Preliminary analysis of mitochondrial genome phylogeography of Blainville's, Cuvier's and Gervais' beaked whales.

Phillip A. Morin[1, 2], Sebastian Duchene[1], Nicole Lee[1, 2], John Durban[1], Diane Claridge[3].

[1] Protected Resources Division, Southwest Fisheries Science Center, National Marine Fisheries Service, National Oceanic and Atmospheric Administration, 3333 North Torrey Pines Ct, La Jolla, CA 92037, USA

[2] Scripps Institution of Oceanography, University of California San Diego, 9500 Gilman Dr., La Jolla, CA 92037, USA

[3] Bahamas Marine Mammal Research Organisation, P.O. Box AB-20714, Marsh Harbour, Abaco, The Bahamas

Keywords: Mitochondrial genome, mitogenome, phylogenetics, phylogeography

ABSTRACT:
Mitochondrial genomes are being used more frequently for phylogenetics, phylogeography and population genetics due to the greater levels of variability and resolution than is obtained from mtDNA fragments such as the control region or cytochrome B. This report describes preliminary analysis of mitogenomic diversity and phylogenetic patterns in three species of beaked whales. The mitogenomic phylogeny of these species and representatives of some other beaked whale species provides strong support for sister group relationships between *H. ampullatus* and the two *Mesoplodon* species, while *Z. cavirostris* was more distantly related to these three species. Within *Mesoplodon densirostris*, haplotypes were divided into two clades representing the western Atlantic and the Pacific. Within *Ziphius*, there were 3 major clades, but samples from the Atlantic were found in all three clades, and samples from the Pacific were found in two of the clades. No haplotypes were shared between ocean basins for either species. This complex pattern in *Ziphius* indicates multiple inter-ocean migration events in recent evolutionary history, or possibly some ongoing gene flow between ocean basins.

INTRODUCTION:

Among the speciose yet poorly known beaked whales (Ziphiidae), only two species are found in multiple oceans: Cuvier's beaked whale (*Ziphius cavirostris*) and Blainsville's beaked whale (*Mesoplodon densirostris*). Beaked whales are typically found in deep offshore waters, spending long periods of time diving to great depths to forage (Hooker and Baird, 1999; Tyack *et al.*, 2006) They have come to greater attention of biologists and conservationists due to their particular sensitivity to military sonar and seismic research and exploration, which have been implicated in several mass strandings (Cox *et al.*, 2006; Jepson *et al.*, 2003; Schrope, 2002).

Substantial effort has gone into using molecular methods to infer phylogenetics, phylogeography, and population structure in ziphiids, particularly in *Mesoplodon* spp and *Z. cavirostris* (Dalebout *et al.*, 2004; Dalebout *et al.*, 2007; Dalebout *et al.*, 2005; Dalebout *et al.*, 2008) to facilitate assessment of risk to potential localized populations from military and other acoustic activities. These studies have made use of various portions of the mitochondrial genome, including the control region and cytochrome B, or nuclear introns, and a combination of "fresh" tissue samples (from biopsies and stranded animals) and DNA extracted from museum bone and tooth specimens. Although these mtDNA fragments have contributed significantly to our understanding of taxonomy and population structure of beaked whales, low levels of variation and conflicting or unresolved relationships remain.

Recent advances in DNA sequencing, however, are making it possible to obtain much longer and more diverse DNA sequences, including whole mitochondrial genomes, from increasingly larger numbers of individuals for use in phylogenetic, phylogeography and population studies (e.g., Hancock-Hanser *et al.*, in revision; Maricic *et al.*, 2010; Morin *et al.*, 2010; Stiller *et al.*, 2009). The advantages of using whole mitochondrial genomes include more accurate mitochondrial tree topologies and estimates of divergence times (Duchene *et al.*, 2011). In particular, use of complete mitogenomes may help to resolve presently unresolved relationships among ziphiids, such as the placement of the single-species genus *Ziphius* relative to the mesoplodont and hyperoodont whales (Dalebout *et al.*, 2004; McGowen *et al.*, 2009). Population mitogenomics might also provide more data on causes and timing of speciation, intra-specific diversity, and potential identification of additional species or subspecies of these very poorly understood odontocetes. It is notable that no subspecies have been described among the currently recognized 21 species of ziphiidae. The species examined here are some of the most likely, given their multi-ocean-basin distributions, to be comprised of multiple unrecognized subspecies.

MATERIALS AND METHODS:

Samples

Samples were collected from stranded animals and by remote biopsy of free-swimming whales (e.g., Hooker *et al.*, 2001) and stored frozen in the Southwest Fisheries Science Center's marine mammal and turtle tissue collection, with or without preservative (e.g., salt-saturated 20% DMSO solution or ethanol). DNA was extracted using silica-based filter purification (Qiaxtractor® DX reagents, Qiagen, Valencia, CA, USA) following manufacturers' instructions, performed on a JANUS® automated workstation (Perkin-Elmer, Waltham, MA, USA). Samples were selected from the SWFSC collection to include multiple samples from three species from The Bahamas in the northwestern Atlantic : *M. europaeus*, *M. densirostris*, and *Z. cavirostris*. Additional samples were selected to maximize geographic distribution of these three species (Supplementary Figure 1). Sample selection was motivated primarily by ascertainment criteria for SNP discovery in nuclear sequences that were generated at the same time as the mitogenome data (Hancock-Hanser *et al.*, in revision), for later use in both local (Bahamas) and global population genetics studies. See Table 1 for sample details.

