J-STAGE Advance published date: May 24, 2021

NEW INSIGHTS INTO THE GENETIC STRUCTURE OF SEI WHALES (*BALAENOPTERA BOREALIS*) AT THE INTER-OCEANIC SCALE

Mioko TAGUCHI[1]*, Mutsuo GOTO[1], Lucas MILMANN[2, 3], Salvatore SICILIANO[4], Ralph TIEDEMANN[5] and Luis A. PASTENE[1, 6]

[1] *Institute of Cetacean Research, Toyomi-cho 4–5, Chuo-ku, Tokyo 104–0055, Japan*
[2] *Grupo de Estudos de Mamíferos Aquáticos do Rio Grande do Sul (GEMARS), Rua Bento Gonçalves 165, sala 1002, Torres, RS, 95560–000, Brazil*
[3] *Applied Ecology and Conservation Lab, Programa de Pos-graduação em Ecologia e Conservação da Biodiversidade, Departamento de Ciencias Biológicas, Universidade Estadual de Santa Cruz (UESC), Rodovia Jorge Amado, km 16, Ilhéus, BA, 45662–900, Brazil*
[4] *Fundação Oswaldo Cruz/Fiocruz, Av. Brasil, 4.365 & Grupo de Estudos de Mamíferos Marinhos da Região dos Lagos (GEMM-Lagos), Rio de Janeiro, RJ 21040–900, Brazil*
[5] *Unit of Evolutionary Biology/Systematic Zoology, Institute of Biochemistry and Biology, University of Potsdam, Karl-Liebknecht-Str. 24–25, Haus 26, 14476 Potsdam, Germany*
[6] *Project Microbiomes as Bioindicators of the Aquatic Ecosystem Health in Chilean Patagonia, Centro de Estudios del Cuaternario de Fuego-Patagonia y Antártica (CEQUA), Punta Arenas, Chile*
* *Corresponding author: taguchi@cetacean.jp*

Abstract

To describe global genetic diversities and genetic structure of sei whales, population genetic and phylogenetic analyses were performed using mitochondrial DNA (mtDNA) control region sequence (480 bp) data from specimens collected in three oceanic regions: North Pacific (NP: $n=39$), North Atlantic (NA: $n=84$) and Southern Hemisphere (SH: $n=6$). Microsatellite DNA (msDNA) analyses were also performed using genotype data at seventeen loci in a sub-set of samples ($n=39$ in NP and $n=4$ in SH). The haplotype (h) and nucleotide (π) diversities were higher in NP ($h=0.92$ and $\pi=0.009$) and SH ($h=1.00$ and $\pi=0.012$) than in NA ($h=0.68$ and $\pi=0.002$). The haplotype frequency was significantly different among the three oceanic regions, and the conventional pairwise F_{ST} estimates support the difference between NA and the other two populations. Furthermore, except for one haplotype, there were no other shared haplotypes among the three oceanic regions, suggesting contemporary migration and gene flow would be strongly restricted at inter-oceanic scales. This inference was also supported by the msDNA analyses. The haplotype genealogy reconstructed by the maximum-likelihood approach strongly supported two clusters, the first consisting of NA haplotypes, and the second consisting of NP and SH haplotypes. This genealogy was supported by the statistical parsimony haplotype network. These results indicated hierarchical genetic structuring of sei whales globally, in which whales in SH are genetically closer to NP whales than to NA whales. Based on a comparison of the inter-oceanic genetic structure and phylogeny of the sei whales with those of fin whales, which is another cosmopolitan baleen whale species, it is suggested that the genetic structure of sei whales reflects occasional gene flow between the Northern and Southern hemispheres and/or incomplete lineage sorting, similar to the case of fin whales.

Key words: sei whale, stock structure, control region, worldwide, phylogeography.

Introduction

The sei whale, *Balaenoptera borealis*, is one of the large baleen whales inhabiting all the major open oceans, except the northern Indian Ocean (Horwood, 1987; Rice, 1998). Sei whales live up to sixty years and their body length reaches up to 20 m. It is believed that they migrate from low-latitudes winter breeding grounds to summer feeding grounds in high-latitudes, although little is known about the migratory routes and the exact location of breeding grounds of this species. This migration pattern, coupled with asynchronous seasonal breeding cycles between the Northern and Southern hemispheres, should favor reproductive isolation and resultant genetic divergence of sei whales between the hemispheres, as observed in other cosmopolitan baleen whales, *e.g.*, fin whale, *Balaenoptera physalus* (Archer *et al.*, 2013) and humpback whale, *Megaptera novaeangliae* (Jackson *et al.*, 2014). Furthermore, the continental masses separating the North Atlantic from the North Pacific have probably prevented gene flow of sei whales between the two oceans since the closure of the Panama Seaway, which is believed to have occurred during the Pliocene (Coates *et al.*, 1992).

Previous published population genetic work of sei whales is limited, despite their global distribution. Pioneering work was carried out by Wada and Numachi (1991) based on allozymes. This study revealed allele frequencies of three polymorphic allozymes to be significantly different between Antarctic and North Pacific sei whales, without any further differentiation within the oceans. The genetic homogeneity within the western North Pacific was supported by the subsequent work by Kanda *et al.* (2006) using microsatellite DNA (msDNA) polymorphisms at seventeen loci. The most recent study used mitochondrial DNA (mtDNA) control region sequences and msDNA genotypes. The study found evidence of genetic structuring between North Pacific and North Atlantic sei whales, but not within the North Atlantic (Huijser *et al.*, 2018). Taking these previous findings together, it is highly possible that sei whales are genetically differentiated among North Pacific (NP), North Atlantic (NA) and oceans of the Southern Hemisphere (SH).

The objective of this study was to describe global genetic diversities and genetic structure of sei whales using mtDNA control region sequences and msDNA genotype data of this species worldwide. This is the first study to incorporate sei whale samples from the NP, NA and SH.

Materials and Methods

Laboratory procedures

Samples and DNA extraction

A total of 44 tissue samples of sei whales (Fig. 1) was collected from three sources: (1) whaling under the second phase of the Japanese Whale Research Program under Special Permit in the western North Pacific (JARPNII) in 2002 ($n=39$); (2) biopsy sampling under the Japanese Whale Research Program under Special Permit in the Antarctic (JARPA) in 2000/01 ($n=3$) and 2002/03 ($n=1$); (3) stranding at southern Brazil in 2015 ($n=1$, export and import CITES permits 18BR030112/DF and 18JP000006/TI, respectively). Note that the 39 samples collected in the western North Pacific were the same as those used in the previous msDNA analyses (Kanda *et al.*, 2006), but their nucleotide sequences at mtDNA control region were determined for the first time in the present study.

Sampled skin tissues were preserved in 99% ethanol or stored frozen at −20°C until use. Total genomic DNA was extracted from 0.05 g of skin tissue using either the standard phenol-chloroform method (Sambrook *et al.*, 1989) or the Gentra Puregene kits (QIAGEN). Extracted DNA was stored in TE buffer (10mM Tris-HCl, 1mM EDTA, pH 8.0).

