



# Characterization of 25 new microsatellite markers for the fin whale (*Balaenoptera physalus*) and cross-species amplification in other cetaceans

Céline Tardy<sup>1,2</sup> · Serge Planes<sup>2,3</sup> · Jean-Luc Jung<sup>4</sup> · Denis Ody<sup>1</sup> · Emilie Boissin<sup>2,3</sup>

Received: 1 April 2020 / Accepted: 28 August 2020 / Published online: 6 September 2020  
© Springer Nature B.V. 2020

## Abstract

Cetaceans are large mammals widely distributed on Earth. The fin whale, *Balaenoptera physalus*, is the second largest living animal. In the 20th century, commercial whaling reduced its global population by 70%, and in the Mediterranean Sea not only was their overall population depleted but the migration between the Mediterranean Sea and the Atlantic Ocean was reduced. Previous genetic studies identified isolation between these two regions, with a limited gene-flow between these adjacent populations based on nuclear and mitochondrial markers. However, only limited information exists for the Mediterranean population as genetic diversity and abundance trends are still unknown. In this study, 39 highly polymorphic microsatellite markers were tested, including 25 markers developed *de novo* together with 14 markers previously published. An average allelic diversity of 8.3 alleles per locus was reported, ranging from 3 to 15 alleles per locus, for *B. physalus*. Expected heterozygosity was variable among loci and ranged from 0.34 to 0.91. Only two markers in the new set were significantly deviant from the Hardy Weinberg equilibrium. Cross-species amplification was tested in four other cetacean species. A total of 27 markers were successfully amplified in the four species (*Balaenoptera acutorostrata*, *Megaptera novaeangliae*, *Physeter macrocephalus* and *Globicephala melas*). A multivariate analysis on the multilocus genotypes successfully discriminated the five species. This new set of microsatellite markers will not only provide a useful tool to identify and understand the genetic diversity and the evolution of the *B. physalus* population, but it will also be relevant for other cetacean species, and will allow further parentage analyses. Eventually, this new set of microsatellite markers will provide critical data that will shed light on important biological data within a conservation perspective.

**Keywords** Microsatellite · Genetic diversity · Fin-whale · Cross-species transfer · Northwestern Mediterranean Sea

## Introduction

Cetaceans are large mammals that are widely distributed throughout the world. All species are impacted by human activities (bycatch, overexploitation of prey resources, ship-strikes, noise and chemical pollution; [1]). Commercial whaling dramatically reduced the population of the

fin whale, *Balaenoptera physalus* (Linnaeus 1758) by over 70% during the twentieth century [2]. In 2018, global abundance was roughly estimated at 100 000 individuals [3] and *B. physalus* remains vulnerable on the IUCN Red List of Threatened Species. However, increases in local population sizes have been recorded for several populations since an international moratorium on commercial whaling was implemented in 1982 [3].

At the beginning of the twentieth century, the Mediterranean population was depleted and migrations between the Atlantic and the Mediterranean were reduced, due to whaling [4]. Within the Mediterranean, different local estimations of abundance were made; in the western basin: 3 583 individuals [5] and in the northern part of this basin: 1 800 individuals [6]. Abundance trends of *B. physalus* in the Mediterranean Sea are still unknown. Previous genetic analyses have suggested a genetic isolation of the Mediterranean

✉ Céline Tardy  
celine.tardy63@gmail.com

<sup>1</sup> WWF-France, Paris, France

<sup>2</sup> PSL Research University: EPHE-UPVD-CNRS, USR 3278 CRIOBE, 66860 Perpignan, France

<sup>3</sup> Laboratoire d'Excellence "CORAIL", Perpignan, France

<sup>4</sup> Univ Brest, Muséum National d'Histoire Naturelle, CNRS, Sorbonne Université, ISYEB, 29200 Brest, France

*B. physalus* population, with a limited gene-flow with the adjacent Atlantic population [7]. To better understand and protect the *B. physalus* population, different parameters (e.g. abundance, effective size, gene-flow, etc.) are required and some should be updated. To do this, genetic markers, such as microsatellite markers, are useful as they enable changes in genetic diversity to be identified and for the structure of natural populations to be studied [8]. Furthermore, a large number of markers (> 20) allow for the identification of individuals, by producing a unique and immutable identification tag for each organism [9]. Through the identification of individuals, the study of both connectivity and the parentage link between groups of individuals is possible, and in turn, the protection of the population may be enhanced.

Microsatellite markers have been widely used since the 1990s [8]. Microsatellites consist of short nucleotide tandem repeats with DNA sequences that are highly variable in length [8] and commonly dispersed throughout the eukaryotic nuclear genome. Microsatellite length determination generally relies on the design of a PCR-primer pair flanking the variable repeat, and is therefore often species-specific. This method is costly and time-consuming, and for these reasons, testing the transferability to related species is important and is of great interest for further study. The cross-species transferability of microsatellites is unevenly distributed across taxa. For example, microsatellites are more conserved between humans and primates [10], than between oyster species [11]. Mutations in the sequence of microsatellite markers are an important component of the variation between individuals and species [12]. Consequently, some markers may be polymorphic in one species but monomorphic in another [10, 12]. In *B. physalus*, some microsatellite markers were previously published [7, 13–16], but were initially developed for the humpback whale, *Megaptera novaeangliae*. Less than ten specific markers have been developed for *B. physalus*, which is an insufficient number of markers for accurate population structure and parentage analyses.

The aim of this study was to develop *de novo* microsatellites for our target species, *B. physalus*, in order to extend the set of available markers for the study of the Mediterranean population during other biological and conservation studies. To take advantage of the time and cost associated with the creation of a new set of markers, we also tested the newly developed and previously published markers for cross-amplification in four other cetacean species.

## Materials and methods

### Biological samples

Fifty samples of our target species *B. physalus* were used in this study. For cross-species amplification, four cetaceans

were selected to test the transferability of microsatellite markers: four minke whales, *Balaenoptera acutorostrata*, three *M. novaeangliae*, ten sperm whales, *Physeter macrocephalus*, and ten long-finned pilot whales, *Globicephala melas*. Samples from *B. physalus*, *P. macrocephalus* and *G. melas* come from biopsy darts fired from a crossbow taken by the NGO WWF in the North-western part of Mediterranean Sea between 2006 and 2018. In the field, samples were stored in 70% ethanol. For *B. acutorostrata*, two samples come from specimens stranded along the Mauritanian coast in 2013 (CITES export permit no. 2014-001 and CITES import permit no. FR1402900008-1 provided by the designated authorities in Mauritania and France, respectively) and two others from whale meat (a steak and a sausage) bought in the fishmarket of Bergen (Norway) in July 2018. The samples of *M. novaeangliae* are biopsies taken by the NGO Cetamada near Sainte Marie Island in the North-eastern coast of Madagascar in 2013 (n = 1) and 2016 (n = 2).

