

Low Diversity in the Mitogenome of Sperm Whales Revealed by Next-Generation Sequencing

Alana Alexander^{1,2,*}, Debbie Steel^{1,2}, Beth Slikas^{1,2}, Kendra Hoekzema², Colm Carragher^{3,4}, Matthew Parks⁵, Richard Cronn⁶, and C. Scott Baker^{1,2,3}

¹Marine Mammal Institute, Hatfield Marine Science Center, Oregon State University

²Department of Fisheries and Wildlife, Oregon State University

³School of Biological Sciences, University of Auckland, Auckland, New Zealand

⁴New Zealand Institute for Plant & Food Research Limited, Auckland, New Zealand

⁵Department of Botany and Plant Pathology, Oregon State University

⁶Pacific Northwest Research Station, USDA Forest Service, Corvallis, Oregon

*Corresponding author: E-mail: alexaala@onid.orst.edu.

Accepted: December 13, 2012

Data deposition: Sequences have been submitted to GenBank and are currently in the process of curation. The GenBank accession numbers are listed in supplementary material S1, Supplementary Material online.

Abstract

Large population sizes and global distributions generally associate with high mitochondrial DNA control region (CR) diversity. The sperm whale (*Physeter macrocephalus*) is an exception, showing low CR diversity relative to other cetaceans; however, diversity levels throughout the remainder of the sperm whale mitogenome are unknown. We sequenced 20 mitogenomes from 17 sperm whales representative of worldwide diversity using Next Generation Sequencing (NGS) technologies (Illumina GAIIx, Roche 454 GS Junior). Resequencing of three individuals with both NGS platforms and partial Sanger sequencing showed low discrepancy rates (454-Illumina: 0.0071%; Sanger-Illumina: 0.0034%; and Sanger-454: 0.0023%) confirming suitability of both NGS platforms for investigating low mitogenomic diversity. Using the 17 sperm whale mitogenomes in a phylogenetic reconstruction with 41 other species, including 11 new dolphin mitogenomes, we tested two hypotheses for the low CR diversity. First, the hypothesis that CR-specific constraints have reduced diversity solely in the CR was rejected as diversity was low throughout the mitogenome, not just in the CR (overall diversity $\pi = 0.096\%$; protein-coding 3rd codon = 0.22%; CR = 0.35%), and CR phylogenetic signal was congruent with protein-coding regions. Second, the hypothesis that slow substitution rates reduced diversity throughout the sperm whale mitogenome was rejected as sperm whales had significantly higher rates of CR evolution and no evidence of slow coding region evolution relative to other cetaceans. The estimated time to most recent common ancestor for sperm whale mitogenomes was 72,800 to 137,400 years ago (95% highest probability density interval), consistent with previous hypotheses of a bottleneck or selective sweep as likely causes of low mitogenome diversity.

Key words: *Physeter macrocephalus*, nucleotide diversity, mitochondrial genome, mtDNA, substitution rates, Bayesian phylogenetics, cetacean, population genetics.

Introduction

Sperm whales (*Physeter macrocephalus*) are distributed throughout all oceans of the world, from the Arctic to the Southern Ocean (Best 1979; Whitehead 2003) and are abundant, with a world wide estimate of approximately 360,000 individuals (Whitehead 2002). Despite being hunted intensively over the last two centuries, sperm whales appear to have suffered only localized declines due to whaling (Best

1979; Whitehead 2002, 2003). Generally, large population sizes and widespread distributions are associated with high mitochondrial DNA (mtDNA) diversity (Mulligan et al. 2006). However, despite their global distribution and high abundance, sperm whales have been noted for extremely low mtDNA control region (CR) diversity relative to other cetaceans (Lyrholm and Gyllensten 1998; Whitehead 1998), both in terms of number of haplotypes and nucleotide diversity (table 1). To date, only 28 mtDNA haplotypes have

© The Author(s) 2012. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Table 1

Estimates of Nucleotide Diversity and Number of mtDNA CR Haplotypes for 26 Cetacean Species (Limited to Studies with Sample Sizes above 100)

| Species | Sampling Location | Sample Size | CR Haplotype Length (bp) | π (%) | No. of Haplotypes | Reference |
|---|-------------------------------------|-------------|--------------------------|-----------|-------------------|-----------------------------------|
| Long-finned pilot whale ^a (<i>Globicephala melas</i>) | Pacific/Atlantic Ocean | 643 | 620 | 0.35 | 13 | Oremus et al. (2009) |
| False killer whale ^a (<i>Pseudorca crassidens</i>) | Worldwide | 124 | 945 | 0.37 | 24 | Chivers et al. (2007) |
| Commerson's dolphin (<i>Cephalorhynchus commersonii</i>) | South America | 196 | 466 | 0.40 | 20 | Pimper et al. (2010) |
| Finless porpoise (<i>Neophocaena phocaenoides</i>) | West Pacific | 386 | 345 | 0.27 | 16 | Li et al. (2011) |
| | China and Japan | 417 | 345 | 0.44 | 18 | Yang et al. (2008) |
| Sperm whale ^a (<i>Physeter macrocephalus</i>) | Worldwide | 1,167 | 399 | 0.51 | 28 | Studies listed in Table 2 |
| Beluga (<i>Delphinapterus leucas</i>) | Arctic | 324 | 410 | 0.51 | 29 | O'Corry-Crowe et al. (1997) |
| Killer whale ^a (<i>Orcinus orca</i>) | Worldwide | 102 | 995 | 0.52 | 13 | Hoelzel et al. (2002) |
| White-beaked dolphin (<i>Lagenorhynchus albirostris</i>) | North Atlantic | 122 | 601 | 0.56 | 18 | Banguera-Hinestroza et al. (2010) |
| Hector's dolphin (<i>C. hectori hectori</i>) | New Zealand | 318 | 650 | 0.79 | 21 | Hamner et al. (2012) |
| Short-finned pilot whale (<i>G. macrorhynchus</i>) | South Hemisphere and Atlantic | 150 | 620 | 0.87 | 14 | Oremus et al. (2009) |
| Common minke whale ^a (<i>Balaenoptera acutorostrata</i>) | North Atlantic | 102 | 500 | 0.60 | 26 | Pastene et al. (2007) |
| | Western North Pacific | 127 | 500 | 1.00 | 34 | Pastene et al. (2007) |
| Bryde's whale ^a (<i>B. brydei</i>) | Pacific/Indian Ocean | 472 | 299 | 1.00 | 51 | Kanda et al. (2007) |
| Dall's porpoise (<i>Phocoenoides dalli</i>) | North Pacific | 103 | 479 | 1.06 | 49 | Hayano et al. (2003) |
| Fin whale ^a (<i>B. physalus</i>) | North Atlantic | 341 | 288 | 1.13 | 48 | Bérubé et al. (1998) |
| Pantropical spotted dolphin ^a (<i>Stenella attenuata</i>) | East Pacific | 225 | 455 | 1.36 | 112 | Escorza-Treviño et al. (2005) |
| Antarctic blue whale (<i>B. musculus intermedia</i>) | South Ocean | 183 | 410 | 1.40 | 52 | Sremba et al. (2012) |
| Harbor porpoise (<i>Phocoena phocoena</i>) | NE Atlantic | 194 | 344 | 0.47 | 37 | Rosel et al. (1999) |
| | North Pacific | 249 | 402 | 1.40 | 88 | Chivers et al. (2002) |
| Southern right whale (<i>Eubalaena australis</i>) | New Zealand | 585 | 500 | 1.43 | 12 | Carroll et al. (2011) |
| Antarctic minke whale (<i>B. bonaerensis</i>) | South Hemisphere | 119 | 500 | 1.50 | 83 | Pastene et al. (2007) |
| Dusky dolphin (<i>L. obscurus</i>) | South Hemisphere | 153 | 591 | 1.63 | 62 | Cassens et al. (2005) |
| Pygmy sperm whale ^a (<i>Kogia breviceps</i>) | Worldwide | 108 | 406 | 1.65 | 74 | Chivers et al. (2005) |
| Atlantic spotted dolphin (<i>S. frontalis</i>) | North Atlantic | 196 | 329 | 1.47 | 34 | Adams and Rosel (2006) |
| | Azores | 144 | 611 | 1.80 | 76 | Quérouil et al. (2010) |
| Short-beaked common dolphin ^a (<i>Delphinus delphis</i>) | North Atlantic | 297 | 360 | 1.80 | 77 | Mirimin et al. (2009) |
| Bottlenose dolphin ^a (<i>Tursiops truncatus</i>) | Eastern North Atlantic | 123 | 630 | 1.6 | 41 | Natoli et al. (2005) |
| | Hawaii | 130 | 400 | 2.20 | 25 | Martien et al. (2012) |
| Western Australia bottlenose dolphin (<i>Tursiops sp.</i>) | West Australia | 220 | 351 | 2.21 | 8 | Krützen et al. (2004) |
| Humpback whale ^a (<i>Megaptera novaeangliae</i>) | Mozambique and Eastern South Africa | 151 | 486 | 1.95 | 65 | Rosenbaum et al. (2009) |
| | North Atlantic and Antarctic | 136 | 288 | 2.60 | 31 | Palsbøll et al. (1995) |

NOTE.—For species with studies from multiple geographic regions, the studies with lowest and highest nucleotide diversity are presented. If multiple studies were available from the same geographic region, only the study with the lowest nucleotide diversity was retained. Species are ordered by nucleotide diversity from low to high (species with multiple estimates are ordered by greatest estimate of nucleotide diversity), with dark blue backgrounds indicating low nucleotide diversity/numbers of haplotypes, through to white indicating higher estimates of nucleotide diversity/numbers of haplotypes. The sperm whale is highlighted in yellow.

^aSpecies that have worldwide ranges.

been reported from 1,167 samples of sperm whales, based on 400 bp of the CR (table 2). The widespread geographic sampling of these haplotypes, and the relatively high abundance of sperm whales, argues against explanations of biased geographic sampling or a whaling-mediated bottleneck (Whitehead 1998). The low CR diversity is also reflected in the weak mtDNA CR phylogeographic structure found in the sperm whale (Lyrholm and Gyllensten 1998): The three most common worldwide haplotypes ("A," "B," and "C": see table 2 for haplotype definitions) are shared in whales from the Pacific and Atlantic Oceans (Richard et al. 1996; Lyrholm and Gyllensten 1998; Whitehead et al. 1998; Engelhaupt et al. 2009; Mesnick et al. 2011; Ortega-Ortiz et al. 2012; Rendell et al. 2012).

A number of hypotheses have been proposed to explain the low diversity in the mtDNA CR of the sperm whale, including 1) an overall slowing of mtDNA substitution rates (Lyrholm et al. 1996; Whitehead 1998), as found for other large whale species (Jackson et al. 2009); 2) a population bottleneck (Lyrholm et al. 1996; Lyrholm and Gyllensten 1998); 3) a selective sweep of mtDNA by "hitchhiking" either with maternal cultural innovations in matrilineal social groups (Whitehead 1998, 2005), or variation in ocean productivity affecting the success of different social groups (Tiedemann and Milinkovitch 1999); and 4) lineage extinctions relating to the mass stranding of matrilineal groups (Whitehead 1998). We suggest a fifth hypothesis: that constraints in the CR can restrict sites that accumulate variation, leading to saturation of sites free to vary in the CR relative to mtDNA protein-coding regions. This hypothesis has been supported in other groups such as killer whales (*Orcinus orca*) and fishers (*Martes pennanti*) (Morin et al. 2010; Knaus et al. 2011), and is also consistent with previous observations of substitutional hot-spots in the sperm whale CR (Lyrholm et al. 1996).