Sequencing and mitogenome assembly

Genomic DNA libraries were prepared and given individual indexing sequences for multiplexing prior to pooling, library enrichment and sequencing as described in Hancock-Hanser *et al.* (in revision). Sample libraries were pooled prior to capture array enrichment, and sample libraries for all species were enriched

using sequence baits from the published mitochondrial genome of *Hyperoodon ampullatus* (Accession No. NC_005273.1, Arnason *et al.*, 2004). The pooled, enriched library was sequenced using single-end sequencing on the HiSeq2000 Analyzer with the TruSeqSR cluster kit (Illumina, Inc, San Diego, CA, USA).

Because we did not have a reference mitogenome of any of the three species, we initially pooled reads from four (*Z. cavirostris, M. densirostris*) or eight (*M. europaeus*) individuals and assembled those to the *H. ampullatus* reference sequence using CLC Genomics Workbench v. 4.0 (CLC Genomics, Muehltal, Germany), using relaxed criteria (Similarity = 0.8, Length fraction = 0.8, Insertion cost = 3, Deletion cost = 3, Mismatch cost = 3). The assembly was visually checked for clusters of differences, indels, and stop codons or frame shifts in coding regions, and manually edited.

The consensus sequences for each species were then used as the reference sequence for assembly of mitogenomes from each individual using a custom pipeline written in R (R Development Core Team, 2006). Reads were first assembled to their respective reference mitogenomes with the program BWA (Li and Durbin, 2009). The mpileup module in SAMTOOLS (Li *et al.*, 2009) was used to convert the resulting BAM-format alignment file into a "pileup" text format that lists the base composition across reads at each site in the reference sequence. This text file was then parsed by custom R code to create the consensus sequence for each individual, using the following rules: If a given site had fewer than three reads covering it, an "N" was placed in the consensus. If the coverage was between three and five, and all reads contained the same nucleotide, then that nucleotide was used in the consensus, otherwise, the consensus received an "N" for that site. If coverage was greater than five, then the nucleotide that occurred in 70% or more of the reads was used in the consensus. If no nucleotide frequency exceeded 70%, then an "N" was inserted. Parameter settings for each step in the pipeline are available from the authors on request. Mitogenomes for each species were originally aligned to each other using MAFFT (Katoh *et al.*, 2005) and assemblies were visually checked in Geneious (V. 5.5.6; Biomatters Ltd, Auckland, NZ).

Gene identification and annotation was performed by importing GenBank sequence annotations for the *H. ampullatus* mitogenome into the newly assembled mitogenomes, followed by a complete inspection of individual gene coverage and reading frame matching in each of the new mitogenomes.

Sequence preparation and evolutionary model selection

The 49 samples were aligned in ClustalW (Thompson *et al.*, 2002) as implemented in Seaview (Gouy *et al.*, 2010), and the number of mitogenomic haplotypes and identical samples were determined so that redundant haplotypes could be removed. In order to perform subsequent phylogenetic analyses, complete mitogenomes of the following taxa were downloaded from NCBI to be used as outgroups and calibration points: *Hyperoodon ampullatus*, *Orcinus orca*, *Monodon monoceros*, *Phocoena phocoena*, and *Lipotes vexillifer* (Accession numbers listed in Table 1).

The outgroups and unique beaked whale haplotypes were then aligned using ClustalW. Then mitogenomic regions *12S* and *16S*, all coding genes, and the *D-loop* were extracted as separate partitions, then the PEGAS package in R (Paradis, 2010) was used to calculate nucleotide diversity (π) and total number of haplotypes in the complete dataset and the beaked whales exclusively.

Evolutionary model selection was performed for each of the three partitions, as implemented in the Phangorn package in R (v1.6-0; Schliep, 2011). All nucleotide substitution models available in the software were tested, and the best model was selected according to the Bayesian Information Criterion (BIC) and a sample size of 100.

Phylogenetic analyses

A Maximum Likelihood (ML) tree was first estimated using GARLI v2.0 (Zwickl, 2006). The software was set for a partitioned analysis, where each partition was assigned its respective BIC model, and a tree topology and branch lengths were then optimized across all partitions, so that a single ML tree was obtained, given that mitochondrial genes in vertebrates are expected to follow the same evolutionary trajectory (Galtier *et al.*, 2009). Branch support was then assessed by performing 100 bootstrap replicates

of all mitogenome sites using the Nearest Neighbor Interchange (NNI) algorithm and likelihood as implemented in Phangorn v1.6-0 (Schliep, 2011).

The BEAST v1.6.2 (Drummond and Rambaut, 2007) was used to co-estimate divergence times and a tree topology under a Bayesian (BI) framework. Settings for the analyses were as follows: Relaxed log-normal molecular clocks and BIC selected substitution models for each of the three partitions, as this method has proved appropriate for complete mitogenomes in cetaceans and other vertebrates (Ho and Lanfear, 2010); and a single tree prior assuming a Yule speciation process, as this model tends to be a good fit for phylogenetic studies including different species (Drummond and Rambaut, 2007). Four prior calibrations from fossils and previous molecular-based dates (McGowen *et al.*, 2009) were used for the radiation times of taxonomic groups; Delphinida and Ziphiidae, Delphinida, Delphinoidea, and Phocoenida and Monodontidae (see Table 2). Calibrations in Table 2 are shown in real scale but were set as priors in lognormal scale in the software. We used a MCMC chain length of 500,000,000, sampling every 5000, and a second identical run to assure convergence of independent analyses. Convergence of parameter estimates was determined in the Tracer program (Rambaut and Drummond, 2003), verifying that sample sizes were at least 500 for all parameters.