### DNA extraction and design of new microsatellite markers

Total genomic DNA was isolated from 12 *B. physalus* individuals using the Genra PureGene tissue kit (Qiagen, Hilden, Germany) and sent to GenoScreen molecular marker services (Lille, France) for microsatellite library preparation and sequencing. A total of 1 µg from an equimolar DNA pool of the 12 specimens was used for the development of Illumina TruSeq Nano library, sequenced on MiSeq platform with Nano v2 500 cycles chip. After de-multiplexing, the resulting paired-end reads quality was checked with FastQC v0.11.15 software. Then, the remaining data were subject to assembly/merging through Velvet v1.2.10 program. Finally, the assembly with the best remapping rate was analysed with the QDD v3.1 bioinformatics program [17]. QDD is designed to deal with all steps of treatment of raw sequences until PCR primers are obtained: removal of adapters/vectors, detection of microsatellites, detection of redundancy/possible mobile element association, selection of sequences with target microsatellites and primer design by using BLAST, ClustalW and Primer3 programs. The default setting used for QDD were 200 bp flanking region length, sequence removal less than 80 bp, a minimum at 95% of pairwise identity between sequences of a contig, a minimum at 66% of sequences that must have the same base at a site to accept it as a consensus. The design of primer was determined with the following setting values: a minimum PCR product size at 90 bp and a maximum at 300 bp, a PCR product size interval for iterative primer design at 50 bp, an optimum length of a primer at 20 bp with a minimum at 18 bp and a maximum at 27 bp. Among 3,609,736 raw sequences, 82,283 merged sequences were obtained and 5

044 primer sets were designed. Based on this dataset, 33 primer pairs were selected, according to the repeat number ( $\geq 9$ ), motif and PCR product size ( $\geq 100$  bp), and tested. Additionally, 14 primer pairs previously published in [7, 13–16] were also tested. Most of them were initially developed for *M. novaeangliae*. To identify redundant sequences, we ran the BLASTN algorithm with default values for the parameters (megablast with: word size: 28; match/mismatch scores: 1/– 2; gap cost: linear; filter: low complexity regions; <http://blast.ncbi.nlm.nih.gov/Blast.cgi>; [18]) for each of the 33 new sequences for which suitable primers were designed and for the 14 previously published on the *B. physalus* genome assembly (Genbank accession number: GCA 008795845.1).

## Molecular analysis

The newly developed loci together with those which were previously published [7, 13–16] were genotyped in our target species and cross-tested in four other cetacean species to verify their transferability. For the characterisation of newly developed microsatellites, DNA was extracted from small tissue fragments ( $< 0.5$  cm<sup>3</sup>) using the Genra PureGene tissue kit (Qiagen, Hilden, Germany).

PCR amplifications were performed using Type-it Microsatellite PCR kit (Qiagen, Hilden, Germany) in 12.5  $\mu$ L total volume reactions containing 4  $\mu$ L Type-it Multiplex PCR Master Mix 2X (contains HotStarTaq<sup>®</sup> Plus DNA Polymerase, Type-it Microsatellite PCR Buffer with 6 mM MgCl<sub>2</sub> and dNTPs), 6  $\mu$ L RNase-free water, 1  $\mu$ L of primers (2  $\mu$ M forward and reverse primers diluted in TE pH 8 buffer) and 1.5  $\mu$ L of DNA template at 50 ng/ $\mu$ L. Forward primers were labelled with a fluorescent dye (PET, NED, VIC or 6-FAM, Applied Biosystems). Amplifications were carried out as follows: 5 min at 95 °C, followed by 40 cycles of 30 s at 95 °C, 1 min 30 s at optimal annealing temperature (53–63 °C, depending on the locus), 30 s at 72 °C and a final extension for 30 min at 60 °C. All loci were combined in six multiplexes according to their size range and primer annealing temperature to perform PCR (Table 1). Each PCR product was run through 1.5% agarose gel and visualized using ethidium bromide to verify amplifications. PCR products were sent to GenoScreen (Lille, France) and allele sizes were assessed using an Applied Biosystems 3730 Sequencer. For accurate sizing, an internal size ladder (GeneScan 500 LIZ, Applied Biosystems) was used.

## Data analysis

Allele sizes were scored and checked manually using GENEMAPPER software v.5 (Applied Biosystems). All ambiguous peak profiles were considered as missing data. Controls for the presence of null alleles, scoring errors and large allele

dropout were performed with MICRO-CHECKER v.2.2.3 [19]. Allele frequencies, total number of alleles ( $N_a$ ) and private allele numbers (PA) were identified and estimated in GenAIEx v.6.503 [20]. Observed ( $H_o$ ) and expected ( $H_e$ ) heterozygosities were estimated for each locus using GENETIX v.4.05.2 [21]. GENETIX was also used to identify the inbreeding coefficient ( $F_{IS}$ ) and linkage disequilibrium ( $LD$ ). Finally, a Principal Coordinates Analysis (PCoA) was computed in GenAIEx with pairwise population matrix unbiased genetic distances in order to explore genetic variance among individuals and species.

## Results

### Marker amplification on *B. physalus* samples

Of the 47 primer pairs (33 newly developed and 14 previously published) tested on *B. physalus* in this study, 32 successfully amplified (70%). For the fifteen markers that did not amplify, eight were newly developed and seven previously published (Table 1). Thus, half of the previously published markers amplified, among which six were initially developed for *M. novaeangliae* and one for *B. physalus* (Table 1). For the seven markers that did not amplify (Table 1), five were developed for *M. novaeangliae*, one for *P. macrocephalus* and one for *B. physalus*.

### Genetic diversity of *B. physalus*

The thirty-two amplified loci were polymorphic (from three to fifteen alleles) and were kept for cross-species transferability tests. Twenty-three loci contained dinucleotide repeats, with the number of alleles ranging from three to fifteen (for a total of 202 alleles), four were trinucleotide repeats with number of alleles ranging from three to six (16 alleles in total), and five contained tetranucleotide repeats with five to fifteen alleles (42 alleles in total).

After analysis with MICROCHECKER, null alleles were detected in three loci (from our new set: Bp1430262, Bp747752, and Bp1003953). There was no evidence of scoring errors or large allele dropout in any other loci. Test for linkage disequilibrium revealed that 14% of the pairwise locus combinations were linked ( $p < 0.05$ ), and distributed among all loci in *B. physalus*. The three loci with null alleles (Bp1430262, Bp747752 and Bp1003953) showed a significant probability of linkage disequilibrium (29%, 35% and 19% respectively). Due to null alleles and linkage disequilibrium, these loci were removed from further analyses in *B. physalus*.

The 29 polymorphic loci that showed no evidence of null alleles revealed distinct allele polymorphisms from three (Bp1311471) to fifteen alleles (Bp740180,

**Table 1** Characterization of the 32 microsatellite markers used for this study on *Balaenoptera physalus* together with the seven loci also tested but that did not amplify

