

# UNIVERSAL BALEEN WHALE MICROSATELLITE PANEL FOR INDIVIDUAL IDENTIFICATION AND POWER TO DETECT PARENTAGE

Marcos Suárez-Menéndez<sup>1\*#</sup>, Martine Bérubé<sup>1,2\*#</sup>, Lutz Bachmann<sup>3</sup>, Peter Best<sup>4£</sup>, Mads Peter Heide-Jørgensen<sup>5</sup>, Veronique Lesage<sup>6</sup>, Tom Oosting<sup>1</sup>, Rui Prieto<sup>8</sup>, Christian Ramp<sup>7</sup>, Jooke Robbins<sup>2</sup>, Richard Sears<sup>7</sup>, Mónica A. Silva<sup>8</sup>, Marc Tollis<sup>9</sup>, Els Vermeulen<sup>4</sup>, Gísli A. Víkingsson<sup>10£</sup>, Øystein Wiig<sup>3</sup>, and Per J. Palsbøll<sup>1,2</sup>

\*Corresponding authors: marcos.sume@gmail.com and martine.c.berube@gmail.com

# Shared first authorships, £deceased

<sup>1</sup> Marine Evolution and Conservation, Groningen Institute of Evolutionary Life Sciences, University of Groningen, Groningen, The Netherlands. <sup>2</sup> Center for Coastal Studies, Provincetown, Massachusetts, USA. <sup>3</sup> Natural History Museum, University of Oslo, Oslo, Norway. <sup>4</sup> Mammal Research Institute Whale Unit, University of Pretoria, South Africa. <sup>5</sup> Greenland Institute of Natural Resources, Copenhagen, Denmark. <sup>6</sup> Maurice Lamontagne Institute, Mont-Joli, Quebec, Canada. <sup>7</sup> Mingan Island Cetacean Study, Saint-Lambert, Quebec, Canada. <sup>8</sup> Institute of Marine Sciences - Okeanos & Institute of Marine Research, University of the Azores, Horta, Portugal. <sup>9</sup> Biodesign Institute, Arizona State University, Tempe, Arizona, USA and School of Life Sciences, Arizona State University, Tempe, Arizona, USA and School of Informatics, Computing, and Cyber Systems, Northern Arizona University, Flagstaff, Arizona, USA. <sup>10</sup> Marine and Freshwater Research Institute, Program for Whale Research, Reykjavík, Iceland.

Keywords: STR loci, Mysticeti, odontoceti, multiplex, paternity

## ABSTRACT

Highly polymorphic single tandem repeat loci (STR, also known as microsatellite loci) remain a familiar, cost efficient class of markers for genetic analyses in ecology, behavior and conservation. We characterize a new universal set of ten STR loci (from 28 potential candidate loci) in seven baleen whale species, which are optimized for PCR amplification in two multiplex reactions along with a Y chromosome marker for sex determination. The optimized, universal set of STR loci provides an ideal starting point for new studies in baleen whales aimed at individual-based and population genetic studies, and facilitates data sharing among research groups. Data from the new STR loci were combined with genotypes from other published STR loci to assess the power to assign parentage (paternity) using exclusion in four species: fin whales, humpback whales, blue whales and bowhead whales. We argue that parentage studies should present a power analysis to demonstrate that the specific data are sufficiently informative to assign parentage with statistical rigor.

