MOLECULAR ECOLOGY

Molecular Ecology (2014) **23**, 2590–2601 doi: 10.1111/mec.12754

# Genetic structure of populations of whale sharks among ocean basins and evidence for their historic rise and recent decline

THOMAS M. VIGNAUD,* JEFFREY A. MAYNARD,*† RAPHAEL LEBLOIS,‡ MARK G. MEEKAN,§ RICARDO VÁZQUEZ-JUÁREZ,¶ DENÍ RAMÍREZ-MACÍAS,¶** SIMON J. PIERCE,†† ‡‡ DAVID ROWAT,§§ MICHAEL L. BERUMEN,¶¶ CHAMPAK BEERAVOLU,‡ SANDRA BAKSAY* and SERGE PLANES*

**Laboratoire d'Excellence «CORAIL» USR 3278 CNRS – EPHE, CRIOBE, Papetoai, Moorea, French Polynesia, †Department of Ecology and Evolutionary Biology, Cornell University, Ithaca, NY 14568, USA, ‡INRA, UMR1062 CBGP, F-34988 Montferrier-sur-Lez, France, §Australian Institute of Marine Science, UWA Oceans Institute (MO96), 35 Stirling Hwy, Crawley, WA 6009, Australia, ¶Centro de Investigaciones Biologicas del Noroeste, Mar Bermejo 195, Col. Playa Palo de Santa Rita, La Paz, B.C.S. 23096, Mexico, **Tiburón Ballena México proyecto de Conciencia Mexico, Manatí 4802, Col. Esperanza III, La Paz, B.C.S. 23090, Mexico, ††Marine Megafauna Foundation, 3024 Frandoras Circle, Oakley, CA 94561, USA, ‡‡Wild Me, Praia do Tofo, Inhambane, Mozambique, §§Marine Conservation Society Seychelles, PO Box 1299, Victoria, Mahe, Seychelles, ¶¶Red Sea Research Center, King Abdullah University of Science and Technology, 23955-6900 Thuwal, Kingdom of Saudi Arabia*

## Abstract

**This study presents genetic evidence that whale sharks, *Rhincodon typus*, are comprised of at least two populations that rarely mix and is the first to document a population expansion. Relatively high genetic structure is found when comparing sharks from the Gulf of Mexico with sharks from the Indo-Pacific. If mixing occurs between the Indian and Atlantic Oceans, it is not sufficient to counter genetic drift. This suggests whale sharks are not all part of a single global metapopulation. The significant population expansion we found was indicated by both microsatellite and mitochondrial DNA. The expansion may have happened during the Holocene, when tropical species could expand their range due to sea-level rise, eliminating dispersal barriers and increasing plankton productivity. However, the historic trend of population increase may have reversed recently. Declines in genetic diversity are found for 6 consecutive years at Ningaloo Reef in Australia. The declines in genetic diversity being seen now in Australia may be due to commercial-scale harvesting of whale sharks and collision with boats in past decades in other countries in the Indo-Pacific. The study findings have implications for models of population connectivity for whale sharks and advocate for continued focus on effective protection of the world's largest fish at multiple spatial scales.**

*Keywords*: demographic history, genetic diversity, microsatellites, molecular ecology mtDNA, population expansion, *Rhincodon typus*

*Received 29 January 2014; revision received 8 April 2014; accepted 13 April 2014*

## Introduction

From science and conservation perspectives, three of the largest sharks – whale (*Rhincodon typus*, Smith, 1828), great white (*Carcharodon carcharias*, Linnaeus, 1758) and tiger (*Galeocerdo cuvier*, Péron & Lesueur, 1822) – create similar challenges for researchers and managers. All of these sharks spend parts of their life cycle in the open or deep oceans where they are difficult to observe and basic aspects of their biology such as breeding and pupping locations are mostly unknown

Correspondence: Thomas M. Vignaud, Fax: (33)(0)4 68 50 36 86; E-mail: contact@thomasvignaud.com

(Musick & Ellis 2005; Carrier *et al.* 2010). Conservation of these sharks is made challenging because of their slow growth, late maturation and resultant low rebound potential, which make them highly vulnerable to overexploitation, and because they move across political boundaries (Cortés 2000; Musick & Ellis 2005; Baum & Worm 2009). These issues are perhaps best exemplified by whale sharks, the largest of the sharks and all extant fish species.

Whale sharks form temporary aggregations of mostly subadult juvenile males near tropical and subtropical coastlines (Nelson 2004; Meekan *et al.* 2006; Riley *et al.* 2010; Rowat & Brooks 2012) that are most likely driven by seasonal blooms in food (Martin 2006; Stevens 2007; Rowat & Brooks 2012). The tendency to aggregate in coastal waters and the approachability of whale sharks makes them easy for fishers to catch and their size and demand for their meat and fins has made them a lucrative target for fisheries (Silas 1986; Norman 2004; Rowat & Brooks 2012). At the peak of the Indian fishery in 1998, over 1000 whale sharks were taken off the Saurashtra coast alone (Pravin 2000). Asian nations such as Taiwan have also been major consumers of whale shark meat, with an estimated 271 sharks taken in these waters in 1997 (Chen & Phipps 2002). Although killing whale sharks commercially is now banned in many countries (e.g. the Philippines, Thailand, Taiwan), fishers continue to try to meet demand in China (Li *et al.* 2012). Whale sharks were added to Appendix II of the Convention on International Trade of Endangered Species (CITES) in 2002 (update in Norman 2005). This listing occurred two years after the International Union for the Conservation of Nature (IUCN) categorized the species as 'Vulnerable to Extinction' based on the probability that 20–50% of the species would be lost over the next three generations (Norman 2000).

Tagging and sighting data (see Sequiera *et al.* 2013 for review) suggest that whale sharks from aggregation sites within ocean basins are connected on at least regional (100s–1000s km) scales. This is supported by genetic evidence from two published studies that sampled locations in three ocean basins (Indian, Pacific and Atlantic). Both Castro *et al.* (2007) and Schmidt *et al.* (2009) found some genetic differentiation of individuals from the Caribbean with those from sites in the Indian and Pacific Oceans. Modelling and genetic evidence suggest broadscale connectivity among populations of the Indo-Pacific with uncertainty as to the degree of mixing between populations in the Atlantic and those of the Indian and Pacific Oceans. Sequiera *et al.* (2013) suggest that whale sharks have the capacity to form a single global metapopulation, given the existing photographic, tagging and genetic evidence.

Although the strongest evidence for broadscale patterns of connectivity of whale shark populations comes from genetic analyses, the generality of the conclusions of these studies is limited by sample sizes. Both Castro *et al.* (2007) and Schmidt *et al.* (2009) sampled a total of <70 sharks, and in these studies, more than ten individuals were sampled in only four or less locations. Additionally, patterns were only analysed in either microsatellite DNA or mtDNA (not both), and both papers acknowledge that greater sampling is required to test the interpreted patterns. Only one recent study has attempted a demographic history analysis of whale sharks (O'Brien *et al.* 2013). These authors suggested that previous climatic events did not appear to have affected population size. For this work, samples were collected from a single location.

