

Population genetic structure of Earth's largest fish, the whale shark (*Rhincodon typus*)

A. L. F. CASTRO,* B. S. STEWART,† S. G. WILSON,† R. E. HUETER,‡ M. G. MEEKAN,§ P. J. MOTTA,* B. W. BOWEN¶ and S. A. KARL¶

*Department of Biology, University of South Florida, SCA110, 4202 E. Fowler Ave., Tampa, FL 33620, USA, †Hubbs-SeaWorld Research Institute, 2595 Ingraham Street, San Diego, CA 92109, USA, ‡Center for Shark Research, Mote Marine Laboratory, 1600 Ken Thompson Parkway, Sarasota, FL 34236, USA, §Australian Institute of Marine Science, PO Box 40197, Casuarina MC, Darwin, NT 0811, Australia, ¶Hawai'i Institute of Marine Biology, University of Hawai'i, Manoa, PO Box 1346, Kane'ohe, HI 96744, USA

Abstract

Large pelagic vertebrates pose special conservation challenges because their movements generally exceed the boundaries of any single jurisdiction. To assess the population structure of whale sharks (*Rhincodon typus*), we sequenced complete mitochondrial DNA control regions from individuals collected across a global distribution. We observed 51 single site polymorphisms and 8 regions with indels comprising 44 haplotypes in 70 individuals, with high haplotype ($h = 0.974 \pm 0.008$) and nucleotide diversity ($\pi = 0.011 \pm 0.006$). The control region has the largest length variation yet reported for an elasmobranch (1143–1332 bp). Phylogenetic analyses reveal no geographical clustering of lineages and the most common haplotype was distributed globally. The absence of population structure across the Indian and Pacific basins indicates that oceanic expanses and land barriers in Southeast Asia are not impediments to whale shark dispersal. We did, however, find significant haplotype frequency differences (AMOVA, $\Phi_{ST} = 0.107$, $P < 0.001$) principally between the Atlantic and Indo-Pacific populations. In contrast to other recent surveys of globally distributed sharks, we find much less population subdivision and no evidence for cryptic evolutionary partitions. Discovery of the mating and pupping areas of whale sharks is key to further population genetic studies. The global pattern of shared haplotypes in whale sharks provides a compelling argument for development of broad international approaches for management and conservation of Earth's largest fish.

Keywords: conservation genetics, control region, marine phylogeography, migration, mitochondrial DNA, tandem repeats

Received 22 May 2007; revision accepted 1 October 2007

Introduction

The vastness of Earth's oceans often conceals regional biological processes particularly for pelagic and highly migratory species. Many species of shark, billfish, and tuna mature and forage far from shore. Some marine mammals and sea turtles approach land to breed or rest, but spend most of their lives beyond shoreline-based jurisdictions. Moreover, large marine vertebrates often have complex migratory behaviours that vary with age and sex (e.g. Brown *et al.* 1995; Pardini *et al.* 2001; Bowen *et al.* 2005; Carlsson *et al.* 2007).

Although the natural histories of many pelagic migrants have been illuminated by genetic markers, little is known about the biology and biogeography of the whale shark (*Rhincodon typus*). Whale sharks appear to be widely distributed in tropical and warm temperate seas (30°N and 35°S) except, perhaps, in the Mediterranean (Compagno 2001). Most information about general distribution, however, is either from seasonal sightings in scattered locations or anecdotal observations (Colman 1997). Aggregations of whale sharks have been routinely reported off Western Australia; Belize; the Yucatan peninsula and Baja California, Mexico; India; Djibouti; Taiwan; Japan; and the Philippines (Colman 1997; Compagno 2001; Stewart & Wilson 2005). Some aggregations occur year-round while others may be

Correspondence: Stephen A. Karl, Fax: 808-236-7443; E-mail: skarl@hawaii.edu

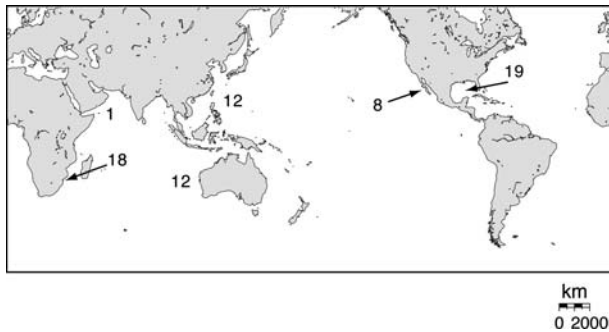


Fig. 1 Geographical distribution and number of whale shark specimens obtained for each geographical location.

associated with seasonal abundance of prey. Most known aggregations are comprised predominantly of immature sharks and segregation by size and sex may occur in some areas (Colman 1997; Compagno 2001). Even though recent studies have demonstrated the ability of this species to migrate long distances (e.g. Eckert & Stewart 2001; Wilson *et al.* 2006), it is not clear whether whale shark populations are panmictic or composed of reproductively isolated subpopulations. Recent evidence of cold-water tolerance when diving (Wilson *et al.* 2006) indicates that low temperatures and perhaps even subpolar waters may not be impediments to whale shark movement. Here, we present a population genetics survey of this widely distributed species using mtDNA control region (CR) sequences. As whale shark numbers appear to be declining in some regions (Stewart & Wilson 2005; Theberge & Dearden 2006), these findings have implications for global management and conservation.

Materials and methods

Sample collection and laboratory procedures

Skin samples from 70 whale sharks were collected by biopsy punch on a pole spear for live animals, or direct sampling of dead specimens, between 1992 and 2005 and were preserved in either salt-saturated dimethyl sulphoxide (DMSO) solution or 95% ethanol and stored at room temperature. Collections were made at Ningaloo Reef, Australia ($N = 12$); Gulf of California, Mexico (8); Pamilacan Island, Philippines (3); Hualien, Taiwan (9); KwaZulu-Natal, South Africa (5); Buinjata, Mozambique (8); Mombasa, Kenya (5); Vattanu Kandu, Maldives (1); Quintana Roo, Mexico (17); and Gulf of Mexico, USA (2) (Fig. 1).

