MOLECULAR ECOLOGY

Molecular Ecology (2017) **26**, 444–456 doi: 10.1111/mec.13929

Inferring contemporary and historical genetic connectivity from juveniles

PIERRE FEUTRY,*† OLIVER BERRY,‡ PETER M. KYNE,† RICHARD D. PILLANS,§ RICHARD M. HILLARY,* PETER M. GREWE,* JAMES R. MARTHICK,¶ GRANT JOHNSON,** RASANTHI M. GUNASEKERA,* NICHOLAS J. BAX*†† and MARK BRAVINGTON*

*CSIRO Oceans and Atmosphere, Castray Esplanade, Hobart, TAS 7000, Australia, †Research Institute for the Environment and Livelihoods, Charles Darwin University, Ellengowan Drive, Darwin, NT 0909, Australia, ‡CSIRO Oceans & Atmosphere Indian Ocean Marine Research Centre, The University of Western Australia, M097, 35 Stirling Highway, Crawley, WA 6009, Australia, §CSIRO Oceans and Atmosphere, 41 Boggo Road, Dutton Park, QLD 4102, Australia, ¶Menzies Institute for Medical Research, University of Tasmania, 17 Liverpool Street, Hobart, TAS 7000, Australia, **Department of Primary Industry and Fisheries, Aquatic Resource Research Unit, GPO Box 3000, Darwin , NT 0801, Australia, ††Institute for Marine and Antarctic Science, University of Tasmania, Private Bag 129, Hobart, TAS 7001, Australia

Abstract

Measuring population connectivity is a critical task in conservation biology. While genetic markers can provide reliable long-term historical estimates of population connectivity, scientists are still limited in their ability to determine contemporary patterns of gene flow, the most practical time frame for management. Here, we tackled this issue by developing a new approach that only requires juvenile sampling at a single time period. To demonstrate the usefulness of our method, we used the Speartooth shark (*Glyphis glyphis*), a critically endangered species of river shark found only in tropical northern Australia and southern Papua New Guinea. Contemporary adult and juvenile shark movements, estimated with the spatial distribution of kin pairs across and within three river systems, was contrasted with historical long-term connectivity patterns, estimated from mitogenomes and genome-wide SNP data. We found strong support for river fidelity in juveniles with the within-cohort relationship analysis. Male breeding movements were highlighted with the cross-cohort relationship analysis, and female reproductive philopatry to the river systems was revealed by the mitogenomic analysis. We show that accounting for juvenile river fidelity and female philopatry is important in population structure analysis and that targeted sampling in nurseries and juvenile aggregations should be included in the genomic toolbox of threatened species management.

Keywords: conservation biology, *Glyphis glyphis*, kinship, marine dispersal, philopatry, population structure, threatened species

Received 1 July 2016; revision received 18 October 2016; accepted 15 November 2016

Introduction

It is well established that genetically and demographically isolated populations are more susceptible to irreversible decline than more connected populations that can be buffered by their connectivity (Fahrig & Merriam 1985). Yet, accurately measuring the level of connectivity between populations has challenged fishery and conservation scientists for many decades (Kalinowski 2004). Methodological advances, particularly in genetics (Gagnaire *et al.* 2015), have gotten closer to answering the question fishery and conservation scientists have been asking: are two geographically separated populations of the same species connected such that a decline in one will affect the other? Or conversely, if one population is reduced to undesirable levels, will the other population help restore it within a practical management time frame (typically a few generations)?

Correspondence: Pierre Feutry, Fax: +61 (0) 3 6232 5000; Email: Pierre.Feutry@csiro.au

Genetic markers are commonly used to identify and measure the strength of population boundaries. Conventionally, the extent of population differentiation between spatially or temporally separated samples is evaluated by quantifying differences in allele frequencies with metrics such as F_{ST} and its analogues (Wright 1951, 1965; Weir & Cockerham 1984; Raymond & Rousset 1995). This indirect approach is a relatively powerful way to detect restricted gene flow but is limited to cases where prior knowledge of putative population boundaries is available and can be tested. As larger genetic data sets and more powerful computers have become available, unsupervised clustering algorithms have increasingly been used to provide indirect delineation of population boundaries (Pritchard *et al.* 2000; Dawson & Belkhir 2001; Corander *et al.* 2004; Alexander *et al.* 2009). Broadly speaking, these methods assign individuals to groups that best meet Hardy–Weinberg and gametic-phase disequilibrium expectations (Pritchard *et al.* 2000). Because they are not reliant on a priori-defined population boundaries, they have the potential to detect cryptic population structure, although prior knowledge of population distribution can help detect structure at low levels of divergence (Hubisz *et al.* 2009). Multivariate analyses such as the Discriminant Analysis of Principal Components (DAPC) offer an alternative to computer-intensive unsupervised clustering algorithms when large data sets are involved (Jombart *et al.* 2010).

One limitation of these approaches is the difficulty of interpreting results in a demographic context, as they reflect processes integrated over evolutionary time frames (Kool *et al.* 2013). A further limitation of these approaches is the upward bias in population subdivision caused by family members within samples used to infer population structure. Sampling a large number of progeny from a small pool of reproducing adults can produce an 'Allendorf–Phelps effect', that is, highly significant measures of population differentiation without reproductive isolation (Allendorf & Phelps 1981; Waples 1998). It can also erroneously produce a signal of population subdivision when clustering algorithms, such as the one implemented in the software package STRUCTURE, are used (Anderson & Dunham 2008; Rodriguez-Ramilo & Wang 2012). To minimize this effect, population samples should adequately represent breeders in the putative populations of interest (Allendorf & Phelps 1981). In practice, however, this can be difficult to achieve, as closely related individuals aggregate in many species (Richard *et al.* 1996; Hansen *et al.* 1997; Oremus *et al.* 2012). This is most problematic if juveniles are sampled because they have had fewer opportunities to disperse, or because they obtain benefits from gregariousness (Wilson 1975).

A more recent genetic approach to connectivity is based on the information contained in the spatial distribution of close relatives (Palsbøll 1999). For example, parent–offspring pairs can provide direct estimates of population connectivity (Jones *et al.* 2005; Peery *et al.* 2008; Planes *et al.* 2009; Saenz-Agudelo *et al.* 2009; Christie *et al.* 2010). In contrast to indirect methods, direct estimates offer a clearly defined brief time frame over which to measure spatial processes because the distance between parents and their offspring must accrue between the offspring's birth and capture (Jones *et al.* 2005). Pushing this idea further, Økland *et al.* (2010) provided a method based on the distribution of first- and second-order relatives to define management units. The use of close relatives is particularly useful for characterizing dispersal kernels, identifying the drivers of dispersal in juveniles, to investigate contemporary recruitment dynamics (Cowen & Sponaugle 2009), or to assess population structure on a demographic timescale (Palsbøll *et al.* 2010). However, direct methods typically cannot determine whether dispersing offspring contribute to subsequent generations, or how consistent the observed movements are over the long term. As both contemporary and historical spatial processes are relevant to species management, the simultaneous application of both direct and indirect methods should be a highly desirable approach (Berry *et al.* 2012), particularly if inferences can be made from the same data set.