Here, we build on earlier work to provide a robust and comprehensive picture of the genetics of whale shark populations at the global scale. We analyse both microsatellite (406 samples) and mtDNA (573 samples) from as many as nine locations that include a minimum of 10 samples (usually many more). Our locations include the Red Sea and the Gulf of California, which are included here in what we call the Indo-Pacific. We use these data to test the hypothesis postulated by Sequiera *et al.* (2013) that whale sharks could form a single global metapopulation. Further, we look for signals of population expansions or reductions to explore the demographic history of the whale shark on evolutionary timescales and more recently when these sharks have been subject to fishing and collision with boats.

## Materials and methods

### *Sample collection and laboratory procedures*

DNA was obtained from skin samples collected from free-swimming sharks and on rare occasion from dead specimens (giving a total of 635 individuals). Additional sequences were also taken from data deposited in GenBank by Castro *et al.* (2007) (68 individuals). Sampling locations were as follows with sample sizes for microsatellite loci (ms) and mitochondrial DNA (mt): Red Sea (ms 46; mt 26), Djibouti (ms 89; mt 79), Seychelles (ms 20; mt 31), Maldives (ms 0; mt 10), Mozambique (ms 26; mt 51), Australia – Ningaloo Reef (ms 128; mt 157), Taiwan and the Philippines (ms 0; mt 57), Mexico – Gulf of California (ms 47; mt 84) and Mexico – Isla Holbox (ms 50; mt 80) (map in Fig. 1). Collection dates ranged from 1992 to 2012 with >95% of samples collected between 2003 and 2012. Two measures were taken to avoid replicate sampling. First, care was taken in the field not to sample the same individual twice, which is easy with this species because individuals can be

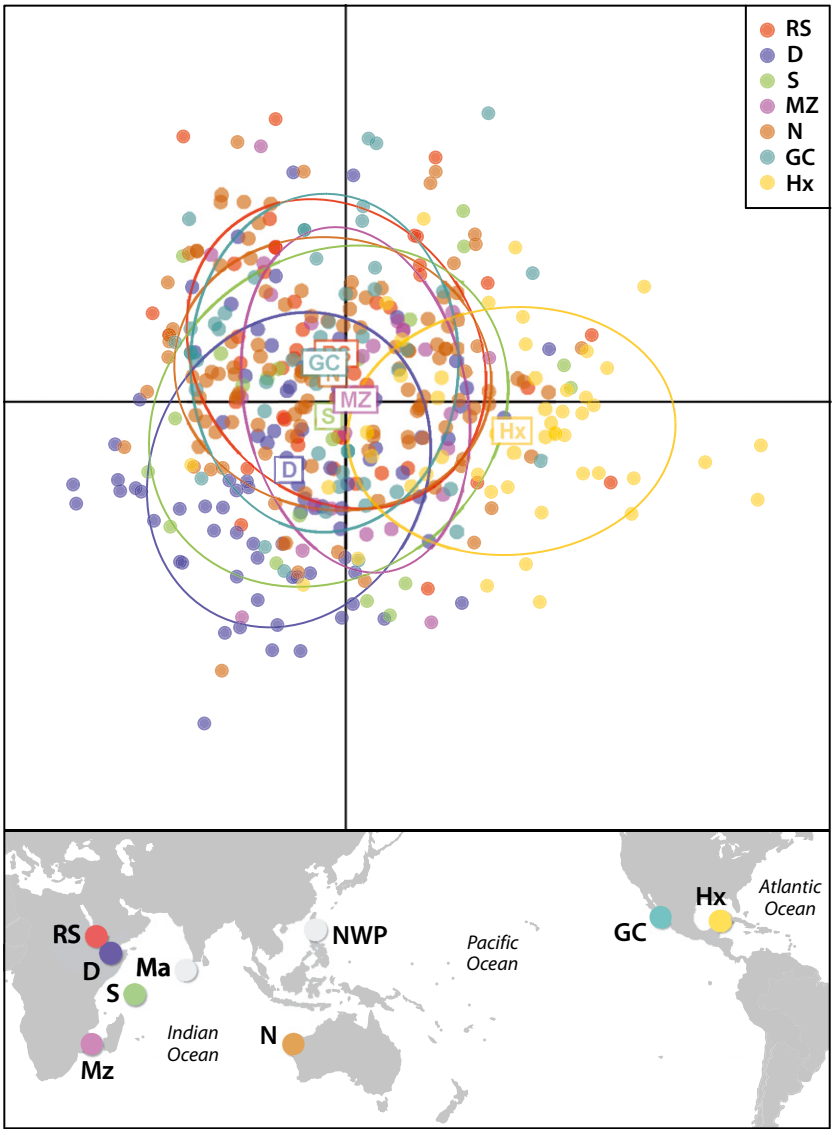


**Fig. 1** Scatterplot output from a discriminant analysis of principal components for genetic signatures from microsatellite DNA ($n$ = 14) of whale shark individuals (based on alpha-score of 26). Dots represent individuals from the seven locations for which microsatellite DNA was available; inertia ellipses centre on the mean for each location and include 67% of the sampling points. Sampling locations are as follows: RS – Red Sea; D – Djibouti; S – Seychelles; Mz – Mozambique; N – Ningaloo; GC – Gulf of California; Hx – Isla Holbox. Only mtDNA was available for Maldives (Ma) and Northwest Pacific (NWP).

recognized using photo-identification techniques that are well developed (e.g. Graham & Roberts 2007; Marshall & Pierce 2012). Second, genetic markers were compared to be sure no individual sample was included more than once. There were a few cases in which the same genotype was found in our body of samples, but in all cases one was removed, and none of these cases were for samples from different years at Ningaloo Reef in Australia. The primers used to isolate part of the mtDNA control region were WSCR1-F and WSCR2-R from Castro *et al.* (2007). All fragments were amplified following the polymerase chain reaction (PCR) protocol as described in Williams *et al.* (2012).

Of the 14 microsatellite loci used by our study, we developed eight with three sourced from each of Schmidt *et al.* (2009) and Ramírez-Macías *et al.* (2009) (Table S1, Supporting information). Details on the multiplex used and the variable quantities of each primer are also shown in Table S1. The mix and PCR protocol used for microsatellites is described in Vignaud *et al.* (2013).

### *Data analysis*

Sizes of microsatellite alleles were read using GENEMAPPER version 3.7 software (Applied Biosystems, Foster City, CA, USA). MICROCHECKER v2.2.3 (van Oosterhoot *et al.* 2004) was used to check potential genotyping errors on the microsatellite data, the presence of null allele(s) and Hardy–Weinberg equilibrium. This led to discarding five of the 19 microsatellite loci originally selected by the study for analysis.

Fragments of mtDNA sequences were read using GENEIOUS 6 (Biomatters, http://www.geneious.com/)

and aligned using the ClustalX method followed by manual corrections (Larkin *et al.* 2007). Two data sets were produced: a raw and a modified data set where gaps/insertions found were replaced with a one-mutation step. Modifying the data avoided losing information or generating misleading results driven solely by different mutation rates for hypervariable regions (Aris-Brosou & Excoffier 1996). All analyses were completed using the modified data set unless noted.