We extracted total genomic DNA using a phenol-chloroform–isoamyl alcohol protocol (Sambrook *et al.* 1989) or 5% Chelex nonboiling protocol (Walsh *et al.* 1991). The mitochondrial CR was amplified using primers developed within the tRNA^{Pro} (WSCR1-F: 5'-TTGGCTCCCAAAGCC-AAGATTCTTC-3') and tRNA^{Phe} (WSCR1-R: 5'-GCATG-

TATAATTTTGGTTACAA-3'). Because of the large size of the CR (~1100–1325 nucleotides), two internal primers were designed to facilitate sequencing of the whole region. Primer WSCR2-R (5'-CTTAATATTTATTGTTCTGTTTCAGCC-3') was paired with WSCR1-F, and primer WSCR2-F (5'-CTATAATTGATTTAAACTGACATTG-3') was paired with WSCR1-R producing two overlapping fragments approximately 950 bp and 700 bp, respectively. Amplification reactions were carried out in 50- μ L volumes consisting of 1 \times Promega buffer (Promega), 1.25 U of IDProof DNA polymerase (ID Laboratories Inc.), 0.8 mM dNTPs, 2 mM MgCl₂, 0.5 μ M of each primer, 6.0 μ g bovine serum albumin, and 1–3 μ L of template. Cycling conditions for all primer pairs consisted of 95 °C 1 min, 35–40 cycles of 95 °C 45 s, 58 °C 60 s, and 72 °C 90 s with a final extension at 72 °C for 7 min. Amplicons were purified with QIAquick kit (QIAGEN) following the manufacturer's instructions. Both strands were sequenced using an ABI 3730XL Genetic analyser (Applied Biosystems, Inc.).

Data analysis

Control region alignments were optimized in SEQUENCHER 4.1 (Gene Codes Corporation) and gaps were introduced to maximize sequence similarity. Contiguous gaps were treated as a single event by omitting all but one of the gapped bases. Gaps were coded as transitions for distance-based analyses. In the case of substitutions within gaps, variable positions were retained and the gap was coded as a single transition. The Akaike information criteria within MODELTEST version 3.06 (Posada & Crandall 1998) was used to determine the best-fit model of evolution. A statistical parsimony network was constructed using tcs 1.2.1 (Clement *et al.* 2000).

Summary statistics (haplotype frequencies, number of polymorphic sites, number of transitions and transversions, and nucleotide composition) were estimated in ARLEQUIN 3.0 (Excoffier *et al.* 2005). Individuals were binned into five groups defined by geographical region: Quintana Roo, Mexico and Gulf of Mexico, USA in the northwestern Atlantic Ocean ($N = 19$); South Africa, Mozambique, and Kenya in the western Indian Ocean (18); Philippines and Taiwan in the northwestern Pacific Ocean (12); Western Australia in the eastern Indian Ocean (12); and the Gulf of California in the northeastern Pacific Ocean (8). Genetic diversity within localities was measured as the number of haplotypes, haplotype diversity (h), and nucleotide diversity (π) estimated with Nei's corrected average genetic divergence (Nei 1987) incorporating Tamura & Nei's (1993) model of sequence evolution with ARLEQUIN.

We used mismatch distributions for each sample to distinguish between population growth models, especially those invoking past exponential growth and historical population stasis (Slatkin & Hudson 1991; Rogers & Harpending 1992). Population parameters τ , θ_v , and θ_1

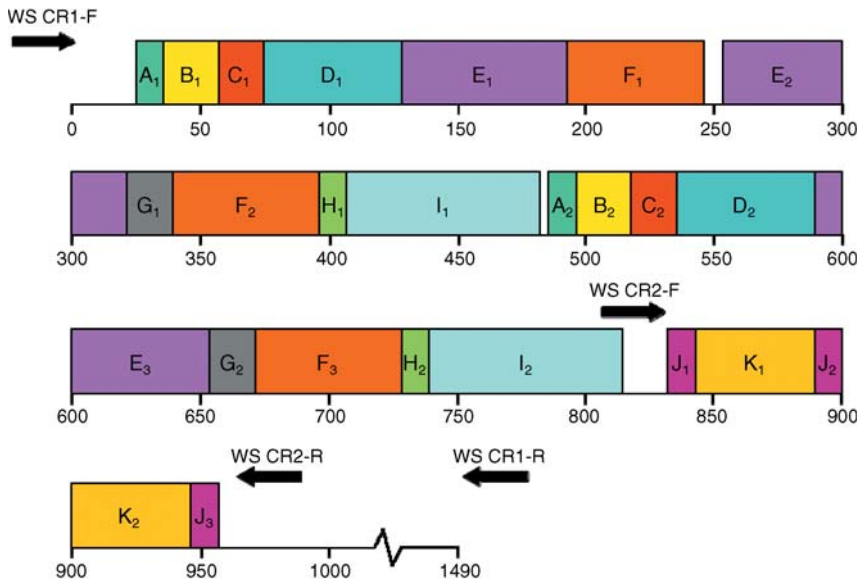


Fig. 2 Schematic diagram showing the consensus of all 44 haplotypes for the complete CR sequences of the whale shark. Coloured blocks represent different repeated fragments along the CR sequence. Similar sequences have the same colour and letter designation and repeats are numbered. Arrows represent primers used in PCR amplification.

were obtained from ARLEQUIN, where τ is the mutational timescale, and θ_0 and θ_1 are the expected pairwise differences before and after a change in population size (growth or contraction), respectively (Harpending 1994). The mutational timescale is $\tau = 2\mu t$, where t is measured in generations and μ is the mutation rate per generation for the entire sequence ($\mu = m_T u$, where m_T = number of nucleotides and u = mutation rate per nucleotide). The expected pairwise differentiation is $\theta = 2N_{ef}\mu$ where N_{ef} is the effective female population size.

Population subdivision and structure were estimated using an analysis of molecular variance (AMOVA, Excoffier *et al.* 1992) and pairwise population Φ_{ST} significance test (Cockerham & Weir 1993) as implemented in ARLEQUIN. Significance of Φ_{ST} was determined via nonparametric permutation (Excoffier *et al.* 1992) with 1000 data permutations. For AMOVA analyses, we used the Tamura & Nei's (1993) model of sequence evolution. Population differentiation also was tested using an exact test based on haplotype frequencies (Raymond & Rousset 1995) in ARLEQUIN.

Results

The mitochondrial CR from a total of 70 individuals ranged from 1143 to 1332 bp with a mean of 1236 bp (GenBank Accession nos EU182401 to EU182444 (see Supplementary material pending)). Nearly all of this size variation was due to indels composed of repeated sequence blocks (Fig. 2). Interestingly, although we did not specifically test for heteroplasmy, none was observed. Considering just the repeat unit structure (i.e. ignoring site substitutions), there were 11 different repeat motifs in the whale shark CR. Repeated blocks ranged in size from 9 (block A) to 64 bp (block E) long. All haplotypes had regions A₁ to D₁, E₂, F₂

E₃, and F₃ to J₃ (sequentially as indicated in Fig. 2) and this was the motif for the smallest haplotype, H19 (see Supplementary material). The largest haplotype, H10, had all the common repeats, some less common ones, and was the only haplotype to have block I1. Haplotypes H11 and H12 were similar to H19 except they possessed blocks E₁ and F₁ (totaling 103 bp) making H11 and H12 the second largest haplotypes.