MtDNA sequencing

For all samples subjected to the DNA extraction, 534 base pairs (bp) of the mtDNA control region were amplified by the polymerase chain reaction (PCR) using the set of primers MT4 (Árnason *et*

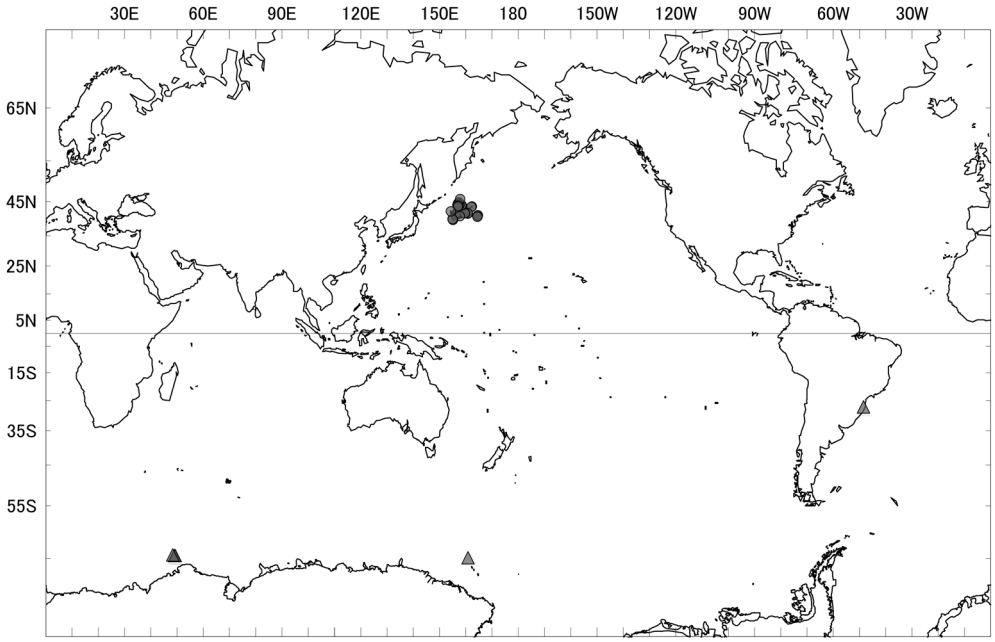

Fig. 1. Geographical position of samples newly sequenced and/or genotyped in the present study. Symbol shape indicates sample population: circle (NP, $n=39$), triangle (SH, $n=5$).

al., 1993) and Dlp5R (5′-CCATCGAGATGTCTTATTTAAGGGGAAC-3′). PCR was carried out in a 25 μL reaction mixture containing 10–100 ng of template DNA, 0.2 mM of dNTPs, 0.1 mM of each primer, 0.5 units of *EX Taq* DNA polymerase (TaKaRa), and 1× PCR buffer. Each reaction was performed with an initial denaturation step at 95°C for 5 minutes, followed by 30 cycles of 30 seconds at 94°C, 30 seconds at 50°C and 30 seconds at 72°C, with a final extension step at 72°C for 10 minutes. PCR products were purified using MicroSpin S-400HR columns (Pharmacia Biotech). Cycle sequencing was performed using BigDye terminator cycle sequence Kit (Applied Biosystems) and the PCR primers, following the protocols of the manufacturer. The cycle sequencing products were purified using AutoSeq G-50 spin Columns (Pharmacia Biotech). The labeled sequencing fragments from tissue samples from JARPNII and JARPA were resolved using an ABI PRISM 377, while the fragment of the stranding sample was sequenced with ABI3500 Genetic Analyzers (Applied Biosystems).

MsDNA genotyping

Three samples collected under the JARPA in 2000/01 and one stranding sample collected in southern Brazil in 2015 were genotyped at 17 nuclear msDNA loci, *i.e.*, EV1, EV14, EV21, EV94, EV104 (Valsecchi and Amos, 1996), GT011 (Bérubé *et al.*, 1998), GT23, GT211, GT271, GT310, GT575 (Bérubé *et al.*, 2000), GATA28, GATA53, GATA98, GATA417, GGAA520 (Palsbøll *et al.*, 1997), and DlrFCB17 (Buchanan *et al.*, 1996), in three multiplex fluorescent PCRs. Each of the multiplex PCRs was carried out in a 10 μL reaction mixture, containing 10–100 ng of template DNA, 5 μL of 2× Type-it Multiplex PCR master mix (QIAGEN), and 1 μL of primer mix (2 μM each primer), at 95°C for 5 minutes, followed by 28 cycles at 95°C for 30 seconds/54, 58 or 59.5°C for 90 seconds/72°C for 30 seconds, and a post-cycling extension at 60°C for 30 minutes. PCR products were electrophoresed on an ABI3500 DNA Analyzer (Applied Biosystems), and allele sizes were determined using a 600 LIZ size standard (Applied Biosystems) and GeneMapper v. 4.0 (Applied Biosystems).

Dataset

In order to investigate the global genetic structure, mtDNA and msDNA data were divided into three oceanic regions: (1) North Pacific (NP); (2) North Atlantic (NA); (3) Southern Hemisphere (SH).

MtDNA

Combining the 44 mtDNA control region sequences newly sequenced in this study with the 85 sequences generated in the previous studies (of which, 84 were from the Gulf of Maine, Iceland and Azores (Huijser *et al.*, 2018) and one from the Antarctic Ocean (Sasaki *et al.*, 2005)) resulted in a total of 129 mtDNA sequences of sei whales from NP (n=39), NA (n=84) and SH (n=6) (Table 1).

MsDNA

Combining msDNA genotypes of the four new individuals from SH (three from the Antarctic and one from Brazil), with a subset of genotype data in NP generated by Kanda *et al.* (2006), resulted in a total of 43 genotype set of sei whales from NP (n=39) and SH (n=4) (Table 1). The newly obtained msDNA scores were standardized to the previous scores, which Kanda *et al.* (2006) generated using the BaseStation100 DNA fragment analyzer (Bio-Rad), by comparing the scores of the same samples between the present and previous platforms.

Data analyses

MtDNA

Haplotype (h) and nucleotide (π) diversities with sample standard deviations (Nei, 1987) were estimated using the program ARLEQUIN v. 3.5.2.2 (Excoffier and Lischer, 2010).

The difference in mtDNA haplotype frequency among oceanic regions was tested using the Monte Carlo simulation-based chi-square test of independence with 10,000 replicates (Roff and Bentzen, 1989) in R (R Core Team, 2016). The conventional F_{ST} estimates between oceanic regions were calculated using 10,000 random permutations of the original dataset, as implemented in ARLEQUIN. The FDR correction (Benjamini and Hochberg, 1995) was used to adjust the statistical significance level in the pairwise estimates.