Recent improvements in sequencing methods now permit the genotyping of hundreds of individuals at thousands of loci (Davey *et al.* 2011) and whole mitogenomes instead of single mitochondrial markers (Feutry *et al.* 2014, 2015). This can benefit both indirect and direct approaches to assessing population connectivity (Palsbøll *et al.* 2010; Gagnaire *et al.* 2015). More markers will for example increase the ability to detect low levels of population differentiation (Waples 1998). The main factor limiting the use of direct estimates of genetic connectivity studies is sampling. Good estimates derived from parent–offspring distribution require the sampling of a significant proportion of the adults and juveniles of each population, which is only possible for small populations with well-defined distributional ranges. With more markers, direct methods can also reveal kinship beyond parent–offspring, potentially removing the need to sample adults (Bravington *et al.* 2016). The spatial distribution of cross-cohort half-sibling pairs for example, provides insight into their parents' breeding movements. Hence, access to adults is not required and sampling can be performed in areas such as nurseries, where juveniles aggregate and boundaries may be understood.

The Speartooth Shark, *Glyphis glyphis* (Carcharhinidae), belongs to a poorly known and highly

threatened group of river sharks, whose taxonomy, distributions, population structure and conservation status are only now beginning to be resolved (Pillans *et al.* 2010; Feutry *et al.* 2014; Li *et al.* 2015; White *et al.* 2015). *Glyphis glyphis* is of high conservation concern and is classified as critically endangered on the Australian *Environment Protection and Biodiversity Conservation Act* 1999. This assessment was mostly based on infrequent collections across a restricted distribution, suggesting low population abundance. Understanding population boundaries and abundance is central to effective management of the species. *Glyphis glyphis* is currently known from three river systems within tropical Australia flowing into Van Diemen Gulf and the Gulf of Carpentaria where they inhabit large tidal river systems, estuaries and coastal environments (Pillans *et al.* 2010; Kyne 2014) (Fig. 1). It is also found in Papua New Guinea (PNG) (White *et al.* 2015).

Until recently, only juveniles and subadults had been observed; the first adults of the species were recorded in 2014 in southern PNG (White *et al.* 2015) and 2015 in Australia (R. D. Pillans, unpublished data). It is suspected that juveniles use rivers as nurseries, whereas adults occur in the marine and coastal zone of northern Australia, possibly entering estuaries and rivers to give birth, as neonates can be reliably found during parturition season from October to December, in upper tidal reaches of rivers (Pillans *et al.* 2010; P. M. Kyne *et al.* unpublished data). Because adults cannot be reliably caught, understanding of the species' biology relies heavily on the study of juveniles (Feutry *et al.* 2014). A recent mitogenomic study suggested female reproductive philopatry in *G. glyphis* (Feutry *et al.* 2014), but the extent of male dispersal remains unknown. Such information is critical to direct management of this threatened species, given its occurrence in only a limited number of river systems. Strong population structure would suggest that management would need to focus at the level of the individual river.

Here, we combined whole mitogenome sequencing and genome scans to investigate the population structure of *G. glyphis*. We infer juvenile and adult contemporary connectivity from the spatial distribution of full- and half-siblings and contrast it with indirect longer term estimates of genetic connectivity to provide management-relevant information on the spatial scale of movement in this threatened species. Specifically, we determine whether juveniles move between river systems (putative populations), whether adults (separately for males and females) breed with adults from more than one river system and the degree of bias in indirect methods caused by the failure to account for familial structure. This is achieved by sampling at a single time period without the need to sample largely inaccessible adults.

Material and methods

Sampling and DNA extraction

Glyphis glyphis samples were collected between January 2012 and December 2014 in the Alligator Rivers system (South Alligator $n = 82$; East Alligator $n = 6$; West Alligator $n = 1$) and the Adelaide River ($n = 142$) of the Northern Territory (NT), and the Wenlock River system ($n = 125$) of Queensland (QLD), northern Australia (Fig. 1). Sharks were caught by rod and line or gillnet. Each shark was measured, sexed and a small fin clip was taken from the inner pectoral fin before it was released at the site of capture. Sampled sharks were

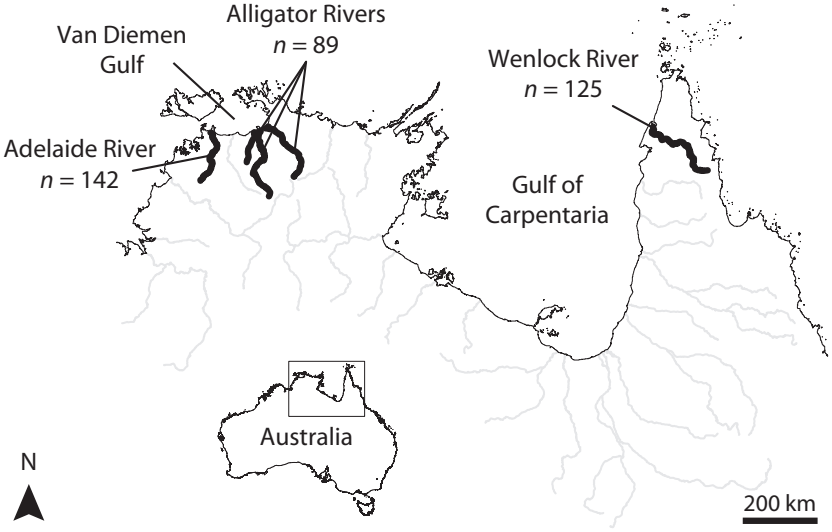

Fig. 1 *Glyphis glyphis* sampling locations and sample size in northern Australia.

from the size range 49–195 cm total length (TL) representing neonates through to subadults. Size at birth is ~50–65 cm TL (Pillans *et al.* 2010) and with a median size of sampled sharks of 82.75 cm TL, most represented juveniles <2 years old. Sharks were sampled under Northern Territory Fisheries Special Permit S17/3252, Kakadu National Park Research Permit RK805, Queensland Fisheries General Research Permit 163582 and Charles Darwin University Animal Ethics Committee A11041. DNA was extracted using either the DNeasy Blood and Tissue kits (Qiagen) or the NucleaMag Tissue kits (Macherey-Nagel).

SNP genotyping

All 356 samples were SNP genotyped. This was performed using DArTseq™, a new implementation of complexity-reduced representations sequencing (Altshuler *et al.* 2000). The protocol used in this study mostly followed that described by Grewe *et al.* (2015), except that to generate more markers, two complexity reduction methods were used, PstI-SphI and PstI-NspI, instead of one. The SNP calling was performed with DArT PLD's proprietary software DArTsoft14. DArTsoft14 uses scoring consistency derived from technical sample replicates (i.e. samples processed twice from DNA library preparation to SNP calling) to optimize its algorithm parameters (Grewe *et al.* 2015).

SNP filtering

The data set used for population analysis consisted of 75-bp fragments containing one or more SNPs. Prior to population analysis, loci were further screened by excluding loci not scored for all individuals (i.e. Call rate = 1), reproducibility lower than 0.99 (approximately 10% of the individuals were genotyped twice, and the reproducibility represented the proportion of the replicate pairs for which the genotyping is consistent), with average sequencing depth lower than 10× and with overall minor allele frequency (MAF) lower than 0.02. When multiple polymorphisms remained on the same 75-bp fragment (i.e. on the same cluster), a single SNP was randomly chosen to represent that fragment avoiding linkage disequilibrium between very close loci.

