

PRIMER NOTE

Characterization of 18 SNP markers for sperm whale (*Physeter macrocephalus*)

PHILLIP A. MORIN,*†NICOLA C. AITKEN,‡NADIA RUBIO-CISNEROS,*‡ANDREW E. DIZON* and SARAH MESNICK*

*Southwest Fisheries Science Center, National Marine Fisheries Service, 8604 La Jolla Shores Drive, La Jolla, CA 92037, USA,

†Laboratory for Conservation Genetics, Max Planck Institute for Evolutionary Anthropology, Deutscher Platz, 04103 Leipzig, Germany,

‡Scripps Institution of Oceanography, UCSD, 9500 Gilman Drive, La Jolla, CA 92093, USA

Abstract

We report the characterization of 18 new single nucleotide polymorphism (SNP) markers for an endangered species, the sperm whale (*Physeter macrocephalus*), developed using a targeted gene approach. SNP markers were derived from autosomal regions of the genome using primers originally characterized for genome mapping in other mammals. These SNP markers are the first to be designed for genotyping sperm whale populations and will provide a necessary addition to the genetic tools employed for understanding population structure on a global scale and for developing a conservation management strategy for this endangered species.

Keywords: single nucleotide polymorphism, SNP, sperm whale, *Physeter macrocephalus*, conservation, genetic

Received 5 September 2006; revision accepted 6 November 2006

The sperm whale has been the target of two large-scale waves of commercial whaling, the most recent of which ended in the 1980s. Given their worldwide pelagic distribution, long submergence times, nomadic lifestyle, and uncertain catch records, current and historical abundances are unknown and difficult to assess (Whitehead 2002). They remain listed as endangered under the US Endangered Species Act and are currently protected from commercial hunting under International Whaling Commission treaties. Despite this, there is no clear picture of the worldwide stock structure (Default *et al.* 1999). Genetic studies to date reveal low genetic differentiation among ocean basins and little evidence of subdivision within ocean basins, with the exception of some isolated basins such as the Mediterranean and Gulf of Mexico (Dillon 1996; Lyrholm & Gyllensten 1998; Bond 1999; Lyrholm *et al.* 1999). More extensive studies using existing microsatellite markers (Lyrholm *et al.* 1999) may provide more power to detect population structure. However, microsatellites are fraught with

known technical difficulties, and complicated evolutionary mechanisms may limit the information obtained (Navidi *et al.* 1992; Callen *et al.* 1993; Balloux & Lugon-Moulin 2002). To complement existing mtDNA and microsatellite markers for this species, we have used a targeted gene approach (Aitken *et al.* 2004) to detect single nucleotide polymorphisms (SNPs) in a panel of sperm whale samples. We have identified a suite of SNP markers, and developed genotyping assays for 18 independent loci using two genotyping technologies.

Two hundred and two comparative anchor tagged sequence (CATS; Lyons *et al.* 1997) primer pairs were used to screen genomic DNA from an individual sperm whale for amplification, following the methods of Aitken *et al.* (2004), but with the exception that only a single magnesium chloride concentration (1.5 mM) was tested. Primers yielding a single polymerase chain reaction (PCR) product were selected for amplification from six individual DNA samples and a pool containing DNA from 20 individuals representing a broad geographical distribution. Although our intent was to use the pooled sample to identify rarer SNPs, or SNPs common in some populations but not others, most SNPs were ascertained using the six individual samples, and the pooled samples served only to confirm the presence of some SNPs.

Correspondence: Phillip Morin, Fax: 858-546-7003;

E-mail: phillip.morin@noaa.gov

‡Current address: Institute for Applied Ecology, University of Canberra, ACT 2601, Australia

Sixty of 84 products amplified reliably across the sample set. We obtained high-quality sequence data from 44 of these loci (Table 1), and screened these for SNPs using the alignment and sequence quality analysis software POLYPHRED in the software package CONSED (Nickerson *et al.* 1997; Gordon *et al.* 1998). SNPs were considered high quality (likely to be real and not sequencing artefacts) if they were seen in both directions, or in only one direction if in a region of good quality sequence and the complementary sequence for that region was not available.

