One panel to rule them all: DArTcap genotyping for population structure, historical demography, and kinship analyses, and its application to a threatened shark

Pierre Feutry[1], Floriaan Devloo-Delva[1,2]*, Adrien Tran Lu Y[3,4]*, Stefano Mona[3,4]*, Rasanthi M. Gunasekera[1], Grant Johnson[5], Richard D. Pillans[6], Damian Jaccoud[7], Andrzej Kilian[7], David L. Morgan[8], Thor Saunders[5], Nicholas J. Bax[1,9], Peter M. Kyne[10]

[1]CSIRO Oceans & Atmosphere, Castray Esplanade, Hobart, TAS 7000, Australia
[2]School of Natural Sciences – Quantitative Marine Science, University of Tasmania, Hobart, TAS 7000, Australia
[3]Institut de Systématique, Évolution, Biodiversité (ISYEB), UMR 7205, MNHN, CNRS, EPHE, Sorbonne Université, 75005 Paris, France
[4]EPHE, PSL University, 75005 Paris, France
[5]Department of Primary Industry and Fisheries, Aquatic Resource Research Unit, GPO Box 3000, Darwin , NT 0801, Australia
[6]CSIRO Oceans and Atmosphere, 41 Boggo Road, Dutton Park, QLD 4102, Australia
[7]Diversity Arrays Technology Pty Ltd, Building 3, University of Canberra, Bruce, ACT 2617, Australia
[8]Centre for Sustainable Aquatic Ecosystems, Harry Butler Institute, Murdoch University, Murdoch, WA 6150, Australia
[9]Institute for Marine and Antarctic Science, University of Tasmania, Hobart, TAS 7000
[10]Research Institute for the Environment and Livelihoods, Charles Darwin University, Ellengowan Drive, Darwin, NT 0909, Australia
* Authors contributed equally to the work

Correspondence
Pierre Feutry, CSIRO Oceans & Atmosphere, Castray Esplanade, Hobart, Australia
Email: Pierre.Feutry@csiro.au

Abstract

With recent advances in sequencing technology, genomic data are changing how important conservation management decisions are made. Applications such as Close-Kin Mark-Recapture demand large amounts of data to estimate population size and structure, and their full potential can only be realised through ongoing improvements in genotyping strategies. Here we introduce DArTcap, a cost-efficient method that combines DArTseq and sequence capture, and illustrate its use in a high resolution population analysis of *Glyphis garricki*, a rare, poorly known and threatened euryhaline shark. Clustering analyses and spatial distribution of kin pairs from four different regions across northern Australia and one in Papua New Guinea, representing its entire known range, revealed that each region hosts at least one distinct population. Further structuring is likely within Van Diemen Gulf, the region that included the most rivers sampled, suggesting additional population structuring would be found if other rivers were sampled. Coalescent analyses and spatially explicit modelling suggest that *G. garricki* experienced a recent range expansion during the opening of the Gulf of Carpentaria following the conclusion of the Last Glacial Maximum. The low migration rates between neighbouring populations of a species that is found only in restricted coastal and riverine habitats show the importance of managing each population separately, including careful monitoring of local and remote anthropogenic activities that may affect their environments. Overall we demonstrated how a carefully chosen SNP panel combined with DArTcap can provide highly accurate kinship inference and also support population structure and historical demography analyses, therefore maximising cost-effectiveness.

KEYWORDS

Glyphis garricki, connectivity, coalescent simulations, RAD, sequence capture, Close-Kin Mark-Recapture

Introduction

Genomic data is changing how wildlife conservation decisions are made (Ovenden et al., 2019). It is now used for species identification or population assignment (Bekkevold et al., 2015; Grewe et al., 2015), inferring sex specific connectivity at both evolutionary and contemporary timescales (Feldheim et al., 2014; Feutry et al., 2017), and for use in kin relationships to estimate population size (Ackerman et al., 2017; Bravington et al., 2016a; Hillary et al., 2018). Genomic data can be mined for sex-linked markers (Anderson et al., 2012), while molecular markers can be used for evolutionary studies investigating the past demography of species (Maisano Delser et al., 2016). Reduced genotyping costs now support the use of large sample sizes and more accurate effective population size (Ne) estimates (Waples et al., 2018).

Sequencing a set portion of the genome is increasingly used in genomic studies, since sequencing the entire genome of many individuals remains cost-prohibitive for most species. Restriction enzyme-based complexity reduction methods and sequence capture are two of the most commonly used methods for the study of non-model organisms (Jones & Good, 2016). Implementation of restriction enzyme-based complexity reduction methods, such as DArT/DArTseq (Jaccoud et al., 2001; Kilian et al., 2012) and restriction site-associated DNA (RAD), allow discovery and genotyping single nucleotide polymorphisms (SNPs) in a single step. These methods are usually much cheaper than whole genome sequencing approaches but a great deal of the sequencing effort is still lost to non-variable regions of the genome and low quality or uninformative SNPs.

Sequence capture is more specific to a particular region of interest than enzyme-based complexity reduction methods, but it suffers from relatively high library preparation costs prior to capture, and low-multiplexing capacity. Recently, Ali et al. (2016) developed RAD capture (Rapture), a combination of RAD and sequence capture techniques. This method is a rapid, flexible and cost effective library preparation from RAD sequencing (RADseq) and includes the ability to restrict sequencing to genomic regions of interest from sequence capture, greatly reducing genotyping costs. Rapture was quickly followed by RADcap (Hoffberg et al., 2016), a variant of Rapture using the 3RAD protocol instead of RAD and which allows the detection of PCR duplicates and reduces the amount of missing data. In this study, we introduce DArTcap, a method based on the same principles as Rapture and RADcap but combining the cost-effective and consistent DArTseq protocol and sequence capture to produce affordable and high-throughput genetic profiles, with negligible amounts of missing data. While the

concept of DArTcap and its methodology was first optimised in 2015 and has been deployed commercially for several years, especially in agriculture, there are still no peer reviewed reports on its use.

Multi-purpose low-cost large SNP panels that enable accurate kinship inference, population structure and demography have been identified as a high priority for biodiversity conservation and management (Ackerman et al., 2017; Hess et al., 2015). Close-Kin Mark-Recapture (CKMR) (Bravington et al., 2016b) is a rapidly expanding procedure for robust population size estimation that has already changed how valuable commercial fish species and rare threatened species are managed and monitored (Bravington et al., 2016a; Hillary et al., 2018). Importantly for this paper, CKMR requires very reliable kinship inference because its success relies on finding a few dozen true kin pairs in large samples (i.e. thousands to millions of comparisons).

In addition to estimates of population size, CKMR can identify population boundaries, which is fundamental to delineating the spatial scale of units for effective conservation and management of threatened species (Feutry et al., 2017; Feutry et al., 2014). In contrast to population genetics, CKMR can provide a direct estimate of connectivity over short timescales (one or two generations), as opposed to long timescales (hundreds or thousands or generations) with population genetics. A threatening process or pressure acting on an isolated part of a species' range has an increased probability of causing local extinction, because there is no buffering by immigration. Estuarine-associated fishes have been shown to have a high incidence of genetic subdivision, and genetic structuring can be especially complex in euryhaline species (Bilton et al., 2002; Feutry et al., 2017; Feutry et al., 2014; Lavergne et al., 2014; Phillips et al., 2011; Watts & Johnson, 2004) which have the ability to move across the fresh-brackish-salt water interface. These species, unlike strictly estuarine species, have the potential to disperse broadly in the marine corridor, but this may also vary between sexes (Feutry et al., 2017), resulting in sex-specific impacts from multiple stressors across distinctly different habitats.

Present population boundaries are the result of demographic and selective processes which have interacted with a species during its evolutionary history. Management needs to operate on these current boundaries and, in particular, boundaries that affect breeding. Modern population genetics can be used to identify migration patterns among sampled and unsampled populations and the change in genetic diversity through recent or more distant generations. It can also be used to reconstruct crucial properties of conservation genetics, such as species origin and any shifts or contractions of range through time. Effective conservation policies depend on an

123 accurate knowledge of current and historical processes, since genetic diversity alone, or
124 changes in effective population size (*Ne*), can be misleading if estimated locally while failing
125 to account for variation at a larger geographic scale using adequate population genetics
126 modelling (Maisano Delser et al., 2018). Integrating classical population genetics and CKMR
127 now provides an accepted approach to clearly identify current and distant properties of a
128 species' biogeographic history (Lowe et al., 2017).

129 Sharks and rays (subclass Elasmobranchii) are a group of global conservation concern, with a
130 quarter of all species estimated to be at risk of extinction (Dulvy et al., 2014). The small
131 proportion (~5%) of this group that lives in freshwater or euryhaline environments is often at
132 an elevated risk of population depletion from overfishing or habitat loss and degradation due
133 to their restricted distributions, intrinsic biological vulnerability, and the escalating intensity of
134 pressures on their aquatic habitats, including climate change (Chin et al., 2010; Dulvy et al.,
135 2014; Lucifora et al., 2015). Despite this elevated level of risk, population boundaries remain
136 undefined for most species, compromising their conservation and management.

137 The river sharks (*Glyphis* spp.) are highly threatened euryhaline sharks of the Indo-West
138 Pacific, characterised by taxonomic uncertainty, poorly-defined distributions, and a lack of
139 ecological data (Li et al., 2015). One species, the Ganges Shark (*Glyphis gangeticus*) faces
140 immense human pressure in Southeast Asia and the Arabian Sea, with only rare contemporary
141 records (Jabado et al., 2018; Li et al., 2015). In contrast, two species, the Speartooth Shark
142 (*Glyphis glyphis*) and the Northern River Shark (*Glyphis garricki*), occur in relatively
143 undisturbed environments of northern Australia where low human population size and the
144 remoteness of the landscape have limited development pressure, and many estuaries are in
145 near-pristine conditions (Pillans et al., 2010; Woinarski et al., 2007). These two species provide
146 ideal case studies in understanding how genomic data can support conservation of rare and
147 threatened aquatic species.

148 Population boundaries have been identified in *G. glyphis* across its limited northern Australian
149 estuarine/riverine range (Feutry et al., 2017; Feutry et al., 2014). Structuring was evident
150 between the three river systems where the species was known to occur, supporting the
151 designation of each river drainage as a discrete management unit (Feutry et al., 2014). It may
152 be hypothesised that its congener, *G. garricki*, a restricted range euryhaline shark found only
153 in northern Australia and Papua New Guinea (Pillans et al., 2010; White et al., 2015), would
154 show similar levels of population structuring. This species' habitat is primarily large tropical
155 rivers and estuaries where it occurs in tidal reaches; however, there are also coastal records

(Pillans et al., 2010), suggesting some level of marine dispersal from estuarine and riverine environments. *Glyphis garricki* is listed as Endangered on Australia's national environmental legislation and is subject to a multi-species Recovery Plan which emphasizes the need to understand population connectivity and population size for its effective management (DoE, 2015). Recent surveys across northern Australia have revealed that *G. garricki* occurs in a wider number of estuarine and river systems than previously documented (e.g. in Pillans et al., 2010). Juveniles in particular are regularly encountered in northern Australian rivers including those flowing into Van Diemen Gulf (Kyne, 2014) and the western Northern Territory, and the Kimberley region of Western Australia (P.M. Kyne et al., unpublished data). This improved understanding of the occurrence and habitat of *G. garricki* has allowed adequate samples to be collected to examine population connectivity.