Locus	Primer sequence	Repeat motif	Source species	Ta (°C)	Size range (bp)	Genome coordinates	GenBank accession
Bp451306	F: GGAAGAAGTCTGGAGTGATGG R: GGGTTTGCAATGCAGGGATG	(AC) <sub>22</sub>	B.p.	53	244–264	29,362–29,793	MT514926
Bp673257	F: CCTTCTCTCAAGAGATTGGG R: GCTGTACAACATGCCAAGAGGA	(AGAT) <sub>16</sub>	B.p.	53	173–197	423,381–423,838	MT585351
Bp1125752	F: ACTGAGGTACCCAGCGTTA R: AGAGCAACATCACATTTGTCCA	(AC) <sub>16</sub>	B.p.	53	120–134	519,907–519,488	MT585352
Bp1178841	F: GGCAGGAAGACATATAATCAGCA R: CCCACTGCCTAGAAAAGCCTT	(AC) <sub>18</sub>	B.p.	53	242–278	631,007–631,464	MT585353
Bp3111392	F: AGAAGGGACAGTGAGAGGGA R: ACAGCCGAACAGAGAATCG	(AC) <sub>14</sub>	B.p.	55	211–221	–	MT585354
Bp1229316	F: ATGCTAAGCTCCAGCCTTGC R: GGTGTGACTACGAACGCTCA	(AGG) <sub>12</sub>	B.p.	55	153–165	316,754–316,921	MT585355
Bp989588	F: AGTTAGTTTCACTCTCCAGGG R: TGGTACAGGAGAGTAAAGCAGA	(ATC) <sub>9</sub>	B.p.	55	262–271	264,038–264,464	MT585356
Bp740180	F: TATCACTCTTCAGCAGCGCA R: AAATGCTGCGGCTCTTAGGT	(AC) <sub>19</sub>	B.p.	57	160–194	–	MT585357
Bp1446664	F: GGTAAGGACGTCGACCAAC R: TGTGGATGCAGACTCAGCAG	(AC) <sub>13</sub>	B.p.	57	224–232	273,322–273,745	MT585358
Bp1043662	F: GGTGTAGCCCACAFCTTGGG R: CAGTTCACTGGGAATGTCTGA	(AGAT) <sub>11</sub>	B.p.	57	126–154	713,761–714,188	MT585359
Bp1311471	F: GGTTAGACTCCACGCTTCCA R: CAGCCATGTAGCCTTGGAA	(AAC) <sub>9</sub>	B.p.	57	260–272	237,046–237,471	MT585360
Bp958405	F: AGACTGTTGCAGTTACAGCT R: ACAAGGAGGGTTTGCATGCT	(AC) <sub>17</sub>	B.p.	60	180–202	686,971–687,410	MT585361
Bp1057120	F: GTTTGTGAAATGTTCCCTTCTCA R: GCAACAAGCCCCATGTACAFATGG	(AC) <sub>20</sub>	B.p.	60	273–305	–	MT585362
Bp1357166	F: ACCACGCAGTTAACAFATTTCTGG R: ACAAGAATGGTGTAGATTTATTGCT	(AC) <sub>17</sub>	B.p.	60	294–306	354,233–354,826	MT585363
Bp1069071	F: CAAGTTCTCTTGTCTGGACCA R: CCTTTGGGAGCACAGTGGTT	(AC) <sub>15</sub>	B.p.	60	174–184	601,048–601,473	MT585364
Bp531923	F: GGCTGCAGTGGTTGATCTGT R: CCATGACTTTACGGTGGGAGA	(AC) <sub>13</sub>	B.p.	60	214–234	15,596–16,021	MT585365
Bp999219	F: GCAGATGCAAGGCCAATCTT R: TATTGCCCAAAATGGTGTGCC	(AC) <sub>15</sub>	B.p.	60	255–287	862,872–863,309	MT585366
Bp779455	F: CCAGAACAGTCCCTTGGCAAC R: ACCACTGTCTGATCAACTTTGT	(AC) <sub>22</sub>	B.p.	60	164–198	697,042–697,559	MT585367
Bp1418059	F: TCCCATCTTGGCATCTGCTT R: CCCAATTAATGATGCCAGATGAT	(AC) <sub>24</sub>	B.p.	63	284–310	2,336,061–2,336,551	MT585368

Table 1 (continued)

Locus	Primer sequence	Repeat motif	Source species	Ta (°C)	Size range (bp)	Genome coordinates	GenBank accession
Bp814420	F: ACATACTTTGGGAGCAGGATTAAGT R: CCTGCAITTAGGCTGTCCCAA	(AC) <sub>15</sub>	B.p.	63	145–155	657,017–657,446	MT585369
Bp564742	F: AGCAGTTATACAAGCCTAGTGAGT R: CTTAAAGGTGTATAGACAGAGACCT	(AC) <sub>16</sub>	B.p.	63	235–255	287,390–287,834	MT585370
Bp1047889	F: ACCAAGAGCCTTAGCATGTCC R: TCCAAGAAAGGATTAAGGCA	(AC) <sub>17</sub>	B.p.	63	265–271	927,811–928,250	MT585371
Bp1430262 <sup>‡</sup>	F: TTCCTCAGTCAGAAAACGGCG R: AGTGGCCATGTACAAGTGCA	(AC) <sub>18</sub>	B.p.	55	270–286	690,203–690,438	MT585372
Bp747752 <sup>‡</sup>	F: GGGACCAATCACCTGTCTCT R: TCGCCGCAACTAGAGAAAGC	(AAAT) <sub>9</sub>	B.p.	60	190–206	583,163–583,488	MT585373
Bp1003953 <sup>‡</sup>	F: GGAACCTGTGGAGATGTGTCT R: ACAATCTAAATTCCTTGCCACAG	(AC) <sub>19</sub>	B.p.	63	186–200	376,344–376,779	MT585374
GATA098 <sup>1</sup>	F: TGTACCCCTGGATGGATAGATT R: TCACCTTATTTGTCTGTCTG	(GATA) <sub>n</sub>	M.n.	53	88–112	41,517–41,782	U93892.1
GT011 <sup>2</sup>	F: CATTTGGGTGGATCAATC R: GTGGAGACCAGGGATATTG	(GT) <sub>n</sub>	M.n.	53	113–127	–	–
EV37 <sup>3</sup>	F: AGCTTGATTTGGAATCATGA R: TAGTAGAGCCGTGATAAAAGTGC	(AC) <sub>24</sub>	M.n.	55	191–223	172,076–172,278	G09081.1
GATA417 <sup>1</sup>	F: CTGAGATAGCATACATGGG R: TCTGCTCAGGAAATTTCAAG	(GATA) <sub>n</sub>	M.n.	57	180–256	719–788	U93893.1
GT122 <sup>4</sup>	F: CCTTTTAAACCCAGAATGTAG R: TGTGGCAGTGACGAATGAAAGG	(GT) <sub>n</sub>	B.p.	57	134–170	1,280,305–1,280,701	DQ151660.1
GT211 <sup>5</sup>	F: GGCACAAATCAGTAAGGTAGG R: CATCTGTCTCCACAAGCCC	(GT) <sub>n</sub>	M.n.	60	197–217	294,505–294,744	AF309693.1
TAA023 <sup>1</sup>	F: CTCGCACAGAAATGAAGACC R: AGAGCCTGAACCCAGACAAGG	(TAA) <sub>n</sub>	M.n.	60	81–99	556,112–556,306	U93896.1
EV1Pm <sup>†3</sup>	F: CCCTGCTCCCAATCTC R: ATAAACTCTAATACACTTCTCCAAC	(AC) <sub>13</sub> (TC) <sub>8</sub>	P.m.	–	115–197	102,101–102,258	G09074.1
EV94Mn <sup>†3</sup>	F: ATCGTATTGGTCTTTTCTGC R: AATAGATAGTATGATGATTCACACC	(TC) <sub>6</sub> [...](AC) <sub>20</sub>	M.n.	–	198–261	410,830–410,984	G09083.1
GATA028 <sup>†1</sup>	F: AAAGACTGAGATCTATAGTTA R: CGCTGATAGATTAGTCTAGG	(GATA)	M.n.	54	184–236	67,071–67,394	U93890.1
GATA053 <sup>†1</sup>	F: ATTTGGCAGTGGCAGGAGACC R: GACACAGAGATGTAGAAGGAG	(GATA)	M.n.	54	180–220	709,309–709,415	U93891.1
GT271 <sup>†5</sup>	F: GCTCACACTGGTAATCTGTGG R: CCCTAGGAAGGATAGACATAG	(GT)	M.n.	62	112–128	202,383–202,553	AF309694.1

Table 1 (continued)