We also found substitutions among repeated blocks within the same sequence. For example, repeat A₁ differed from A₂ by a substitution of one nucleotide in haplotype H23. Other examples included substitutions shared between different haplotypes including block B, which was repeated twice in nearly all haplotypes. For some haplotypes, these were perfect repeats whereas there were single transitional changes in others. Clearly, both larger indel changes and smaller substitutional changes are common in the evolution of the whale shark CR.

To maximize sequence similarity among specimens, the complete DNA sequence alignment required multiple gaps of sizes ranging from 1 to 163 bp. There were 49 single site substitutions (38 transitions and 11 transversions) and eight gaps resolving 44 haplotypes. Five of the sequences in the gapped regions were monomorphic, while the other three each had substitutions. Overall, the haplotype diversity (h), and nucleotide diversity (π) were relatively high ($h = 0.90$ – 1.0 , $\pi = 0.007$ – 0.016 ; Table 1). Among the 44 observed haplotypes, only nine were observed in more than a single shark (Table 2). Five of those haplotypes were found in only a single geographical region (H4 and H6 in the Atlantic Ocean, H5 and H7 in the western Indian Ocean, and H9 in the northwest Pacific Ocean), with the remainder being found in two or more locations. Except for some of the Atlantic Ocean haplotypes, there appeared

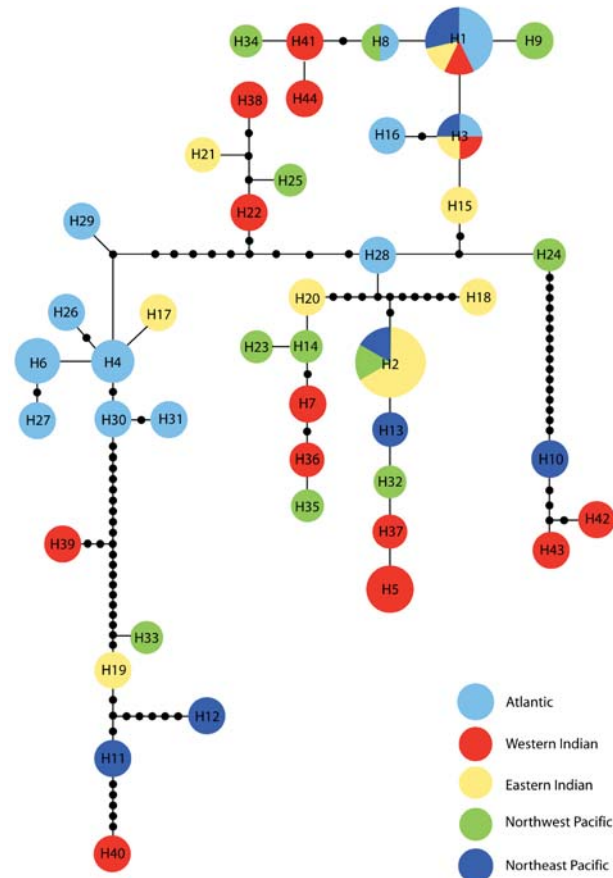


Fig. 3 Statistical parsimony network of haplotypes. All haplotypes are separated by one mutation and solid black circles represent hypothetical haplotypes not observed in this study. The size of the circle is proportional to the frequency of that haplotype.

to be no phylogeographical clustering of haplotypes (Fig. 3). There was statistically significant structure in whale shark populations with overall $\Phi_{ST} = 0.107$ ($P < 0.001$). The Atlantic Ocean population was significantly different from all other populations (Table 3), although small sample sizes temper this conclusion. Moreover, an exact test of haplotype frequencies shows divergence primarily between the Atlantic and Indian/Pacific Ocean populations (with the exception of the northeast Pacific Ocean; Table 3). There are two exceptions to the finding of genetic homogeneity across the Indo-Pacific region both involving the western Indian Ocean population. Neither of those, however, makes phylogeographical sense and both oppose the general trend of homogeneity. Hence, we consider these findings to be the result of stochastic sampling error.

The observed haplotype mismatch distribution was not significantly different from expectations under constant population size ($P = 0.90$; Fig. 4). Haplotypes H10 (northeast Pacific Ocean) and H42 and H43 (western Indian Ocean) contributed to this conclusion being distinct from all other

Table 1 Location, number of individuals (N), number of haplotypes (n), haplotype (h) and nucleotide (π) diversity estimates and standard deviations observed in the CR of the whale shark within five major ocean basins. Single individual from Maldives not included

| Geographical location | N | n | h | π |
|-----------------------|-----|-----|-----------------|-------------------|
| Atlantic | 19 | 12 | 0.93 ± 0.04 | 0.007 ± 0.002 |
| Western Indian | 18 | 14 | 0.95 ± 0.04 | 0.005 ± 0.003 |
| Eastern Indian | 12 | 9 | 0.91 ± 0.08 | 0.004 ± 0.002 |
| Northwest Pacific | 12 | 11 | 1.00 ± 0.04 | 0.005 ± 0.003 |
| Northeast Pacific | 8 | 7 | 0.96 ± 0.08 | 0.006 ± 0.004 |
| TOTAL | 69 | 44 | 0.97 ± 0.01 | 0.011 ± 0.006 |

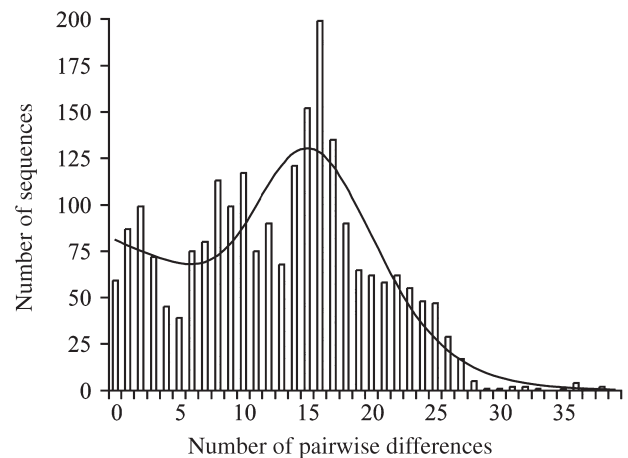


Fig. 4 Haplotype mismatch distribution. Note that nearly all comparisons with 11 or more differences between the sequences involve haplotypes H10, H42, and H43. The line is the expected frequency given a demographically stable population.

haplotypes by 8–18 substitutions (Fig. 4). There was no clear geographical clustering, however, and several haplotypes were shared among regions. Indeed, we detected haplotypes H1 and H3 in every region except the northwest Pacific Ocean.