### *Genetic diversity and structure*

Indices of diversity were analysed using GENEPOP 4.2 (Rousset 2008) for microsatellites and DNASP v5.10.01 (Librado & Rozas 2009) for mtDNA. For microsatellites, the rarefaction method was used in the software HP-rare (Kalinowski 2005) to calculate the allelic richness as this method accounts for differences in sample size. AMOVA and pairwise $F_{ST}$ (Weir & Cockerham 1984) values for microsatellites were calculated using ARLEQUIN 3.5 (Excoffier & Lischer 2010). The genotypic differentiation test (G-based, Goudet *et al.* 1996) and associated significance were computed using GENEPOP 4.2 software. For the mtDNA control region, pairwise $F_{ST}$ (Slatkin 1995) values were calculated using ARLEQUIN 3.5. Adegenet (Jombart 2008) for R (R Development Core Team 2013) was used to perform discriminant analysis of principal components (DAPC, Jombart *et al.* 2010) with the number of principal components set to 26, following alpha-score indication. For the DAPC plot, inertia ellipses were generated encompassing the conventional ~67% of the cloud of points for each sampling location. Ellipse centres are at the gravity centre of the cloud of points for each sampling location.

### *Demographic history*

Analyses of demographic history used mtDNA, except where indicated, and were performed on all individuals from the Indo-Pacific (Isla Holbox was excluded for reasons presented in the results). Neutrality analysis *Fs* (Fu 1997), $R^2$ (Ramos-Onsins & Rozas 2002) and *D* (Tajima 1989) and the associated *P*-values (using empirical distribution from coalescent simulations) were performed using DNAsp v5.10.01. A population expansion is indicated when *Fs* is a large negative value, when $R^2$ is a small positive value and when *D* is a small negative value. Mismatch analysis was performed following the method implemented in ARLEQUIN 3.5, which infers ancestral and actual $\theta$ values along with $\tau$ and computes sum of square deviations and associated *P*-values, assuming a sudden population expansion. $\tau$ can give the timing of expansion (if found, noted *T*) as $\tau = T \times 2 \times \mu$. Similarly, $\theta$ can give the number of genes (to be converted in effective number of individual depending on marker used, noted *Ng*) as $\theta = 2 \times Ng \times \mu$ (see Excoffier & Lischer 2011). Calculating the timing of expansions and effective population sizes is thus highly dependent on the chosen mutation rate ($\mu$). Mutation rates are unknown for whale sharks, and those used in other studies come from very distantly related sharks and other species. No mutation rate was selected here, and the reasoning behind and implications of this decision are discussed.

Raw control region models were tested using JMODELTEST2 (Guindon & Gascuel 2003; Darriba *et al.* 2012), and mutation models were ranked using BIC values. The Bayesian skyline plot (BSP), which infers historical population sizes, was then performed using BEAST2 (Ho & Shapiro 2011; Bouckaert *et al.* 2013) and associated software (Beauti and Tracer). For the BSP, a HKY model was used, with a chain length of 20 000 000 iterations with thinning every 20 000 iterations.

Demographic history was also explored using the MIGRAINE software (http://kimura.univ-montp2.fr/~rousset/Migraine.htm) and the newly developed model of a single population with past variations in population size (Leblois *et al.* in review) on both microsatellite and mtDNA data. To infer model parameters, MIGRAINE uses the class of importance sampling algorithms developed by de Iorio & Griffiths (2004a,b) and de Iorio *et al.* (2005) and extended in Leblois *et al.* in review. MIGRAINE was used to estimate ancestral $\theta$ and actual $\theta$ values and *D*, which operates like $\tau$ in the mismatch analysis described earlier, except that past variation in population size is exponential and not discrete/sudden. Like $\tau$, *D* is an indicator of population expansions and reductions and can be used to calculate expansion/reduction timing if a mutation rate is chosen. Here, the formula to obtain the timing of the expansion (if found) in generation is $T = 2 \times D \times Ng$. Actual and ancestral number of genes follows $Ng = \theta/(2 \times \mu)$. A benefit of using MIGRAINE is that it allows departure from use of the strict stepwise mutation model (i.e. using a generalized stepwise mutation model). Because MIGRAINE is based on the infinitely many-site model (ISM) for analysis of sequence data, two data sets were produced for the mtDNA control region to fit this model. There are two reasons that sequence data sets may not fit the ISM: sites can show more than two nucleotidic states, or pairwise comparisons of sites may not comply to the four-gamete test (Hudson & Kaplan 1985). For one data set, we systematically removed incompatible sites for all individuals (resulting in 511-bp fragment; $n = 493$), and for the second, we removed haplotypes with incompatible sites (resulting in 608-bp fragment; $n = 370$). All runs in MIGRAINE were made for microsatellites using 20 000 trees, 2400–5000 points and 3–10 iterations, and

for mtDNA using 200 000 trees, 2400 points and 2–4 iterations.

Annual genetic diversity indices were calculated for Ningaloo Reef (N), the only location where at least 10 individuals were sampled for 6 years in succession. The indices were calculated using ARLEQUIN 3.5 for haplotype diversity and θ(Hom) on mtDNA; HP-Rare (Kalinowski 2005) was used again for allelic richness ($N_e$) on microsatellites.

## Results

### *Genetic diversity and structure*

All indices suggest high diversity of mtDNA with very similar levels of diversity for all locations, but lower diversity in samples from Isla Holbox (Table 1). Sample sizes for microsatellites ranged from 20 (Seychelles) to 128 (Ningaloo) so allelic richness was considered to be a good indicator of differences in genetic diversity among the sampling locations, given that the rarefaction method accounted for the large differences in *N* (Kalinowski 2005). Allelic richness ranged from 4.37 in the Gulf of California to 4.82 in Djibouti with Isla Holbox having the only value below 4 (3.95). Isla Holbox was also the only location with a mean number of allele over loci lower than 6.00 (5.71) and had the lowest expected diversity at 0.60 (Table 1). Similar patterns among localities were found in the 608-bp control region fragment. Haplotype diversity (H) was above 0.90 at all locations, again with the exception of Isla Holbox where H was 0.752. Nucleotide diversity and θ (Hom) results were also lowest at Isla Holbox (Table 1).

Very little genetic structure was detected by the analyses for the sharks sampled from the Indo-Pacific. Greater structure was seen for all comparisons associated with Isla Holbox. AMOVA percentages of variation were 0.55% for microsatellites and 1.06% for mtDNA among Indo-Pacific locations, but increased to 2.08 and 7.50% for microsatellites and mtDNA, respectively, when samples from Isla Holbox were included in the analysis. Similarly, pairwise $F_{ST}$ values for comparisons of microsatellite DNA between sampling locations were ≤0.13, with the exception of comparisons that included Isla Holbox, which were all >0.2 excepting in the comparison with Mozambique. A test of genotypic differentiation for microsatellite DNA produced highly significant (<0.001) results only for comparisons of samples from locations that included Isla Holbox (Table 2); the two exceptions include Djibouti which had preserved samples with lower-quality microsatellite DNA (see **x in Table 2).