Locus	Primer sequence	Repeat motif	Source species	Ta (°C)	Size range (bp)	Genome coordinates	GenBank accession	
GT575 <sup>15</sup>	F: TATAAGTGAATACAAAAGACCC R: ACCATCAACTGGAAAGICTTTC	(GT)	M.n.	60	140–154	1,178,324–1,178,546	AF309698.1	
GT142 <sup>14</sup>	F: CTGAGTAATATCCACCATAC R: ATGGATAAAGAAGATGTGGGG	(GT)	B.p.	60	72–90	–	DQ151658.1	
Locus	Multiplex number	N	Na	PA	H <sub>0</sub>	H <sub>E</sub>	F <sub>IS</sub> (W&C)	Cross species amplification (success/total)
Bp451306	1	50	10	5	0.7400	0.7424	0.013	4/4
Bp673257	1	50	7	0	0.7600	0.7918	0.050	4/4
Bp1125752	1	50	7	4	0.7400	0.7162	–0.023	4/4
Bp1178841	1	50	13	10	0.8800	0.8454	–0.031	4/4
Bp311392	2	50	6	2	0.6600	0.6998	0.067	4/4
Bp1229316	2	50	4	1	0.5800	0.5670	–0.013	4/4
Bp989588	2	50	3	0	0.5200	0.4864	–0.059	4/4
Bp740180	3	50	15	3	0.9184	0.9017	–0.008	4/4
Bp1446664	3	50	5	1	0.7400	0.6934	–0.057	4/4
Bp1043662	3	50	8	3	0.7600	0.7800	0.036	3/4
Bp1311471	3	50	3	1	0.3400	0.3394	0.008	4/4
Bp958405	4	50	8	3	0.8000	0.8016	0.012	4/4
Bp1057120	4	50	11	2	0.8000	0.8548	0.074	4/4
Bp1357166	4	50	7	2	0.8600	0.7442	–0.146 *	4/4
Bp1069071	5	50	5	2	0.8400	0.7582	–0.098	4/4
Bp531923	5	50	8	1	0.7400	0.7528	0.027	4/4
Bp999219	5	50	9	2	0.7800	0.7854	0.017	4/4
Bp779455	5	50	10	4	0.7200	0.8262	0.138 *	4/4
Bp1418059	6	50	10	4	0.8200	0.8582	0.055	4/4
Bp814420	6	50	6	1	0.6600	0.6234	–0.049	3/4
Bp564742	6	50	10	4	0.8800	0.8492	–0.026	4/4
Bp1047889	6	50	4	0	0.6400	0.5866	–0.081	2/4
Bp1430262 <sup>‡</sup>	2	50	8	3	0.7800	0.8290	0.069	4/4
Bp747752 <sup>‡</sup>	4	50	5	1	0.2449	0.4744	0.492 ***	3/4
Bp1003953 <sup>‡</sup>	6	50	7	6	0.7600	0.7920	0.050	4/4
GATA098 <sup>1</sup>	1	50	7	4	0.8000	0.8060	0.018	4/4

Table 1 (continued)

Locus	Multiplex number	N	Na	PA	H <sub>0</sub>	H <sub>E</sub>	F <sub>IS</sub> (W&C)	Cross species amplification (success/total)
GT011 <sup>2</sup>	1	50	7	1	0.8200	0.7888	-0.029	4/4
EV37 <sup>3</sup>	2	50	12	6	0.9000	0.8412	-0.060	4/4
GATA417 <sup>1</sup>	3	50	15	9	0.9800	0.9064	-0.071	4/4
GT122 <sup>4</sup>	3	50	15	13	0.8200	0.8480	0.043	4/4
GT211 <sup>5</sup>	4	50	9	6	0.8200	0.8142	0.003	4/4
TAA023 <sup>1</sup>	5	50	6	1	0.5714	0.7341	0.231 **	3/4
EVIPm <sup>†3</sup>	-	-	-	-	-	-	-	-
EY94Mn <sup>†3</sup>	-	-	-	-	-	-	-	-
GATA028 <sup>†1</sup>	-	-	-	-	-	-	-	-
GATA053 <sup>†1</sup>	-	-	-	-	-	-	-	-
GT271 <sup>†5</sup>	-	-	-	-	-	-	-	-
GT575 <sup>†5</sup>	-	-	-	-	-	-	-	-
GT142 <sup>†4</sup>	-	-	-	-	-	-	-	-

Ta, annealing temperature in degrees; N, number of samples; Na, number of alleles; PA, private alleles; H<sub>0</sub>, observed heterozygosities; H<sub>E</sub>, expected heterozygosities; F<sub>IS</sub>, fixation index following Weir and Cockerham; B.p., *Balaenoptera physalus*; M.n., *Megaptera novaeangliae*

\*indicates significant departures from Hardy–Weinberg Equilibrium (p < 0.05); † loci presenting null alleles; ‡ loci that did not amplify in this study

<sup>1</sup>[14]; <sup>2</sup>[7]; <sup>3</sup>[13]; <sup>4</sup>[16]; <sup>5</sup>[15]

**Table 2** Results of the cross-species amplifications using the fin-whale loci

Locus		Common minke whale ( <i>Balaenoptera acutorostrata</i> )	Humpback whale ( <i>Megaptera novaeangliae</i> )	Sperm whale ( <i>Physeter macrocephalus</i> )	Long-finned pilot whale ( <i>Globicephala melas</i> )
Bp451306	N/Ntotal	4/4	3/3	10/10	10/10
	Size range	242	246–250	264–288	258–276
	No. of alleles	1	3	8	4
	No. of PA	1	1	6	1
Bp673257	N/Ntotal	3/4	3/3	9/10	10/10
	Size range	173–189	161–205	169–181	145
	No. of alleles	5	4	4	1
Bp1125752	No. of PA	0	2	1	1
	N/Ntotal	4/4	3/3	10/10	10/10
	Size range	136–142	114–120	122–124	112
Bp1178841	No. of alleles	2	2	2	1
	No. of PA	1	1	0	1
	N/Ntotal	3/4	3/3	10/10	10/10
Bp311392	Size range	246	246–252	234–256	228–248
	No. of alleles	1	4	7	2
	No. of PA	0	0	4	1
Bp1229316	N/Ntotal	4/4	2/3	9/10	8/10
	Size range	203–205	205–207	201–219	201–223
	No. of alleles	2	2	5	3
Bp989588	No. of PA	1	1	0	0
	N/Ntotal	4/4	3/3	6/10	5/10
	Size range	150–156	156–159	159–162	144–165
Bp740180	No. of alleles	2	2	2	4
	No. of PA	1	0	0	1
	N/Ntotal	3/4	3/3	10/10	6/10
Bp1446664	Size range	280	271	262–271	256–271
	No. of alleles	1	1	2	3
	No. of PA	1	0	0	1
Bp1043662	N/Ntotal	3/4	3/3	9/10	10/10
	Size range	168–188	172–184	172–192	158–166
	No. of alleles	6	5	9	3
Bp1311471	No. of PA	0	0	1	2
	N/Ntotal	3/4	3/3	10/10	5/10
	Size range	226–238	246–282	240–248	228–232
Bp958405	No. of alleles	3	4	5	3
	No. of PA	1	2	3	0
	N/Ntotal	4/4	3/3	10/10	4/10
Bp1311471	Size range	130–138	162–178	126–134	126–150
	No. of alleles	2	3	3	4
	No. of PA	0	0	4	3
Bp1311471	N/Ntotal	4/4	3/3	9/10	10/10
	Size range	266	266–269	245–260	263
	No. of alleles	1	2	2	1
Bp958405	No. of PA	0	0	1	1
	N/Ntotal	3/4	3/3	10/10	10/10
	Size range	174–178	174–192	196–214	186–210
Bp958405	No. of alleles	3	2	7	5
	No. of PA	2	0	3	1