The mutational timescale $\tau = 2\mu t$ can be used to estimate coalescence times for populations if generation time and mutation rate (μ) are available. Moreover, the initial and current effective population sizes (N_{f0} and N_{f1}) can be estimated from the pairwise differences θ_0 and θ_1 , if a mutation rate is available or estimated. Based on the observation of an adolescent female with a vertebral age estimate of 20 years (Wintner 2000), we provisionally apply a generation estimate of 25 years. The control region clock for the scalloped hammerhead, *Sphyrna lewini*, is 0.8% divergence between lineages per million years (Duncan *et al.* 2006) and is similar to a rate derived from control regions of lemon sharks (*Negaprion brevirostris*; J. Schultz, personal communication). In contrast, Keeney & Heist (2006) reported a rate of

| Haplotype | Geographical location | | | | | Total |
|-----------|-----------------------|----------------|----------------|-------------------|-------------------|-------|
| | Atlantic | Western Indian | Eastern Indian | Northwest Pacific | Northeast Pacific | |
| H1 | 3 | 1 | 1 | — | 2 | 7 |
| H2 | — | — | 4 | 1 | 1 | 7* |
| H3 | 1 | 1 | 1 | — | 1 | 4 |
| H4 | 4 | — | — | — | — | 4 |
| H5 | — | 4 | — | — | — | 4 |
| H6 | 3 | — | — | — | — | 3 |
| H7 | — | 2 | — | — | — | 2 |
| H8 | 1 | — | — | 1 | — | 2 |
| H9 | — | — | — | 2 | — | 2 |
| H10 | — | — | — | — | 1 | 1 |
| H11 | — | — | — | — | 1 | 1 |
| H12 | — | — | — | — | 1 | 1 |
| H13 | — | — | — | — | 1 | 1 |
| H14 | — | — | — | 1 | — | 1 |
| H15 | — | — | 1 | — | — | 1 |
| H16 | 1 | — | — | — | — | 1 |
| H17 | — | — | 1 | — | — | 1 |
| H18 | — | — | 1 | — | — | 1 |
| H19 | — | — | 1 | — | — | 1 |
| H20 | — | — | 1 | — | — | 1 |
| H21 | — | — | 1 | — | — | 1 |
| H22 | — | 1 | — | — | — | 1 |
| H23 | — | — | — | 1 | — | 1 |
| H24 | — | — | — | 1 | — | 1 |
| H25 | — | — | — | 1 | — | 1 |
| H26 | 1 | — | — | — | — | 1 |
| H27 | 1 | — | — | — | — | 1 |
| H28 | 1 | — | — | — | — | 1 |
| H29 | 1 | — | — | — | — | 1 |
| H30 | 1 | — | — | — | — | 1 |
| H31 | 1 | — | — | — | — | 1 |
| H32 | — | — | — | 1 | — | 1 |
| H33 | — | — | — | 1 | — | 1 |
| H34 | — | — | — | 1 | — | 1 |
| H35 | — | — | — | 1 | — | 1 |
| H36 | — | 1 | — | — | — | 1 |
| H37 | — | 1 | — | — | — | 1 |
| H38 | — | 1 | — | — | — | 1 |
| H39 | — | 1 | — | — | — | 1 |
| H40 | — | 1 | — | — | — | 1 |
| H41 | — | 1 | — | — | — | 1 |
| H42 | — | 1 | — | — | — | 1 |
| H43 | — | 1 | — | — | — | 1 |
| H44 | — | 1 | — | — | — | 1 |
| TOTAL | 19 | 18 | 12 | 12 | 8 | 70 |

Table 2 Geographical distribution of haplotypes found in 70 whale sharks from five major ocean basins

*One individual with H₂ was sampled in the Maldives but not listed by location in the table.

0.4% per million years for the control region in the blacktip shark (*Carcharhinus limbatus*). We provisionally applied both rates to whale sharks, with the caution that these three species are tens of millions of years divergent from *Rhincodon typus*. When analysed across all samples, these

data indicated coalescence times on the order of 1 600 000–3 200 000 years (Pliocene–Pleistocene boundary), with founding effective population sizes of $N_{f0} = 13\ 000$ –26 000 individuals, and current effective population size $N_{f1} = 119\ 000$ –238 000 individuals (Table 4).

| Populations | Atlantic | Western Indian | Eastern Indian | Northwest Pacific | Northeast Pacific |
|-------------------|----------|----------------|----------------|-------------------|-------------------|
| Atlantic | — | 0.004 | 0.011 | 0.008 | 0.235 |
| Western Indian | 0.215* | — | 0.037 | 0.028 | 0.258 |
| Eastern Indian | 0.196* | 0.072* | — | 0.173 | 0.768 |
| Northwest Pacific | 0.163* | 0.001 | 0.002 | — | 0.499 |
| Northeast Pacific | 0.208* | 0.037 | 0.000 | 0.000 | — |

Table 4 Time to common ancestor (τ), founding effective female population size (from θ_0), and current effective female population size (from θ_1) based on a coalescence approach

| Parameter | Divergence rate | |
|------------|------------------------|------------------------|
| | 0.8% per million years | 0.4% per million years |
| τ | 14.300 | 1 600 000 years |
| θ_0 | 3.011 | 13 000 |
| θ_1 | 27.107 | 119 000 |

Discussion

Our global survey of whale sharks indicates unusual size polymorphism in the CR, significant population structure between the Atlantic and Indian-Pacific ocean basins, and coalescence times on the order of 2–3 million years. Before interpreting these results, we address two caveats:

- 1 Sample size is small and lapses in coverage include the South Atlantic, Central Pacific, and South Pacific oceans. Sample sizes clearly limit inferences and we temper our conclusions accordingly. There are no directed oceanic surveys for whale sharks, as there are for tuna, billfish, and sea turtles, and the species occurs at low densities even in regional aggregates. The sample size of 70 represents over a decade of effort and this study is the most comprehensive genetic evaluation of this rare and enigmatic species (see Ramírez-Macías *et al.* 2007 for an analysis of whale sharks in the Gulf of California, Mexico). The observation of shared haplotypes (e.g. H1–H3) across the extremes of the geographical range is a robust finding that will not change regardless of sample sizes.
- 2 Estimates of generation time and mutation rate are provisional, and the latter is derived from distantly related sharks. The corresponding estimates of coalescence times and effective population sizes should be regarded as general indicators. Shark mtDNA appears to evolve about an order of magnitude slower than for bony fishes (Martin *et al.* 1992), consistent with our clock estimates. Consequently, corresponding estimates are useful in a qualitative sense for determining whether, for example, population histories coalesce at 10^5 , 10^6 , or 10^7 years. The

Table 3 Estimate of pairwise Φ_{ST} values of whale sharks from five major ocean basins using Tamura & Nei (1993) genetic distances (below diagonal) and exact test significance values (above). The Φ_{ST} values marked with * are significant ($P = 0.05$)

estimation of generation time is also somewhat controversial. The generation time of 25 years was based on the size–age relationship of a single immature female (Wintner 2000). Hoelzel *et al.* (2006) used a generation time of 16 years for the basking shark. If the generation time for the whale shark is similar (i.e. shorter than we estimate), our population estimates can be considered conservative. Nonetheless, we suggest caution in accepting genetically effective population sizes as being accurate within no less than an order of magnitude.