RESULTS

Complete mitogenome sequences were obtained from 8 *M. europaeus*, 19 *M. densirostris*, and 22 *Z. cavirostris*, with sequence lengths between 16330 and 16352 bp. The 49 aligned samples resulted in 43 unique beaked whale haplotypes, for a total of 48 sequences in the analysis, including outgroups and calibrations (Table 1), and an alignment length of 17632 bp. Identical mitogenome haplotypes corresponded to samples within species and found at the same geographic region or within a short distance (Table 1).

Genetic diversity in terms of nucleotide diversity and number of haplotypes was different among the *12S+16S*, coding sites, and the *D-loop* partitions (Table 3). The *12S+16S* had the lowest diversity, whereas coding sites had the highest, in contrast to other studied cetaceans (Duchene *et al.*, 2011) and other mammals (Willerslev *et al.*, 2009) where the *D-loop* has been shown to be highly polymorphic. The preferred nucleotide substitution model according to the BIC was GTR for all partitions, with an additional gamma rate parameter for the coding sites only.

Tree topologies for the ML and BI analyses were identical, with branch support between 95 and 99% for the ML bootstrap, and 0.96 and 0.99 for the BI posterior probabilities, for most branches (Figure 1). The only poorly supported clades displayed a bootstrap of 40 and 30% and posterior probabilities of 0.22 and 0.21, and corresponded to two nodes within Clade 2 of *Z. cavirostris* as shown as a sub-tree close-up of these taxa in Figure 2.

Phylogenetic relationships among beaked whale species were highly supported and consistent with previous studies using diverse sampling of cetacean mitochondrial and nuclear loci (McGowen *et al.*, 2009). *H. ampullatus* was a sister taxon to the two *Mesoplodon* species, and *Z. cavirostris* was more distantly related to these three species. However, these results are inconsistent with *D-loop* (437 bp), Cytochrome b (*Cyt b*) (384 bp) and nuclear actin intron (925 bp) phylogenies (Dalebout *et al.*, 2004) which place *H. ampullatus* as a sister taxon to *Ziphius* instead of *Mesoplodon,* though support for those groupings is low.

Different patterns were obtained for species with sampling of Pacific and Atlantic oceans (*M. densirostris* and *Z. cavirostris*). In the case of *M. densirostris*, two highly supported clades were identified, corresponding to Atlantic and Pacific oceans, as shown in Figure 1 and Table 2. On the contrary, in *Z. cavirsotris* there was no reciprocal monophyly for all Atlantic and Pacific haplotypes in the identified clades (Figures 1 and 2).

Clades 1 and 2 in *Z. cavirostris* (Figure 2) contained haplotypes from both oceans, but within Clade 2 there was a clear grouping of haplotypes by ocean basin, where samples from the Atlantic (Bahamas and Puerto

Rico) were in a highly supported group and apart from those from the Pacific. Clade 1 did not display this pattern, and Clade 3 contained samples from the Atlantic only (Bahamas).

Times to the most recent common ancestor (TMRCA), expressed in Million Years Before Present (MYBP), from the Bayesian analysis are reported in Table 2. There were no conflicting calibrations, given that prior and posterior TMRCA distributions for calibrated nodes (nodes I, II, III, and VI) overlapped almost entirely (Table 2). The origin of all beaked whale haplotypes in this study was estimated at 20.06 MYBP (HPD: 16.76 – 23.53), and TMRCA for species with multiple haplotypes (*M. europaeus*, *M. densirostris*, and *Z. cavirostris*) ranged from 2.51 (HPD: 1.94 – 3.09) in *Z. cavirostris* to 0.88 (HPD: 0.66 – 1.13) in *M. europaeus*. Divergence times between the *Mesoplodon* species were consistent with previous studies in this genus based on nuclear introns (Dalebout *et al.*, 2008).

Radiation times of haplotype clades within species overlapped considerably, showing a similar diversification timeframe in these species. Clades within *Z. cavirostris* were estimated to have radiated (i.e., the most recent common ancestor for members of a clade) between 1.89 (HPD: 1.45 – 2.37) for Clade 1+2, to 0.054 (0.013 – 0.11) for Clade 3. In *M. densirostris* Atlantic and Pacific clades were estimated to have radiated 1.15 (HPD: 0.73 – 2.74) and 0.6 (HPD: 0.33 – 1.81) MYBP, respectively.

DISCUSSION AND CONCLUSIONS:

One of the main limitations of phylogenies based on short sequence fragments such as the mtDNA control region is that they often have limited phylogenetic resolution and haplotypes may be shared between widely separated populations and even subspecies or species in cetaceans. Previous work based on a variety of mitochondrial and nuclear sequences has provided strong support for the species-level phylogeny of beaked whales with deep interspecific divergence times (Dalebout *et al.*, 2008; McGowen *et al.*, 2009), yet intra-specific patterns indicate that diversity is too low for resolution of population diversification except by differences in haplotype frequencies (e.g., Dalebout *et al.*, 2005). The data presented here provide substantially greater genetic diversity within the rapidly evolving mitochondrial genome, such that samples from different geographic regions have unique haplotypes and highly resolved phylogenetic trees can be produced to investigate intra-specific phylogeography and divergence times.

