J Mol Evol (2009) 68:97–111
DOI 10.1007/s00239-008-9193-2

Comparing Evolutionary Patterns and Variability in the Mitochondrial Control Region and Cytochrome *b* in Three Species of Baleen Whales

S. Elizabeth Alter · Stephen R. Palumbi

Received: 17 October 2007 / Accepted: 9 December 2008 / Published online: 31 December 2008

Abstract The rapidly evolving mitochondrial control region remains an important source of information on phylogeography and demographic history for cetaceans and other vertebrates, despite great uncertainty in the rate of nucleotide substitution across both nucleotide positions and lineages. Patterns of variation in linked markers with slower rates of evolution can potentially be used to calibrate the rate of nucleotide substitution in the control region and to better understand the interplay of evolutionary and demographic forces across the mitochondrial genome above and below the species level. We have examined patterns of diversity within and between three baleen whale species (gray, humpback, and Antarctic minke whales) in order to determine how patterns of molecular evolution differ between cytochrome *b* and the control region. Our results show that cytochrome *b* is less variable than expected given the diversity in the control region for gray and humpback whales, even after functional differences are taken into account, but more variable than expected for minke whales. Differences in the frequency distributions of polymorphic sites and in best-fit models of nucleotide substitution indicate that these patterns may be the result of hypervariability in the control region in gray and humpback whales but, in minke whales, may result from a large, stable or expanding population size coupled with saturation at the control region. Using paired cytochrome *b* and control region data across individuals, we show that the average rate of nucleotide substitution in the control region may be on average 2.6 times higher than phylogenetically derived estimates in cetaceans. These results highlight the complexity of making inferences from control region data alone and suggest that applying simple rules of DNA sequence analyses across species may be difficult.

Keywords Substitution rate · Mammalian · Mitochondrial genome · Marine mammal · Cetacea

S. E. Alter · S. R. Palumbi
Department of Biological Sciences, Hopkins Marine Station, Stanford University, 120 Oceanview Boulevard, Pacific Grove, CA 93950, USA

S. E. Alter (✉)
NRDC, 40 W 20th Street, New York, NY 10011, USA
e-mail: lalter@nrdc.org; sealter@stanford.edu

Introduction

The control region is a noncoding segment of the mitochondrial genome that is highly variable in many vertebrate species. Because of this variability, the control region has frequently been used to estimate population divergence and gene flow (e.g., reviewed by Hoelzel 1994), effective population size (Rooney et al. 2001; Roman and Palumbi 2003), and population expansion or decline (e.g., Bradley et al. 1996) in natural populations. However, in many organisms such as primates and rodents, several features of the control region complicate genealogical analyses. These features include rate heterogeneity (Meyer et al. 1999), bias in nucleotide composition (Raina et al. 2006), higher frequency of transitions than transversions (Wakeley 1994), and the violation of clock-like behavior in evolutionary rate (Aquadro and Greenberg 1983; Kocher and Wilson 1991; Tamura and Nei 1993). Because a good proportion of the population genetic analyses performed on control region data are genealogy based, an improved understanding of evolutionary patterns and rates at the intraspecific level is critical to validating studies of within-species genetic variation that rely on this marker.

Understanding the patterns and rate of substitution in the control region is particularly important for baleen whales. Because of the slow rate of evolution of whale DNA, perhaps due to their large body size (Martin and Palumbi 1993; Gillooly et al. 2005), few genetic markers have been found to be suitably variable to study demographic processes with the exception of mitochondrial and microsatellite markers. In addition, many tissue samples used in genetic studies of baleen whales come from stranded animals. When such samples are highly degraded, mitochondrial markers tend to be more easily amplified than nuclear markers due to the high copy number in the cell. Because of its high variability and ease of amplification, the control region has been the genetic marker of choice in many cetacean studies, with implications for baleen whale stock determination, migration, historical demography, and forensics (e.g., Baker et al. 1993; Rosenbaum et al. 2000; Roman and Palumbi 2003; Dalebout et al. 2005).

In most baleen whale species, the control region exhibits higher than expected levels of polymorphism, given the current estimates of population sizes (Roman and Palumbi 2003). Surprisingly high genetic diversity and deep genetic divergences, as observed in species including North Atlantic minke (Roman and Palumbi 2003), gray (LeDuc et al. 2002), humpback (Baker and Medrano-Gonzalez 2002), and bowhead (Rooney et al. 2001) whales could result from demographic features such as historical abundance and population structure, but could also result at least in part from the complicated features of control region evolution mentioned above. Assessing whether the high observed genetic diversity in baleen whales is caused by unusual features of molecular evolution in the control region requires both quantification of rate heterogeneity in the control region and information from additional genetic regions, such as other mitochondrial genes. However, despite the importance of control region data in marine mammal research, no studies to date have quantified and compared rate heterogeneity and other features of molecular evolution in the control region across whale species, and studies on mitochondrial regions other than the control region have been rare in cetaceans.

Coding genes in the mitochondrial genome such as cytochrome *b* may have fewer problems of rate heterogeneity overall due to functional constraints, though this claim has not previously been tested in cetaceans. Because the control region sequence and cytochrome *b* gene are linked on the same nonrecombining mitochondrial genome and thus have the same underlying genealogies, a comparison of substitution patterns between gene regions among individuals may help determine more precisely how the control region evolves, increasing the accuracy of intraspecific estimations. Another possibility is using neutral variation in cytochrome *b* sequences with different numbers of substitutions to define shorter intervals of time in which to measure control region evolution. Thus, cytochrome *b* data could potentially be used to address some of the issues that make genealogical studies of control region data problematic in a broad range of taxa. There are reasons to approach this idea with caution, however. Both control region and cytochrome *b* accumulate substitutions via a stochastic process (Gillespie 2004), and therefore using stochastic cytochrome *b* changes on the temporal framework to evaluate control region evolution may add so much variance to the analysis that firm conclusions are impossible.

The goal of this study is to compare the rate and pattern of nucleotide substitution in the control region and cytochrome *b* regions of the same mitochondrial genomes in order to understand whether high variability at the control region in whales is typical of other markers in the mitochondrial genome or whether it results from gene-specific features of molecular evolution. We utilized and augmented previously collected datasets from three baleen whale species: gray whales (*Eschrichtius robustus*), humpback whales (*Megaptera novaeangliae*), and Antarctic minke whales (*Balaenoptera bonaerensis*). These species differ in their current abundance, degree of population structure, and whaling history. Gray and humpback whales have been commercially exploited since the 19th century and populations were heavily reduced by whaling (e.g., Henderson 1984), but Antarctic minke whales were not targeted by commercial whalers until 1971 (Laws 1977) and remain highly abundant. Humpback whales show extensive population structure across and sometimes within ocean basins (Baker et al. 1993). No strong genetic differentiation has been found within eastern Pacific gray whales (Swartz et al. 2006), though the lack of strong structure might reflect whaling history (Alter et al. 2008). Population structure within Antarctic minke whales is largely unknown, but recent studies have indicated some differentiation between whales in the Pacific and those in the Indian Ocean sectors of Antarctic waters (Pastene 2006). Specifically, we sought (1) to test whether cytochrome *b* data in these three whale species exhibit similar levels of polymorphism to the control region, after functional constraints on cytochrome *b* are taken into account, and (2) to determine whether rate heterogeneity and other patterns of nucleotide evolution differ between the control region and cytochrome *b* in whales. We explore a new method for estimating evolutionary rate in the control region by comparing control region data across individuals with identical cytochrome *b* haplotypes. Finally, we investigate whether differences in polymorphism patterns are observed among the three whale species and explore the evolutionary and demographic forces that could potentially explain these differences.

Materials and Methods

Obtaining DNA Sequence Data

To compare cytochrome *b* and control region data from the same individuals, we utilized or augmented four previously collected datasets and added additional cytochrome *b* datasets for gray and humpback whales. A complete list of the species datasets used in this study, accession numbers, and sequence lengths is reported in Table 1. To the best of our knowledge, these datasets currently represent the only intraspecific data for both control region and cytochrome *b* available for baleen whales.

Control region data for 120 gray whales were obtained from NCBI (LeDuc et al. 2002), in addition to 10 individuals amplified and sequenced as part of this study. Primers for gray whale control region amplification were 5′-TACC AAATGTAT GAAACCTCAG-3′ (Rosel et al. 1995) and 5′-CCTCCCTAAGACTCAAGGAAG-3′ (LeDuc et al. 2002). Amplification conditions and cycling were as follows: denaturation for 2 min at 94°C, followed by 30 cycles of denaturation at 95°C for 30 s, annealing at 56°C for 30 s, and a 1.5-min extension at 72°C. Control region data for humpback whales were obtained from C. S. Baker (Baker and Medrano-Gonzalez 2002), with the exception of five individuals sequenced for this study using the same primers and conditions described for gray whales above. Both control region and cytochrome *b* data for Antarctic minke whales were also provided by C. S. Baker (unpublished data). All Antarctic minke whale samples were obtained in Japanese markets from 1991 to 1993. Because of this unusual source of samples, some chance exists that individuals may have been sampled more than once. Few identified pairs of individuals were identical at both cytochrome *b* and control region (29 pairs of 2485 possible pairings, falling into nine different haplotypes), suggesting that resampling was minimal. We discuss the potential impacts of low-level resampling under Results, below.

