Conserv Genet (2010) 11:1143–1146
DOI 10.1007/s10592-009-9903-3

TECHNICAL NOTE

Anonymous nuclear markers for cetacean species

A. R. Amaral · M. C. Silva · L. M. Möller · L. B. Beheregaray · M. M. Coelho

Received: 19 March 2009 / Accepted: 21 March 2009 / Published online: 2 April 2009

Abstract Here we report the development and characterization of 17 anonymous nuclear markers for cetacean species. These markers were isolated from a genomic library built from a common dolphin (genus *Delphinus*), and tested across several families within Cetacea. An average of 1 SNP per 272 bp was found in 10 anonymous markers screened for polymorphism within the genus *Delphinus* (total of 6,537 bp sequenced). These markers represent a significant addition to the set of tools used in genetic studies of cetaceans where population and species boundaries have to be inferred in order to implement proper conservation strategies.

Keywords SNPs · Anonymous loci · Common dolphins · Delphinids

Mitochondrial DNA (mtDNA) and microsatellites are amongst the most common classes of markers in ecological and conservation genetic studies (Beheregaray 2008). However, some caveats exist when mtDNA and microsatellite data are compared. The mtDNA genome evolves as a single unit, which yields a single gene tree, no matter how many base pairs or genes are sequenced. This warrants the need to use multiple nuclear sequence loci in studies where parameters such as effective population sizes and coalescent times are to be estimated (Ballard and Whitlock 2004). Microsatellites, on the other hand, are very popular because of their high variability and power to resolve population structure. However, the mutation models associated with these markers are not well understood in some cases and are not comparable to the mutation model of single nucleotide substitutions per nucleotide of mtDNA genes (Takezaki and Nei 1996). Additionally, using microsatellite frequencies to infer the phylogeny as part of a phylogeographic study involves making a number of heroic assumptions involving clustering individuals into populations for analysis, and rooting the resulting distance trees (Zink and Barrowclough 2008).

Recently, with an increasing number of genome sequencing projects underway for model organisms, single nucleotide polymorphisms (SNPs) have become markers of choice for a number of studies due to several advantages compared to microsatellites (Brumfield et al. 2003; Morin et al. 2004). In non-model organisms, SNPs have been traditionally obtained following two different approaches: a targeted gene approach in which primers can be designed from conserved regions of aligned sequences of at least two species in order to amplify a less conserved region [CATS, (Lyons et al. 1999)]; and an alternative approach where a genomic library is constructed with randomly sheared DNA and loci obtained by cloning sequencing (Rosenblum et al. 2007). This latter approach is more appealing for non-model organisms where hardly any sequence information exists.

The Family Delphinidae is the largest and most diverse family within the Order Cetacea, with currently 37 species (Caballero et al. 2008). It is a group with a complex

A. R. Amaral (✉) · M. C. Silva · M. M. Coelho
Centro de Biologia Ambiental, Departamento de Biologia Animal, Faculdade de Ciências da Universidade de Lisboa, Campo Grande, 1749-016 Lisbon, Portugal
e-mail: aramaral@fc.ul.pt

A. R. Amaral · L. M. Möller · L. B. Beheregaray
Department of Biological Sciences, Macquarie University, Sydney, NSW 2109, Australia

L. M. Möller · L. B. Beheregaray
School of Biological Sciences, Flinders University, Adelaide, SA 5001, Australia

Table 1 Primer sequences (5′–3′) for 17 anonymous nuclear loci and PCR annealing temperatures (°C) across several cetacean species