This preliminary analysis, while providing the largest number of ziphiid mitogenomes to date, is limited by the number and geographic diversity of samples sequenced. For all three species studied, intra-specific diversity is high and suggests that radiations within the species date to between approximately 1 and 2.5 MY. The phylogeographic patterns indicate that the tropical *M. densirostris* have diverged in separate ocean basins for approximately 2 MY, suggesting that inter ocean-basin gene flow is unlikely and further genetic and morphological research could suggest separation into Atlantic and Pacific subspecies or species. In contrast, the widely distributed *Z. cavirostris* shows limited clustering of haplotypes by ocean basin, and is likely to have experienced multiple episodes of gene flow between ocean basins over the last approximately 2.5 MY. The lack of strong phylogeographic patterns in *Z. cavirostris* suggests that further research is needed before drawing inferences about contemporary gene flow. Nevertheless, previous studies that indicated high probability of isolation among regional populations (Dalebout *et al.*, 2005; McSweeney *et al.*, 2007) are supported, as the most similar haplotypes tend to be geographically clustered as well.

These preliminary studies indicate that the deep diversity within species can be captured using next-generation sequencing methods to sequence large numbers of mitogenomes and provides a useful tool for better understanding of population structure and evolutionary patterns in beaked whales. Further geographic sampling of mitogenomes along with studies of nuclear loci at the population level should help to clarify the scale of population structure and levels of gene flow.

ACKNOWLEDGEMENTS:
Sample collection from free-ranging beaked whales in The Bahamas and generation of genetic data were supported by an Office of Naval Research grant (N00014-07-1-0120). Laboratory analysis was supported by the SWFSC, and data analysis of mitogenomes was supported by a grant from the Scientific Committee of the International Whaling Commission. We are grateful to all of the people and organizations who collected and donated samples used in this study: Leigh Hickmott, Charlotte Dunn and other field assistants from the Bahamas Marine Mammal Research Organisation , The Bahamas Marine Mammal Stranding Network, Susan Chivers, Kerri Danil and the California Stranding network, the NMFS Southwest Region observer program managers and drift gill net observers, Cascadia Research, Jay Barlow, Tim Gerrodette, Bob Pitman, Luella Dolar, Daniel Palacios, Jan Straley, Scott Baker, Merel Dalebout, the New Zealand Cetacean Tissue Archive curated at the University of Auckland, The New Zealand Department of Conservation (DOC), the Texas Stranding Network, Hawaii Pacific University Stranding Program, NOAA Pacific Islands Observer Program, Greg Schorr, Erin Falcone, and John Calambokidis and the Cascadia Research Collective, Marlo Mota, Joanne Whaley, Judy St. Leger and the Hubbs-SeaWorld Research Institute, the Alaska marine mammal stranding network. We are also grateful to Barbara Taylor, Bob Pitman, and Bill Perrin for comments on drafts of the manuscript.

Table 1.

Laboratory sample ID or Genbank accession ID, haplotype names, species, and geographic location of all the samples used in the analysis.

Sample ID	Haplotype name in analysis	Species	Geographic Location
NC5273.1+AJ554056.1	Hyperoodon_ampullatus	*Hyperoodon ampullatus*	-
gi|270301949|gb|GU187185.1|	Orcinus_orca	*Orcinus orca*	-
gi|38602571|emb|AJ554062.1|	Monodon_monoceros	*Monodon monoceros*	-
gi|38602585|emb|AJ554063.1|	Phocoena_phocoena	*Phocoena phocoena*	-
gi|55274416|gb|AY789529.1|	Lipotes_vexillifer	*Lipotes vexillifer*	-
z0002698	MeuropFLUSA1	*Mesoplodon europaeus*	FL, USA
z0002817	MeuropNCUSA	*Mesoplodon europaeus*	NC, USA
z0003853	MeuropFLUSA2	*Mesoplodon europaeus*	FL, USA
z0004010	MdenSBCA	*Mesoplodon densirostris*	Santa Barbara, CA, USA
z0004120	MeuropFLUSA3	*Mesoplodon europaeus*	FL, USA
z0004472	ZcavFLUSA1	*Ziphius cavirostris*	FL, USA
z0004967	ZcavHIUSA1	*Ziphius cavirostris*	HI, USA
z0005565	ZcavPhillipines	*Ziphius cavirostris*	Philippines
z0007445	ZcavFLUSA2	*Ziphius cavirostris*	FL, USA
z0008681	MdenFlUSA	*Mesoplodon densirostris*	FL, USA
z0009110	MdenStrndNZ	*Mesoplodon densirostris*	New Zealand (Stranding)
z0009122	ZcavNZ	*Ziphius cavirostris*	New Zealand
z0009561	ZcavTaiwan	*Ziphius cavirostris*	Taiwan
z0014950	ZcavPR	*Ziphius cavirostris*	Puerto Rico