We used the F_{ST} outlier approach developed by Beaumont & Nichols (1996) as implemented in LOSITAN (Antao *et al.* 2008) and the R package OutFLANK (Whitlock & Lotterhos 2015) to identify outlier loci putatively under the influence of directional selection. The approach implemented in OutFLANK is based on an improved method for deriving the null distribution of population differentiation for neutral loci. It results in fewer false positives than other outlier tests, which are more influenced by the effects of demographic history (Lotterhos & Whitlock 2015). We ran OutFLANK with 5% left and right trim for the null distribution of F_{ST}, minimum heterozygosity for loci of 0.1, and a 5% false discovery rate (*q* value).

Mitogenome sequencing

The mitogenomes of 92 *G. glyphis* included in this study were sequenced as part of previous work (Feutry *et al.* 2014). Another 81 were amplified and sequenced following the same protocol (GenBank Accession, KY039188-KY039268). In short, the 173 mitogenomes were amplified in two overlapping fragments. The PCR products were then purified with Agencourt AMPure XP magnetic beads (Beckman Coulter) and prepared with Nextera XT DNA Sample Preparation kits (Illumina) for sequencing on a Miseq (Illumina). Reads were trimmed, filtered and mapped onto the reference sequence (Chen *et al.* 2014) using default parameters for the low sensitivity and no fine-tuning options in GENEIOUS PRO (v. 8.1.7).

Sibship analyses and fish filtering

COLONY (v. 2.0.5.8) (Jones & Wang 2010) was used to identify full-sibling (FS) and half-sibling (HS) relationships from the nuclear DNA data. Analysis parameters are provided in Appendix S1 (Supporting information). After assessment of the probability distribution (Appendix S2, Supporting information), only pairs of FS or HS with probabilities above 0.95 were considered true sibships. Cross-cohort HS were determined by comparing capture dates and fish length to growth rate estimates derived from recaptures.

To address potential bias from family sampling (Allendorf & Phelps 1981; Anderson & Dunham 2008), identical population analyses were carried out on both the all individuals (ALL) and without FS or HS (NoSib) sample sets. To create the NoSib data set, one individual from each sibling pair was randomly discarded from the ALL data set. When some individuals belong to more than one pair of FS or HS, those discarded were chosen so as to maximize the number of individuals preserved.

Population structure analysis

ARLEQUIN (v. 3.5.1.3) was used to calculate pairwise fixation indexes (Φ_{ST}) between each pair of rivers and test for reproductive female philopatry. Tamura-Nei was used as the model of nucleotide evolution in the AMOVA and to calculate Φ_{ST} values. Contemporary

female reproductive philopatry was tested using an approximate likelihood ratio test based on cross-cohort HS mitogenome haplotypes. Details for this test are provided in Appendix S3 (Supporting information).

Pairwise F_{ST} (Weir & Cockerham 1984) and associated p-values were derived from the SNP data using the R package STAMPP and 10 000 bootstraps (Pembleton *et al.* 2013).

To further evaluate whether the nuclear genetic variation was partitioned geographically, a model-based clustering approach was used as implemented in STRUCTURE (v. 2.3.4) (Pritchard *et al.* 2000). STRUCTURE analyses were performed on the CSIRO Accelerator Cluster 'Bragg', which consists of 128 Dual Xeon 8-core E5-2650 compute nodes. STRUCTURE seeks to group individuals in such a way that the groups maximize conformity to Hardy–Weinberg and linkage equilibrium. We ran STRUCTURE across values for *K* (number of clusters) between 1 and 8, and evaluated the fit of the data to different values of *K*. The fits of alternative models were evaluated with the Delta *K* method (Evanno *et al.* 2005) implemented in CLUMPAK (Kopelman *et al.* 2015) and based on 20 independent runs for each value of *K*. All runs incorporated a 200 000 iterations burn-in followed by 500 000 clustering iterations. We ensured the adequacy of the run length by checking the runtime likelihood and alpha for stability. For all runs, we assumed that allele frequencies were correlated between sampling sites and allowed for admixture. All runs were completed with and without inclusion of prior location information (LOCPRIOR).

Finally, the genetic structure was analysed with a Discriminant Analysis of Principle Components (DAPC), as implemented in the R package ADEGENET (Jombart 2008; Jombart *et al.* 2010). In the first DAPC analysis, the *K*-means method was used to identify the optimal number of clusters in the data. In the second DAPC analysis, a priori grouping based on sampling locations was investigated. Cross-validation, with 30 replicates and a 90/10 ratio for the training/validation sets, was used as an optimization procedure to select the adequate number of principal components to retain in the analysis.

Results

SNP filtering

The DArTsoft14 pipeline delivered 2191 and 1944 SNPs for the PstI-SphI and PstI-NspI complexity reduction methods, respectively (Appendix S4, Supporting information). These SNPs were then combined into a single SNP data set for quality filtering and analysis. A total of 1330 SNPs passed all quality control filtering steps (Appendix S5, Supporting information). No outlier SNP was detected using either LOSITAN or OUTFLANK. Descriptive statistics including allelic richness (A_R), observed heterozygosity (H_o), expected heterozygosity (H_e) and inbreeding coefficient (F_{is}) are given in Appendix S6 (Supporting information).

Distribution of sib pairs

A total of 72 FS pairs (94 unique individuals) were identified, of which 12, 11 and 49 originated from the Adelaide River, Alligator rivers and Wenlock Rivers, respectively. No cross-river FS pairs were identified. All FS pairs identified in the Alligator rivers system were caught in the South Alligator River. A total of 145 HS pairs (179 unique individuals) were identified, 44 within the Adelaide River, 14 within the Alligator rivers, 69 within the Wenlock River and 18 split across the Adelaide and Alligator Rivers (Table 1). Most likely, these 18 HS pairs were paternally related (Appendix S3, Supporting information). Within the Alligator system, one HS pair was split across the East and the South Alligator Rivers.

Growth rate derived from recapture data ranged from 18.2 to 36.5 cm year^{-1} for fish smaller than 85 cm TL (n = 4) and from 6.3 to 7.4 cm year^{-1} for fish larger than 85 cm TL (n = 2). Fish from 18 HS pairs with length differences less than 7 cm were captured fewer than 150 days apart and classified as same cohort. Fish from another six pairs of HS with length differences ranging 14–19 cm and captured between 200 and 400 days apart were also classified as same cohort. None of these 24 same-cohort HS pairs had fish captured in different rivers. Given the amount of time between captures, the length difference and the growth rate observed, fish from all other HS pairs were unlikely to be born at the same time and were thus considered cross-cohort (Table 1).