# INTRODUCTION

Conservation genetic assessments typically draw inferences from the degree of genetic diversity within and among populations or individuals. Estimating the degree of relatedness among individuals can yield insights into the social system, mating strategy and population structure of endangered species. During recent years, single nucleotide polymorphisms (SNPs) have gained in popularity relative to short tandem repeat loci (STR, also known as microsatellites (Tautz 1989), due to their greater abundance in genomes (Morin, Luikart, and Wayne 2004). However, genotyping STR loci in conservation is still common, and often the only viable alternative when resources are scarce or the research objective requires genotyping large sample sizes and highly polymorphic loci. The specific set of STR loci genotyped is often identified and characterized in the targeted species or from published resources in closely related species. As a result, different combinations of STR loci are genotyped among individual laboratories and studies (with some notable exceptions, such as salmonids, Moran et al. 2006). Although there are obvious downstream advantages to genotype “universal” sets of STR loci, such optimization is tedious, due to experimental issues (e.g., cross-species amplification, null alleles and ascertainment bias Primmer et al. 1996) and varying levels of polymorphism among species. However, the DNA sequences flanking STR loci are often conserved among closely related taxa, which facilitates genotyping of homologous loci in closely related species. We developed a set of ten STR loci and a sexing marker that can be genotyped in seven baleen whale species in two multiplex polymerase chain reactions (PCR, Saiki et al. 1986). Identifying and optimizing cross-species amplifying STR loci would not only facilitate the development of new studies, but also enable data sharing among laboratories and the re-use of published data.

A main use of STR loci is the identification of related individuals in a population (Blouin 2003). Employing different numbers of loci and the varying level of polymorphism among

loci has a stark effect on the power to infer accurate relationships between individuals. For example, individuals that are closely related to a putative offspring (e.g. siblings or its own offspring) may incorrectly be assigned as a parent if the number of STR loci is insufficient given the sample size. Most parentage assessments in wild populations studies were aimed at terrestrial mammals, such as brown bears (*Ursus arctos*, Shimozuru et al. 2022) or animals in captivity, such as the cultured giant groupers (*Epinephelus lanceolatus*, Weng et al. 2021). In *Mysticeti* (baleen whales), a few parentage studies have been performed in humpback whales (*Megaptera novaeangliae*, Cerchio et al. 2005; Cypriano-Souza et al. 2010; Clapham and Palsbøll 1997; Nielsen et al. 2001), minke whales (*Balaenoptera acutorostrata*, Skaug, Bérubé, and Palsbøll 2010), North Atlantic right whales (*Eubalaena glacialis*, Frasier et al. 2007) and southern right whales (*E. australis*, Carroll et al. 2012). Most studies do not assess the informativeness of the STR loci employed prior to performing parentage analysis and thus have little or no insight into the error rate (i.e., false positives). Here we present a procedure (and supply the associated code) to determine how many STR loci are needed for a rigorous assignment of parentage using parentage exclusion probabilities. Towards this specific goal, we selected STR loci from published sources and characterized 28 new STR loci, which we employed with previously published STR loci to assess the number of STR loci required for rigorous parentage assignment in fin whales (*Balaenoptera physalus*), humpback whales, blue whales (*Balaenoptera musculus*) and bowhead whales (*Balaena mysticetus*).

## MATERIALS AND METHODS

All tissue samples were collected as skin biopsy from free-ranging whales as described by Palsbøll et al. (1991). Samples were stored at -20 or -80 degrees Celsius (°C) in saturated NaCl with 25% dimethylsulphoxide (Amos and Hoelzel 1991).

## MOLECULAR ANALYSES

Total-cell DNA was extracted using either standard phenol-chloroform extractions (Sambrook and Russell 2001) or QIAGEN DNEasy™ extraction columns for animal tissue, following the manufacturer's instruction (QIAGEN Inc.). The quality of the DNA was assessed visually by electrophoresis through 0.7% agarose gel and the amount quantified with a Qubit™ following the manufacturer's instructions (Thermo Fisher Scientific Inc.). DNA extractions were adjusted to a final concentration at 10ng DNA/μL.

Candidate tetramer STR loci with the repeat motif GATA were identified using the software SCIROKO (ver. 3.4, Kofler et al. 2007) and the humpback whale genome assembly (Tollis et al. 2019). The search was conducted with the parameter settings: search mode (perfect repeats), minimum number of repeats (5), upper and lower bound of motif length (4), SSR-couple considerations (all).

Oligo-nucleotides for PCR amplification of each STR locus were designed using PRIMER3PLUS (ver. 3.2.6, Untergasser et al., 2012) with default parameter settings, except fixing the annealing temperature at 57 °C and the oligo-nucleotide length at 21 nucleotides (Table 1). For each pair of oligo-nucleotides, the forward oligo-nucleotide was extended with either a universal T7 or M13 DNA sequence (Schuelke 2000). The T7/M13 extension facilitated labeling of the amplification products with a fluorophore (6-FAM or HEX) of the complementary T7/M13 primer during PCR and hence detection during capillary electrophoresis on ABI 3730 Genetic Analyzer (Applied Biosystems Inc.).