z0024816	ZcavAKUSA1	*Ziphius cavirostris*	AK, USA
z0026279	ZcavWestCoastUSA	*Ziphius cavirostris*	West Coast, USA
z0030065	ZcavHIUSA2	*Ziphius cavirostris*	HI, USA
z0033737	MdenHI1	*Mesoplodon densirostris*	HI, USA
z0050723	MdenHINorth	*Mesoplodon densirostris*	North HI, USA
z0061950	ZcavCAUSA1	*Ziphius cavirostris*	CA, USA
z0068606	ZcavBajaMX	*Ziphius cavirostris*	Baja California, Mexico
z0072279	ZcavETPMX	*Ziphius cavirostris*	ETP, Mexico
z0074252	MeuropBahamas1	*Mesoplodon europaeus*	Bahamas
z0074259	MeuropBahamas2	*Mesoplodon europaeus*	Bahamas
z0074265	MeuropBahamas3	*Mesoplodon europaeus*	Bahamas
z0074425+z0074424	MdenBahamas7	*Mesoplodon densirostris*	Bahamas
z0079632	ZcavAKUSA2	*Ziphius cavirostris*	AK, USA
z0079824	MdenBahamas1	*Mesoplodon densirostris*	Bahamas
z0079835	MdenBahamas2	*Mesoplodon densirostris*	Bahamas
z0079837	MdenBahamas3	*Mesoplodon densirostris*	Bahamas
z0079838	MdenBahamas4	*Mesoplodon densirostris*	Bahamas
z0079840+z0079839	MdenBahamas5	*Mesoplodon densirostris*	Bahamas
z0079887	ZcavBahamas1	*Ziphius cavirostris*	Bahamas
z0079893	ZcavBahamas2	*Ziphius cavirostris*	Bahamas
z0079898+z0074262	ZcavBahamas3	*Ziphius cavirostris*	Bahamas
z0079900	ZcavBahamas4	*Ziphius cavirostris*	Bahamas
z0087481	ZcavCAUSA2	*Ziphius cavirostris*	CA, USA
z0094563+z0074263+z0	MdenBahamas6	*Mesoplodon*	Bahamas

074264		*densirostris*	
z0094578	MeuropBahamas4	*Mesoplodon europaeus*	Bahamas
z0094591	ZcavBahamas5	*Ziphius cavirostris*	Bahamas
z0094595	ZcavBahamas6	*Ziphius cavirostris*	Bahamas
z0094785	MdenHIMAUI	*Mesoplodon densirostris*	Maui, HI, USA
z0098730+z0033736	MdenHI2	*Mesoplodon densirostris*	HI, USA

Table 2: Prior and posterior TMRCA, and posterior probability branch support for nodes labeled in Figure 1.

Taxonomic group	Node label Node	Prior TMRCA (MYBP) Median	s.d.	Posterior TMRCA Median	HPD	Branch support Posterior probability
Delphinida+Ziphiidae [m]	I	34.08	1.08	34.74	31.26 – 38.42	1
Delphinida [m]	II	29.08	1.08	31.23	28.28 – 32.33	1
Delphinoidea [m]	III	22.27	1.11	21.80	19.54 – 24.14	1
M. densirostris	IV	Tree prior	Tree prior	2.13	1.74 – 2.81	1
Ziphiidae	V	Tree prior	Tree prior	20.06	16.76 – 23.53	1
Phocoenidae+Monodontidae [f]	VI	17.68	1.13	16.35	15.29 – 19.44	1
Z. cavirostris	VII	Tree prior	Tree prior	2.51	1.94 – 3.09	1
Z. cavirostris clade 1+2	VIII	Tree prior	Tree prior	1.89	1.45 – 2.37	0.99
Z. cavirostris clade 1	IX	Tree prior	Tree prior	0.98	0.60 – 4.39	1
Z. cavirostris clade 2	X	Tree prior	Tree prior	0.47	0.34 – 0.64	0.96
Z. cavirostris clade 3	XI	Tree prior	Tree prior	0.054	0.013 – 0.11	0.96
M. europaeus	XII	Tree prior	Tree prior	0.88	0.66 – 1.13	1
M. densirostris Atlantic	XIII	Tree prior	Tree prior	1.15	0.73 – 2.74	0.99
M. densirostris Pacific	XIV	Tree prior	Tree prior	0.6	0.33 – 1.81	0.99
Mesoplodon+Hyperoodon	XV	Tree prior	Tree prior	16.64	13.55 – 19.87	0.99
Mesoplodon	XVI	Tree prior	Tree prior	12.54	10.06 – 15.18	0.99

[m] Molecular based calibrations; [f] fossil based calibrations (McGowen *et al.*, 2009).

Table 3: Substitution models, median rate and highest posterior density (HPD), nucleotide and mitogenomic haplotype diversities for the three partitions in the Bayesian phylogenetic analysis.

Partition	BIC selected substitution model	Median rate 10^{-3} (substitutions/site/MY)	HPD	Nucleotide diversity	Number of haplotypes
12S+16S	GTR	0.87	0.66 – 1.12	0.0504	12
Coding sites	GTR+G	3.88	3.50 – 4.27	0.085	19
D-loop	GTR	3.05	2.66 – 3.46	0.052	12

Figure 1. Chronogram for the highest clade credibility tree. Clades are colored to distinguish intra-species groupings, the scale bar corresponds to million years before present (MYBP), and node bars (in blue) represent uncertainty of node ages. Posterior probabilities for all nodes were >0.96 in all cases, except for nodes labeled with red and yellow dots, with posterior probabilities of 0.22 and 0.21, respectively. Node labels correspond to information in Table 2.