Control region structure

The CR in whale sharks (1143–1332 bp) is larger than that observed in most cartilaginous fishes. Amplification of the CR in 52 elasmobranch species has indicated a length of 1030–1050 bp except for the barndoor skate (*Dipturus laevis*), which is ~1200 bp (Stoner *et al.* 2003). Other sharks have a shorter CR (spiny dogfish, *Squalus acanthias*, 1080 bp, Rasmussen & Arnason 1999; starspotted smooth-hound, *Mustelus manazo*, 1068 bp, Cao *et al.* 1998; horn shark, *Heterodontus francisi*, 1068 bp, Arnason *et al.* 2001; lesser spotted dogfish, *Scyliorhinus canicula*, 1050 bp, Delarbre *et al.* 1998), or one comparable in length (white shark, *Carcharodon carcharias*, 1146 bp, Pardini *et al.* 2001) to the smallest whale shark CR. Size variation in the CR of whale sharks is also higher than that reported for other sharks (Kitamura *et al.* 1996; Pardini *et al.* 2001; Keeney *et al.* 2005), with a 189-bp difference between the largest and smallest amplicon.

Variation in the size of the control region has been reported for a substantial number of bony fishes (Lee *et al.* 1995; Brown *et al.* 1996; Bentzen *et al.* 1998; Hoarau *et al.* 2002). In bony fishes, the CR typically consists of tandem repeats, as observed in the whale shark (Fig. 2). Our initial attempts to polymerase chain reaction (PCR)-amplify the CR of whale sharks using a variety of published shark primers failed, probably because of the highly duplicated nature of the CR. Because the rate and pattern of these mutations are unknown, most studies have not used size variants as population markers. Insertions and deletions of repeat blocks may be relatively common, and homoplasy (convergence on the same number of repeats) is likely to confound any genealogical analysis.

Genetic diversity and effective population size

Despite an apparent decline in whale sharks abundance in some regions (e.g. Stewart & Wilson 2005; Theberge & Dearden 2006; Bradshaw *et al.* 2007), there is still relatively high genetic diversity in the species. Declining populations are expected, however, to retain historic levels of genetic diversity if the decline has occurred only recently (Roman & Palumbi 2003; Bowen *et al.* 2007). In the only other global surveys of shark CRs, the blacktip shark (*Carcharhinus limbatus*), yielded $h = 0.843 \pm 0.015$ and $\pi = 0.004 \pm 0.002$ (Keeney & Heist 2006), and the scalloped hammerhead (*Sphyrna lewini*) had $h = 0.800 \pm 0.020$ and $\pi = 0.013 \pm 0.007$ (Duncan *et al.* 2006), compared to $h = 0.970 \pm 0.010$ and $\pi = 0.011 \pm 0.006$ for whale sharks. These values are typical of abundant, geographically widespread shark species (cf. Heist 1999; Heist 2004).

The extant population size of whale shark is unknown. The nucleotide diversity values for blacktip sharks and scalloped hammerheads correspond to $N_{ef} = 140\,000$ compared to our estimates of $N_{ef} = 119\,000\text{--}238\,000$ for whale sharks. This is surprising given that blacktip sharks and hammerheads are globally distributed, abundant, coastal species, whereas the known dozen or so aggregates of whale sharks typically consist of tens to a few hundred individuals (Bradshaw *et al.* 2007). Two general processes might contribute to the relatively high genetic diversity that we have documented in whale sharks: (i) secondary contact between divergent allopatric lineages, or (ii) large stable populations. Except perhaps for haplotypes H10, H42, and H43, the mtDNA phylogeny reveals no evidence of distinct evolutionary lineages now in sympatry and the mismatch distribution indicates a relatively large, stable population.

The large N_{ef} of 119 000–238 000 females indicates that the surface waters supporting transient feeding aggregations (nearly all of the living *Rhincodon typus* observations) are not the sole, or perhaps even principal, habitat of adult whale sharks. Recent telemetry studies demonstrate that whale sharks occupy habitat far from shore and often in relatively deep, cold water for transient periods (Wilson *et al.* 2006). Although whale sharks are not known to possess anatomical, physiological or behavioural adaptations to conserve heat, the large body mass of adults may provide sufficient thermal inertia to allow extended cold-water exposure (Sims 2003; Wilson *et al.* 2006). Regardless of the extent of geographical and vertical population movements, it is clear that much of the habitat for this species is still unknown, and population sizes may be considerably larger than previously assumed.

Population structure

Our genetic studies indicate that whale shark aggregations within ocean basins are substantially interconnected on an

evolutionary timescale. Because a majority of our samples were collected from seasonal feeding aggregations (seven were from stranded or fishery-caught sharks), we cannot determine whether this pattern is due to ocean-wide interbreeding or to physical mixing in seasonal foraging areas of sharks from different breeding populations. In genetic studies of organisms with complex, multiphase life histories, conclusions regarding population subdivisions are fundamentally influenced by the particular life-history phase sampled (Bowen & Karl 2007). It is possible that whale shark populations consist of discrete, differentiated breeding units, which would not be detected in a survey of feeding aggregates. Regardless, our finding of haplotype sharing between ocean basins is consistent with the potential for global migrations allowing for population mixing over large distances. If this is the case, any breeding population subdivision would likely be primarily based on behavioural attributes. Clearly, more research on the basic life history of whale sharks is needed before questions of complex population structure in whale sharks can be addressed.

Whale shark population structure is low, even against the standards of large migratory fishes and whales. Bluefin tuna (*Thunnus thynnus*) show subtle ($\Phi_{ST} = 0.013$) but significant population structure between the western Atlantic Ocean (Gulf of Mexico) and the Mediterranean Sea, separated by ~11 000 km (Carlsson *et al.* 2007). The sailfish *Istiophorus platypterus* also is divided among ocean basins with significant population structure within the Pacific Ocean (Graves & McDowell 2003). Blue marlin (*Makaira nigricans*) are strongly divided among ocean basins ($\Phi_{ST} = 0.217$, Buonaccorsi *et al.* 2001). Whales show similar patterns of inter-ocean differentiation. Humpback (*Megaptera novaeangliae*, Baker *et al.* 1994), minke (*Balaenoptera acutorostrata*, van Pijlen *et al.* 1995), fin (*Balaenoptera physalus*, Bérubé *et al.* 1998), and Cuvier's beaked (*Ziphius cavirostris*, Dalebout *et al.* 2005) whales all have pronounced inter-ocean subdivision and some population structure within ocean basins. Barriers to movement within and between ocean basins generally appear to be stronger for most whales and large pelagic fishes than for whale sharks. These comparisons indicate that pelagic expanses can be barriers to highly mobile species, whereas the only apparent barriers to whale sharks may be geological and thermal (see below).