Population structure

Measures of population differentiation based on whole mitogenomes and nuclear SNPs are given in Tables 2

Table 1 Intra- and inter-river number of full-sibling pairs + cross-cohort half-sibling pairs + same-cohort half-sibling pairs

Rivers	Adelaide N = 142	Alligators N = 89	Wenlock N = 125
Adelaide	12 + 42 + 2		
Alligators	0 + 18 + 0	11 + 9 + 5	
Wenlock	0 + 0 + 0	0 + 0 + 0	49 + 52 + 17

Table 2 Mitogenome-based pairwise Φ_{ST} for all individuals (above; Adelaide $N = 74$, Alligators $N = 60$, Wenlock $N = 15$) and the data set without full-sibling and half-sibling pairs (below; Adelaide $N = 41$, Alligators $N = 35$, Wenlock $N = 14$)

Rivers	Adelaide	Alligators	Wenlock
Adelaide		0.24705**	0.70517**
Alligators	0.24352**		0.23768*
Wenlock	0.67673**	0.25928*	

*P-value <0.01; **P-value <0.0001.

and 3, respectively. All pairwise mitogenome-based measures of population differentiation were statistically significant, independent of whether the FS and HS were included in the analyses or not (Table 2). Private haplotypes were found in each river, but at least one haplotype per river was found at another sampling site (Appendix S7, Supporting information). Population differentiation between each river pair was also supported by nuclear SNPs, except for the Adelaide and Alligator Rivers after the FS and HS were discarded (Table 3). SNP-based pairwise F_{ST} were higher for the ALL data set than the NoSib data set. The F_{ST} between Adelaide and Alligator Rivers was an order of magnitude lower and became nonsignificant, whereas F_{ST} between Adelaide/Alligator and Wenlock Rivers decreased by a factor of about two but remained significantly different from zero (Table 3).

Only the ALL data set showed clear evidence of genetic differentiation among the rivers, and this was manifest as a division between Adelaide/Alligator Rivers and Wenlock River. The delta K analysis indicated $K = 7$ as the best fit (delta $K = 2.61$), but five small clusters consisted of full and half-siblings (Fig. 2a). These results were consistent whether location priors were included or not (Appendix S8, Supporting information). The only signal of population structure remaining in the NoSib data set was the distribution of q-values at $K = 2$, which distinguished Wenlock samples from Adelaide and Alligator samples when location information was included as prior (Fig. 2b). This signal disappeared when location priors were not included in the analyses. L(K) was stable and did not support $K = 2$ as the best fit whether the location information was included as prior or not (Appendix S8, Supporting information).

Table 3 Nuclear SNP-based pairwise F_{ST} for all individuals (above; Adelaide $N = 142$, Alligators $N = 89$, Wenlock $N = 125$) and the data set without full-sibling and half-sibling pairs (below; Adelaide $N = 99$, Alligators $N = 58$, Wenlock $N = 59$)

Rivers	Adelaide	Alligators	Wenlock
Adelaide		0.00095**	0.00458**
Alligators	0.00008^{NS}		0.00493**
Wenlock	0.00279**	0.00285**	

^{NS}P-value >0.05; **P-value <0.0001.

K-mean-based DAPC analyses did not suggest the presence of any substructure in either the ALL or the NoSib data sets (Appendix S9, Supporting information). DAPC analyses based on a priori grouping supported the same structure pattern as the F_{ST} analyses. Clear evidence of genetic heterogeneity was found between each river system in the ALL data set (Fig. 3a), whereas only two distinct gene pools remained in the NoSib data set, one in NT and one in QLD (Fig. 3b).

Discussion

For the first time in any elasmobranch species, whole mitogenome sequencing and genotyping-by-sequencing genome scans have been used in combination to characterize population connectivity at multiple spatial and temporal scales. Our results reveal that a significant fraction of the *G. glyphis* individuals analysed from all three rivers were close relatives (26% FS; 50% HS). Their spatial distribution permits direct estimation of contemporary sex-specific adult (breeding) and juvenile movements in this threatened species. In addition, the identification of kin means that historical connectivity estimated from population subdivision can be made from juveniles only, without the family sampling bias that may occur in population genetic data sets (Allendorf & Phelps 1981).

Direct estimate of contemporary connectivity

The spatial distribution of FS pairs has previously been used to infer the movements of juvenile fishes. This is the first time sex-specific adult movements are inferred from the spatial distribution of juvenile HS pairs. This is a valuable contribution for connectivity studies, and for threatened species in particular, where adults are rare and/or not easily sampled. In the case of *G. glyphis*, only two adults have been caught in Australia as part of a scientific study (R. D. Pillans, unpublished data).

We identified over 200 *G. glyphis* full- and half-sibling pairs with a high degree of certainty (Appendix S2, Supporting information), made possible by the large number of SNP loci analysed. Full-sibling pairs were only captured within the same river suggesting that juveniles remain in the natal river for some time. Because age and growth data are not available for *G. glyphis*, the age-at-length of juvenile Bull Shark *Carcharhinus leucas* reported by Tillett *et al.* (2011) is the best proxy available. These two species are sympatric in

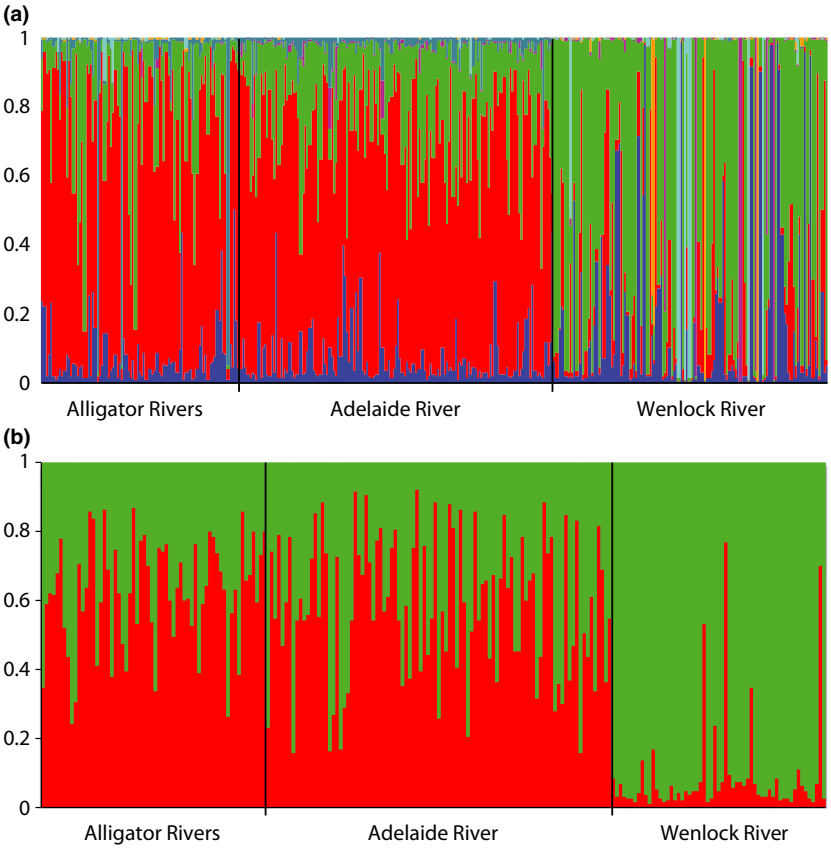

Fig. 2 *Glyphis glyphis* STRUCTURE admixture analysis. Each cluster (*K*) is designated by a different colour. Each vertical bar represents one individual, partitioned according to estimated membership to each cluster. (a) Analysis of data set with all samples, most likely $K = 7$. (b) Analysis of data set without full-sibling and half-sibling pairs, most likely $K = 7$.

northern Australian rivers, have a similar life history including the use of river systems as nursery areas and have similar size at birth and maximum sizes. The largest *G. glyphis* full-sibling identified in the current study was thus estimated to be 6 years old, suggesting that the use of river nurseries last several years for juveniles. Age data for *G. glyphis* would be required to estimate more accurately the extent of their presence in natal rivers.