The genomic coordinates of each STR locus were determined by aligning the oligo-nucleotides against the reference blue whale genome at the chromosome-level scaffolds (Bukhman et al. 2022) using BOWTIE2 (ver 2.3.5.1, Langmead and Salzberg 2012) with the parameter settings defined by the preset *–very-sensitive*. PCR amplification were conducted

under the following conditions in 10 $\mu$ L volumes, each with 10ng of genomic DNA, 67mM Tris-HCl, pH 8.8, 2mM MgCl<sub>2</sub>, 16.6mM (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, 10mM  $\beta$ -mercaptoethanol, 0.2mM dNTPs, 1 $\mu$ M of each oligo-nucleotides as well as 0.4 units of *Taq* DNA polymerase (New England Biolabs Inc.). PCR reactions were subjected to two minutes at 94 °C followed by 35 cycles each with 30 seconds at 94 °C, 90 seconds at 57 °C and 30 seconds at 72 °C; followed by 10 minutes at 68 °C. The initial quality of the PCR amplification products was first assessed by gel electrophoresis in 2% agarose and 1xTBE at 175 volts for 25 minutes. After staining with ethidium bromide, amplification products were visualized under UV light at 260 nm. All candidate STR loci were amplified by as described above except for the three oligo-nucleotides which were added in the following concentrations: 1 $\mu$ M of the unlabeled locus-specific oligo-nucleotides, 0.5 $\mu$ M of the 5'end-labeled (HEX or 6-FAM) oligo-nucleotide and 0.5 $\mu$ M of the M13/T7-extended (see Table S2), unlabeled locus-specific oligo-nucleotide. PCR reactions were subjected to two minutes at 94 °C followed by 10 cycles each consisting of 30 seconds at 94 °C, one minute at 57 °C, and 30 seconds at 72 °C followed by an additional 27 cycles each with 30 seconds at 94 °C, one minute at 55 °C and 30 seconds at 72 °C, followed by 10 minutes at 72 °C. The amplification products were separated by capillary electrophoresis on an ABI 3730 Genetic Analyzer (Applied Biosystems Inc.). The size of the amplification products was estimated using the size standard GeneScan™ 500 ROX™ dye size standard (Applied Biosystems Inc.) with GENEMAPPER ver 4.0 (Applied Biosystems Inc.). One of each pair of oligo-nucleotide pairs was labeled with either 6-FAM, HEX or NED for those candidate STR loci selected for further analysis (Table S1) and genotyped in 48 individuals from each of seven mysticete species; humpback whale, fin whale, minke whale, sei whale (*B. borealis*), blue whale, southern right whale and bowhead whale.

Multiplex PCRs were performed using the Qiagen™ Multiplex Kit Plus (Qiagen Inc.) in 5μL reaction volumes following the manufacturer's recommendations. PCR reactions were subjected to two minutes at 94 °C, followed by 35 cycles each with 30 seconds at 94 °C, 90 seconds at 57 °C and 30 seconds at 72 °C, followed by 10 minutes at 68 °C. PCR amplification products were separated and detected by capillary gel electrophoresis on an ABI 3730 Genetic Analyzer (Applied Biosystems Inc.) including a size standard GeneScan™-500 ROX (Applied Biosystems Inc.). The length of each PCR product was determined with GENEMAPPER™ (ver. 4.1, Applied Biosystems Inc.).

Finally, a random set of 60 DNA extractions collected from different individual fin, humpback, bowhead and blue whales were genotyped at 30, 32, 22 and 20 loci, respectively, and the sex determined by co-amplification of a Y chromosome specific marker (Table 2) for the parent and offspring (PO) and sire-dam-offspring trio (SDO) assignments.