The mtDNA haplotype genealogy of sei whales was reconstructed using the maximum-likelihood approach (ML-tree) with 10,000 bootstrap resampling in the program MEGA ver. 10.0.5 (Kumar *et al.*, 2018). The best-fit nucleotide substitution model was determined based on the Bayesian Information Criterion using MEGA, and Tamura 3-parameter model (Tamura, 1992) with G=0.487 and I=0.747 was selected. Two sequences of *Balaenoptera edeni* (GenBank accession number: X72196) and *Balaenoptera brydei* (unpublished data) were used as outgroups in the analysis. The statistical parsimony network (Clement *et al.*, 2000) of mtDNA control region haplotypes was also depicted using the program PopART (Leigh and Bryant, 2015).

MsDNA

The departure from Hardy–Weinberg equilibrium (*HWE*) was tested and the inbreeding coefficient (F_{IS}; Weir and Cockerham, 1984) was estimated, using GENEPOP (Raymond and Rousset, 1995; Rousset, 2008). The number of alleles (A), allelic richness (AR) and expected heterozygosity (H_E)

Table 1. Numbers of mitochondrial (mt) and microsatellite (ms) DNA data used in this study. The mtDNA nucleotide sequences in NP were newly obtained from the same samples used for msDNA analyses in Kanda *et al.* (2006).

Oceanic region		MtDNA		MsDNA	
		n	Reference	*n*	Reference
North Pacific	(NP)	39	This study	39	Kanda *et al*. (2006)
North Atlantic	(NA)	84	Huijser *et al*. (2018)	–	
Southern Hemisphere	(SH)	5 1	This study Sasaki *et al*. (2005)	4	This study
Total		129		43	

Table 2. Summary of mtDNA variations of sei whales and haplotype frequency in each of and across the oceanic regions. Numbers in 'Variable site' represent nucleotide positions corresponding to 'H01' deposited in DDBJ under the accession no. LC629100. See Table 1 for abbreviations of the oceanic region.

Haplotype	Variable site																																Accession No.	Reference	Haplotype frequency			
	64	78	92	103	107	110	111	112	139	191	207	247	251	257	260	264	270	272	283	289	294	295	296	297	299	303	339	340	371	404	407	408			NP	NA	SH	Global
H01	T	T	G	C	T	C	A	C	A	T	T	T	A	A	T	C	C	A	C	G	C	T	G	T	C	A	T	T	C	T	C		LC629100	This study	1	0	0	1
H02	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	T	G	.	A	.	.	.	C	.	.	C	.	.	C	.		LC629101	This study	1	0	0	1
H03	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	T	G	.	A	.	.	.	C	.	.	C	.	.	.	.		LC629102	This study	8	0	0	8
H04	.	.	.	.	.	.	.	.	G	.	C	.	.	.	.	.	T	.	.	.	T	.	A	C	T	.	C	.	.	.	.		LC629103	This study	2	0	0	2
H05	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	T	.	.	A	.	.	.	C	T	.	C	C	.	.	.		LC629104	This study	2	0	0	2
H06	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	T	.	.	.	T	.	A	.	T	G	.	.	.	.	.		LC629105	This study	1	0	0	1
H07	.	.	.	.	.	.	.	.	G	.	C	.	.	.	.	.	T	G	.	.	.	.	A	.	.	.	C	.	.	.	.		LC629106	This study	7	0	0	7
H08	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	T	.	.	.	.	.	A	.	.	G	.	.	.	.	.		LC629107	This study	4	0	0	4
H09	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	T	.	.	A	.	.	.	C	.	.	C	.	.	.	.		LC629108	This study	3	0	0	3
H10	.	.	.	.	.	.	.	.	G	.	C	.	.	.	.	.	T	.	.	.	T	.	A	.	T	.	C	.	.	.	.		LC629109	This study	1	0	0	1
H11	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.		LC629110	This study	1	0	0	1
H12	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	T	.	.	.	.	.	.	C	.	.	C	.	T	.	.		LC629111	This study	1	0	0	1
H13	.	.	.	.	.	.	.	.	G	.	.	.	.	G	.	.	.	.	.	A	.	.	.	.	.	.	C	.	.	.	.		LC629112	This study	1	0	0	1
H14	.	.	.	.	C	.	.	.	G	.	.	.	.	.	.	.	T	.	.	A	.	.	.	.	.	.	G	.	.	.	.		LC629113	This study	1	0	0	1
H15	.	.	.	.	.	.	.	.	G	C	C	.	.	.	.	.	T	G	.	.	.	.	A	.	.	.	C	.	.	.	.		LC629114	This study	1	0	0	1
H16	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	A	.	.	.	C	.	.	C	.	.	.	.		LC629115	This study	1	0	0	1
H17	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	T	.	.	A	.	.	.	.	.	.	C	.	.	.	.		LC629116	This study	1	0	0	1
H18	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.		LC629117	This study	1	0	0	1
H19	.	.	.	.	.	.	.	.	G	.	C	.	.	.	.	.	T	.	.	.	.	.	A	.	.	.	C	.	.	.	.		LC629118	This study	1	0	0	1
H87	.	.	.	.	.	T	.	.	G	.	.	.	.	.	.	.	T	.	.	A	.	C	A	.	.	.	C	.	.	.	T		LC629119	This study	0	0	1	1
H88	.	.	.	.	.	T	.	.	G	.	.	.	.	.	.	T	T	.	.	A	T	.	A	.	.	.	C	.	.	.	.		LC629120	This study	0	0	1	1
H89	.	.	.	.	.	.	C	.	G	.	.	C	G	.	.	.	T	.	.	A	T	C	A	.	.	.	.	.	.	.	.		LC629121	This study	0	0	1	1
H90	.	.	.	.	.	T	.	.	G	.	.	.	.	.	.	T	.	.	.	A	T	.	A	.	T	.	C	.	.	.	.		LC629122	This study	0	0	1	1
SH1	.	.	.	.	.	.	.	T	G	.	.	.	.	.	.	.	T	.	.	.	.	.	A	.	T	.	C	.	.	.	.		NC_006929.1	Sasaki *et al*. (2005)	0	0	1	1
NA1	C	C	C	T	.	.	.	.	G	C	.	.	.	.	G	.	T	.	.	C	.	C	A	.	T	.	C	C	.	.	.		MH035689	Huijser *et al*. (2018)	0	9	0	9
NA2	.	C	C	T	.	.	.	.	G	C	.	.	.	.	G	.	.	.	.	C	.	C	A	.	T	.	C	C	.	.	.		MH035690	Huijser *et al*. (2018)	0	10	0	10
NA3	.	C	C	T	.	.	.	.	G	C	.	.	.	.	G	.	T	.	.	C	.	C	A	.	T	.	C	C	.	.	.		MH035691	Huijser *et al*. (2018)	0	43	0	43
NA4	.	C	C	T	.	.	.	.	G	.	.	.	.	.	G	.	T	.	.	C	.	C	A	.	T	.	C	C	.	.	.		MH035692	Huijser *et al*. (2018)	0	16	0	16
NA5	.	C	C	T	.	.	.	.	G	C	.	.	.	.	G	.	T	.	T	C	.	C	A	.	T	.	C	C	.	.	.		MH035693	Huijser *et al*. (2018)	0	2	0	2
NA6	.	.	.	.	.	T	.	.	G	.	.	.	.	.	.	T	.	.	.	A	T	C	A	.	.	.	C	.	.	.	.		MH035694	Huijser *et al*. (2018)	0	1	1	2
NA7	.	C	C	T	.	.	.	.	G	C	.	.	.	.	G	.	T	.	.	C	.	C	A	.	.	.	C	C	.	.	.		MH035695	Huijser *et al*. (2018)	0	3	0	3
																																		Total		39	84	6	129
																																		Number of haplotypes		19	7	6	31
																																		Haplotype diversity *(s.d.)*		0.92 (0.03)	0.68 (0.04)	1.00 (0.10)	0.86 (0.02)
																																		Nucleotide diversity *(s.d.)*		0.009 (0.005)	0.002 (0.002)	0.012 (0.008)	0.013 (0.007)