We identified 39 SNPs in 23 of the 44 loci encompassing a total length of 21 063 bp of high quality DNA sequence. This equates to an average of one SNP every 540 bp, though SNPs were not evenly distributed and tended to be clustered in some loci (Table 1). To avoid linked loci, we selected one SNP from each locus and designed assays (Table 2) for allele-specific amplification using either Luminex X-map or Amplifluor technologies (or both in some cases; see Table 2), following the manufacturer's design protocols. SNPs were selected based on a variety of practical factors, such as whether sufficient nonvariable flanking sequence was available for primer design, and whether high quality primers could be designed for genotyping. In a few cases where multiple suitable SNPs existed in a fragment, an assay was designed for an arbitrarily chosen SNP, or a SNP with more balanced frequencies of the two alleles in the sequenced samples.

The Luminex system requires pre-amplification of each locus using flanking PCR primers, followed by allele-specific primer extension (ASPE) with two allele-specific primers and incorporation of biotin-labelled nucleotides, with subsequent binding of ASPE products to streptavidin phycoerythrin for fluorescent detection (Taylor *et al.* 2001). Allele-specific primers each have a unique oligonucleotide tag on the 5' end that binds to complementary tags on coloured microspheres for flow-cytometric separation of beads in a multiplex assay, and simultaneous measurement of allele-specific fluorescence. We followed standard Luminex protocols for amplification and ASPE. Five assays were optimized for Luminex genotyping, with all but PND being PCR amplified in a single multiplex reaction, and PND being amplified separately and then combined with the multiplex PCR products prior to ASPE.

Amplifluor technology is simpler, but doesn't allow multiplexing. Briefly, genomic DNA or pre-amplified PCR product is amplified using two allele-specific primers containing 5'-Amplifluor-complimentary oligonucleotide tags and a common reverse primer, plus two universal Amplifluor primers, which contain a hairpin loop and a fluorescent reporter and quencher (Myakishev *et al.* 2001; Giancola *et al.* 2006). After initial allele-specific amplification using the tagged primers and common reverse primer, the Amplifluor primers can bind to the template and produce PCR products that incorporate and linearize the Amplifluor

sequences, causing an increase in fluorescence. Fluorescence can be detected in real time using a quantitative PCR instrument, or after completion of the PCR using a fluorometer. We used standard amplification protocols for all assays, varying only the annealing temperatures in the first or second PCR cycle set, the magnesium chloride concentration of the Amplifluor 10X buffer (Chemicon International), and the primer concentrations (Table 1). Pre-amplification was carried out for several assays that generally performed poorly with genomic DNA (see Table 1), or for samples that performed poorly for all assays. When pre-amplified product was used as the template for the Amplifluor reaction, only the second set of PCR cycles was used for allele-specific amplification and detection.

We genotyped DNA samples of sperm whales from the eastern Pacific Ocean, ranging from regions near the continental margins of South, Central, and North America westward to Hawaii. We also included five or six previously sequenced samples as controls in most genotyping batches. Of 21 assays tested (five with Luminex, 16 with Amplifluor), 18 were optimized and successfully produced genotypes. Of the three remaining, two (TS1 and IFN1) produced predominantly heterozygote-like signals, which were probably due to co-amplification of duplicated genes in the genome that differed at the SNP site. The third (NPPA) appeared homozygous, and re-analysis of the DNA sequences indicated that the 'SNP' was probably a sequencing artefact.

Frequencies of the minor allele ranged from zero to 50% in the eastern Pacific (> 0% including the worldwide control samples) (Table 1). No loci were found to exhibit significant linkage disequilibrium. As would be expected when relatively few samples are used to ascertain SNPs, the majority of our SNP loci had a relatively high frequency of the minor allele. This ascertainment bias will need to be kept in mind when using these SNPs for some types of population analysis (e.g. analyses of demographic patterns; Morin *et al.* 2004).