Here, we extend estimates of sperm whale mtDNA CR diversity using Next Generation Sequencing (NGS) technologies (Illumina GAIIx and Roche GS Junior 454) to sequence 17 sperm whale mitogenomes chosen to represent the worldwide diversity of CR haplotypes (table 2). Given the expectation of low diversity in the sperm whale mitogenome, and the potential for NGS error to inflate diversity estimates (Shen et al. 2010), we first investigated discrepancy rates between the two NGS platforms by resequencing three mitogenomes. We also calculated NGS discrepancy rates with targeted comparison to Sanger sequencing for all sperm whale mitogenomes included in this study. Additionally, we developed quality control (QC) criteria to validate variable sites among the mitogenomes, and took a number of precautions against co-amplification of numts. We then aligned these sperm whale mitogenomes to the available NCBI Reference Sequence (RefSeq) cetacean mitogenomes and to the mtDNA protein-coding genes and CR from a further 11 dolphin species/subspecies contributed by this study (including

8 previously unpublished species/subspecies) for a total of 42 cetacean species included in our analyses.

Using this data set, we investigated two of the hypotheses for low sperm whale mtDNA CR diversity: 1) constraints acting solely on the CR resulting in reduced variation relative to the remainder of the mitogenome, or 2) slow substitution rates resulting in low diversity either in the CR or throughout the entire sperm whale mitogenome. To investigate the constraint hypothesis (hypothesis 1), we quantified intraspecific sperm whale diversity in the mtDNA CR relative to other regions of the mitogenome. We also compared intraspecific phylogenetic signal from the CR with that of the combined protein-coding regions and tested for phylogenetic congruence (Farris et al. 1995).

To test the hypothesis of slow substitution rates in the CR and protein-coding regions of the sperm whale mitogenome (hypothesis 2), we calculated Bayesian estimates of mitogenomic substitution rates utilizing one of the largest cetacean mitogenome phylogenies constructed to date: 44 mitogenomes representing 42 species (supplementary material S1, Supplementary Material online). The sperm whale CR and protein-coding substitution rates were then contrasted with other cetacean species. Finally, we also investigated the likelihood of population bottlenecks or selective sweeps acting on this species by calculating the time to most recent common ancestor (TMRCA) of sperm whale mitogenomes.

Materials and Methods

Sample Collection and Polymerase Chain Reaction Amplification of Mitogenomes

The 17 sperm whale skin samples were collected from stranded individuals represented in the New Zealand Cetacean Tissue Archive (CeTA) at the University of Auckland (Thompson et al. forthcoming) and from the Oregon Marine Mammal Stranding Network (OMMSN) Tissue Archive (supplementary material S2, Supplementary Material online). DNA was extracted from these tissue samples using a standard phenol/chloroform method (Sambrook et al. 1989), modified by Baker et al. (1994) for smaller samples. Sequences of the CR and mtDNA protein-coding genes from 11 dolphin species/subspecies described in Carraher (2004), but previously unpublished, were also generated from skin samples curated in CeTA using the same DNA extraction methods.

The mitogenome of each sperm whale was amplified in five overlapping fragments ranging from 3.0 to 4.3 kbp in long-range polymerase chain reaction (LR-PCR) reactions using high fidelity Phusion Polymerase (New England Biolabs, USA). Thermoprofiles consisted of an initial denaturation step of 98°C for 30 s followed by 35 cycles of 98°C for 8 s, the specific annealing temperature for each fragment for 30 s, and 72°C for 1 min 15 s; followed by a final extension of 72°C for 10 min. Reagent concentrations for each reaction

Table 2

The Diversity and Oceanic Distribution of Published mtDNA CR Haplotypes from the Sperm Whale, Defined by DQ512921 (Haplotype A) and Reflecting a Consensus Length of 400 bp

| Haplotypes | 58 | 62 | 105 | 107 | 121 | 150 | 184 | 200 | 207 | 208 | 211 | 235 | 243 | 260 | 272 | 273 | 286 | 287 | 288 | 305 | 308 | 319 | 324 | 350 | Ocean | | Grand Total | |
|------------|----|----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|----------------|----------------------|-------|-------------|-----|
| | | | | | | | | | | | | | | | | | | | | | | | | North Atlantic | South Indian/Pacific | | | |
| A | T | C | C | A | C | C | T | T | A | A | C | A | G | A | A | C | A | A | A | C | A | G | C | C | 66 | 12 | 258 | 336 |
| B | . | T | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | 55 | 12 | 187 | 254 |
| C | . | T | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | G | . | . | . | . | . | 115 | | 117 | 232 |
| D | . | . | . | . | . | . | C | . | . | T | . | . | . | . | . | . | . | . | . | . | . | . | . | . | | | 11 | 11 |
| E | C | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | | 1 | 28 | 29 |
| F | C | . | . | . | T | . | . | . | . | . | . | . | . | . | T | . | . | . | . | . | . | . | . | . | | | 6 | 6 |
| G | C | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | T | . | | | 5 | 5 |
| H | . | . | . | . | . | . | . | . | . | . | . | . | . | . | T | . | G | . | . | . | . | . | . | . | | | 8 | 8 |
| I | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | T | . | | | 10 | 10 |
| J | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | G | . | . | . | . | . | 9 | 29 | 38 | |
| K | . | T | . | . | . | . | . | . | . | . | . | . | . | . | T | . | . | G | . | . | . | . | T | . | | | 2 | 2 |
| L | . | T | . | . | . | . | . | . | . | . | A | . | . | . | . | . | . | G | . | . | . | . | . | . | | | 1 | 1 |
| M | . | T | . | . | . | . | . | . | . | . | . | . | . | . | . | G | . | . | . | . | . | . | . | . | | | 7 | 7 |
| N | . | T | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | A | . | . | 3 | 2 | 13 | 18 |
| O | . | T | . | . | . | . | C | . | . | . | . | . | G | . | . | . | . | . | . | . | . | A | T | . | | 2 | 12 | 14 |
| P | . | T | . | . | . | . | . | . | . | T | . | . | . | . | . | . | . | . | . | . | . | . | . | . | | | 2 | 2 |
| Q | . | T | . | . | . | . | C | . | . | . | . | . | G | G | . | . | . | . | . | . | . | A | T | . | | | 1 | 1 |
| R | . | T | . | . | . | . | . | . | . | . | . | . | G | . | . | . | . | . | . | . | . | A | T | . | | 1 | | 1 |
| S | . | T | T | . | . | . | C | . | . | . | . | . | G | G | . | . | . | . | . | . | . | A | T | . | 1 | 1 | 5 | 6 |
| T | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | | | 3 | 3 |
| U | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | | | 1 | 1 |
| X | . | T | . | . | . | . | . | G | . | . | . | . | . | . | . | . | . | G | . | . | . | . | . | . | 157 | | | 157 |
| Y | . | T | . | . | . | . | G | . | G | . | . | . | . | . | . | . | . | G | . | . | . | . | . | . | 19 | | | 19 |
| Z | . | T | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | G | . | G | . | . | . | . | | | 1 | 1 |
| AA | C | . | . | . | . | T | . | . | . | . | G | . | . | . | . | . | . | . | . | T | . | . | . | . | | 2 | 2 | 2 |
| BB | . | . | . | . | . | T | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | | | 2 | 2 |
| EE | . | . | G | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | | | | |
| II | . | T | . | . | . | . | . | . | . | . | . | . | . | . | T | . | . | . | . | . | . | . | . | . | | 1 | | 1 |
| | | | | | | | | | | | | | | | | | | | | | | | 417 | 41 | 709 | 1,167 | | |

NOTE.—Haplotypes of mtDNAs sequenced in this study are shaded in gray. The numbers of samples with each haplotype were taken from previous studies by ocean basin: the North Atlantic—including the Gulf of Mexico and Mediterranean (Lyholm and Gyllenstein 1998; Engelhaupt et al. 2009; Ortega-Ortiz et al. 2012), South Indian/Pacific Oceans (Lyholm and Gyllenstein 1998), and Pacific Ocean (Richard et al. 1996; Lyholm and Gyllenstein 1998; Whitehead et al. 1998; Mesnick et al. 2011; Rendell et al. 2012). Haplotype EE was first characterized in this study. For each haplotype, “.” indicates a nucleotide identical to the reference at that given site.

were as follows: 1× Phusion HF Buffer (NEB, USA); 0.5 μM of each primer (IDT); 2% DMSO (NEB, USA); 15–30 ng of template DNA, dNTP (Promega, USA); and Phusion Polymerase (NEB, USA) concentration varied by fragment as described in [supplementary material S3, Supplementary Material](#) online (primer sequences and fragment-specific annealing temperatures are also detailed in this [supplementary material, Supplementary Material](#) online) and ddH₂O to 20 μL.

As described by Carraher (2004), the CR and mtDNA protein-coding genes from 11 dolphin species/subspecies were amplified in 11 shorter (1–2.5 kbp), and partially overlapping, fragments ([supplementary material S4, Supplementary Material](#) online). Fragments were amplified using the Expand Long Template PCR System (Roche Diagnostics) as per the manufacturer's recommendations. Thermoprofiles consisted of an initial denaturation step of 93°C for 2 min followed by 10 cycles of 93°C for 30 s, 60°C for 30 s, 68°C for 45 s/kb, then 24 cycles of 93°C for 30 s, 60°C for 30 s, 68°C for 45 s/kb, increased by 20 s/cycle, with a final extension of 68°C for 10–15 min.

Illumina Sequencing, Assembly, and QC

Thirteen sperm whale mitogenomes ([supplementary material S2, Supplementary Material](#) online) were sequenced using a single-end 40 bp run on one lane of an Illumina GAIIx. The input library was prepared using multiplex identifier (MIDs) adapters to identify each sample (Cronn et al. 2008), following standard Illumina library preparation v1.12, combined in equimolar quantities and run on the Illumina at a combined concentration of 5 pM. Output data were filtered using Illumina's GA pipeline v1.5.0. Reads were then sorted by MIDs using the program bcsort (Jennings et al. 2011) before being converted to the PHRED scoring system using a custom PERL script (Davies 2009). Reads were trimmed from the 3'-end, using BWA v0.5.7 (Li and Durbin 2009), until a base with a PHRED quality above 20 was reached, to ensure only high-quality fragments of reads were included. Reads were then assembled to the sperm whale mitogenome available from NCBI (NC_002503: Arnason et al. 2000) using the program BWA. To compensate for the artifact of low coverage at the beginning and end of the arbitrarily linearized alignment (as the mitogenome is circular), a second assembly was conducted with the reference genome relinearized at 8,213 bp relative to the original reference. Samtools v0.1.7 (Li et al. 2009) was used to generate the consensus sequence (>51%) from the BWA assembly.

Roche 454 Sequencing, Assembly, and QC

Mitogenomes of seven sperm whales, including three that had been sequenced using the Illumina technology, were sequenced on a Roche GS Junior 454 ([supplementary material S2, Supplementary Material](#) online). Sample libraries were prepared using MIDs and the standard Roche GS Junior 454 rapid

library preparation, emulsion PCR, and sequencing steps (June 2010 version). Reads were separated by their MIDs using the program sftools (Roche, USA). For each sample, reads were then trimmed and assembled against the sperm whale mitogenome (NC_002503) using the default settings of GS Reference Mapper (Roche, USA). As for the Illumina sequencing described earlier, combined assemblies were generated by relinearization of the reference and the consensus base call at each site was used.