were estimated using the program FSTAT ver. 2.9.3.2 (Goudet, 1995).

Bayesian clustering analysis was performed to infer the most likely number of clusters using STRUCTURE 2.3.4 (Pritchard *et al.*, 2000). The analysis was conducted with ten independent runs for $K=2$–3. All runs were performed with 100,000 Markov chain Monte Carlo repetitions and 10,000 burn-in length using the admixture model with correlated allele frequencies. The web-based program STRUCTURE HARVESTER (Earl and von Holdt, 2012) was used to estimate the mean posterior probability of the data. Additionally, principal component analysis (PCA) was performed in R using the '*dudi.pca*' function in the *ade4* package (Chessel *et al.*, 2004; Dray *et al.*, 2007).

Results

Genetic variations

MtDNA

The mtDNA control region sequences (480 bp) of 129 sei whales from the three oceanic regions contained 31 variable nucleotide sites and a single indel (1 bp) defining 31 haplotypes, of which 23 were novel (deposited in GenBank with accession numbers: LC629100–LC629122) (Table 2). Apart from haplotype 'NA6', none of the haplotypes were shared among oceanic regions. Haplotype 'NA6' was shared between NA and SH (Table 2). The h and π for the total samples were 0.86 and 0.013, respectively (Table 2). These estimates were higher in SH (1.00 and 0.012) and NP (0.92 and 0.009) than in NA (0.68 and 0.002).

MsDNA

All 17 msDNA loci were polymorphic in a total of 43 sei whales in NP and SH, with locus-specific allele numbers from 3 alleles at GATA53 and GT271 to 17 alleles at DlrFCB17 (Table 3). The *AR* was higher in SH than in NP. No significant deviations from *HWE* were observed in each locus or across the loci of NP and SH (Table 3). Although the significant departure from *HWE* was detected at GATA417 when the two populations were combined (Table 3), the global test did not support the statistical significance.

Table 3. Summary statistics of 17 msDNA loci of sei whales: *A*, the number of alleles; *AR*, allelic richness; H_E, expected heterozygosity; F_{IS}, inbreeding coefficient; *HWE*, *p*-value of the Hardy-Weinberg equilibrium test. F_{IS} and *HWE* at GT011 in SH were incomputable since one of two alleles was represented by only one copy. See Table 1 for abbreviations of the oceanic region. Bold text indicates the statistical significance at $\alpha=0.05$.

Locus	NP (n = 39)					SH (n = 4)					Total (n = 43)				
	A	*AR*	H_E	F_{IS}	*HWE*	*A*	*AR*	H_E	F_{IS}	*HWE*	*A*	*AR*	H_E	F_{IS}	*HWE*
DlrFCB17	15	5.42	0.87	0.026	0.686	6	6.00	0.92	-0.091	1.000	17	5.45	0.87	0.010	0.565
EV1	12	4.65	0.81	0.015	0.455	6	6.00	0.92	0.182	0.466	15	4.99	0.84	0.054	0.225
EV14	14	5.35	0.87	-0.063	0.913	7	7.00	0.96	-0.043	1.000	14	5.49	0.88	-0.057	0.749
EV21	6	3.35	0.64	-0.085	0.754	4	4.00	0.75	0.000	1.000	7	3.59	0.68	-0.029	0.831
EV94	6	3.20	0.61	-0.087	1.000	6	6.00	0.92	0.182	0.446	8	3.59	0.67	-0.015	0.443
EV104	5	3.80	0.75	-0.064	0.730	5	5.00	0.83	-0.200	1.000	5	3.90	0.76	-0.069	0.681
GATA28	10	4.73	0.82	0.001	0.749	6	6.00	0.92	-0.091	1.000	11	4.85	0.83	-0.005	0.761
GATA53	3	2.17	0.35	0.123	0.241	2	2.00	0.58	0.571	0.426	3	2.25	0.37	0.191	0.075
GATA98	6	3.70	0.73	0.119	0.454	2	2.00	0.50	-0.500	1.000	6	3.60	0.71	0.084	0.482
GATA417	7	4.00	0.76	0.023	0.143	5	5.00	0.79	0.053	0.775	10	4.49	0.80	0.072	**0.007**
GGAA520	8	4.26	0.80	0.098	0.190	4	4.00	0.79	0.368	0.302	8	4.33	0.80	0.132	0.062
GT011	4	2.59	0.47	0.021	0.572	2	2.00	0.25	–	–	5	2.59	0.46	0.028	0.660
GT023	6	2.80	0.49	-0.040	0.380	5	5.00	0.88	-0.143	1.000	7	3.01	0.54	-0.029	0.177
GT211	4	2.11	0.33	0.069	0.249	4	4.00	0.79	-0.263	1.000	5	2.39	0.41	0.081	0.141
GT271	3	1.55	0.15	-0.051	1.000	2	2.00	0.58	0.571	0.429	3	1.78	0.23	0.297	0.073
GT310	3	2.14	0.39	-0.256	0.275	4	4.00	0.71	-0.412	1.000	5	2.42	0.43	-0.245	0.499
GT575	4	2.51	0.56	-0.006	0.913	4	4.00	0.79	-0.263	1.000	6	2.85	0.61	0.011	0.422
Average			0.61	-0.004	0.884			0.76	-0.010	1.000			0.68	0.020	0.070

Table 4. Conventional pairwise F_{ST} estimates between oceanic regions for mtDNA. Asterisks show the statistical significance after FDR correction: *$p<0.05$, **$p<0.01$, ***$p<0.001$. See Table 1 for abbreviations of the oceanic regions.