We amplified and sequenced the cytochrome *b* gene from a subset of previously collected and extracted samples of gray and humpback whales for which control region data were available. Samples from individual gray whales were taken between 1991 and 2004 from the eastern North Pacific population via biopsy and necropsy sampling (see LeDuc et al. 2002, for additional sampling information), which is thought to be panmictic (Swartz et al. 2006). DNA extraction of seven additional eastern North Pacific gray whale muscle and blubber tissues was performed using a QIAamp kit (QIAGEN, Valencia, CA). Samples of humpbacks were taken between 1989 and 1992 from both North Atlantic and Pacific populations (see Baker et al. 1993, for additional sampling information). Extracted DNA was amplified using cytochrome *b* primers with the following sequence: cytb-F, 5′-CCTCATGAT GAAACTTCGGTTCCC-3′ and cytb-R, 5′-AAGAGGAA GTAGAGGATGGATGCG-3′ (Arnason and Gullberg 1994). Amplification conditions were as follows: denaturation for 2 min at 94°C, followed by 30 cycles of denaturation at 95°C for 30 s, annealing at 65°C for 30 s, and a 1.5-min extension at 72°C. Sequencing reactions were performed with the primers given above and two internal primers: 5′-ATATCATTCTGAGGCGCAACCG TCA-3′ and 5′-CCCAGATTCATTCGACTAGGGTAG-3′ (designed as part of this study). Amplified products were sequenced in both directions on an automated sequencer (ABI 3100;Applied Biosystems). Sequence data were edited and aligned in Sequencher 4.5 (Gene Codes Corp).

Comparing Genetic Diversity at Cytochrome *b* and Control Region

We measured standard genetic diversity indices in order to compare polymorphism between cytochrome *b* and control

Table 1 Species and accession numbers for unique haplotypes in each dataset, in addition to geographic origin of samples (GO), number of samples (*N*), and length of sequence used (bp)

Organism/marker	GO	*N*	bp	Source(s)	GenBank accession nos.
Gray whale					
Cytochrome *b*	NEP	42	852	This study	EF165336–EF165341
Control region		130	523	This study; LeDuc et al. (2002)	AF326789–AF326824
Minke whale					
Cytochrome *b*	ANT	71	475	C. S. Baker	Available from C. S. Baker
Control region		71	494	C. S. Baker	Available from C. S. Baker
Humpback whale					
Cytochrome *b*	NA/NP	36	751	This study	EF165342–EF165348
Control region		67	289	This study; Baker et al. (1993)	AF068067–AF068114

Note: *NEP* Northeast Pacific; *ANT* Antarctica; *NA/NP* North Atlantic, North Pacific. Number of control region samples includes previously produced sequences in addition to those produced as part of this study (gray whale, $N = 10$; humpback whale, $N = 5$)

region markers for each species. Values of Watterson's Θ (Watterson 1975, Eq. 1.4a) were obtained using the maximum-likelihood coalescent-based program LAMARC, version 2.0 (Kuhner et al. 2005). In order to ensure consistency of results, these datasets were run through LAMARC five times each using different starting parameters, and the search strategy was modified (number of replicates increased to five). Further detail on parameters used in LAMARC runs is given by Alter et al. (2007). Haplotype diversity for both cytochrome *b* and control region was obtained using DnaSP v 4.0 (Rozas et al. 2003).

To determine whether values of Θ differ significantly between cytochrome *b* and the control region, we compared probability distributions of Θ (generated using LAMARC). We first scaled all Θ values for cytochrome *b* and the control region by their appropriate (phylogenetically derived) mutation rates in order to account for functional constraints on evolutionary rate (resulting in units of *Ne*), since any relative constraints due to function should be reflected in substitution rates. The evolutionary rate of different segments of the mitochondrial genome in whales and other mammals has previously been estimated by comparing genetic distance between species at a given marker against divergence dates for those species (Pesole et al. 1999; Roman and Palumbi 2003). Phylogenetic analyses indicate that the control region is evolving at a rate of roughly 1.5–2% per million years (Myr) in whales (Pesole et al. 1999; Roman and Palumbi 2003), and cytochrome *b* is evolving at a rate of about 0.4% per Myr (Alter et al. 2007). Therefore, we scaled Θ values using rates of 0.4% per Myr per bp for cytochrome *b* and a rate of 2% per Myr per bp for the control region (Roman and Palumbi 2003). We then tested whether rate-adjusted cytochrome *b* and control region Θ values were significantly different within each species, using a two-tailed Wilcoxon rank-sum test.

Determining the Best-Fit Models of Nucleotide Substitution

We sought to determine whether rate heterogeneity and other features of molecular evolution differ between cytochrome *b* and the control region in whales. To determine the best-fit model of nucleotide substitution for each marker in the three whale species, we used a nested model selection framework (Posada and Crandall 1998). First, a neighbor-joining tree was built for each dataset using PAUP* (Swofford 2003). Likelihood scores were then calculated for each dataset using the neighbor-joining tree for 56 evolutionary models. Three classes of parameters varied in the models: (1) nucleotide base frequencies, (2) substitution rates between bases, and (3) rate variation among sites. Likelihood scores of progressively complex models were calculated and compared using the hierarchical likelihood ratio test (hLRT) implemented in MODELTEST (Posada and Crandall 1998). The "best-fit" model of DNA substitution (e.g., the model for which the likelihood score is significantly higher than that of the next more complex model), as well as maximum likelihood estimates for model parameters including ti/tv (transition/transversion ratio) and α (shape parameter of the gamma distribution), was estimated for each dataset.

Assessing Deviation from Equilibrium Conditions

Differences in patterns of polymorphism between species may be the result of evolutionary or demographic forces, some of which leave characteristic signatures of deviation from expected patterns given selective neutrality and demographic equilibrium. To assess overall signatures of deviation from equilibrium conditions (such as selection or demographic expansion), we calculated basic statistics, including Tajima's *D*, (Tajima 1989, Eq. 38), Fu and Li's *D** (Fu and Li 1993, Eq. 32), and Fu and Li's *F** statistic (Fu and Li 1993), and assessed whether they were significantly different from neutral expectations. Hitchhiking and selective sweeps are common in animal mitochondrial genomes, but the pattern of polymorphism resulting from selective sweeps is also produced by population expansion. However, Fay and Wu (2000) found that hitchhiking (as opposed to expansion) leaves a unique pattern, namely, an excess of derived variants at high frequency. Thus, in order to detect a signature of hitchhiking specifically, we applied Fay and Wu's *H* statistic (Fay and Wu 2000) to the three intraspecific datasets, using *Balaenoptera acutorostrata* (northern minke whale) as an outgroup. The significance of each test statistic was determined by comparing the calculated statistics to a distribution of values generated using 100,000 neutral coalescent simulations.

For cytochrome *b*, we conducted several additional analyses to detect the pattern of selection operating on this gene in each whale population. First, we measured Ka/Ks in each species and for an alignment of all baleen whales. Second, we conducted a McDonald and Kreitman (1991) test between each of the three whale datasets. The McDonald–Kreitman test compares the ratio of synonymous-to-nonsynonymous polymorphisms and divergent sites, under the neutral hypothesis that the ratio of nonsynonymous-to-synonymous fixed differences between species should be the same as this ratio within species. All tests were conducted using DnaSP v.4.0.

Analysis of Site-Specific Variation Within and Between Baleen Whale Species at the Control Region

Hypervariable sites or "hotspots" may cause inflated diversity estimates and rate underestimation in baleen

whale control regions. One signature of hotspots is a correlation in site variability both between and within species (Galtier et al. 2006). To better understand the degree to which hypervariable sites observed in intraspecific control region data are shared across and within whale species, we compared site-specific rate variation between intraspecific control region data (gray, humpback, and minke) and control region data for all baleen whales (data from NCBI). Alignments for interspecific data were performed using CLUSTAL (Higgins et al. 1991). In order to determine whether site-specific rate variation was correlated across whale species, we implemented a maximum likelihood approach to estimate relative rates at individual sites for the gray, humpback, minke, and baleen whale alignments. First, we obtained neighbor-joining trees in PAUP* for each dataset. These trees were used in the BASEML program implemented in PAML (Yang 2003) to estimate relative rates across sites, using a model allowing rate heterogeneity with 20 possible rate categories. Site-specific correlation between datasets was assessed using the Pearson product-moment correlation coefficient. Because of the nonnormality of the data, Spearman's ρ was used to test whether correlation coefficients were significantly different from zero.