**Table 1** Indices of genetic diversity for whale sharks from each sampling site for both microsatellites and mtDNA

| | Microsatellites | | | | |
|---|---|---|---|---|---|
| | *N* | Mean number of allele over loci | Expected heterozygocity (*He*) | Observed heterozygocity (*Ho*) | Allelic richness ($N_e$) |
| Red Sea | 46 | 7.28 | 0.65 | 0.67 | 4.60 |
| Djibouti | 89 | 9.14 | 0.65 | 0.50 | 4.82 |
| Seychelles | 20 | 6.00 | 0.66 | 0.67 | 4.57 |
| Maldives | — | — | — | — | — |
| Mozambique | 26 | 6.64 | 0.64 | 0.67 | 4.52 |
| Ningaloo | 128 | 8.57 | 0.65 | 0.62 | 4.49 |
| Northwest Pacific | — | — | — | — | — |
| Gulf of California | 47 | 7.14 | 0.63 | 0.62 | 4.37 |
| Isla Holbox | 50 | 5.71 | 0.60 | 0.58 | 3.95 |

| | mtDNA | | | | |
|---|---|---|---|---|---|
| | *N* | Number of Haplotypes (NH) | Haplotype diversity (H) | Nucleotide diversity (π) | θ(Hom) (SD) |
| Red Sea | 26 | 12 | 0.923 | 0.01041 | 10.57 (4.38) |
| Djibouti | 79 | 30 | 0.925 | 0.01069 | 10.95 (2.61) |
| Seychelles | 31 | 18 | 0.948 | 0.01062 | 16.77 (8.87) |
| Maldives | 10 | 9 | 0.978 | 0.01868 | 42.17 (108.92) |
| Mozambique | 51 | 29 | 0.96 | 0.01360 | 22.31 (8.58) |
| Ningaloo | 157 | 41 | 0.907 | 0.01300 | 14.03 (2.67) |
| Northwest Pacific | 57 | 36 | 0.964 | 0.01323 | 24.80 (10.75) |
| Gulf of California | 83 | 25 | 0.921 | 0.00991 | 10.48 (2.27) |
| Isla Holbox | 80 | 15 | 0.752 | 0.00755 | 2.35 (0.51) |

**Table 2** Pairwise $F_{ST}$ values for all sampling locations. Values for microsatellites (left) are one-locus estimates following standard ANOVA (as in Weir & Cockerham 1984), and values for mtDNA (right) are based on the distance method. Significant (<0.05*) and highly significant (<0.001**) results are shown in bold for a test of genotypic differentiation for microsatellite DNA and are for *P*-values associated with the distance method for mtDNA. The significant results shown for comparisons that include Maldives and Djibouti are probably driven by low sample sizes and the low quality of DNA in preserved samples, respectively

| | Red Sea | Djibouti | Seychelles | Maldives | Mozambique | Ningaloo | NW Pacific | G California | Isla Holbox |
|---|---|---|---|---|---|---|---|---|---|
| Red Sea | | −0.0124 | −0.0154 | **0.1019*** | −0.0006 | 0.0049 | −0.0068 | −0.0013 | **0.2429**** |
| Djibouti | **0.011**x** | | −0.0060 | **0.1086*** | −0.0050 | **0.0130*** | −0.0032 | 0.0028 | **0.2162**** |
| Seychelles | 0.0025 | −0.0022 | | **0.0912*** | −0.0102 | 0.0059 | −0.0085 | −0.0030 | **0.2269**** |
| Maldives | — | — | — | | **0.0735*** | **0.1216**** | **0.0724*** | **0.1019*** | **0.3506**** |
| Mozambique | **0.0091*** | 0.0080 | −0.0031 | — | | **0.0214*** | −0.0048 | 0.0040 | **0.1959**** |
| Ningaloo | 0.0038 | **0.0069**x** | −0.0032 | — | 0.0026 | | 0.0113 | **0.0123*** | **0.2359**** |
| NW Pacific | — | — | — | — | — | — | | 0.0028 | **0.1940**** |
| G California | 0.0015 | **0.0076*** | 0.0009 | — | −0.0024 | **0.0013*** | — | | **0.2532**** |
| Isla Holbox | **0.0303**** | **0.0367**** | **0.0253**** | — | **0.0132**** | **0.0252**** | — | **0.0253**** | |

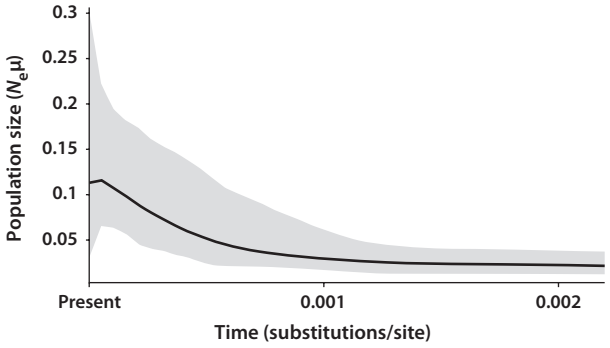


**Fig. 2** Bayesian skyline plot of variation in effective population size through time based on the control region of the mtDNA for individuals ($n$ = 493) from all sampling locations (see Fig. 1), except Isla Holbox. Grey shade indicates the 95% highest posterior density.

Pairwise $F_{ST}$ values for mtDNA between localities produced similar results. Comparisons that included Isla Holbox were also the only $F_{ST}$ values for mtDNA that were greater than 0.15 (Table 2). There were some exceptions to the finding of genetic homogeneity across the Indo-Pacific locations (see Table 2). Those that include Maldives are not reliable because of the low sample sizes there. The other two suggest that structure is greatest in the Indo-Pacific for comparisons that include Ningaloo Reef, Australia. However, these pairwise $F_{ST}$ values were an order of magnitude lower in all cases than for comparisons that included Isla Holbox.

Discriminant analysis of principal components (DAPC) showed that none of the samples were tightly grouped (Fig. 1). Ellipses encompassing ~67% of the cloud of points for each sampling location all overlapped although samples from Isla Holbox were clearly most different from the other locations.

### *Demographic history*

As structured populations can produce discrepancies in demographic history analyses (Chikhi *et al.* 2010; Wegmann & Excoffier 2010; Leblois *et al.* in review), Isla Holbox was excluded and the demographic history analyses were conducted on samples from the Indo-Pacific only. A significant ($P < 0.05$) population expansion was indicated by neutrality and by mismatch analysis ($P = 0.00241$), as well as by visual interpretation of a Bayesian skyline plot based on the mtDNA (Fig. 2). A significant expansion was also indicated by MIGRAINE software, which used both microsatellite and mtDNA (Table 3).

Values for the neutrality analysis were −94.17 for *Fs*, 0.0291 for $R^2$ and −1.8765 for *D*, all of which indicate a significant population expansion ($n$ = 493 samples from the Indo-Pacific of mtDNA). The mismatch analysis produced an actual θ of 14.9 and an ancestral θ of 2.9 with a τ value of 5.7. The large differences between ancestral and actual θ result in a large θ ratio (e.g. 5.14), which is also indicative of an expansion.