Locus ID	Accession numbers	Primer sequence (5′–3′)	PCR annealing temperature			
			Del.	*P.pho.*	*K.bre*	*B.acu*
Del_01	FJ490557	GAGCCTCACTTGGAAACTGG GCTGGTGGATAGGCAAAATG	64	64	64	64
Del_02	FJ490558	TGACTCCATGCCTCCTCTCT AGACGGTGAGGCCAATTTTT	64	64	64	64
Del_03	FJ490559	TAGGGAGTGAGGGAGCTCAG TCTTCACCAACCCTTCCAGT	64	64	64	64
Del_04	FJ490560	GCTGTATAACAAATGACCCCAGT CAGATCACATCTGGGGGAAC	64–60	64–60	n/a	n/a
Del_05	FJ490561	TACAGAAAGCCCATGTGCAG CGGTGGCATTTCTAAAAGGA	60	60	n/a	60
Del_06	FJ490562	TAAAGCCCCAGAGATTTGGA CGAATTCGCCCTTCACTTA	64	64	64	64
Del_07	FJ490563	TCGCAGCTGCTGTTTGTTAG TGGCTTGGTAGTTCAGAGACC	64	64	64	64
Del_08	FJ490564	TAGCTCTTGAGCGAATGCAA TGGACCTAGCCTTGTTAATGC	64–60	64–60	64–60	n/a
Del_09	FJ490565	TTCAAATTGGAAAGGAAGAGG GTGGAATTGGGAGCAATGAT	60	60	60	60
Del_10	FJ490566	CAGATATTGGAACTTCCCTGGT TTTCCAAAAAGCCAGATGGT	60	60	60	n/a
Del_11	FJ490567	CACAAATCTGAGGAACACACAAA TTGTAAAGCCTTATAAATTTCAGGTTA	64–60	64–60	64–60	64–60
Del_12	FJ490568	GGAGGTAGGGACCACACTGA AGAATGATCCGCTCCAAATG	60	60	60	n/a
Del_13	FJ490569	ACAAAATGTCCCACAGCGTA TTAATAGCTTCCGGGGATGG	60	60	60	60
Del_14	FJ490570	TGGGTCCCAGAAGAAGAACA TCTCTTAGCTTTTGCTTGTCTGT	64	64	64	64
Del_15	FJ490571	ACAAAACCTCGTTGGTCCAG GGTTGACAGCTTGCCATGT	64	64	64	64
Del_16	FJ490572	TCTATATAAAATCTGTTGAGTCCCTTT CAGAGCAACAACACATTTAGGG	60	60	60	60
Del_17	FJ490573	TTCTCTGTCTGACTGACTTCACTG CCATCCTGTTAAATGCCTTG	60	60	60	n/a

Del. Delphinidae, *P.pho Phocoena phocoena*, *K.bre Kogia breviceps*, *B.acu Balaenoptera acutorostrata*, *n/a* not applicable

evolutionary history, with many phylogenetic relationships still unresolved. Several species and/or populations are facing threats such as pollution, by-catch, food depletion and global warming, which warrant the need to develop new molecular markers that can help to understand and define biological boundaries, so that proper conservation strategies can be designed and implemented. Here we describe the development and characterization of multiple anonymous nuclear markers that have the potential to be used for phylogenetic, phylogeographic and population genetic structure studies of several cetacean species.

We built a genomic library for an individual short-beaked common dolphin (*Delphinus delphis*), which stranded in the Portuguese coast, using the TOPO® Shotgun Subcloning Kit (Invitrogen). Genomic DNA was extracted following standard phenol–chloroform procedures and was then sheared with a nebulizer. DNA fragments, 2 to 4 kb in size, were then blunted with T4 DNA and Klenow polymerases, dephosphorylated with calf intestinal phosphatase, ligated into a vector (pCR®4Blunt-TOPO®) and then transformed into *Escherichia coli* competent cells.

We sequenced 30 random clone inserts with vector primers, used Sequencher (version 4.2 Gene Codes Corporation) to visualize sequences and performed a BLAST search in GenBank to confirm their suitability for population genetic studies by ruling out the possibility they encode proteins or other conserved regions. Most loci remained anonymous with the exception of one that exhibited high percentage match to a known gene in other mammals. We also used the RepeatMasker program to screen sequences for interspersed repeats (http://www.repeatmasker.org). In several loci, repetitive elements were found, namely short interspersed elements (SINEs) and long interspersed elements (LINEs). Roughly half of the higher eukaryotic genome is composed of a variety of repetitive sequences with no obvious function (Ray 2007), so this finding was not surprising.

