Microsatellites in Cetaceans: An Overview

Vincent Bourret, Matthias Macé, Maxime Bonhomme and Brigitte Crouau-Roy[*]

Laboratoire Evolution & Diversité Biologique Bâtiment 4R3 Université Paul Sabatier 118, Route de Narbonne 31062 Toulouse Cedex 9, France

Abstract: This paper presents a comprehensive overview of the published literature on microsatellites studied in cetaceans from 1989 to 2007. We inventoried 246 loci isolated from 18 and amplified in 51 cetacean species representing 11 families. The majority of loci (68%) were dinucleotide (CA) repeats, which were also shown to be on average more variable than tetranucleotide repeats. For each of these loci and the 1610 locus/species combinations, we present the species in which the markers were isolated and tested, specific diversity parameters (number of alleles, expected and observed heterozygosities), together with the primer sequences, the size range of the PCR products and the GenBank accession number. Simple analyses were performed on the assembled data and the widespread use of cross-species amplification, an important source of microsatellites in cetaceans, is discussed using laboratory data from *Stenella cœruleoalba*. No significant ascertainment bias was detected when considering all dinucleotide or tetranucleotide loci. This comprehensive database should help to inform those working on population and conservation genetic studies in most cetacean species.

Keywords: Cetacea, microsatellite, genetic diversity, conservation genetics, cross-species amplification, ascertainment bias, *Stenella cœruleoalba*.

INTRODUCTION

Several cetacean populations and species have been severely depleted during the last century and some are today considered at risk of extinction [1]. Until the mid-late 20th century, intensive commercial whaling was a major factor in the collapse of a number of stocks worldwide. At present, the whaling effort has been markedly reduced, but anthropogenic degradation of cetacean habitats is increasing, causing effects that are difficult to quantify. Such changes include the over-exploitation of marine resources, chemical and acoustic pollution, and heavy maritime traffic (see [2] for a review). The conservation of species or stocks living at the edge of extinction has raised international concern for several decades; however, consideration should also be given to the management of threatened populations not yet critically endangered [3].

Evaluating the structure and diversity of animal populations in conservation biology today involves the use of accurate genetic tools that support the planning of management schemes. Microsatellites are short, tandemly repeated DNA sequences (e.g. CACACACA) that are widely dispersed throughout the eukaryotic nuclear genome. Due to their high polymorphism and codominance, they are considered cornerstone molecular markers to address wildlife conservation issues [4]. They are among the most variable genetic markers available today and hence are particularly suitable to detect low levels of genetic diversity that are often found in vulnerable, declining or isolated populations [3]. They also allow investigations into the demographic history of populations and potentially correlated anthropogenic factors [5].

Furthermore, these markers are essential for forensic investigations, allowing the reliable identification of individuals [6-8] and estimation of the numbers of individuals entering illegal or unreported trades [9]. In this context, an increasing number of population structure inference tools allow the assignment of individuals to populations [10-12]. Finally, microsatellites are helpful for elucidating breeding strategies and social relationships which are characterized by complex behaviors in cetaceans [13], and for the fine-scale description of kinship that is of great interest in conservation biology. From a practical standpoint, microsatellite genotyping can be readily achieved from tiny quantities of even poor quality DNA. This allows conservative sampling of living as well as dead individuals, such as stranded, by-caught, harvested, or museum cetaceans [14, 15].

Specific microsatellites can be isolated by screening a species' genomic library with any repetitive probe (e.g. $(CA)_n$); however, this method is both costly and time-consuming. An alternative approach is to test loci that have been previously characterized in related species. Generally, the more closely related the species, the more successful the cross-amplification [16, 17]. However, Rico *et al.* [18] reported successful cross-amplifications between different fish Superclasses believed to have diverged some 470 million years (My) ago, suggesting variable conservation patterns among different taxonomic groups.

This paper presents the microsatellite markers isolated from cetacean genomic libraries from 1989 to 2007, and their polymorphism across a wide range of cetacean species. An extensive search of the published literature was carried out to create a synthetic database that can be used for population or conservation genetic studies in cetaceans.