In this study we demonstrate how a carefully selected DArTcap SNP panel allows for cost-effective CKMR grade kinship inference and other population analyses. We illustrate the benefits of this approach with a practical example, developing robust population structure and historical demographic analyses for *G. garricki*, that provides important information for the conservation and management of this species.

Material and methods

Sample collection and DNA extraction

A total of 468 *G. garricki* were collected and genotyped from 11 rivers, large marine embayments or estuaries (thereafter referred to as sampling locations) in 5 different regions covering the entire known geographic range (Fig. 1) (Pillans et al., 2010; White et al., 2015; P.M. Kyne et al. unpublished data) between 2012 and 2016. One sample was identified genetically as a recapture and two individuals of the sister species *G. glyphis* were also genotyped to allow the polarisation of markers in historical demographic analyses, for a total of 469 unique genotypes. Each shark from Australia was measured, sexed, and sampled for genetic material before it was released at the site of capture. Total lengths (TL) of all sharks ranged from 52 to 182 cm; most sharks were juveniles or sub-adults, with 26 males >141 cm TL assessed as being sexually mature (possessing calcified claspers). Sexual maturity in female sharks cannot be assessed externally in sharks, but 10 females >142 cm TL were assumed to be mature based on the established male size-at-maturity (Supplemental Information S1 section 2).

Sharks were sampled under Northern Territory Fisheries Special Permits S17/3252 and S17/3364, Kakadu National Park Research Permit RK805, Western Australian Department of Fisheries Exemption No. 2630, Western Australian Department of Parks and Wildlife Permit SF010485, and Charles Darwin University Animal Ethics Committee Approval A11041. Samples from Papua New Guinea (PNG) were obtained through artisanal fisheries (White et al., 2015). DNA was extracted with the DNeasy Blood and Tissue kits (Qiagen) following standard protocol.

SNP selection for bait design

In order to minimise ascertainment bias, samples from 93 sharks, with a minimum of eight individuals from each of the 11 sampling locations, were included in the SNP discovery phase. Samples were genotyped using DArTseq™ as described by (Feutry et al., 2017; Grewe et al., 2015). DArTseq™ is a combination of complexity reduction methods and next generation sequencing platforms with each complexity reduction method tailored to the organism under study. In the absence of a reference genome a de novo approach was used for SNP calling with DArTsof14. For *G. garricki*, the first set of restriction enzymes (PstI-SphI) did not yield enough SNPs and it was therefore necessary to apply a second set (PstI-NspI) in order to reach a target of ~2,000 high-quality markers that we had previously estimated necessary to accurately resolve the second order relationships through simulations (data not shown). The selection process for the DArTcap SNP panel was as follows: i) SNP counts were normalised for each individual and SNPs with normalised counts below 6 and above 80 were discarded; ii) SNPs with more than 5% ambiguous genotypes were also discarded (genotypes were considered ambiguous if the count proportion of one of the alleles fell between 0 and 0.1); iii) SNPs with a call rate (i.e. proportion of individuals scored) <0.7 were discarded; iv) SNPs with minor allele frequency (MAF) <0.01 were discarded; and finally, v) a chi-square test to detect deviation from Hardy-Weindberg equilibrium (HWE) was performed and SNPs with *p*-values <0.05 discarded (the thresholds for these filters were defined after plotting the data, Supplemental Information S1, section 4.1). From the remaining set of 2,094 SNPs, we selected 2,007 for the DArTcap panel based on their power to best resolve kinship by calculating a pseudo-likelihood (PLOD) score (Hillary et al., 2018).

In order to assess if the HWE filter would have much impact on the DArTcap panel's ability to detect population structure, we calculated pairwise fixation indices by sampling location and carried out a Discriminant Analysis of Principal Components (DAPC) analysis with K-means clustering based on the SNPs discarded by that filter with similar parameters to those of

DArTcap dataset described below. The results were consistent with those obtained from the SNPs that were retained (Supplemental Information S1, sections 4.2 and 4.3).

DArTcap genotyping

DArTcap involves adding a hybridisation-based enrichment step before the DArTseq libraries are sequenced. The hybridisation step uses custom synthesised biotinylated RNA MYbaits (Arbor Bioscience) designed based on the chosen DArTseq markers. The 2,007 DArTSeq markers short-listed for the DArTcap panel were put through a selection process using a proprietary algorithm based on assessing sequence length and complexity in order to limit non-specific capture. Markers with known sequences shorter than 40bp were removed, as were those with low complexity. Sequence complexity was assessed by calculating a score based on median levels of GC, number of sequence variants at the locus and length of homo-polymers. This reduced the number to 1709 sequences to use for the enrichment and one bait was designed based on the sequence of the most common allele. DArTcap hybridisation and washing used the protocols based on Version 3 of the MYBaits manual (https://arborbiosci.com/wp-content/uploads/2017/10/MYbaits-manual-v3.pdf). These DArTcap enriched libraries, one per sample, were sequenced on a HiSeq 2500 (Illumina). The instrument was setup for 1x77 bp per run and the libraries were spread over 1.8 lanes of Hiseq flowcells, giving approximately 270 million clusters worth of sequence data for the samples used in the study.

SNP and individual filtering

The DArTcap baits, as with any other enrichment system, are not 100% specific to the SNP-bearing restriction fragment targeted (i.e new regions of the genome are captured) and some loci that passed QC with DArTseq may be troublesome when genotyped with DArTcap. Consequently, we added several filtering steps prior to doing any analysis (Supplemental Information S2.1).

Filtering for population structure and kinship analyses was as follows. In order to avoid short distance linkage disequilibrium when multiple SNPs were present on the same 75bp fragment, we only retained the one with the highest polymorphism information content (PIC), which measures the probability of identifying which of the two alleles at a single locus is transmitted from a parent to an offspring (Botstein et al., 1980). SNPs and sharks were sequentially filtered by increasing the proportion of missing data over 100 iterations to a maximum threshold of 0.85 for both the SNPs and sharks. After removing individuals, monomorphic loci were

252 removed. Loci with low reproducibility (<0.98; the proportion of identical genotypes
253 calculated from technical replicates routinely included by DArT P/L for routine quality control)
254 and low sequencing coverage (<10), subject to higher genotyping error rates, were removed.
255 Poorly informative loci with low MAF (<0.01) were also removed. SNPs with high counts
256 (>300) and high heterozygosity (>0.6) were deleted to account for potential paralogous loci
257 (error during sequence clustering due to high sequence similarity). Again, the thresholds for
258 these filters were defined after plotting the data (Supplemental Information S1, section 5).
259 Finally, the R packages Radiator (Gosselin et al., 2020) and OutFLANK (Whitlock &
260 Lotterhos, 2015) were used to detect and remove sex-linked and outlier loci respectively. Since
261 unequal sample sizes could introduce bias in clustering algorithms (Foster et al., 2018), filtering
262 was applied to the full dataset and a subsampled dataset including only 30 randomly chosen
263 sharks from Van Diemen Gulf (VDG) was selected so that putative populations would roughly
264 be of equal size (Supplemental Information S1, section 6).

265 The SNP and individual filtering for detecting the origin of the range expansion and the
266 historical demographic analyses were similar to the one used for population structure analyses
267 except for the following steps. All loci with more than 3 SNPs, which are more likely to be
268 paralogous, were removed and the reproducibility threshold was set at 0.99. SNPs and
269 individuals with more than 15% of missing data were also eliminated. Rare variants are
270 extremely important for demographic inferences, both to detect variation in *Ne* and to trace the
271 spread of the species spatially, hence the increased filtering stringency. Finally, when
272 computing the unfolded site frequency spectrum (SFS) and further downstream analyses we
273 removed all SNPs with missing data (reproducing a pattern of missing data during the
274 modelling process would be difficult if not impossible) and SNPs heterozygous in the sister
275 species *G. glyphis* for which polarization would have been uncertain (Supplemental
276 Information S2.1).

277 *Kinship inference*

278 Kinship inference was conducted using a log-likelihood-ratio (LLR) approach developed by
279 (Bravington et al., 2016b) and applied previously to sharks by (Hillary et al., 2018). This
280 approach has the advantage of offering statistical error control to minimize false positives while
281 controlling for false negatives, which is critical when the number of comparisons is large and
282 the goal is to estimate connectivity and/or abundance. This method, while suitable for large
283 sample sizes, relies on accurate estimation of allele frequencies. Given the relatively low

sample size in all populations but the VDG, we restricted our search of kin pairs to the VDG population.

Population structure analyses

Allelic richness (*Ar*), observed heterozygosity (*Ho*) and expected heterozygosity (*He*) were calculated for each sampling location using the R package diveRsity v1.9.90, while inbreeding coefficients (F_{IS}) were calculated with hierfstat v0.04-22. Pairwise fixation indices (Weir & Cockerham, 1984) were calculated on the full dataset for each sampling location separately using the R package *stAMPP v1.5.1* with 10,000 bootstraps (Pembleton et al., 2013). Next, model-based and dimensionality-reduction clustering analyses were conducted on the full and subsampled datasets with *ADMIXTURE v1.3* (Alexander & Lange, 2011) software and *Adegenet v2.1.1* (Jombart & Ahmed, 2011). A hierarchical approach was used for the clustering methods (Vähä et al., 2007). A first round of *ADMIXTURE* and DAPC analyses was carried out on the full and subsampled datasets followed by a second round on the groups identified in the first round that included more than one sampling location and at least 10 samples in each of these (i.e. Cambridge Gulf (CG) and VDG). *ADMIXTURE* was used to investigate the genetic ancestry of each individual. The algorithm was run for K = 1-8 with a 100-fold cross-validation and 20,000 bootstraps. The dimensionality-reduction clustering was performed with the DAPC; (Jombart et al., 2010)). Initially, individuals were grouped according to the data itself, using the successive K-means algorithm implemented in the *find.clusters()* function. The goodness of fit, determined by the Bayesian information criterion (BIC), was employed to find the best number of clusters (K). In order to avoid over-fitting, the optimal number of principal components was selected through cross-validation with a 10% hold-out set and 100 replicates for all DAPC analyses.

Detecting the origin of the range expansion

Significant positive correlation between genetic and geographic distance is indicative of isolation by distance (IBD), but can also be the result of an equilibrium stepping stone model. Range expansions generate IBD, but they also leave characteristic footprints in patterns of genetic diversity within species. Theoretical predictions can be used to both test for the occurrence of a range expansion and to estimate its centre of origin. Indeed, shared derived alleles are expected to be at low frequency near the centre of origin of the expansion but reach higher frequencies in demes with increasing geographic distance from the origin due to serial founder effects (Slatkin & Excoffier, 2012). The directionality index, Ψ, is the average

316 difference in the shared derived allele frequency between two populations (computed only on
317 alleles where the ancestral copy is not fixed in either of the two populations), and is expected
318 to be around 0 in an equilibrium stepping stone model but significantly different from 0 in a
319 range expansion model (Peter & Slatkin, 2013). We polarized SNPs to detect the ancestral
320 variant using *G. glyphis* as the outgroup and further computed the matrix of the pairwise Ψ,
321 testing for its significance using a permutation approach (i.e., whether Ψ is significantly
322 different from 0). Finally, the origin of the expansion was identified using the Time Difference
323 of Arrival (TDOA) algorithm (Gustafsson & Gunnarsson, 2003) as implemented in the
324 *rangeExpansion* library (Peter & Slatkin, 2013) in the R environment.