A Bayesian skyline plot (BSP, Fig. 2) showed effective population size to be an order of magnitude higher at present than in the past. The timing of this expansion was unknown because we did not select a mutation rate, but the small *D* values from the MIGRAINE analysis suggest that the expansion was recent (few hundred generations). In Fig. 2, the values from the coalescence inference used to produce the plot were bounded by a 95% highest posterior density. The value at the lowest boundary of the effective population size at present was higher than the value at the highest boundary of the population size prior to expansion, increasing confidence that the expansion is real. Outputs from the MIGRAINE software using microsatellite and mtDNA were aligned with those of the mismatch analysis. Actual θ

**Table 3** Inferences on demographic history by the software MIGRAINE and from the mismatch analysis. Values are shown above the full data range outputted in brackets. Ngene = $\theta/2\mu$; Tgeneration = $2 \times D \times N_e$ for MIGRAINE and Tgeneration = $\tau/2\mu$ for the mismatch analysis

| *N* | Ancestral θ | Actual θ | *D* | θ ratio |
|---|---|---|---|---|
| Microsatellites | | | | |
| 356 | 2.5 [0.0091–3.4] | 7.5 [3.9–17.9] | 0.0088 [0.000984–1.5] | 3.03 [1.7–545] |
| mtDNA | | | | |
| Deleted problematic sites (511 bp left) | | | | |
| 493 | 1.55 [0.000078–4.77] | 56 [24.6–123] | 0.0054 [0.00137–0.118] | 36.21 [9.5–460 635] |
| Deleted problematic individuals (370 individuals left) | | | | |
| 370 | 0.0227 [0.000085–4.2] | 24.3 [14.0–46.2] | 0.107 [0.0069–0.61] | 1071 [5.9–472 553] |

| Mismatch analysis | | | | |
|---|---|---|---|---|
| *N* | Ancestral θ | Actual θ | τ | θ ratio |
| 493 | 2.9 [0–2.41] | 14.92 [9.67–67.73] | 5.7 [3.95–11.28] | 5.15 [?–28.1] |

was larger than ancestral θ for microsatellites and mtDNA, and the θ ratio confidence interval spaces were similar and relatively large in all cases (Table 3).

For Ningaloo Reef, three indicators of genetic diversity – haplotype diversity and θ(Hom) for mtDNA and allelic richness for microsatellites – decreased each year from 2007 to 2012 (Table 4). Haplotype diversity decreased from 1 in 2007 (and 0.922 in 2008) to 0.877 in 2012. θ(Hom) decreased from 10.35 to 5.99 during the same period. Allelic richness decreased from 4.52 in 2010 to 4.29 in 2012. These indicators decreasing between 2007 and 2012 at Ningaloo Reef suggest that there could be a recent reverse in the longer-term trend of an expanding population indicated by all four of the analyses of demographic history.

## Discussion

Our global analyses of whale shark genetics, the largest of its kind to date, suggest that whale sharks exist in two distinct populations with minimal connectivity between the Indo-Pacific and the Atlantic Ocean. This is in contrast to the recent suggestion that there could be a single global metapopulation of whale sharks (Sequiera *et al.* 2013). Our findings are consistent with the suggestion made in Castro *et al.* (2007) and Schmidt *et al.* (2009), the authors of which were uncertain of their interpretation of their results because of low sample sizes. Microsatellite and mtDNA analyses also found evidence of genetic structure in the Indo-Pacific, where comparisons included either the Maldives or Ningaloo Reef. For the former locality, the result is probably attributable to sampling as Maldives had the lowest sample size of any of our locations ($n = 10$, mtDNA only). Further sampling at Maldives will help assess connectivity between Maldives and other locations in the region. The significant pairwise $F_{ST}$ values found when comparing Ningaloo Reef with Mozambique, Djibouti and the Gulf of California indicate that mixing may be limited among sharks contributing juveniles to these aggregation areas.

However, it makes little phylogeographical sense for some locations in the Indo-Pacific to have structure and others to not. Movement among all locations in the Indo-Pacific is within the known physical abilities of whale sharks. There are no physical or environmental barriers in the Indo-Pacific that compare to the southern tip of Africa or South America, which extend into cool temperate waters and restrict migration of many tropical species. Genetic structure analyses can be influenced by factors such as the properties of the markers used, departure from classical mutation models, selection, population size and/or the presence of ghost populations (Slatkin 2005; Lowe & Allendorf 2010; Marko & Hart 2011). Thus, the possibility remains that some level of structure exists between Indo-Pacific aggregation sites, and more comprehensive sampling, markers and analytical techniques may reveal this. Most importantly, pairwise $F_{ST}$ values were an order of magnitude higher for comparisons between Isla Holbox and locations in the Indo-Pacific than for any of the comparisons of locations within the Indo-Pacific.

Analyses of teeth suggest that whale sharks first appear in the fossil record at least 30 million years ago (Bourdon, unpubl. data). Whale sharks probably formed a single population in the past that was divided when the Isthmus of Panama closed during the Pliocene (Coates & Obando 1996; Haug & Keigwin 2004) as this ended the pan-global connection of tropical seas across what are now the Atlantic, Indian and Pacific Ocean basins. There has been some connection of tropical fish faunas of the Atlantic and Indo-Pacific since this time,

**Table 4** Indices of genetic diversity for whale sharks sampled at Ningaloo Reef (see Fig. 1) from 2007 to 2012. Groups are shown in the second θ(Hom) row of 2007 + 2008, 2009 + 2010 and 2011 + 2012

| | | 2007 | 2008 | 2009 | 2010 | 2011 | 2012 |
|---|---|---|---|---|---|---|---|
| mtDNA | *N* | 12 | 39 | 13 | 33 | 18 | 31 |
| | Haplotype diversity (H) | 1 | 0.922 | 0.897 | 0.894 | 0.889 | 0.877 |
| | θ(Hom) (se) | — | 10.35 (4.17) | 7.47 (5.94) | 7.17 (2.84) | 6.77 (4.83) | 5.99 (2.11) |
| | θ(Hom) (se) | 13.41 (4.74) | | 7.85 (2.58) | | 6.56 (1.96) | |
| Microsatellites | *N* | | | | 34 | 24 | 60 |
| | Allelic richness ($N_e$) | — | — | — | 4.52 | 4.43 | 4.29 |

presumably due to dispersal around the tip of southern Africa (Bowen *et al.* 1997, 2001). These events are thought to occur very infrequently, on scales of $10^5$ to $10^6$ years (Roberts *et al.* 2004; Rocha *et al.* 2005; Bowen *et al.* 2006). Dispersal is most likely in the warmer periods between glacial epochs, which characterizes present day, with the last glaciation ceasing around 11 700 years ago (Walker *et al.* 2009). Tropical plankton have been found in sediment cores off southwestern Africa (Peeters *et al.* 2004), and this is thought to indicate a temporary hiatus in the Benguela upwelling system during interglacial warming periods that resulted in a westward flow of warm water around southern Africa (reviewed in Castro *et al.* 2007). The hiatus in the Benguela upwelling may have facilitated movement between the two ocean basins of many tropical fauna and is part of the basis for the suggestion by Sequiera *et al.* (2013) that movements of whale sharks between the Indian and Atlantic Oceans may be occurring in the present day. Our data suggest that mixing between the Indian and Atlantic was and is rare, or that if migrations do occur they very rarely involve breeding or pupping away from the natal oceans. The level of mixing is clearly not sufficient to counter the effects of genetic drift.

The exact number of migrants required to counter drift is a contentious issue in molecular ecology and is likely to vary among species (review in Waples & Gaggiotti 2006). The minimum generation time of whale sharks is estimated to be at least 25 years and potentially more (Wintner 2000; Castro *et al.* 2007; Rowat & Brooks 2012). Given this, only a few individuals dispersing between the Indian and Atlantic Oceans every 2–4 years and then breeding with resident populations would be sufficient to create lower levels of genetic structure than we recorded.