Acknowledgements

We are grateful to the Max Planck Institute for Evolutionary Anthropology, the Marine Mammal Commission, and the Southwest Fisheries Science Center for support of this research. We thank Amanda Bowman for assistance with alignment and editing of DNA sequences, and Lara Asato for technical support. Sperm whale DNA samples were obtained from many sources, and without their generous donation of samples we could not do this research. Samples were generously supplied by Chief Scientists Jay Barlow, Lisa Balance, Barbara Taylor, Susan Chivers, and Tim Gerrodette of the SWFSC cetacean cruises; John Calambokidis/Cascadia Research; Nancy Black and Richard Ternullo; John E. Heyning, Dave Janiger and Jim Dines; Daniel Palacios/the crew of Ocean Alliance's R/V Odyssey (D.M. Palacios, I. Kerr, K. Marshall, E. Lettevall, K. Rogers, D. Day, B. Brennan, P. Rodriguez, R. Wallace, and L. Galley); Jan Hodder and Michael Graybill; Craig Matkin,

628 PRIMER NOTE

Table 1 List of 44 CATS loci sequenced, the number of SNPs identified, and GenBank Accession numbers for consensus sequences. N, number of samples genotyped. Two PCR cycle sets are used for Amplifluor genotyping (see Amplifluor kit protocols). All 20x primer mixes for Amplifluor genotyping contained 0.5 μ M of each allele-specific (AS) primer and 7.5 μ M of the common reverse (R) primer unless stated otherwise. T_a , annealing temperature

Locus	Sequence length	No. of SNPs found	N	Minor allele frequency	Amplifluor buffer [MgCl ₂] mM	Cycle 1 set T_a (in °C)	Cycle 2 set T_a (in °C)	20x primer mix, comments	GenBank Accession no.
ACTC	466	1	89	0.6%	1.5	61	55		EF087933
ADH2	285								EF087934
ADRB2	454								EF087935
ADRBK1	311	1	134	0.8%					EF087936
AMBP	428								EF087937
C5	595								EF087938
CAT	670	2	142	24.7%	1.5	55	55	1 μ M G-allele AS, 0.5 μ M A-allele AS, 7.5 μ M R	EF087939
CHRNA1	372	2	144	22.6%	1.8	55	55		EF087940
CHY	674								EF087941
CK	541	1	144	43.1%	1.8	55	55	0.5 μ M T-allele AS, 2 μ M G-allele AS, 7.5 μ M R	EF087942
CKMM	595								EF087943
COL3A1	198								EF087944
COL9A1	292								EF087945
CSF2	881	3	155	4.84%†	1.8	55	58		EF087946
DRD2	1193	4	109	47.3%	2.5	55	55		EF087947
ELN	239	1	146	40.1%	1.5	55	55		EF087948
EPO	428	3	150	20.7%	2.5	55	55		EF087949
F9	339	1	141	39.7%	1.5	55	55		EF087950
FES	465								EF087951
G6PD	518								EF087952
GAPD	141								EF087953
GLB1	233								EF087954
GLUT2	344								EF087955
GRP	677	1	118	3.4%	2.5	55	55	PCR mix includes 0.5 μ M GC-rich additive from Amplifluor kit	EF087956
HOX7	149								EF087957
IFN1	355	5	—	—	1	63	55	Assay failed to resolve genotypes (see text)	EF087959
IFNG	320	1	159	9.4%	1.5	55	58		EF087958
INT	512	3	146	50.0%†	2.5	55	55		EF087960
MGF†	588	(> 6)							EF087961
MPO	615								EF087962
MYL4	458								EF087963
NF1	377								EF087964
NPPA	295	1	—	—	1.8	61	55	Assay failed to resolve genotypes (see text)	EF087965
P4HB	290								EF087966
PIT1	756								EF087967
PKM	577	1	158	18.0%	1.5	55	58		EF087968
PND	373	1	142	47.9%	1.5		55		EF087969
RDS	505	1	141	15.3%	2.5	57	55		EF087970
RYR2	470	1	127	35.0%	1.8		58	0.5 μ M A-allele AS, 1 μ M G-allele AS, 7.5 μ M R	EF087971
SPTBN1	568	2	148	21.6%	1.8	55	55		EF087972
SST	855								EF087973
TCRA	149								EF087974
TS	757	2	—	—	1	55	55	Assay failed to resolve genotypes (see text)	EF087975
WT1*	755	1							EF087976

*SNP located between single and dinucleotide repeats, preventing assay design; †SNPs in one individual; possible technical artefact; ‡significant deviation from Hardy–Weinberg expectations.