325 *Historical demographic inferences: unstructured models*

326 We first investigated the demographic history of each population by considering them as fully
327 isolated (unstructured demographic model). We used the software *stairwayplot v2.0* (Liu and
328 Fu, 2015) which investigates the *Ne* and its changes through time using a composite likelihood
329 approach to find the values that best reproduce the observed SFS. The *stairwayplot* was run on
330 the unfolded SFS with a mutation rate per base per generation of $7*10^{-8}$ and a 7 year generation
331 time (following the genomic estimates obtained for *Carcharhinus melanopterus* (Maisano
332 Delser et al., 2016). The rational for these choices is: i) *C. melanopterus* has a similar size to
333 *G. garricki*, which is generally a good predictor of the molecular clock; ii) both species are
334 structured over restricted areas, iii) Maisano Delser et al. (2016) used exon capture which is
335 more biased towards slower evolving genes than DArTseq and DArTcap (S. Mona,
336 Unpublished data). The *stairwayplot* fits *n-1 Ne* parameters, with *n* being the number of
337 sampled chromosomes. When *n* is large (and so are the classes of the SFS), it may be difficult
338 to correctly fit the demography if the number of SNP is relatively low. We subsampled the
339 VDG population down to 30 individuals to resolve this issue (in this case, having too many *Ne*
340 parameters and too few SNPs). To this end, we randomly sampled 30 individuals 10 times and
341 computed the average *stairwayplot* and its confidence interval over the 10 runs. Finally, we
342 analysed a *scatter sample* sensu Wakeley (1999) by randomly pooling one individual per
343 population. We repeated the process 100 times and computed an average *stairwayplot* and its
344 confidence interval.

345 *Historical demographic inferences: structured models*

346 We developed an Approximate Bayesian Computation (ABC) approach to compare four
347 structured non-equilibrium demographic models (Fig. 2). The four models were devised to

answer two specific questions: a) estimating the migration matrix connecting the five populations as defined by population structure analyses (which corresponded to the five regions sampled, see the population structure results section) and its changes through time; b) estimating the time of colonization of the habitat. First, all models are characterized by an ancestral population (with an effective size N_{anc}) splitting into the five modern populations at time T_i. This corresponds to an instantaneous colonization of the habitat from *G. garricki*. After the split, all populations exchange migrants (Nm) following a linear stepping stone (LSS_1 and LSS_2) or an island matrix (FIM_1 and FIM_2). The parameters of the migration matrix are constant through time after the split (LSS_1 and FIM_1) or they change instantaneously at T_m (LSS_2 and FIM_2). Every Nm is extracted independently from prior distributions (Table 1), resulting in 8 or 20 parameters for any linear stepping stone or islands matrix respectively.

We generated 100,000 coalescent simulations for each demographic model with *fastsimcoal v2.5.4* (Excoffier et al., 2013), extracting parameters from prior distributions using an in-house *R* script. We used a mutation rate of $7*10^{-8}$ per base per generation and a generation time of 7 years as for the *stairwayplot* analyses. We computed the following summary statistics to estimate the demographic parameters: within population nucleotide diversity (MPD), the pairwise unfolded site frequency spectrum (2D-SFS), and the pairwise fixation index (F_{ST}). We reduced the number of classes of the 2D-SFS by using a Partial Least Square (PLS) approach (Wegmann et al., 2009) and retained the first 15 components, reaching a total of 30 summary statistics when adding the 5 MPD and the 10 pairwise F_{ST}. Parameter estimation was computed using a local linear regression (Beaumont et al., 2002) on the closest 5,000 simulations to the vector of observed summary statistics. Similarly, to the unstructured models, we subsampled 30 individuals from VDG to avoid having an unbalanced sample, to reduce the number of classes when computing the 2D-SFS and to counteract the positive F_{IS} found.

To compute the posterior probability of each model we used the MPD and the pairwise F_{ST} only, since the PLS cannot be applied in the context of model selection. We applied a weighted multinomial logistic regression (Beaumont, 2008) in which we retained the closest 40,000 or 80,000 simulations to the vector of observed summary statistics. We performed a cross-validation test to check for the validity of our model selection procedure: we randomly generated 1,000 pseudo-observed data set *(pods)* from the prior distributions of each model and then we applied the same model selection procedure used for the real data.

Finally, we performed a Bayesian posterior predictive test (Gelman et al., 2013) to check if the estimated model is able to reproduce observed data. Briefly, we extracted parameter values

from the posterior distributions of the model under examination and computed summary statistics on the simulated dataset (of the same size of the real data). We used global F_{ST} as a predictive statistic, since we did not use it in the estimation process (Bertorelle et al., 2010).

Results

Marker discovery

On average a total of 3,979,746 reads per sample were obtained from the DArTseq sequencing. A total of 3,530 SNPs, found on 3,423 unique contigs ,were called by the DArTsoft14 pipeline for the combined *PstI-SphI* and *PstI-NspI* complexity reductions. Filters on read depth, ambiguous genotype calls, call rate, MAF and HWE successively discarded 375, 14, 509, 78, and 460, respectively, leaving a total of 2,094 high quality SNPs for bait design.

Genotyping and filtering

On average 766,621 reads per sample were obtained from the DArTcap sequencing. After clustering and SNP calling, we obtained 9,111 SNPs found on 7,853 unique contigs. The bait efficiency was 80.2% with 1,370 SNPs from the original DArTcap panel recovered. Filtering was applied separately to the full dataset (467 sharks) and a subsampled dataset (113 sharks; Supplemental Information S2.1). After applying the filtering steps for population structure analyses, 1,729 SNPs (of which 1,113 were in the DArTcap panel) for 461 sharks, and 1,731 SNPs (of which 1,115 were in the DArTcap panel) for 111 sharks remained for the full and subsampled datasets, respectively (Supplemental Information S2.1). All sampling areas exhibited similar genetic diversity, except for King Sound (KS), and to a lesser extent PNG, that exhibited lower diversity (Table 2).

After all the filtering steps (Supplemental Information S2.1), the final dataset for detecting the origin of the range expansion and the historical demographic analyses was composed of 461 *G. garricki* and one *G. glyphis* and comprised 1,822 loci harbouring 1,850 SNPs. We note that *G. glyphis* was used only to compute the unfolded SFS needed to detect the origin of the expansion and the demographic reconstruction (structured and unstructured models).

Genetic diversity and fixation index

Close to half of the loci were monomorphic in the KS population and its heterozygosity was about half of what was observed in the other populations. All samples combined exhibited a positive F_{IS}, whereas it did not differ significantly from 0 or was slightly negative for each sampling region taken separately (Table 2). Fixation indexes ranged from 0.001 (between some

412 rivers within the VDG) to 0.404 (KS vs PNG). Population differentiation between KS,
413 Cambridge Gulf (West Cambridge Gulf/Ord River), Daly River, VDG (Adelaide, Wildman,
414 West Alligator, South Alligator, and East Alligator Rivers, and Sampan Creek) and PNG was
415 about two orders of magnitude higher than between rivers within these regions. King Sound
416 hosted the most differentiated population overall, almost twice as differentiated as PNG, the
417 second highest differentiated population. Within the VDG, the Adelaide River was the most
418 differentiated, but still one order of magnitude less than between the sampling regions defined
419 above (Table 3). Consequently, we analysed the subsampled dataset with the Adelaide River
420 both within and distinct from the VDG population.

421 *Population structure*

422 Both Admixture and DAPC analyses of the full dataset without using *a priori* information on
423 sampling location revealed five distinct groups - KS, CG, Daly River, VDG and PNG -
424 although it was necessary to investigate up to K=8 to find these five groups because of some
425 apparent heterogeneity within VDG (Supplemental Information S1, sections 5.10 and 5.11).
426 The subsampled dataset did not suffer the same issue and the five groups are clear at K=5 (Fig.
427 3). The Admixture analysis showed that some individuals from the Daly River may have
428 inherited DNA from the adjacent populations (CG and VDG), whereas the DAPC analysis
429 showed one individual caught in the Daly River seemed to belong to the VDG gene pool (Fig.
430 3).

431 The second round of the hierarchical analysis generated more contrasted results. Further
432 genetic heterogeneity was evident in the VDG, although the differentiation was not quite as
433 clear as between regions (Supplemental Information S1, sections 7 and 8). No sign of
434 population differentiation was observed between sampling locations in CG (Supplemental
435 Information S1, section 9).

436 *Kin finding*

437 The accuracy of the parent-offspring, full-sibling, and half-sibling pair identification was
438 adequate for CKMR. Indeed, the LLR of the unrelated pairs was well separated from the half-
439 sibling pairs. Both are predicted by theory to be normally distributed and it was therefore easy
440 to visually define a cut-off that would eliminate all false-positives (i.e. high enough that no
441 unrelated pairs are expected above the cut-off) while retaining a large number of kin pairs.
442 There was also a clear gap between half-sibling and full-sibling/parent-offspring pairs (Fig. 4).
443 Four parent-offspring pairs were identified using the exclusion principle (Thompson, 2000)

444 and from their sizes it was easy to determine which one was the parent, in each case a male in
445 the size range 150–160 cm TL (i.e. at a size > size-at-maturity). For all pairs, each member was
446 caught in a different river, either East and South Alligator, or Wildman and South Alligator. A
447 total of 34 full-sibling and 130 half-sibling pairs were also identified. It was calculated that just
448 under 30% of the true half-sibling pairs did not pass the cut-off. The distribution of full and
449 half-sibling pairs across the different VDG rivers is shown in Table 4. All full-sibling pairs
450 except one, with one member in Sampan Creek and the other in the South Alligator River were
451 found in the same river.

452 Importantly, only nine full-sibling pairs and 34 half-sibling pairs were caught within two weeks
453 and, out of these, 40 had differences in total length over 140mm, which is more that the average
454 yearly growth rate (Bravington et al., 2019) suggesting the siblings belonged to different
455 cohorts. Also, no kin pair was present in the dataset with equal sample size (Supplemental
456 Information S1 section 6). We therefore believe that our sampling wasn't biased toward the
457 capture of litter mates and did not impact our population analyses.