## DATA ANALYSIS

Allele frequencies, the observed ( $H_O$ ) and expected ( $H_E$ ) heterozygosity, as well as the probability of identity for pairs of unrelated individuals ( $P(I)$ , Paetkau and Strobeck 1994) corrected for low sample sizes were implemented in GIMLET (ver. 1.3.3, Valière 2002). The deviations from the expected Hardy-Weinberg genotype frequencies (HW) was estimated using GENALEX (ver. 6.5, Peakall and Smouse 2012; 2006). The non-exclusion probability of the first parent ( $P_{\text{NON-EXCL}}$ , i.e., the probability of failing to exclude an unrelated, non-parent, Selvin 1980) was estimated using CERVUS (ver. 3.0.7, Kalinowski, Taper, and Marshall 2007).

## PARENT-OFFSPRING ASSIGNMENT

PO pairs were identified using custom Python (ver. 3) scripts (see Data availability). The scripts identify all pairs of multilocus genotypes that segregate according to Mendelian

expectations for parent and offspring. SDO trios were identified in a similar manner, i.e., by first identifying putative dam and offspring pairs among all possible pairs of multilocus genotypes requiring at least one sample to be from a female. Subsequently, putative sires were identified among male genotypes that were consistent with the putative dam and offspring genotypes and Mendelian segregation. The assessment did not allow for genotyping errors, i.e., putative parent and offspring pairs with loci that did not segregate according to Mendelian expectations were rejected.

KININFOR (ver. 2, Wang 2006) was employed to assess the informativeness of the different relatedness and relationship estimators. Since most relatedness estimators correlate (Table S4), the “informativeness of relationship” (IR) criterion was used to rank the STR loci in terms of their statistical power to discern among different degrees of relatedness. Two assessments were conducted to assess the power of a given number of STR loci to infer PO pairs. The first assessment was based upon the most or least informative STR loci (ranked by their IR value, see above); the second on a random sample of loci sampled among all loci genotyped and without replacement. In both assessments, data sets from eight to the maximum number of loci genotyped were generated and for each set of loci PO pairs and SDO trios identified as described above. The assessment of the effect of locus IR values on the ability to identify PO pairs and SDO trios was conducted in two ways, starting with the loci with the highest and lowest IR values, respectively.

#### **ASSESSMENTS OF PATERNITY ASSIGNMENTS**

Close relatives to the putative offspring are most likely to be incorrectly assigned as the sire. Behavioral observations (e.g., Frasier et al. 2007) and genetic studies (e.g., Clapham and Palsbøll 1997) suggest that baleen whales are promiscuous and mating randomly. Although there is a possibility that a dam might mate with the same sire more than once and thus



produce offspring that are related as full-siblings, such instances are likely rare unless the overall population size is very small. This implies that first (the offspring's own offspring) and second order relatives (the offspring's grandparents, grandchildren and half-siblings) are those individuals that are most likely to be incorrectly assigned as the sire. Since each offspring has more second order than first order relatives, we focused our assessment of the probability of incorrect paternity assignments on second order relatives. Using half-siblings as a representative of a second order relative, we assessed the probability of incorrectly assigning a half-sibling as the sire by generating simulated data sets during which *in silico* pairs of half-siblings were generated by randomly sampling alleles at each STR and sex locus from *in silico* sires and dams. The probability of an incorrect paternity assignment in each assessment was estimated as the proportion of half-siblings without any loci violating the Mendelian expectations of being a putative sire (allowing for a maximum of two Mendelian violations). The assessment was conducted with a maximum of 50 STR sampled at random, with replacement, from the observed loci. For each set of STR loci, 5,000 calves and half-siblings were simulated, and the assessment repeated 50 times. The median probability of an incorrect paternity assignment for each set of STR loci was smoothed using a Savitzky-Golay filter (Savitzky and Golay 1964) from which the “knee of the curve”, i.e. the point where adding additional STR loci had a diminished effect, was estimated using the Python package *kneed* (ver. 0.8.2, Satopaa et al. 2011).