**Table 2** (continued)

Locus		Common minke whale ( <i>Balaenoptera acutorostrata</i> )	Humpback whale ( <i>Megaptera novaeangliae</i> )	Sperm whale ( <i>Physeter macrocephalus</i> )	Long-finned pilot whale ( <i>Globicephala melas</i> )
Bp1057120	N/Ntotal	3/4	2/3	10/10	10/10
	Size range	285–287	273	291–307	277–289
	No. of alleles	2	1	8	2
	No. of PA	1	0	2	1
Bp1357166	N/Ntotal	3/4	3/3	10/10	10/10
	Size range	278	288–294	292–302	296–300
	No. of alleles	1	3	6	3
Bp1069071	N/Ntotal	3/4	2/3	10/10	10/10
	Size range	174	178–182	166	204
	No. of alleles	1	2	1	1
Bp531923	N/Ntotal	4/4	3/3	8/10	10/10
	Size range	220–240	222–236	220–230	212–246
	No. of alleles	4	4	5	6
Bp999219	N/Ntotal	2/4	2/3	6/10	10/10
	Size range	271–283	263–265	255–285	275–283
	No. of alleles	3	2	5	3
Bp779455	N/Ntotal	3/4	3/3	8/10	10/10
	Size range	182–198	168–172	182–194	164–202
	No. of alleles	5	3	4	8
Bp1418059	N/Ntotal	3/4	2/3	10/10	10/10
	Size range	284–296	286–294	256	264
	No. of alleles	5	3	1	1
Bp814420	N/Ntotal	1/4	2/3	10/10	10/10
	Size range	167–169	149–153	155–167	145–159
	No. of alleles	2	3	5	5
Bp564742	N/Ntotal	2/4	3/3	6/10	10/10
	Size range	231–251	227–235	233–239	253–255
	No. of alleles	3	4	3	2
Bp1047889	N/Ntotal	1/4	2/3	2/10	8/10
	Size range	275–277	263–277	265–271	273–295
	No. of alleles	2	4	3	3
Bp1430262	N/Ntotal	3/4	3/3	10/10	10/10
	Size range	264–266	268–288	278–292	298–308
	No. of alleles	2	5	7	4
	No. of PA	2	1	2	4

**Table 2** (continued)

Locus		Common minke whale ( <i>Balaenoptera acutorostrata</i> )	Humpback whale ( <i>Megaptera novaeangliae</i> )	Sperm whale ( <i>Physeter macrocephalus</i> )	Long-finned pilot whale ( <i>Globicephala melas</i> )
Bp747752	N/Ntotal	4/4	2/3	2/10	6/10
	Size range	190	214–222	202–206	190–202
	No. of alleles	1	3	2	3
	No. of PA	0	2	0	0
Bp1003953	N/Ntotal	3/4	3/3	10/10	10/10
	Size range	168	184–186	162	176
	No. of alleles	1	2	1	1
	No. of PA	1	1	1	1
GATA098	N/Ntotal	4/4	3/3	5/10	10/10
	Size range	80–92	72–100	80–96	68
	No. of alleles	4	2	5	1
	No. of PA	0	1	0	1
GT011	N/Ntotal	3/4	3/3	10/10	10/10
	Size range	119–133	123–137	107–131	121–129
	No. of alleles	3	4	7	5
	No. of PA	0	1	1	0
EV37	N/Ntotal	3/4	3/3	10/10	10/10
	Size range	207–209	197–211	213–239	183–197
	No. of alleles	2	5	5	4
	No. of PA	0	0	4	3
GATA417	N/Ntotal	3/4	3/3	10/10	10/10
	Size range	184–240	180–228	168–184	104
	No. of alleles	4	5	3	1
	No. of PA	0	0	1	1
GT122	N/Ntotal	4/4	3/3	10/10	10/10
	Size range	134–140	128	126–138	122–152
	No. of alleles	4	1	5	4
	No. of PA	1	1	2	1
GT211	N/Ntotal	3/4	3/3	10/10	10/10
	Size range	195–203	201–207	187–201	183–187
	No. of alleles	4	2	4	3
	No. of PA	0	1	0	2
TAA023	N/Ntotal	4/4	2/3	3/10	6/10
	Size range	87	81	87–150	81–96
	No. of alleles	1	1	5	3
	No. of PA	0	0	1	0

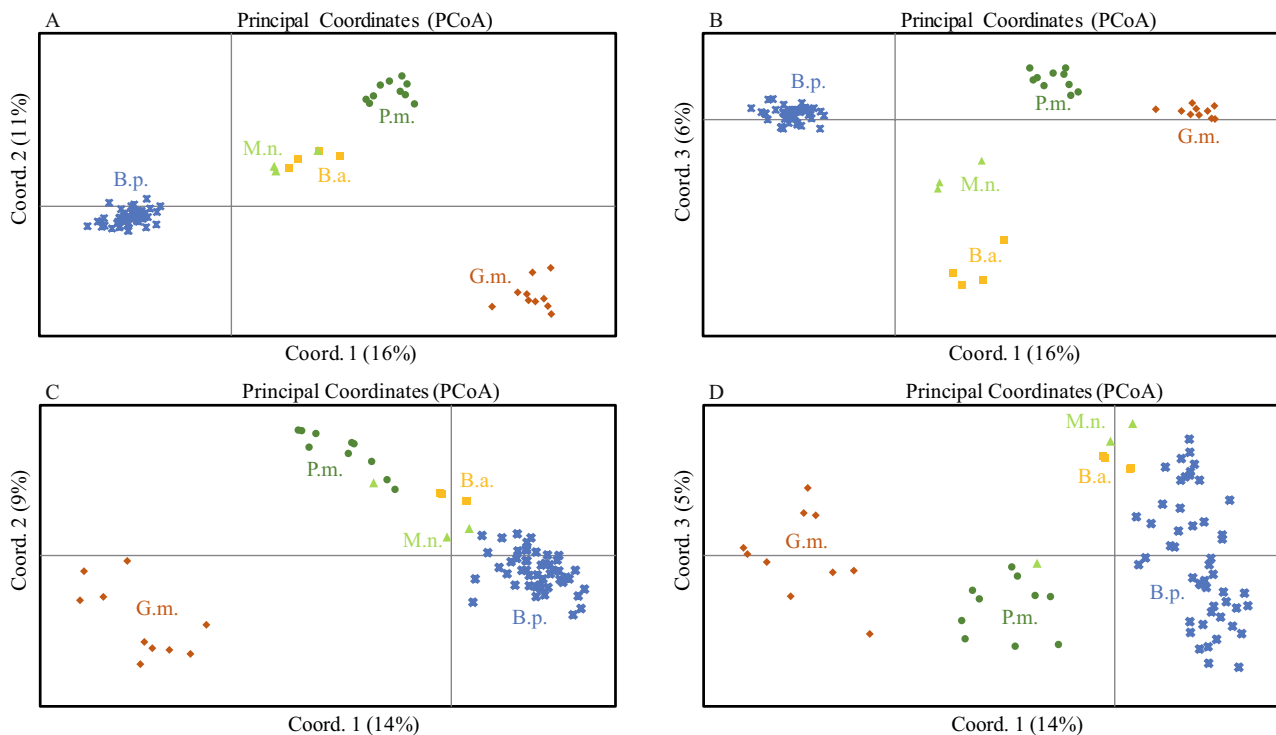
Size range is in base pairs; PA, private alleles

GATA417, and GT122; Table 1) per locus, with an average of 8.3 alleles per locus. In total, for all loci, 240 alleles were detected in *B. physalus*. Expected heterozygosity ( $H_E$ ) varied between 0.34 and 0.91 (mean  $\pm$  standard deviation values:  $0.75 \pm 0.13$ ) whereas observed heterozygosity ( $H_0$ ) ranged from 0.34 to 0.98 (mean  $\pm$  standard deviation values:  $0.75 \pm 0.13$ ; Table 1). The inbreeding coefficient varied from  $-0.146$  to  $0.231$ , with three loci significantly deviant from the Hardy Weinberg equilibrium

(HWE; loci TAA023, Bp1357166 and Bp779455) and with a significant probability of linkage disequilibrium too (13%, 13% and 23% respectively).