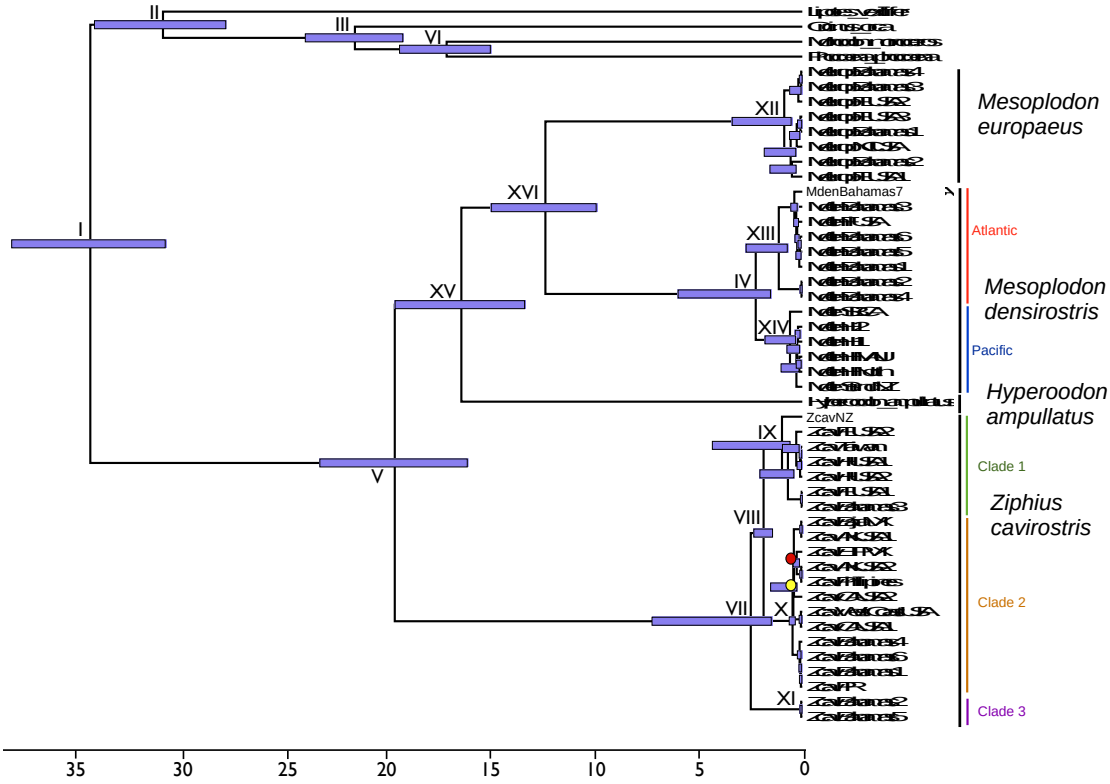
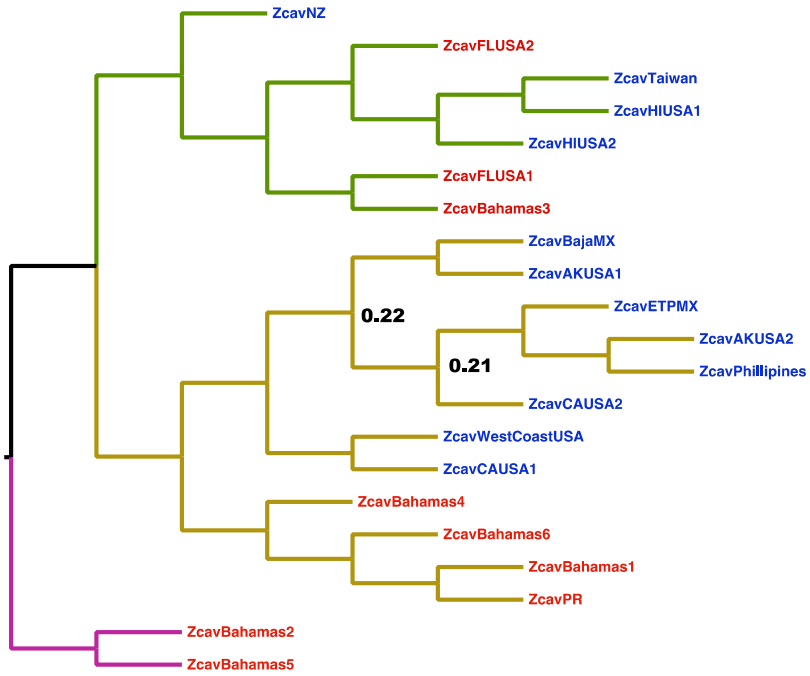
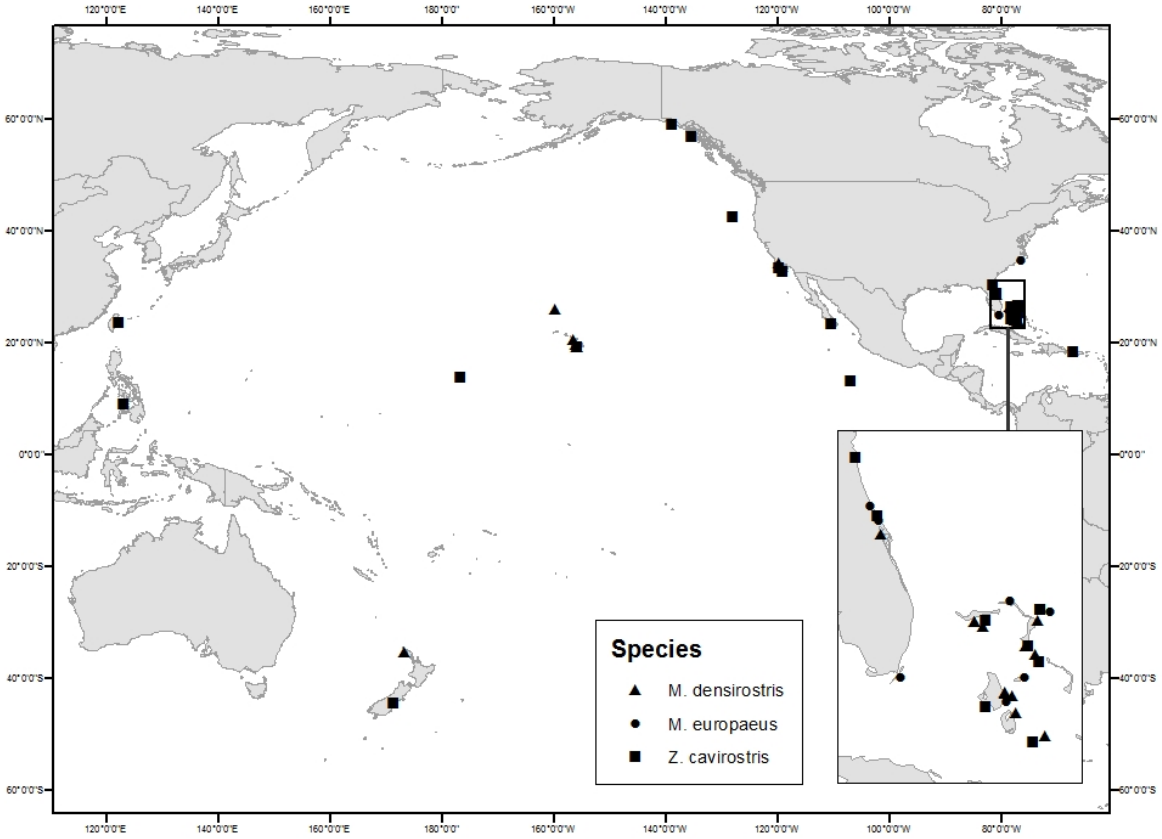