Both microsatellite and mtDNA analyses of demographic history of the Indo-Pacific population indicate a population expansion. The expansion found is very likely to be recent (few hundred generations) given the small *D* values in the MIGRAINE analysis. The recent expansion indicated potentially coincides with the early Holocene (approximately 7000–11 000 years ago) and the rise in sea levels that increased both suitable oceanic surface for the whale shark and plankton productivity (Marx & Uhen 2010). Estimating the exact timing of the start of the expansion and the effective population sizes associated with the ancestral and actual θ values all require selecting a mutation rate. We have not used a mutation rate here for the following reasons. The mutation rate for microsatellites applied to whale sharks in a previous study was borrowed from only very distantly related species, which were not con-familial (0.001 mutations/generation/locus; Schmidt *et al.* 2009). The accuracy of this approach is questionable (Ho *et al.* 2011; Grant *et al.* 2012; Shapiro & Ho 2014). In another previous whale shark study, a mutation rate is applied to the analysis of mtDNA (Castro *et al.* 2007). These authors use a mutation rate from very distant relatives of whale sharks, and the authors of the original studies that reported these rates claimed to be uncertain as to the reliability of the rate estimates (Duncan *et al.* 2006; Keeney & Heist 2006). Lastly, the hypervariability of the whale shark's control region probably means that mutation rates are higher than the rates used previously. Irrespective of the debate about whether a rate should be applied, selecting a rate would not alter the magnitude of the difference between actual and ancestral θ. Consequently, selecting a rate would not change that our results indicate an expansion and, indeed, that the actual effective whale shark population size is indicated to be far greater than the ancestral effective population size. Ongoing and future research will improve the accuracy of estimated mutation rates that at present remain controversial for many genes and species (Baer *et al.* 2007; Ho *et al.* 2011; Grant *et al.* 2012; Shapiro & Ho 2014).

The finding of a recent population expansion suggests that the Indo-Pacific population has and can continue to grow under the current climatic conditions. However, there are numerous reports showing that whale sharks have been exploited over the last century, sometimes heavily. For this reason, not finding a recent bottleneck signal can be seen as surprising although there is a slight trend down in the very recent past in

the BSP in Fig. 2. Recent bottlenecks are known to sometimes go undetected by demographic history inference due to variability in sampling and sampling time (Heller *et al.* 2013). There is especially great potential for this with whale sharks because of the long generation times of ~25 years. Aside from the recent downward trend in the BSP for recent years, measures of genetic diversity from Ningaloo Reef in Western Australia suggest that the effective population contributing individuals to Ningaloo may have declined in recent decades. We documented decreases in genetic diversity for five consecutive years with mtDNA (2007–2012) and two consecutive years with microsatellites (2010–2012) for individuals aggregating at Ningaloo. There are several potential explanations for the declines in genetic diversity at Ningaloo we document that do not relate to the sharks being exploited. Breeding and pupping locations may change regularly, mating may be nonrandom, and the sampling may not allow for realistic estimation of diversity. These explanations are all plausible although sampling follows a strict protocol, so sampling issues are unlikely. The decline in diversity could also possibly be due to fewer females breeding and/or less females being available to breed when the individuals sampled at the aggregation site were born (approximately 15–25 years ago). This time period coincides with when whale sharks were heavily exploited throughout India and Asia (Speed *et al.* 2008; WildLife-Risk Report 2014). Whether the decline in diversity we document is real is unknown, but it is curious that the decline is seen for so many consecutive years. The successive declines in genetic diversity documented are small even over the 6-year period, but this represents only ~20% of one generation for this species. The result is certainly concerning enough to warrant further research and adds circumstantial evidence to the growing body of evidence that whale sharks need to continue to be conserved at multiple spatial scales.

Holmberg *et al.* (2009) used capture–mark–recapture to calculate changes in the abundance of whale sharks at Ningaloo during years that include our study period. These authors suggest that the abundance is either stable or slightly increasing and that median size is decreasing. In contrast, others have suggested that both mean size and abundance have been decreasing (Bradshaw *et al.* 2008; Meekan *et al.* unpublished data). Abundance at the aggregation site and the population's genetic diversity can be independent; the former could possibly be increasing, while the latter is declining (and vice versa, or both could be stable). Essentially, the strength of the link between abundance at the aggregation site and the population size (effective and total) for whale sharks in the Indo-Pacific is tenuous and unknown. It could be that the larger number of smaller individuals being seen is a sign of recovery and that the decline in genetic diversity we find is the signature or mark left of the levels of exploitation common in past decades, as we suggest is possible above.

In scope and scale, this is the largest study on the genetics of whale sharks conducted to date. There are three primary results and conclusions, each of which informs critical future research. The first is that there are high levels of genetic structure between whale sharks in the Atlantic and Indo-Pacific Oceans, suggesting mixing is not sufficient to counter genetic drift. Only one aggregation site from the Atlantic is included here so levels of mixing between sites within the Atlantic are unknown as is the level of mixing between aggregation sites in western Africa and those just east of Africa such as Maldives. Such research is now possible and would create a more complete picture of whale shark connectivity among and between ocean basins. The second is finding a significant and likely recent population expansion, the exact timing of which is unknown. Actual and ancestral effective population sizes can be calculated using the values published here once mutation rates for whale sharks can be estimated with greater confidence. The third is that despite finding an expanding whale shark population, a very recent bottleneck might have gone undetected as genetic diversity at Ningaloo Reef in Australia has declined during 5 consecutive recent years. The booming ecotourism industries at whale shark aggregation sites around the world enable more frequent and consistent sampling. In future years, genetic analyses can greatly increase our still very limited understanding of whale shark ecology and the status of what appears for now to be at least two populations.

## Acknowledgements

All of the following provided funding for the research presented here (no particular order after the first organization): Labex CORAIL, Ministère de l'Ecologie du Développement Durable et de l'Energie, Ministère de l'Outre Mer, Fonds Pacifique, IFRECOR, Délégation à la recherche de Polynésie, the Agence Nationale de la Recherche, Institut National de Recherche en Agronomie, the King Abdullah University of Science and Technology, Apache Energy Ltd, SeaWorld Research, Rescue Foundation Inc, Save our Seas Foundation, Australian Institute of Marine Science, Western Australian Department of Parks and Wildlife, EMARNAT-CONACYT, Save Our Seas Foundation, Whale Shark México, Shark Foundation, Rufford Small Grant Foundation and the PADI Foundation. Part of the MIGRAINE work was undertaken using the resources of the INRA MIGALE and GENOTOUL bioinformatics platform and the computing grids of ISEM and CBGP labs. In total, the sampling across all aggregation sites required hundreds of colleagues, students and volunteers. We thank everyone that assisted with sampling.

## References


Aris-Brosou S, Excoffier L (1996) The impact of population expansion and mutation rate heterogeneity on DNA sequence polymorphism. *Molecular Biology and Evolution*, **13**, 494–504.

Baer CF, Miyamoto MM, Denver DR (2007) Mutation rate variation in multicellular eukaryotes: causes and consequences. *Nature Reviews Genetics*, **8**, 619–631.

Baum JK, Worm B (2009) Cascading top-down effects of changing oceanic predator abundances. *Journal of Animal Ecology*, **78**, 699–714.