### Cross-species amplification

All 32 loci were also tested for amplification in four other cetacean species (*M. novaeangliae*, *B. acutorostrata*, *P. macrocephalus* and *G. melas*; Table 2). The success of the



**Fig. 1** Principal coordinates analysis on 5 cetacean species genotyped with newly developed and previously published microsatellite markers (32 microsatellite loci; **a** and **b**) or only with previously published (7 microsatellite loci; **c** and **d**). **a** Axis 1 (16%) and 2 (11%). **b** Axis

1 (16%) and 3 (6%). **c** Axis 1 (14%) and 2 (9%). **d** Axis 1 (14%) and 3 (5%). B.p., *Balaenoptera physalus*; M.n., *Megaptera novaeangliae*; B.a., *Balaenoptera acutorostrata*; P.m., *Physeter macrocephalus*; G.m., *Globicephala melas*

amplification varied among species, and 27 loci were amplified with success in all four species. Cross-amplification revealed a variation in polymorphism depending on the locus and the species from one to nine alleles per locus. The lowest percentages of polymorphic loci were for *B. acutorostrata* and *G. melas* with 72% and 75%, respectively. Furthermore, the five species presented private alleles, with numbers ranging from one to thirteen. *Balaenoptera physalus* presented a large number of private alleles, in total 105, with an average of 3.28 per locus. *Physeter macrocephalus* and *G. melas* revealed 39 and 31 private alleles, respectively; with an average of 1.22 and 0.97 private alleles per locus, while *M. novaeangliae* and *B. acutorostrata* showed 28 and 19 private alleles, respectively; with an average of 0.88 and 0.59 per locus. The number of private alleles was dependent on the locus. Loci Bp451306 and Bp1430262 allowed for the species to be distinguished, due to a large number of private alleles per species. Regarding the PCoA, the second axis separated the species into four groups (Fig. 1a). *Balaenoptera physalus*, *P. macrocephalus* and *G. melas* were isolated whereas *M. novaeangliae* and *B. acutorostrata* were congregated. The third axis clearly divided the specimens into five groups (Fig. 1b), corresponding to the five species. In contrast, the PCoA with the seven previously published microsatellite markers allowed odontocetes (*G. melas* and

*P. macrocephalus*) to be differentiated from mysticetes (*M. novaeangliae*, *B. acutorostrata* and *B. physalus*), but did not allow differentiation between the three mysticete species (Fig. 1c–AD). The addition of the new set of microsatellite markers thus provides a finer resolution and allows for all species to be differentiated.

## Discussion

In this study, 32 microsatellite markers were tested on five cetacean species. Twenty-five were newly developed for *B. physalus*. The 32 loci were amplified in five species, providing a valuable tool for future studies of cetacean populations.

On the 14 previously published markers, only seven could be amplified in this study, while all were verified by their authors on *B. physalus* with success. Locus GT142 was tested at five temperatures (53, 55, 57, 60 and 63 °C) for amplification, like all microsatellite markers, but none of the temperatures showed an amplification. The non-amplification of this locus can be due to mutations in the primer sites in our population of *B. physalus* [22], and the displacement of the primer could resolve the non-amplification of the locus GT142. For loci GATA053, GT271 and GT575, the non-amplification appeared when we added the fluorescent

dye, for each locus we tested only the temperature of their multiplex. A solution for these loci would have been to test all of the temperatures again, but at this stage, the multiplexes were already established and because we had enough loci to study the population, no further tests were performed. The last three loci that did not amplify (GATA028, EV01 and EV94) were lost during the last step, corresponding to the PCR multiplex with fluorescent dye on the primer. One reason for this may be the interaction between primers of two loci in multiplex PCR, which can lead to poor PCR sensitivity and specificity and/or the preferential amplification of non-target loci [23, 24].

The new set of microsatellite markers displayed numerous dinucleotide repeats (76%), compared to tri- (12%) and tetranucleotide repeats (12%), as also seen in others studies [14, 25]. In general, tri- and tetranucleotides are preferred to dinucleotides, since variations in the number of repeat units are easier to detect between individuals due to the larger number of base pair differences between alleles [26]. These differences remove potential biases such as stuttering and ambiguous allele sizing [27]. However, dinucleotides are also very informative since they usually exhibit higher number of alleles and higher polymorphism. In cetaceans, dinucleotide repeats were found to be significantly more polymorphic than tetranucleotide repeats [25]. In our case, dinucleotides and tetranucleotides have the same level of polymorphism, approximately 8 alleles per locus. Twenty-nine loci were used resulting in a total of 240 alleles in the dataset, which seems sufficiently robust to study the structure, parentage and history of the *B. physalus* population [28].

The heterozygosity levels were coherent with results published for *B. physalus* populations from other regions [29, 30]. Some of the previously published loci (EV37 [13], GATA417 [14] and GT211 [15]) displayed higher heterozygosity levels in our population (EV37: 0.9000; GATA417: 0.9800; GT211: 0.8200) than in other studies (e.g. EV37: 0.75; GATA417: 0.70; GT211: 0.64 in [31]). Only four loci presented significant heterozygote deficiency. The percentage of heterozygote deficiency in our population (14%) is less than in other cetacean species for the same region (*Stenella cæruleoalba*: 33% [32]; *P. macrocephalus*: 25% [33]). A high level of heterozygote deficiency could be explained by a significant inbreeding within the population or a 'Wahlund effect' due to sampling distinct populations [34]. The level of heterozygosity is linked to the level of linkage disequilibrium (LD), and markers with low heterozygosity have been demonstrated to show less LD than those with high heterozygosity [12]. Locus TAA023 was significantly deviant from HWE in our population in the Mediterranean Sea, but was not deviant in other regions (Gulf of California: [30]), contrary to loci GATA098, EV37 and GT211 which were not deviant in our population but were deviant in other regions.

The level of LD (14%) was similar to the one estimated for *B. physalus* in different regions (13% in [7]). This level depends on species and populations, and while some cetacean species show a level of LD below 2% [35] as in the present study, many species show no LD [36]. Globally, cetaceans have limited LD in comparison to other species: 78% for Hula painted frog [37], 35% for chimpanzee [12].

The new set of microsatellites included three loci with null alleles. One of them presented a departure from HWE that could be explained by null alleles [38]. Bp747752 showed null alleles, a departure from HWE and linkage disequilibrium, and should be removed for all further analyses. Regarding loci Bp1430262 and Bp1003953, analyses should be performed with and without them to detect any biases that they may trigger in the results. However, the presence of null alleles are dependent on the *B. physalus* population, and while some loci were detected as likely having null alleles in the Gulf of California [30] this was not the case in our region. The design of new primers could reveal new alleles for many of the homozygous individuals of these three loci.

The 32 microsatellites used in the present study showed a high probability of being polymorphic in different cetacean species (> 72%). Cetaceans have already demonstrated a success greater than 70% for cross-amplification between suborders, Odontoceti and Mysticeti [13]. Cetaceans, birds, and frogs are known to have microsatellites with a high degree of success with respect to cross-species amplification [39]. The successful cross-species amplification provides evidence of conserved DNA sequences bordering microsatellite regions for these taxa, confirming previous studies [40, 41]. The success or failure of cross-amplification does not depend on the taxonomic classification system [42]. However, some taxa show a low probability of having polymorphic loci through cross-amplification, such as fishes [43] and bivalves [11]. With regards to our results, the new set of loci could be useful for further studies on other cetacean species.