Sanger Sequencing of Dolphin Mitogenomes and Sperm Whale CRs

As described earlier, the CR and mtDNA protein-coding genes from 11 dolphin species/subspecies were amplified using LR-PCR. Excess primers and nucleotides were removed using a standard SAP dephosphorylation protocol (USB, USA). Products were then sequenced with BigDye dye terminator chemistry (Applied Biosystems, USA) with 43 internal primers detailed in [supplementary material S4, Supplementary Material](#) online, followed by dye-terminator removal using Agencourt CleanSEQ beads (Beckman Coulter, USA). Sequences were run on an ABI3730xl and assembled manually with a published dolphin mitogenome sequence, the white-beaked dolphin (*Lagenorhynchus albirostris*, NC_005278: Arnason et al. 2004). To evaluate the risk of numts affecting the analysis, sequences were reviewed to ensure overlap of long-range fragments.

Initial CR haplotype identity of sperm whales used in this study was based on PCR amplification and Sanger sequencing using the primers dlp1.5 (Baker et al. 1998) and tpe as described by Carroll et al. (2011). Substitutions in the first (5') 400 bp of sequence were used to define the CR haplotypes referred to in previous publications (see table 2 for haplotype definitions and haplotype distributions).

Estimation of NGS Discrepancy Rates

To estimate NGS discrepancy rates, three sperm whale samples were independently amplified and sequenced on both the Illumina and 454 platforms, and limited Sanger resequencing was carried out using all sperm whales included in this study. Final consensus sequences were aligned using Sequencher v4.6 (Gene Codes Corporation, USA) and discrepancy rates calculated by comparing all individuals with sequence available from two or more sequencing (NGS or Sanger) platforms.

QC and Validation of Variable Sites

Coverage plots were examined for gaps in coverage indicative of the incorporation of linear numts. Mitogenomes were then aligned using Sequencher v4.6, and variable sites in the multiple sperm whale alignment were accepted if all mitogenomes met a minimum sequencing depth of 15× (Smith et al. 2008) and the minimum measure of mapping quality

determined for each platform, as described later (mapping quality is the probability that reads are correctly placed during the assembly process, as opposed to the quality of individual reads assessed during the assembly process). At variable sites, Illumina-sequenced mitogenomes were required to have a mapping quality of at least 20 at the variable site (determined by BWA using a PHRED scale), with ≥ 15 for the 5 bp stretch of nonvariable sites adjacent to the variable site (Li et al. 2008, 2010). If an Illumina-sequenced mitogenome had a mixed-base signal present at a variable site, identical (duplicate) reads were removed from that assembly to minimize the potential impact of PCR oversampling. In addition, for Illumina-sequenced mitogenomes, the possibility of strand bias was tested using a Fisher's exact test to test for a difference in the distribution of reads for each nucleotide in the forward and reverse directions of variable sites. For 454-sequenced mitogenomes variable sites had to fulfill the 454HCDiffs (Roche, USA) criteria: The variable site was required to be covered by both forward and reverse reads (which addressed the problem of strand bias), or at least five reads with quality scores over 20 (using the GS Reference Mapper scale), or 30 if the difference between reads involved a homopolymer ≥ 5 bp in length.

If mapping quality criteria were not met, or NGS coverage was $< 15\times$ at a variable site for a particular mitogenome, Sanger sequencing was used to confirm the sequence using the primers listed in [supplementary materials S3](#) or [S5](#), [Supplementary Material](#) online. Because of the known error rate of 454 in homopolymeric regions (Kircher and Kelso 2010), Sanger and Illumina sequencing were used to confirm sequences wherever variable homopolymers ≥ 5 bp in length were present. All variable sites within primer regions were masked.

Annotation and Analysis of Intraspecific Diversity in the Sperm Whale

The sperm whale mitogenomes were annotated using the reference sequence sperm whale mitogenome (NC_002503) available from NCBI (Arnason et al. 2000). Nucleotide diversity analyses (including codon- and gene-specific analyses) were conducted in MEGA v5.0.5 (Tamura et al. 2011) on the Sequencher v4.6 alignment of the 17 sperm whale mitogenomes using a standard error computation of 10,000 bootstrap replicates, the maximum composite likelihood substitution model, homogenous rates among lineages, and different rates among sites (with a gamma parameter of 1.5, as indicated by Bayesian modeling). The relative levels of diversity in the CR and other mitogenome regions were compared using the standard error computations calculated above to provide 95% confidence intervals (CIs). The same parameters were also used in MEGA v5.0.5 to calculate the maximum pairwise divergence by gene between the sperm whale mitogenomes generated in this study to examine numt influence. To test for

congruence in phylogenetic signal of the CR and protein-coding regions, we used an incongruence length difference (ILD) test (Farris et al. 1995), implemented with 1,000 replicates and the default search parameters in TNT v1.1 (Willi Hennig Society version; Goloboff et al. 2008) using a customized script (Siddall). Phylogenetic reconstructions were also carried out to directly visualize the comparison of CR and protein-coding gene phylogenetic signal, as described later.

The number of heteroplasmies in the sperm whale mitogenomes was assessed by reviewing all positions where the minority base represented at least 30% of reads at a site. In addition, the Illumina mitogenomes were required to have a minimum coverage of two reads in each direction for each alternate base (Li et al. 2010), and the 454 mitogenomes had to satisfy the 454HCDiffs criteria (Roche, USA).

Interspecies Alignment of Cetacean Mitogenomes

Reference sequence mitogenomes of the 33 cetacean species available from GenBank were downloaded to complement the dolphin species and sperm whales sequenced in this study, leading to a total of 42 cetacean species available for phylogenetic analyses ([supplementary material S1](#), [Supplementary Material](#) online). Interspecies alignments of the GenBank sequences, the dolphin sequences generated by Carraher (2004), and the sperm whale sequences generated in this study were constructed using MUSCLE v3.6 (Edgar 2004) and checked by eye. *ND6* was excluded from phylogenetic analyses for the same reasons given in Ho and Lanfear (2010): distinctive patterns of evolution and location on the light strand opposite all other mtDNA protein-coding genes. Where the start and end of protein-coding genes overlapped, the overlapping portions were excluded. The MEGA v5.0.5 data viewer was then used to examine protein-coding regions of mitogenomes generated in this study for premature stop codons indicative of numts.

Bayesian Phylogenetic Estimation of Substitution Rates and TMRCA

Species-specific substitution rates were estimated using Bayesian phylogenetic reconstructions conducted in BEAST v1.6.2 (Drummond and Rambaut 2007) with a GTR+I+ Γ model of nucleotide substitution and estimated base frequencies. Two independent runs of 90,000,000 states, sampling every 3,000 states, were completed for each analysis, using an uncorrelated lognormal relaxed clock (Drummond et al. 2006) and a linked Yule tree prior across the CR and protein-coding region partitions. The rate analyses were calibrated using the fossil calibration dates given by Ho and Lanfear (2010). Unlike Ho and Lanfear (2010), a lognormal distribution, with the means represented in real space and a standard deviation of 1 Myr, was used as the prior shape for all calibration points

with the exception of the root (note that this contrasts with use of minimum-age priors, e.g., Ho and Lanfear 2010). The root node was calibrated with a uniform prior of 34 to 46 Ma as in Ho and Lanfear (2010). Independent runs were checked for effective sample sizes (ESS) of more than 200 and convergence of posterior values in Tracer v1.5 (Rambaut and Drummond 2007), and convergence of tree topologies in FigTree v1.3.1 (Rambaut 2009) before combining.

Phylogenetic relationships for 42 cetacean species (including the sperm whale) were reconstructed and lineage-specific substitution rates were estimated for the two separate partitions of the CR and codon-partitioned protein-coding genes using TreeAnnotator v1.6.2 visualized in FigTree. The sperm whale substitution rates were then compared with the 41 other cetacean species included in this study. Each species was represented once, with the exception of the bottlenose dolphin (*Tursiops truncatus*), where a previous mitogenome was available from GenBank, and the Hector's dolphin (*Cephalorhynchus hectori*), where the two subspecies *C. hectori hectori* and *C. hectori maui* were both included. The sperm whale was represented by a randomly selected mitogenome sequenced in this study, PmaNZ005. The median values for point estimates are reported in this article with associated 95% highest probability density (HPD) intervals. Copies of the *.xml files used in all BEAST analyses, containing the priors defined, are given in [supplementary material S6](#), [Supplementary Material](#) online.

The TMRCA of sperm whale mitogenomes was estimated using sequences of mtDNA protein-coding regions in BEAST. The sperm whale-specific substitution rate for the protein-coding regions calculated in the analysis described earlier was used, along with a skyline tree prior as this model is independent of the demographic history of the population (Drummond et al. 2005). The intraspecific protein-coding tree constructed in this analysis was also used to further investigate CR-specific constraints. The 400 bp mtDNA CR haplotype for each sperm whale mitogenome was mapped on to the intraspecific protein-coding tree to visually assess whether the mtDNA CR was reflective of underlying mitogenomic patterns of evolutionary relatedness.

Results

Summary of NGS and Sanger Sequencing Coverage

Over 93 Mbp of NGS sequence data was generated for the 20 sperm whale mitogenomes representing 17 individuals. For the 13 mitogenomes sequenced with Illumina, this provided an average sequencing depth of 359 \times . For the seven mitogenomes sequenced with 454, this provided an average sequencing depth of 174 \times . Average mapping quality exceeded 36 (BWA: PHRED quality) for sperm whale mitogenomes sequenced with Illumina, and exceeded 63.5 (GS Reference Mapper 454 quality) for samples sequenced with

454. An additional 43 kbp of Sanger sequence was used to validate variable sites in the multiple sperm whale alignment and to estimate sequencing error of the NGS platforms. As described in Carraher (2004), the CR and protein-coding genes of 11 dolphin mitogenomes were successfully amplified and sequenced to an average depth of 2.64 \times by conventional Sanger methods.

Discrepancy Rates in NGS Sequencing

Mitogenomes of the three sperm whales sequenced on both the Illumina and 454 platforms showed good agreement. After independent assembly, the alignment of the 3 pairs revealed only 41 inconsistencies in the total comparison of 98,568 bp: This included 34 inconsistencies in homopolymers of ≥ 5 bp in length, a known source of error in 454 sequencing. Of the remaining seven inconsistencies, three were resolved when duplicate Illumina reads were removed; these presumably reflected PCR re-amplification of a minority nucleotide for a low-frequency heteroplasmy, or PCR-based polymerase error. Three additional inconsistencies occurred between 454 and Illumina in the CR of PmaNZ034. At all three positions, the 454 sequence showed mixed-bases (between 50% and 59%) in comparison with the Illumina sequence, possibly indicative of a 454 library artifact. The final inconsistency in 454 sequence relative to the Illumina sequence was also in PmaNZ034, located adjacent to a 20 bp homopolymer. After excluding homopolymer inconsistencies, which were verified using Sanger sequencing (see below), the 454-Illumina discrepancy rate was seven in 98,568 bp, or 0.0071%.

Of the 43 kbp of Sanger sequence, 28,620 bp was available for comparison with the 13 Illumina-sequenced sperm whale mitogenomes, and 21,971 bp was available for the seven 454-sequenced mitogenomes. Comparisons of Illumina and Sanger sequencing revealed two inconsistencies: site 23 in PmaNZ058, and site 12,982 in PmaNZ005 (no longer inconsistent when duplicate Illumina reads were removed), giving a total Illumina-Sanger discrepancy rate of two inconsistencies in 57,240 bp compared or 0.0034%. Comparisons of 454 and Sanger found eight inconsistencies, all of which were associated with homopolymers. After excluding homopolymer inconsistencies that were verified using Sanger sequencing (discussed later), the 454-Sanger discrepancy rate was <0.0023%, that is, no errors detected in 43,942 bp compared.