Bouckaert R, Heled J, Kuhnert D *et al.* (2013) *BEAST2: A software platform for Bayesian evolutionary analysis*. Available at http://beast2.cs.auckland.ac.nz.

Bowen BW, Clark AM, Abreu-Grobois FA, Chaves A, Reichart HA, Ferl RJ (1997) Global phylogeography of the ridley sea turtles (*Lepidochelys* spp.) as inferred from mitochondrial DNA sequences. *Genetica*, **101**, 179–189.

Bowen BW, Bass AL, Rocha LA, Grant WS, Robertson DR (2001) Phylogeography of the trumpetfishes (*Aulostomus*): ring species complex on a global scale. *Evolution*, **55**, 1029–1039.

Bowen BW, Muss A, Rocha LA, Grant WS (2006) Shallow mtDNA coalescence in Atlantic pygmy angelfishes (genus *Centropyge)* indicates a recent invasion from the Indian Ocean. *Journal of Heredity*, **97**, 1–12.

Bradshaw CJA, Fitzpatrick BM, Steinberge CC, Brook BB, Meekan MG (2008) Decline in whale shark size and abundance at Ningaloo Reef over the past decade: the world's largest fish is getting smaller. *Biological Conservation*, **141**, 1894–1905.

Carrier JC, Musick JA, Heithaus MR, eds (2010) *Sharks and their Relatives II: Biodiversity, Adaptive Physiology, and Conservation*, pp. 608–633. CRC Press, Boca Raton, Florida, USA.

Castro ALF, Stewart BS, Wilson SG *et al.* (2007) Population genetic structure of Earth's largest fish, the whale shark (*Rhincodon typus*). *Molecular Ecology*, **16**, 5183–5192.

Chen VY, Phipps MJ (2002) *Management and Trade of Whale Sharks in Taiwan*. TRAFFIC East Asia, Taipei.

Chikhi L, Sousa VC, Luisi P, Goossens B, Beaumont MA (2010) The confounding effects of population structure, genetic diversity and the sampling scheme on the detection and quantification of population size changes. *Genetics*, **186**, 983–995.

Coates AG, Obando JA (1996) The geologic evolution of the Central American isthmus. In: *Evolution and Environment in Tropical America* (eds Jackson J, Budd AF, Coates AG), pp. 21–56. University of Chicago Press, Chicago, Illinois.

Cortés E (2000) Life history patterns and correlations in sharks. *Reviews in Fisheries Science*, **8**, 299–344.

Darriba D, Taboada GL, Doallo R, Posada D (2012) jModelTest 2: more models, new heuristics and parallel computing. *Nature Methods*, **9**, 772.

Duncan KM, Martin AP, Bowen BW, de Couet GH (2006) Global phylogeography of the scalloped hammerhead shark (*Sphyrna lewini*). *Molecular Ecology*, **15**, 2239–2251.

Excoffier L, Lischer HEL (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, **10**, 564–567.

Excoffier L, Lischer HEL (2011) *Arlequin 3.5 An Integrated Software Package for Population Genetics Data Analysis – Manual Ver 2.0*. Available at http://cmpg.unibe.ch/software/arlequin35/man/Arlequin35.pdf.

Fu YX (1997) Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics*, **147**, 915–925.

Goudet J, Raymond M, de Meeüs T, Rousset F (1996) Testing differentiation in diploid populations. *Genetics*, **144**, 1931–1938.

Graham R, Roberts CM (2007) Assessing the size, growth and structure of a seasonal population of whale sharks (*Rhincodon typus* Smith 1828) using conventional tagging and photo identification. *Fisheries Research*, **84**, 71–80.

Grant WS, Liu M, Gao T, Yanagimoto T (2012) Limits of Bayesian skyline plot analysis of mtDNA sequences to infer historical demographies in Pacific herring (and other species). *Molecular Phylogenetics and Evolution*, **65**, 203–212.

Guindon S, Gascuel O (2003) A simple, fast and accurate method to estimate large phylogenies by maximum-likelihood. *Systematic Biology*, **52**, 696–704.

Haug GH, Keigwin LD (2004) How the Isthmus of Panama Put Ice in the Arctic: Drifting continents open and close gateways between oceans and shift Earth's climate. *Oceanus magazine*, Woods Hold Oeanographic Institution, Vol. 42.

Heller R, Chikhi L, Siegismund HR (2013) The confounding effect of population structure on Bayesian skyline plot inferences of demographic history. *PLoS One*, **8**, e62992.

Ho SYW, Shapiro B (2011) Skyline-plot methods for estimating demographic history from nucleotide sequences. *Molecular Ecology Resources*, **11**, 423–434.

Ho SYW, Lanfear R, Bromham L *et al.* (2011) Time-dependent rates of molecular evolution. *Molecular Ecology*, **20**, 3087–3101.

Holmberg J, Norman B, Arzoumanian Z (2009) Robust, comparable population metrics through collaborative photo-monitoring of whale sharks *Rhincodon typus*. *Ecological Applications*, **18**, 222–233.

Hudson RR, Kaplan NL (1985) Statistical properties of the number of recombination events in the history of a sample of DNA sequences. *Genetics*, **111**, 147–164.

de Iorio M, Griffiths RC (2004a) Importance sampling on coalescent histories. *Advances in Applied Probability*, **36**, 417–433.

de Iorio M, Griffiths RC (2004b) Importance sampling on coalescent histories II. Subdivided population models. *Advances in Applied Probability*, **36**, 434–454.

de Iorio M, Griffiths RC, Leblois R, Rousset F (2005) Stepwise mutation likelihood computation by sequential importance sampling in subdivided population models. *Theoretical Population Biology*, **68**, 41–53.

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.

Kalinowski ST (2005) HP-RARE 1.0: a computer program for performing rarefaction on measures of allelic richness. *Molecular Ecology Notes*, **5**, 187–189.

Keeney DB, Heist EJ (2006) Worldwide phylogeography of the blacktip shark (*Carcharhinus limbatus*) inferred from mitochondrial DNA reveals isolation of western Atlantic

