology* , *16* (24), 5183–5192. doi: 10.1111/j.1365-294x.2007.03597.x

Kerdoncuff, E., Lambert, A., & Achaz, G. (2020). Testing for population decline using maximal linkage disequilibrium blocks. *Theoretical Population Biology* , *134* (xxxx), 171–181. doi: 10.1016/j.tpb.2020.03.004

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

Lapierre, M., Lambert, A., & Achaz, G. (2017). Accuracy of Demographic Inferences from the Site Frequency Spectrum: The Case of the Yoruba Population. *Genetics* , *206* (1), 439–449. doi: 10.1534/genetics.116.192708

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

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

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

Mazet, O., Rodriguez, 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, Rodriguez, 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

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

Paris, J. R., Stevens, J. R., & Catchen, J. M. (2017). Lost in parameter space: a road map for stacks. *Methods in Ecology and Evolution* , *8* (10), 1360–1373. doi: 10.1111/2041-210X.12775

Parsons, G. R. (1990). Metabolism and swimming efficiency of the bonnethead shark Sphyrna tiburo. *Marine Biology* , *104* (3), 363–367. doi: 10.1007/BF01314338

Peter, B. M., Wegmann, D., & Excoffier, L. (2010). Distinguishing between population bottleneck and population subdivision by a Bayesian model choice procedure. *Molecular Ecology* , *19* (21), 4648–4660. doi: 10.1111/j.1365-294X.2010.04783.x

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). doi: 10.1371/journal.pone.0037135

Pirog, A., Jaquemet, S., Ravigne, 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

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

Rochette, N. C., Rivera-Colon, 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

Rodriguez, 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

Stadler, T., Haubold, B., Merino, C., Stephan, W., & Pfaffelhuber, P. (2009). The impact of sampling schemes on the site frequency spectrum in nonequilibrium subdivided populations. *Genetics* , *182* (1), 205–216. doi: 10.1534/genetics.108.094904

Tajima, F. (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* , *123* (3), 585–595. doi: 10.1093/genetics/123.3.585

Wakeley, J. (1998). Segregating Sites in Wright's Island Model.*Theoretical Population Biology* , *53* , 166–174.

Wakeley, J. (1999). Nonequilibrium migration in human history.*Genetics* , *153* (4), 1863–1871.

Warmuth, V. M., & Ellegren, H. (2019). Genotype-free estimation of allele frequencies reduces bias and improves demographic inference from RADSeq data. *Molecular Ecology Resources* , *19* (3), 586–596. doi: 10.1111/1755-0998.12990

Watterson, G. A. A. (1975). On the number of segregating sites in genetical models without recombination. *Theoretical Population Biology* , *7* (2), 256–276. doi: 10.1016/0040-5809(75)90020-9

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.

Table 1. Summary statistics and ABC estimation. Number of loci and SNPs after filtering, mean pairwise difference (ϑ_{π}), Watterson theta (ϑ_{w}), Tajima's D (TD), posterior probability of the most supported model and its parameters (median value and 95% credible interval in parentheses).

	N° Loci	N° SNP	ϑ_{π}	ϑ_{w}	TD	Model (probability)[+]	*Nm*	*Tcol* [++]	*Na*
G. cuvier[§] (N=8)	117976	25785	0.00057	0.00051	-0.03	NS (0.84)	-	-	-
C. amblyrhynchos (N=12)	69490	68355	0.00216	0.00229	-0.23[*]	SST (0.85)	11.5 (3.0-22.0)	20456 (12567-75649)	409 (13 492
C. limbatus (N=13)	60812	43449	0.00180	0.00166	0.43[*]	SST (0.55)	6.6 (1.5-15.4)	50198 (475-245440)	255 (19 528
C. melanopterus[¶] (N=8)	926	784	0.00040	0.00030	0.691[*]	SST (0.89)	1.8 (0.7-3.0)	91719 (5000-291341)	340 (27 953
						Priors [a]	U: 0.001 - 100	U: 1 - 300000	U: – 100

[*] Tajima's D values are significant (p <0.001).

[+] Most supported model and its posterior probability.

[++] *Tcol* is expressed in generations.