*Address correspondence to this author at the Laboratoire Evolution & Diversité Biologique Bâtiment 4R3 Université Paul Sabatier 118, Route de Narbonne 31062 Toulouse Cedex 9, France; Tel: +33 (0) 561 557 328; Fax: +33 (0) 561 556 259; Email: bcrouau@cict.fr

METHODS

Database assembly. Ninety papers published between 1989 and 2007 were assembled, reporting either the characterization of new microsatellite loci or their use in cetacean species through specific or cross-species amplifications. From this material, a database was constructed including 246 microsatellites with the following information for each locus: i) name, repeat type, GenBank accession number, and sequence of each primer; ii) species in which the locus was tested; iii) specific genetic diversity parameters (number of individuals tested, number of alleles, expected (H_e) and observed (H_o) heterozygosities) and size range of the PCR products, whenever available. This database consists of four tables, one for each type of repeat: di-, tri-, tetranucleotide and complex repeats (Tables **A-D**, Supplementary Material).

Statistical analyses. We investigated which type of locus (di-, tri-, tetranucleotide, complex) is the most variable, and whether the variability is greater in species where the loci were isolated as opposed to species where the loci were cross-amplified (ascertainment bias). For this purpose, permutation tests were performed, which are suitable for analyzing the assembled database since they make no particular assumption of the distribution of the data or the characteristics of the datasets to be compared. The aim of these permutation tests is to test for a significant difference between means. The permutation procedure first involves calculating the difference between the observed means of two vectors of the H_e values (d_{obs}). These two vectors are then concatenated to create a new vector. Two vectors of the same respective lengths as the original two vectors to be compared are randomly sampled into this new vector. The difference between the estimated means of the two new vectors of H_e values is recorded (d_{est}). The procedure is repeated 10,000 times. The p-value of the permutation test is estimated as the proportion d_{est} superior to d_{obs} for positive d_{obs}, and the proportion d_{est} inferior to d_{obs} for negative d_{obs}. This permutation procedure was performed using R statistical package (version 2.0.1).

RESULTS & DISCUSSION

Microsatellites in cetacean species. Microsatellite data are reported from 51 cetacean species encompassing 11 families. The 244 species-specific loci have been isolated from genomic libraries in four mysticete and 14 odontocete species from 1989 to 2007 (Table **1** and Tables **A-D**). These 18 species represent six families (Balænidæ, Balænopteridæ, Delphinidæ, Monodontidæ, Phocœnidæ and Physeteridæ) among the 14 recognized by Berta *et al.* [19]. In particular, the family Delphinidæ was largely studied with 91 specific microsatellites isolated from nine delphinid species. This was not unexpected as this family is comprised of 34 species and is largely represented worldwide [19], with some species such as *Tursiops truncatus* (Montagu) being among the most studied and best-known cetacean models. Five families (Eschrichtiidæ, Iniidæ, Kogiidæ, Pontoporiidæ and Ziphiidæ) were studied only through cross-species amplifications. No microsatellite was isolated nor amplified in three cetacean families, namely the poorly known Neobalænidæ, endangered Platanistidæ, and critically endangered Lipotidæ [1]. Two additional microsatellites used in cetaceans were primarily isolated from the cattle *Bos taurus* (Linnæus), the total number of loci reported therefore being 246.

Table 1. Cetacean Species for which Microsatellites Could be Amplified. Bold Characters Indicate the Taxa from which Specific Microsatellites were Isolated. Other Taxa Reported here were Studied Only Through Cross-Species Amplifications