458 *Detecting the origin of the range expansion*

459 The serial founder effects that characterize range expansions create a pattern of neutral shared
460 derived alleles that increase in frequency as one progresses away from the centre of origin. We
461 calculated the matrix of pairwise Ψ and tested for its significance using a permutation approach.
462 The equilibrium IBD model was barely rejected (p-value ~0.05), suggesting a range expansion
463 model is more likely to explain the observed data. The two peripheral populations, namely KS
464 and PNG, displayed the highest frequency of shared derived alleles and the lower genetic
465 diversity (Supplemental Information S2.2 and S2.6). The TDOA algorithm identifies the Gulf
466 of Carpentaria (that is, the area laying spatially between PNG and the VDG) as the most likely
467 origin of the expansion (Supplemental Information S2.6) consistently with the higher genetic
468 diversity found in its proximity (Supplemental Information S2.2). The probability of the
469 emplacement of the origin of the expansion decreases symmetrically east and west of the Gulf
470 (Supplemental Information S2.6), suggesting that two independent waves of colonization
471 occurred, one towards southern Papua New Guinea and the other following the northern coast
472 of Australia.

473 *Historical demographic inferences: unstructured models*

474 The *stairwayplot* is a non-parametric model that makes no assumptions over the change in *Ne*
475 through time, being able to recover complex demography (Liu & Fu, 2015). We ran several

replicates to explore the historical demography of each population to find the parameters that better reproduce the observed unfolded SFS. Bearing in mind that this approach considers the sample under investigation as coming from a panmictic population fully isolated (exchanging no migrants with any other populations), we found a strong abrupt bottleneck in all five populations as well as a more gradual decrease of *Ne* through time in the scatter sample (Fig. 5 and Supplemental Information S2.7). The two peripheral populations (KS and PNG) showed the signal of the bottleneck slightly more recent than the three others and of lower intensity (considering the ratio between the *Ne* at the time to the most recent common ancestor to the modern *Ne*). Repeating the analyses without singletons obtained consistent results (Supplemental Information S2.8).

Historical demographic inferences: structured models

We applied four structured demographic models to investigate the migration patterns between the five sampled populations and the time of origin of the colonization of the habitat (Fig. 2). First, we performed an ABC model selection procedure (Supplemental Information S2.3). Independently of the number of simulations retained to perform the logistic regression, LSS_1 was largely supported with a posterior probability of ~0.80. The two linear stepping stone models, sum up to a posterior probability of ~0.92, suggesting that an island migration matrix is highly unlikely. We checked the validity of our model selection procedure by performing a cross-validation experiment (Supplemental Information S2.4). No *pods* simulated under LSS_1 or LSS_2 were wrongly attributed to any island models with a probability higher than 0.80. Conversely, only two *pods* simulated under either FIM_1 or FIM_2 were attributed to LSS_1 or LSS_2 with the same threshold. This suggests that it is possible to carefully distinguish the two migration patterns (stepping stone vs island). Moreover, only few datasets (67) simulated under LSS_2 were wrongly attributed to LSS_1 with a probability equal or higher than 0.80 (Supplemental Information S2.4). Given the results of posterior probability obtained in real data, this shows that it is highly unlikely that our populations experienced a change in connectivity through time. We then focused on the demographic parameters estimated under LSS_1. First, we found that the modes of all *Nm* parameters range between 0.37 and 2.5, suggesting low connectivity (Table 1). Moreover, PNG and KS, which are the two most peripheral populations of the linear stepping stone system, are indeed the less connected with their respective neighbours, further suggesting their isolation. We note that all distributions are well peaked and different from the priors, suggesting that the data contains enough information to correctly estimate these parameters (Supplemental Information S2.9). The time of the

instantaneous colonization T_i is very recent, with a mode of 2,000 generations (95% credible interval: 554 – 45,700), showing a well peaked distribution (Table 1 and Supplemental Information S2.9). Finally, we checked the validity of LSS_1 by means of a posterior predictive test. The observed value of the global F_{ST} falls within the 95% of the posterior predictive distribution simulated under LSS_1, suggesting that this model cannot be rejected and it is able to reproduce our data (Supplemental Information S2.10).

Discussion

Population structure

Inferring population structure and reconstructing the historical demography of a species is essential to better establish conservation priorities and management policies. This study provides the first insight into the genetic population structure of the threatened shark *G. garricki*. From our analyses, it is clear that this species has very limited reproductive dispersal. All five sample regions host at least one distinct population, with possible substructure within the VDG. Indeed, the clustering analyses and the distribution of kin pairs, mostly found within the same river, suggest the gene pool is not even homogeneous at that scale. Such fine-scale structure is uncommon in sharks and as far as we know, *G. glyphis* is the only other shark species presenting similar levels of genetic differentiation over just a few hundred kilometres of coastline (Feutry et al., 2017; Feutry et al., 2014). These two species make slightly different use of rivers, *G. glyphis* adults never being found in them, whereas *G. garricki* adults are found in tidal reaches of rivers (unpublished data), which may result in different dispersal capabilities.

Interestingly, 24 out of 130 half-sibling pairs were found in different rivers, whereas all 34 full-sibling pairs but one were found in the same river. This indicates fairly restricted juvenile movements, with most of the dispersal being undertaken by larger individuals. Also, all parents involved in a parent-offspring pair were males and the parent was always found in a different river to the offspring. Although the number of observations is small, this tends to support the idea that adult males frequently move from one river to another to breed. Alternatively, mating may occur outside rivers with females always pupping in their natal river and males mating with females from different rivers. While future analysis of mitochondrial DNA combined with kin data has the potential to reveal sex-specific structuring at even finer scale within the VDG (Feutry et al., 2017), our results already demonstrate that reproductive philopatry in both sexes is strong enough in this species to generate highly differentiated populations over distances as short as 200 km.

Historical demography

Both the variation of the *Ne* and migration rate through time among sampled populations need to be critically evaluated in order to evaluate potential threats affecting the target species. Population genetics is a powerful tool to infer these parameters, in particular when NGS data are available, offering a large number of (mostly) unlinked SNPs. Having large inferential power comes with a drawback: a wrong model will give a wrong answer with high degree of confidence, implying that careful attention is warranted when choosing the model that best explains the data. Here we followed recommendations from Maisano Delser et al. (2018) by first testing the spatial structure of the data to detect if a range expansion occurred in *G. garricki*.

We found lower diversity in the more isolated populations at the western and eastern edge of their range (KS and PNG, Table 3). The occurrence of a range expansion and the large values of F_{ST} found at such a small geographical scale both suggest that metapopulation models should be applied to best explain the observed data. Nevertheless, contrasting unstructured models (i.e., models which assume that the population under examination has never exchanged migrants with other populations) at a different sampling level is a simple approach to provide a first inference on the history of the metapopulation (Maisano Delser et al., 2016; Städler et al., 2009; Wakeley, 1999).

We first computed the *stairwayplot*, which considers populations unstructured, in our five populations and the *scatter* sample. Generally speaking, if population structure is suspected (as in our case), results obtained from unstructured models cannot be interpreted as simple variation in *Ne* through time but as the consequence of the interaction between *Ne* and *m* (Maisano Delser et al., 2018; Rodríguez et al., 2018). Typically, populations belonging to a metapopulation characterized by low *Nm* show a signature of decline even if demographically stable (Chikhi et al., 2010). The *stairway* plot reconstructed in the five populations and in the *scatter* sample showed in all cases a dramatic decline of *Ne*. We therefore interpreted this variation in *Ne* as a consequence of the low *Nm* of the metapopulation rather than a demographic bottleneck. This is crucial from a conservation genetics perspective, suggesting in our case, few exchanges between regions but possibly genetically healthy populations within (i.e. not at risk of inbreeding depression).

The analysis of the *stairwayplot* (or any unstructured methods exploring the variation of *Ne* through times) conveys important details on the temporal dynamics of *Nm* but its interpretation is not straightforward (Rodríguez et al. 2018). To confirm these intuitions and deeply

574 investigate the evolutionary history of this species, we further applied structured models to
575 explicitly infer the parameters of interests (i.e., the *Nm* between the *G. garricki* populations
576 and the expansion time). By testing four complex demographic models devised to study the
577 migration matrix linking sampled populations (Fig. 2) and to detect temporal changes in *Nm*,
578 we showed that: i) the populations are connected through a linear stepping stone given the
579 higher posterior probability received by LSS when compared against FIM models (>0.90 and
580 well supported by the cross validation experiment); ii) the connectivity is very low, with the
581 mode of *Nm* values never higher then 2.5 (Table 1, Supplemental Information S2.9), consistent
582 with the results from the *stairwayplot*; iii) the migration rates have not changed since the time
583 of the colonization of the habitat (model LSS_1 has the highest posterior probability); iv) the
584 expansion time is very recent, with a mode of ~14,000 years B.P. (Table 1, Supplemental
585 Information S2.9). These results are biologically reasonable since this species prefers highly
586 turbid coastal, estuarine and tidal riverine environments (Pillans et al., 2010), (P.M. Kyne,
587 (unpublished data) which may restrict individuals to particular habitats, decreasing the
588 likelihood of long distance migrations.

589 Importantly, we did not detect a change in the *Nm* through time and LSS_1 was largely
590 preferred over LSS_2 (Supplemental Information S2.3). This finding seems robust given the
591 results of the cross-validation test, where LSS_1 is generally well differentiated from LSS_2
592 (Supplemental Information S2.4). This is reassuring from a conservation genetics perspective
593 as it highlights that this species has not declined significantly, consistent with the fact that it
594 inhabits mostly pristine environments.

595 When we combined evidence from the estimated expansion time and the present distribution
596 of both *G. garricki* and *G. glyphis*, the evolutionary history of *G. garricki* appears even more
597 intriguing. Considering a generation time of 7 years, the estimated expansion time of ~14,000
598 years B.P. is compatible with the opening of the Gulf of Carpentaria (Yokoyama et al., 2001),
599 the area where we inferred the origin of the range expansion using the directionality index of
600 Peter and Slatkin (2013) (Supplemental Information S2.6). The most parsimonious explanation
601 is that this shark species started expanding during the opening of the Gulf of Carpentaria,
602 somewhere in between PNG and VDG, tracking patches of suitable habitat becoming
603 progressively available after the Last Glacial Maximum (Yokoyama et al. 2001). There are no
604 historical or contemporary records of *G. garricki* in the Gulf of Carpentaria, meaning that either
605 we have not found it yet (the region is remote and many rivers are poorly surveyed) or that
606 environmental conditions became progressively unsuitable during the opening of the sea

607 expanse, in agreement with the wave of colonization towards newly established areas. This
608 suggests that the history of this species is extremely recent, but whether *G. garricki* was present
609 elsewhere before the Last Glacial Maximum or speciated soon after from *G. glyphis* remains
610 to be discovered. In both cases, these two species represent model species to investigate
611 speciation in sharks.

612 The KS population stands out, both in terms of genetic differentiation (about twice as high as
613 between any other two populations) and heterozygosity (about twice as low as any other
614 population). *Glyphis garricki* are rare in KS and the high incidence of skeletal deformities
615 reported from there was suspected to be due to inbreeding in a small gene pool (Thorburn &
616 Morgan, 2004). Fast genetic drift leading to higher genetic differentiation and low
617 heterozygosity (mostly due to the high proportion of monomorphic loci) is also consistent with
618 the presence of a small population in KS. These observations are supported by our historical
619 demographic results: KS lies at the western edge of the range expansion starting from the Gulf
620 of Carpentaria, therefore experiencing more drift than the other populations. Consistently with
621 theoretical (Peischl et al., 2015) and empirical observations (Willi et al., 2018) on the dynamics
622 of range expansion, KS has likely accumulated a larger mutational load, which provides an
623 alternative explanation to inbreeding for the observed morphological anomalies. The inferred
624 small population size, low heterozygosity, and relatively high genetic differentiation compared
625 to other Australian populations, indicates that the KS population in particular needs to be
626 managed without any expectation that any local declines as a result of threatening processes
627 can be balanced by immigration. It is reasonable to expect this population, with its low
628 heterozygosity, rarity, and high incidence of skeletal deformities is likely to be more
629 susceptible to anthropogenic change, including those due to climate, and our results suggest
630 the need for additional caution in its management.