Extended residency within the limited spatial habitat of these natal rivers may increase susceptibility to anthropogenic impacts. However, neither of the NT river systems in this study has commercial line or net fisheries, and therefore, pressure is greatly reduced in comparison with some adjacent coastal areas. In Queensland, commercial net and crab fisheries, which are known to capture juvenile *G. glyphis*, overlap with the species distribution in the Wenlock River system as well as in coastal environments. The extent of capture of juveniles in rivers by recreational fishers is unknown, but illegal captures of this protected species have been recorded in the NT (P. M. Kyne and P. Feutry, unpublished data) and Queensland (R. D. Pillans, unpublished data). Furthermore, the scale of Indigenous harvest is unknown. Future plans for further agricultural development of northern Australia and associated increased water demand (Australian Government 2015) will likely have implications for the riverine habitats of this, and other, threatened species.

Given that the juveniles do not or very rarely move between rivers, the distribution of HS provides insight into the movements of adults between reproductive events. Of the 121 cross-cohort HS pairs, 103 (85%) were captured within the same river system, indicating that in most cases, at least one parent returned to reproduce in the same river system across breeding seasons. Despite very limited sample sizes for the East and West Alligator Rivers, one HS pair split across the East and South Alligator Rivers was found, demonstrating parental movement within the Alligator Rivers system (straight-line distance between river mouths c. 15 km). The remaining 18 (15%) cross-cohort HS pairs were shared between the Adelaide and the Alligator Rivers. In these cases, at least one parent, most likely the male from the cross-cohort HS mitogenome haplotype analysis (Appendix S3, Supporting information), had moved between these rivers (or their associated mating aggregation areas if gamete exchange occurs outside the river) to reproduce. Van Diemen Gulf is a relatively small and shallow system, and it is possible that adults from different rivers flowing in the gulf mix in this area. In contrast to the cross-cohort HS pairs,

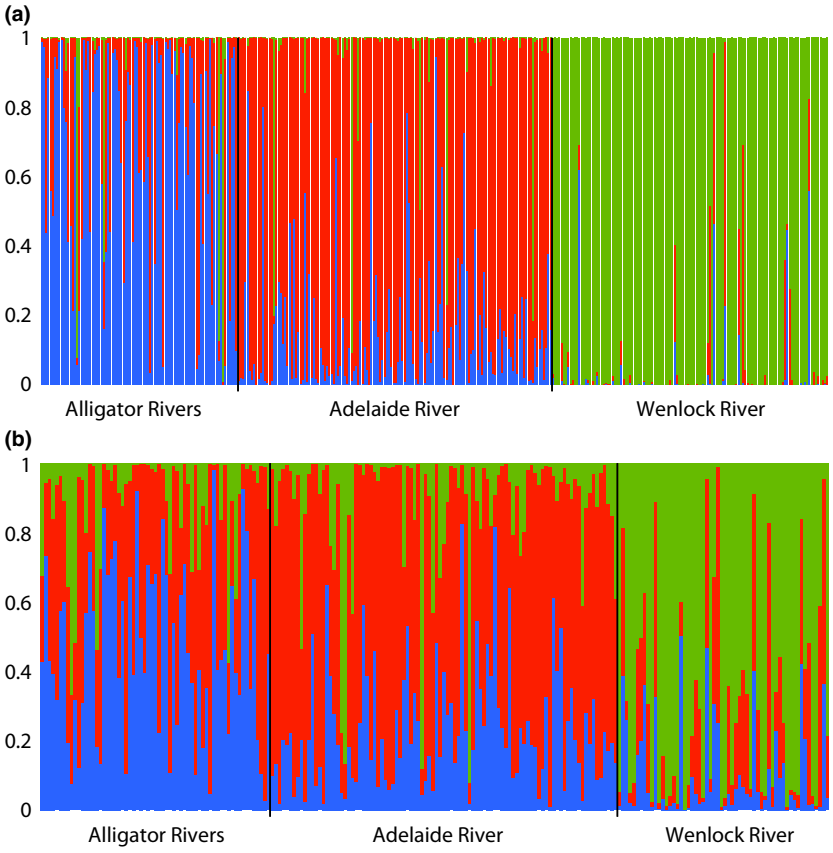

Fig. 3 *Glyphis glyphis* DAPC analysis. Each vertical bar represents one individual, partitioned according to membership probability for each a priori-defined group. (a) Analysis of data set with all samples. (b) Analysis of data set without full-sibling and half-sibling pairs.

same-cohort HS pairs were never captured between rivers. Assuming females only breed once a year, this suggests that males do not reproduce with females going to pup in different rivers within the same year. Based on the variability in reproductive periodicity of Australian carcharhinids of similar or smaller size, minimum reproductive periodicity would be annual (Chin *et al.* 2013; Harry *et al.* 2013), or potentially biennial given large size at maturity (Mcauley *et al.* 2007). Hence, it is likely that the Adelaide and Alligator Rivers' populations have different mating aggregation areas. Once fish can be aged accurately, reproductive periodicity could be determined by examining the time gap between HS pairs.

It is significant that no cross-cohort HS pairs were shared between the Alligator/Adelaide Rivers emptying into Van Diemen Gulf and the more distant Wenlock River emptying into the eastern Gulf of Carpentaria. Adult breeding movements on scales of ~150 km therefore seem commonplace in *G. glyphis*, but nonexistent or very rare over distance an order of magnitude higher.

Population structure when sampling families

Previously, whole mitogenome sequencing of *G. glyphis* had revealed female philopatry (Feutry *et al.* 2014) which is common in sharks (Dudgeon *et al.* 2012), but had not provided insight into the movements of males, nor been able to discount the effects of sampling kin. Nuclear markers provide the ability to take the understanding of population structure of *G. glyphis* a step further because they reflect both male- and female-mediated gene flow. In addition, they permit identification of kin, whose presence has the potential to drive an upward bias in apparent population subdivision (Allendorf & Phelps 1981; Anderson & Dunham 2008), including in a previous study on *G. glyphis* by Feutry *et al.* (2014). In all finite-sized populations, there is a real possibility of randomly sampling related individuals, especially in small populations of rare and threatened species. If these sampling events were independent, then excluding one individual of each pair prior to population structure analysis would be incorrect. In the present case, the removal of FS and HS pairs is justified given the limited dispersal abilities of juveniles; the sampling of two close relatives in each river does not represent independent events. Doing so did not greatly affect pairwise fixation indexes Φ_{ST} (average <5% absolute difference in Φ_{ST}), demonstrating the population differentiation observed in the mitochondrial DNA was due to female reproductive philopatry and not bias from family sampling.