Oceanic region	NP	NA	SH
NP			
NA	***0.210		
SH	0.048	**0.206	

Table 5. Summary of STRUCTURE analyses showing the average ln Pr(X|K) for each of *K*, and respective probability.

K	ln Pr(X|K)	Variance	P(K|X)
1	-2122.47	53.79	~1.00
2	-2138.42	193.24	~0.00
3	-2233.04	436.19	~0.00

Genetic differentiation and structure

MtDNA

The haplotype frequency was significantly different among the three oceanic regions ($\chi^2=246.48$, $p<0.001$). Conventional pairwise F_{ST} estimates were significantly different from zero in the pairs of NP and NA as well as SH and NA, but no significant difference was seen between SH and NP (Table 4).

MsDNA

The STRUCTURE analyses conducted for different sampling partitions without information on their geographic origins presented the highest probability at $K=1$ (Table 5). However, the four sei whales from the SH were distinguishable from NP whales by high posterior probabilities ($q>0.8$) of belonging to a particular cluster at $K=2$ (Fig. 2).

In PCA analysis, the first (PC1) and second (PC2) principal components explained 9.81% and 6.95% of the total variation, respectively. This analysis suggests that the 43 sei whales could be divided into two distinct clusters (NP and SH) along the PC1 axis (Fig. 3).

Phylogenetic analyses

The ML-tree showed two clusters supported by high bootstrap values (Fig. 4a), one which consisted of haplotypes found in NA and the other of haplotypes found in NP and SH. Within the NP/SH cluster, the ML-tree also showed a sub-cluster containing four of the six haplotypes found in SH, *i.e.*, 'NA6', 'H87', 'H88' and 'H90'. However, this cluster had a low bootstrap value (Fig. 4a).

The statistical parsimony network showed the two clusters were separated by six mutational steps, which consisted of a star-like genealogy of haplotypes found in NA, and more complex structures in NP and SH (Fig. 4b). The haplotype network also showed that all haplotypes found in SH were separated from those in NP by several mutational steps.

The haplotype 'NA6' found in both NA and SH was located in the NP/ SH cluster in each topology (Fig. 4).

Discussion

This is the first population genetic study of sei whales worldwide. In particular, analyses using msDNA has never been reported in SH sei whales. Notwithstanding the small number of samples used from SH, the preliminary findings of this study provide some new insights into the genetic structure and phylogeography of this species.

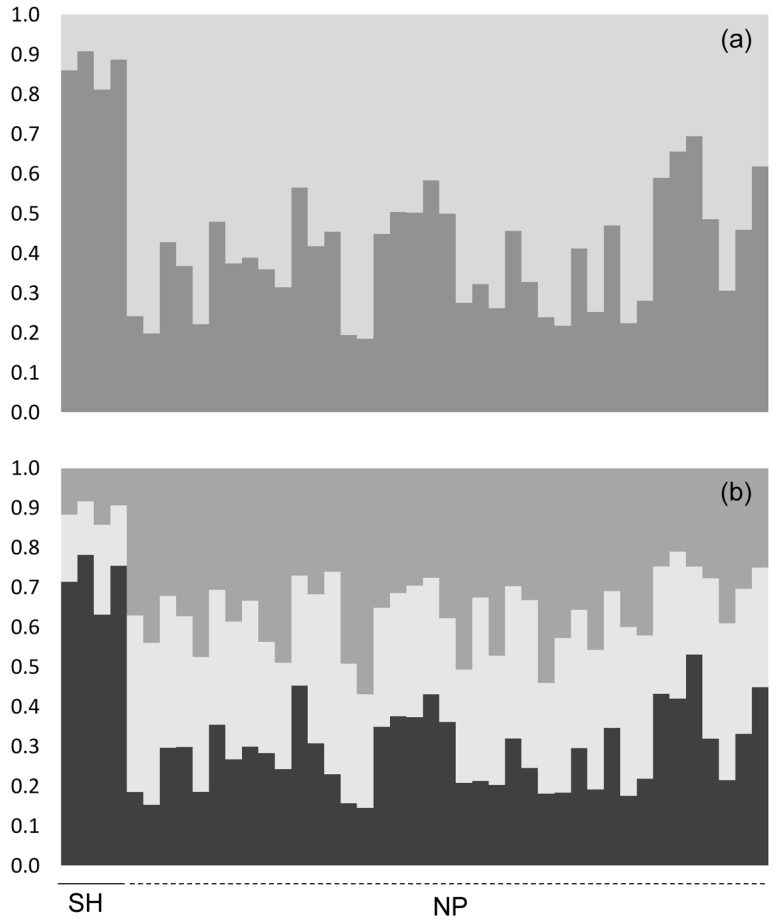

Fig. 2. Bar plot of posterior probabilities in STRUCTURE analyses for sei whales in oceans of the Southern Hemisphere (SH) and the North Pacific (NP): (a) $K=2$ and (b) $K=3$. Each individual is characterized by a thin vertical line, which is divided into K colored segments on the basis of the individual's membership fractions in K clusters.

Genetic structure

The haplotype frequency of sei whales was significantly different among the three oceanic regions, and the conventional pairwise F_{ST} estimates suggested the genetic differentiation of this species between NA and the other two oceanic regions. This observation is consistent with the pattern of genetic differentiation of sei whales between NP and NA presented by Huijser *et al.* (2018). Although the F_{ST} estimates did not support a genetic difference between NP and SH, given no haplotype sharing, the insignificant result was likely to be caused by the small sample size in SH. In fact, Wada and Numachi (1991) using a larger number of samples demonstrated the significant genetic differentiation between NP and SH sei whales. Furthermore, this inference was supported by the msDNA analyses. The STRUCTURE analysis suggested that the most likely number of populations at Hardy-Weinberg/linkage equilibrium in the data set consisting of sei whales in SH and NP was one, which was consistent with no deviations from *HWE* in each of and across populations. However, despite the small sample size in SH, the four SH sei whales were distinguishable from the NP whales by high posterior probabilities ($q>0.8$) when $K=2$ was assumed. This was also supported by the result of multivariable analysis, *i.e.*, PCA, separating the four sei whales from other NP whales along the first PC axis.