631 *DArTcap performances and kinship inference*

632 The DArTcap method proved effective in isolating a sufficient number of informative SNPs
633 for CKMR analysis to estimate population parameters that will influence how this threatened
634 species, or pressures acting upon it, are managed. On-target efficiency seems somewhat lower
635 than reported for other approaches combining RAD and sequence capture (Ali et al., 2016;
636 Hoffberg et al., 2016). Several factors can explain this: (i) in this study over 1,700 SNPs were
637 targeted as opposed to 500 and 964 for Rapture and RADcap, respectively; (ii) there was no
638 redundancy in our bait design, we had only one bait per locus (redundancy increases efficiency
639 but also increases costs); (iii) genetic diversity is extremely low in *G. garricki* and we had a

limited choice of loci to choose from and GC content was not always optimal; and, (iv) *G. garricki* has a relatively large genome (~5Gb, P. Feutry, unpublished data), which is likely to increase the competition between the targeted sequences and their paralogs. DArTcap performance on species with smaller genomes suggests efficiency probably decreases when genome size increases (unpublished data), due to an increasing risk of capturing sequences with some similarity with the panel. Despite these difficulties, the combination of on-target and off-target loci obtained with DArTcap resulted in a number of high-quality SNP equivalent to the number of loci we had baits designed for and at a very reasonable cost (~AU$15 per individual including bioinformatic support). In other less demanding species, more loci could probably be included in the panel, for a similar cost.

While there is no direct attempt to select makers from the 'functional' part of the genome, in many organisms (especially plants), the DArTseq method selects genic regions with very high efficiency. One of the most important selection criteria for DArTcap assay markers is the size of sequence clusters in which the marker is identified. DArTseq marker identification (through the DArTsoft14 program) involves clustering sequences with a defined distance threshold and parsing larger clusters into SNP loci. The smaller the cluster the more likely the marker is coming from a single copy sequence in the genome. In any DArTcap panel there is a definite enrichment for low/single copy sequences and therefore is likely to enrich for the functional fraction of the genome. The selection of markers and design of the capture baits (oligonucleotides) excludes low complexity repetitive regions of the genome, thereby effectively eliminating the issue of paralogous sequences affecting allele calling.

Potential problems with estimating diversity is not a feature unique to DArTcap, but of any technology which selects specific sets of markers, which leads to potential ascertainment bias (Lachance & Tishkoff, 2013). We were fully aware of such risk and our SNP panel was selected after genotyping a large population of samples on the DArTseq platform which is free from ascertainment bias. Analyses conducted on the subset of samples that was used to discover the SNPs showed ascertainment bias is likely limited in this study (Supplemental Information S1 section 4).

The SNP panel designed for this study was perfectly adequate for CKMR. Given the distinct LLR distributions of unrelated versus kin pairs, it was possible to retain a large number of kin pairs without having to worry about false-positives (Fig. 4). The panel can also be used to infer self-identity (i.e. recaptures) or species identification by adopting a similar approach to the one taken by Kyne and Feutry (2017). For species without a clear external indicator of sex, like

673 claspers in sharks, the SNP panel could be further optimized to include sex-specific markers
674 such as those identified in this study.

675 *Conclusions*

676 DArTcap is a new, cost-effective, high-throughput option in the growing market of complexity
677 reduction sequencing methods and this study demonstrates how its efficiency can be
678 maximised by carefully designing the SNP panel. This enabled cost-efficient and highly
679 accurate identification of first and second-degree relatives, which is critical for downstream
680 applications such as CKMR. In addition to kinship analysis, we demonstrated that the SNP
681 panel could be used to investigate population structure and historical demography in great
682 detail, providing important information for the management of threatened species at no extra
683 cost. This is a significant improvement on earlier methods, but does depend on a well-designed
684 SNP panel. For our case study of *G. garricki,* five distinct populations were detected across the
685 known species range, with extremely low inter-population geneflow, and evidence of further
686 intra-population structuring. Overall, populations are believed to be genetically healthy, but
687 small, isolated and confined to rivers and coastal embayments where anthropogenic pressures
688 could result in rapid declines. The KS population may be especially susceptible to
689 anthropogenic change. While much of the species' range is currently subject to low human
690 interference, excepting possibly the future impacts of climate change and localised mortality
691 due to fishing activities, opportunities for increasing development in northern Australia are
692 under active consideration (Commonwealth of Australia, 2015). This suggests that future
693 pressures on these isolated populations will increase. Sampling undertaken for this study, and
694 others on threatened river sharks, has revealed the occurrence of the species in many locations
695 not previously documented. Not all rivers containing suitable habitat have been surveyed for
696 this species, and it is possible that additional sampling would reveal additional populations.
697 Each population should be considered an independent unit for management purposes given
698 gene flow is extremely low.
699

700 ## Acknowledgements

701 This work was undertaken for the Marine Biodiversity Hub, a collaborative partnership
702 supported through funding from the Australian Government's National Environmental Science
703 Program. This research was funded in part through an Ord River Research Offset grant through
704 CSIRO. We thank the many Traditional Owners, assistants, and volunteers that assisted with
705 field work. We are grateful to the genotoul bioinformatics platform Toulouse Midi-Pyrenees

(Bioinfo Genotoul) for providing computing resources (www.bioinfo.genotoul.fr) and to Mark Bravington, Paige Evenson and Shane Baylis for assistance with the DArTseq marker selection and kinship analysis. Finally, we would like to thank the five anonymous reviewers who helped improve the quality of this manuscript.

References

Ackerman, M. W., Hand, B. K., Waples, R. K., Luikart, G., Waples, R. S., Steele, C. A., . . . Campbell, M. R. (2017). Effective number of breeders from sibship reconstruction: Empirical evaluations using hatchery steelhead. *Evolutionary Applications, 10*(2), 146-160.

Alexander, D. H., & Lange, K. (2011). Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. *BMC Bioinformatics, 12*(1), 246.

Ali, O. A., O'Rourke, S. M., Amish, S. J., Meek, M. H., Luikart, G., Jeffres, C., & Miller, M. R. (2016). RAD capture (Rapture): flexible and efficient sequence-based genotyping. *Genetics, 202*(2), 389-400.

Anderson, J. L., Mari, A. R., Braasch, I., Amores, A., Hohenlohe, P., Batzel, P., & Postlethwait, J. H. (2012). Multiple sex-associated regions and a putative sex chromosome in zebrafish revealed by RAD mapping and population genomics. *Plos One, 7*(7), e40701.

Beaumont, M. A. (2008). Joint determination of topology, divergence time, and immigration in population trees. In S. Matsumura, P. Forster, & C. Renfrew (Eds.), *Simulation, genetics, and human prehistory* (pp. 135-154). Cambridge: McDonald Institute for Archaeological Research.

Beaumont, M. A., Zhang, W., & Balding, D. J. (2002). Approximate Bayesian computation in population genetics. *Genetics, 162*(4), 2025-2035.

Bekkevold, D., Helyar, S. J., Limborg, M. T., Nielsen, E. E., Hemmer-Hansen, J., Clausen, L. A., & Carvalho, G. R. (2015). Gene-associated markers can assign origin in a weakly structured fish, Atlantic herring. *ICES Journal of Marine Science, fsu247*.

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.

Bilton, D. T., Paula, J., & Bishop, J. D. D. (2002). Dispersal, genetic differentiation and speciation in estuarine organisms. *Estuarine, Coastal and Shelf Science, 55*(6), 937-952.

Botstein, D., White, R. L., Skolnick, M., & Davis, R. W. (1980). Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *American Journal of Human Genetics, 32*(3), 314.

Bravington, M. V., Feutry, P., Pillans, R. D., Hillary, R. M., Johnson, G., Saunders, T., . . . Kyne, P. M. (2019). Close-Kin Mark-Recapture population size estimate of *Glyphis garricki* in the Northern Territory. *Report to the National Environmental Science Program, Marine Biodiversity Hub. CSIRO Oceans & Atmosphere, Hobart.*

Bravington, M. V., Grewe, P. M., & Davies, C. R. (2016a). Absolute abundance of southern bluefin tuna estimated by close-kin mark-recapture. *Nature Communications, 7*, 13162.

Bravington, M. V., Skaug, H. J., & Anderson, E. C. (2016b). Close-kin mark-recapture methods. *Statistical Science, 31*, 259-274.

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.

Chin, A., Kyne, P. M., Walker, T. I., & McAuley, R. B. (2010). An integrated risk assessment for climate change: analysing the vulnerability of sharks and rays on Australia's Great Barrier Reef. *Global Change Biology, 16*(7), 1936-1953.

DoE. (2015). *Sawfish and river sharks multispecies recovery plan*. Retrieved from Canberra:

Dulvy, N. K., Fowler, S. L., Musick, J. A., Cavanagh, R. D., Kyne, P. M., Harrison, L. R., . . . Francis, M. P. (2014). Extinction risk and conservation of the world's sharks and rays. *elife, 3*, e00590.

Excoffier, L., Dupanloup, I., Huerta-Sánchez, E., Sousa, V. C., & Foll, M. (2013). Robust demographic inference from genomic and SNP data. *PLoS Genetics, 9*(10), e1003905.

Feldheim, K. A., Gruber, S. H., DiBattista, J. D., Babcock, E. A., Kessel, S. T., Hendry, A. P., . . . Chapman, D. D. (2014). Two decades of genetic profiling yields first evidence of natal philopatry and long-term fidelity to parturition sites in sharks. *Molecular Ecology, 23*(1), 110-117.

Feutry, P., Berry, O., Kyne, P. M., Pillans, R. D., Hillary, R. M., Grewe, P. M., . . . Bravington, M. V. (2017). Inferring contemporary and historical genetic connectivity from juveniles. *Molecular Ecology, 26*, 444-456.

Feutry, P., Kyne, P. M., Pillans, R. D., Chen, X., Naylor, G. J. P., & Grewe, P. M. (2014). Mitogenomics of the Speartooth Shark challenges ten years of control region sequencing. *BMC Evolutionary Biology, 14*, 232. doi:10.1186/s12862-014-0232-x

Foster, S. D., Feutry, P., Grewe, P. M., Berry, O., Hui, F. K., & Davies, C. R. (2018). Reliably discriminating stock structure with genetic markers: Mixture models with robust and fast computation. *Molecular Ecology Resources, 18*(6), 1310-1325.

Gelman, A., Stern, H. S., Carlin, J. B., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). *Bayesian data analysis* (3rd Edition ed.). New York: Chapman and Hall/CRC.