[§] *G. cuvier* is best represented by the NS model: its demography is depicted through the *stairwayplot*algorithm (see discussion).

[¶] Data from Maisano Delser et al. (2019).

[a] Uniform prior distribution. The prior distribution of *Nm* is the product of two uniforms (one for *N* and one for *m*).

Table 2. Coalescent simulations of 50,000 Rad-loci under SST model, with mutation rate fixed to $1.93*10^{-8}$ per site per generation and *Nanc* fixed to 50,000. Mean pairwise difference (ϑ_{π}), Watterson theta (ϑ_{w}), Tajima's D (TD), and number of segregating sites (S) are averaged over 100 replicates.

Nm	Tcol	$\boldsymbol{\vartheta_{\pi}}$ $^{++}$	$\boldsymbol{\vartheta_{\varsigma}}$ $^{++}$	**TD**	**S**
1	5000	0.0013	0.0011	0.531	23599
	15000	0.0013	0.0012	0.405	24094
	50000	0.0017	0.0016	0.406	32201
	[?][+]	0.0161	0.0139	0.669	283564
5	5000	0.0017	0.0016	0.361	32443
	15000	0.0019	0.0018	0.191	37712
	50000	0.0028	0.0028	0.035	56474
	[?]	0.0177	0.0150	0.749	306786
10	5000	0.0019	0.0018	0.180	36561
	15000	0.0021	0.0022	-0.087	44380
	50000	0.0031	0.0034	-0.364	69436
	[?]	0.0180	0.0158	0.585	321619
15	5000	0.0019	0.0019	0.048	38919
	15000	0.0022	0.0024	-0.274	48479

Nm	Tcol	ϑ_π $^{++}$	ϑ_ς $^{++}$	TD	S
	50000	0.0032	0.0038	-0.608	77391
	[?]	0.0181	0.0163	0.465	331816

$^{+}$ Equilibrium model obtained by simulating $Tcol$ =[?].

$^{++}$ Theta values are expressed per site per generation.

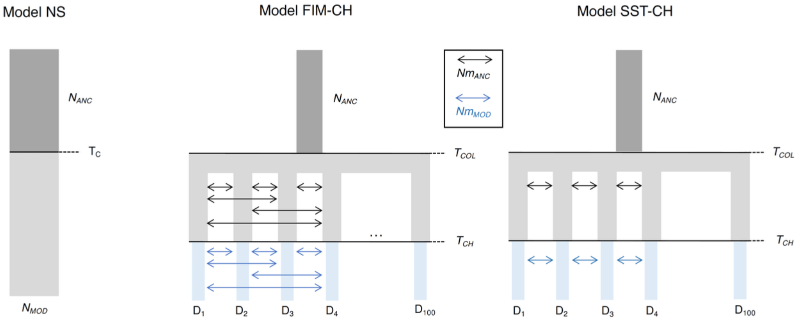

Figure 1. Evolutionary scenarios considered in this study (to both infer parameters in real data under an ABC framework and to perform coalescent simulations). SST (FIM) model is a simplified version of SST-CH (FIM-CH) in which connectivity Nm is constant after Tcol. Details on each parameter are presented in the main text.

Hosted file

`image2.emf` available at https://authorea.com/users/410702/articles/709538-coalescent-times-life-history-traits-and-conservation-concerns-an-example-from-four-shark-species-from-the-indo-pacific

Figure 2. Posterior distribution of the number of migrants per generation Nm (panel A) and of the colonisation time of the array of deme $Tcol$ (panel B) estimated under the stepping stone model (SST) for *Carcharhinus amblyrhynchos* (red), *Carcharhinus limbatus* (green) and *Carcharhinus melanopterus* (blue).