Table 2 SNP genotyping primers. The polymorphic nucleotide is shown underlined in each allele-specific primer. Note: (i) the polymorphic nucleotide is not always at the most 3' site, (ii) sometimes primers were designed for the complementary strand. Absence of flanking primers for the Amplifluor assay means that pre-amplification was not used for this locus. Only RYR2 required pre-amplification for all samples. Reverse primer used in pre-amplification is sometimes the same or similar to the reverse primer used in the Amplifluor reaction

Locus	SNP location	SNP type	Flanking primers	Allele-specific primers	Reverse primer
Amplifluor assays					
ACTC	392	C/A		#GAAAGTCTCAAACATGATCTGCAAT *AAAGTCTCAAACATGATCTGCAAG	CTGGGCCCTTCCCTTTGATTAA
CAT	456	A/G	TCCCTTCAGGTTCCGTGTTTG GGTGAGTGCATTGGTTGAAA	#TTTTCGCCAGGAAAGACTCA <u>A</u> *TGCCAGGAAAGACTCA <u>T</u>	GTGAGTGCATTGGTTGAAACTTT
CHRNA1	111	C/T	AACGTAAGCTCTGTGGCTTGA GATGACTTGCTAAGATGGCCTA	*GTTTGGCTTGATGGGAAGGC #GGTTTGGCTTGATGGGAAGG <u>T</u>	GGCACATCTTACGTACAGTAGCT
CK	273	G/T		#TTAAGAACCGCCCTTCCC <u>G</u> *CTAAGAACCGCCCTTCCC <u>T</u>	ACAGGTGGTGGCAGGATT
CSF2	278	A/G	CATCTCAAGAGGGTCCCAGT CTTTGCTCACACAGCAGGTC	#GAGGACAGTGACCTCTGTTTTCT <u>T</u> *AGGACAGTGACCTCTGTTTTCC	AGCACAGCCACATTTCCCTT
DRD2	679	C/T	GAGCCTTTTACTAATTTGCACAC GGTTTAGGTCTCGTTCAGCA	*TGTAATTTGTTTATGATGCCACACG #ATGGTAATTTGTTTATGATGCCACAC <u>A</u>	TACACCCTTTGCCGGATTCTTT AGCGGCACCTGAGAAGAG
ELN40	209	G/T		#GTCTGACCCCATCC <u>G</u> *CGTCTGACCCCATCC <u>T</u>	
EPO	292	C/T		*TGAAGCTAGGTAAGGGG <u>C</u> #GGAAGCTAGGTAAGGGG <u>T</u>	GAAATGCACAAGCCTGGAGT
F9	80	T/C	CGAAAAAGAAGACAGGTTAATG CCATTTCAGATGCAGAGCAAA	#TCGATTTGATTCCCTTCTCTATTGTAACATT <u>C</u> *CGATTTGATTCCCTTCTCTATTGTAACATT <u>T</u>	GGGAACCATACTTGCCTTTGGAA
GRP	190	T/C		#TCCTTTCAAGGTTTCGACCTGGT <u>C</u> *CCTTTCAGGTTTCGACCTGGT <u>T</u>	CCCCCCCCACTTTCTTTTTTT
IFN1	93	G/C	GACAATCTCCCAGGCACAAAG AAGCCTGTCTGATGCAGGA	*GAAGTATTTCTTCACAGCCAGC #TGAAGTATTTCTTCACAGCCAGG	CTGCTGAAGGAGGACTCCAT