QC and Validation of Variable Sites

After alignment of the 17 mitogenomes, additional QC measures were applied to variable sites, including verification of low coverage/quality sequence with Sanger sequencing, removal of duplicate Illumina reads and testing for strand bias. After application of a Bonferroni correction (by sample), no significant strand bias was detected at variable sites in Illumina

samples. Over 354 instances of high NGS QC for variable sites were covered by Sanger sequencing, showing complete agreement. In contrast, 54 instances of low NGS QC were found over 92 putative variable sites. Sanger sequencing confirmed 44 of these 54 to be identical to the nucleotide called by NGS. Two sites in Illumina sequences (site 23 in PmaNZ058 and site 12,982 in PmaNZ005: as described earlier) were corrected based on Sanger sequencing/removing Illumina duplicate reads (assuming that Sanger sequencing was correct). An additional eight sites in 454 sequences, all at homopolymer sites (described earlier), were corrected based on Sanger sequencing. There were also two additional lengths of homopolymer that could not be resolved using Sanger sequencing for the 454-generated mitogenomes included in this study: a C_{12–21} homopolymer at position 1,129 (located in the 16S rRNA) from four mitogenomes (PmaNZ013, PmaNZ076, PmaNZ082, and PmaOR001) and a C_{9–10} homopolymer at position 16,270 (located in the CR) from three mitogenomes (PmaNZ013, PmaNZ076, and PmaNZ082). Because of the uncertainty of the true length of these homopolymer regions, potential differences in homopolymer length were masked at these positions. A summary of sites that differed between sequencing platforms in the error-checking and validation of variable sites sections of this article are available in [supplementary material S7, Supplementary Material](#) online. Average NGS sequencing depth on validated variable sites is given for each sperm whale in [supplementary material S8, Supplementary Material](#) online.

We found no evidence of numts: All sperm whale mitogenomes assembled with no gaps in coverage, as did the protein-coding region fragments for the dolphin mitogenomes sequenced in this study (Carragher 2004). There were no abrupt or discordant regions of high divergence among the sperm whale mitogenomes, which would indicate inadvertent assembly of numt sequences (maximum pairwise divergence observed was 1.65% for tRNA-Lys: [supplementary material S9, Supplementary Material](#) online). No premature stop codons in protein-coding regions were observed for any mitogenome (sperm whale or dolphin) generated for this study.

Mitogenome Haplotypes

After completion of QC review, the multiple alignment of 17 sperm whale mitogenomes showed only 82 variable sites, 76 of which were transitions, and 6 of which were transversions. No sites showed both transitions and transversions, and no indels were required in the alignment after masking the small number of homopolymers that could not be successfully Sanger sequenced (table 3). Despite the low number of variable sites, all sperm whale mitogenomes in this study were resolved as having unique haplotypes, even those with identical CRs (fig. 1). Pairwise differences between mitogenomes ranged from 2 to 33. The largest number of singleton substitutions (sites differing in only one individual) found among the 17 sperm whale mitogenomes sequenced in this study was 11 ([supplementary material S10, Supplementary Material](#) online). By comparison, a large number of singleton substitutions ($n=27$) were found in the only sperm whale mitogenome available from NCBI, NC_002503 (Arnason et al. 2000). Given the potential for cloning or sequencing errors relative to mitogenomes sequenced in this study, NC_002503 was excluded from further intraspecific diversity analyses in this article.

Sperm Whale Mitogenome Diversity by Gene and Codon Position

The overall mitogenome diversity among the 17 sperm whales sequenced in this study was low ($\pi=0.096\%$). In contrast to expectations under the CR-specific constraint hypothesis, the CR (previously noted for its low diversity in comparison with other species: table 1) actually showed the highest diversity of any mitogenomic partition ($\pi=0.35\%$). CR nucleotide diversity was more than 1.5 \times greater than that of the combined 3rd codon protein-coding positions ($\pi=0.22\%$), despite the 3rd codon position and diversity significantly exceeding all other remaining mitogenomic partitions (as determined by nonoverlapping 95% CIs, fig. 2). As the 95% CIs of the CR and 3rd codon position partitions overlap, we could not conclude that the CR has significantly higher diversity than the 3rd codon position. The percentage of variable sites in the CR

Table 3

Substitutions Characterized in Terms of Their Mitogenome Region for the Alignment of 17 Sperm Whales, Including Codon Position for the Protein-Coding Regions, Type (Ts, Transition; Tv, Transversion), and Whether They Were Nonsynonymous (NS) or Synonymous (S)

| Mitogenome Region | NS | S | Ts Total | Tv Total | Ts/Tv | Total |
|------------------------------------|----|----|----------|----------|-------|-------|
| CR | | 15 | 14 | 1 | 14 | 15 |
| rRNA regions | | | 1 | 1 | 0.5 | 2 |
| tRNA regions | | | 4 | | | 4 |
| Total protein coding | 15 | 46 | 57 | 4 | 14.25 | 61 |
| Protein coding: 1st codon position | 10 | 1 | 11 | | | 11 |
| Protein coding: 2nd codon position | 5 | | 4 | 1 | | 5 |
| Protein coding: 3rd codon position | | 45 | 42 | 3 | | 45 |
| Overall mitogenome | | | 76 | 6 | 12.5 | 82 |

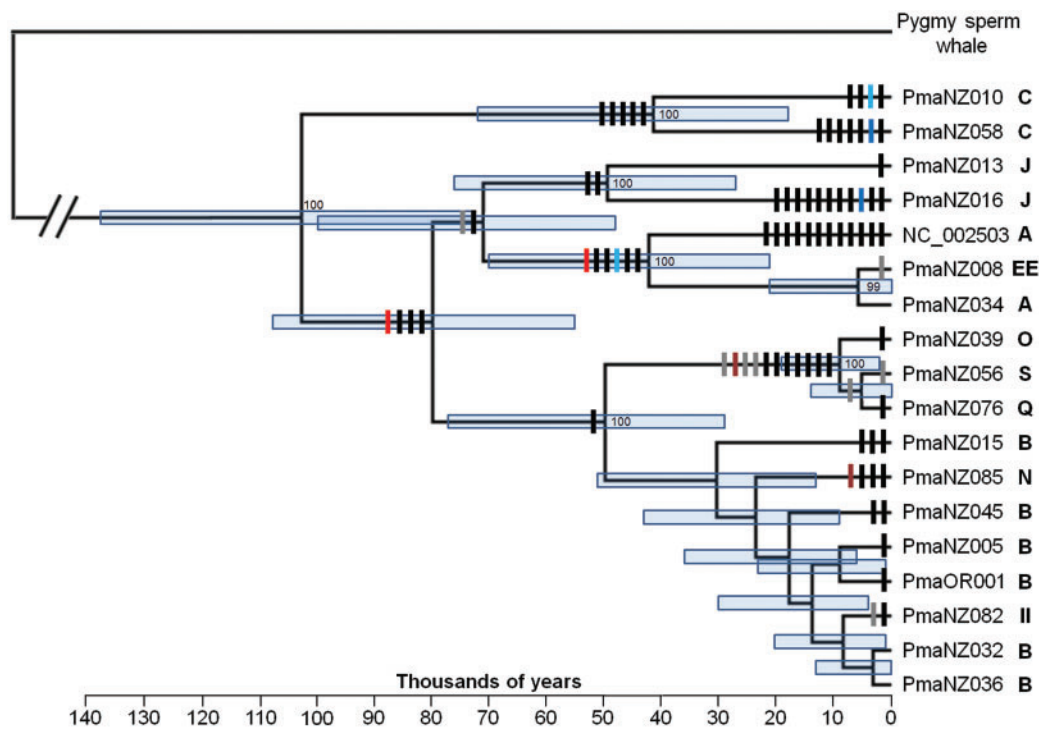


Fig. 1.—Bayesian phylogenetic reconstructions of the relationship between the 17 sperm whale mitogenomes in this study based on protein-coding genes (10,815 bp), including the NCBI GenBank sample, NC_002503 (Arnason et al. 2000). Terminal taxa are labeled by their individual sample codes and CR haplotypes (400 bp) are listed in bold after sample codes. Clades with more than 95% posterior probability are indicated on the tree, and 95% HPD intervals for node ages are shown by the horizontal blue boxes. Parsimony-inferred character changes (using *Mesquite* v2.75; Maddison and Maddison 2011) in the protein-coding regions are shown by black cross bars. Inferred character changes in the 400 bp region of the CR used to define haplotypes are shown in gray. Homoplastic characters are shown in shades of blue for the protein-coding region (two sites requiring two changes each), and in shades of red for the CR (also two sites requiring two changes each). Consistency Indices: 0.9722 for protein-coding regions and 0.8333 for CR. Retention Indices: 0.9565 for protein-coding regions and 0.8333 for CR.

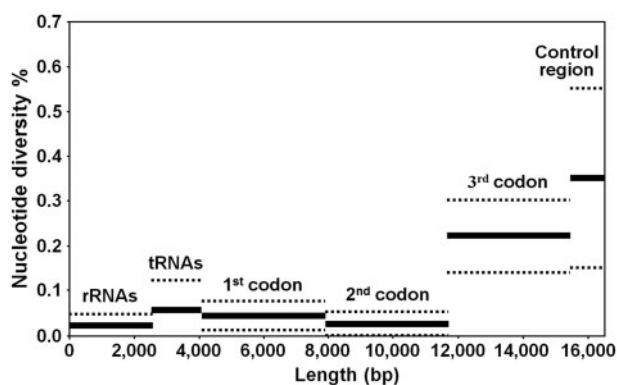


Fig. 2.—Nucleotide diversity (%) for the concatenated rRNA regions; concatenated tRNA regions; 1st, 2nd, and 3rd codon positions of the protein-coding regions, and the CR of the sperm whale mitogenome ($n = 17$). 95% confidence limits (constructed with bootstrap standard error computation) are shown by the dashed lines. For the rRNAs and tRNAs, lower 95% confidence limits overlap zero.

(1.57%, or 15 out of 954 sites), exceeded that of the 3rd codon (including *ND6*; 1.18%, or 45 out of 3,803 sites), although again not significantly based on a Fisher's exact test ($P = 0.3315$). In a gene by gene comparison, only 3 out of the 13 protein-coding genes (*COX2*, *ND3*, and *ND4L*) had 3rd codon diversity that exceeded CR diversity, and these differences were not statistically significant based on overlapping 95% CIs (supplementary material S11, Supplementary Material online). We note that the estimates of nucleotide diversity presented here are likely to be biased upwards because of the nonrandom selection of samples based on CR haplotypes. Nucleotide diversity for a random sample of sperm whales is likely to be considerably lower as even distantly related individuals could inherit identical mitogenomes.

On the basis of our criteria of at least 30% of reads representing the secondary base at a site, we detected between 1 and 6 heteroplasmies in three mitogenomes, located in the 12srRNA, 16srRNA; protein-coding genes, and the CR. Sanger sequencing was available for one of these sites which confirmed the heteroplasmy present (supplementary

material S12, Supplementary Material online). Heteroplasmies occurring in the protein-coding regions were either synonymous or required only the substitution of an amino acid with similar properties (Adachi and Hasegawa 1996; Gilis et al. 2001). The other 14 mitogenomes showed no heteroplasmies that satisfied the detection criteria. Two of these mitogenomes were sequenced with both Illumina and 454 and showed no evidence of heteroplasmies with either NGS platform (supplementary material S12, Supplementary Material online). These overall rates of heteroplasmy are consistent with studies on humans (Sosa et al. 2012) and other cetacean species (Vollmer et al. 2011).