Estimating Control Region Substitution Rate Using Individuals Identical at Cytochrome *b*

One method of estimating substitution rate in a rapidly evolving genetic marker is to use a linked marker with a slower rate to define the interval over which changes must have occurred. We used data from the cytochrome *b* gene in three species of baleen whales to calculate the relative substitution rate in control region. Because cytochrome *b* is a coding region in which many positions are evolutionarily constrained, we restricted our analysis to silent variation in cytochrome b.

To estimate the silent substitution rate at cytochrome *b* in whales, we obtained published data for the entire cytochrome *b* gene ($n = 1140$ bp) for 10 species of baleen whales from the NCBI database (accession nos. AP006466–AP006475) and aligned the data in Sequencher 4.5 (Gene Codes Corp.). We then performed a pairwise distance calculation for these data in MEGA version 3.1 (Kumar et al. 2004) under the appropriate mutational model as determined by MODELTEST. Synonymous pairwise distances were calculated using the method of Li et al. (1985), including only synonymous sites. Substitution rate was estimated by regressing the silent pairwise distances against estimated divergence dates for each species pair (Sasaki et al. 2005) and calculating the rate as half the slope of the relationship. The waiting time (w) for each silent substitution in cytochrome *b* is $(\mu \times n)^{-1}$ where $\mu =$ substitutions per base pair per year and $n =$ the number of fourfold degenerate sites plus one-thir of twofold degenerate sites (Li et al. 1985). This estimate will give an upper bound for the waiting time to the next mutation in cytochrome *b* because transition bias is known to occur in mammalian mitochondrial genes and all transitions are silent at twofold degenerate sites. The mutation bias for transitions in mtDNA ($\mu s/\mu v$) should be roughly equal to the transition-transversion ratio at fourfold degenerate sites (Ina 1995; Xia et al. 1996). Using the proportion of transitions at fourfold degenerate sites, we estimated the ratio of mutational bias toward transitions in baleen whales as 4.56 (82% of mutations are transitions). This estimate of bias in mutation implies that $\sim$82% of all mutations at twofold degenerate sites are silent, so we also calculated n as the number of fourfold degenerate sites in addition to 82% of twofold degenerate sites in cytochrome *b* and used a modified Nei-Gojobori model (Nei and Kumar 2000) to calculate synonymous pairwise distances.

From the intraspecific datasets for each of the three whale species as described above, we used PAUP* to construct pairwise genetic distance matrices to identify those individuals that were identical at cytochrome *b*. We then calculated the number of pairwise changes in the control region for those same individuals using a neighbor-joining tree (assuming mutational models as specified in Table 3). We tested for significant differences in the mean number of changes between species using a Kruskal-Wallis test implemented in JMP (SAS Institute). We assume that, on average, two individuals will diverge at cytochrome *b* at half the waiting time until the next cytochrome *b* mutation (though this parameter may vary depending on the distribution of divergences; see Discussion). Therefore, control region substitution rate per base pair was estimated as $x/((1/2) \times w \times n)$, where x is the mean number of differences, w is the waiting time to the next mutation in cytochrome *b* in years, and n is the number of base pairs in the section of the control region used.

Results

Comparing Genetic Diversity at Cytochrome *b* and the Control Region

Diversity indexes (Table 2) are higher for control region than cytochrome *b* as expected, due to selective forces acting on cytochrome *b*, but the ratio of intraspecific diversity differed from the expected value based on the ratio of control region-to-cytochrome *b* evolutionary rate (3.75–5). Mean values of Θ differ between the control region and cytochrome *b* within a single species by roughly 3.4 to 15.7 times but are more consistent among species at

Table 2 Haplotype diversity (Hd) and DNA polymorphism for each dataset estimated using LAMARC ($\Theta_{(L)}$)

Organism/marker	Hd	$\theta_{(L)}$	*D*	FL-D*	FL-F*	FW-H
Gray whale						
Cytochrome *b*	0.679	0.0028	−0.8451	0.8053	0.3192	**−7.9380**
Control region	0.95	0.0330	0.0183	0.9415	0.7296	−9.3000
Minke whale						
Cytochrome *b*	0.687	0.0274	**−1.7626**	**−3.7248**	**−3.6054**	**−6.6213**
Control region	0.956	0.0929	−1.6176	−0.6862	−1.2447	**−10.2108**
Humpback whale						
Cytochrome *b*	0.756	0.0022	−0.3645	−1.2026	−1.1062	−2.3546
Control region	0.889	0.0345	0.0184	−0.0002	0.0700	−6.4250

Note: *D* Tajima's D; *FL-D** Fu and Li's D*; *FL-F** Fu and Li's F*; *FW-H* Fay and Wu's H. Boldface values denote significance at $p < 0.05$

each locus. The ratio between Θ (control region) and Θ (cytochrome b) is highest for humpback whales, where control region diversity exceeds cytochrome *b* diversity by 15.7 times. By contrast, this ratio (3.4) is slightly lower than expected for minke whales.

Statistical comparisons between average Θ values for cytochrome *b* and control region for each whale species, after functional constraints were accounted for by scaling by evolutionary rate, show significant differences based on a two-tailed Wilcoxon rank-sum test ($p < 0.0001$) in all cases (Fig. 1). However, the direction of the difference varied between species. Whereas for gray and humpback whales, the control region showed significantly greater diversity than cytochrome *b*, in minke whales cytochrome *b* was found to have significantly greater diversity. We also compared Θ values for a given marker between species and found all comparisons to be significantly different after adjustment for multiple comparisons (Bonferroni-corrected *p*-value of 0.017).

Selecting Models of Nucleotide Substitution

The best-fit model of evolution differed between both markers and species (Table 3). For cytochrome *b*, gray and humpback whale datasets were best fit by the HKY model (Hasegawa et al. 1985), which allows unequal nucleotide frequencies and unequal rates for transitions and transversions, with no rate variation across sites. Minke whale cytochrome *b* evolution was also best explained by the HKY model, but a model incorporating rate variation across sites was found to be a better fit ($\alpha = 0.0018$, $p = 0.006$). This may result from stronger negative selection on cytochrome *b* in minkes, such that evolution occurs at a much more rapid pace at silent sites. Humpback whales showed no transversions for the cytochrome *b* gene, resulting in a ti/tv value of infinity.

At the control region, the best-fit model of evolution differed substantially across the three whale species. Gray and humpback whale control region data are best explained by the HKY model ($p < 10^{-6}$), although a hierarchical likelihood ratio test showed that for gray whales, the TrN model (Tamura and Nei 1993; unequal nucleotide frequencies, rate differences between two types of transitions allowed) is only slightly inferior to the HKY model ($p = 0.06$). Gray and humpback whale control region data show significant rate heterogeneity ($p < 10^{-5}$). Estimates for the shape parameter for the gamma distribution (α) were < 1 (0.0018 to 0.823), consistent with the presence of numerous hypervariable sites (Excoffier and Yang 1999). In contrast, minke whale control region data are best explained by the TIM model (Rodriquez et al. 1990; unequal nucleotide frequencies and five substitution types), and no rate heterogeneity was detected in this dataset. The transition/transversion ratio (ti/tv) in the control region in the three whale species ranged between 7.5 and 47. These numbers roughly match transition-transversion ratios in primate mitochondrial genes, which range from 12 to 37 (Meyer et al. 1999).

Assessing Deviation from Equilibrium Conditions

Significantly negative values of Tajima's *D*, Fu and Li's *D**, and Fu and Li's *F** were observed for minke whale cytochrome *b*, indicating a significant excess of rare polymorphisms, but no significant values were observed for gray or humpback whale cytochrome *b* (Table 2). No significant values were observed for these statistics at the control region for any species. Fay and Wu's *H*, which indicates evidence of hitchhiking or selective sweeps, was highly negative and significant for minke whale control region data and moderately but still significantly negative for cytochrome *b*. Humpback whales showed negative but nonsignificant values for this statistic at both markers, and gray whale results were not significant for the control region but moderately negative and significant for cytochrome *b*.