Figure 2.

Close-up of phylogenetic relationships in *Z. cavirostris* from Figure 1. Posterior probabilities are displayed only for nodes with posterior probabilities below 0.99. Branch colors correspond to Clades 1-3 in Figure 1, and sample names are color coded by ocean (Red=Atlantic, Blue=Pacific). Branch lengths were assigned a single arbitrary value for a better display of the topology, so their magnitude is not relevant.



Supplementary Figure 1: Map of sampling locations



References:

Arnason U., Gullberg A. and Janke A. 2004. Mitogenomic analyses provide new insights into cetacean origin and evolution. *Gene* 333:27-34.

Cox T.M., Ragen T.J., Read A.J., Vos E., Baird R.W., Balcomb K., Barlow J., Caldwell J., Cranford T., Crum L., D'Amico A., D'Spain G., Fernández A., Finneran J., Gentry R., Gerth W., Gulland F., Hildebrand J., Houserp D., Hullar T., Jepson P.D., Ketten D., Macleod C.D., Miller P., Moore S., Mountain D.C., Palka D., Ponganis P., Rommel S., Rowles T., Taylor B., Tyack P., Wartzok D., Gisiner R., Meads J. and Benner L. 2006. Understanding the impacts of anthropogenic sound on beaked whales. *Journal of Cetacean Research and Management* 7:177-187.

Dalebout M.L., Baker C.S., Mead J.G., Cockcroft V.G. and Yamada T.K. 2004. A comprehensive and validated molecular taxonomy of beaked whales, family Ziphiidae. *Journal of Heredity* 95:459-473.

Dalebout M.L., Baker C.S., Steel D., Robertson K.M., Chivers S.J., Perrin W.F., Mead J.G., Grace R.V. and Schofield T.D. 2007. A divergent mtDNA lineage among Mesoplodon beaked whales: Molecular evidence for a new species in the tropical Pacific? *Marine Mammal Science* 23:954-966.

Dalebout M.L., Robertson K.M., Frantzis A., Engelhaupt D., Mignucci-Giannoni A.A., Rosario-Delestre R.J. and Baker C.S. 2005. Worldwide structure of mtDNA diversity among Cuvier's beaked whales (Ziphius cavirostris): implications for threatened populations. *Mol Ecol* 14:3353-3371.

Dalebout M.L., Steel D. and Baker C.S. 2008. Phylogeny of the beaked whale genus *Mesoplodon* (Ziphiidae: Cetacea) revealed by nuclear introns: implications for the evolution of male tusks. *Systematic Biology* 57:857-875.