IFNG	234	C/T	CCTGTGACTATTTCACTTGACCC CCTAGTTGGCCCTGAGATA	#GCCCTGAGATAAAGCCPTG <u>T</u> *TTGGCCCTGAGATAAAGCCPT <u>A</u>	ACTCACTAGGCAAGTCTATGTGATT
INT	368	G/C	TCAAAAATCCAGGTGGAGAGA CAGGGCACTGTCAGAGTCAG	*GCACAGAGGAAAGGGAGG #GCACAGAGGAAAGGGAGG	CAAATCTCTCGGGACAGTT
NPPA	253	T/C	GACGAGCGTCTTATCAAGC GTCCGTTGGTGCTGAAGTTT	#CCATGCGTTAAATTTTAAATCACCTGTACT <u>T</u> #CCCATGCGTTAAATTTTAAATCACCTGTACT <u>T</u>	GTCCGTTGGTGCTGAAGTTTAT
PKM	237	C/T	CTTCCTTAGCAGAGCGTCTCA AGACTTGGCCAGCCCTCTAT	#CTTTTTGACATGCTCTGTAC <u>A</u> *CTTTTTGACATGCTCTGTAC <u>G</u>	AGACTTGGCCAGCCCTCTAT
PND	111	G/A	GTAATAGCTGGACCTCCGCA CCCAGAGAGATGGGGTG	*AGGAGGGCAGATCTATCGG <u>A</u> #TTCAGGAGGGCAGATCTATCG <u>A</u>	ACCCAGCCCAGAGAGAT
RDS	456	G/T		*CCATCGACATGCTGCAAATTGAG #ACCATCGACATGCTGCAAATTGAT	TTGTTGCCACAGCACTTGAA
RYR2	327	G/A	GGTGGAAAGGATGAGCAGAA CACGTATCTCTAGGGAGCAGC	#AATTCCTGTTTCGGAATGGAGAACAT *TTCCTGTTTCGGAATGGAGAAC <u>C</u>	GGTGGAAAGGATGAGCAGAA
SPTBN1	279	C/G	ATGGGTTATGTGTGGCAGTG CCAAGCAGCACTCAAGTACG	*AGGTCCATTTTCAGAAATAGCAAGACAG #TAGGTCCATTTTCAGAAATAGCAAGAC <u>C</u>	GAGATGCAGTAGGGTAGCCCTT
TS1	308	T/C	GCGGGCTTTCTCTAGTTGC CCCGTGTGCCACAACACTACT	#AACTACTGAAGCCCGTG <u>C</u> *CAACTACTGAAGCCCGTG <u>C</u>	CTCATTTGCTGTGGCTTCTCTTT
Luminex assays					
ADRBK1	131	A/G	CTTGAATCCTAGGTTCGGCTG CACGAGGAAGGTGAGGGTC	LUA57-CTGTTTCCGATCTTTAAAGC <u>A</u> LUA8-CTGTTTCCGATCTTTAAAGC <u>G</u>	
CSF2	278	T/C	CATCTCAAGAGGGTCCCAGT CTTTGCTCACACAGCAGGTC	LUA65-GAGGACAGTGACCTCTGTTTTCT <u>T</u> LUA64-GGACAGTGACCTCTGTTTTCC	
IFNG	234	T/C	CCTGTGACTATTTCACTTGACCC CCTAGTTGGCCCTGAGATA	LUA12-CTCACTAGGCAAGTCTATGTGATTAT LUA20-CTCACTAGGCAAGTCTATGTGATTAC	
PKM	237	A/G	CTTCCTTAGCAGAGCGTCTCA AGACTTGGCCAGCCCTCTAT	LUA33-TCTTTTGGACATGCTCTGTAC <u>A</u> LUA44-TCTTTTGGACATGCTCTGTAC <u>G</u>	
PND	111	G/A	GTAATAGCTGGACCTCCGCA CCCAGAGAGATGGGGTG GAAGGTTCGGAGTCAACGGATT GAAGGTGACCAAGTTCATGCT	LUA6-AGGAGGGCAGATCTATCGG LUA28-CAGGAGGGCAGATCTATCG <u>A</u>	

*JOE tail.