We designed primers for 17 clones using the program Primer 3 (Rozen and Skaletsky 2000), which were then tested using a polymerase chain gradient (PCR) gradient thermocycle (MyCycler, Biorad) with annealing temperatures ranging from 55 to 64°C. Primer sequences, annealing temperatures and approximate fragment length can be found in Table 1. Standard PCR conditions were applied to all reactions: 25-µl reactions containing 10–100 ng DNA, 0.2 mM each dNTP, 0.3 µM each primer, 1 U *Taq* Polymerase and 1× *Taq* buffer. Thermocycle profiles included 5 min at 94°C, followed by 35 cycles of 30 s at 94°C, 30 s at specified annealing temperature, 45 s at 72°C and a final extension step of 10 min at 72°C. Amplification was tested in a panel of 41 individuals from several delphinid cetacean species: *D. delphis*, *D. capensis*, *Stenella coeruleoalba*, *S. frontalis*, *S. attenuata*, *S. longirostris*, *Tursiops truncatus T. aduncus*, *Lagenodelphis hosei*, *Sousa chinensis*, *Sotalia fluviatilis* and *Globicephala melas*. We also tested amplification in representatives of other cetacean families: *Phocoena phocoena* (Phocoenidae), *Kogia breviceps* (Kogidae) and *Balaenoptera acutorostrata* (Balaenopteridae), to determine how primers would perform across families. When specific annealing temperatures failed to amplify in more distantly related species, a touchdown PCR profile was used [5 min at 94°C, followed by 10 cycles of 30 s at 94°C, 30 s at 64–60°C (decrease of 0.5°C per cycle) and 45 s at 72°C, followed by 35 cycles of 30 s at 94°C, 30 s at 60°C and 45 s at 72°C; and final extension step of 10 min at 72°C]. We obtained high levels of cross-amplification success, with almost all loci being amplified in all species tested. Ten loci were randomly chosen to screen for a panel of individuals from the genus *Delphinus*. This panel consisted of 12 *D. delphis* (4 from NE Atlantic, 2 from NW Atlantic, 2 from West Pacific and 4 from Eastern Tropical Pacific), 4 *D. capensis* (from the Eastern Tropical Pacific) and 1 *D. capensis tropicalis* (from the Arabian Sea). These loci were directly sequenced in both direction (BigDye Terminator CycleSequencing: Applied Biosystems) on an ABI 3730xl automated sequencer (Applied Biosystems). Sequencher (version 4.2, Gene Codes Corporation) was used to visualize sequences. Gametic phase was resolved computationally using PHASE v. 2.1.1 (Stephens and Donnelly 2003).

Table 2 Summary statistics for the 10 anonymous nuclear loci screened for variation in 17 *Delphinus* species

Locus ID	Product size (bp)	*S*	π	GC content (%)
Del_02	829	2	0.00030	31.26
Del_04	636	0	0.0000	47.67
Del_05	723	1	0.00016	41.35
Del_08	768	3	0.00106	28.94
Del_10	401	3	0.00180	35.97
Del_11	571	1	0.00021	32.95
Del_12	736	5	0.00148	53.73
Del_15	356	0	0.0000	36.40
Del_16	782	7	0.00350	56.97
Del_18	735	2	0.00090	46.75
Total	6,537	24	0.00101	41.20

S segregating sites, π nucleotide diversity

A total of 24 variable sites, or SNPs, were found in 6,537 bp sequenced with an average of 1 SNP/272 bp (Table 2). This level of nucleotide variation is higher to the one described for sperm whales [average of 1 SNP/540 bp (Morin et al. 2007)], but it falls within the range of 1 SNP in every 200–500 bp given by Morin et al. (2004) for terrestrial mammals. However it is lower than the levels described for birds [1 SNP/175 bp, (Primmer et al. 2002)] and for reptiles [1 SNP/30 bp, (Rosenblum et al. 2007)]. Polymorphism was not uniformly distributed across all loci, with some loci showing no polymorphism (Table 2). This pattern may be due to differences in nucleotide composition and genomic location, which may influence substitution rates among loci. The two loci showing higher levels of polymorphism also had the higher percentage of CG content (Table 2). Genomic regions rich in CpG islands are highly susceptible to mutation through methylation, where the cytosine mutates to a thymine, as opposed to genomic regions rich in A/T which are less prone to mutations (Han et al. 2008).

The minimum number of recombination events within loci (Hudson and Kaplan 1985) was evaluated using DNAsp v. 4.5 (Rozas et al. 2003), but no recombination was detected. The non-random association between polymorphisms at different loci was measured by the degree of linkage disequilibrium (LD), also using DNAsp. After correcting for multiple comparisons, no significant LD was observed among loci, indicating that the 10 loci analyzed are not likely to be linked.