## RESULTS

### NEWLY CHARACTERIZED STR LOCI

A total of 5,047 STR loci with the repeat motif GATA and minimum of five repeats was detected. We selected 28 STR loci which were deemed (a) of sufficient quality on the basis of the gel electrophoresis (above) and (b) likely polymorphic in mysticetes. These 28 selected

STR loci were genotyped in eight DNA extractions in each species (Table 1). The majority of the STR loci amplified in the *Balaenopteridae* species, but less so within *Eubalaena* and *Balaena* (Table S1). Among the 28 candidate loci, ten STR loci that were polymorphic in most baleen whales were selected for further characterization. The ten candidate STR loci and a sex specific Y chromosome-specific locus were optimized for multiplexed PCR amplification. Amplification was conducted in two reactions; one with five STR loci and a second with five STR loci and the Y chromosome-specific locus (Table 2). In total, 48 DNA extractions from each mysticete species were genotyped with these two panels. Among the ten new loci, three loci were monomorphic in some species: GATA25072 (bowhead whale), GATA5947654 (southern right whale) and GATA6237777 (blue whale).  $H_E$  was similar among species (Table 3). The allele sizes at the loci GATA5947654 differed in the blue whale, and GATA5984139 differed in the bowhead whale and the southern right whale compared to the other species (Table S1). For the seven sets of 48 samples, the  $P(I)$  was estimated and ranged from  $4.17 \times 10^{-3}$  to  $8.10 \times 10^{-1}$ . Samples with identical genotypes were inferred as duplicate samples. One pair of duplicates was found in the blue, sei, humpback and right whales. Only one sample was kept for the data analysis. The above aspects are tabulated in Table 3.

## PARENTAGE ASSIGNMENTS

In order to evaluate the impact of both the quantity and quality of genetic markers on the parentage inference, a total of 1,300 datasets were generated by selecting various combinations of loci from a dataset of 60 humpback whales individuals (30 females and 30 males) typed at 32 loci with no missing data. The selected 60 samples contained no duplicates.  $P_{\text{NON-EXCL}}$  was estimated from samples of 60 randomly selected individuals in each of four species; fin, humpback, bowhead and blue whale (Table S2). The total of 32 STR loci genotyped in humpback whales represented a comparatively high level of polymorphisms with

a mean locus  $H_E$  and  $P_{\text{NON-EXCL}}$  at 0.75 and  $2.6 \times 10^{-7}$ , respectively. The same statistics were estimated for 30 loci in fin whales (mean locus  $H_E$  and  $P_{\text{NON-EXCL}}$  at 0.77 and  $7 \times 10^{-8}$ , respectively), 20 loci in bowhead whales (mean locus  $H_E$  and  $P_{\text{NON-EXCL}}$  at 0.74 and  $1.3 \times 10^{-5}$ , respectively) and 22 loci in blue whales (mean locus  $H_E$  and  $P_{\text{NON-EXCL}}$  at 0.68 and  $6.2 \times 10^{-4}$ , respectively, Table S2).

PO pairs were identified among the 60 humpback whales genotyped at 32 STR loci as pairs of samples with the expected Mendelian segregation of alleles. If no missing data or mismatches were allowed, 17 PO pairs in total were detected. No additional (incorrect) PO assignments were detected if the assessment was based on 14 or more STR loci with the highest IR values, whereas a total of 27 loci among the least informative loci was required to achieve the same result (Table 4 and S3).

The number of detected putative PO pairs using the minimum number of loci (eight) selected at random among the 32 loci varied between 34 to 104 (median 57). Both the median and range of the number of putative PO pairs declined with increasing number of loci. At 14 loci, a single of the 50 datasets yielded the same 17 PO pairs identified with the full data set (i.e., 32 loci). In this instance, the 14 loci at random happened to include all the most informative loci (as described above). At 24 or more loci, the median number of detected PO pairs was 17 PO (Figure 1, top panel).

A similar pattern was observed when detecting SDO trios. Three SDO trios were identified among the 60 humpback whales in the full data set (32 STR loci). Datasets with only eight loci detected between three and 16 trios (median of eight) with an outlier at 24 trios. As expected the minimum of three trios was detected in all datasets. At 20 loci or higher, the median number of SDO trios detected was three (Figure 1, bottom panel).