#FAM tail.

Eva Saulitis and Harald Kalve; Daniel Odell and William B. Brooks; Sylviane Jaume, Ursula Gonzalez and Jorge Urbán; Diane Gendron; Frances Gulland; Deborah Thiele; Luella Dolar; and Lars Kleivane. All samples were collected by or transferred to the SWFSC under NMFS/MMPA and CITES permits.

References

- Aitken N, Smith S, Schwarz C, Morin PA (2004) Single nucleotide polymorphism (SNP) discovery in mammals: a targeted-gene approach. *Molecular Ecology*, **13**, 1423–1431.
- Balloux F, Lugon-Moulin N (2002) The estimation of population differentiation with microsatellite markers. *Molecular Ecology*, **11**, 155–165.
- Bond J (1999) Genetical analysis of the sperm whale using microsatellites. PhD Dissertation, Cambridge University.
- Callen DF, Thompson AD, Shen Y *et al.* (1993) Incidence and origin of 'null' alleles in the (AC)_n microsatellite markers. *American Journal of Human Genetics*, **52**, 922–927.
- Dillon MC (1996) *Genetic structure of sperm whale populations assessed by mitochondrial DNA sequence variation*. PhD Dissertation, Dalhousie University.
- Dufault S, Whitehead H, Dillon M (1999) An examination of the current knowledge on the stock structure of sperm whales (*Physeter macrocephalus*) worldwide. *Journal of Cetacean Research and Management*, **1**, 1–10.
- Giancola S, McKhann HI, Berard A *et al.* (2006) Utilization of the three high-throughput SNP genotyping methods, the GOOD assay, Amplifluor and TaqMan, in diploid and polyploid plants. *Theoretical and Applied Genetics*, **112**, 1115–1124.
- Gordon D, Abajian C, Green P (1998) CONSED: a graphical tool for sequence finishing. *Genome Research*, **8**, 195–202.
- Lyons LA, Laughlin TF, Copeland NG *et al.* (1997) Comparative anchor tagged sequences (CATS) for integrative mapping of mammalian genomes. *Nature Genetics*, **15**, 47–56.
- Lyrholm T, Gyllensten U (1998) Global matrilineal population structure in sperm whales as indicated by mitochondrial DNA sequences. *Proceedings of the Royal Society of London. Series B, Biological Sciences*, **265**, 1679–1684.
- Lyrholm T, Leimar O, Johannesson B, Gyllensten U (1999) Sex-biased dispersal in sperm whales: contrasting mitochondrial and nuclear genetic structure of global populations. *Proceedings of the Royal Society of London. Series B, Biological Sciences*, **266**, 347–354.
- Morin PA, Luikart G, Wayne RK, SNP-Workshop_Group (2004) SNPs in ecology, evolution and conservation. *Trends in Ecology & Evolution*, **19**, 208–216.
- Myakishev MV, Khripin Y, Hu S, Hamer DH (2001) High-throughput SNP genotyping by allele-specific PCR with universal energy-transfer-labeled primers. *Genome Research*, **11**, 163–169.
- Navidi W, Arnheim N, Waterman MS (1992) A multiple-tubes approach for accurate genotyping of very small DNA samples by using PCR: statistical considerations. *American Journal of Human Genetics*, **50**, 347–359.
- Nickerson DA, Tobe VO, Taylor SL (1997) POLYPHRED: automating the detection and genotyping of single nucleotide substitutions using fluorescence-based resequencing. *Nucleic Acids Research*, **25**, 2745–2751.
- Taylor JD, Briley D, Nguyen Q *et al.* (2001) Flow cytometric platform for high-throughput single nucleotide polymorphism analysis. *Biotechniques*, **30** (661–666), 668–669.
- Whitehead H (2002) Estimates of the current global population size and historical trajectory for sperm whales. *Marine Ecology-Progress Series*, **242**, 295–304.