Sperm Whale mtDNA Protein-Coding Genes and CR Phylogenetic Signal

As a further test of the CR-specific constraint hypothesis, we compared the phylogenetic signal of the sperm whale CR and protein-coding regions. From the hypothesis, we expected there would be a lack of congruence between the CR and protein-coding regions due to homoplasy in the CR. In fact, the phylogenetic signal from the CR and protein-coding regions was found to be congruent with a significant correlation between pairwise differences over the 400 bp CR haplotype and pairwise differences over the rest of the mitogenome (Pearson's product-moment correlation = 0.54, $P < 0.001$). CR haplotypes were also generally congruent with phylogenetic reconstructions based on protein-coding regions within the sperm whale, albeit with less resolution in the CR haplotypes (fig. 1). When either protein-coding or CR characters were traced on to the phylogenetic tree, only a small number of characters showed evidence of homoplasy, reflected by the high consistency (Kluge and Farris 1969) and retention indices (Farris 1989) when either partition was mapped on to the protein-coding tree (fig. 1). Finally, the ILD test detected no phylogenetic incongruence between the CR and protein-coding regions ($P = 1.000$).

Patterns of Purifying Selection in the Sperm Whale Mitogenome

The sperm whale mitogenomes showed the expected pattern of purifying selection reported in other mammalian mitogenomes (Stewart et al. 2008). Of the 61 variable sites occurring in protein-coding regions, 46 were synonymous, and 15 were nonsynonymous, with twice as many replacement substitutions found in 1st versus 2nd codon positions (10 vs. 5; table 3). Over the combined protein-coding genes, the dN/dS ratio was 0.25; significantly lower than the value expected under neutral evolution (Fisher's exact test, $P < 0.001$). The transitional bias observed in the coding regions (57 transitions, 4 transversions) was similar to that of both the CR and RNA partitions (table 3). The 15 amino acid substitutions present in this alignment (substitutions shown in supplementary material

S10, Supplementary Material online) generally had a low cost of replacement based on several substitution-cost matrices (Adachi and Hasegawa 1996; Gilis et al. 2001).

Cetacean Bayesian Phylogenetic Reconstructions and Substitution Rates

The two interspecific BEAST runs for estimating substitution rates showed ESS values over 200 for all parameter estimates, and convergence for both parameters and tree topologies. Given this agreement, runs were combined, which gave ESS values over 400 for all parameter estimates. The phylogenetic tree obtained was well supported, with 38/43 clades supported by posterior probabilities exceeding 95% (fig. 3). Estimates of the age of the Mysticeti clade were younger in our analyses (95% HPD: 11.96–17.86 Ma) than in other reconstructions (Jackson et al. 2009; Ho and Lanfear 2010; Dornburg et al. 2012), likely due to the differences in prior shape for date calibration at this node. In addition, both our phylogeny and previous studies showed uncertain relationships within Delphinidae, with the main difference being positioning of the white-beaked dolphin (*L. albirostris*) between reconstructions (Ho and Lanfear 2010; Vilstrup et al. 2011; Dornburg et al. 2012). Apart from these differences, our tree showed very similar relationships in comparison with previous publications utilizing mitogenomes (fig. 3) and the two samples of the two species with multiple representatives in the phylogeny (Hector's dolphin: *C. hectori*; bottlenose dolphin: *T. truncatus*) grouped together with high confidence. As well as phylogenetic concordance, overall cetacean substitution rates obtained from this analysis were broadly comparable with those obtained from previous studies (table 4). Over all cetaceans, substitution rate estimates were 1.12%/Myr or 1.12×10^{-2} substitutions/site/Myr (95% HPD: 0.99–1.26%/Myr) for the CR, and 1.08%/Myr (95% HPD: 0.99–1.18%/Myr) for the coding region. First codon positions evolved at $0.45 \times$ this average protein-coding rate (95% HPD: 0.42–0.48), 2nd codon positions at $0.12 \times$ (95% HPD: 0.11–0.14), and 3rd codon positions at $2.4 \times$ this rate (95% HPD: 2.39–2.46).

To test the hypothesis of a slow substitution rate in the sperm whale, estimates of substitution rates for the sperm whale obtained from age-calibrated Bayesian phylogenetic reconstructions in BEAST were compared to estimates from the other 41 cetacean species included in the analysis. Under the hypothesis of reduced substitution rates in the sperm whale, the expectation was slow rates in the sperm whale relative to other cetaceans. All species-specific substitution rates fell within 0.5%/Myr of each other in the protein-coding regions and CR, with the exception of the sperm whale (an outlier with a fast CR substitution rate) and the Franciscana (*Pontoporia blainvillei*; an outlier with fast CR and protein-coding region rates) (fig. 4). Sperm whales ranked 25th out of 42 in terms of protein-coding region

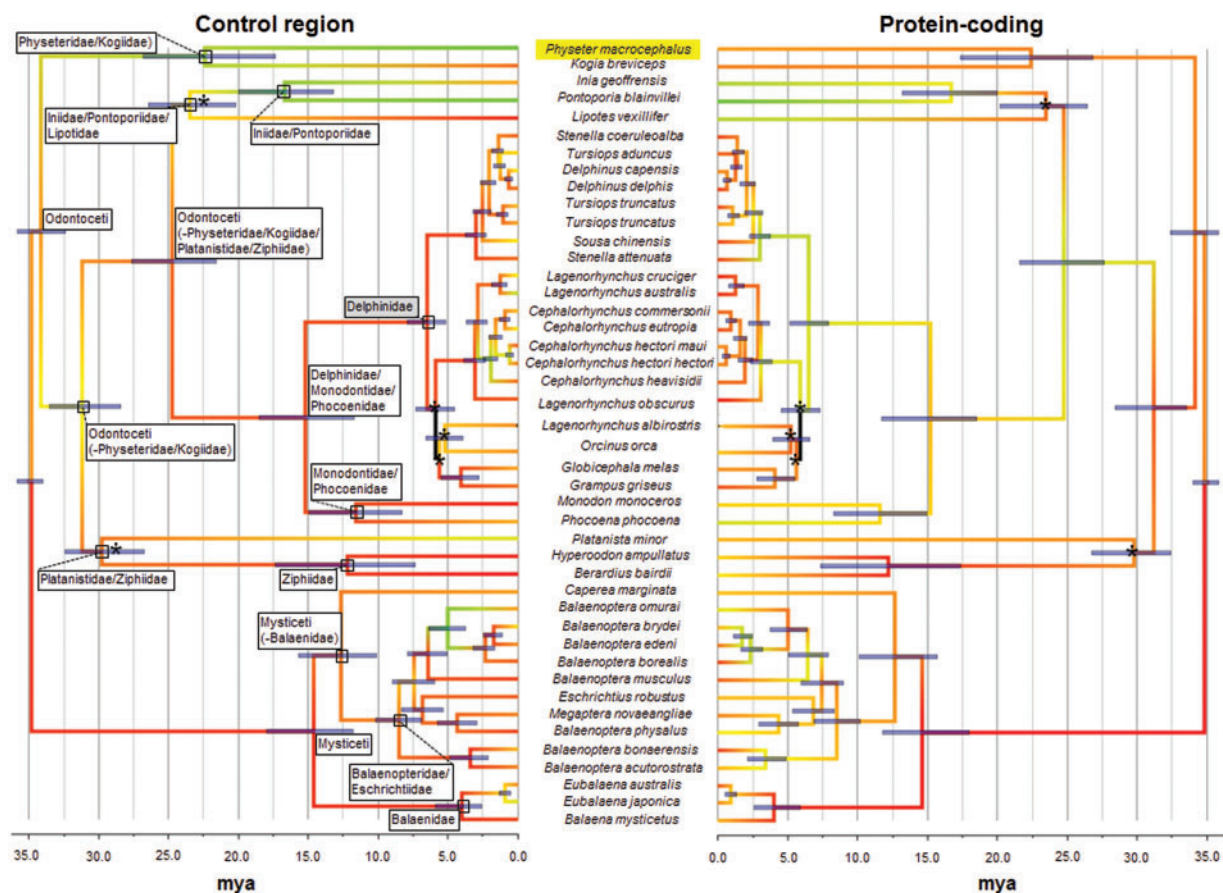


Fig. 3.—Estimates of substitution rates for the CR and protein-coding regions using a Bayesian tree reconstruction. Tree priors for these partitions were linked, so the topology is the same. The sperm whale taxon label is highlighted in yellow. Branches are colored by rate, ranging from slow in red (0.41%/Myr for the CR, 0.59%/Myr for the protein-coding regions), to fast in green (2.60%/Myr for the CR, 2.12%/Myr for the protein-coding regions). 95% higher posterior distributions for the age of nodes are indicated by the blue bars. All nodes had greater than 95% posterior probability support with the exception of the clades indicated by asterisks. With the exception of some of the relationships within Delphinidae (clade label shaded in gray), all clades labeled on the tree were supported by the previous cetacean mitogenome studies of Ho and Lanfear (2010) and Dornburg et al. (2012). The Mysticeti and Balaenopteridae/Eschrichtiidae clades were also supported by a Mysticeti-only study, Jackson et al. (2009). A Delphinidae-only study, Vilstrup et al. (2011) also partially supported relationships in Delphinidae. Discrepancies within Delphinidae are mentioned in greater detail in the text. Clade nomenclature from Perrin (2009).

rates (ranking from fast to slow), implying an average (but not slow) substitution rate for this region relative to other cetaceans. Surprisingly, sperm whales had the fastest rate of substitutions in the CR for any of the 42 cetaceans included in the analysis (2.60%/Myr; fig. 4).

Bayesian Estimates of Sperm Whale Mitogenome TMRCA

In addition to investigating the hypotheses of CR-specific constraints and rate variation in the sperm whale mitogenome, the TMRCA of the protein-coding mtDNA genes of the sperm whale was estimated to investigate the potential for a population bottleneck or selective sweep. Assuming clock-like intraspecific substitution rates within the protein-coding regions, and using the sperm-whale protein-coding

substitution rate estimated from the interspecies phylogeny, the TMRCA for the combined sperm whale mtDNA protein-coding regions was estimated as 103,000 years ago (95% HPD: 72,800–137,400 years ago).

Discussion

This study represents the first population level survey of mitogenome diversity in the sperm whale. We found that the low diversity previously characterized for the sperm whale CR (Lyrholm and Gyllensten 1998; Whitehead 1998) is a feature of the entire sperm whale mitogenome. Furthermore, intra-specific phylogenetic signals from the protein-coding region and CR were congruent. Overall, these data indicate that evolution of the CR has not been significantly constrained

Table 4

Estimated Substitution Rates for Mitogenomes of Cetaceans

| Mitogenome Partition | Taxa | Estimate from This Study | Previous Estimates |
|------------------------------------|-------------|--------------------------|--|
| CR | Cetaceans | 0.99–1.26 | 0.28–1.04 ^a |
| | Odontocetes | 1.09 | 0.23–0.76 ^b |
| | Mysticetes | 0.99 | 0.70–1.00, ^c 1.20–3.70, ^d 1.67–9.32 ^e |
| Protein-coding region | Cetaceans | 0.99–1.18 | 1.80–2.20 ^f |
| | Odontocetes | 1.00 | 0.15–0.48, ^{b,g} 0.85–1.13, ^{h,i} 1.23–1.54 ^j |
| | Mysticetes | 0.92 | 0.45–0.67 ^j |
| Protein coding: 1st codon position | Cetaceans | 0.49 | 0.35 ^f |
| Protein coding: 2nd codon position | Cetaceans | 0.14 | 0.11 ^f |
| Protein coding: 3rd codon position | Cetaceans | 2.62 | 1.00, ^{k,l} 2.40 ^f |
| | Mysticetes | | 0.08–2.46, ^e 0.40, ^{l,m} 0.70–0.80 ^{k,l} |

NOTE.—Rates are in % divergence/Myr (%/Myr) and ranges show 95% HPD intervals.