Gosselin, T., Lamothe, M., Devloo-Delva, F., & Grewe, P. (2020). radiator: RADseq Data Exploration, Manipulation and Visualization using R. R package version 1.1.5 https://thierrygosselin.github.io/radiator/. doi : 10.5281/zenodo.3687060.

Grewe, P. M., Feutry, P., Hill, P. L., Gunasekera, R. M., Schaefer, K. M., Itano, D. G., . . . Davies, C. R. (2015). Evidence of discrete yellowfin tuna (*Thunnus albacares*) populations demands rethink of management for this globally important resource. *Scientific Reports, 5*, 16916.

Gustafsson, F., & Gunnarsson, F. (2003). *Positioning using time-difference of arrival measurements.* Paper presented at the IEEE (Institute of Electrical and Electronics Engineers) International Conference on Acoustics, Speech, and Signal Processing, 2003. Proceedings. (ICASSP'03). Hong Kong.

Hess, J. E., Campbell, N. R., Docker, M. F., Baker, C., Jackson, A., Lampman, R., . . . Young, W. P. (2015). Use of genotyping by sequencing data to develop a high-throughput and multifunctional SNP panel for conservation applications in Pacific lamprey. *Molecular Ecology Resources, 15*(1), 187-202.

Hillary, R. M., Bravington, M. V., Patterson, T. A., Grewe, P. M., Bradford, R., Feutry, P., . . . Francis, M. P. (2018). Genetic relatedness reveals total population size of white sharks in eastern Australia and New Zealand. *Scientific Reports, 8*(1), 2661.

798 Hoffberg, S. L., Kieran, T. J., Catchen, J. M., Devault, A., Faircloth, B. C., Mauricio, R., & Glenn,
799 T. C. (2016). RAD cap: sequence capture of dual-digest RAD seq libraries with
800 identifiable duplicates and reduced missing data. *Molecular Ecology Resources, 16*(5), 1264-1278.
801
802 Jabado, R. W., Kyne, P. M., Nazareth, E., & Sutaria, D. N. (2018). A rare contemporary
803 record of the Critically Endangered Ganges shark *Glyphis gangeticus*. *Journal of*
804 *Fish Biology, 92*(5), 1663-1669.
805 Jaccoud, D., Peng, K., Feinstein, D., & Kilian, A. (2001). Diversity arrays: a solid state
806 technology for sequence information independent genotyping. *Nucleic Acids*
807 *Research, 29*(4), e25-e25.
808 Jombart, T., & Ahmed, I. (2011). adegenet 1.3-1: new tools for the analysis of genome-
809 wide SNP data. *Bioinformatics, 27*(21), 3070-3071.
810 Jombart, T., Devillard, S., & Balloux, F. (2010). Discriminant analysis of principal
811 components: a new method for the analysis of genetically structured populations.
812 *BMC Genetics, 11*(1), 94.
813 Jones, M. R., & Good, J. M. (2016). Targeted capture in evolutionary and ecological
814 genomics. *Molecular Ecology, 25*(1), 185-202.
815 Kilian, A., Wenzl, P., Huttner, E., Carling, J., Xia, L., Blois, H., . . . Hopper, C. (2012). Diversity
816 arrays technology: a generic genome profiling technology on open platforms. In
817 *Data production and analysis in population genomics* (pp. 67-89): Springer.
818 Kyne, P. M. (2014). Threatened fishes and marine turtles of Kakadu National Park (with
819 notes on marine mammals). In S. Winderlich & J. Woinarski (Eds.), *Kakadu*
820 *National Park Landscape Symposia Series. Symposium 7: Conservation of*
821 *Threatened Species, 26–27 March 2013, Bowali Visitor Centre, Kakadu National*
822 *Park* (pp. 58-74). Darwin: Internal Report 623, Supervising Scientist.
823 Kyne, P. M., & Feutry, P. (2017). Recreational fishing impacts on threatened river sharks:
824 A potential conservation issue. *Ecological Management & Restoration, 18*(3), 209-
825 213.
826 Lachance, J., & Tishkoff, S. A. (2013). SNP ascertainment bias in population genetic
827 analyses: why it is important, and how to correct it. *Bioessays, 35*(9), 780-786.
828 Lavergne, E., Calvès, I., Meistertzheim, A. L., Charrier, G., Zajonz, U., & Laroche, J. (2014).
829 Complex genetic structure of a euryhaline marine fish in temporarily open/closed
830 estuaries from the wider Gulf of Aden. *Marine Biology, 161*(5), 1113-1126.
831 Li, C., Corrigan, S., Yang, L., Straube, N., Harris, M., Hofreiter, M., . . . Naylor, G. J. P. (2015).
832 DNA capture reveals transoceanic gene flow in endangered river sharks.
833 *Proceedings of the National Academy of Sciences, 112*(43), 13302-13307.
834 Liu, X., & Fu, Y.-X. (2015). Exploring population size changes using SNP frequency spectra.
835 *Nature Genetics, 47*(5), 555.
836 Lowe, W. H., Kovach, R. P., & Allendorf, F. W. (2017). Population genetics and demography
837 unite ecology and evolution. *Trends in Ecology & Evolution, 32*(2), 141-152.
838 Lucifora, L. O., de Carvalho, M. R., Kyne, P. M., & White, W. T. (2015). Freshwater sharks
839 and rays. *Current Biology, 25*(20), R971-R973.
840 Maisano Delser, P., Corrigan, S., Duckett, D., Suwalski, A., Veuille, M., Planes, S., . . . Mona,
841 S. (2018). Demographic inferences after a range expansion can be biased: the test
842 case of the blacktip reef shark (*Carcharhinus melanopterus*). *Heredity, 122*, 759-
843 769.
844 Maisano Delser, P., Corrigan, S., Hale, M., Li, C., Veuille, M., Planes, S., . . . Mona, S. (2016).
845 Population genomics of *C. melanopterus* using target gene capture data:

846 demographic inferences and conservation perspectives. *Scientific Reports, 6*,
847 33753.
848 Ovenden, J. R., Dudgeon, C., Feutry, P., Feldheim, K., & Maes, G. E. (2019). Genetics and
849 genomics for fundamental and applied research on elasmobranchs. In J. C. Carrier,
850 M. R. Heithaus, & C. A. Simpfendorfer (Eds.), *Shark Research: Emerging*
851 *Technologies and Applications for the Field and Laboratory* (pp. 235-253). Boca
852 Raton: Taylor and Francis Group.
853 Peischl, S., Kirkpatrick, M., & Excoffier, L. (2015). Expansion load and the evolutionary
854 dynamics of a species range. *The American Naturalist, 185*(4), E81-E93.
855 Pembleton, L. W., Cogan, N. O., & Forster, J. W. (2013). StAMPP: an R package for
856 calculation of genetic differentiation and structure of mixed - ploidy level
857 populations. *Molecular Ecology Resources, 13*(5), 946-952.
858 Peter, B. M., & Slatkin, M. (2013). Detecting range expansions from genetic data. *Evolution,*
859 *67*(11), 3274-3289.
860 Phillips, N. M., Chaplin, J. A., Morgan, D. L., & Peverell, S. C. (2011). Population genetic
861 structure and genetic diversity of three critically endangered *Pristis* sawfishes in
862 Australian waters. *Marine Biology, 158*(4), 903-915.
863 Pillans, R. D., Stevens, J. D., Kyne, P. M., & Salini, J. (2010). Observations on the distribution,
864 biology, short-term movements and habitat requirements of river sharks *Glyphis*
865 spp. in northern Australia. *Endangered Species Research, 10*, 321-332.
866 doi:10.3354/esr00206
867 Rodríguez, W., Mazet, O., Grusea, S., Arredondo, A., Corujo, J. M., Boitard, S., & Chikhi, L.
868 (2018). The IICR and the non-stationary structured coalescent: towards
869 demographic inference with arbitrary changes in population structure. *Heredity,*
870 *121*(6), 663.
871 Slatkin, M., & Excoffier, L. (2012). Serial founder effects during range expansion: a spatial
872 analog of genetic drift. *Genetics, 191*(1), 171-181.
873 Städler, T., Haubold, B., Merino, C., Stephan, W., & Pfaffelhuber, P. (2009). The impact of
874 sampling schemes on the site frequency spectrum in nonequilibrium subdivided
875 populations. *Genetics, 182*(1), 205-216.
876 Thompson, E. (2000). *Statistical inference from genetic data on pedigrees* (Vol. 6): Institute
877 of Mathematical Statistics.
878 Thorburn, D. C., & Morgan, D. L. (2004). The northern river shark *Glyphis* sp. *C*
879 (Carcharhinidae) discovered in Western Australia. *Zootaxa, 685*(1), 1-8.
880 Vähä, J. P., Erkinaro, J., Niemelä, E., & Primmer, C. R. (2007). Life - history and habitat
881 features influence the within - river genetic structure of Atlantic salmon. *Molecular*
882 *Ecology, 16*(13), 2638-2654.
883 Wakeley, J. (1999). Nonequilibrium migration in human history. *Genetics, 153*(4), 1863-
884 1871.
885 Waples, R. S., Grewe, P. M., Bravington, M. W., Hillary, R., & Feutry, P. (2018). Robust
886 estimates of a high Ne/N ratio in a top marine predator, southern bluefin tuna.
887 *Science advances, 4*(7), eaar7759.
888 Watts, R. J., & Johnson, M. S. (2004). Estuaries, lagoons and enclosed embayments:
889 habitats that enhance population subdivision of inshore fishes. *Marine and*
890 *Freshwater Research, 55*(7), 641-651.
891 Wegmann, D., Leuenberger, C., & Excoffier, L. (2009). Efficient approximate Bayesian
892 computation coupled with Markov chain Monte Carlo without likelihood. *Genetics,*
893 *182*(4), 1207-1218.

894 Weir, B. S., & Cockerham, C. C. (1984). Estimating F-statistics for the analysis of
895 population structure. *Evolution, 38*(6), 1358-1370.
896 White, W. T., Appleyard, S. A., Sabub, B., Kyne, P. M., Harris, M., Lis, R., . . . Corrigan, S.
897 (2015). Rediscovery of the Threatened River Sharks, *Glyphis garricki* and *G. glyphis*,
898 in Papua New Guinea. *Plos One, 10*(10), e0140075.
899 Whitlock, M. C., & Lotterhos, K. E. (2015). Reliable detection of loci responsible for local
900 adaptation: inference of a null model through trimming the distribution of Fst. *The*
901 *American Naturalist, 186*, S24-36.
902 Willi, Y., Fracassetti, M., Zoller, S., & Van Buskirk, J. (2018). Accumulation of mutational
903 load at the edges of a species range. *Molecular Biology and Evolution, 35*(4), 781-
904 791.
905 Woinarski, J., Mackey, B., Nix, H., & Traill, B. (2007). *The nature of Northern Australia: it's*
906 *natural values, ecological processes and future prospects*: ANU E Press.
907 Yokoyama, Y., Purcell, A., Lambeck, K., & Johnston, P. (2001). Shore-line reconstruction
908 around Australia during the last glacial maximum and late glacial stage.
909 *Quaternary International, 83*, 9-18.