Marine phylogeography

In recent years there has been renewed interest in the biogeographical barrier between the Indian and Pacific oceans, enhanced by substantially lower sea levels during glacial maxima. While this barrier is consistent with evolutionary separations in small marine invertebrates (Barber *et al.* 2000), it is a less substantial, albeit still significant, population barrier to marine fishes and sharks

(Chenoweth & Hughes 2003; Duncan *et al.* 2006; Keeney & Heist 2006; Craig *et al.* 2007). Whale shark dispersal ability appears to be unimpeded by this intermittent barrier. It is not clear why some species are affected more than others by historical barriers across the Indo-Pacific region, although habitat preference is likely a dominant influence (Rocha *et al.* 2002) and pelagic species may disperse more effectively than reef-associated organisms. Over evolutionary time, whale shark migratory routes may be flexible enough to accommodate newly submerged habitats or perhaps connectivity can be quickly re-established after glacial intervals of tens of thousands of years.

The last tropical connection between the Atlantic and Indo-Pacific Oceans ended with the rise of the isthmus of Panama, about 3.5 million years ago (Coates & Obando 1996). In contemporary biogeography, the southern extensions of Africa and South America are regarded as formidable impediments to tropical connectivity. Faunas of the Atlantic and Indo-Pacific oceans share connections on a scale shorter than 3.5 million years, however, indicating dispersal around southern Africa (Bowen *et al.* 1998, 2001). Such exchanges are rare because of the cold Benguela Current along western South Africa (Gibbons & Thibault-Botha 2002), and might occur on a scale of 10^5 – 10^6 years (Roberts *et al.* 2004; Rocha *et al.* 2005; Bowen *et al.* 2006).

In a compilation of whale shark strandings and sightings in South Africa, Beckley *et al.* (1997) confirmed the occurrence of whale sharks along the frigid Atlantic coast. To explain the sporadic stranding in this area, they suggested that sharks arriving from the Indian Ocean succumb to the cold upwelling water and quickly perish. In contrast, Wilson *et al.* (2006) demonstrated that whale sharks can inhabit cold water, although perhaps not indefinitely. Whale shark dives have recently been recorded to ~1400 m (R.E. Hueter, J. Tyminski, C. Simpfendorfer, R. de la Parra, M. Trigo Mendoza, unpublished data). A deep, hour-long, cold-water dive in the tropics can be offset with a return to warm surface waters. In the Benguela upwelling system, however, cold water extends to the surface and no such relief is possible. Nonetheless, the sharing of haplotypes between Atlantic, Indian, and Pacific Ocean locations indicates a relatively recent connection via southern Africa. Whale sharks could have moved between the Atlantic and Indian oceans during a hiatus of Benguela upwelling that occurred between Pleistocene glacial epochs (Chang *et al.* 1999; Flores *et al.* 1999). Immediately following each ice age (100 000 to 400 000 years ago, but most recently 10 000 to 20 000 years ago), tropical plankton appears in sediment cores off southwestern Africa, indicating an avenue of warm water into the south Atlantic (Peeters *et al.* 2004). Contemporary movement is also possible. Warm-core gyres from the Indian Ocean occasionally become entrained in the northward moving Benguela Current, feeding into the central Atlantic Ocean (Flores *et al.* 1999; Penven *et al.* 2001).

Warm core gyres originating from the North Atlantic Gulf Stream are thought to be responsible for sightings of whale sharks in cold temperate areas such as the Bay of Fundy (44°15'N; Turnbull & Randell 2006). Regardless, historic or ongoing gene flow is apparently limited as indicated by the significant global $\Phi_{ST} = 0.107$.

Conservation implications

This first worldwide genetic survey of whale sharks indicates significant population structure on a global scale. Management units for whale sharks may span over 8000 km in the Atlantic Ocean, and over 16 000 km in the Indian-Pacific Ocean. Any management plan for whale sharks must consider that feeding aggregations may draw individuals from an ocean-wide population to a single location. Unilateral management in any single political jurisdiction will be inadequate for a highly mobile species that travels through multiple political jurisdictions. Indeed, tracking studies and our mtDNA data both indicate that management plans for the Earth's largest fish will, at a minimum, require ocean-basin-wide cooperation and governance.

Acknowledgements

We thank the major contributors to this study including the organizations and persons who provided tissue samples (D. Abercrombie, J. Brunnschweiler, R. Pittman, M. Shivji, T. Thys, J. Tyminski, S. Winter, Georgia Aquarium, and Mote Marine Laboratory) as well as those that provided laboratory assistance and data analysis (A.L. Bass, C. Curtis, K. Hayes, C. Puchulutegui, and T. Schwartz). We also thank A. Martin, J. Zamzow, and three anonymous reviewers for helpful comments on the manuscript. Funding for this project was provided in part by a CAPES Fellowship (BEX 1277-02/2) to A.L.F.C., an NSF grant DEB 03-21924 to S.A.K., a Georgia Aquarium grant to R.E.H., and an NSF grant OCE-0454873 to B.W.B., with additional funding provided by Hubbs-Sea World Research Institute, the Hubbs Society, and Mote Marine Laboratory. This is SOEST contribution no. 7195, and HIMB contribution no. 1295.

References

- Arnason U, Gullberg A, Janke A (2001) Molecular phylogenetics of gnathostomous (jawed) fishes: old bones, new cartilage. *Zoologica Scripta*, **30**, 249–255.
- Baker CS, Slade RW, Bannister JL *et al.* (1994) Hierarchical structure of mitochondrial DNA gene flow among humpback whales *Megaptera novaeangliae*, world-wide. *Molecular Ecology*, **3**, 313–327.
- Barber PH, Palumbi SR, Erdmann MV, Moosa MK (2000) A marine Wallace's line? *Nature*, **406**, 692–693.
- Beckley LE, Cliff G, Smale MJ, Compagno LJV (1997) Recent strandings and sightings of whale sharks in South Africa. *Environmental Biology of Fishes*, **50**, 343–348.
- Bentzen P, Wright JL, Bryden LT, Sargent M, Zwanenburg KCT (1998) Tandem repeat polymorphism and heteroplasmy in the mitochondrial control region of redfishes (*Sebastes*: Scorpaenidae). *Journal of Heredity*, **89**, 1–7.