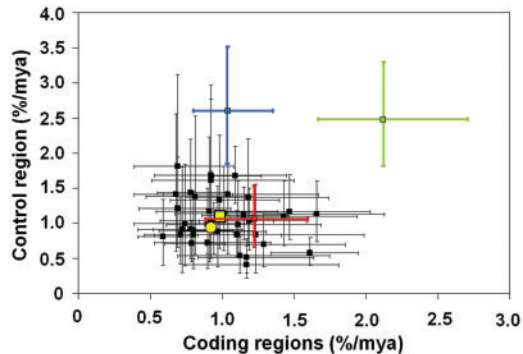
^aHoelzel et al. (1991).^bDuchêne et al. (2011).^cBaker et al. (1993).^dRooney et al. (2001).^eJackson et al. (2009).^fHo and Lanfear (2010).^gEntire mitogenome used instead of protein-coding region.^hVilstrup et al. (2011).ⁱProtein-coding genes plus tRNA genes used.^jDornburg et al. (2012).^kNabholz et al. (2008).^lEstimates available for *CytB* only.^mAlter et al. (2007).

FIG. 4.—CR and protein-coding region lineage specific substitution rates for the 42 cetacean species included in this study. Error bars associated with points indicate the 95% HPD associated with each species. The sperm whale is shown in blue, the Franciscana (*Pontoporia blainvillei*) in green and the killer whale (*Orcinus orca*) in red. The median odontocete rates are shown in the yellow square, the median mysticete rates in the yellow circle.

compared with the rest of the sperm whale mitogenome. Although substitution rates of sperm whale mtDNA protein-coding regions were similar to rates in other cetacean species, the sperm whale CR substitution rate was the fastest of all 42 cetacean species included in this analysis. This suggests that slow sperm whale-specific substitution rates are not responsible for low CR or overall mitogenomic diversity.

Sources of Sequencing Error

Studies of population-level diversity can be sensitive to NGS sequencing errors or lack of QC enforced on variable sites (Shen et al. 2010), and there are few standardly reported QC measures in NGS sequences (Goto et al. 2011). Given the previously reported low levels of CR diversity in the sperm whale (Lyrholm and Gyllensten 1998), it was important to ensure that sequencing errors did not impact our mitogenomic diversity estimates. We carried out a three-way comparison of NGS and Sanger sequencing technologies and found relatively low discrepancy rates. The highest discrepancy rate (454 to Illumina: 0.0071%) was an order of magnitude below the overall estimate of diversity obtained for the sperm whale mitogenome ($\pi = 0.096\%$). We compensated for the small number of discrepancies, including apparent PCR duplicates; polymerase, library and assembly artifacts; and differences in homopolymer lengths (a known artifact of 454 technology; Kircher and Kelso 2010) by Sanger sequencing variable homopolymers and low QC NGS sites. The complete agreement between high QC NGS and Sanger sequence validates the QC threshold used in this study.

Another source of potential error in mtDNA sequencing is the inadvertent inclusion of numts (Zhang and Hewitt 1996; Sorenson and Quinn 1998; Dunshea et al. 2008). As a precaution against this, we used LR-PCR with fragment sizes that exceeded the typical length of mammalian numts, as well as high annealing temperatures to reduce nonspecific

amplification (Richly and Leister 2004; Li et al. 2012). Assemblies showed no evidence of gaps in coverage that could be indicative of linear numts incorporating into the alignment (Sorenson and Quinn 1998; Thalmann et al. 2004), nor were there anomalously divergent sperm whale mitogenome regions or premature stop codons that would indicate the incorporation of ancient pseudogenes (Sorenson and Quinn 1998; Nabholz et al. 2010). Along with the low discrepancy rate, robust QC at variable sites, and lack of detectable strand bias, the absence of detectable numts provided confidence in the estimates of diversity obtained from our sperm whale mitogenome alignments.

Are CR-Specific Constraints Operating on the Sperm Whale Mitogenome?

Like sperm whales, killer whales and fishers also have very low levels of mitogenomic diversity relative to other mammal species for which estimates are available (supplementary material S13, Supplementary Material online). In killer whales and fishers, the CR presents misleading phylogeographic interpretations, as deep divergences among non-CR regions of the mitogenome are not accurately represented by the CR (Morin et al. 2010; Knaus et al. 2011). Instead, the CR seems to have become saturated with substitutions due to constraints, thus making it less accurate at predicting intraspecific relationships than the full mitogenome (Knaus et al. 2011). In sperm whales, this is not the case. Not only does the CR accurately reflect the intraspecific phylogeny shown by the protein-coding genes, but the CR (previously noted for its low diversity compared with other cetacean species [Lyrholm and Gyllenstein 1998; Whitehead 1998]) actually showed the greatest nucleotide diversity among the partitions compared. Only a modest increase in diversity was recovered by sequencing the full mitogenome in comparison with the CR (10 haplotypes distinguished from 400 bp of CR; 14 haplotypes distinguished from 954 bp of CR; 17 haplotypes from the full 16,428 bp mitogenome). This further indicates that the CR does not have a large number of sites placed under mutational constraints, leading to saturation, in comparison with the protein-coding regions in the sperm whale.

Are Substitution Rates Slow in the Sperm Whale Mitogenome?

Although tests of substitution rates in the sperm whale mitogenome have been conducted previously (Lyrholm et al. 1996; Whitehead 1998), these tests were limited to only a fraction of the mitogenome. They were also conducted assuming Kimura's (1980) nucleotide substitution model and no variation in substitution rates among sites. This might not accurately model evolution in cetacean mitogenomes. The use of the domestic cow (*Bos taurus*), a relatively distant outgroup, could also have limited the power to detect significant rate variation (Bromham et al. 2000). We investigated variation in

substitution rates across the CR and all protein-coding genes (with the exception of *ND6*); and used BEAST, as this allowed our analyses to accommodate differences in base composition, substitution biases, and rate heterogeneity (Drummond et al. 2006). We also calibrated our rate analyses with cetacean fossil calibration points (in comparison to the domestic cow outgroup used in previous analyses). We found no evidence of a pervasive pattern of slow substitution rates in the sperm whale mitogenome relative to other cetaceans. This supports previous studies that found no evidence for sperm whale-specific slowing of substitution rates in the limited number of genes examined (Lyrholm et al. 1996; Whitehead 1998). In fact, our study indicated that the sperm whale CR had a significantly elevated rate of evolution, a finding that is also consistent with previous studies utilizing different relative rate tests (Lyrholm et al. 1996). These results suggest that a slow substitution rate is not responsible for the low mitogenome diversity seen in the sperm whale.

Estimates of cetacean substitution rates in this study were also broadly comparable with previous studies of cetaceans (table 4), including supporting previous findings of slow rates in cetaceans compared with other mammals such as primates and rodents (Martin and Palumbi 1993; Nabholz et al. 2008; Jackson et al. 2009). The correlation between CR and coding region rates among cetaceans was weak and not statistically significant (Pearson's product-moment correlation = 0.155, $P=0.326$) probably reflecting substitutional saturation of the CR between species, or different selective constraints on specific genes, across cetaceans.

Can a Genetic Bottleneck or Selective Sweep Explain Low Sperm Whale mtDNA Diversity?

The Bayesian analyses employed in this study estimated a TMRCA for the sperm whale mitogenome of 103,000 years ago, consistent with previous research based on the CR (Lyrholm et al. 1996). It is likely that the actual age of the TMRCA is younger than this estimate due to the use of external calibration points to estimate the substitution rate within the sperm whale (Ho et al. 2008). This would most likely put the age of the TMRCA in the Pleistocene, a period characterized by its glaciation (Steeman et al. 2009). We consider this TMRCA to be younger than expected considering the age of the sperm whale lineage (divergence between the sperm whale and pygmy sperm whale 95% HPD: 17.6–27.9%/Ma) and the sperm whale's abundance and worldwide distribution (Best 1979; Whitehead 2002, 2003). Further evidence for a relatively recent TMRCA is the low number of transversions (Ts/Tv ratio = 12.5) and complete lack of Sanger/Illumina-confirmed indels in the sperm whale mitogenomic sequence. The Ts/Tv ratio is far higher than seen in mtDNA (mostly based on CR and *CytB*) from other cetacean species such as the common minke whale (*Balaenoptera acutorostrata*; Ts/Tv = 5; Pastene et al. 2007), the killer whale

($Ts/Tv = 3.75$; Hoelzel et al. 2002), and even the Hector's dolphin (*C. hectori*) ($Ts/Tv = 3.3$; Pichler et al. 1998), a species that is known to have undergone a population reduction.

Our estimate of the TMRCA is consistent with a population bottleneck or selective sweep having acted on the sperm whale in the past, but is not consistent with the impact of commercial whaling that exploited sperm whales from the 18th century onwards (Best 1979; Whitehead 2003). A whaling-induced sperm whale population bottleneck has also previously been discounted due to estimates of current sperm whale abundance (Whitehead 1998). However, sperm whales occupy a very specialized niche (Best 1979; Whitehead 2003), and past climatic influences, particularly glaciation during the Pleistocene, may have altered prey distribution (Steeman et al. 2009), or otherwise impacted on the abundance of the sperm whale.

Selective sweeps could also have reduced sperm whale mitogenomic diversity through the previously proposed hypothesis of "hitchhiking" with maternal cultural innovation (Whitehead 1998). In addition, sperm whales make routine foraging dives in excess of 1,000 m (Watkins et al. 1993). Unique physiological adaptations in the sperm whale appear to have resulted in response to the selective pressure of deep diving including increases in muscle myoglobin levels; greater utilization of blood and muscle versus lung storage of oxygen; and collapsing of lungs at shallow depths (Kooyman and Ponganis 1998). Additional adaptations might include positive selection on mtDNA-encoded proteins due to the mitochondria's role in oxidative phosphorylation (Ballard and Dean 2001). Mutations with novel adaptive properties could lead to a selective sweep of mitogenomes. However, in order for a selective sweep to globally reduce mtDNA diversity, there must be inter-ocean dispersal of female sperm whales. Despite female social units being relatively philopatric, they are known to carry out large-scale dispersal events in apparent response to poor ocean conditions (e.g., Galapagos to Ecuador/Gulf of California) over short time scales of less than a decade (Whitehead et al. 1997). These documented movements and the shared mtDNA haplotypes between ocean basins confirms at least occasional inter-ocean dispersal in female sperm whales, providing the potential for a selective sweep to reduce mtDNA diversity on a global scale.

Although a population bottleneck or selective sweep are both plausible causes of the low mitogenomic diversity, distinguishing between these hypotheses will require comparative population-level nuclear DNA data sets (which are not yet available for the sperm whale or its sister taxon, the pygmy sperm whale). A population bottleneck would be expected to reduce mtDNA and nuclear diversity in the sperm whale relative to the pygmy sperm whale, whereas a selective sweep should reduce only mtDNA diversity in the sperm whale (Rokas et al. 2001). Within the sperm whale, a population bottleneck would be expected to reduce the time to coalescence for both mtDNA and nuclear DNA, whereas a selective

sweep should only reduce the time to mtDNA coalescence (Charlesworth et al. 2003), although these inferences would depend heavily on an accurate assessment of mutation rates (Karl et al. 2012).