910

Data accessibility:

DArTseq/DArTcap genotypes and associated metadata have been deposited on Dryad DOI: 10.5061/dryad.hqbzkh1ch

Data availability:

The data that supports the findings of this study are openly available, see Data accessibility section above for DOI.

Author contributions:

PF and PMK designed the research. PMK, GJ, RDP, DLM, PF collected samples. PF, RMG, DJ, AK performed research. DJ and AK contributed new genotyping method. PF, FD, ATLY, SM analysed data. PF and SM wrote the initial draft and all other authors provided feedback. TS, NJB, RDP, PMK secured funding.

Tables

Table 1: Parameters estimated under the linear stepping stone model with 1 migration matrix (model LSS_1). Number of migrants per generation between populations (*Nm*) are expressed in reverse, following coalescent norm (i.e. number of immigrants received by the population left of the arrow from the one at its right). T_i is in generations. U: uniform distribution. U^2: combination of two independent prior distributions (for *N* and *m*). Population abbreviation: King Sound (KS), Cambridge Gulf (CG), Daly River (DR), Van Diemien Gulf (VDG), Papua New Guinea (PNG). T_i is in generations (generation time is estimated as 7 years, see main text). U: uniform distribution. U^2: combination of two independent prior distributions (for *N* and *m*).

Parameter	Prior	Median	Mode	2.5% CI	97.5% CI
Nm KS → CG	U^2 : 0-50	0.446	0.372	0.087	1.087
Nm CG → KS	U^2 : 0-50	1.167	0.822	0.326	4.177
Nm CG → DR	U^2 : 0-50	3.650	2.330	1.106	24.089
Nm DR → CG	U^2 : 0-50	4.126	2.279	0.747	36.034
Nm DR → VDG	U^2 : 0-50	4.009	2.490	1.083	25.945
Nm VDG → DR	U^2 : 0-50	3.050	2.103	1.156	13.530
Nm VDG → PNG	U^2 : 0-50	0.785	0.565	0.285	4.999
Nm PNG → VDG	U^2: 0-50	2.423	1.528	0.578	25.866
N_{anc}	U: 100-50000	35363	36359	25790	38673
T_i	U: 100-50000	4151	2002	554	45713

Table 2: Genetic diversity indices for the global dataset (all regions), and for each of the five populations, comprising 461 *G. garricki* and 1,734 SNPs. Number of monomorphic loci (*Mo*), Allelic richness (*Ar*), observed heterozygosity (*Ho*), unbiased expected heterozygosity (*uHe*), and inbreeding coefficient (F_{IS} [95 % confidence intervals]).

Diversity Indices	**All Regions**	**King Sound**	**Cambridge Gulf**	**Daly River**	**Van Diemen Gulf**	**Papua New Guinea**
	N=461	**N=19**	**N=30**	**N=29**	**N=379**	**N=4**
Mo	0	814	175	139	149	595
Ar	1.591	1.378	1.663	1.658	1.693	1.565
Ho	0.264	0.171	0.285	0.281	0.292	0.289
uHe	0.261	0.165	0.287	0.283	0.293	0.280
F_{IS}	0.057	-0.026	0.006	0.005	0.005	-0.058
	[0.051,0.060]	[-0.052,-0.014]	[-0.005,0.017]	[-0.004,0.020]	[-0.001,0.007]	[-0.069,-0.006]

Table 3: Pairwise F_{ST} values between all sampled locations for 461 *G. garricki*, with a bootstrap of 10,000.

***: p<0.0001, **: p<0.01*: p<0.05

	King Sound N= 19	**West Cambridge Gulf N= 15**	**Ord River N= 15**	**Daly River N= 29**	**Adelaide River N= 32**	**Sampan Creek N= 30**	**Wildman River N= 47**	**West Alligator River N= 41**	**South Alligator River N= 159**	**East Alligator River N= 70**	**Papua New Guinea N=4**
West Cambridge Gulf	0.312***										
Ord River	0.317***	0.008***									
Daly River	0.302***	0.093***	0.096***								
Adelaide River	0.293***	0.121***	0.128***	0.090***							
Sampan Creek	0.287***	0.122***	0.128***	0.089***	0.014***						
Wildman River	0.280***	0.119***	0.126***	0.088***	0.016***	0.006***					
West Alligator River	0.283***	0.121***	0.126***	0.089***	0.015***	0.004***	0.007***				
South Alligator River	0.263***	0.121***	0.127***	0.089***	0.014***	0.002**	0.007***	0.004***			
East Alligator River	0.275***	0.123***	0.128***	0.091***	0.014***	0.001*	0.006***	0.004***	0.001***		
Papua New Guinea	0.404***	0.155***	0.153***	0.179***	0.172***	0.169***	0.171***	0.173***	0.170***	0.171***	

Table 4: Intra and inter-river number of *G. garricki* full-sibling pairs (N = 34, upper triangle and second number along the diagonal) and half-sibling pairs (N = 130, lower triangle and first number along the diagonal) within Van Diemen Gulf.

	Adelaide River N= 32	**Sampan Creek N= 30**	**Wildman River N= 47**	**West Alligator River N= 41**	**South Alligator River N= 159**	**East Alligator River N= 70**
Adelaide River	8 \ 1	0	0	0	0	0
Sampan Creek	0	1 \ 1	0	0	1	0
Wildman River	1	1	28 \ 8	0	0	0
West Alligator River	1	0	0	8 \ 2	0	0
South Alligator River	1	0	0	6	54 \ 19	0
East Alligator River	1	6	0	3	4	7 \ 2

Figures

Figure 1: Sampling map of *G. garricki* in northern Australia and Papua New Guinea. Sampling regions are listed in bold and sampling locations are: (WC), West Cambridge Gulf (i.e. Durack and Pentecost Rivers, and the West Arm of Cambridge Gulf); (O) Ord River; (A), Adelaide River; (S), Sampan Creek; (W), Wildman River; (WA), West Alligator River; (SA), South Alligator River; and, (EA), East Alligator River.

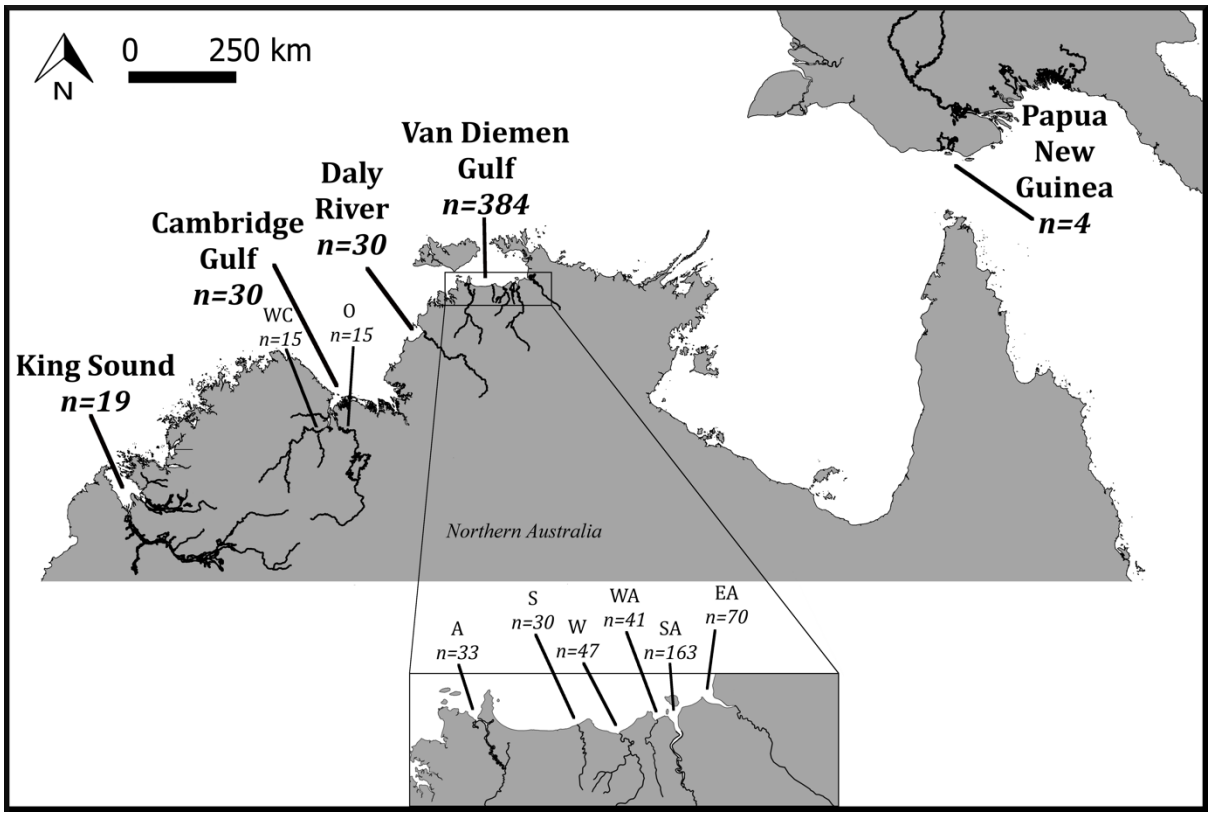

Figure 2: Alternative scenarios of *G. garricki* evolution, tested under an ABC framework on the filtered genetic dataset. (a) Linear stepping-stone model with 1 migration matrix (LSS_1) and (b) 2 migration matrices (LSS_2), (c) non-equilibrium finite island model with 1 migration matrix (FIM_1) and (d) 2 migration matrices (FIM_2). Detailed description of each model and their associated parameters is presented in the main text. Nanc is the ancestral effective population size of the founding deme. Ti is the instantaneous colonization / expansion time, when *G. garricki* colonized the available habitat. Tm is the instantaneous time change of the migration matrix (models LSS_2 and FIM_2). Population abbreviations are as in Table 1.