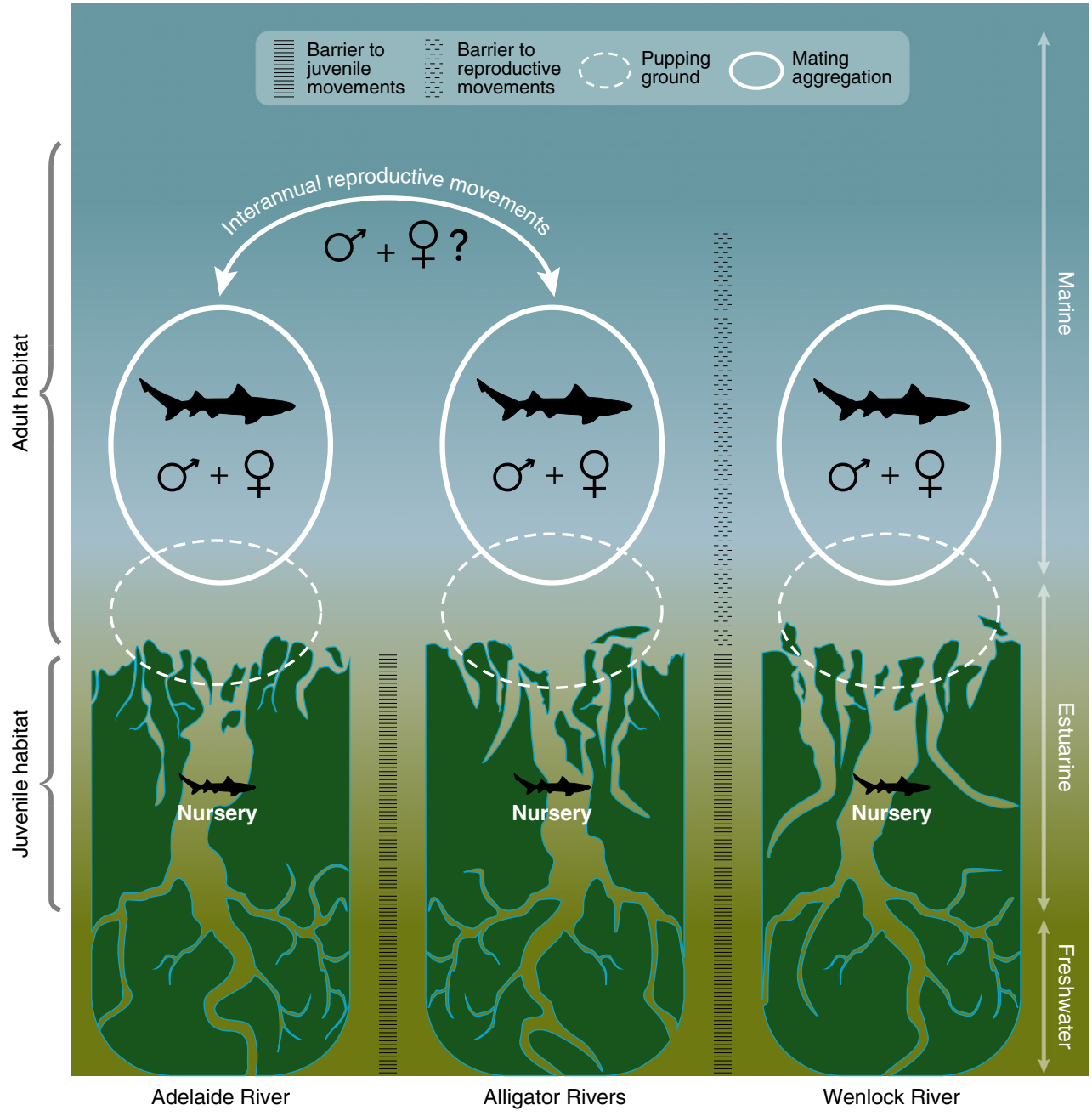

Fig. 4 Schematic representation of *Glyphis glyphis* movements as inferred from spatial distribution of full- and half-sibling pairs, and population structure analyses of whole mitogenomes and nuclear genome scans.

In contrast to the mitogenome data, the presence of close relatives in the nuclear SNP data set substantially increased the signal of population subdivision revealed by F_{ST}, STRUCTURE and DAPC analyses. Most likely, there are only two distinct gene pools in Australian *G. glyphis* over the long term, one in NT and one in QLD. All additional grouping reflects family rather than population structure. The presence of FS and HS in the sample created an upwards bias in the estimation of F_{ST} between the Adelaide and Alligator Rivers and an overestimation of the number of populations identified by STRUCTURE, as predicted by Allendorf & Phelps (1981) and Anderson & Dunham (2008), respectively. The significant population differentiation initially identified in the ALL data set between the Adelaide and Alligator Rivers was due entirely to allele frequency bias from FS and HS. Similarly, the STRUCTURE analysis overestimated the sample partitioning with groups of FS and HS forming independent clusters (Anderson & Dunham 2008), and this bias was also evident in the DAPC analysis.

The barrier to gene flow between Wenlock and Adelaide/Alligator Rivers was more evident in F_{ST} and DAPC analyses compared with STRUCTURE. This highlights the limited ability of the STRUCTURE clustering compared to a priori grouping-based methods to detect subtle levels of genetic differentiation (Waples & Gaggiotti 2006).

The contrast between nuclear and mtDNA markers indicates sex-biased dispersal, with males' reproductive movements greatly exceeding those of females. Sex-biased dispersal has previously been reported in sharks (Pardini *et al.* 2001; Daly-Engel *et al.* 2012) and has important implications for management. Daly-Engel *et al.* (2012) noted that the use of female or biparentally inherited loci individually can mislead conclusions with regard to management units. While mitochondrial markers showed structuring between the Adelaide and Alligator Rivers, the use of nuclear SNP loci indicated that these rivers are part of the same gene pool. Importantly, as females exhibit river-specific reproductive philopatry, this gene flow could not compensate for the loss of females from a specific river, so the female population of each river still needs to be managed as though it is an isolated population. The Van Diemen Gulf population should be managed as a separate unit to the isolated Wenlock River population. The relationship of these populations to the species in Papua New Guinea (PNG) also needs to be examined.

Direct versus indirect connectivity estimates and management implications

Both direct and indirect estimates of population connectivity support the Adelaide and Alligator Rivers as part of the same nuclear gene pool, whereas the Wenlock River likely has a strong degree of demographic independence, at least for the generation of adults who produced the juveniles included in this study. Direct estimates of connectivity have two main advantages over indirect methods. The first one is a known time frame for the movements; FS and same-cohort HS providing information for the current generation of juveniles. The exact period of time covered depends on the age of the juveniles. Given appropriate sampling, potential between river movements could be inferred for each year class. Cross-cohort HS provide sex-specific information about their parents' movements between breeding events and has the potential to reveal very recent weak barriers to gene flow.

The second advantage is the information about contemporary migration rates between populations that lies in the distribution of HS pairs although an appropriate statistical framework remains to be developed in order to make use of it (Palsbøll *et al.* 2010). In the nonspatial context, Bravington *et al.* (2016) have outlined how these data can be used to estimate sex-specific abundance and survival rates in a modified mark–recapture framework called close-kin mark–recapture (CKMR). An extension of this framework into the spatial domain would utilize the migratory- and abundance-related information in these data to separate the two, and obtain quantitative estimates of between river migration rates.

Conclusion

In this study, we demonstrated how combined information from direct and indirect connectivity estimates can be used to detect historical and intergenerational between-river movements and mating and breeding patterns from a single contemporaneous sample of juveniles only. This represents a significant addition to the toolbox of threatened species management. For *G. glyphis* in northern Australia: (i) juveniles do not move between river systems during riverine residencies (possibly >6 years); (ii) females predominantly return to a single river to pup; but, (iii) reproducing males likely move between breeding aggregations for river systems closer than 150 km apart (although data on where breeding aggregations occur are lacking) (Fig. 4). This has implications for the conservation of this critically endangered species, in both the management and potential mitigation of increasing demands on their environment.