Conclusions

The low diversity of sperm whale mitogenomes found in this study, and consistency of the CR with protein-coding intraspecific reconstructions, refutes CR-specific constraints as an explanation for low CR diversity. Furthermore, the pervasive low mitogenome diversity requires a hypothesis that can explain the low mitogenome-wide variation. In this study, we eliminated slow substitution rates in the sperm whale mitogenome as an explanation for low diversity. The lack of "rare" substitutions such as indels and transversions, and the recent TMRCA for sperm whale mitogenomes suggests that the previously proposed hypotheses of selective sweeps or population bottlenecks are the most likely candidates for explaining the low mitogenome diversity seen in the sperm whale.

Supplementary Material

Supplementary materials S1–S13 are available at *Genome Biology and Evolution* online (<http://www.gbe.oxfordjournals.org/>).

Acknowledgments

The authors thank R. Constantine and K. Thompson for curation of the CeTA database, and New Zealand Department of Conservation staff for collecting NZ samples used in this study; J. Rice for curation of the OMMSN database that supplied the Oregon sample; M. Vant for primer development; D. Jacobsen for assistance with 454 prep; K. Jones for assistance with 454 analysis; the National Oceanic and Atmospheric Administration Southwest Fisheries Science Center and A. Liston for consultation on cetacean mitogenomics; M. Dasenko and C. Sullivan (Oregon State University Center for Genomics Research and Biocomputing) for Illumina sequencing and computational support; S. Wu and W.L.S. Li for consultation on Bayesian analysis; Oregon State University Cetacean Conservation and Genomics Laboratory for additional lab support; and A. Sremba, A. Hickey, E. Carroll, E. Harvey, P. Alexander, K. Makova, and two anonymous referees for providing valuable comments on this manuscript. This work was supported by a Mamie Markham Award and a Lylian Brucefield Reynolds Award from the Hatfield Marine Science Center; and a 2008–2011 International Fulbright Science & Technology award to A.A; and co-funded by the ASSURE program of the Department of Defense in partnership with the National Science Foundation REU Site program to K.H and C.S.B. [grant number NSF OCE-1004947]. Publication of this paper was supported, in part, by the Thomas G. Scott Publication Fund.

Literature Cited

- Adachi J, Hasegawa M. 1996. Model of amino acid substitution in proteins encoded by mitochondrial DNA. *J Mol Evol.* 42:459–468.
- Adams L, Rosel P. 2006. Population differentiation of the Atlantic spotted dolphin (*Stenella frontalis*) in the western North Atlantic, including the Gulf of Mexico. *Mar Biol.* 148:671–681.
- Alter SE, Rynes E, Palumbi SR. 2007. DNA evidence for historic population size and past ecosystem impacts of gray whales. *Proc Natl Acad Sci U S A.* 104:15162–15167.
- Arnason U, Gullberg A, Gretarsdottir S, Ursing B, Janke A. 2000. The mitochondrial genome of the sperm whale and a new molecular reference for estimating Eutherian divergence dates. *J Mol Evol.* 50:569–578.
- Arnason U, Gullberg A, Janke A. 2004. Mitogenomic analyses provide new insights into cetacean origin and evolution. *Gene* 333:27–34.
- Baker CS, et al. 1993. Abundant mitochondrial DNA variation and world-wide population structure in humpback whales. *Proc Natl Acad Sci U S A.* 90:8239–8243.
- Baker CS, et al. 1994. Hierarchical structure of mitochondrial DNA gene flow among humpback whales *Megaptera novaeangliae*, world-wide. *Mol Ecol.* 3:313–327.
- Baker CS, et al. 1998. Population structure of nuclear and mitochondrial DNA variation among humpback whales in the North Pacific. *Mol Ecol.* 7:695–707.
- Ballard JWO, Dean MD. 2001. The mitochondrial genome: mutation, selection and recombination. *Curr Opin Genet Dev.* 11:667–672.
- Banguera-Hinestroza E, Bjørge A, Reid R, Jepson P, Hoelzel A. 2010. The influence of glacial epochs and habitat dependence on the diversity and phylogeography of a coastal dolphin species: *Lagenorhynchus albirostris*. *Conserv Genet.* 11:1823–1836.
- Bérubé M, 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. *Mol Ecol.* 7:585–599.
- Best PB. 1979. Chapter 7: Social organization in sperm whales, *Physeter macrocephalus*. In: Winn HE, Olla BL, editors. *Behavior of marine animals, Vol. 3. Cetaceans*. New York: Plenum Press.
- Bromham L, Penny D, Rambaut A, Hendy MD. 2000. The power of relative rates tests depends on the data. *J Mol Evol.* 50:296–301.
- Carraher CJF. 2004. Comparative mitogenomics of the Southern Hemisphere dolphin genus *Cephalorhynchus* [MSc thesis]. [Auckland (New Zealand)]: University of Auckland.
- Carroll E, et al. 2011. Population structure and individual movement of southern right whales around New Zealand and Australia. *Mar Ecol Prog Ser.* 432:257–268.
- Cassens I, et al. 2005. Evidence for male dispersal along the coasts but no migration in pelagic waters in dusky dolphins (*Lagenorhynchus obscurus*). *Mol Ecol.* 14:107–121.
- Charlesworth B, Charlesworth D, Barton NH. 2003. The effects of genetic and geographic structure on neutral variation. *Annu Rev Ecol Syst.* 34:99–125.
- Chivers SJ, Dizon AE, Gearin PJ, Robertson KM. 2002. Small-scale population structure of eastern North Pacific harbour porpoises (*Phocoena phocoena*) indicated by molecular genetic analyses. *J Cetacean Res Manage.* 4:111–122.
- Chivers SJ, Leduc RG, Robertson KM, Barros NB, Dizon AE. 2005. Genetic variation of *Kogia* spp. with preliminary evidence for two species of *Kogia sima*. *Mar Mammal Sci.* 21:619–634.
- Chivers SJ, et al. 2007. Genetic variation and evidence for population structure in eastern North Pacific false killer whales (*Pseudorca crassidens*). *Can J Zool.* 85:783–794.
- Cronn R, et al. 2008. Multiplex sequencing of plant chloroplast genomes using Solexa sequencing-by-synthesis technology. *Nucleic Acids Res.* 36:e122.
- Davies RM. 2009. Custom PERL script [Internet]. Available from: <http://seqanswers.com/forums/showthread.php?s=361b236c525031c8219a97238f053b06&t=5192> (last accessed January 8, 2012).
- Dornburg A, Brandley MC, McGowen MR, Near TJ. 2012. Relaxed clocks and inferences of heterogeneous patterns of nucleotide substitution and divergence time estimates across whales and dolphins (Mammalia: Cetacea). *Mol Biol Evol.* 29:721–736.
- Drummond A, Rambaut A. 2007. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol.* 7:214.
- Drummond AJ, Ho SYW, Phillips MJ, Rambaut A. 2006. Relaxed phylogenetics and dating with confidence. *PLoS Biol.* 4:e88.
- Drummond AJ, Rambaut A, Shapiro B, Pybus OG. 2005. Bayesian coalescent inference of past population dynamics from molecular sequences. *Mol Biol Evol.* 22:1185–1192.
- Duchêne S, Archer FI, Vilstrup J, Caballero S, Morin PA. 2011. Mitogenome phylogenetics: the impact of using single regions and partitioning schemes on topology, substitution rate and divergence time estimation. *PLoS One* 6:e27138.
- Dunshea G, et al. 2008. Pseudogenes and DNA-based diet analyses: a cautionary tale from a relatively well sampled predator-prey system. *Bull Entomol Res.* 98:239–248.
- Edgar R. 2004. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* 5:113.
- Engelhaupt D, et al. 2009. Female philopatry in coastal basins and male dispersion across the North Atlantic in a highly mobile marine species, the sperm whale (*Physeter macrocephalus*). *Mol Ecol.* 18:4193–4205.
- Escorza-Treviño S, Archer F, Rosales M, Lang A, Dizon A. 2005. Genetic differentiation and intraspecific structure of Eastern Tropical Pacific spotted dolphins, *Stenella attenuata*, revealed by DNA analyses. *Conserv Genet.* 6:587–600.
- Farris JS. 1989. The retention index and homoplasy excess. *Syst Biol.* 38:406–407.
- Farris JS, Källersjö M, Kluge AG, Bult C. 1995. Testing significance of incongruence. *Cladistics* 10:315–319.
- Gilis D, Massar S, Cerf NJ, Rooman M. 2001. Optimality of the genetic code with respect to protein stability and amino-acid frequencies. *Genome Biol.* 2:research0049.
- Goloboff PA, Farris JS, Nixon KC. 2008. TNT, a free program for phylogenetic analysis. *Cladistics* 24:774–786.
- Goto H, et al. 2011. Dynamics of mitochondrial heteroplasmy in three families investigated via a repeatable re-sequencing study. *Genome Biol.* 12:R59.
- Hammer R, Pichler F, Heimeier D, Constantine R, Baker C. 2012. Genetic differentiation and limited gene flow among fragmented populations of New Zealand endemic Hector's and Maui's dolphins. *Conserv Genet.* 13:987–1002.
- Hayano A, Amano M, Miyazaki N. 2003. Phylogeography and population structure of the Dall's porpoise, *Phocoenoides dalli*, in Japanese waters revealed by mitochondrial DNA. *Genes Genet Syst.* 78:81–91.
- Ho SYW, Lanfear R. 2010. Improved characterisation of among-lineage rate variation in cetacean mitogenomes using codon-partitioned relaxed clocks. *Mitochondr DNA.* 21:138–146.
- Ho SYW, Saarma U, Barnett R, Haile J, Shapiro B. 2008. The effect of inappropriate calibration: three case studies in molecular ecology. *PLoS One* 3:e1615.
- Hoelzel AR, Hancock JM, Dover GA. 1991. Evolution of the cetacean mitochondrial D-loop region. *Mol Biol Evol.* 8:475–493.
- Hoelzel AR, et al. 2002. Low worldwide genetic diversity in the killer whale (*Orcinus orca*): implications for demographic history. *Proc Roy Soc B Biol Sci.* 269:1467–1473.
- Jackson JA, et al. 2009. Big and slow: phylogenetic estimates of molecular evolution in baleen whales (Suborder Mysticeti). *Mol Biol Evol.* 26:2427–2440.