The primers presented here, obtained with a random fragment genomic library approach, represent a good alternative for rapid marker development and SNP discovery in cetaceans and will help addressing questions in phylogenetics and population structure in cetaceans.

Acknowledgments We thank Kelly Robertson (Southwest Fisheries Research Center, NOAA) for providing samples of several delphinid cetacean species. A. R. Amaral is supported by a PhD grant (SFR/BD/27245/2006) and M. C. Silva is supported by a post-doctoral grant (SFRH/BPD/22276/2005), both funded by Fundação para a Ciência e Tecnologia (Portugal).

References

Ballard JWO, Whitlock MC (2004) The incomplete natural history of mitochondria. Mol Ecol 13:729–744. doi:10.1046/j.1365-294X.2003.02063.x

Beheregaray LB (2008) Twenty years of phylogeography: the state of the field and the challenges for the Southern Hemisphere. Mol Ecol 17:3754–3774

Brumfield RT, Beerli P, Nickerson DA, Edwards SV (2003) The utility of single nucleotide polymorphisms in inferences of population history. Trends Ecol Evol 18:249–256. doi:10.1016/S0169-5347(03)00018-1

Caballero S, Jackson J, Mignucci-Giannoni AA, Barrios-Garrido H, Beltran-Pedreros S, Montiel-Villalobos MG, Robertson KM, Baker CS (2008) Molecular systematics of South American dolphins Sotalia: sister taxa determination and phylogenetic relationships, with insights into a multi-locus phylogeny of the Delphinidae. Mol Phylogenet Evol 46:252–268. doi:10.1016/j.ympev.2007.10.015

Han L, Su B, Li WH, Zhao ZM (2008) CpG island density and its correlations with genomic features in mammalian genomes. Genome Biol 9:R79. doi:10.1186/gb-2008-9-5-r79

Hudson RR, Kaplan NL (1985) Statistical properties of the number of recombination events in the history of a sample of DNA-sequences. Genetics 111:147–164

Lyons LA, Kehler JS, O'Brien SJ (1999) Development of comparative anchor tagged sequences (CATS) for canine genome mapping. J Hered 90:15–26. doi:10.1093/jhered/90.1.15

Morin PA, Luikart G, Wayne RK (2004) SNPs in ecology, evolution and conservation. Trends Ecol Evol 19:208–216. doi:10.1016/j.tree.2004.01.009

Morin PA, Aitken NC, Rubio-Cisneros N, Dizon AE, Mesnick S (2007) Characterization of 18 SNP markers for sperm whale (*Physeter macrocephalus*). Mol Ecol Notes 7:626–630. doi:10.1111/j.1471-8286.2006.01654.x

Primmer CR, Borge T, Lindell J, Saetre GP (2002) Single-nucleotide polymorphism characterization in species with limited available sequence information: high nucleotide diversity revealed in the avian genome. Mol Ecol 11:603–612. doi:10.1046/j.0962-1083.2001.01452.x

Ray DA (2007) SINEs of progress: mobile element applications to molecular ecology. Mol Ecol 16:19–33. doi:10.1111/j.1365-294X.2006.03104.x

Rosenblum EB, Belfiore NM, Moritz C (2007) Anonymous nuclear markers for the eastern fence lizard, *Sceloporus undulatus*. Mol Ecol Notes 7:113–116. doi:10.1111/j.1471-8286.2006.01547.x

Rozas J, Sanchez-DelBarrio JC, Messeguer X, Rozas R (2003) DnaSP, DNA polymorphism analyses by the coalescent and other methods. Bioinformatics 19:2496–2497. doi:10.1093/bioinformatics/btg359

Rozen S, Skaletsky HJ (2000) Primer3 on WWW for general users and for biologist programmars. In: Misener S, Krawetz S (eds) Bioinformatics methods protocols: methods in molecular biology. Humana Press, Totowa, NJ, pp 365–386

Stephens M, Donnelly P (2003) A comparison of Bayesian methods for haplotype reconstruction from population genotype data. Am J Hum Genet 73:1162–1169. doi:10.1086/379378

Takezaki N, Nei M (1996) Genetic distances and reconstruction of phylogenetic trees from microsatellite DNA. Genetics 144:389–399

Zink RM, Barrowclough GF (2008) Mitochondrial DNA under siege in avian phylogeography. Mol Ecol 17:2107–2121. doi:10.1111/j.1365-294X.2008.03737.x