Suborder	Family	Genus	Species
Mysticeti	**Balænidæ**	***Balæna***	***B. mysticetus***
		Eubalæna	*E. australis*
			E. glacialis
			E. japonica
			E. spp.
	Balænopteridæ	***Balænoptera***	*B. acutorostrata*
			B. borealis
			B. brydei
			B. musculus
			B. physalus
		Megaptera	***M. novæangliæ***
	Eschrichtiidæ	*Eschrichtius*	*E. robustus*
Odontoceti	**Delphinidæ**	*Cephalorhynchus*	*C. commersonii*
		Delphinus	***D. delphis***
			D. spp.
		Globicephala	*G. macrorhynchus*
			G. melas
		Grampus	*G. griseus*
		Lagenorhynchus	*L. acutus*
			L. albirostris
			L. obliquidens
			L. obscurus
		Orcælla	*O. brevirostris*
		Orcinus	***O. orca***
		Pseudorca	*P. crassidens*
		Sotalia	*S. fluviatilis*
			S. guianensis
		Sousa	*S. chinensis*
		Stenella	*S. attenuata*
			S. clymene
			S. cœruleoalba
			S. frontalis
			S. longirostris
		Steno	*S. bredanensis*
		Tursiops	***T. aduncus***
			T. spp.
			T. truncatus
	Monodontidæ	***Delphinapterus***	***D. leucas***
		Monodon	*M. monoceros*
	Kogiidæ	*Kogia*	*K. simus*
	Phocoenidæ	***Neophocæna***	***N. phocænoides***
		Phocœna	*P. dioptrica*
			P. phocæna
			P. sinus
			P. spinipinnis
		Phocœnoides	*P. dalli*
	Physeteridæ	***Physeter***	***P. macrocephalus***
	Iniidæ	*Inia*	*I. geoffrensis*
	Pontoporiidæ	*Pontoporia*	*P. blainvillei*
	Ziphiidæ	*Berardius*	*B. bairdii*
		Hyperoodon	*H. ampullatus*
		Mesoplodon	*M. bidens*
			M. stejnegeri
		Ziphius	*Z. cavirostris*

Cross-amplifications have been performed in a number of cetacean species as shown in Table **1** and Tables **A-D**, resulting in a total of 1610 locus/species combinations. As an example of cross-amplification successes, 41 markers isolated from odontocetes were cross-amplified in mysticetes. Conversely, 28 markers isolated from mysticetes were cross-amplified in odontocetes, the two cetacean Suborders being presumed to have diverged ~35 My ago [19]. This confirms a high degree of conservation of the sequences flanking the repeats in cetaceans, as has been reported in earlier studies [20, 21].

Although cross-species amplifications are a common source of microsatellite markers in cetaceans, some caution is needed in their use as the evolution scheme of the repeated sequences can result in size homoplasy. For instance, a nucleotide substitution, shortening a perfect repeat array, can tend to prevent mutations [22]. Furthermore, there is a relative instability around the boundary between the repeated and flanking regions revealed by insertions/deletions (indels) or substitutions [23, 24].

To investigate the conservation of the repeats and their flanking sequences through different cetacean species, five microsatellites were sequenced in the Striped dolphin *Stenella cœruleoalba* (Meyen) that had originally been isolated from the humpback whale *Megaptera novœangliœ* (Borowski): GATA098, GATA053 [25], EV92*Mn* [26] or from the Indo-Pacific bottlenose dolphin *Tursiops aduncus* (Ehrenberg): MK6, MK9 [27]. One homozygous individual was sequenced for each marker, the PCR conditions being the same as those described in [28]. Four of these sequences are deposited in GenBank under accession numbers EF641271-EF641274 (GATA098 sequence was less than 50 base pairs in length and therefore was not submitted to GenBank). The five repeated motifs (three di- and two tetranucleotides) were conserved through evolution between *Stenella* and both *Tursiops* (divergence time: <12 My [29]) and

Megaptera (divergence time: ~35 My [19]). The orthologous sequences from *Stenella* and *Megaptera*/*Tursiops* were similar for two of the markers (GATA098 and MK9; data not shown). However, in the cases of GATA053, MK6, and EV92*Mn*, indel events and substitutions were observed (Fig. **1**). The indels were mostly located in the sequences immediately flanking the repeats (44 contiguous bases in GATA053, six bases in MK6, and four + 16 bases in EV92*Mn*), while the substitutions were found in the repeated motifs as well as the flanking sequences. Such observations emphasize the importance of a sequence-based allele determination when using a marker designed for a related species for the first time.