The relative percentage of half-siblings assigned as sire was consistent with the combined exclusion probabilities for the different datasets in each species, with higher probabilities of incorrect assignments in blue whales and humpbacks compared to bowhead whales (Figure 2). The “knee” of the curve was at 15 and 17 loci at zero or one Mendelian violation in the bowhead whale; 15 and 17 in the fin whale; 16 and 17 in the humpback; and 15 and 17 in the blue whale. Allowing two Mendelian violations (or missing data) increased the critical number of loci to 21 in all but the fin whales for which it was 19. Allowing a few Mendelian violations (or missing data) had a significant impact on the probability of incorrect assignments, e.g., a 32 to 79-fold difference between zero and two Mendelian violations (or missing data) at 18 loci. The probability of an incorrect assignment approached zero at 24 loci in the fin whale, 26 in the humpback whale, 28 loci in the bowhead and at 30 loci in the blue whale.

## DISCUSSION AND CONCLUSIONS

Multi-allelic STR loci remains a cost-efficient alternative to genomic approaches compared to whole genome sequencing and reduced representation methods, such as genotype-by-sequencing (e.g., Peterson et al. 2012). The low cost per sample makes STR genotyping an efficient means by which to screen large sample sizes for duplicate samples and pairs of first order relatives. PO pairs, and especially full pedigrees (i.e., SDO trios), are interesting for many purposes, such as kinship mark-recapture (Palsbøll 1999; Skaug 2001).

The increasing access to reference genomes further facilitates identification of STR loci in a cost effective manner. In this study, 28 “novel” STR loci were identified among ~5,000 candidate STR loci in the humpback reference genome from which a genotyping assay comprising ten STR loci and a Y chromosome-specific locus were amplified in two multiplex PCR amplifications in most baleen whale species. Standardization of a basic set of STR loci applicable to a group of closely related species, simplifies initial genotyping within single

research groups (incl. species identification as shown here) and enables sharing of datasets among laboratories working on the same species. Although the idea of developing multiplex panels to facilitate data sharing among laboratories working within the same group of species is not novel (Chambers et al. 2004) and has obvious long-term advantages, they remain rare (Beugin et al. 2017; Moran et al. 2006).

The second goal in this study was to facilitate the identification of PO pairs and SDO trios with a high level of confidence. The power of STR loci to discern between PO pairs and other close relationship categories is a function of the level of polymorphism at each included locus, the mating system of the targeted species, as well as reproductive rates. Some studies have inferred parentage employing only three STR loci (Zane et al. 1999) and among baleen whales assessments have been conducted with as few as nine to 13 STR loci (Cypriano-Souza et al. 2010; Carroll et al. 2012). In this study we demonstrate that the number of STR loci required to identify PO pairs with statistical rigor (using parentage exclusion probabilities) can be assessed in a relatively manner, hence maximizing the power of downstream parentage analysis. Unsurprisingly, we show that it is necessary to not only consider the number of STR loci but also the degree of polymorphism at each locus. The difference, in terms of loci to be genotyped to make rigorous inferences, were substantial between the loci with the lowest and highest IR values. Homologous STR loci differ in informativeness among species and populations, e.g., due to a high degree of relatedness among sampled individuals, which has a detrimental effect on the power to identify PO pairs and SDO trios. Thus, studies should be accompanied by a power analysis of the kinds shown here, for each specific dataset and reported when publishing the results. Currently only few parentage assessments undertake such a power analysis or accept high false positive rates (e.g., the 80% or 95% "confidence" options available in CERVUS). Some studies reject PO pairs and SDO trios with two or more putative parent/sire, but accept PO pairs and SDO trios where only a single putative parent/sire

was identified (e.g., Carroll et al. 2012; Frasier et al. 2007), thus ignoring the unsampled part of the population and the possibility that the identified parent/sire could be a non-parental close relative.

Genotyping errors might result in incorrect parentage/paternity exclusions and thus it is tempting to allow for a few Mendelian violations when assigning parentage/paternity.