The PCA analysis also showed a large genetic variation of SH cluster consisting of samples collected across the vast oceanic region, *i.e.*, the western South Atlantic and the Antarctic Ocean. This, coupled with high h in SH, might be suggestive of an inclusion of samples derived from multiple breeding populations in a single sample population. Given the genetic structuring shown in other SH

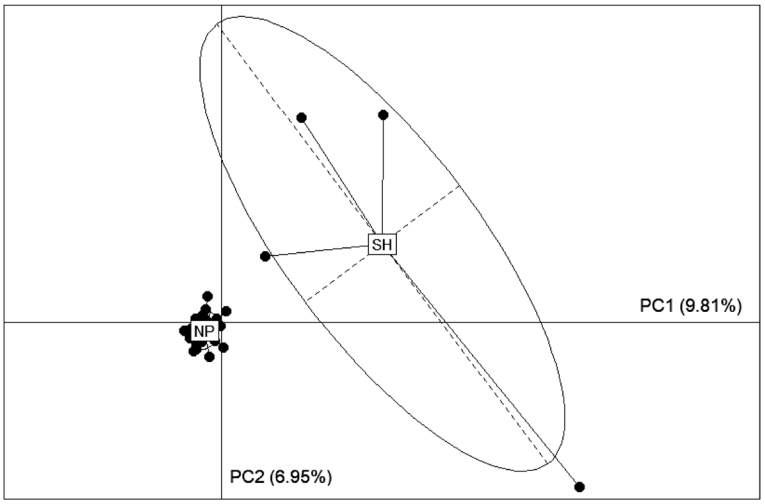

Fig. 3. Result of PCA analysis on genotypes at seventeen msDNA loci. See Table 1 for abbreviations of the oceanic regions.

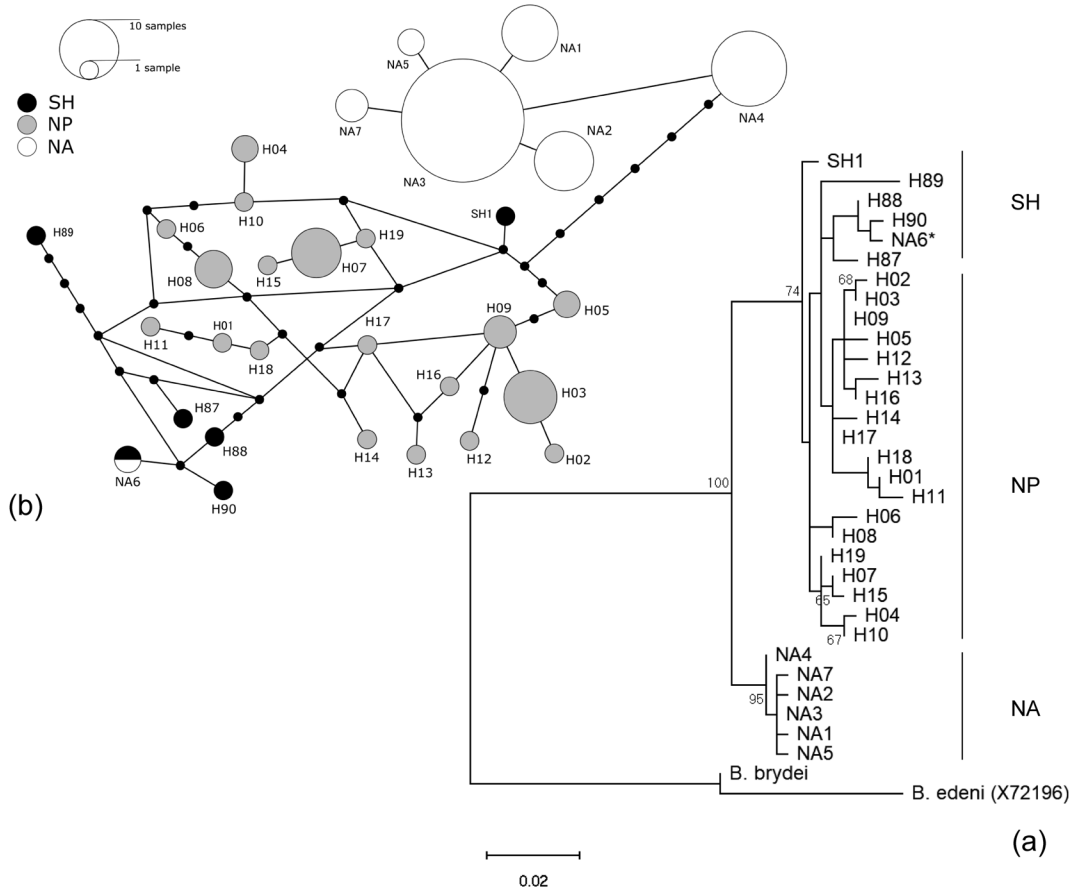

Fig. 4. Genealogy reconstructed by the maximum-likelihood approach (a), and statistical parsimony network (b) of mtDNA control region haplotypes of sei whales. The phylogenetic tree was rooted by outgroup of *B. brydei* and *B. edeni*, which was drawn to scale with branch lengths measured in the number of substitutions per site. Only bootstrap values above 60% are shown. Asterisk shows haplotypes shared among oceanic regions. Circles of the haplotype network represent different haplotypes. The colors and sizes of circles refer to the geographical origin and abundance of haplotypes, respectively. Small black circles indicate intermediate haplotypes not found in this study. See Tables 1 and 2 for abbreviations of the geographical origin of haplotypes and haplotype information, respectively.

baleen whales (*e.g.*, humpback whales, Olavarría *et al.*, 2007; Antarctic minke whales, Pastene and Goto, 2016), that possibility could not be excluded at this stage.

The haplotype network and ML-tree strongly supported the divergence of haplotypes found in NA from those in NP and SH, as shown by Huijser *et al.* (2018). The haplotype network further showed that all haplotypes found in SH were separated from those in NP by several mutational steps. The ML-tree revealed a sub-cluster containing most of the haplotypes found in SH, however, there was no significant bootstrap support. These results are indicative of ocean-specific mitochondrial lineages, which are compatible with the observed difference in the degrees of the pairwise F_{ST} estimates among sei whales from the three oceanic regions.

Taking all the present and previous findings together, sei whales appear to be significantly differentiated among oceanic regions hierarchically, and whales in SH are more closely related to whales in NP than to whales in NA. Except for haplotype 'NA6', there were no other haplotypes that were shared among the three oceanic regions, which suggests that contemporary migration and gene flow would be strongly restricted at inter-oceanic scales.

Interestingly, the pattern of inter-oceanic genetic structuring observed in the present study was similar to that in fin whales with the lower F_{ST} estimates between NP and SH (mtDNA, 0.005–0.106; single-nucleotide polymorphisms (SNPs), 0.098) than between NA and NP (mtDNA, 0.018–0.198; SNPs, 0.1668) as well as between NA and SH (mtDNA, 0.017–0.121; SNPs, 0.1447) (Archer *et al.*, 2019). This, coupled with genealogical concordance between fin and sei whales described below, suggests that the pattern of genetic structuring in sei whales could be attributed to historical events, *i.e.*, recent occasional gene flow between the Northern and Southern hemispheres and/or incomplete lineage sorting (ILS).

Phylogeographic inferences

The ML-tree found no evidence of divergence of SH haplotypes from NP ones, despite the inference of genetic differentiation between the two oceanic regions. ILS and/or recent occasional gene flow may explain this observation, because they can be the cause of shared genetic variations among populations after their divergence. The mtDNA phylogenetic study of fin whales demonstrated a polyphyletic pattern of NP and SH haplotypes (Archer *et al.*, 2013), which was similar to that of the sei whale. The study also suggests the possibilities of a relatively recent introgression between the two oceans and/or retention of ancestral polymorphisms due to ILS of the large whale species.