Drummond A. and Rambaut A. 2007. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC evolutionary biology* 7:214.

Duchene S., Archer F.I., Vilstrup J., Caballero S. and Morin P.A. 2011. Mitogenome phylogenetics: the impact of using single regions and partitioning schemes on topology, substitution rate and divergence time estimation. *PLoS One* 6:e27138.

Galtier N., Nabholz B., Glemin S. and Hurst G. 2009. Mitochondrial DNA as a marker of molecular diversity: a reappraisal. *Molecular Ecology* 18:4541-4550.

Gouy M., Guindon S. and Gascuel O. 2010. SeaView version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Molecular biology and evolution* 27:221-224.

Hancock-Hanser B., Frey A., Leslie M., Dutton P.H., Archer E.I. and Morin P.A. in revision. Targeted multiplex next-generation sequencing: Advances in techniques of mitochondrial and nuclear sequencing for population genomics *Mol Ecol Res*.

Ho S.Y.W. and Lanfear R. 2010. Improved characterisation of among-lineage rate variation in cetacean mitogenomes using codon-partitioned relaxed clocks. *Mitochondrial DNA* 21:138-146.

Hooker S.K. and Baird R.W. 1999. Deep-diving behaviour of the northern bottlenose whale, Hyperoodon ampullatus (Cetacea : Ziphiidae). *Proceedings of the Royal Society of London Series B-Biological Sciences* 266:671-676.

Hooker S.K., Iverson S.J., Ostrom P. and Smith S.C. 2001. Diet of northern bottlenose whales inferred from fatty-acid and stable-isotope analyses of biopsy samples. *Canadian Journal of Zoology-Revue Canadienne De Zoologie* 79:1442-1454.

Jepson P.D., Arbelo M., Deaville R., Patterson I.A., Castro P., Baker J.R., Degollada E., Ross H.M., Herraez P., Pocknell A.M., Rodriguez F., Howie F.E., Espinosa A., Reid R.J., Jaber J.R., Martin V., Cunningham A.A. and Fernandez A. 2003. Gas-bubble lesions in stranded cetaceans. *Nature* 425:575-576.

Katoh K., Kuma K., Toh H. and Miyata T. 2005. MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic Acids Res* 33:511-518.

Li H. and Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25:1754-1760.

Li H., Handsaker B., Wysoker A., Fennell T., Ruan J., Homer N., Marth G., Abecasis G. and Durbin R. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25:2078-2079.

Maricic T., Whitten M. and Paabo S. 2010. Multiplexed DNA sequence capture of mitochondrial genomes using PCR products. *PLoS One* 5:e14004.

McGowen M.R., Spaulding M. and Gatesy J. 2009. Divergence date estimation and a comprehensive molecular tree of extant cetaceans. *Molecular phylogenetics and evolution* 53:891-906.

McSweeney D.J., Baird R.W. and Mahaffy S.D. 2007. Site fidelity, associations, and movements of Cuvier's (*Ziphius cavirostris*) and Blainville's (*Mesoplodon densirostris*) beaked whales off the island of Hawai'i. *Marine Mammal Science* 23:666-687.

Morin P.A., Archer F.I., Foote A.D., Vilstrup J., Allen E.E., Wade P., Durban J., Parsons K., Pitman R., Li L., Bouffard P., Nielsen S.A., Rasmussen M., Willerslev E., Gilbert M.T.P. and Harkins T. 2010. Complete mitochondrial genome phylogeographic analysis of killer whales (*Orcinus orca*) indicates multiple species. *Genome Research* 20:908-916.

Paradis E. 2010. pegas: an R package for population genetics with an integrated modular approach. *Bioinformatics* 26:419-420.

R Development Core Team R. 2006. R: a language and environment for statistical computing. *R foundation for Statistical Computing*.

Rambaut A. and Drummond A. 2003. Tracer: a program for analysing results from Bayesian MCMC programs such as BEAST & MrBayes. Oxford, UK. http://evolve. zoo. ox. ac. uk/software. html.

Schliep K.P. 2011. phangorn: Phylogenetic analysis in R. *Bioinformatics* 27:592-593.

Schrope M. 2002. Whale deaths caused by US Navy's sonar. *Nature* 415:106.

Stiller M., Knapp M., Stenzel U., Hofreiter M. and Meyer M. 2009. Direct multiplex sequencing (DMPS)--a novel method for targeted high-throughput sequencing of ancient and highly degraded DNA. *Genome Res* 19:1843-1848.

Thompson J.D., Gibson T.J. and Higgins D.G. 2002. Multiple sequence alignment using ClustalW and ClustalX. *Current protocols in bioinformatics*.

Tyack P.L., Johnson M., Soto N.A., Sturlese A. and Madsen P.T. 2006. Extreme diving of beaked whales. *J Exp Biol* 209:4238-4253.

Willerslev E., Gilbert M.T.P., Binladen J., Ho S.Y.W., Campos P.F., Ratan A., Tomsho L.P., Da Fonseca R.R., Sher A. and Kuznetsova T.V. 2009. Analysis of complete mitochondrial genomes from extinct and extant rhinoceroses reveals lack of phylogenetic resolution. *BMC evolutionary biology* 9:95.

Zwickl D.J. 2006. Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion.