MOLECULAR ECOLOGY

Molecular Ecology (2009) **18**, 3379–3393 doi: 10.1111/j.1365-294X.2009.04283.x

Assessing substitution patterns, rates and homoplasy at HVRI of Steller sea lions, *Eumetopias jubatus*

C. D. PHILLIPS,*† R. G. TRUJILLO,‡ T. S. GELATT,§ M. J. SMOLEN,* C. W. MATSON,¶ R. L. HONEYCUTT,** J. C. PATTON*† and J. W. BICKHAM*†

**Center for the Environment, Purdue University, West Lafayette, IN 47907, USA, †Department of Forestry and Natural Resources, Purdue University, West Lafayette, IN 47907, USA, ‡USDA Forest Service, Northern Research Station, 202 Natural Resources, University of Missouri, Columbia, MO 65211-7260, USA, §National Marine Fisheries Service, Alaska Fisheries Science Center, National Marine Mammal Laboratory, 7600 Sand Point Way, NE Seattle, WA 98114, USA, ¶Nicholas School of the Environment, Duke University, Durham, NC 27708, USA, **Natural Science Division, Pepperdine University, 24255 Pacific Coast Highway, Malibu, CA 90263, USA*

Abstract

Despite the widely recognized incidence of homoplasy characterizing this region, the hypervariable region I (HVRI) of the mitochondrial control region is one of the most frequently used genetic markers for population genetic and phylogeographic studies. We present an evolutionary analysis of HVRI and cytochrome *b* sequences from a range-wide survey of 1031 Steller sea lions, *Eumetopias jubatus*, to quantify homoplasy and substitution rate at HVRI. Variation in HVRI was distributed across 41 variable sites in the 238-bp segment examined. All variants at HVR1 were found to be transitions. However, our analyses suggest that a minimum of 101 changes have actually occurred within HVRI with as many as 18 substitutions occurring at a single site. By including this hidden variation into our analyses, several instances of apparent long-range dispersal were resolved to be homoplasies and 8.5–12% of observed HVRI haplotypes were found to have geographic distributions descriptive of convergent molecular evolution rather than identity by descent. We estimate the rate of substitution at HVRI in Steller sea lions to be ~24 times that of cytochrome *b* with an absolute rate of HVRI substitution estimated at 27.45% per million years. These findings have direct implications regarding the utility of HVRI data to generate a variety of evolutionary genetic hypotheses.

Keywords: convergent molecular evolution, *Eumetopias jubatus*, homoplasy, HVRI, Steller sea lion, substitution rate

Received 17 February 2009; revision received 20 May 2009; accepted 29 May 2009

Introduction

The mitochondrial genome (mtDNA) is a nearly universally used tool for addressing evolutionary genetic questions in a variety of contexts. The reasons for the wide use of mtDNA sequence data are a reflection of inherent properties of the molecule including its haploid inheritance, high cellular copy number, elevated mutation rate relative to the nuclear genome, strict maternal inheritance, absence of recombination and contrasting rates of substitution across various components of the genome (Pakendorf & Stoneking 2005). These characteristics make acquisition and interpretation of data relatively straightforward, provide ample polymorphisms, and enable questions to be posed over a diversity of timescales.

The mtDNA control region, a site containing the origin of heavy strand replication, is noncoding and displays a mosaic of conserved and highly variable regions. Rather than selective constraints based on codon position as in the protein coding genes, portions of the control region that are highly conserved are probably stabilized by the maintenance of secondary structural characteristics necessary for DNA replication (Saccone *et al.* 1991). In contrast, the hypervariable

Correspondence: Caleb D. Phillips, Fax: 765-496-1369; E-mail: phillip6@purdue.edu

© 2009 Blackwell Publishing Ltd

regions I and II (HVRI, HVRII) of the control region exhibit extensive genetic variability (Pakendorf & Stoneking 2005) and contain mutational hot spots subject to high rates of nucleotide substitution (Excoffier & Yang 1999; Meyer *et al.* 1999; Meyer & von Haeseler 2003; Malyarchuk & Rogozin 2004). However, it is unknown whether these hot spots are the consequence of site-specific positive selection or an increased rate of evolution at the molecular level influenced by sequence context (Meyer *et al.* 1999; Stoneking 2000; Howell *et al.* 2007). Because of these substitution properties, the control region is currently best described using a finite alleles model with consideration to rate heterogeneity among sites (Wakeley 1993; Excoffier & Yang 1999; Meyer *et al.* 1999; Meyer & von Haeseler 2003).

The high level of genetic variability often observed at HVRs (HVRI in particular) has led to the extensive use of this locus in population studies. Nevertheless, except for humans, where the depth of archaeological and pedigree information allows for reasonable rate estimates (Pakendorf & Stoneking 2005), little is known about the underlying substitution patterns and rates of nucleotide substitution in HVRs. Because substitution rate is critical to the calculation of theta (θ), a measure of diversity, as well as several other population parameters, and because of the broad exploration of HVRs observed in the literature, our lack of understanding about HVRs substitution rate represents a significant gap in our knowledge. In addition to the difficulty of calculating reliable substitution rates, recurrent substitutions at unidentified mutational hotspots reduces the phylogenetic utility of HVR data sets by increasing the number of reticulations in haplotype networks (Heyer *et al.* 2001). Finally and importantly, convergent molecular evolution at the haplotype level, caused by underlying homoplasy leads to observed phylogeographic patterns discordant with the evolutionary history of the gene.

In contrast to HVRs, the mitochondrial cytochrome *b* gene is protein coding and patterns of nucleotide substitution within this gene vary across codon position. This gene, like HVRs, also is widely used in evolutionary genetic studies, and reveals less haplotype diversity than is typical of HVRs. Because mitochondrial DNA is inherited as a single locus, parallel comparisons of both control region and cytochrome *b* sequences offer an opportunity to investigate how the substitution processes of HVRs influence patterns of haplotype distribution and diversity. Selective pressure, phylogeographic and demographic events, and mutation affect sequence diversity and population genetic structure and in part is the basis for why a comparison between HVRs and cytochrome *b* can provide new information about substitution processes at HVRs. The nonrecombining nature of the mtDNA genome dictates that all of these influences act on the entire genome rather than on individual genes, and as such, gene trees derived from different regions of the mtDNA genome should be similar (Pakendorf & Stoneking 2005). Related, each mutational event, recurrent or unique, is hypothetically traceable through its linkage with other apomorphic mutations occurring in other regions of the mtDNA genome. Because of this, more information about substitution processes responsible for patterns of HVR haplotype diversity can be recovered.

The Steller sea lion, *Eumetopias jubatus*, is distributed along the North Pacific Rim from the Sea of Okhotsk to central California (Loughlin *et al.* 1987; Fig. 1). Currently, field work has resulted in the accumulation of over 7000 tissue samples taken from pups at their natal rookeries. Since 1992, these samples have provided a resource for detailed studies of the population genetic structure of *E. jubatus*, especially as it relates to mitochondrial phylogeography. For example, throughout the range of *E. jubatus*, HVRI (238 bp) has revealed high levels of genetic variation (Bickham *et al.* 1996) and broad-scale geographic structuring revealed by analysis of HVRI justified the separation of the *E. jubatus* population into eastern and western stocks near Prince William Sound, Alaska (144°W). Further studies indicated another subdivision within the western stock (Bickham *et al.* 1998; Baker *et al.* 2005), and resulted in the recognition of an Asian stock. Using information from 531 bp of control region sequence (including HVR1), O'Corry-Crowe *et al.* (2006) discovered another less dramatic, albeit biologically significant, divergence within the western stock that correlates with differences in population trends, diet and the physical environment. Additionally, population genetic patterns identified using nuclear markers corroborate the general patterns of population subdivision derived from analyses of mtDNA sequence variation in this species (Hoffman *et al.*, 2009).

Hypervariable region I sequence data combined with full cytochrome *b* sequence data have also been used to investigate the phylogeographic history of the species (Harlin-Cognato *et al.* 2006). Those authors suggested that the temporal history of *E. jubatus* probably consists of repeated periods of range expansion followed by population contractions and fragmentation caused by glacial episodes. In recent times, the Western and Asian stocks of *E. jubatus* experienced severe population declines (Kenyon & Rice 1961; Loughlin *et al.* 1992; Trites & Donnelly 2003), resulting in the protection of the species under the US Endangered Species Act. The above described genetic and demographic data in conjunction with a morphological data set were used to justify recognition of two subspecies; *Eumetopias jubatus jubatus* (western and Asian stocks) and *Eumetopias jubatus monteriensis* (eastern stock) (Phillips *et al.* 2009;

© 2009 Blackwell Publishing Ltd

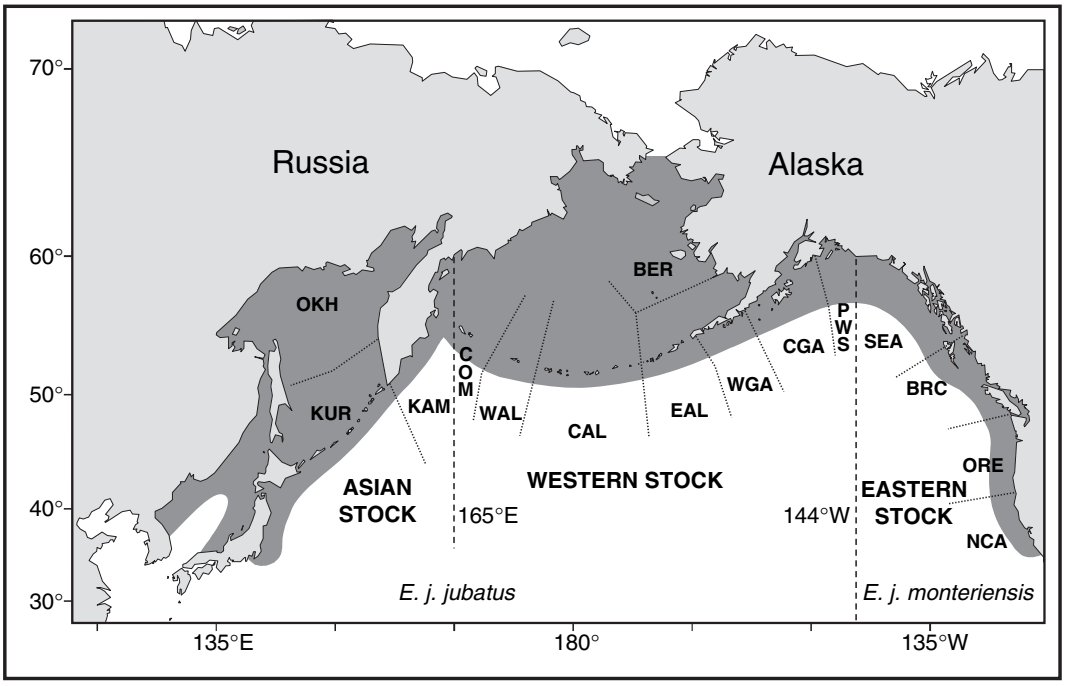

Fig. 1 Map of the distribution of *Eumetopias jubatus*. Region designations are as follows: OKH, Sea of Okhotsk; KUR, Kuril Islands; KAM, Kamchatka Peninsula; COM, Commander Islands; WAL, Western Aleutian Islands; CAL, Central Aleutian Islands; EAL, Eastern Aleutian Islands; WGA, Western Gulf of Alaska; BER, Bering Sea; CGA, Central Gulf of Alaska; PWS, Prince William Sound; SEA, Southeastern Alaska; BRC, British Columbia; ORE, Oregon; NCA, Northern California.

Fig. 1). All these studies have resulted in a clear picture of *E. jubatus* population structure as well as the accumulation of a large genetic data set that enables the two gene comparison conducted in the present study.

The current research builds on findings of previous studies by investigating how the observed distribution of HVRI haplotypes in *E. jubatus* reflects the evolutionary history of substitution at this locus. Initially, genetic linkages and geographic distributions of HVRI and cytochrome *b* haplotypes were defined and examined to make inferences about the evolutionary history having produced the observed patterns. Site-specific rates of substitution and sites of recurrent substitution were identified to make site by site nucleotide comparisons as well as to extrapolate rates of substitution at HVRI relative to cytochrome *b*. In addition, information obtained from the various observed HVRI/cytochrome *b* haplotype linkages was used to estimate the absolute substitution rate of HVRI. By combining all available information, a clearer picture of HVRI evolution emerges.

Materials and methods

Sequence data acquisition

Hypervariable region I haplotypes for most individuals included in this study were reported in previous studies (Baker *et al.* 2005), and a portion of the animals was previously reported for cytochrome *b* (Harlin-Cognato *et al.* 2006). Baker *et al.* (2005) reported HVRI sequence data from 1568 individuals whereas Harlin-Cognato *et al.* (2006) reported cytochrome *b* and HVRI sequence data for 336 individuals. In this study, we used 1031 individuals (pups sampled at their natal rookeries) sampled from across the species distribution (including both subspecies and the three inclusive stocks) from which both mitochondrial regions were sequenced (Appendix I). For samples from which either (but not both) HVRI or cytochrome *b* sequence data were previously obtained, the missing gene sequence was obtained following the methodologies described by Baker *et al.* (2005) and Harlin-Cognato *et al.* (2006). Using different sequencing platforms, various primer combinations were required to obtain sequence data across the genes. Primers LGL 560 and LGL 610 (Harlin-Cognato *et al.* 2006) were sufficient to obtain sequence data for cytochrome *b* when automated sequencing was performed on ABI BigDye version 1.1. However, to obtain full length sequences for cytochrome *b* using ABI BigDye version 3.1 technologies, it was necessary to move the reverse primer 100 bp down stream as version 3.1 consistently gave longer reads but with quality sequence starting on average about 40 bp more distant from the primer. The new primer, Ejub710,

© 2009 Blackwell Publishing Ltd

has the sequence GCATTAAGATTAGTAGGAGGGTT. Similarly, primers LGL 283 and LGL 1115 (Baker *et al.* 2005) were used for amplifying and sequencing HVRI using ABI BigDye version 1.1 technologies. Using ABI BigDye version 3.1 technologies, primers LGL 283 and Ejub-R reported by O'Corry-Crowe *et al.* (2006) were used for amplification followed by the use of a new primer CRseq (CCAAATGCATGAMACCWCAG) as an internal sequencing primer. For quality control and to confirm the authenticity of all sequence variation, at least one individual representing each unique observed HVRI/cytochrome *b* linkage was re-sequenced for both genes.

All sequences were aligned by eye in Sequencher version 4.8 (Gene Codes Corporation). Neighbour-joining trees were constructed from Kimura (1980) 2-parameter distances in PAUP version 4.0b10 (Swofford 2003) and used to resolve the haplotype identity of sequences. Subsequently, names for HVRI haplotypes were assigned letter designations following Bickham *et al.* (1996), whereas cytochrome *b* haplotypes were assigned Arabic numerals following Harlin-Cognato *et al.* (2006). The concatenation of both haplotypes observed in each individual was performed to establish the most robust phylogeny from which to infer the prevalence of homoplasy at HVRI as well as to provide a basis for characterizing gene and site-specific substitution patterns and rates. Composite haplotypes resulting from the concatenation of linked HVRI and cytochrome *b* variants were identified by combining the name of the HVRI and cytochrome *b* haplotypes. Nucleotide position designation follows the numbering given for the *Eumetopias jubatus* mitochondrial genome sequence (Arnason *et al.* 2002).

Characterizing the evolutionary history of observed haplotypes

Hypervariable region I haplotypes observed linked to multiple cytochrome *b* haplotypes within the sample of 1031 individuals (Table 1) were included in an analysis designed to investigate the role of homoplasy vs. haplotype descent in shaping the distribution of haplotype diversity. Haplotype frequencies were calculated and haplotype networks for cytochrome *b* and HVRI were constructed separately following the rules of parsimony as implemented in TCS version 1.3 (Clement *et al.* 2000) and overlaid to map substitution patterns. Specifically, by comparing these haplotype networks in light of the various linkage combinations observed between the two loci, additional information about the substitution processes that have led to the observed frequencies and geographic distributions of HVRI haplotypes was provided. The major clarification provided by this method was discrimination between identity by descent and identity by state (homoplasy) at HVRI. These comparisons were made with the following assumptions:

1 The haplotype network traces its ancestry to a single ancestor possessing a linkage of the plesiomorphic HVRI and cytochrome *b* haplotypes.
2 New linkages appear as a result of new mutations at either locus.

Table 1 The count of observed occurrences of 13 HVRI haplotypes (columns) with their respective multiple cytochrome *b* haplotype linkages

	BB	YYYY	A	FF	FFF	S	T	CC	H	Q	Z	MMM	FFFF	C*b* Totals
Cyt *b* 1	169	*	34	9	*	93	2	34	*	*	5	10	9	365
Cyt *b* 2	46	8	31	*	*	*	*	*	*	14	*	*	*	99
Cyt *b* 3	48	1	3	*	*	*	*	*	62	*	39	*	1	154
Cyt *b* 4	44	*	2	2	*	*	*	*	*	*	*	*	*	48
Cyt *b* 6	5	*	*	4	*	*	*	*	*	*	*	*	*	9
Cyt *b* 9	3	*	*	*	*	*	*	*	*	*	*	*	*	3
Cyt *b* 10	6	4	*	2	*	*	*	*	*	*	*	*	*	12
Cyt *b* 16	*	*	*	*	*	*	*	1	*	*	*	*	*	1
Cyt *b* 19	*	*	*	*	1	*	*	*	*	*	*	*	*	1
Cyt *b* 20	*	*	*	*	*	1	2	*	*	*	*	*	*	3
Cyt *b* 22	*	*	*	*	*	3	*	*	*	*	*	*	*	3
Cyt *b* 23	*	*	*	*	1	21	*	*	*	*	*	*	*	22
Cyt *b* 27	*	1	*	*	*	*	*	*	*	*	*	*	*	1
Cyt *b* 31	*	1	*	*	*	*	*	*	*	3	*	*	*	4
Cyt *b* 32	*	*	*	*	*	*	*	*	3	*	*	*	*	3
Cyt *b* 36	*	*	*	*	*	*	*	*	*	*	*	*	1	1
Cyt *b* 40	*	*	*	*	*	*	*	*	*	*	*	1	*	1
HVRI totals	321	15	70	17	2	118	4	35	65	17	44	11	11	730

© 2009 Blackwell Publishing Ltd

3 Because of the hypervariable nature of HVRI, homoplastic mutations are much more likely in HVRI than in cytochrome *b*.
4 Plesiomorphic (ancestral) haplotypes are common, geographically widespread, and exhibit greater numbers of haplotype linkages than derived or tip haplotypes.
5 Two HVRI haplotypes can have no more than one cytochrome *b* linkage in common except by homoplasy.
6 Linkages of a given HVRI haplotype to multiple cytochrome *b* haplotypes involve homoplasy if the HVRI haplotype is not linked to all cytochrome *b* haplotypes in the series of mutational steps at cytochrome *b* (except when involving an inferred cytochrome *b* haplotype).

Describing rates and patterns of substitution at HVRI

All composite haplotypes were used to calculate site-specific rates of substitution to both assess patterns of rate variation among sites as well as to calculate the relative rates of substitution for HVRI and cytochrome *b*. First, the program ModelTest 3.7 (Posada & Crandall 1998) was used to determine the best model of substitution for concatenated HVRI and cytochrome *b* sequences as well as the optimal number of discrete Γ categories (to describe rate heterogeneity), which were evaluated using both the Akaike Information Criterion (AIC) and hierarchical likelihood ratio tests (hLRTs). Next, the program TREE-PUZZLE (Schmidt *et al.* 2002) was implemented for maximum-likelihood tree construction using the quartet puzzling algorithm (50 000 quartet puzzling steps) of Strimmer and von Haeseler (1996). For this, assignment to site-specific rate categories were determined by a discrete Γ-distribution, approximating a continuous distribution. The number of categories corresponded to that determined in ModelTest. A python script was used to extract Γ rate category assignments of sites from TREE-PUZZLE's output. Because individual rates of substitution per site within HVRI and cytochrome *b* were expressed as rates relative to the average substitution rate of HVRI and cytochrome *b* combined, the relative rate of substitution of HVRI to cytochrome *b* was calculated by taking the ratio of the average substitution rate for each gene.

In addition to estimating site-specific nucleotide substitution rate by Γ rate category assignment, further analysis was directed towards quantifying rate variation within rate categories by estimating actual numbers of substitutions per site, thus more accurately describing rate heterogeneity. To do this, a neighbour-joining tree was constructed in PAUP version 4.0b10 (Swofford 2003) from the concatenated sequences. Protein coding nucleotides were preferentially up-weighted relative to HVRI nucleotides to ensure that tree topology, especially among basal nodes, was determined by substitution events at cytochrome *b*. This method is similar to that employed by Bandelt *et al.* (2002) for the identification of mutational hotspots. Although recurrent substitution at cytochrome *b* within the *E. jubatus* lineage is possible, preliminary analysis of haplotype network structure, Γ category assignments (see Fig. 4a), and substitution pattern across codon positions (9 of 19 substitutions occur at the first or second codon position) present no signs of homoplasy at cytochrome *b*, thus strengthening the rationale for the use of this procedure. Nucleotide substitutions occurring at HVRI were then mapped on the neighbour-joining tree, and sites having experienced recurrent substitution within the *E. jubatus* lineage were identified. The quantification of site-specific numbers of recurrent substitution at HVRI through this method enabled an independent estimate of substitution rate at HVRI relative to cytochrome *b*.

Recently, Alter & Palumbi (2009) have presented a method for estimating substitution rates of rapidly evolving markers by using the substitution rate at a linked and more slowly evolving marker for calibration. Using this approach, a third and final estimate of the HVRI substitution rate in *E. jubatus* was calculated. The substitution rate per base pair at HVRI was estimated as $x/((1/2)\cdot w\cdot n)$, where x is the mean number of pairwise differences at HVRI among individuals identical at cytochrome *b*, w is the estimated waiting time until the next substitution at cytochrome *b*, and n is the length of the HVRI sequence used. For this analysis, complete cytochrome *b* sequences (1140 bp) were obtained from nine otariid (eared seals) species from the NCBI database (Accession nos: *Arctocephalus gazella*, 693951; *Arctocephalus* forsteri, 693953; *Eumetopias jubatus*, 34577124; *Arctocephalus townsendi*, 115344885; *Arctocephalus pusillus*, 115344843; *Callorhinus ursinus*, 115344815; *Zalophus californicus*, 115344829; *Neophoca cinerea*, 115344871; *Phocarctos hookeri*, 115344857). From these sequences, the synonymous pairwise distance was calculated using the method of Li *et al.* (1985) in MEGA version 4 (Tamura *et al.* 2007). The silent substitution rate at cytochrome *b* was then estimated by half the slope of the relationship of the regression of synonymous pairwise distances against divergence times (Higdon *et al.* 2007). From this value, w was calculated as $(\mu\cdot n)^{-1}$, where μ is the substitutions per base pair per year and n is the number of fourfold degenerate sites plus one-third the number of twofold degenerate sites (Li *et al.* 1985).

Demographic phenomena such as population expansions and contractions as well as selection can modify the distribution of haplotype diversity, and hence can

© 2009 Blackwell Publishing Ltd

affect calculations directed towards assessing substitution pattern and rate. To identify any patterns of variation indicative of non-neutral evolution present in the *E. jubatus* data set that could potentially affect the previous analyses, Tajima's *D** (Tajima 1989), Fu and Li's *D** and Fu and Li's *F** (Fu & Li 1993), as well as Fay and Wu's *H* (Fay & Wu 2000) were calculated for both HVRI and cytochrome *b*. These statistics were calculated at all levels including the total population (species), subspecies and stocks.

Results

The 1031 individuals examined for both HVRI and cytochrome *b* yielded 83 HVRI and 18 cytochrome *b* haplotypes. Thirteen HVRI haplotypes found in 70.8% ($n = 730$) of individuals were linked to multiple (17) cytochrome *b* haplotypes and are thus included in this study (Table 1). The number of linkages each of the 13 HVRI haplotypes had with different cytochrome *b* haplotypes ranged from 2 to 7, totalling 41 different haplotype linkages (Fig. 3). Thirteen of these haplotype linkages were found exclusively in *E. j. monteriensis*. Twenty haplotype linkages were unique to *E. j. jubatus*, with seven and eight of these being unique to the western and Asian stocks, respectively (Figs 2 and 3). The haplotype network constructed from the 13 HVRI haplotypes linked to multiple cytochrome *b* haplotypes when considered alone is explained by 13 substitutions. However, incorporating information from the cytochrome *b* network revealed a HVRI network explained by 24–29 substitutions.

Characterizing the evolutionary history of observed haplotypes

The geographic distributions of HVRI haplotypes in conjunction with their observed linkages to cytochrome *b* haplotypes provided insight regarding the unique evolutionary pathways leading to the observed distributions of HVRI haplotypes. Inferences regarding specific haplotypes are outlined below and summarized in Table 2.

Hypervariable region I haplotype BB. This haplotype is the hub of the HVRI network and is a potential direct ancestor for eight of the remaining 12 HVRI haplotypes that are linked to multiple cytochrome *b* haplotypes. This abundant and geographically widespread haplotype was found associated with the greatest number of cytochrome *b* haplotypes ($n = 7$). Mapping this HVRI haplotype on the cytochrome *b* network (Fig. 3) revealed its seven cytochrome *b* haplotypes link ages to BB are separated by one or two mutational changes covering both halves of the network. In addition, 15.28% of 2317 pups previously examined for HVRI exhibit the BB haplotype (Baker *et al.* 2005 and unpublished data). Although homoplasy has probably increased the BB haplotype's distribution and frequency, it is reasonable to assume that BB probably represents the pleisiomorphic HVRI haplotype.

Hypervariable region I haplotypes CC, H and MMM. These three haplotypes are linked to two cytochrome *b* haplotypes each. In each case, one of the cytochrome *b* haplotypes is of low frequency and not linked to any other HVRI haplotypes. In all cases, these unique combinations are separated by a single mutational step at cytochrome *b*. These multiple haplotype linkages are probably the result of mutation at cytochrome *b*.

Hypervariable region I haplotypes A, FF, FFF and Z. These four haplotypes are each linked to multiple cytochrome *b* haplotypes and are probably the consequence of homoplastic mutations at HVRI. The number of mutational events necessary to produce each HVRI/cytochrome *b* linkages is 4, 4, 2 and 2, respectively (Fig. 3). Cytochrome *b* haplotypes with which HVRI haplotypes A and Z are linked also are locally observed in association with BB (Fig. 2). FFF is linked to two cytochrome *b* haplotypes that differ by four substitutions. FF occurs with four cytochrome *b* haplotypes that differ by one to four substitutions. In addition, all of these haplotypes occur on both halves of the cytochrome *b* network (Fig. 3).

Hypervariable region I haplotypes FFFF, YYYY, and S. FFFF is linked to three cytochrome *b* haplotypes through two mutational events at HVRI (Fig. 3). This haplotypes' linkage to cytochrome *b* haplotype 36 was the result of mutation at cytochrome *b*; however, the linkage to cytochrome *b* haplotype 1 or 3 is due to a homoplastic mutation at HVRI. YYYY is linked with five cytochrome *b* haplotypes, by an undetermined three or four mutations at HVRI. YYYY was retained with concurrent mutation at cytochrome *b* resulting in a new HVRI/cytochrome *b* linkage (YYYY3–YYYY27). Two other linkages are the result of homoplasy at HVRI (YYYY2, YYYY10), whereas the linkage to cytochrome *b* haplotype 31, the fifth event, could be the result of retention of or homoplasy to YYYY. The linkages of HVRI haplotype S with multiple cytochrome *b* haplotypes also are the result of a complex evolutionary history. Haplotypes S22 and S23 are clearly the result of substitution at cytochrome *b*; however, it is not clear which process led to S20.

© 2009 Blackwell Publishing Ltd

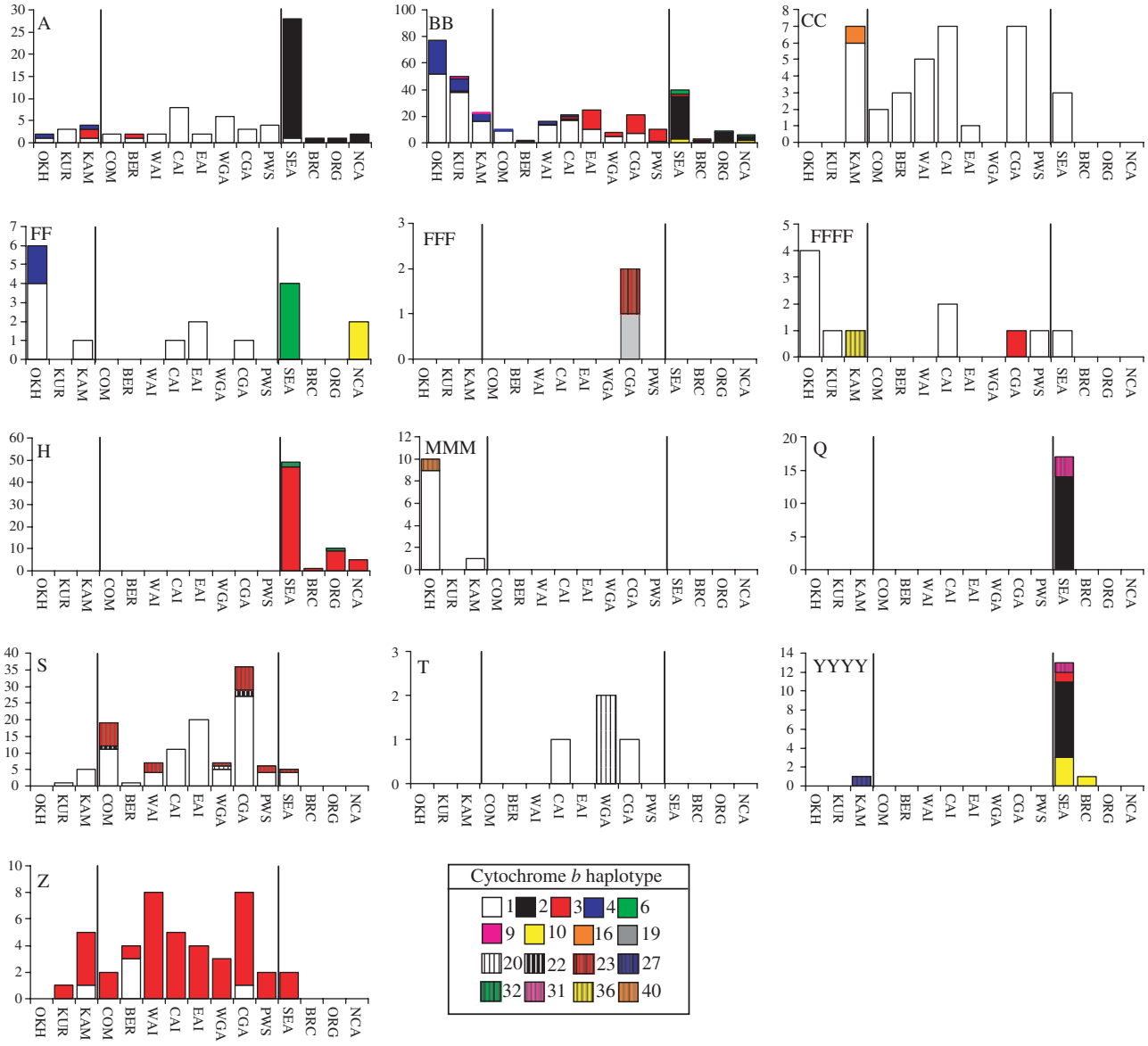

Fig. 2 Histograms displaying the regional distribution of HVRI/cytochrome *b* linkages. The data on a particular HVRI haplotype displayed by each histogram are indicated in the top left corner. Cytochrome *b* linkages for each HVRI haplotype are color coded. Thin vertical lines partition regions into their respective stocks.

Hypervariable region I haplotypes T and Q. Both of these haplotypes have linkages to two different cytochrome *b* haplotypes. It is unclear whether this is the result of retention of, or homoplasy to, the particular HVRI haplotype. HVRI haplotypes T and Q are each linked to common, interior cytochrome *b* haplotypes in one instance (1 and 2, respectively), and terminal, less frequent cytochrome *b* haplotypes in the other instance (20 and 31, respectively). However, these terminal cytochrome *b* haplotypes are also linked to other HVRI haplotypes. Necessarily, one of the HVRI haplotypes linked to each of these terminal cytochrome *b* haplotypes is pleisiomorphic, while the other arose through homoplasy.

Cytochrome b *network.* Despite the complexity of the cytochrome *b* network relative to the network based on the included HVRI haplotypes, it is not necessary to assume any homoplasies to explain this network. Of 19 segregating sites identified at cytochrome *b*, nearly half (9) were located in the first or second codon position. The right side of the cytochrome *b* network consists of haplotypes 1, 9, 16, 20, 22, 23, 36 and 40 (Fig. 3). All of

© 2009 Blackwell Publishing Ltd

Table 2 Summary of the observed linkages of HVRI haplotypes to multiple cytochrome *b* haplotypes (cytb) detailing whether specific linkages arose through identity by descent (IBD), identity by state (IBS, i.e. homoplasy), or was unresolved. HVRI haplotype BB was considered the pleisiomorphic haplotype for this assessment

HVRI haplotype	Cytb linkages	IBD	IBS	Unresolved
BB	7	7		
CC	2	2		
H	2	2		
MMM	2	2		
A	4	1	3	
FF	4	1	3	
FFF	2	1	1	
Z	2	1	1	
FFFF	3	2	1	
YYYY	5	2	2	1
S	4	2		2
T	2	1		1
Q	2	1		1

these haplotypes were exclusively found in *E. j. jubatus* except for 24 *E. j. monteriensis* individuals identified with haplotype 1 and one individual identified with haplotype 23. The left side of the network includes haplotypes 2, 3, 4, 6, 10, 19, 27, 31 and 32. Except for the most interior haplotype (3) on the left side of the network and two *E. j. jubatus* individuals observed with haplotype 2, all haplotypes were observed in one subspecies or the other, but not both. The remaining observed cytochrome *b* haplotype (35) is excluded from Fig. 3 because its only HVRI haplotype linkage was unique. Because of its distribution and frequency, haplotype 1 is considered the pleisiomorphic cytochrome *b* haplotype. The pleisiomorphic HVRI/cytochrome *b* haplotype is therefore considered to be BB/1 and is thus used to root the network.

Describing rates and patterns of substitution at HVRI

Both AIC and hLRT selected HKY + Γ + I (−ln L = 2612.0374, AIC = 5236.0747, six rate categories, $\alpha = 0.8793$) as the optimal model of sequence evolution for all concatenated HVRI and cytochrome *b* haplotypes. As expected, sites with the highest substitution rates were found exclusively within HVRI (Fig. 4a). Specifically, 5 of 41 variable HVRI sites were placed in the highest rate category, a category with an estimated rate 39.27 times the average of the concatenated sequence. In addition, all five of these sites were variable within the set of 13 HVRI haplotypes associated with multiple cytochrome *b* haplotypes. The substitution rate for 238 bp of HVRI based on Γ rate category assignments was calculated as 25.46 times greater than that of the 1140 bp cytochrome *b*.

By comparison, the tree-weighting and substitution mapping procedure identified 19 sites (46% of all variable HVRI sites) where multiple substitutions have putatively occurred within the *E. jubatus* lineage (Fig. 4b). This analysis identified two outlier sites, 15 466 and 15 536, with a much higher number of estimated substitutions, 18 and 11, respectively (sites that also were placed in the highest rate categories). Accounting for recurrent substitutions, the estimated relative-rate of substitution at HVRI using these data was 23.52 times greater than cytochrome *b*.

The synonymous substitution rate at cytochrome *b* was estimated as 3.26% per million years (Myr; Fig. 5a). The mean pairwise difference at HVRI among individuals identical at cytochrome *b* was 3.69 (SD = 1.55) and the distribution was normally distributed (Kolmogorov–Smirnov Test, $D = 0.15429$, $\Pr > D = <0.010$; Fig. 5b). From these estimates and following the methods of Alter & Palumbi (2009), the substitution rate at HVRI was estimated at 27.45% (15.92–38.97%) per Myr.

The various tests implemented to assess non-neutral substitution process at HVRI and cytochrome *b* (Tajima's *D*, Fu and Li's *D*, Fu and Li's *F*, and Fay and Wu's *H*) yielded nonsignificant test statistics at all levels considered except within *E. j. jubatus* in which Tajima's *D* ($D = -1.376$, $P = 0.041$) and Fu and Li's *F* ($F = -2.036$, $P = 0.021$) were significantly negative.

Discussion

Population genetic signal obtained from intraspecific HVRI sequence data sets has largely been interpreted following the assumption of identity by descent among observed haplotypes. In this study, based on comparison with cytochrome *b,* the number of substitutions occurring at HVRI is at least 2.46 times greater than the number of segregating sites observed at this locus when considered alone. In other words, while 41 variable sites were identified at HVR1, 101 total substitutions were also identified. Consequently, geographic distributions of HVRI haplotypes are shaped in part by convergent molecular evolution. That is to say, convergent haplotypes appear in different geographic areas that can only be detectable as homoplasies upon comparison with another linked locus. We estimate that in this data set the geographic distributions of ∼8.5–12% of the observed HVRI haplotypes were incorrectly understood to be broader than is actually the case. Once the homoplastic mutations were identified, more accurate estimations of the distributions and frequencies of maternal lineages were obtained.

© 2009 Blackwell Publishing Ltd

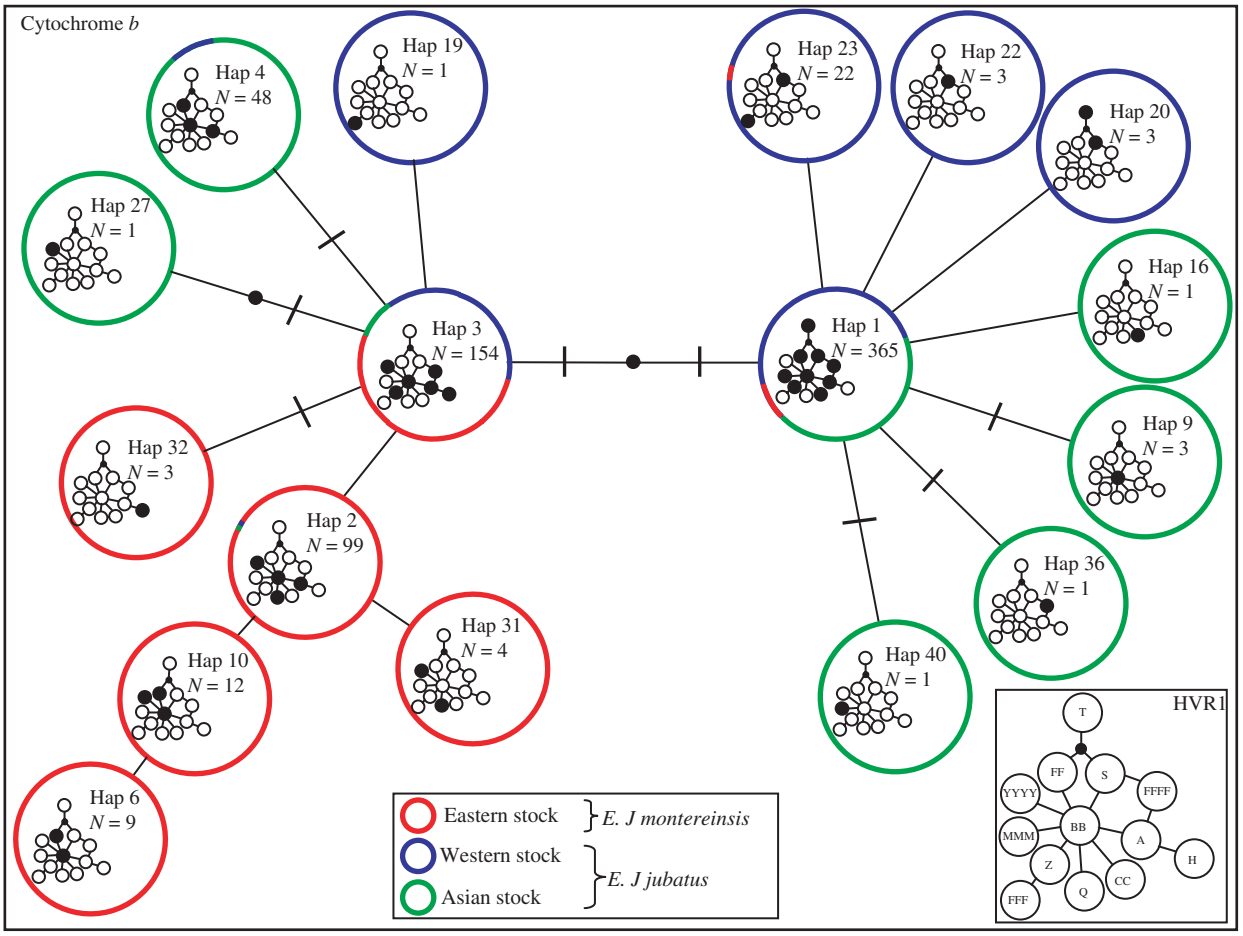

Fig. 3 In the cytochrome *b* haplotype network (large main network), each large circle represents one of 17 observed haplotypes. Lines connecting the haplotypes represent mutational steps. Small filled circles on these lines represent missing inferred haplotypes. Hash marks on lines connecting haplotypes indicate a mutation resulting in an amino acid replacement. The small network in the lower right of the figure and reproduced inside each cytochrome *b* haplotype is the network of 13 HVRI haplotypes linked to multiple cytochrome *b* haplotypes. Filled circles on the HVRI networks inside each cytochrome *b* haplotype circle indicate those particular haplotypes were linked to the cytochrome *b* haplotype in which they appear. The colorings of the perimeter of the circles representing cytochrome *b* haplotypes indicate the geographic distribution among the three stocks.

As an example, consider the two most abundant and geographically widespread HVRI haplotypes in this data set, A and BB. From analysis of HVRI it appears that both of these haplotypes range-wide distribution and abundance is a reflection of their ancient genesis. However, by considering their linkages to different cytochrome *b* haplotypes across the species' range, a more likely scenario explaining their range-wide distribution emerges. Initially, it is clear that convergence has occurred by noting that these two haplotypes share four cytochrome *b* haplotype linkages in common. Minimally, two HVRI haplotypes being linked to the same two cytochrome *b* haplotypes can only be the result of homoplasy (assumption 5 from Materials and methods). Between the A and BB haplotypes, there have been three detectable homoplasies. Considering that in each stock, the A haplotype is only associated with cytochrome *b* haplotypes that also are common and locally associated with BB indicate that multiple convergences have occurred independently in each stock. In addition, haplotypes A and BB differ only by one transitional change (A<->G) at nucleotide site 15 536 that was placed in the highest Γ rate category, and 11 substitutions were inferred to have occurred at this site through the tree-weighting procedure. In this example, haplotype convergence has produced an incorrect pattern of range-wide distribution that overestimates population genetic similarity.

There are other instances where homoplasy has produced misleading population genetic patterns. For

© 2009 Blackwell Publishing Ltd

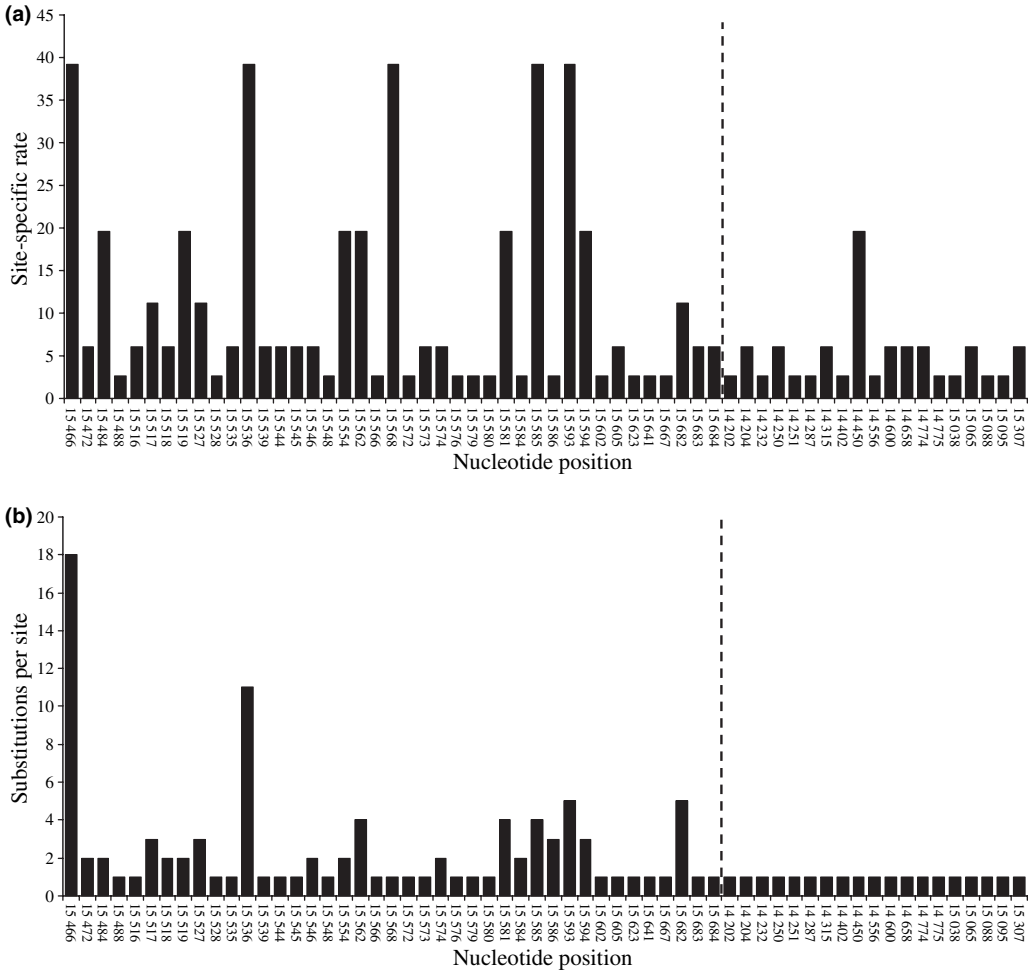

Fig. 4 Histograms of (a) site-specific substitution rates of 238 bp of HVRI (left of hashed line) and the full cytochrome *b* gene (right of hashed line) for all observed and verified haplotypes relative to the average substitution rate of both genes and (b) the inferred number of substitution per site for the same data. Nucleotide position numbering corresponds to that of the *Eumetopias jubatus* mtDNA genome sequence (Arnason *et al.* 2002).

example, FF is associated with four different cytochrome *b* haplotypes that differ by three or four substitutions in distinct geographic and genetically identifiable regions (Figs 2 and 3). At each of these regions, BB is associated with the same cytochrome *b* haplotypes as FF. BB and FF differ only at position 15 466; this site was placed in the highest Γ rate category and was inferred to be a site of 18 recurrent substitutions. Similarly, three of four cytochrome *b* associations of YYYY observed in southeastern Alaska also are found with BB at this location. The perceived distributions and frequencies of S, T, Q, FFF and FFFF have also similarly been shaped by homoplasy. One of the most striking characteristics of these observations is how the observed distribution of HVRI haplotypes created by homoplasy often appears identical to the patterns long-distance dispersal events create and would be undetectable without a second gene comparison to identify this homoplasy (the distributions of HVRI haplotypes FF and YYYY exemplify this phenomenon). As such, we have demonstrated that long-distance dispersal is likely less common in Steller sea lions than was previously estimated based on HVRI sequences alone. While Harlin-Cognato *et al.* (2006) identified three long-distance dispersal events in their Nested Clade analysis of this species, a Nested Clade analysis performed using the current data set in which recurrent substitutions were accounted for yielded no inferences of long distance dispersal (data not shown).

Although phylogenetic and familial based estimates of substitution rate have been calculated for humans (Parsons *et al.* 1997; Sigurðardóttir *et al.* 2000; Heyer *et al.* 2001; Howell *et al.* 2003; Henn *et al.* 2009), the majority of data sets for other species lack the necessary

© 2009 Blackwell Publishing Ltd

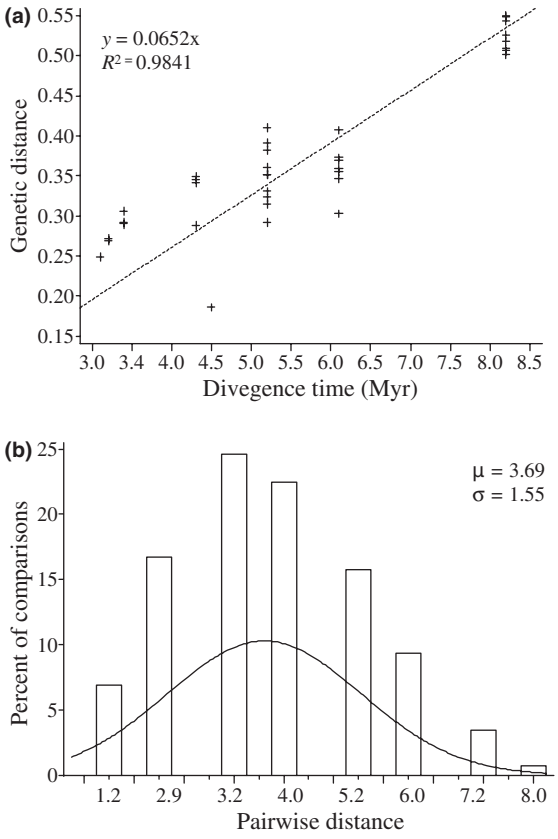

Fig. 5 The (a) plot of the regression of synonymous pairwise distance at cytochrome *b* for nine Otariid species using the method of Li *et al.* (1985) vs. divergence times (Higdon *et al.* 2007) by forcing the regression line through the origin and (b) the distribution of pairwise differences at HVRI among individuals identical at cytochrome *b*.

breadth of archaeological and pedigree information to make accurate estimates of these parameters. Alternatively, rates of substitution for HVRI relative to cytochrome *b* were calculated in this study using two different methods that account for variability in substitution rate among sites. While these two methods differ in their approach to estimating the substitution rate of HVRI relative to cytochrome *b*, they provided very similar estimates (25.46 using Γ rate category assignments and 23.52 using the tree-weighting character mapping procedure). The minor difference in these estimates is attributable to at least a few influences. First, because a discrete rather than continuous Γ-distribution was used to model rate heterogeneity, individual site rates were systematically over- and underestimated, thus influencing the estimation of relative-rate. In addition, Γ rate category assignments in TREE-PUZZLE were made by incorporating tree topology information from a tree constructed with equal weighting to HVRI and cytochrome *b*. Putatively incorrect taxonomic relationships in this tree explain why some sites in cytochrome *b* and HVRI were placed in Γ rate categories inconsistent with the estimated number of substitutions per site. Regardless, these estimates were very similar, complimentary, and demonstrate the highly elevated rate of substitution at HVRI.

The third substitution rate estimate for *Eumetopias jubatus* HVRI of 27.45% per Myr (15.92–38.97%) is large, yet is reasonable in comparison to the estimated HVRI to cytochrome *b* relative substitution rate (23.52–25.46). Previous estimates of control region substitution rate in humans calibrated using ancient DNA (33–44% per Myr; Kemp *et al.* 2007) yielded rates similar to that currently obtained for *E. jubatus*. This comparison not only indicates a similar rate of control region evolution between humans and *E. jubatus* (on a phylogenetic timescale), but also indicates that by defining shortened intervals in which to measure control region substitution (either by using ancient samples or a substitution rate from a linked and more slowly evolving gene) the effects of homoplasy on rate estimation are reduced (Alter & Palumbi 2009). In contrast, pedigree based substitution rate estimates for control region in humans have yielded much higher values (32–250% per Myr; Parsons *et al.* 1997; Sigurðardóttir *et al.* 2000) describing an extremely rapid substitution rate at control region that is not observed at intraspecific timescales likely due to the rapid removal of new control region variation by genetic drift in a maternally inherited, haploid element.

The acquisition of a reasonable substitution rate estimate is critical for meaningful estimation of important population genetic parameters that describe diversity, population size, migration rates and divergence times. In addition, coalescent-based simulation methods designed to model empirical data for evolutionary reconstruction require knowledge of both substitution rate and the Γ-distribution shape parameter, α, to simulate correctly the substitution process. Applying a correction for multiple substitutions to data sets is only useful to the extent that recurrent substitutions are visible in a sequence alignment given a defined phylogenetic tree. When substitutions are cryptic they are unaccounted for in substitution rate estimation and phylogenetic tree estimation, leading to both an underestimation of substitution rate as well as incorrect tree topology. Hence, the substitution rate at HVRI extrapolated from the synonymous substitution rate at cytochrome *b* and the observed HVRI linkages provides a valuable estimate. It should be noted that this method (Alter & Palumbi 2009) is only appropriately applied when sites of silent substitution are not saturated across the taxonomic group used to calculate synonymous pairwise distances. Furthermore, the accuracy of divergence

© 2009 Blackwell Publishing Ltd

dates used for synonymous rate estimation highly affects rate estimation, and w (waiting time to the next mutation) can be influenced by demographic history as pointed out by Alter & Palumbi (2009).

Non-neutral substitution processes do not appear to have strongly influenced the above substitution rate estimates. Test statistics computed at the various levels were generally nonsignificant; however Tajima's D^* and Fu and Li's F^* were significantly negative within *E. j. jubatus* for HVRI. Negative values of these statistics indicate an excess of low frequency polymorphisms suggestive of a population expansion and/or purifying selection. However, no test statistics were significant for cytochrome *b* at this level of population structure. This comparison suggests that any demographic or selective influences producing significant statistics were recent events. In other words, recent events are most detectable in rapidly evolving loci.

Of particular interest is whether the identified mutational hotspots (15 466, 15 536) occur due to selection or some unidentified cellular process. The substitution rate at fourfold degenerate sites is generally considered to be the rate of neutral evolution (Howell *et al.* 2007). In this study, we do not report site rate comparisons to fourfold sites due to the small number of fourfold sites obtained from one 1140 bp gene with only 10 third-position variable sites. Interestingly, Howell *et al.* (2007), using a complete mtDNA genome data set on humans, observed 3% of control region sites with a rate of sequence evolution exceeding the rate of neutral evolution by an order of magnitude. Although in this study a comparison to fourfold degenerate sites is not warranted, sites 15 466 and 15 536 in all likelihood exhibit substitution rates greatly exceeding the neutral rate of sequence evolution, however, the reason for this is not directly identifiable. However, results of multiple studies have provided evidence supporting the role of sequence context in shaping substitution processes at the control region (Meyer *et al.* 1999; Howell and Bogolin Smejkal 2000; Malyarchuk & Rogozin 2004). Indeed, the estimation of 18 substitutions at site 15 466 within *E. jubatus* indicates that factors other than random mutation are influencing substitution at this site.

The observed rapid mutation rate at the HVRs, which served as the impetus for extensive sequencing efforts for many taxa, also bestows a substitution pattern more akin to identity by state rather than identity by descent. In light of this, researchers should employ methods of detection and compensation for recurrent mutation. The additional information from cytochrome *b* in the *E. jubatus* data set enabled the identification of mutational hotspots and the characterization of HVRI substitution patterns and rates. Because most data sets on wildlife populations to date do not consist of multiple mtDNA genes, an alternative method for identifying mutational hotspots at HVRI consists of simply mapping nucleotide substitutions onto a phylogenetic tree constructed from the HVRI data. Conducting this procedure using the *E. jubatus* HVRI data set led to the identification of 17 of the 19 sites of recurrent substitution that were identified by incorporating cytochrome *b* data (data not shown). This approach is useful for gaining a better general understanding about substitution patterns and the location of hotspots in specific HVRI data sets. However, only the inclusion of coding region data provides detailed insight about the role of homoplasy in modifying observed haplotype distributions and allows for more accurate substitution rate estimation, the acquisition of which translates to a better understanding of species' evolution.

Acknowledgements

We thank the numerous biologists who provided tissue samples, especially V. Burkanov (coordinator for the collecting trips in Russia), T. Loughlin and D. Calkins, the NOAA cruises in the Aleutians and the Gulf of Alaska, and the Alaska Department of Fish and Game. Thanks to K. M. Nichols, O. E. Rhodes, J. A. DeWoody, as well as three anonymous referees for their valuable insights and suggestions. Also, special thanks to David Horner for providing the python script used to retrieve site-specific rate assignments from the TREE-PUZZLE output. R. G. Trujillo would like to thank the Hispanic Leadership Program in Agriculture and Natural Resources at Texas A&M University. This study was conducted in partial fulfilment of C. D. Phillips' dissertation. Funding was provided by the National Marine Fisheries Service and the Wildlife and Contaminants Program of the World Wildlife Fund.

References

Alter SE, Palumbi SR (2009) Comparing evolutionary patterns of variability in the mitochondrial control region and cytochrome *b* in three species of baleen whales. *Journal of Molecular Evolution*, **68**, 97–111.

Arnason U, Adegoke JA, Bodin K (2002) Mammalian mitogenomic relationships and the root of the eutherian tree. *Proceedings of the National Academy of Sciences, USA*, **99**, 8151–8156.

Baker AR, Loughlin TR, Burkanov VN *et al.* (2005) Variation of mitochondrial control region sequences of Steller sea lions, *Eumetopias jubatus*: the three-stock hypothesis. *Journal of Mammalogy*, **86**, 1075–1084.

Bandelt HJ, Quintana-Murci L, Salas A, Macaulay V (2002) The fingerprint of phantom mutations in mitochondrial DNA data. *American Journal of Human Genetics*, **71**, 1150–1160.

Bickham JW, Patton JC, Loughlin TR (1996) High Variability for control-region sequences in a marine mammal: implications for conservation and biogeography of Steller sea lions (*Eumetopias jubatus*). *Journal of Mammalogy*, **77**, 95–108.

Bickham JW, Loughlin TR, Wickliffe JK, Burkanov VN (1998) Geographic variation in the mitochondrial DNA of Steller

© 2009 Blackwell Publishing Ltd

sea lions: haplotype diversity and endemism in the Kuril Islands. *Biosphere Conservation*, **1**, 107–117.

Clement MD, Posada D, Crandall KA (2000) TCS: a computer program to estimate gene genealogies. *Molecular Phylogenetics and Evolution*, **3**, 102–113.

Excoffier L, Yang Z (1999) Substitution rate variation among sites in mitochondrial hypervariable region I of humans and chimpanzees. *Molecular Biology and Evolution*, **16**, 1357–1368.

Fay JC, Wu CI (2000) Hitchhicking under positive Darwinian selection. *Genetics*, **155**, 1405–1413.

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

Harlin-Cognato A, Bickham JW, Loughlin TR, Honeycutt RL (2006) Glacial refugia and the phylogeography of Steller's sea lion (*Eumatopias jubatus*) in the North Pacific. *Journal of Evolutionary Biology*, **19**, 955–969.

Henn BM, , Gignoux CR, Feldman MW, Mountain JL (2009) Characterizing the time dependency of human mitochondrial DNA mutation rate estimates. *Molecular Biology and Evolution*, **26**, 217–230.

Heyer E, Zietkiewicz E, Rochowski A *et al.* (2001) Phylogenetic and familial estimates of mitochondrial substitutions rates: study of control region mutations in deep-rooting pedigrees. *American Journal of Human Genetics*, **69**, 1113–1126.

Higdon JW, Bininda-Emonds ORP, Beck RMD, Ferguson SH (2007) Phylogeny and divergence of the pinnipeds (Carnivora: Mammalia) assessed using a multigene dataset. *BMC Evolutionary Biology*, **7**, 216.

Hoffman JI, Matson CW, Amos W, Loughlin TR, Bickham JW (2006) Deep genetic subdivision within a continuously distributed and highly vagile marine mammal, the Steller's sea lion (*Eumetopias jubatus*). *Molecular Ecology*, **15**, 2821–2832.

Hoffman JI, Dasmahapatra KK, Amos W, Phillips CD, Gelatt TS, Bickham JW (2009) Contrasting patterns of genetic diversity at three different genetic markers in a marine mammal metapopulation. *Molecular Ecology* (in press).

Howell N, Bogolin Smejkal C (2000) Persistent heteroplasmy of a mutation in the human mtDNA control region: hypermutation as an apparent consequence of simple-repeat expansion/contraction. *American Journal of Human Evolution*, **66**, 1589–1598.

Howell N, Bogolin Smejkal C, Mackey DA *et al.* (2003) The pedigree rate of sequence divergence in the human mitochondrial genome: there is a difference between phylogenetic and pedigree rates. *American Journal of Human Genetics*, **72**, 659–670.

Howell N, Elson JL, Howell C, Turnbull DM (2007) Relative rates of evolution in the coding and control regions of African mtDNAs. *Molecular Biology and Evolution*, **24**, 2213–2221.

Kemp BM, Malhi RS, McDonough J *et al.* (2007) Genetics analysis of early Holocene skeletal remains from Alaska and its implications for the settlement of the America. *American Journal of Phylogenetic Anthropology*, **132**, 605–621.

Kenyon KW, Rice DW (1961) Abundance and distribution of the Steller sea lion. *Journal of Mammalogy*, **42**, 223–234.

Kimura M (1980) A simple model for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *Journal of Molecular Evolution*, **16**, 111–120.

Li WH, Chu CI, Luo CC (1985) A new method for estimating synonymous and nonsynonymous rates of nucleotide substitution considering the relative likelihood of nucleotide and codon changes. *Molecular Biology and Evolution*, **2**, 150–174.

Loughlin TR, Perez MA, Merrick RL (1987) Eumetopias jubatus. *Mammalian Species*, **283**, 1–7.

Loughlin TR, Perlov AS, Vladimirov VA (1992) Range-wide survey and estimation of total number of Steller sea lions in 1989. *Marine Mammal Science*, **8**, 220–239.

Malyarchuk BA, Rogozin IB (2004) Mutagenesis by transient misalignment in human mitochondrial DNA. *Genetics Research*, **68**, 324–339.

Meyer S, von Haeseler A (2003) Identifying site-specific substitution rates. *Molecular Biology and Evolution*, **20**, 182–189.

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

O'Corry-Crowe G, Taylor BL, Gelatt T *et al.* (2006) Demographic independence along ecosystem boundaries in steller sea lions revealed by mtDNA analysis: implications for management of an endangered species. *Canadian Journal of Zoology*, **84**, 1796–1809.

Pakendorf B, Stoneking M (2005) Mitochondrial DNA and human evolution. *Annual Review of Genomics and Human Genetics*, **6**, 165–183.

Parsons TJ, Muniec DS, Sullivan K *et al.* (1997) A high observed substitution rate in the human mitochondrial DNA control region. *Nature*, **15**, 363–368.

Phillips CD, Bickham JW, Patton JC, Gelatt TS (2009) Systematics of Steller sea lions (*Eumetopias jubatus*): subspecies recognition based on concordance of genetics and morphometrics. *Occasional Papers, Museum of Texas Tech University*, **283**, 1–15.

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

Saccone C, Pesole G, Sbisa E (1991) The main regulatory region of mammalian mitochondrial DNA: structure-function model and evolutionary pattern. *Journal of Molecular Evolution*, **33**, 83–91.

Schmidt HA, Strimmer K, Vingron M, Haeseler A (2002) TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing. *Bioinformatics*, **18**, 502–504.

Sigurðardóttir S, Helgason A, Gulcher JR, Stefansson K, Donnelly P (2000) The mutation rate in the human mtDNA control region. *American Journal of Human Genetics*, **66**, 1599–1609.

Stoneking M (2000) Hypervariable sites in the mtDNA control region are mutational hotspots. *American Journal of Human Genetics*, **67**, 1029–1032.

Strimmer K, von Haeseler A (1996) Quartet puzzling: A quartet maximum likelihood method for reconstructing tree topologies. *Molecular Biology & Evolution*, **13**, 964–969.

Swofford DL (2003) PAUP*: *Phylogenetic Analysis Using Parsimony (* and Other Methods)*, Version 4.0b 10. Sinauer Associates, Sunderland, MA.

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

Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. *Molecular Biology and Evolution*, **24**, 1596–1599.

© 2009 Blackwell Publishing Ltd

Trites AW, Donnelly CP (2003) The decline of Steller sea lions Eumetopias jubatus in Alaska: a review of the nutritional stress hypothesis. *Mammal Review*, **33**, 3–28.

Wakeley J (1993) Substitution rate variation among sites in hypervariable region 1 of human mitochondrial DNA. *Journal of Molecular Evolution*, **37**, 613–623.

Caleb Phillips' research interests include mtDNA molecular evolution, phylogeography, speciation, hybridization, and systematics. Robert Trujillo's research interests focus on molecular systematics, phylogeography, and population genetics of mammals. Tom Gelatt and his research group are responsible for investigating various questions pertaining to the declining populations of Steller sea lions contributing a wide assortment of projects involving animal foraging behaviors, demographics, and abundance estimation to collectively address these questions and provide fisheries management direction. Mike Smolen is interested in conservation and Alaskan community involvement in regional science studies. Cole Matson's research focuses on aquatic and evolutionary toxicology. He also has additional interests in phylogeography and conservation genetics. Rodney Honeycutt is a University Professor, whose research interests include population genetics, molecular evolution, and molecular phylogenetics. John Patton's major interests involve the processes of genomic differentiation among intraspecific and interspecific populations and how reticulate evolution affects these processes of differentiation. To achieve this in mammals his investigations focus on the evolutionary patterns of clonally inherited mtDNA, Y, and effectively haplodiploid X chromosomes and the subsequent contrasts of those patterns with patterns observed among autosomally inherited genes both above and below the level of species. John Bickham's research interests focus on genetic mutations and how they are produced and transmitted in individuals, populations, species and the evolutionary processes that affect genetic change. His current sponsored research projects include population genetics of Steller sea lions and bowhead whales, biodiversity studies in bats and ecotoxicological studies in contaminated environments in Azerbaijan. This paper is one in a series resulting from a long-term study (since 1992) on the population genetics and systematics of Steller sea lions sponsored mainly by NMFS. The genetic studies are part of an overall program studying the population trends, basic biology, and conservation status of this endangered species.

Appendix I

Frequency of HVRI/cytb haplotype linkages across the three genetically identifiable stocks. GenBank Accession numbers for HVRI haplotypes: AY340876–AY340937, FJ948491–FJ948546. GenBank Accession numbers for cytochrome *b* haplotypes: DQ144995–DQ145021, FJ948486–FJ948490

HVRI/cytb haplotype	Asian stock	Western Stock	Eastern Stock	Total
A1	6	23	5	34
A2			31	31
A3	3			3
A4	2			2
AAA27	6			6
AAAA1	6			6
AAAAAA1		3		3
AAAAAAA3			2	2
BB1	107	61	1	169
BB10			6	6
BB2	1	1	44	46
BB3	1	35	12	48
BB4	40	4		44
BB6			5	5
BB9	3			3
BBBB1	1			1
BBBBB3			1	1
BBBBBBB2			1	1
CC1	9	22	3	34
CC16	1			1
CCCCC3			1	1
CCCCCC3			4	4
CCCCCCC3			1	1
D3			1	1
DD1		7	2	9
DDD3		2		2
DDDD3	1	1		2
DDDDDD3			4	4
DDDDDDD1		1		1
E1	3	15	1	19
EE1		1		1
EEEE3	1			1
EEEEEEE3			1	1
FF1	5	4		9
FF10			2	2
FF4	2			2
FF6			4	4
FFF19		1		1
FFF3		1		1
FFFF1	5	2	2	9
FFFF23		1		1
FFFF36	1			1
FFFFFFF3			1	1
G1	3	6	1	10
GGG3		1		1
GGGG3			4	4
H3			62	62
H32			3	3
HH1		9		9
HHH1		1		1

© 2009 Blackwell Publishing Ltd

Appendix I *Continued*

HVRI/cytb haplotype	Asian stock	Western Stock	Eastern Stock	Total
HHHHHH35	2			2
I3			2	2
II1		1		1
III1	1	1		2
IIIII3			3	3
J3			1	1
JJJ3		2		2
JJJJJJ3	1			1
K3			26	26
KKK1	4	1		5
L3			8	8
LL3			7	7
LLL1	5			5
M3			6	6
MMM1	10			10
MMM40	1			1
N3			33	33
NNNNNN3			1	1
O3			2	2
OOO1		2		2
P2			6	6
Q2			14	14
Q31			3	3
QQQ1	1			1
R3			1	1
RRRR3			2	2
S1	7	78	8	93
S20		1		1
S22		3		3
S23		20	1	21
SS3			2	2

Appendix I *Continued*

HVRI/cytb haplotype	Asian stock	Western Stock	Eastern Stock	Total
SSSS2			1	1
SSSSS3	5			5
SSSSSS3			1	1
T1		2		2
T20		2		2
TTT4	1			1
TTTT1		1		1
U1	1	2		3
UUU27	4			4
UUUU3		1		1
VVVV3			10	10
VVVVV4	2			2
W1		7	1	8
WWW3	2			2
WWWW3			11	11
WWWWW1	2			2
X1	1			1
XXXX1		1		1
YY3			18	18
YYYY10			4	4
YYYY2			8	8
YYYY27	1			1
YYYY3			1	1
YYYY31			1	1
YYYYY1	1			1
Z1	4	1		5
Z3	6	29	4	39
ZZ3			3	3
ZZZ1	3	2		5
ZZZZZZ3			6	6
Total	272	359	400	1031

© 2009 Blackwell Publishing Ltd