Huijser *et al.* (2018) inferred that an expansion of the North Atlantic sei whale population occurred after the last glacial maximum (LGM, 26.5–19 thousand years ago; Clark *et al.*, 2009) based on phylogeographic analyses. The star-like genealogy of haplotypes and lower genetic diversities observed in NA in the present study would be attributed to this historical event in the case of the NA sei whales. On the other hand, SH and NP haplotypes showed more complex structures with a high number of haplotypes of intermediate frequency on the network. Given the inferences of ILS with high h in NP and SH, the result could reflect ancestral polymorphisms of NP and SH whales in a large long term effective population size. It seems that they have been separated long enough to be void of shared haplotypes, but not long enough to have evolved reciprocal mitochondrial monophyly. Such reciprocal monophyly is indeed observed among NA and the combined NP/SH cluster, suggesting a considerably more ancient divergence.

In the present study, there are also signs of recent or contemporary occasional inter-oceanic migration/dispersal events, one from SH to NA (NA6) and one from NP to SH (SH1). This inference is compatible with the restricted but rare inter-equatorial gene flow inferred in humpback whales (*e.g.*, Jackson *et al.*, 2014) and fin whales (Cabrera *et al.*, 2019). So far, there is no direct evidence for trans-equatorial gene flow in sei whales. However, this seems probable in this species, given its high mobility as demonstrated by tagging studies (*e.g.*, Olsen *et al.*, 2009) and continuous distribution across the equatorial waters inferred from stranding records along the northern coast of Brazil (*e.g.*, Costa *et al.*, 2017; Mayorga *et al.*, 2020; Milmann *et al.*, 2020).

Conclusion

Notwithstanding the small sample size used from SH, the present study demonstrated the hierarchical genetic structuring of sei whales globally for the first time, with high genetic diversity in SH/NP, in which whales in SH are genetically closer to NP whales than to NA whales. These findings, coupled with the previous findings in sei whales as well as in other cosmopolitan baleen whales, suggested the following phylogeographical scenarios of this species: (1) a shorter history of divergence among SH and NP, leading to retention of ancestral polymorphisms due to ILS in SH and NP, (2) occasional gene flow between the Northern and Southern hemispheres.

Extended genetic analyses using larger sample sizes across the oceans and more genetic loci have to be conducted in this species to investigate finer genetic structure and demographic estimates, *e.g.*, migration rate and effective population size.

Acknowledgements

We thank all researchers, crew members and colleagues who contributed to the collection of genetic samples used in this study. We are grateful to Hiroyuki Oikawa (ICR) for his assistance in the laboratory work, and to Naohisa Kanda (JANUS) for preparing the msDNA data used in this study. We would also like to thank the anonymous reviewers for their helpful comments on the manuscript.

References

Archer, F.I., Brownell Jr., R.L., Hancock-Hanser, B.L., Morin, P.A., Robertson, K.M., Sherman, K.K., Calambokidis, J., Urbán, R.J., Rosel, P.E., Mizroch, S.A., Panigada, S. and Taylor, B.L. 2019. Revision of fin whale *Balaenoptera physalus* (Linneaus, 1758) subspecies using genetics. *J. Mammal.* 100: 1653–1670.

Archer, F.I., Morin, P.A., Hancock-Hanser, B.L., Robertson, K.M., Leslie, M.S., Bérubé, M., Panigada, S. and Taylor, B.L. 2013. Mitogenomic phylogenetics of fin whales (*Balaenoptera physalus* spp.): Genetic evidence for revision of subspecies. *PLoS One* 8: e63396.

Árnason, Ú., Gullberg, A. and Widegten, B. 1993. Cetacean mitochondrial DNA control region: Sequences of all extant baleen whales and two sperm whale species. *Mol. Biol. Evol.* 10: 960–970.

Benjamini, Y. and Hochberg, Y. 1995. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. R. Statist. Soc. B* 57: 289–300.

Bérubé, M., Aguilar, A., Dendanto, D., Larsen, F., Notarbartolo-di-Sciara, G., Sears, R., Sigurjónsson, J., Urban-Ramirez, J. and Palsbøll, P.J. 1998. Population genetic structure of North Atlantic, Mediterranean Sea and Sea of Cortez fin whales, *Balaenoptera physalus* (Linnaeus, 1758); analysis of mitochondrial and nuclear loci. *Mol. Ecol.* 7: 585–599.

Bérubé, M., Jørgensen, H., Mcewing, R. and Palsbøll, P.J. 2000. Polymorphic di-nucleotide microsatellite loci isolated from the humpback whale, *Megaptera novaeangliae*. *Mol. Ecol.* 9: 2181–2183.

Buchanan, F.C., Friesen, M.K., Littlejohn, R.P. and Clayton, J.A. 1996. Microsatellites from beluga whale *Delphinapterus leucas*. *Mol. Ecol.* 5: 571–575.

Cabrera, A.A., Hoekendijk, J.P.A., Aguilar, A., Barco, S.G., Berrow, S., Bloch, D., Borrell, A., Cunha, H.A., Dalla Rosa, L., Dias, C.P., Gauffier, P., Hao, W., Landry, S., Larsen, F., Martín, V., Mizroch, S., Oosting, T., Øien, N., Pampoulie, C., Panigada, S., Prieto, R., Ramp, C., Rivera-Léon, V., Robbins, J., Ryan, C., Schall, E., Sears, R., Silva, M.A., Urbán, J., Wenzel, F.W., Palsbøll, P.J. and Bérubé, M. 2019. Fin whale (*Balaenoptera physalus*) mitogenomics: A cautionary tale of defining sub-species from mitochondrial sequence monophyly. *Mol. Phylogenet. Evol.* 135: 86–97.

Chessel, D., Dufour, A. and Thioulouse, J. 2004. The ade4 Package – I: One-Table Methods. *R News* 4: 5–10.

Clark, P.U., Dyke, A.S., Shakun, J.D., Carlson, A.E., Clark, J., Wohlfarth, B., Mitrovica, J.X., Hostetler, S.W. and McCabe, A.M. The last glacial maximum. *Science* 325: 710–714.

Clement, M., Posada, D. and Crandall, K.A. 2000. TCS: a computer program to estimate gene genealogies. *Mol. Ecol.* 9: 1657–1659.

Costa, A.F., Siciliano, S., Emin-Lima, R., Martins, B.M.L., Sousa, M.E.M., Giarrizzo, T. and Silva Júnior, J.S. 2017. Stranding survey as a framework to investigate rare cetacean records of the north and north-eastern Brazilian coasts. *ZooKeys* 688: 111–134.