Furthermore, all loci revealed at least one private allele for one or several species. Some species have numerous private alleles for a single locus. *Balaenoptera physalus* presented the highest number of private alleles, due to the development of microsatellite markers specifically for this species. Next, *P. macrocephalus* and *G. melas*, the two species most phylogenetically distant, displayed the highest number of private alleles. Identification of private alleles showed that the microsatellite markers could be useful to distinguish species. The number of private alleles is a measure of genetic distinctiveness [44]. However, a large number of private alleles could also be an artefact of low sample size. Increasing the number of samples decreases the average frequency of private alleles in the species [45]. Nonetheless, the number of private alleles for *B. physalus*, which presents a higher sampling effort compared to the other species,

suggests that these microsatellite markers are very informative for the study of population structure in this species.

Microsatellite markers can help to detect hybridization events, when markers are numerous and informative. In the infraorder Cetacea, 62 putative hybridization events have been reported in captive and natural populations [46]. In captivity, all hybridizations occurred only in Odontoceti species and more precisely in the *Delphinoidea* superfamily while in the natural population, 13 concerned Mysticeti species. Our target species *B. physalus* is involved in 11 hybridizations with *Balaenoptera musculus*. A complementary use of nuclear and mitochondrial markers makes it possible to better identify the parental species [47]. Our microsatellite markers will likely enable the detection and study of hybridization within the *Balaenoptera* genus.

The cross-amplification revealed some genetic differentiation between species (Fig. 1). When considering only axes 1 and 2, two species (*M. novaeangliae* and *B. acutorostrata*) are grouped together (Fig. 1a), but are separated when considering the third axis (Fig. 1b). Cross-amplification was tested on two Odontoceti species (*P. macrocephalus* and *G. melas*) and three Mysticeti species (*B. physalus*, *B. acutorostrata* and *M. novaeangliae*). The emergence and diversification of Mysticeti and Odontoceti occurred 35 Ma ago [48]. This divergence time did not prevent most of the loci developed for *B. physalus* (Mysticeti) from amplifying in Odontoceti species.

Phylogenetic studies have incorporated *M. novaeangliae* within the *Balaenoptera* genus [49]. The proximity of *M. novaeangliae* and *B. acutorostrata* with our target species is congruent with the phylogeny. The two Odontoceti species are the most distant on the PCoA, with *G. melas*, the most isolated, belonging to Delphinidae. Overall, for cetaceans, the cross-amplification was successful regardless of genetic distances, but the success seems to be species-dependent in other groups [50].

The success of amplification (25 *de novo* and seven previously published) and the high degree of polymorphism detected suggests that these newly developed microsatellite markers will be suitable to infer genetic diversity, population and family structures. Furthermore, the cross-species transferability was successful and provides a new set of markers for four other cetacean species spread across the phylogeny, and will likely be useful for many other cetacean species. The availability of numerous microsatellite markers is critical for studying and understanding such ecologically important migratory populations.

**Acknowledgements** We thank WWF-France for providing samples from *Balaenoptera physalus*, *Physeter macrocephalus*, *Globicephala melas*. We are also grateful to all the volunteers of WWF-France who participated in the campaigns at sea. We thank W. Mullié in Mauritania for the collection of *Balaenoptera acutorostrata* samples. Thank you to J. Almany for english corrections.

## Compliance with ethical standards

**Conflict of interest** The authors declare that they have no conflict of interest.

**Ethical approval** The sampling was non-lethal and *a priori* approved by the DREAL (Direction Régionale de l'Environnement, de l'Aménagement et du Logement) of Provence-Alpes Côte d'Azur.

## References

1. Harwood J (2001) Marine mammals and their environment in the twenty-first century. *J Mammal* 82:630–640
2. Brownell RL, Yablokov AV (2009) Whaling, illegal and pirate. In: *Encyclopedia of marine mammals*, 2nd edn. Elsevier, Amsterdam, pp 1235–1239
3. Cooke JG (2018) *Balaenoptera physalus*. IUCN Red List Threat Species 2018 eT2478A50349982. <https://doi.org/10.2305/IUCN.UK.2018-2.RLTS.T2478A50349982.en>
4. Sanpera C, Aguilar A (1992) Modern whaling off the Iberian Peninsula during the 20th century. *Rep Int Whal Comm* 42:723–730
5. Forcada J, Aguilar A, Hammond PS et al (1996) Distribution and abundance of fin whales (*Balaenoptera physalus*) in the western Mediterranean sea during the summer. *J Zool* 238:23–34. <https://doi.org/10.1111/j.1469-7998.1996.tb05377.x>
6. Notarbartolo-di-Sciara G, Zanardelli M, Jahoda M et al (2003) The fin whale *Balaenoptera physalus* (L. 1758) in the Mediterranean Sea. *Mamm Rev* 33:105–150
7. Bérubé M, Aguilar A, Dendanto D et al (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. <https://doi.org/10.1046/j.1365-294x.1998.00359.x>
8. Weber JL, May PE (1989) Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. *Am J Hum Genet* 44:388–396
9. Lamb CT, Ford AT, Proctor MF et al (2019) Genetic tagging in the Anthropocene: scaling ecology from alleles to ecosystems. *Ecol Appl* 29:e01876. <https://doi.org/10.1002/eap.1876>
10. Clisson I, Lathuillière M, Crouau-Roy B (2000) Conservation and evolution of microsatellite loci in primate taxa. *Am J Primatol* 50:205–214. [https://doi.org/10.1002/\(SICI\)1098-2345\(20003\)50:3%3c205:AID-AJP3%3e3.0.CO;2-Y](https://doi.org/10.1002/(SICI)1098-2345(20003)50:3%3c205:AID-AJP3%3e3.0.CO;2-Y)
11. Hedgecock D, Li G, Hubert S et al (2004) Widespread null alleles and poor cross-species amplification of microsatellite DNA loci cloned from the pacific oyster, *Crassostrea gigas*. *J Shellfish Res* 23:379–385
12. Crouau-Roy B, Service S, Slatkin M, Freimer N (1996) A fine-scale comparison of the human and chimpanzee genomes: linkage, linkage disequilibrium and sequence analysis. *Hum Mol Genet* 5:1131–1137. <https://doi.org/10.1093/hmg/5.8.1131>
13. Valsecchi E, Amos W (1996) Microsatellite markers for the study of cetacean populations. *Mol Ecol* 5:151–156. <https://doi.org/10.1111/j.1365-294X.1996.tb00301.x>
14. Palsbøll PJ, Bérubé M, Larsen AH, Jørgensen H (1997) Primers for the amplification of tri- and tetramer microsatellite loci in baleen whales. *Mol Ecol* 6:893–895. <https://doi.org/10.1111/j.1365-294X.1997.tb00146.x>
15. Bérubé M, Jørgensen H, McEwing R, Palsbøll PJ (2000) Polymorphic di-nucleotide microsatellite loci isolated from the humpback whale, *Megaptera novaeangliae*. *Mol Ecol* 9:2181–2183. <https://doi.org/10.1046/j.1365-294X.2000.105315.x>