- Bérubé M, Aguilar A, Dendanto D *et al.* (1998) Population genetic structure of North Atlantic, Mediterranean Sea and Sea of Cortez fin whales, *Balaenoptera physalus* (Linnaeus 1758): analysis of mitochondrial and nuclear loci. *Molecular Ecology*, **7**, 585–599.
- Bowen BW, Karl SA (2007) Population genetics and phylogeography of sea turtles. *Molecular Ecology*, doi: 10.1111/j.1365-294X.2007.03542.X.
- Bowen BW, Clark AM, Abreu-Grobois FA *et al.* (1998) 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, Bass AL, Soares L, Toonen RJ (2005) Conservation implications of complex population structure: lessons from the loggerhead turtle (*Caretta caretta*). *Molecular Ecology*, **14**, 2389–2402.
- 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.
- Bowen BW, Grant WS, Hillis-Starr Z *et al.* (2007) Mixed stock analysis reveals the migrations of juvenile hawksbill turtles (*Eretmochelys imbricata*) in the Caribbean Sea. *Molecular Ecology*, **16**, 49–60.
- Bradshaw CJA, Mollet HF, Meekan MG (2007) Inferring population trends for the world's largest fish from mark-recapture estimates of survival. *Journal of Animal Ecology*, **73**, 480–489.
- Brown JR, Beckenbach K, Beckenbach AT, Smith MJ (1996) Length variation, heteroplasmy and sequence divergence in the mitochondrial DNA of four species of sturgeon (*Acipenser*). *Genetics*, **142**, 525–535.
- Brown MR, Corkeron PJ, Hale PT, Schultz KW, Bryden MM (1995) Evidence for a sex-segregated migration in the humpback whale (*Megaptera novaeangliae*). *Proceedings of the Royal Society B: Biological Sciences*, **259**, 229–234.
- Buonaccorsi VP, McDowell JR, Graves JE (2001) Reconciling patterns of inter-ocean molecular variance from four classes of molecular markers in blue marlin (*Makaira nigricans*). *Molecular Ecology*, **10**, 1179–1196.
- Cao Y, Waddell PJ, Okada N, Hasengawa M (1998) The complete mitochondrial DNA sequence of the shark *Mustelus manazo*: evaluating rooting contradictions to living bony vertebrates. *Molecular Biology and Evolution*, **15**, 1637–1646.
- Carlsson J, McDowell JR, Carlsson JEL, Graves JE (2007) Genetic identity of YOY bluefin tuna from the eastern and western Atlantic spawning areas. *Journal of Heredity*, **98**, 23–28.
- Chang Y-P, Chang C-C, Wang L-W, Chen M-T, Wang C-H, Yu E-F (1999) Planktonic foraminiferal sea surface temperature variation in the Southeast Atlantic Ocean: a high-resolution record MD962085 of the past 400 000 years from the IMAGES II – NAUSICAA cruise. *Terrestrial, Atmospheric and Oceanic Sciences*, **10**, 185–200.
- Chenoweth SF, Hughes JM (2003) Oceanic interchange and nonequilibrium population structure in the estuarine dependent Indo-Pacific tasselfish *Polynemus sheridani*. *Molecular Ecology*, **12**, 2387–2397.
- Clement M, Posada D, Crandall KA (2000) TCS: a computer program to estimate gene genealogies. *Molecular Ecology*, **9**, 1657–1660.
- 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.
- Cockerham CC, Weir BS (1993) Estimation of gene flow from *F*-statistics. *Evolution*, **47**, 855–863.
- Colman JG (1997) A review of the biology and ecology of the whale shark. *Journal of Fish Biology*, **51**, 1219–1234.
- Compagno LJV (2001) *Sharks of the World: An Annotated and Illustrated Catalogue of Shark Species Known to Date. Volume 2 Bullhead, Mackerel and Carpet Sharks (Heterodontiformes, Lamniformes and Orectolobiformes)*. Food and Agriculture Organization of the United Nations, species catalogue for fishery purposes, Rome, Italy.
- Craig MT, Eble JA, Robertson DR, Bowen BW (2007) High genetic connectivity across the Indian and Pacific Oceans in the reef fish *Myripristis berndti* (Holocentridae). *Marine Ecology Progress Series*, **334**, 245–254.
- Dalebout ML, Robertson KM, Frantzis A *et al.* (2005) Worldwide structure of mtDNA diversity among Cuvier's beaked whales (*Ziphius cavirostris*): implications for threatened populations. *Molecular Ecology*, **14**, 3353–3371.
- Delarbre C, Spruyt N, Delmarre C *et al.* (1998) The complete nucleotide sequence of the mitochondrial DNA of the dogfish, *Scyliorhinus canicula*. *Genetics*, **150**, 331–344.
- Duncan KM, Martin AP, Bowen BW, de Couet GH (2006) Global phylogeography of the scalloped hammerhead shark (*Sphyrna lewini*). *Molecular Ecology*, **15**, 2239–2251.
- Eckert SA, Stewart BS (2001) Telemetry and satellite tracking of the whale shark, *Rhincodon typus*, in the Sea of Cortez, Mexico, and the North Pacific Ocean. *Environmental Biology of Fishes*, **60**, 299–308.
- Excoffier L, Smouse P, Quattro J (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics*, **131**, 479–491.
- Excoffier L, Laval G, Schneider S (2005) ARLEQUIN ver. 3.0: an integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online*, **1**, 47–50.
- Flores J-A, Gersonde R, Sierro FJ (1999) Pleistocene fluctuations in the Agulhas Current Retroflection based on the calcareous plankton record. *Marine Micropaleontology*, **37**, 1–22.
- Gibbons MJ, Thibault-Botha D (2002) The match between ocean circulation and zoogeography of epipelagic siphonophores around southern Africa. *Journal of the Marine Biological Association of the UK*, **82**, 801–810.
- Graves JE, McDowell JR (2003) Stock structure of the world's istiophorid billfishes: a genetic perspective. *Marine and Freshwater Research*, **54**, 287–298.
- Harpending RC (1994) Signature of ancient population growth in a low-resolution mitochondrial DNA mismatch distribution. *Human Biology*, **66**, 591–600.
- Heist EJ (1999) A review of population genetics in sharks. *American Fishery Society Symposium*, **23**, 161–168.
- Heist EJ (2004) Genetics of sharks, skates and rays. In: *Biology of Sharks and Their Relatives* (eds Carrier JC, Musick JA, Heithaus MR), pp. 471–485. CRC Press, Boca Raton, Florida.
- Hoarau G, Holla S, Lescasse R, Stam WT, Olsen JL (2002) Heteroplasmy and evidence for recombination in the mitochondrial control region of the flatfish *Platichthys flesus*. *Molecular Biology and Evolution*, **19**, 2261–2264.
- Hoelzel AR, Shiviyi MS, Magnussen J, Francis MP (2006) Low worldwide genetic diversity in the basking shark. *Biology Letters*, **2**, 630–642.
- Keeney DB, Heist EJ (2006) Worldwide phylogeography of the blacktip shark (*Carcharhinus limbatus*) inferred from mitochondrial DNA reveals isolation of West Atlantic populations coupled with recent Pacific dispersal. *Molecular Ecology*, **15**, 3669–3679.