However, allowing a few Mendelian violations may result in a significant number of incorrect parentage/paternity assignments, especially when the number of loci or degree of polymorphism is low as shown here. As it is straightforward to assess the consequences of allowing some Mendelian violations on the rate of incorrect parentage/paternity assignments, the effect should be assessed and reported when interpreting the results. In the same vein, missing genotypes are common and often not reported or under-reported in many studies (e.g., Gerber et al. 2022). Putative genotyping errors are readily re-genotyped and checked.

Consequently, the optimal approach is to check as many Mendelian violations as possible in the base data, rather than including such likely errors in the subsequent data analysis.

Null-alleles (Paetkau and Strobeck 1995) could be the cause of loci violating Mendelian segregation, but are easily detected (both individuals homozygous for different alleles) and fixed (redesign PCR primers, and re-genotype all homozygous individuals).

Currently, missing genotypes are usually simply reported as the overall percentage of missing genotypes in the dataset, or as individuals with more than a minimum number loci. Since the effect of employing a reduced number of loci is likely to increase the number of incorrect parentage/paternity assignments, the effect of the inclusion of samples with the minimum allowed number genotypes (and their level of polymorphism) should be assessed as well to ascertain that the minimum number of genotypes is sufficient to identify PO pairs and SDO trios with a reasonable power.

In conclusion, we presented a new set of 10 STR loci and one Y chromosome marker to be amplified in two multiplexes in baleen whales. This set of markers serves as an easy, optimized starting point to conduct individual-based and population genetic studies in baleen whales, producing data that later can be shared among research groups. We also provide extended sets of STR loci (from previously published sources) with which to conduct rigorous parentage/paternity assignments along with Python scripts to assess the statistical power of specific sets of STR loci in order to ascertain the power to exclude non-parental genotypes.

#### **DATA AVAILABILITY STATEMENT**

All scripts are available at <https://github.com/MSuarezMenendez/PaternityExclusion>. The genotype data are available upon request.

#### **AUTHOR CONTRIBUTIONS**

MSM conducted the data analysis and drafted the manuscript. MB performed DNA extractions, selected markers, optimized PCR reactions, conducted data analysis and drafted the manuscript. PJP designed primers, discussed results and revised all drafts of the manuscripts. LB and ØW provided extracted DNA. TO genotyped the selected markers in DNA samples from different species. MT provided access to genome data. PB, MPHJ, VL, CR, JR, RS, MAS and GV provided samples. All co-authors revised and approved the final manuscript.

#### **ACKNOWLEDGEMENTS**

MSM was supported by a doctorate fellowship from the University of Groningen. MAS was supported by AZORES 2020, through the EU Fund 01-0145-FEDER-000140. The authors would like to thank Mariel T.I ten Doeschate for samples contribution and W. Hao for her technical assistance.