Acknowledgements

This study was supported by the Marine Biodiversity Hub, a collaborative partnership supported through funding from the Australian Government's National Environmental Research Program (NERP). PF was partly supported by the North Australia Marine Research Alliance (NAMRA). PMK was partly supported by the NERP Northern Australia Hub. We thank the Traditional Owners and staff of Kakadu National Park, especially Anne O'Dea and Khan Spokes, and Northern Territory Department of Primary Industry and Fisheries, especially Thor Saunders and Mark Grubert for field assistance. The contribution of a large number of field volunteers is also acknowledged. Kate Buckley, Karen Devitt, Rhian Evans, Sharon Every, Charlotte Klempin and Roy Tipiloura, in particular provided significant field assistance. Thanks to CSIRO Scientific Computing for the use of the Bragg Cluster and to Bob Ward and the three anonymous reviewers who significantly contributed to improving the quality of this manuscript.

References

Alexander DH, Novembre J, Lange K (2009) Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*, **19**, 1655–1664.

Allendorf FW, Phelps SR (1981) Use of allelic frequencies to describe population structure. *Canadian Journal of Fisheries and Aquatic Sciences*, **38**, 1507–1514.

Altshuler D, Pollara VJ, Cowles CR *et al.* (2000) An SNP map of the human genome generated by reduced representation shotgun sequencing. *Nature*, **407**, 513–516.

Anderson EC, Dunham KK (2008) The influence of family groups on inferences made with the program Structure. *Molecular Ecology Resources*, **8**, 1219–1229.

Antao T, Lopes A, Lopes RJ, Beja-Pereira A, Luikart G (2008) LOSITAN: a workbench to detect molecular adaptation based on a Fst-outlier method. *BMC Bioinformatics*, **9**, 323.

Australian Government (2015) Our north, our future: white paper on developing Northern Australia. Available at: http://industry.gov.au/ONA/WhitePaper/Documents/northern_australia_white_paper.pdf.

Beaumont MA, Nichols RA (1996) Evaluating loci for use in the genetic analysis of population structure. *Proceedings of the Royal Society of London, Series B: Biological Sciences*, **263**, 1619–1626.

Berry O, England P, Marriott RJ, Burridge CP, Newman SJ (2012) Understanding age-specific dispersal in fishes through hydrodynamic modelling, genetic simulations and microsatellite DNA analysis. *Molecular Ecology*, **21**, 2145–2159.

Bravington MV, Skaug HJ, Anderson EC (2016) Close-kin mark-recapture methods. *Statistical Science*, **31**, 259–274.

Chen X, Liu M, Grewe PM, Kyne PM, Feutry P (2014) Complete mitochondrial genome of the Critically Endangered speartooth shark *Glyphis glyphis* (Carcharhiniformes: Carcharhinidae). *Mitochondrial DNA*, **25**, 431–432.

Chin A, Simpfendorfer C, Tobin A, Heupel M (2013) Validated age, growth and reproductive biology of *Carcharhinus melanopterus*, a widely distributed and exploited reef shark. *Marine and Freshwater Research*, **64**, 965–975.

Christie MR, Tissot BN, Albins MA *et al.* (2010) Larval connectivity in an effective network of marine protected areas. *PLoS ONE*, **5**, e15715.

Corander J, Waldmann P, Marttinen P, Sillanpää MJ (2004) BAPS 2: enhanced possibilities for the analysis of genetic population structure. *Bioinformatics*, **20**, 2363–2369.

Cowen RK, Sponaugle S (2009) Larval dispersal and marine population connectivity. *Marine Science*, **1**, 443–466.

Daly-Engel TS, Seraphin KD, Holland KN *et al.* (2012) Global phylogeography with mixed-marker analysis reveals male-mediated dispersal in the endangered scalloped hammerhead shark (*Sphyrna lewini*). *PLoS ONE*, **7**, e29986.

Davey JW, Hohenlohe PA, Etter PD, *et al.* (2011) Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nature Reviews: Genetics*, **12**, 499–510.

Dawson KJ, Belkhir K (2001) A Bayesian approach to the identification of panmictic populations and the assignment of individuals. *Genetical Research*, **78**, 59–77.

Dudgeon CL, Blower DC, Broderick D *et al.* (2012) A review of the application of molecular genetics for fisheries management and conservation of sharks and rays. *Journal of Fish Biology*, **80**, 1789–1843.

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology*, **14**, 2611–2620.

Fahrig L, Merriam G (1985) Habitat patch connectivity and population survival. *Ecology*, **66**, 1762–1768.

Feutry P, Kyne PM, Pillans RD *et al.* (2014) Mitogenomics of the Speartooth Shark challenges ten years of control region sequencing. *BMC Evolutionary Biology*, **14**, 232.

Feutry P, Kyne PM, Pillans RD *et al.* (2015) Whole mitogenome sequencing refines population structure of the Critically Endangered sawfish *Pristis pristis*. *Marine Ecology Progress Series*, **533**, 237–244.

Gagnaire PA, Broquet T, Aurelle D *et al.* (2015) Using neutral, selected, and hitchhiker loci to assess connectivity of marine populations in the genomic era. *Evolutionary Applications*, **8**, 769–786.

Grewe PM, Feutry P, Hill PL *et al.* (2015) Evidence of discrete yellowfin tuna (*Thunnus albacares*) populations demands rethink of management for this globally important resource. *Scientific Reports*, **5**, 16916.

Hansen MM, Nielsen EE, Mensberg KLD (1997) The problem of sampling families rather than populations: relatedness among individuals in samples of juvenile brown trout *Salmo trutta* L. *Molecular Ecology*, **6**, 469–474.

Harry AV, Tobin AJ, Simpfendorfer CA (2013) Age, growth and reproductive biology of the spot-tail shark, *Carcharhinus sorrah*, and the Australian blacktip shark, *C. tilstoni*, from the Great Barrier Reef World Heritage Area, north-eastern Australia. *Marine and Freshwater Research*, **64**, 277–293.

Hubisz MJ, Falush D, Stephens M, Pritchard JK (2009) Inferring weak population structure with the assistance of sample group information. *Molecular Ecology Resources*, **9**, 1322–1332.

Jombart T (2008) ADEGENET: a R package for the multivariate analysis of genetic markers. *Bioinformatics*, **24**, 1403–1405.

Jombart T, Devillard S, Balloux F (2010) Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genetics*, **11**, 94.

Jones OR, Wang J (2010) COLONY: a program for parentage and sibship inference from multilocus genotype data. *Molecular Ecology Resources*, **10**, 551–555.

Jones GP, Planes S, Thorrold SR (2005) Coral reef fish larvae settle close to home. *Current Biology*, **15**, 1314–1318.

Kalinowski ST (2004) Genetic polymorphism and mixed-stock fisheries analysis. *Canadian Journal of Fisheries and Aquatic Sciences*, **61**, 1075–1082.

Kool JT, Moilanen A, Treml EA (2013) Population connectivity: recent advances and new perspectives. *Landscape Ecology*, **28**, 165–185.

Kopelman NM, Mayzel J, Jakobsson M, Rosenberg NA, Mayrose I (2015) CLUMPAK: a program for identifying clustering modes and packaging population structure inferences across K. *Molecular Ecology Resources*, **15**, 1179–1191.