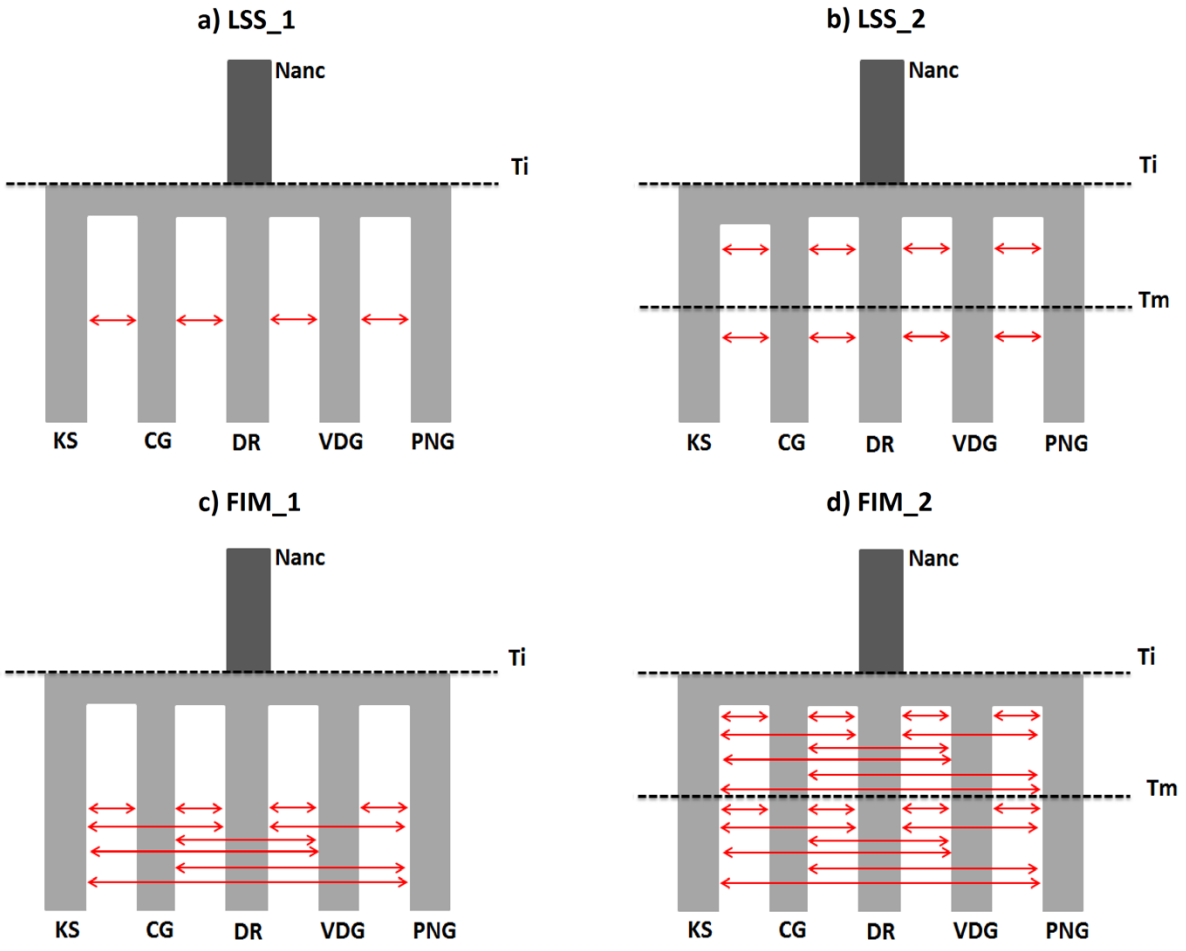

Figure 3: *Glyphis garricki* individual clustering for the subsetted dataset. A) ADMIXTURE ancestry based on posterior membership probabilities. B) DAPC assignment of the subsetted based on posterior membership probabilities.

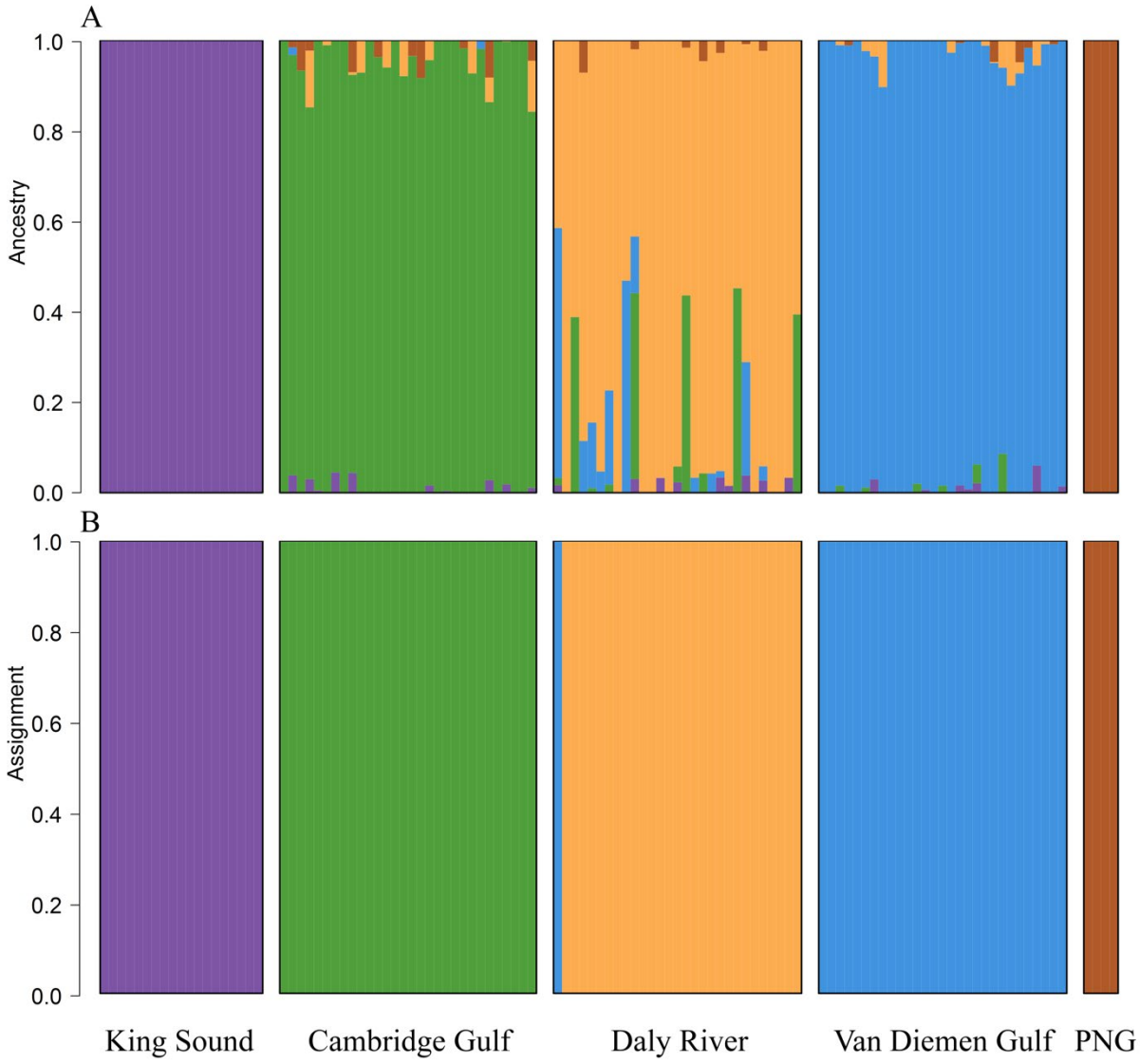

Figure 4: *Glyphis garricki* pairwise log-likelihood ratios in Van Diemen Gulf. Top: all pairwise comparisons, the histogram has been cropped at y = 30 for improved visualisation of kin pair frequencies. Magenta line indicates expected mean for unrelated pairs, orange line indicates expected mean for half-sibling pairs and dash line indicates false-positive cut-off (pairs retained as true kin are on its right side). Orange curve shows the expected distribution of the half-sibling pairs. Bottom: comparisons between pairs retained as true kin. Half-sibling pairs have LLR values below 0, full sibling or parent-offspring pairs have LLR values above 0.

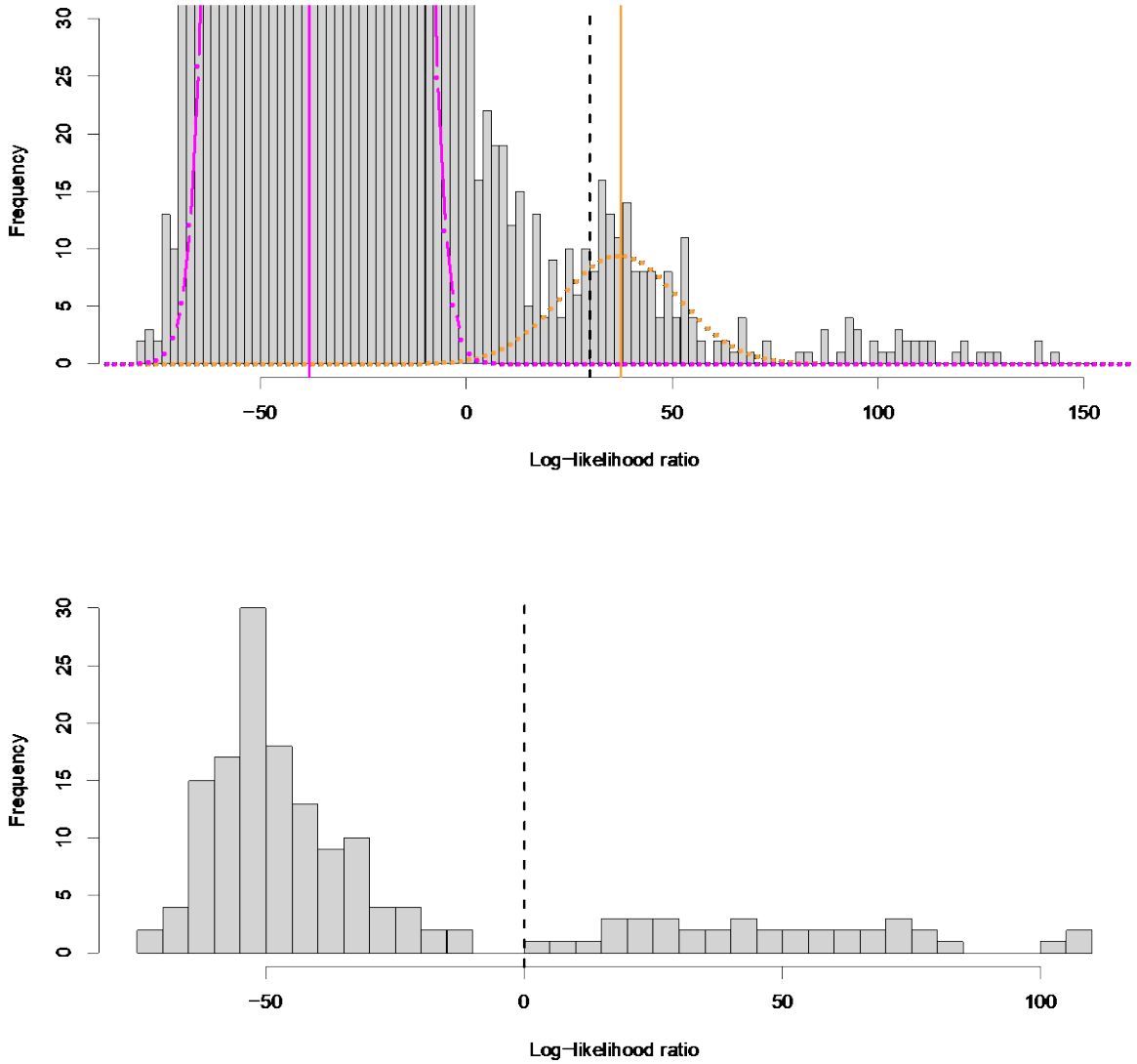

Figure 5: Estimate of *Ne* variation in *G. garricki* through time obtained with the *stairwayplot* method on the unfolded SFS. Maximized composite likelihood for each population and the scatter sample is presented, confidence intervals are reported separately for each population in Fig. S5 and Fig. S6, analysed with and without singletons respectively.