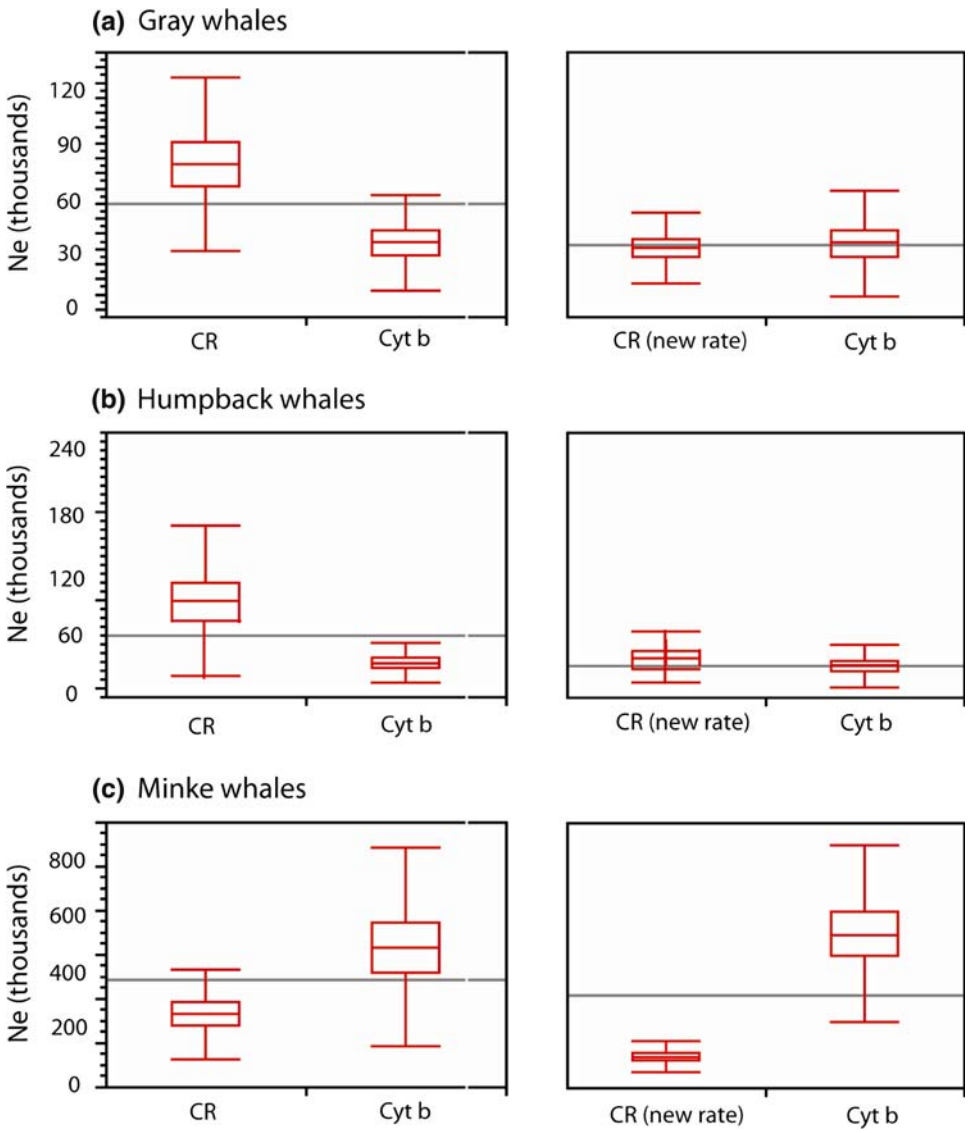

Fig. 1 Comparisons between *Ne* values ($\Theta\mu$ gen^{-1}) on the *y*-axis for each marker. CR, control region (using a rate of 2% per my); CR (new rate), control region (using cytochrome *b* calibrated species-specific rates from 5% to 5.4% per Myr). Box plots show median values (middle line), upper and lower quartiles (box edges), and 95% tails of the distribution (whiskers). Values for species-specific generation times were obtained from Roman and Palumbi (2003) and Alter et al. (2007). Using a cytochrome *b* calibrated rate for the control region brings *Ne* estimates for the two markers into better agreement for gray and humpback whales but not for minke whales

Table 3 Results of hierarchical likelihood ratio tests (best-fit model of nucleotide substitution) for each species and marker

Marker/organism	Model	*n*	α	Ti/Tv
Cytochrome *b*				
Minke whale	HKY	71	0.0018	30.87
Gray whale	HKY	42	Equal rates	9.03
Humpback whale	HKY	36	Equal rates	∞
Control region				
Minke whale	TIM	71	Equal rates[a]	7.45
Gray whale	HKY	120	0.823	47[b]
Humpback wale	HKY	67	0.1410	8.05

Note: HKY (Hasegawa et al. 1985); TIM (Rodriguez et al. 1990). α = gamma shape parameter. (∞) No transversions observed in dataset

[a] Although a model of rate heterogeneity was originally selected for this dataset, α was estimated as ∞ indicating equal rates

[b] Empirically observed ratio (calculated directly from gray whale dataset)

For cytochrome *b*, Ka/Ks values were 0.229, 0.172, and 0.105 for gray, humpback, and minke whales, respectively. Of the three McDonald–Kreitman tests (humpback-minke, gray-minke, and humpback-gray), the humpback-gray comparison showed a significant excess of nonsynonymous polymorphisms using Fisher's exact test and a Bonferroni-corrected *p*-value of 0.016. All tests showed an excess of nonsynonymous polymorphisms (Table 4), but because the three tests are not independent, they cannot be combined to produce an overall statistical result.

Analysis of Site-Specific Rates Within and Between Baleen Whale Species at the Control Region

We analyzed differences in site-specific rates in control region, for each of the three whale species and for an alignment of all baleen whales. Relative rates and

Table 4 Results of McDonald–Kreitman test among humpback, gray, and minke whales; significance of values tested using Fisher's exact test

Divergent		Polymorphic
Gray-humpback		
Synonymous	45	9
Nonsynonymous	5	6
p-value	0.014	
Gray-minke		
Synonymous	12	11
Nonsynonymous	1	4
p-value	NS	
Minke-humpback		
Synonymous	8	7
Nonsynonymous	1	2
p-value	NS	

Note: *NS* not significant

correlation coefficients are shown in Fig. 2 and Table 5. Elevated rates of mutation were moderately correlated across all comparisons, with correlation coefficients between the baleen alignment and gray, humpback, and minke of 0.429, 0.388, and 0.341, respectively. All correlation coefficients were found to be significantly greater than zero based on Spearman's ρ ($p < 0.0001$).

Estimating Control Region Rate Using Individuals Identical at Cytochrome b

Under the Li et al. (1985) model, the silent substitution rate for cytochrome *b* in baleen whales was estimated as 1% per Myr, with a resulting waiting time to the next mutation of 4.26×10^5 years for the entire locus (Fig. 3). This silent rate is thus roughly 2.5 times faster than the substitution rate for the entire cytochrome *b* gene in whales (0.4% per Myr; Alter et al. 2007). If the Ina (1995) model is used, the waiting time increases slightly, to 4.89×10^5 years.

The distribution of pairwise substitutions in the control region for individuals identical at cytochrome *b* for each species (Fig. 4) shows high mean values and high variability (Table 6), but data were distributed nonnormally in each case. Individual gray whales identical at cytochrome *b* differed by a mean of 6.19 ± 0.67 changes in the control region, though the tail of the frequency distribution is long: 40 pairwise comparisons showed 19 or more control region changes between individuals identical at cytochrome *b* (Fig. 4). The mean number of changes in the control region was 3.2 ± 0.68 for humpback whales and 5.3 ± 0.36 for minke whales. The mean number of changes was found to be significantly different between species using a Kruskal-Wallis test ($p < 0.0001$). Median values were similar to, but slightly lower than, average values.

When the method of Li et al. (1985) (which uses the sum of all fourfold and one-third of twofold degenerate sites) was used to estimate the *w* of cytochrome *b*, estimates of substitution rate in the control region were 5.4% (4.96–6.16%), 5.2% (4.12–6.32%), and 5.0% (4.7–5.39%) per Myr in gray, humpback, and minke whales respectively, more than four times as fast as the silent substitution rate in cytochrome *b* in baleen whales (mean rate, with 95% confidence intervals in parentheses). Rates were slightly lower when *w* was calculated using the method of Ina (1995) (using the sum of all fourfold and 82% of twofold sites). Using this method, substitution rates were 4.8% (4.32–5.36%), 4.6% (3.59–5.5%), and 4.4% (4.0–4.69%) per Myr for gray, humpback, and minke, respectively. Regardless of the method used, all rates are significantly higher than control region substitution rates estimated from divergence dates using the fossil record (e.g., Pesole et al. 1999; Rooney et al. 2001; Roman and Palumbi 2003), which range from 1.5% to 2% per Myr. Although no error estimates are available for earlier rates, confidence intervals for cytochrome *b* calibrated rates do not approach these earlier values.