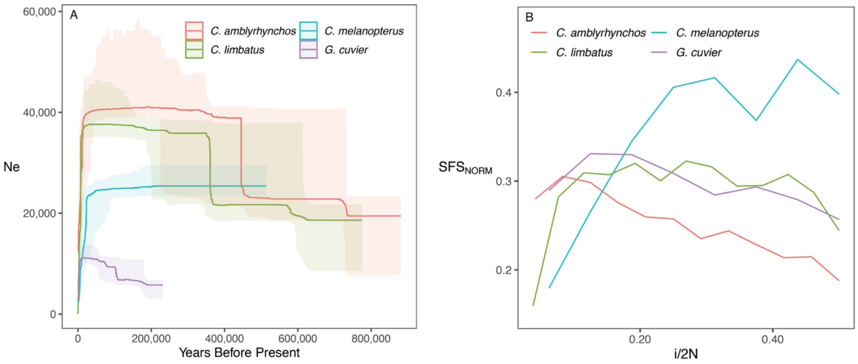

Figure 3. Panel A: variation of the effective population size (Ne) through time and its 75% confidence interval estimated by the *stairwayplot*. Panel B: normalized SFS computed as in (Lapierre et al., 2017). *Car-*

charhinus amblyrhynchos is represented in red, *Carcharhinus limbatus* in green,*Carcharhinus melanopterus* in blue, and *Galeocerdo cuvier*in purple.

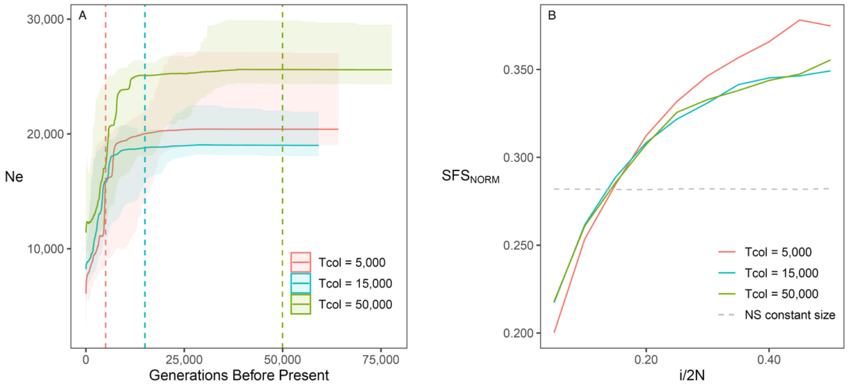

Figure 4. *stairwayplot* (maximum likelihood *Ne* and 75% confidence interval) (panel A) and normalized SFS (panel B) computed in simulated non-equilibrium SST scenarios with Nm=1, averaged over 100 replicates. Colonisation time of the array of deme *Tcol* occurred 5,000 (red), 15,000 (blue), and 50,000 (green) generations B.P., visually represented by the vertical dashed lines in panel A. The normalized SFS expected under a constant size non-structured model (NS constant size) is also shown (grey dashed line in panel B).

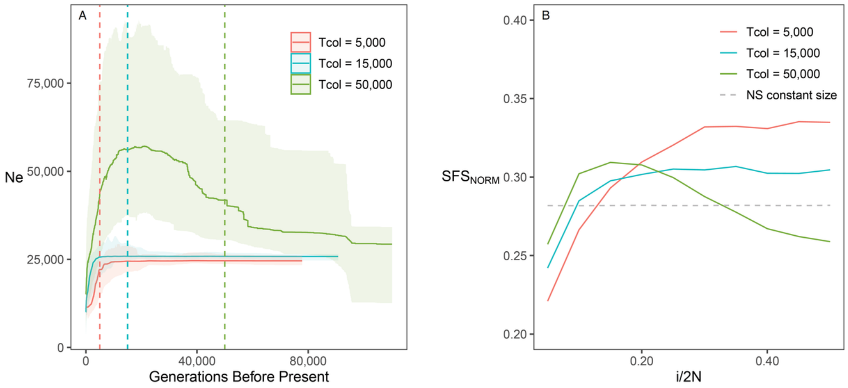

Figure 5. *stairwayplot* (maximum likelihood *Ne* and 75% confidence interval) (panel A) and normalized SFS (panel B) computed in simulated non-equilibrium SST scenarios with Nm=5, averaged over 100 replicates. Colonisation time of the array of deme *Tcol* occurred 5,000 (red), 15,000 (blue), and 50,000 (green) generations B.P., visually represented by the vertical dashed lines in panel A. The normalized SFS expected under a constant size non-structured model (NS constant size) is also shown (grey dashed line in panel B).

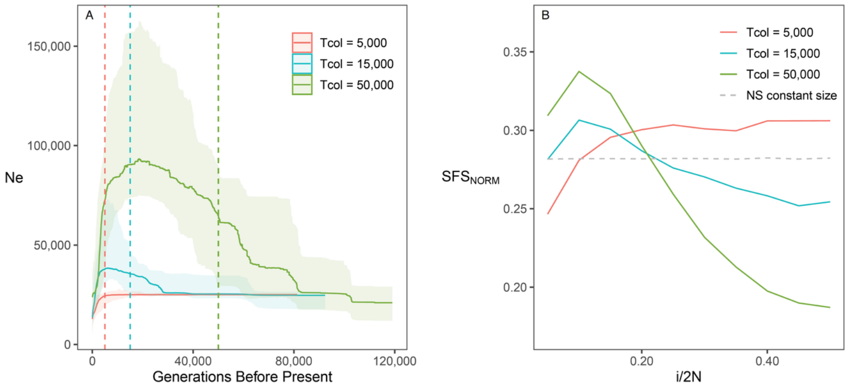

Figure 6. *stairwayplot* (maximum likelihood *Ne* and 75% confidence interval) (panel A) and normalized SFS (panel B) computed in simulated non-equilibrium SST scenarios with *Nm*=10, averaged over 100 replicates. Colonisation time of the array of deme *Tcol* occurred 5,000 (red), 15,000 (blue), and 50,000 (green) generations B.P., visually represented by the vertical dashed lines in panel A. The normalized SFS expected under a constant size non-structured model (NS constant size) is also shown (grey dashed line in panel B).

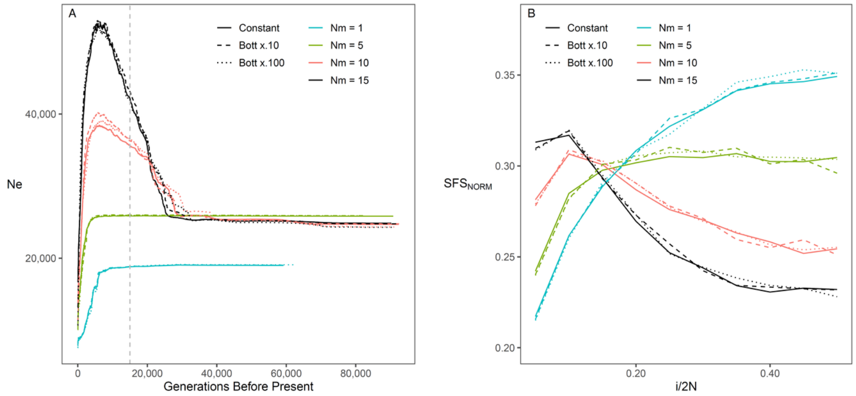

Figure 7. *stairwayplot* (maximum likelihood *Ne*) (panel A) and normalized SFS (panel B) computed in simulated non-equilibrium SST scenarios with *Tcol*=15,000 generations B.P. and an instantaneous decrease of the migration rate (*m*) forward in time at *Tch*=10 generations B.P. Colours represent the long-term connectivity values: *Nm*=1 (blue), *Nm*=5 (green),*Nm*=10 (red), *Nm*=15 (black). Line style represents the 10-fold (small dashes) or 100-fold (dots) reduction of *m*, or constant *Nm* (continuous line). The vertical grey dashed line in panel A represents the simulated colonisation time *Tcol.*

Hosted file

`image8.emf` available at https://authorea.com/users/410702/articles/709538-coalescent-times-life-history-traits-and-conservation-concerns-an-example-from-four-shark-species-from-the-indo-pacific

Figure 8. *stairwayplot* (maximum likelihood *Ne*) (panel A) and normalized SFS (panel B) computed in simulated non-equilibrium SST scenarios with *Tcol*=15,000 generations B.P. and an instantaneous decrease of the deme size (*N*) forward in time at *Tch*=10 generations B.P. Colours represent the long-term connectivity

values: Nm=1 (blue), Nm=5 (green),Nm=10 (red), Nm=15 (black). Line style represents the 10-fold (small dashes) or 100-fold (dots) reduction of N, or constant Nm (continuous line). The vertical grey dashed line in panel A represents the simulated colonisation time $Tcol$.