Nature of the repeated motifs. Among the total of 246 microsatellite loci reported in Tables **A**, **B**, **C** and **D** (Supplementary Material), dinucleotides were the most frequent, with 197 loci; five trinucleotides, 38 tetranucleotides, and six complex repeats are also reported (Fig. **2a**). The dinucleotide CA repeat is the most frequent of all isolated motifs, with a frequency of 68%. GATA (13% of the total) is the most frequent tetranucleotide motif (82%). The ratio between the number of loci of a certain type (e.g., number of CA loci) and the number of studies in which a genomic library was screened specifically for that repeated motif, gives a rough indication of the relative abundance of that motif in the genome. Such ratios indicate that CA and GATA would be the most common microsatellite repeats in the cetacean genome, with, on average, 11.2 CA and 7.8 GATA loci isolated per study (Fig. **2b**). It should be noted, however, that only a few studies specifically attempted to isolate trinucleotide repeats, making the ratios less informative for this type of loci.

Genetic diversity of the repeated motifs. The genetic diversity of each type of marker was estimated through the average of all H_e values reported in Tables **A-D**. The mean H_e values were 0.63 (± 0.23 SD) for dinucleotide (573 values ranging from 0.00 to 0.98), 0.62 (± 0.24) for trinucleotide

Fig. (1). Examples of microsatellite sequences in *Stenella cœruleoalba* aligned against homologous sequences in the species of origin (*Megaptera novœangliœ* or *Tursiops aduncus*). Substitutions and repeated motifs are in bold.
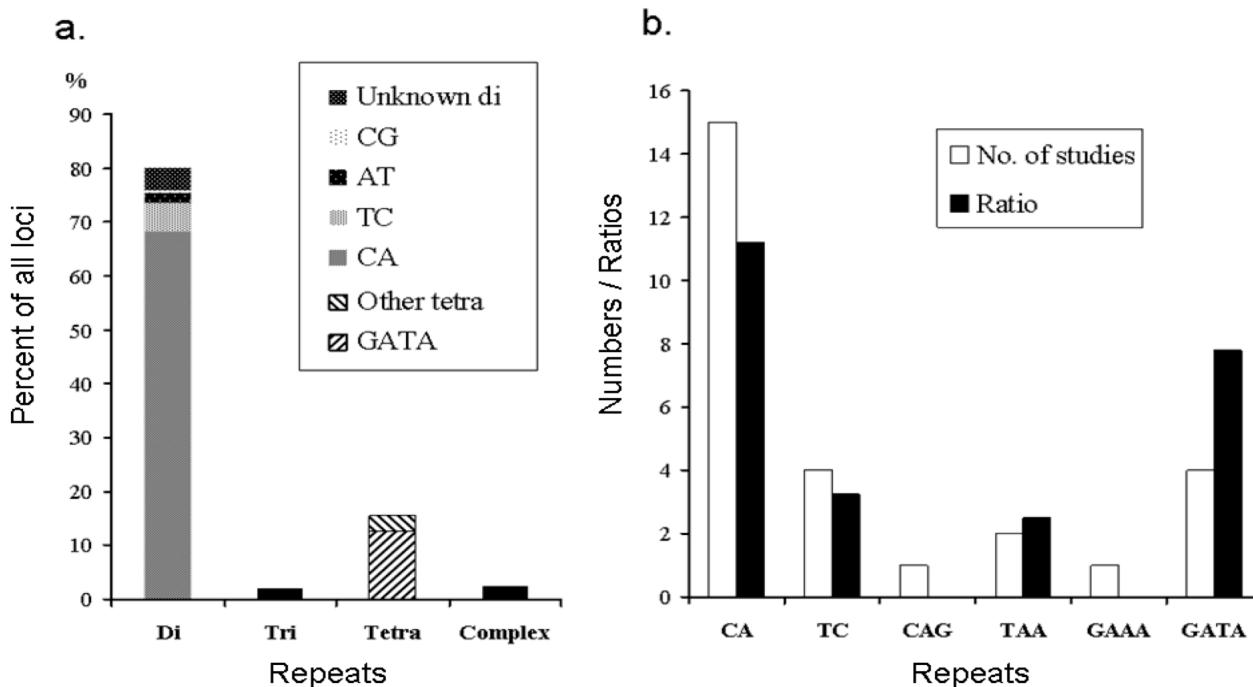
Fig. (2). Numbers and proportions of the different microsatellite motifs isolated in cetaceans from 1989 to 2007: (**a**) relative frequency of each type of repeat (percent), (**b**) numbers of published studies in which a cetacean genomic library was screened for a specific microsatellite motif (white bars), and ratios between the number of loci found and the corresponding number of studies (black bars) (see text for details).