Effects of Low-Level Resampling of Minke Whale Individuals on Analyses

Because the samples of minke whales used in this study originated in Japanese markets, the possibility exists that a small number of individuals may be represented in the dataset twice. Levels of genetic diversity in minke whales are high and few individuals show the same sequence at both cytochrome *b* and the control region; we estimate that the maximum number of duplicated sequences is 14. To determine the effects of including potential duplicates on the results of the analyses presented above, we repeated those analyses that would be affected with modified datasets in which all 14 potential duplicates were removed. As expected, diversity values were found to be slightly higher (by ~8–9%) when duplicates were excluded (Θ [control region; CR] = 0.1013, Θ[cyt-*b*] = 0.0296), but the ratio of diversity between the two markers did not change. Parameters designed to test for signatures of nonequilibrium became slightly more negative (and therefore more significant): for example, values of Tajima's *D* were −1.723 (CR) and −1.849 (cyt-*b*). The best-fit model of nucleotide substitution did not change for either marker when potential duplicates were excluded. Finally, cytochrome *b* calibrated rate estimates were elevated slightly by excluding duplicates: using the method of Li et al. (1985), the evolutionary rate for the minke whale control region was estimated to be 5.3% rather than 5% per Myr.

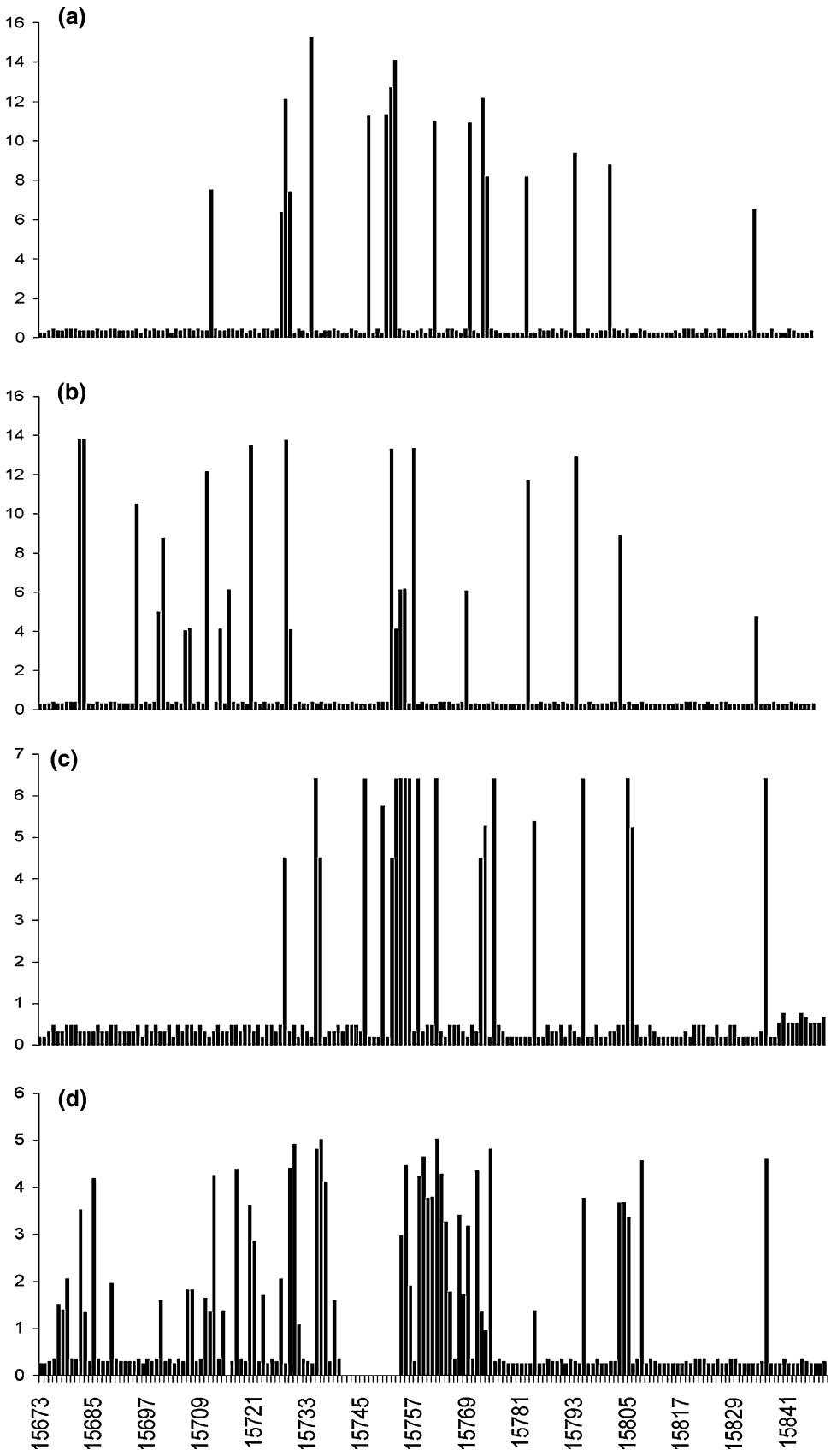

Fig. 2 Relative substitution rates (scaled such that the average rate is 1) along the length of a section of the control region (positions 15673 to 15848) for **a** gray, **b** minke, and **c** humpback whales and **d** an alignment of all baleen whales (a deletion in some whale species between 15741 and 15753 prevents relative rate estimation at those sites)

Table 5 Correlation coefficients between site-specific rates

Dataset 1	Dataset 2	Coefficient
All baleen whales	Gray	0.429514856
All baleen whales	Humpback	0.387720351
All baleen whales	Minke	0.341540674
Gray	Minke	0.250046113
Gray	Humpback	0.418812374
Humpback	Minke	0.34759775

Discussion

Intraspecific patterns of diversity differed significantly between cytochrome *b* and control region datasets in whales, and the way these two regions compared differed from one whale species to another. Many of the observed patterns were roughly similar between gray and humpback whales but differed strongly in Antarctic minke whales (summarized in Table 7). The control region evolves much more rapidly than the rest of the mitochondrial genome in many vertebrates, with estimates of 5- to 10-fold faster rates in primates (Aquadro and Greenberg 1983), and this pattern appears to hold true for baleen whales as well. Given the complete linkage between these markers, and provided that interspecific rates apply to polymorphism data, we expect that levels of genetic diversity in a given population should be statistically indistinguishable between the two markers after adjusting for locus-specific rates of evolution (and thus for major selective constraints). However, rate-adjusted intraspecific diversity in the control region was higher than expected for humpback and gray whales, and a signal of rate heterogeneity was found in the control region, whereas no rate heterogeneity was detected in cytochrome *b*. In contrast, rate-adjusted intraspecific diversity in the control region was lower than expected for Antarctic minke whales and rate heterogeneity was only detected in cytochrome *b* (Table 3). Patterns of deviation from equilibrium conditions also differed between species, with minke whales showing an excess of rare haplotypes, indicative of population expansion or purifying selection.

Several types of evolutionary forces may have caused these disparate patterns across marker and species, including gene-specific evolutionary features (e.g., hypervariability, secondary structure) and forces affecting all markers (e.g., demography). While it can be difficult to distinguish these processes given control region data alone, adding polymorphism data from cytochrome *b* allows further insight into the evolutionary mechanisms that created the observed patterns.

The Influence of Demography and Gene-Specific Processes on Polymorphism Patterns

Demographic forces, including population size, growth or decline, and substructure, exert a strong influence over the amount and distribution of polymorphism and could potentially explain several of the observed patterns. For example, the differences in overall diversity levels between whale species at both markers are most easily explained by differences in long-term population sizes. Indeed, the higher observed diversity in Antarctic minke whales compared to gray and humpback whales accords with the larger size of this population, both now and probably across history.

However, the relationship between markers within each species is more difficult to explain with demographic

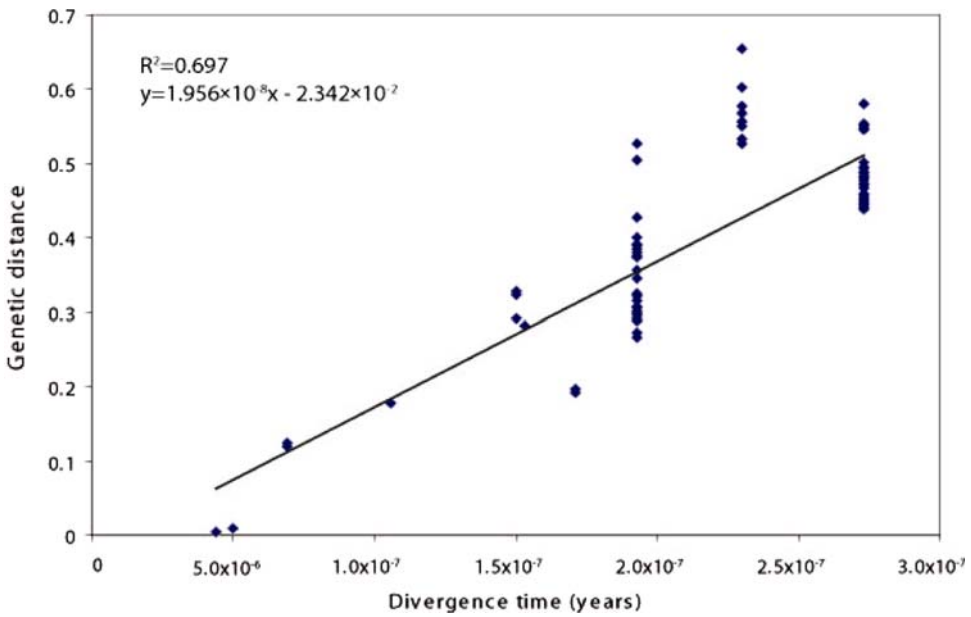

Fig. 3 Silent substitution rate for cytochrome *b* in baleen whales using the model of Li et al. (1985). Divergence times are based on estimates from Sasaki et al. (2005)