Kyne PM (2014) Threatened fishes and marine turtles of Kakadu National Park (with notes on marine mammals). In: *Kakadu National Park Landscape Symposia Series. Symposium 7: Conservation of Threatened Species, 26–27 March 2013, Bowali Visitor Centre, Kakadu National Park* (eds. Winderlich S, Woinarski J), pp. 58–74. Internal Report 623, Supervising Scientist, Darwin.

Li C, Corrigan S, Yang L *et al.* (2015) DNA capture reveals transoceanic gene flow in endangered river sharks. *Proceedings of the National Academy of Sciences*, **112**, 13302–13307.

Lotterhos KE, Whitlock MC (2015) The relative power of genome scans to detect local adaptation depends on sampling

design and statistical method. *Molecular Ecology*, **24**, 1031–1046.
Mcauley RB, Simpfendorfer C, Hyndes G, Lenanton R (2007) Distribution and reproductive biology of the sandbar shark, *Carcharhinus plumbeus* (Nardo), in Western Australian waters. *Marine and Freshwater Research*, **58**, 116–126.
Økland J-M, Haaland ØA, Skaug HJ (2010) A method for defining management units based on genetically determined close relatives. *ICES Journal of Marine Science: Journal du Conseil*, **67**, 551–558.
Oremus M, Poole MM, Albertson GR, Baker CS (2012) Pelagic or insular? Genetic differentiation of rough-toothed dolphins in the Society Islands, French Polynesia. *Journal of Experimental Marine Biology and Ecology*, **432–433**, 37–46.
Palsbøll PJ (1999) Genetic tagging: contemporary molecular ecology. *Biological Journal of the Linnean Society*, **68**, 3–22.
Palsbøll PJ, Zachariah Peery M, Berube M (2010) Detecting populations in the 'ambiguous' zone: kinship-based estimation of population structure at low genetic divergence. *Molecular Ecology Resources*, **10**, 797–805.
Pardini AT, Jones CS, Noble LR *et al.* (2001) Sex-biased dispersal of great white sharks. *Nature*, **412**, 139–140.
Peery MZ, Beissinger SR, House RF *et al.* (2008) Characterizing source–sink dynamics with genetic parentage assignments. *Ecology*, **89**, 2746–2759.
Pembleton LW, Cogan NO, Forster JW (2013) STAMPP: an R package for calculation of genetic differentiation and structure of mixed-ploidy level populations. *Molecular Ecology Resources*, **13**, 946–952.
Pillans RD, Stevens JD, Kyne PM, Salini J (2010) Observations on the distribution, biology, short-term movements and habitat requirements of river sharks *Glyphis spp.* in northern Australia. *Endangered Species Research*, **10**, 321–332.
Planes S, Jones GP, Thorrold SR (2009) Larval dispersal connects fish populations in a network of marine protected areas. *Proceedings of the National Academy of Sciences*, **106**, 5693–5697.
Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945–959.
Raymond M, Rousset F (1995) An exact test for population differentiation. *Evolution*, **49**, 1280–1283.
Richard KR, Dillon MC, Whitehead H, Wright JM (1996) Patterns of kinship in groups of free-living sperm whales (*Physeter macrocephalus*) revealed by multiple molecular genetic analyses. *Proceedings of the National Academy of Sciences*, **93**, 8792–8795.
Rodriguez-Ramilo ST, Wang J (2012) The effect of close relatives on unsupervised Bayesian clustering algorithms in population genetic structure analysis. *Molecular Ecology Resources*, **12**, 873–884.
Saenz-Agudelo P, Jones GP, Thorrold SR, Planes S (2009) Estimating connectivity in marine populations: an empirical evaluation of assignment tests and parentage analysis under different gene flow scenarios. *Molecular Ecology*, **18**, 1765–1776.
Tillett BJ, Meekan MG, Field IC, Hua Q, Bradshaw CJ (2011) Similar life history traits in bull (*Carcharhinus leucas*) and pigeye (*C. amboinensis*) sharks. *Marine and Freshwater Research*, **62**, 850–860.
Waples RS (1998) Separating the wheat from the chaff: patterns of genetic differentiation in high gene flow species. *Journal of Heredity*, **89**, 438–450.
Waples RS, Gaggiotti O (2006) What is a population? An empirical evaluation of some genetic methods for identifying the number of gene pools and their degree of connectivity. *Molecular Ecology*, **15**, 1419–1439.
Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population structure. *Evolution*, **38**, 1358–1370.
White WT, Appleyard SA, Sabub B et al. (2015) Rediscovery of the threatened river sharks, *Glyphis garricki* and *G. glyphis*, in Papua New Guinea. *PLoS ONE*, **10**, e0140075.
Whitlock MC, Lotterhos KE (2015) Reliable detection of loci responsible for local adaptation: inference of a null model through trimming the distribution of F_{ST}. *The American Naturalist*, **186**, S24–S36.
Wilson EO (1975) *Sociobiology: The New Synthesis*. Harward University Press, Cambridge.
Wright S (1951) The genetical structure of populations. *Annals of Eugenics*, **15**, 323–354.
Wright S (1965) The interpretation of population structure by F-statistics with special regard to systems of mating. *Evolution*, **19**, 395–420.

P.F. wrote the paper with input from O.B., P.M.K., N.J.B., M.B. All authors provided editorial comments. P.M.K., R.D.P., G.J. collected samples. P.F., O.B., R.H., M.B. analysed the data. P.F., P.M.G., M.B. designed research. P.F., P.M.G., J.R.M., R.M.G. performed research.

Data accessibility

Mitogenome DNA sequences: GenBank accessions KY039188-KY039268. R script associated with Appendix S3 (Supporting information) is available on DRYAD: doi:10.5061/dryad.86mh8. Appendix S8 (Supporting information) is available on DRYAD: doi:10.5061/dryad.86mh8.

Supporting information

Additional supporting information may be found in the online version of this article.

Appendix S1 Parameters for COLONY analysis.

Appendix S2 Distribution of full sibling and half sibling probabilities for each pair of individuals as estimated with COLONY.

Appendix S3 Table S3a: Haplotypes for the XHSP and associated probability to observe the data if the pair was paternally or maternally related. The first column indicates whether the individuals were found in a different or the in same river.

Table S3b: Haplotype distribution.

Appendix S4 DArTseq SNP genotypes.

Appendix S5 Number of remaining SNP after each filtering steps for each complexity reduction method.

Appendix S6 Diversity indices for the nuclear SNPs. Ar = Allelic Richness; Ho = Observed heterozygosity, He = Expected Heterozygosity and Fis = Inbreeding Coefficient.

Appendix S7 Median-joining haplotype network inferred from whole mitogenome of northern Australian *Glyphis glyphis*. A circle depicts each haplotype, with size proportional to the number of copies found in this study. Each black circle on connecting lines between haplotypes represents one mutation step. a) Analysis of dataset with all samples, $n = 149$. b) Analysis of dataset without full-sibling and half-siblings pairs, $n = 90$.

Appendix S8 The file is available for download on DRYAD: doi:10.5061/dryad.86mh8

Appendix S9 Bayesian Information Criterion values for each number of clusters, a) Analysis of dataset with all individuals. a) Analysis of NoSib dataset.