(10 values ranging from 0.12 to 0.85), 0.48 (± 0.32) for tetranucleotide (68 values ranging from 0.00 to 0.90), and 0.72 (± 0.13) for complex repeats (5 values ranging from 0.58 to 0.87). High SD values show that the genetic diversity levels are heterogeneous among cetacean populations and species due notably to differences in effective population size and population history.

We investigated which type of locus (di-, tri-, tetranucleotide, complex) is the most variable, and whether the variability is greater in species where the loci were isolated as opposed to species where the loci were cross-amplified (ascertainment bias). Dinucleotide repeats were found to be significantly more polymorphic than tetranucleotide repeats (p-value < 0.0001). The CA and GATA repeats were the most common repeat types and were responsible for this difference in polymorphism (mean H_e 0.63 ± 0.23 and 0.49 ± 0.32, respectively, p = 0.0002). This is consistent with studies in humans showing that dinucleotide loci evolve at a rate 1.5-2 times faster than tetranucleotide loci [30]. Similar results are found in a range of organisms, and can be expected when considering the mutation mechanism in microsatellite regions [31]. These observations suggest that, where possible, variability comparisons between different populations should be made using similar sets of markers.

No significant ascertainment bias was detected when considering all dinucleotide (p = 0.24) or tetranucleotide (p = 0.16) loci. This result indicates that the use of cross-amplified loci does not *a priori* bias the estimation of genetic diversity in cetaceans, suggesting that comparisons are possible between studies based on cross-amplification and studies using specifically isolated markers.

CONCLUSION

The legacy of past intensive whaling, combined with an increasing anthropogenic pressure on the marine environment where human and animal communities have diverging needs and interests, make cetacean conservation problematic. In this context, a thoughtful use of molecular genetic markers will help in defining conservation priorities. This database presents a comprehensive overview of the microsatellites studied in cetaceans from 1989 to 2007 with their technical characteristics and genetic diversity, and includes data from 246 loci studied among 51 cetacean species (1610 locus/species combinations). Dinucleotide repeats, and especially (CA) repeats, are the most frequent and the most variable loci, as observed in other species. The widespread use of cross-species amplification appears to be a convenient approach due to the high degree of conservation of the repeats and flanking sequences and the absence of significant ascertainment bias in cetacean species. Such genetic information should fruitfully be integrated with more traditional approaches for the design of realistic and efficient cetacean management policies.

SUPPLEMENTARY MATERIAL

This article is accompanied by supplementary tables and it can be viewed at www.bentham.org/open/tombj

REFERENCES

[1] International Union for the Conservation of Nature (IUCN). 2006 IUCN Red List of Threatened Species [Downloaded on May 10, 2007]. Available from: www.iucnredlist.org

[2] Reeves RR, Smith BD, Crespo EA, di Sciara GN. Dolphins, whales and porpoises: 2002-2010 conservation action plan for the World's cetaceans. Gland and Cambridge: IUCN; 2003.

[3] Beebee T, Rowe G. An introduction to molecular ecology. New York: Oxford University Press; 2004.

[4] Beaumont MA, Bruford MW. In: Goldstein DB, Schlötterer C, Eds. Microsatellites - Evolution and Applications. New York: Oxford University Press 1999; 165-82.

[5] Goossens B, Chikhi L, Ancrenaz M, Lackman-Ancrenaz I, Andau P, Bruford MW. Genetic signature of anthropogenic population collapse in orang-utans. PloS Biol 2006; 4: 285-91.

[6] Baker CS, Cipriano F, Palumbi SR. Molecular genetic identification of whale and dolphin products from commercial markets in Korea and Japan. Mol Ecol 1996; 5: 671-85.

[7] Waits LP, Luikart G, Taberlet P. Estimating the probability of identity among genotypes in natural populations: cautions and guidelines. Mol Ecol 2001; 10: 249-56.

[8] Baker CS, Lukoschek V, Lavery S, *et al.* Incomplete reporting of whale, dolphin and porpoise 'bycatch' revealed by molecular monitoring of Korean markets. Anim Conserv 2006; 9: 474-82.