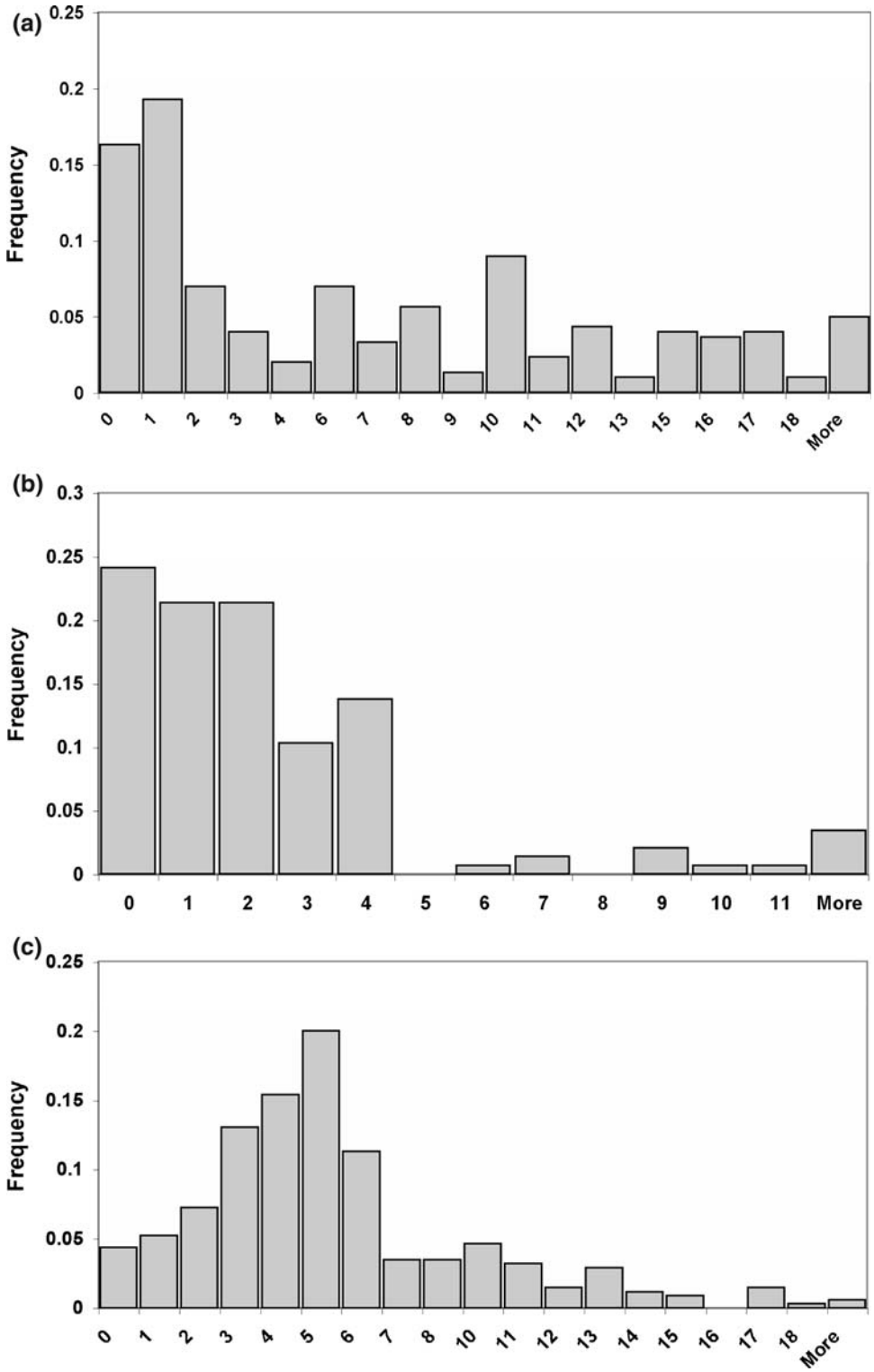

Fig. 4 Frequency distribution for number of changes in control region for pairwise individuals identical at cytochrome *b* for **a** gray, **b** humpback, and **c** minke whales

processes alone. Because demographic forces should affect genetic diversity at linked genes in the same direction (either increasing or decreasing diversity), we can utilize comparisons between patterns in cytochrome *b* and the control region to determine whether patterns result from gene-specific molecular processes rather than demography. For instance, high diversity in the control region of humpback whales may be explained solely by population divergence and/or long-term effective size, but if this were the case, we would also expect to see relatively high diversity in cytochrome *b* after selection is accounted for. In other words, while in the absence of data from another gene, we might

Table 6 Descriptive statistics for number of changes in control region in three whale species for pairwise individuals identical at cytochrome *b*

	Minke	Humpback	Gray
Mean	5.31	3.21	6.19
SE	0.19	0.35	0.34
Median	5.00	2.00	5.00

Table 7 Summary of intraspecific patterns of diversity that differed between gray/humpback and minke whales

	Gray & humpback	Minke
Ratio of CR:cyt-*b* diversity	Higher than expected	Lower than expected
Tajima's *D* (cyt-*b*)	Nonsignificant	Significant, negative
Fu and Li's *D* (cyt-*b*)	Nonsignificant	Significant, negative
Fu and Li's *F* (cyt-*b*)	Nonsignificant	Significant, negative
CR rate variation detected	Yes	No
Cyt-*b* rate variation detected	No	Yes
Ka/Ks	Higher	Lowest
Site-specific rate correlation	Higher	Lowest

Note: *CR* control region

attribute all of the diversity at the control region to demographic causes, when these data are put within the context of cytochrome *b* diversity, gene-specific molecular forces are needed to fully explain the observed patterns. In particular, hypervariability and saturation in the control region could result in the underestimation of the true mutation rate in these lineages, and as a result, the observed ratio between control region and cytochrome *b* diversity could be higher than expected. The stochastic nature of mtDNA evolution (e.g., Tajima 1983) must also be considered as a potential explanation for these patterns. However, the saturation hypothesis is supported by strong signatures of hypervariability in the gray and humpback control region, as indicated by the best-fit model of nucleotide evolution and correlation between hotspots across whale species. In addition, the ratio of diversity in cytochrome *b* and control region in humpback and gray whales is reconciled with the ratio of rates when the higher control region rate estimated using the cytochrome *b* clock is used (Fig. 1).

In the case of Antarctic minke whales, the observed patterns (lower-than-expected diversity in the control region relative to cytochrome *b*) require an explanation more complex than rate heterogeneity and poor phylogenetic calibrations. These patterns might be caused by a relaxation of selection on cytochrome *b* and/or a slowdown in evolutionary rate in the control region. However, several lines of evidence suggest that neither of these scenarios is likely. Purifying selection in minke whale cytochrome *b* appears to be stronger than in humpback or gray whales; despite the high diversity in minke whale cytochrome *b*, there is a paucity of nonsynonymous changes relative to the two other whale species, as evidenced by its much lower Ka/Ks value. Likewise, there is no direct evidence for a slowdown in control region rate; diversity values are high and no additional selective constraints on secondary structure are apparent. We hypothesize that this pattern may instead result from the retention of high cytochrome *b* diversity due to a large, demographically stable or expanding minke whale population, coupled with saturation at the control region. If the last selective sweep in the minke whale mitochondrial genome happened a longer time ago than for gray or humpback whales, higher diversity could build in both cytochrome *b* and the control region, but genetic diversity may have built to the maximum level at hotspots in the control region, masking the accumulation of additional diversity. After hypervariable sites became saturated, diversity may have begun to build slowly at additional sites across the control region, including sites that are slow to mutate in other whale populations, eliminating the expected signal of hypervariability. This is perhaps reflected in the site-specific comparisons between Antarctic minke whales and other whales: Fig. 2 shows that in minke whales, mutations are more evenly spread across the control region rather than being concentrated in areas defined as hotspots in other whales, and of the three species, minke whales show the lowest spatial correlation of rates across this marker. The relatively high number of transversions in the control region (Table 3) is further evidence that a long interval has likely occurred since the last selective sweep, since transversions are expected to accumulate slowly in mammalian mtDNAs.