16. Bérubé M, Rew MB, Skaug H et al (2005) Polymorphic microsatellite loci isolated from humpback whale, *Megaptera novaeangliae* and fin whale, *Balaenoptera physalus*. *Conserv Genet* 6:631–636. <https://doi.org/10.1007/s10592-005-9017-5>
17. Megléc E, Pech N, Gilles A et al (2014) QDD version 3.1: a user-friendly computer program for microsatellite selection and primer design revisited: experimental validation of variables determining genotyping success rate. *Mol Ecol Resour* 14:1302–1313. <https://doi.org/10.1111/1755-0998.12271>
18. Zhang Z, Schwartz S, Wagner L, Miller W (2000) A greedy algorithm for aligning DNA sequences. *J Comput Biol* 7:203–214. <https://doi.org/10.1089/10665270050081478>
19. Van Oosterhout C, Hutchinson WF, Wills DPM, Shipley P (2004) MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. *Mol Ecol Notes* 4:535–538. <https://doi.org/10.1111/j.1471-8286.2004.00684.x>
20. Peakall R, Smouse PE (2012) GenALEX 6.5: genetic analysis in Excel. Population genetic software for teaching and research—an update. *Bioinformatics* 28:2537–2539. <https://doi.org/10.1093/bioinformatics/bts460>
21. Belkhir K, Borsa P, Chikhi L, et al (2004) GENETIX: Windows TM software for population genetics. Lab. Génome, Popul. Interact. CNRS Umr 5000, Univ. Montpellier II, Montpellier (France)
22. Paetkau D, Strobeck C (1995) The molecular basis and evolutionary history of a microsatellite null allele in bears. *Mol Ecol* 4:519–520. <https://doi.org/10.1111/j.1365-294X.1995.tb00248.x>
23. Hayden MJ, Nguyen TM, Waterman A et al (2008) Application of multiplex-ready PCR for fluorescence-based SSR genotyping in barley and wheat. *Mol Breed* 21:271–281. <https://doi.org/10.1007/s11032-007-9127-5>
24. Edwards MC, Gibbs RA (1994) Multiplex PCR: advantages, development, and applications. *Genome Res* 3:S65–S75. <https://doi.org/10.1101/gr.3.4.S65>
25. Bourret V, Macé M, Bonhomme M, Crouau-Roy B (2008) Microsatellites in cetaceans: an overview. *Open Mar Biol J* 2:38–42. <https://doi.org/10.2174/1874450800802010038>
26. Liu Z-W, Jarret RL, Kresovich S, Duncan RR (1995) Characterization and analysis of simple sequence repeat (SSR) loci in seashore paspalum (*Paspalum vaginatum Swartz*). *Theor Appl Genet* 91:47–52. <https://doi.org/10.1007/BF00220857>
27. Edwards A, Civitello A, Hammond HA, Caskey CT (1991) DNA typing and genetic mapping with trimeric and tetrameric tandem repeats. *Am J Hum Genet* 49:746–756
28. Rosenberg NA, Burke T, Elo K et al (2001) Empirical evaluation of genetic clustering methods using multilocus genotypes from 20 chicken breeds. *Genetics* 159:699–713
29. Bérubé M, Daniëlsdottir AK, Aguilar A, et al (2006) High rates of gene flow among geographic locations in North Atlantic fin whales (*Balaenoptera Physalus*)
30. Rivera-León VE, Urbán J, Mizroch S et al (2019) Long-term isolation at a low effective population size greatly reduced genetic diversity in Gulf of California fin whales. *Sci Rep* 9:12391. <https://doi.org/10.1038/s41598-019-48700-5>
31. Bérubé M, Urbán J, Dizon AE et al (2002) Genetic identification of a small and highly isolated population of fin whales (*Balaenoptera physalus*) in the Sea of Cortez, Mexico. *Conserv Genet* 3:183–190
32. Bourret V, Macé M, Crouau-Roy B (2007) Genetic variation and population structure of western Mediterranean and northern Atlantic *Stenella coeruleoalba* populations inferred from microsatellite data. *J Mar Biol Assoc UK* 87:265–269. <https://doi.org/10.1017/S0025315407054859>
33. Engelhaupt D, Hoelzel AR, Nicholson C et al (2009) Female philopatry in coastal basins and male dispersion across the North Atlantic in a highly mobile marine species, the sperm whale (*Physeter macrocephalus*). *Mol Ecol* 18:4193–4205
34. Crouau-Roy B (1988) Genetic structure of cave-dwelling beetles populations: significant deficiencies of heterozygotes. *Heredity* 60:321–327. <https://doi.org/10.1038/hdy.1988.49>
35. Frasier TR, Rastogi T, Brown MW et al (2006) Characterization of tetranucleotide microsatellite loci and development and validation of multiplex reactions for the study of right whale species (genus *Eubalaena*). *Mol Ecol Notes* 6:1025–1029. <https://doi.org/10.1111/j.1471-8286.2006.01417.x>
36. Cunha HA, Watts PC (2007) Twelve microsatellite loci for marine and riverine tucuxi dolphins (*Sotalia guianensis* and *Sotalia fluviatilis*). *Mol Ecol Notes* 7:1229–1231. <https://doi.org/10.1111/j.1471-8286.2007.01839.x>
37. Perl RGB, Geffen E, Malka Y et al (2018) Population genetic analysis of the recently rediscovered Hula painted frog (*Latonia nigriventer*) reveals high genetic diversity and low inbreeding. *Sci Rep* 8:5588. <https://doi.org/10.1038/s41598-018-23587-w>
38. Panova M, Makinen T, Fokin M et al (2008) Microsatellite cross-species amplification in the genus *Littorina* and detection of null alleles in *Littorina saxatilis*. *J Molluscan Stud* 74:111–117. <https://doi.org/10.1093/mollus/eym052>
39. Primmer CR, Painter JN, Koskinen MT et al (2005) Factors affecting avian cross-species microsatellite amplification. *J Avian Biol* 36:348–360. <https://doi.org/10.1111/j.0908-8857.2005.03465.x>
40. Coughlan J, Mirimin L, Dillane E et al (2006) Isolation and characterization of novel microsatellite loci for the short-beaked common dolphin (*Delphinus delphis*) and cross-amplification in other cetacean species. *Mol Ecol Notes* 6:490–492. <https://doi.org/10.1111/j.1471-8286.2006.01284.x>
41. Chen L, Yang G (2009) A set of polymorphic dinucleotide and tetranucleotide microsatellite markers for the Indo-Pacific humpback dolphin (*Sousa chinensis*) and cross-amplification in other cetacean species. *Conserv Genet* 10:697–700. <https://doi.org/10.1007/s10592-008-9618-x>
42. Avise JC, Johns GC (1999) Proposal for a standardized temporal scheme of biological classification for extant species. *Proc Natl Acad Sci* 96:7358–7363. <https://doi.org/10.1073/pnas.96.13.7358>
43. Barbara T, Palma-Silva C, Paggi GM et al (2007) Cross-species transfer of nuclear microsatellite markers: potential and limitations. *Mol Ecol* 16:3759–3767. <https://doi.org/10.1111/j.1365-294X.2007.03439.x>
44. Kalinowski ST (2004) Counting alleles with rarefaction: private alleles and hierarchical sampling designs. *Conserv Genet* 5:539–543. <https://doi.org/10.1023/B:COGE.0000041021.91777.1a>
45. Slatkin M (1985) Rare alleles as indicators of gene flow. *Evolution* 39:53–65
46. Bérubé M, Palsbøll PJ (2018) Hybridism. In: *Encyclopedia of marine mammals*. Elsevier, Amsterdam, pp 496–501
47. Amaral AR, Lovewell G, Coelho MM et al (2014) Hybrid speciation in a marine mammal: the clymene dolphin (*Stenella clymene*). *PLoS ONE* 9:e83645. <https://doi.org/10.1371/journal.pone.0083645>
48. Fordyce RE (2018) Cetacean evolution. In: *Encyclopedia of marine mammals*. Elsevier, Amsterdam, pp 180–185
49. Sasaki T, Nikaido M, Hamilton H et al (2005) Mitochondrial phylogenetics and evolution of mysticete whales. *Syst Biol* 54:77–90. <https://doi.org/10.1080/10635150590905939>
50. Peyran C, Planes S, Tolou N et al (2020) Development of 26 highly polymorphic microsatellite markers for the highly endangered fan mussel *Pinna nobilis* and cross-species amplification. *Mol Biol Rep*. <https://doi.org/10.1007/s11033-020-05338-1>

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.