[9] Baker CS, Cooke JG, Lavery S, *et al.* Estimating the number of whales entering trade using DNA profiling and capture-recapture analysis of market products. Mol Ecol 2007; 16: 2617-26.

[10] Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. Genetics 2000; 155: 945-59.

[11] Corander J, Waldmann P, Sillanpää MJ. Bayesian analysis of genetic differentiation between populations. Genetics 2003; 163: 367-74.

[12] Guillot G, Mortier F, Estoup A. GENELAND: a computer package for landscape genetics. Mol Ecol Notes 2005; 5: 712-5.

[13] Gaspari S, Azzellino A, Airoldi S, Hoelzel AR. Social kin associations and genetic structuring of striped dolphin populations (*Stenella cœruleoalba*) in the Mediterranean Sea. Mol Ecol 2007; 16: 2922-33.

[14] Parsons KM. Reliable microsatellite genotyping of dolphin DNA from faeces. Mol Ecol Notes 2001; 1: 341-4.

[15] Krützen M, Barré LM, Möller LM, Heithaus MR, Simms C, Sherwin WB. A biopsy system for small cetaceans: darting success and wound healing in *Tursiops* spp. Mar Mammal Sci 2002; 18: 863-78.

[16] Primmer CR, Møller AP, Ellegren H. A wide-range survey of cross-species microsatellite amplification in birds. Mol Ecol 1996; 5: 365-78.

[17] Clisson I, Lathuillière M, Crouau-Roy B. Conservation and evolution of microsatellite loci in primate taxa. Am J Primatol 2000; 50: 205-14.

[18] Rico C, Rico I, Hewitt G. 470 million years of conservation of microsatellite loci among fish species. P Roy Soc B-Biol Sci 1996; 263: 549-57.

[19] Berta A, Sumich JL, Kovacs KM. Marine mammals: evolutionary biology. Burlington and San Diego: Academic Press; 2006.

[20] Schlötterer C, Amos B, Tautz D. Conservation of polymorphic simple sequence loci in cetacean species. Nature 1991; 354: 63-5.

[21] Primmer CR, Painter JN, Koskinen MT, Palo JU, Merilä J. Factors affecting avian cross-species microsatellite amplification. J Avian Biol 2005; 36: 348-60.

[22] Crouau-Roy B, Service S, Slatkin M, Freimer N. A fine-scale comparison of the human and chimpanzee genomes: linkage, linkage disequilibrium and sequence analysis. Hum Mol Genet 1996; 5:1131-7.

[23] Brohede J, Ellegren H. Microsatellite evolution: polarity of substitutions within repeats and neutrality of flanking sequences. P Roy Soc B-Biol Sci 1999; 266: 825-33.

[24] Crouau-Roy B, Clisson I. Evolution of an Alu DNA element of type Sx in the lineage of primates and the origin of an associated tetranucleotide microsatellite. Genome 2000; 43: 642-8.

[25] Palsbøll PJ, Bérubé M, Larsen AH, Jørgensen H. Primers for the amplification of tri- and tetramer microsatellite loci in baleen whales. Mol Ecol 1997; 6: 893-5.

[26] Valsecchi E, Amos W. Microsatellite markers for the study of cetacean populations. Mol Ecol 1996; 5: 151-6.

[27] Krützen M, Valsecchi E, Connor RC, Sherwin WB. Characterization of microsatellite loci in *Tursiops aduncus*. Mol Ecol Notes 2001; 1: 170-2.

[28] Bourret VJR, Macé MRJM, Crouau-Roy B. Genetic variation and population structure of western Mediterranean and northern Atlantic *Stenella cœruleoalba* populations inferred from microsatellite data. J Mar Biol Assoc UK 2007; 87: 265-9.

[29] Barnes LG. In: Leatherwood S, Reeves RR, Eds. The bottlenose dolphin. San Diego: Academic Press 1990; 3-25.

[30] Chakraborty R, Kimmel M, Stivers DN, Davison LJ, Deka R. Relative mutation rates at di-, tri-, and tetranucleotide microsatellite loci. Proc Natl Acad Sci USA 1997; 94: 1041-6.

[31] Goldstein DB, Schlötterer C, Eds. Microsatellites - evolution and applications. New York: Oxford University Press; 1999.