Coates, A.G., Jackson, J.B.C., Collins, L.S., Cronin, T.M., Dowsett, H.J., Bybell, L.M., Jung, P. and Obando, J.A. 1992. Closure of the Isthmus of Panama: The near-shore marine record of Costa Rica and western Panama. *Geol. Soc. Am. Bull.* 104: 814–828.

Dray, S., Dufour, A. and Chessel, D. 2007. The ade4 Package – II: Two-Table and K-Table Methods. *R News* 7: 47–52.

Earl, D.A. and von Holdt, B.M. 2012. STRUCTURE HARVESTER: A website and program for visualizing STRUCTURE

output and implementing the Evanno method. *Conserv. Genet. Resour.* 4: 359–361.
Excoffier, L. and Lischer, H.E.L. 2010. Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Mol. Ecol. Resour.* 10: 564–567.
Goudet, J. 1995. FSTAT (Version 1.2): A computer program to calculate F-statistics. *J. Hered.* 86: 485–486.
Horwood, J. 1987. *The Sei Whale: Population Biology, Ecology and Management.* Croom Helm, London. 375 pp.
Huijser, L.A.E., Bérubé, M., Cabrera, A.A., Prieto, R., Silva, M.A., Robbins, J., Kanda, N., Pastene, L.A., Goto, M., Yoshida, H., Vikingsson, G.A. and Palsbøll, P.J. 2018. Population structure of North Atlantic and North Pacific sei whales (*Balaenoptera borealis*) inferred from mitochondrial control region DNA sequences and microsatellite genotypes. *Conserv. Genet*. 19: 1007–1024.
Jackson, J.A., Steel, D.J., Beerli, P., Congdon, B.C., Olavarría, C., Leslie, M.S., Pomilla, C., Rosenbaum, H. and Baker, C.S. 2014. Global diversity and oceanic divergence of humpback whales (*Megaptera novaeangliae*). *Proc. Biol. Sci.* 281: 20133222.
Kanda, N., Goto, M. and Pastene, L.A. 2006. Genetic characteristics of western North Pacific sei whales, *Balaenoptera borealis*, as revealed by microsatellites. *Mar. Biotechnol*. 8: 86–93.
Kumar, S., Stecher, G., Li, M., Knyaz, C. and Tamura, K. 2018. MEGA X: Molecular Evolutionary Genetics Analysis across computing platforms. *Mol. Biol. Evol.* 35: 1547–1549.
Leigh, J.W. and Bryant, D. 2015. POPART: Full-feature software for haplotype network construction. *Methods Ecol. Evol.* 6: 1110–1116.
Mayorga, L.F.S.P., Vanstreels, R.E.T., Bhering, R.C.C., Mamede, N., Costa, L.M.B., Pinheiro, F.C.F., Reis, L.W.D., Trazzi, A., Meirelles, W.L.C., Ribeiro, A.M. and Siciliano, S. 2020. Strandings of cetaceans on the Espírito Santo coast, southeast Brazil, 1975–2015. *ZooKeys* 948: 129–152.
Milmann, L., Siciliano, S., Morais, I., Tribulato, A.S., Machado, R., Zerbini, A.N., Baumgarten, J.E. and Paulo Henrique, O. 2020. A review of Balaenoptera strandings along the east coast of South America. *Reg. Stud. Mar. Sci.* 37: 101343.
Nei, M. 1987. *Molecular Evolutionary Genetics.* Columbia University Press, New York. 512 pp.
Olavarría, C., Baker, C.S., Garrigue, C., Poole, M.M., Hauser, N., Caballero, S., Florez-Gonzalez, L., Brasseur, M., Bannister, J., Capella, J., Clapham, P.J., Dodemont, R., Donoghue, M., Jenner, C., Jenner, M.N., Moro, D., Oremus, M., Paton, D., Rosembaum, H.C. and Russell, K. 2007. Population structure of South Pacific humpback whales and the origin of the eastern Polynesian breeding grounds. *Mar. Ecol. Prog. Ser.* 330: 257–268.
Olsen, E., Budgell, W.P., Head, E., Kleivane, L., Nøttestad, L., Prieto, R., Silva, M., Skov, H., Víkingsson, G., Waring, G. and Øien, N. 2009. First satellite-tracked long-distance movement of a sei whale (*Balaenoptera borealis*) in the North Atlantic. *Aquat. Mamm.* 35: 313–318.
Palsbøll, P.J., Bérubé, M., Larsen, A.H. and Jørgensen, H. 1997. Primers for the amplification of tri-and tetramer microsatellite loci in baleen whales. *Mol. Ecol.* 6: 893–895.
Pastene, L.A. and Goto, M. 2016. Genetic characterization and population genetic structure of the Antarctic minke whale *Balaenoptera bonaerensis* in the Indo-Pacific region of the Southern Ocean. *Fish. Sci.* 82: 873–886.
Pritchard, J.K., Stephens, M. and Donnelly, P. 2000. Inference of population structure using multilocus genotype data. *Genetics* 155: 945–959.
R Core Team. 2016. R: *A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna.
Raymond, M. and Rousset, F. 1995. GENEPOP (Version 1.2): Population genetics software for exact tests and ecumenicism. *J. Hered.* 86: 248–249.
Rice, D.W. 1998. *Marine Mammals of the World: Systematics and Distribution.* Society for Marine Mammalogy, Lawrence, KS. 231 pp.
Roff, D.A. and Bentzen, P. 1989. The statistical analysis of mtDNA polymorphisms: χ^2 and the problem of small samples. *Mol. Biol. Evol.* 6: 539–545.
Rousset, F. 2008. Genepop'007: A complete re-implementation of the genepop software for Windows and Linux. *Mol. Ecol. Resour.* 8: 103–106.
Sambrook, J., Fritsch, E.F. and Maniatis, T. 1989. *Molecular Cloning: A Laboratory Manual.* 2nd Ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY. 1546 pp.
Sasaki, T., Nikaido, M., Hamilton, H., Goto, M., Kato, H., Kanda, K., Pastene, L.A., Cao, Y., Fordyce, R.E., Hasegawa, M. and Okada, N. 2005. Mitochondrial phylogenetics and evolution of mysticete whales. *Syst. Biol*. 54: 77–90.
Tamura, K. 1992. Estimation of the number of nucleotide substitutions when there are strong transition-transversion and G+C-content biases. *Mol. Biol. Evol.* 9: 678–687.
Valsecchi, E. and Amos, W. 1996. Microsatellite markers for the study of cetacean populations. *Mol. Ecol.* 5: 151–156.
Wada, S. and Numachi, K. 1991. Allozyme analyses of genetic differentiation among the populations and species of the *Balaenoptera*. *Rep. Int. Whal. Comm.* (Special Issue) 13: 125–154.
Weir, B.S. and Cockerham, C.C. 1984. Estimating F-statistics for the analysis of population structure. *Evolution* 38: 1358–1370.

Received: November 18, 2020
Accepted: March 22, 2021
Published online: May 24, 2021