populations coupled with recent Pacific dispersal. *Molecular Ecology*, **15**, 3669–3679.
Larkin MA, Blackshields G, Brown NP *et al.* (2007) Clustal W and clustal X version 2.0. *Bioinformatics*, **23**, 2947–2948.
Li W, Wang Y, Norman B (2012) A preliminary survey of whale shark *Rhincodon typus* catch and trade in China: an emerging crisis. *Journal of Fish Biology*, **80**, 1608–1616.
Librado P, Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*, **25**, 1451–1452.
Lowe WH, Allendorf FW (2010) What can genetics tell us about population connectivity? *Molecular ecology*, **19**, 3038–3051.
Marko PB, Hart MW (2011) Retrospective coalescent methods and the reconstruction of metapopulation histories in the sea. *Evolutionary Ecology*, **26**, 291–315.
Marshall AD, Pierce SJ (2012) The use and abuse of photographic identification in sharks and rays. *Journal of Fish Biology*, **80**, 1361–1379.
Martin RA (2006) A review of behavioural ecology of whale sharks (*Rhincodon typus*). *Fisheries Research*, **84**, 10–16.
Marx FG, Uhen MD (2010) Climate, critters, and cetaceans: cenozoic drivers of the evolution of modern whales. *Science*, **327**, 993.
Meekan MG, Bradshaw CJA, Press M *et al.* (2006) Population size and structure of whale sharks (*Rhincodon typus)* at Ningaloo Reef, Western Australia. *Marine Ecology Progress Series*, **319**, 275–285.
Musick JA, Ellis J (2005) Reproductive evolution of chondrichthyans. In: *Reproductive Biology and Phylogeny of Chondrichthyans* (ed. Hamlett WS), pp. 45–79. Science Publishers Inc., Enfield, New Hampshire.
Nelson JD (2004) *Distribution and foraging ecology by whale sharks (*Rhincodon typus*) within Bahia de los Angeles, Baja California Norte, Mexico*. MSc Thesis, University of San Diego, 118 pp.
Norman B (2000) *Rhincodon typus*. In: IUCN 2007. *2007 IUCN Red List of Threatened Species*. www.iucnredlist.org.
Norman B (2004) *Review of the Current Conservation Concerns for the Whale Shark (Rhincodon typus), a Regional Perspective*. Coast and Clean Seas Project 2127. Australian Government, Department of the Environment and Heritage, Canberra.
Norman B (2005) Rhincodon typus. In: IUCN 2013. *IUCN Red List of Threatened Species. Version 2013.1*. www.iucnredlist.org.
O'Brien SM, Gallucci VF, Hauser L (2013) Effects of species biology on the historical demography of sharks and their implications for likely consequences of contemporary climate change. *Conservation Genetics*, **14**, 125–144.
van Oosterhoot C, Hutchinson W, Wills D, Shipley P (2004) MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes*, **4**, 535–538.
Peeters FJC, Acheson R, Brummer GJA *et al.* (2004) Vigorous exchange between the Indian and Atlantic oceans at the end of the past five glacial periods. *Nature*, **430**, 661–665.
Pravin P (2000) Whale shark in the Indian coast—need for conservation. *Current Science*, **79**, 310–315.
R Development Core Team (2013) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
Ramírez-Macías D, Shaw K, Ward R, Galván-Magaña F, Vázquez-Juárez R (2009) Isolation and characterization of microsatellite loci in the whale shark (*Rhincodon typus*). *Molecular Ecology Resources*, **9**, 798–800.
Ramos-Onsins SE, Rozas J (2002) Statistical properties of new neutrality tests against population growth. *Molecular Biology and Evolution*, **19**, 2092–2100.
Riley MJ, Hale MS, Harman A, Rees RG (2010) Analysis of whale shark *Rhincodon typus* aggregations near South Ari Atoll, Maldives Archipelago. *Aquatic Biology*, **8**, 145–150.
Roberts MA, Schwartz TS, Karl SA (2004) Global population genetic structure and male mediated gene flow in the green sea turtle (*Chelonia mydas*): analysis of microsatellite loci. *Genetics*, **166**, 1857–1970.
Rocha LA, Robertson DR, Rocha CR, Van Tassell JL, Craig MT, Bowen BW (2005) Recent invasion of the tropical Atlantic by an Indo-Pacific coral reef fish. *Molecular Ecology*, **14**, 3921–3928.
Rousset F (2008) Genepop'007: a complete reimplementation of the Genepop software for Windows and Linux. *Molecular Ecology Resources*, **8**, 103–106.
Rowat D, Brooks KS (2012) A review of the biology, fisheries and conservation of the whale shark *Rhincodon typus*. *Journal of Fish Biology*, **80**, 1019–1056.
Schmidt JV, Schmidt CL, Ozer F *et al.* (2009) Low genetic differentiation across three major ocean populations of the whale shark, *Rhincodon typus*. *PLoS One*, **4**, e4988.
Sequiera AMM, Mellin C, Meekan MG, Sims DW, Bradshaw CJA (2013) Inferred global connectivity of whale shark *Rhincodon typus* populations. *Journal of Fish Biology*, **82**, 367–389.
Shapiro B, Ho SYW (2014) Ancient hyaenas highlight the old problem of estimating evolutionary rates. *Molecular Ecology*, **23**, 499–501.
Silas EG (1986) The whale shark (*Rhincodon typus* Smith) in Indian coastal waters: is the species endangered or vulnerable? *CMFRI, Marine Fisheries Information Service* **66**, 1–19. Available at http://eprints.cmfri.org.in/2774/1/MFIS_66-1.pdf/.
Slatkin M (1995) A measure of population subdivision based on microsatellite allele frequencies. *Genetics*, **139**, 457–462.
Slatkin M (2005) Seeing ghosts: the effect of unsampled populations on migration rates estimated for sampled populations. *Molecular Ecology*, **14**, 67–73.
Speed CW, Meekan MG, Rowat D, Pierce S, Marshall AD, Bradshaw CJA (2008) Scarring patterns and relative mortality rates of Indian Ocean whale sharks. *Journal of Fish Biology*, **72**, 1488–1503.
Stevens JD (2007) Whale shark (*Rhincodon typus)* biology and ecology: a review of the primary literature. *Fisheries Research*, **84**, 4–9.
Tajima F (1989) Statistical-method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, **123**, 585–595.
Vignaud T, Clua E, Mourier J, Maynard J, Planes S (2013) Microsatellite analyses of blacktip reef sharks (*Carcharhinus melanopterus*) in a fragmented environment show structured clusters. *PLoS One*, **8**, e61067.

Walker M, Johnsen S, Rasmussen SO *et al.* (2009) Formal definition and dating of the GSSP (Global Stratotype Section and Point) for the base of the Holocene using the Greenland NGRIP ice core, and selected auxiliary records. *Journal of Quaternary Science*, **24**, 3–17.

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.

Wegmann D, Excoffier L (2010) Bayesian inference of the demographic history of chimpanzees. *Molecular Biology and Evolution*, **27**, 1425–1435.

Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population structure. *Evolution*, **38**, 1358–1370.

WildLifeRisk Report (2014) *Planet's Biggest Slaughter of Whale Sharks Exposed in PuQi*. Paul Hilton & Alex Hofford, Zhejiang, China.

Williams JT, Delrieu-Trottin E, Planes S (2012) A new species of Indo-Pacific fish, *Canthigaster criobe*, with comments on other *Canthigaster* (Tetraodontiformes: Tetraodontidae) at the Gambier Archipelago. *Zootaxa*, **3523**, 80–88.

Wintner SP (2000) Preliminary study of vertebral growth rings in the whale shark, *Rhin- codon typus,* from the East Coast of South Africa. *Environmental Biology of Fishes*, **59**, 441–451.

T.M.V, S.P and M.G.M designed the study. J.M. and T.M.V. wrote the manuscript with input from S.P., M.G.M. and all other authors. S.P. is the lead P.I. on the grant with the primary funding source. D.R., D.R.-M., M.G.M., M.L.B., R.V.-J. and S.J.P. provided samples. T.M.V., C.B. and R.L. analysed the data and collaborated with JM to present the findings. S.B. and T.M.V. conducted laboratory procedures.

## Data accessibility

Microsatellites genotype and raw sequences with sampling locations are both accessible at the following doi:10.5061/dryad.489s0.

## Supporting information

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

**Table S1** PCR details for whale shark microsatellite DNA.