We observed highly negative values of Tajima's *D* and Fu and Li's *D* and *F* in minke whale mitochondrial genes, consistent with a signal of population expansion over evolutionary time, a scenario hypothesized to have occurred in minke whales during the Pliocene (Pastene et al. 2007). However, other features of the minke whale polymorphism data remain difficult to explain fully. In particular, contrary to what we might expect if saturation at the control region has occurred, the relationship between control region and cytochrome *b* distances remains linear even at high levels of divergence, and the maximum genetic distance between pairs of individual minke whales is low relative to the other two species (Fig. 3). Indeed, the maximum genetic distance within gray whales (>9%) is roughly equal to the distance between Antarctic minke whales (*B. bonaerensis*) and its sister species in the Atlantic, *B. acutorostrata* (Pastene et al. 2007). One possibility is that the high values of maximum genetic distance observed in other whale species are reflective of deep splits

due to interoceanic population structure, whereas Antarctic minke whales may have always been well-mixed and therefore do not show the same deep genetic divergences. In addition, mutation hits at sites outside of the usual hypervariable sites may be responsible for the linearity of the relationship between the control region and cytochrome *b* even at large cytochrome *b* distances. The fact that the best-fit model of evolution in minke whales does not require rate heterogeneity may also result from this extra diversity outside of hotspots.

If the last selective sweep in fact occurred longer ago in minke whales than in grays and humpback whales, this difference in intervals might be due to chance, or could possibly result from differences in life histories between species. The effect of matrilineally transmitted cultural traits on the speed of selective sweeps has been proposed as a hypothesis to explain differences in mitochondrial diversity among some cetacean species (Whitehead 1998). Both humpback and gray whales are thought to exhibit strong fidelity to breeding and possibly feeding grounds, and calves are thought to learn migration routes from their mothers. In contrast, such patterns have not been observed for Antarctic minke whales (though behavioral ecology in this species is poorly known compared to that of gray and humpback whales). If following particular migratory routes (or adhering to other maternally inherited customs or behaviors) results in selective advantages for particular maternal families and mitochondrial haplotypes, we might expect that selective sweeps would occur more frequently for humpback and gray as opposed to minke whales. This hypothesis remains speculative at present, but could be evaluated through the collection and comparison of nuclear DNA diversity from these and additional cetacean species.

Estimating Control Region Rate Using Cytochrome *b* Data

The problem of estimating a rate of substitution for genetic loci that display rate heterogeneity has been well studied in human datasets (e.g., Meyer et al. 1999; Excoffier and Yang 1999). Human genealogical datasets over a few generations suggest control region rates of 32% (Sigurdardottir et al. 2000) and up to 250% per Myr (Parsons et al. 1997), and calibrations based on ancient DNA in humans have yielded rates of 33–44% per Myr (Kemp et al. 2007), but these rates are far too high to bring mitochondrial and nuclear data into agreement about genealogical history. Yet such discrepancies may indicate that rate heterogeneity and other features of the control region in humans rapidly generate homoplasy that complicates estimates of rates. Recent studies on time dependence of molecular rates indicate that substitution rates are generally underestimated in a phylogenetic comparison of old divergences, and that comparing more recently diverged taxa provides better estimates of instantaneous substitution rates (Ho et al. 2005; Ho and Larson 2006; Burridge et al. 2008). For whales, fossil dates for divergences are old because the fossil record of modern species is poor. Thus, fossil calibrations depend on the divergence of genera or families more than 10 Myr ago.

To address this problem, we used a new method to estimate rate of nucleotide substitution in the control region, using individuals with identical cytochrome *b* haplotypes. The advantage of this method is that it provides a shortened time interval over which accumulating changes in the control region can be measured and, therefore, should reduce the problem of homoplasy. When the waiting time until the next mutation for cytochrome *b* was calculated using the method of Li et al. (1985), control region rates were on average 2.6 times higher than the previously cited rate of 2% per Myr. If the method of Ina (1995) is used, the waiting time to the next mutation is increased and the rate of control region substitution is slightly decreased. A higher control region rate could explain much of the difference in genetic diversity that we observed between cytochrome *b* and the control region in gray and humpback whales (Fig. 1).

In this analysis, we have assumed that the average divergence time at cytochrome *b* for any two individuals will be half the waiting time to the next mutation as calculated from the fossil-calibrated silent rate. However, this assumption depends on the distribution of divergences in each species, which in turn depends on the demographic history and also has a large stochastic element. In populations that have experienced recent bottlenecks and are now expanding, for example, the time separating individuals with the same haplotypes may be much shorter than in a stable population. In this context, we would expect that gray or humpback individuals that are identical at cytochrome *b* are separated by, on average, a shorter time interval than are minke whales (resulting in higher estimated rates of control region evolution based on the cytochrome *b* clock). This pattern is evident from control region distributions (Fig. 4), which are skewed toward zero and one change in gray and humpback whales but centered around five changes in minke whales. Data from additional rapidly evolving markers may aid in determining how the distribution of cytochrome *b* divergences differs across populations and species. In the absence of additional datasets, we have reported rates calculated using the mean and the 95% confidence intervals, with the caveat that the distribution of divergence times in cytochrome *b* will influence the mean rate.

This analysis also presumes that the phylogenetically measured silent rate of cytochrome *b* is reliable at the intraspecific level. For gray whales, data from multiple

unlinked introns (Alter et al. 2007) indicate that this is the case: genetic diversity at cytochrome *b* in gray whales matches that of nuclear introns when all markers are scaled by their respective mutation rates and a heritability scalar. Using a control region rate of 5.4% per Myr brings *Ne* estimates calculated from the control region into excellent agreement with those calculated from multiple nuclear introns in gray whales (Alter et al. 2007). In addition, this rate falls within the 95% highest posterior density rate estimates for bowhead whales based on Bayesian analysis of ancient and modern DNA samples (Ho et al. 2007, 2008). However, without an equivalent nuclear dataset for the other whale species, we cannot yet generalize about the reliability of control region rates calculated using a cytochrome *b* clock across all baleen whales.

Conclusions

Patterns of diversity compared between the mitochondrial control region and cytochrome *b* differ substantially among three whale species. Gray and humpback whales display a deficiency of genetic diversity in cytochrome *b* relative to the control region even after selective forces are accounted for, perhaps due to hypervariability at several sites in the control region that results in underestimation of the true mutation rate. Estimates of control region rate for these two species based on individuals with identical cytochrome *b* sequences confirm phylogenetic underestimation of intraspecific control region rates and suggest values on average 2.6 times higher. For Antarctic minke whales, however, cytochrome *b* shows an excess of diversity relative to the control region, and this pattern occurs despite evidence of a higher degree of selective constraint on cytochrome *b* in this species. This may be the result of a large, stable or expanding population in minke whales combined with longer intervals between selective sweeps in the mitochondrial genome, resulting in saturation at the control region.

These results suggest that phylogenetically determined rates and patterns of nucleotide substitution do not always adequately capture patterns of diversity or evolutionary parameters at the intraspecific level for mitochondrial data because of substantial variation across species. Given the ease of data collection and its rich density of variable sites, the control region will no doubt continue to be used frequently in ecological and evolutionary studies. However, these results demonstrate that where the accurate estimation of control region substitution rate is critical to the conclusions drawn in a study, using the phylogenetic rate alone may be inadequate, even if that rate is calibrated using closely related species. Finally, although variation in rate across species can add a layer of uncertainty and complication to phylogeographic studies based on control region, the comprehensive analysis of such variation where it occurs may eventually lead to a better understanding of the evolutionary and demographic forces shaping genetic variation at both inter- and intraspecific levels.

Acknowledgments We thank C. S. Baker, Colm Carraher, and A. Lang for providing sequences used in this analysis. We are grateful to four anonymous reviewers for helpful comments that improved the manuscript.