- Jennings TN, Knaus BJ, Mullins TD, Haig SM, Cronn RC. 2011. Multiplexed microsatellite recovery using massively parallel sequencing. *Mol Ecol Resour.* 11:1060–1067.
- Kanda N, Goto M, Kato H, McPhee M, Pastene L. 2007. Population genetic structure of Bryde's whales (*Balaenoptera brydei*) at the inter-oceanic and trans-equatorial levels. *Conserv Genet.* 8:853–864.
- Karl SA, Toonen RJ, Grant WS, Bowen BW. 2012. Common misconceptions in molecular ecology: echoes of the modern synthesis. *Mol Ecol.* 21:4171–4189.
- Kimura M. 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J Mol Evol.* 16:111–120.
- Kircher M, Kelso J. 2010. High-throughput DNA sequencing—concepts and limitations. *BioEssays* 32:524–536.
- Kluge AG, Farris JS. 1969. Quantitative phyletics and the evolution of Anurans. *Syst Biol.* 18:1–32.
- Knaus BJ, Cronn R, Liston A, Pilgrim K, Schwartz MK. 2011. Mitochondrial genome sequences illuminate maternal lineages of conservation concern in a rare carnivore. *BMC Ecol.* 11:10.
- Kooyman GL, Ponganis PJ. 1998. The physiological basis of diving to depth: birds and mammals. *Annu Rev Physiol.* 60:19–32.
- Krützen M, Sherwin WB, Berggren P, Gales N. 2004. Population structure in an inshore cetacean revealed by microsatellite and mtDNA analysis: bottlenose dolphins (*Tursiops* sp.) in Shark Bay, Western Australia. *Mar Mammal Sci.* 20:28–47.
- Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25:1754–1760.
- Li H, Ruan J, Durbin R. 2008. Mapping short DNA sequencing reads and calling variants using mapping quality scores. *Genome Res.* 18:1851–1858.
- Li H, et al. 2009. The sequence alignment/map format and SAMtools. *Bioinformatics* 25:2078–2079.
- Li M, Schroeder R, Ko A, Stoneking M. 2012. Fidelity of capture-enrichment for mtDNA genome sequencing: influence of NUMTs. *Nucleic Acids Res.* 40:e137.
- Li M, et al. 2010. Detecting heteroplasmy from high-throughput sequencing of complete human mitochondrial DNA genomes. *Am J Hum Genet.* 87:237–249.
- Li X, et al. 2011. Analysis of global and local population stratification of finless porpoises *Neophocaena phocaenoides* in Chinese waters. *Mar Biol.* 158:1791–1804.
- Lyrholm T, Gyllensten U. 1998. Global matrilineal population structure in sperm whales as indicated by mitochondrial DNA sequences. *Proc Roy Soc B Biol Sci.* 265:1679–1684.
- Lyrholm T, Leimar O, Gyllensten U. 1996. Low diversity and biased substitution patterns in the mitochondrial DNA control region of sperm whales: implications for estimates of time since common ancestry. *Mol Biol Evol.* 13:1318–1326.
- Maddison WP, Maddison DR. 2011. Mesquite: a modular system for evolutionary analysis. Version 2.75 [Internet]. Available from: <http://mesquiteproject.org> (last accessed January 8, 2012).
- Martien KK, et al. 2012. Population structure of island-associated dolphins: evidence from mitochondrial and microsatellite markers for common bottlenose dolphins (*Tursiops truncatus*) around the main Hawaiian Islands. *Mar Mammal Sci.* 28:E208–E232.
- Martin AP, Palumbi SR. 1993. Body size, metabolic rate, generation time, and the molecular clock. *Proc Natl Acad Sci U S A.* 90:4087–4091.
- Mesnick SL, et al. 2011. Sperm whale population structure in the eastern and central North Pacific inferred by the use of single-nucleotide polymorphisms, microsatellites and mitochondrial DNA. *Mol Ecol Resour.* 11:278–298.
- Mirimin L, et al. 2009. Population structure of short-beaked common dolphins (*Delphinus delphis*) in the North Atlantic Ocean as revealed by mitochondrial and nuclear genetic markers. *Mar Biol.* 156:821–834.
- Morin PA, et al. 2010. Complete mitochondrial genome phylogeographic analysis of killer whales (*Orcinus orca*) indicates multiple species. *Genome Res.* 20:908–916.
- Mulligan CJ, Kitchen A, Miyamoto MM. 2006. Comment on “Population size does not influence mitochondrial genetic diversity in animals.” *Science* 314:1390.
- Nabholz B, Glémin S, Galtier N. 2008. Strong variations of mitochondrial mutation rate across mammals—the longevity hypothesis. *Mol Biol Evol.* 25:120–130.
- Nabholz B, Jarvis ED, Ellegren H. 2010. Obtaining mtDNA genomes from next-generation transcriptome sequencing: a case study on the basal Passerida (Aves: Passeriformes) phylogeny. *Mol Phylogenet Evol.* 57:466–470.
- Natoli A, Birkun A, Aguilar A, Lopez A, Hoelzel AR. 2005. Habitat structure and the dispersal of male and female bottlenose dolphins (*Tursiops truncatus*). *Proc Roy Soc B Biol Sci.* 272:1217–1226.
- O’Corry-Crowe GM, Suydam RS, Rosenberg A, Frost KJ, Dizon AE. 1997. Phylogeography, population structure and dispersal patterns of the beluga whale *Delphinapterus leucas* in the western Nearctic revealed by mitochondrial DNA. *Mol Ecol.* 6:955–970.
- Oremus M, et al. 2009. Worldwide mitochondrial DNA diversity and phylogeography of pilot whales (*Globicephala* spp.). *Biol J Linn Soc.* 98:729–744.
- Ortega-Ortiz JG, Engelhaupt D, Winsor M, Mate BR, Hoelzel AR. 2012. Kinship of long-term associates in the highly social sperm whale. *Mol Ecol.* 21:732–744.
- Palsbøll PJ, et al. 1995. Distribution of mtDNA haplotypes in North Atlantic humpback whales: the influence of behaviour on population structure. *Mar Ecol Prog Ser.* 116:1–10.
- Pastene LA, et al. 2007. Radiation and speciation of pelagic organisms during periods of global warming: the case of the common minke whale, *Balaenoptera acutorostrata*. *Mol Ecol.* 16:1481–1495.
- Perrin WF. 2009. World cetacea database [Internet]. Available from: <http://www.marinespecies.org/cetacea> (last accessed June 14, 2012).
- Pichler FB, Dawson SM, Slooten E, Baker CS. 1998. Geographic isolation of Hector's dolphin populations described by mitochondrial DNA sequences. *Conserv Biol.* 12:676–682.
- Pimper L, Baker C, Goodall R, Olavarria C, Remis M. 2010. Mitochondrial DNA variation and population structure of Commerson's dolphins (*Cephalorhynchus commersonii*) in their southernmost distribution. *Conserv Genet.* 11:2157–2168.
- Quérouil S, et al. 2010. Molecular insight into the population structure of common and spotted dolphins inhabiting the pelagic waters of the Northeast Atlantic. *Mar Biol.* 157:2567–2580.
- Rambaut A. 2009. FigTree [Internet]. Available from: <http://tree.bio.ed.ac.uk/software/figtree/> (last accessed January 8, 2012).
- Rambaut A, Drummond A. 2007. Tracer v1.4 [Internet]. Available from: <http://beast.bio.ed.ac.uk/Tracer> (last accessed January 8, 2012).
- Rendell L, Mesnick S, Dalebout M, Burtenshaw J, Whitehead H. 2012. Can genetic differences explain vocal dialect variation in sperm whales, *Physeter macrocephalus*? *Behav Genet.* 42:332–343.
- Richard KR, Dillon MC, Whitehead H, Wright JM. 1996. Patterns of kinship in groups of free-living sperm whales (*Physeter macrocephalus*) revealed by multiple molecular genetic analyses. *Proc Natl Acad Sci U S A.* 93:8792–8795.
- Richly E, Leister D. 2004. NUMTs in sequenced eukaryotic genomes. *Mol Biol Evol.* 21:1081–1084.
- Rokas A, Atkinson RJ, Brown GS, West SA, Stone GN. 2001. Understanding patterns of genetic diversity in the oak gallwasp *Biorhiza pallida*: demographic history or a *Wolbachia* selective sweep? *Heredity* 87:294–304.
- Rooney AP, Honeycutt RL, Derr JN. 2001. Historical population size change of bowhead whales inferred from DNA sequence polymorphism data. *Evolution* 55:1678–1685.

- Rosel PE, France SC, Wang JY, Kocher TD. 1999. Genetic structure of harbour porpoise *Phocoena phocoena* populations in the northwest Atlantic based on mitochondrial and nuclear markers. *Mol Ecol*. 8: 541–554.
- Rosenbaum HC, et al. 2009. Population structure of humpback whales from their breeding grounds in the South Atlantic and Indian Oceans. *PLoS One* 4:e7318.
- Sambrook J, Fritsch EF, Maniatis T. 1989. *Molecular cloning: a laboratory manual*. Cold Spring Harbor (New York): Cold Spring Harbor Laboratory Press.
- Shen Y, et al. 2010. A SNP discovery method to assess variant allele probability from next-generation resequencing data. *Genome Res*. 20: 273–280.
- Siddall ME. 2012. *ild.run* [Internet]. Available from: <http://tnt.insectmuseum.org/index.php/Scripts/ild.run> (last accessed August 10, 2012).
- Smith DR, et al. 2008. Rapid whole-genome mutational profiling using next-generation sequencing technologies. *Genome Res*. 18: 1638–1642.
- Sorenson MD, Quinn TW. 1998. Numts: a challenge for avian systematics and population biology. *Auk* 115:214–221.
- Sosa MX, et al. 2012. Next-generation sequencing of human mitochondrial reference genomes uncovers high heteroplasmy frequency. *PLoS Comput Biol*. 8:e1002737.
- Sremba AL, Hancock-Hanser B, Branch TA, LeDuc RL, Baker CS. 2012. Circumpolar diversity and geographic differentiation of mtDNA in the critically endangered Antarctic Blue whale (*Balaenoptera musculus intermedia*). *PLoS One* 7:e32579.
- Steeleman ME, et al. 2009. Radiation of extant cetaceans driven by restructuring of the oceans. *Syst Biol*. 58:573–585.
- Stewart JB, et al. 2008. Strong purifying selection in transmission of mammalian mitochondrial DNA. *PLoS Biol*. 6:e10.
- Tamura K, et al. 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol*. 28:2731–2739.
- Thalmann O, Hebler J, Poinar HN, Pääbo S, Vigilant L. 2004. Unreliable mtDNA data due to nuclear insertions: a cautionary tale from analysis of humans and other great apes. *Mol Ecol*. 13:321–335.
- Thompson KF, et al. 2013. A novel conservation approach provides insights into the management of rare cetaceans. *Biol Conserv* 157: 331–340.
- Tiedemann R, Milinkovitch MC. 1999. Culture and genetic evolution in whales. *Science* 284:2055.
- Vilstrup J, et al. 2011. Mitogenomic phylogenetic analyses of the Delphinidae with an emphasis on the Globicephalinae. *BMC Evol Biol*. 11:65.
- Vollmer NL, Viricel A, Wilcox L, Moore KM, Rosel PE. 2011. The occurrence of mtDNA heteroplasmy in multiple cetacean species. *Curr Genet*. 57: 115–131.
- Watkins WA, Daher MA, Frstrup KM, Howald TJ, Notarbartolo di Sciarra G. 1993. Sperm whales tagged with transponders and tracked underwater by sonar. *Mar Mammal Sci*. 9:55–67.
- Whitehead H. 1998. Cultural selection and genetic diversity in matrilineal whales. *Science* 282:1708–1711.
- Whitehead H. 2002. Estimates of the current global population size and historical trajectory for sperm whales. *Mar Ecol Prog Ser*. 242: 295–304.
- Whitehead H. 2003. *Sperm whales: social evolution in the ocean*. Chicago (IL): University of Chicago Press.
- Whitehead H. 2005. Genetic diversity in the matrilineal whales: models of cultural hitchhiking and group-specific non-heritable demographic variation. *Mar Mammal Sci*. 21:58–79.
- Whitehead H, Christal J, Dufault S. 1997. Past and distant whaling and the rapid decline of sperm whales off the Galápagos Islands. *Conserv Biol*. 11:1387–1396.
- Whitehead H, Dillon M, Dufault S, Weigart L, Wright J. 1998. Non-geographically based population structure of South Pacific sperm whales: dialects, fluke-markings and genetics. *J Anim Ecol*. 67:253–262.
- Yang G, Guo LI, Bruford MW, Wei F, Zhou K. 2008. Mitochondrial phylogeography and population history of finless porpoises in Sino-Japanese waters. *Bio J Linn Soc*. 95:193–204.
- Zhang D-X, Hewitt GM. 1996. Nuclear integrations: challenges for mitochondrial DNA markers. *Trends Ecol Evol*. 11:247–251.

Associate editor: Kateryna Makova