- Keeney DB, Heupel MR, Hueter RE, Heist EJ (2005) Microsatellite and mitochondrial DNA analyses of the genetic structure of blacktip shark (*Carcharhinus limbatus*) nurseries in the north-western Atlantic, Gulf of Mexico, and Caribbean Sea. *Molecular Ecology*, **14**, 1911–1923.
- Kitamura T, Takemura A, Watabe S, Taniuchi T, Shimizu M (1996) Mitochondrial DNA analysis for the cytochrome *b* gene and D-loop from the bull shark *Carcharhinus leucas*. *Fisheries Science*, **62**, 21–27.
- Lee WJ, Conroy J, Howell WH, Kocher TD (1995) Structure and evolution of teleost mitochondrial control region. *Journal of Molecular Evolution*, **41**, 54–66.
- Martin AP, Naylor GJP, Palumbi SR (1992) Rates of mitochondrial DNA evolution in sharks are slow compared with mammals. *Nature*, **357**, 153–155.
- Nei M (1987) *Molecular Evolutionary Genetics*. Columbia University Press, New York.
- Pardini AT, Jones CS, Noble LR *et al.* (2001) Sex-biased dispersal of great white sharks. *Nature*, **412**, 139–140.
- Peeters JC, Acheson R, Brummer G-JA *et al.* (2004) Vigorous exchange between the Indian and Atlantic Oceans at the end of the past five glacial periods. *Nature*, **430**, 661–665.
- Penven P, Lutjeharms JRE, Marchesiello P, Roy C, Weeks SJ (2001) Generation of cyclonic eddies by the Agulhas Current in the lee of the Agulhas Bank. *Geophysical Research Letters*, **28**, 1055–1058.
- van Pijlen IA, Amos B, Burke T (1995) Patterns of genetic variability at individual mini-satellite loci in minke whale *Balaenoptera acutorostrata* populations from three different oceans. *Molecular Biology and Evolution*, **12**, 459–472.
- Posada D, Crandall KA (1998) MODELTEST: testing the model of DNA substitution. *Bioinformatics*, **14**, 817–818.
- Ramírez-Macías D, Vázquez-Juárez R, Galván-Magaña F, Munguía-Vega A (2007) Variations of the mitochondrial control region sequence in whale sharks (*Rhincodon typus*) from the Gulf of California, Mexico. *Fisheries Research*, **84**, 87–95.
- Rasmussen AS, Arnason U (1999) Phylogenetic studies of complete mitochondrial DNA molecules place cartilaginous fishes within the tree of bony fishes. *Journal of Molecular Evolution*, **48**, 118–123.
- Raymond M, Rousset F (1995) An exact test for population differentiation. *Evolution*, **49**, 1280–1283.
- 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, Bass AL, Robertson DR, Bowen BW (2002) Adult habitat preferences, larval dispersal, and the comparative phylogeography of three Atlantic surgeonfishes (Teleostei: Acanthuridae). *Molecular Ecology*, **11**, 243–252.
- Rocha LA, Robertson DR, Rocha CR *et al.* (2005) Recent invasion of the tropical Atlantic by an Indo-Pacific coral reef fish. *Molecular Ecology*, **14**, 3921–3928.
- Rogers AR, Harpending H (1992) Population growth makes waves in the distribution of pairwise genetic differences. *Molecular Biology and Evolution*, **9**, 552–569.
- Roman J, Palumbi SR (2003) Whales before whaling in the North Atlantic. *Science*, **301**, 508–510.
- Sambrook J, Fritsch EF, Maniatis T (1989) *Molecular Cloning: A Laboratory Manual*, 2nd edn. Cold Spring Harbor Laboratory Press, New York.
- Sims DW (2003) Tractable models of testing theories about natural strategies: foraging behaviour and habitat selection of free-ranging sharks. *Journal of Fish Biology*, **63** (Suppl. A), 53–73.
- Slatkin M, Hudson RH (1991) Pairwise comparisons of mitochondrial DNA sequences in stable and exponentially growing populations. *Genetics*, **129**, 555–562.
- Stewart BS, Wilson SG (2005) Threatened fishes of the world: *Rhincodon typus* (Smith 1828) (Rhincodontidae). *Environmental Biology of Fishes*, **74**, 184–185.
- Stoner DS, Grady JM, Priede KA, Quattro JM (2003) Amplification primers for the mitochondrial control region and sixth intron of the nuclear-encoded lactate dehydrogenase A gene in elasmobranch fishes. *Conservation Genetics*, **4**, 805–808.
- Tamura K, Nei M (1993) Estimation of the number of nucleotide substitutions in the control region of the mitochondrial DNA in humans and chimpanzees. *Molecular Biology and Evolution*, **10**, 512–526.
- Theberge MM, Dearden P (2006) Detecting a decline in whale shark *Rhincodon typus* sightings in the Andaman Sea, Thailand, using ecotourist operator-collected data. *Oryx*, **40**, 337–342.
- Turnbull SD, Randell JE (2006) Rare occurrence of a *Rhincodon typus* (whale shark) in the Bay of Fundy Canada. *Northeastern Naturalist*, **13**, 57–58.
- Walsh HE, Metzger DA, Higuchi R (1991) Chelix medium for simple extraction of DNA for PCR-based typing from forensic material. *Biology Techniques*, **10**, 506–513.
- Wilson SG, Polovina JJ, Stewart BS, Meekan MG (2006) Movement of whale sharks (*Rhincodon typus*) tagged at Ningaloo Reef, Western Australia. *Marine Biology*, **148**, 1157–1166.
- Wintner SP (2000) Preliminary study of vertebral growth rings in the whale shark, *Rhincodon typus*, from the east coast of South Africa. *Environmental Biology of Fishes*, **59**, 441–451.

A. L. F. Castro is a Ph.D. student in the Biology Department of the University of South Florida. This study represents a portion of his doctoral research on phylogeography of elasmobranchs. B. S. Stewart studies population biology, foraging and physiological ecology, and behaviour of a variety of marine mammals, sea birds, sea turtles, and sharks. R. E. Hueter is director of Mote Marine Laboratory's Centre for Shark Research where he conducts research on anatomy, physiology, behavior, ecology and conservation of sharks. P. Motta is interested in the functional morphology and evolution of feeding in sharks and bony fishes. B. W. Bowen and S. A. Karl are researchers at the Hawaii Institute of Marine Biology, a premier marine molecular genetics facility sitting atop a coral reef in Kaneohe Bay, Hawaii.

Supplementary material

The following supplementary material is available for this article:

Figure S1 This figure shows the presence and absence of repeat units for each haplotype.

This material is available as part of the online article from:

<http://www.blackwell-synergy.com/doi/abs/10.1111/j.1365-294X.2007.03597.x>

(This link will take you to the article abstract).

Please note: Blackwell Publishing are not responsible for the content or functionality of any supplementary materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.