References

Alter SE, Rynes E, Palumbi SR (2007) DNA evidence for historic population size and past ecosystem impacts of gray whales. Proc Natl Acad Sci USA 104:15162–15167

Alter SE, Flores Ramirez S, Nigenda S, Urbán Ramirez J, Rojas Bracho L, Palumbi SR (2008) Mitochondrial and nuclear genetic variation across calving lagoons in Eastern North Pacific gray whales (*Eschrichtius robustus*). J Hered. doi:10.1093/jhered/esn090

Aquadro CF, Greenberg BD (1983) Human mitochondrial DNA variation and evolution: analysis of nucleotide sequences from seven individuals. Genetics 103:287–312

Arnason U, Gullberg A (1994) Relationship of baleen whales established by cytochrome b gene sequence comparison. Nature 367:726–728

Baker CS, Medrano-Gonzalez L (2002) World-wide distribution and diversity of humpback whale mitochondrial DNA lineages. In: Pfeiffer CJ (ed) Cell and molecular biology of marine mammals. Kreiger, Melbourne, FL, pp 84–99

Baker CS, Perry A, Bannister JL, Weinrich MT, Abernethy RB, Calambokidis J, Lien J, Lambertsen RH, Ramirez JU, Vasquez O (1993) Abundant mitochondrial DNA variation and world-wide population structure in humpback whales. Proc Natl Acad Sci USA 90:8239–8243

Bradley DG, MacHugh DE, Cunningham P, Loftus RT (1996) Mitochondrial diversity and the origins of African and European cattle. Proc Natl Acad Sci USA 93:5131–5133

Burridge CP, Craw D, Fletcher D, Waters JM (2008) Geological dates and molecular rates: fish DNA sheds light on time dependency. Mol Biol Evol 25:624–633

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

Excoffier L, Yang Z (1999) Substitution rate variation among sites in mitochondrial hypervariable region I of humans and chimpanzees. Mol Biol Evol 16:1357–1368

Fay JC, Wu C-I (2000) Hitchhiking under positive Darwinian selection. Genetics 155:1405–1413

Fu YX, Li WH (1993) Statistical tests of neutrality of mutations. Genetics 133:693–709

Galtier N, Enard D, Radondy Y, Bazin E, Belkhir K (2006) Mutation hot spots in mammalian mitochondrial DNA. Genome Res 16:215–222

Gillespie H (2004) Population genetics: a concise guide, 2nd edn. Johns Hopkins University Press, Baltimore, MD

Gillooly JF, Allen AP, West GB, Brown JH (2005) The rate of DNA evolution: effects of body size and temperature on the molecular clock. Proc Natl Acad Sci USA 102:140–145

Hasegawa M, Kishino H, Yano T (1985) Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. J Mol Evol 22:160–174

Henderson DA (1984) Nineteenth century gray whaling: grounds, catches and kills, practices and depletion of the whale population. In: Jones ML, Swartz SL, Leatherwood S (eds) The Gray Whale, Eschrichtius robustus. Academic Press, Orlando, FL, pp 159–186

Higgins DG, Bleasby AJ, Fuchs R (1991) Clustal V: improved software for multiple sequence alignment. CABIOS 8:189–191

Ho SYW, Larson G (2006) Molecular clocks: when times are a-changin. Trends Genet 22:79–83

Ho SYW, Phillips MJ, Cooper A, Drummond AJ (2005) Time dependency of molecular rate estimates and systematic overestimation of recent divergence events. Mol Biol Evol 22:1561–1568

Ho SYW, Kolokotronis S, Allaby RG (2007) Elevated substitution rates estimated from ancient DNA sequences. Biol Lett 3:702–705

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. doi:10.1371/journal.pone.0001615

Hoelzel AR (1994) Genetics and ecology of whales and dolphins. Annu Rev Ecol Syst 25:377–399

Ina Y (1995) New methods for estimating the numbers of synonymous and non-synonymous substitutions. J Mol Evol 40:190–226

Kemp BM, Malhi RS, McDonough J, Bolnick DA, Eshleman JA, Rickards O, Martinez-Lbarga C, Johnson JR, Lorenz JG, Dixon EJ, Fifield TE, Heaton TH, Worl R, Smith DG (2007) Genetic analysis of early Holocene skeletal remains from Alaska and its implications for the settlement of the Americas. Am J Phy Anthropol 132:605–621

Kocher TD, Wilson AC (1991) Sequence evolution of mitochondrial DNA in human and chimpanzees: control region and protein coding region. In: Osawa S, Honjo T (eds) Evolution of life: fossils molecules and culture. Springer, Tokyo, pp 391–413

Kuhner MK, Yamato J, Beerli P, Smith LP, Rynes E, Walkup E, Li C, Sloan J, Colacurcio P, Felsenstein J (2005) LAMARC v 2.0. University of Washington. Available at http://evolution.gs.washington.edu/lamarc.html

Kumar S, Tamura K, Nei M (2004) MEGA3: Integrated software for molecular evolutionary genetics analysis and sequence alignment. Brief Bioinform 5:150–163

Laws RM (1977) The significance of vertebrates in the Antarctic marine ecosystem. In: Llano GA (ed) Adaptations within Antarctic ecosystems. Smithsonian Institution, Washington DC, pp 411–438

LeDuc R, Weller DW, Hyde J, Burdin AM, Rosel PE, Brownell RL Jr, Wursig B, Dizon AE (2002) Genetic differences between western and eastern gray whales (*Eschrichtius robustus*). J Cetac Res Manage 4:1–5

Li W-H, Chu C-I, Luo C-C (1985) A new method for estimating synonymous and nonsynonymous rates of nucleotide substitution considering the relative likelihood of nucleotide and codon changes. Mol Biol Evol 2:150–174

Martin AP, Palumbi SR (1993) Body size, metabolic rate, generation time, and the molecular clock. Proc Natl Acad Sci USA 90:4087–4091

McDonald JH, Kreitman M (1991) Adaptive protein evolution at the Adh locus in Drosophila. Nature 351:652–654

Meyer S, Weiss G, von Haeseler A (1999) Pattern of nucleotide substitution and rate heterogeneity in the hypervariable regions I and II of human mtDNA. Genetics 152:1103–1110

Nei M, Kumar S (2000) Molecular evolution and phylogenetics. Oxford University Press, New York

Parsons TJ et al (1997) A high observed substitution rate in the human mitochondrial DNA control region. Nat Genet 15:363–368

Pastene LA (2006) What do we know about the stock structure of the Antarctic minke whale? A summary of studies and hypotheses. Paper SC/D06/J12. Presented at the 58th meeting of the international whaling commission, St. Kitts and Nevis

Pastene LA, Got M, Kanda N, Zerbini AN, Kerem D 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

Pesole G, Gissi C, DeChirico A, Saccone C (1999) Nucleotide substitution rate of mammalian mitochondrial genomes. J Mol Evol 48:427–434

Posada D, Crandall KA (1998) MODELTEST: testing the model of DNA substitution. Bioinformatics 14:817–818

Raina SZ, Faith JJ, Disotell TR, Seligmann H, Stewart C, Pollock DD (2006) Evolution of base-substitution gradients in primate mitochondrial genomes. Genome Res 15:665–673

Rodriguez F, Oliver JL, Marin A, Median JR (1990) The general stochastic model of nucleotide substitution. J Theor Biol 142:485–501

Roman J, Palumbi SR (2003) Whales before whaling. Science 301:508–510

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, Dizon AE, Haygood MG (1995) Variability of the mitochondrial control region in populations of the harbour porpoises, *Phocoena phocoena*, on interoceanic and regional scales. Can J Fish Aquat Sci 52:1210–1219

Rosenbaum HC, Egan MG, Clapham PJ, Brownell RL Jr, Malik S, Brown MW, White BN, Walsh P, DeSalle R (2000) Utility of North Atlantic right whale museum specimens for assessing changes in genetic diversity. Conserv Biol 17:1837

Rozas J, Sanchez-Delbarrio JC, Messeguer X, Rozas R (2003) DnaSP, DNA polymorphism analysis by the coalescent and other methods. Bioinformatics 19:2496–2497

Sasaki T, Nikaido M, Hamilton H, Goto M, Kato H, Kanda N, Pastene LA, Cao Y, Fordyce RE, Hasegawa M, Okada N (2005) Mitochondrial phylogenetics and evolution of mysticete whales. Syst Biol 54:77–90

Siguroardottir S, Helgason A, Gulcher JR, Stefansson K, Donnelly P (2000) The mutation rate in the human mtDNA control region. Am J Hum Genet 66:1599–1609

Swartz SL, Taylor B, Rugh D (2006) Gray whale *Eschrichtius robustus* population and stock identity. Mammal Rev 36:66–84

Swofford DL (2003) PAUP*. Phylogenetic Analysis Using Parsimony (*and other methods), version 4. Sinauer Associates, Sunderland, MA

Tajima F (1983) Evolutionary relationship of DNA sequences in finite populations. Genetics 105:437–460

Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123:585–595

Tamura K, Nei M (1993) Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Mol Biol Evol 10:512–526

Wakeley J (1994) Substitution-rate variation among sites and the estimation of transition bias. Mol Biol Evol 11:436–442

Watterson GA (1975) On the number of segregation sites. Theor Pop Biol 7:256–276

Whitehead H (1998) Cultural selection and genetic diversity in matrilineal whales. Science 282:1708–1711

Xia X, Hafner MS, Sudman PD (1996) On transition bias in mitochondrial genes of pocket gophers. J Mol Evol 43:32–40

Yang Z (2003) Phylogenetic analysis by maximum likelihood (PAML), version 3.15. Institute of Molecular Evolutionary Genetics, Pennsylvania State University, Philadelphia