## REFERENCES

- Amos, William, and A R Hoelzel. 1991. 'Long-Term Preservation of Whale Skin for DNA Analysis'. *Reports of the International Whaling Commission Special Issue* 13: 99–104.
- Beugin, Marie-Pauline, Jérôme Letty, Cécile Kaerle, Jean-Sébastien Guitton, Lina Muselet, Guillaume Queney, and Dominique Pontier. 2017. 'A Single Multiplex of Twelve Microsatellite Markers for the Simultaneous Study of the Brown Hare (*Lepus Europaeus*) and the Mountain Hare (*Lepus Timidus*)'. *Ecology and Evolution* 7 (11): 3931–39. <https://doi.org/10.1002/ece3.2943>.
- Blouin, Michael S. 2003. 'DNA-Based Methods for Pedigree Reconstruction and Kinship Analysis in Natural Populations'. *Trends in Ecology & Evolution* 18 (10): 503–11. [https://doi.org/10.1016/S0169-5347\(03\)00225-8](https://doi.org/10.1016/S0169-5347(03)00225-8).
- Bukhman, Yury V, Phillip A. Morin, Susanne Meyer, Li-Fang (Jack) Chu, Jeff K. Jacobsen, Jessica Antosiewicz-Bourget, Daniel Mamott, et al. 2022. 'A High-Quality Blue Whale Genome, Segmental Duplications, and Historical Demography'. Preprint. In Review. <https://doi.org/10.21203/rs.3.rs-1910240/v1>.
- Carroll, Emma L., Simon J. Childerhouse, Mark Christie, Shane Lavery, Nathalie Patenaude, Alana Alexander, Rochelle Constantine, Debbie Steel, Laura Boren, and C. Scott Baker. 2012. 'Paternity Assignment and Demographic Closure in the New Zealand Southern Right Whale'. *Molecular Ecology* 21 (16): 3960–73. <https://doi.org/10.1111/j.1365-294X.2012.05676.x>.
- Cerchio, Salvatore, Jeff K. Jacobsen, Danielle M. Cholewiak, Erin A. Falcone, and D. Andrew Merriwether. 2005. 'Paternity in Humpback Whales, *Megaptera Novaeangliae*: Assessing Polygyny and Skew in Male Reproductive Success'. *Animal Behaviour* 70 (2): 267–77. <https://doi.org/10.1016/j.anbehav.2004.10.028>.
- Chambers, Karen E., Ulrich H. Reichard, Asja Möller, Katrin Nowak, and Linda Vigilant. 2004. 'Cross-Species Amplification of Human Microsatellite Markers Using Noninvasive Samples from White-Handed Gibbons (*Hylobates Lar*)'. *American Journal of Primatology* 64 (1): 19–27. <https://doi.org/10.1002/ajp.20058>.
- Clapham, P. J., and P. J. Palsbøll. 1997. 'Molecular Analysis of Paternity Shows Promiscuous Mating in Female Humpback Whales (*Megaptera Novaeangliae*, Borowski)'. *Proc. R. Soc. Lond. Ser. B* 264: 95–98.
- Cypriano-Souza, Ana Lúcia, Gabriela P. Fernández, Carlos André V. Lima-Rosa, Márcia H. Engel, and Sandro L. Bonatto. 2010. 'Microsatellite Genetic Characterization of the Humpback Whale (*Megaptera Novaeangliae*) Breeding Ground off Brazil (Breeding Stock A)'. *Journal of Heredity* 101 (2): 189–200. <https://doi.org/10.1093/jhered/esp097>.
- Frasier, Timothy R, P K Hamilton, M W Brown, L A Conger, A R Knowlton, Marilyn Marx, C K Slay, S D Kraus, and Bradley N White. 2007. 'Patterns of Male Reproductive Success in a Highly Promiscuous Whale Species: The Endangered North Atlantic Right Whale'. *Molecular Ecology* 16 (24): 5277–93. <https://doi.org/10.1111/j.1365-294X.2007.03570.x>.
- Gerber, Livia, Richard C. Connor, Simon J. Allen, Kay Horlacher, Stephanie L. King, William B. Sherwin, Erik P. Willems, Samuel Wittwer, and Michael Krützen. 2022. 'Social Integration Influences Fitness in Allied Male Dolphins'. *Current Biology* 32 (7): 1664–1669.e3. <https://doi.org/10.1016/j.cub.2022.03.027>.
- Kalinowski, Steven T., Mark L. Taper, and Tristan C. Marshall. 2007. 'Revising How the Computer Program Cervus Accommodates Genotyping Error Increases Success in



- Paternity Assignment'. *Molecular Ecology* 16 (5): 1099–1106.  
<https://doi.org/10.1111/j.1365-294X.2007.03089.x>.
- Kofler, Robert, Christian Schlötterer, Tamas Lelley, and Affiliations. 2007. 'SciRoKo: A New Tool for Whole Genome Microsatellite Search and Investigation'. *Bioinformatics* 23 (13): 1683–85. <https://doi.org/doi:10.1093/bioinformatics/btm157>.
- Langmead, Ben, and Steven L. Salzberg. 2012. 'Fast Gapped-Read Alignment with Bowtie 2'. *Nature Methods* 9 (4): 357–59. <https://doi.org/10.1038/nmeth.1923>.
- Moran, P., D. J. Teel, E. S. LaHood, J. Drake, and S. Kalinowski. 2006. 'Standardising Multi-Laboratory Microsatellite Data in Pacific Salmon: An Historical View of the Future'. *Ecology of Freshwater Fish* 15 (4): 597–605.  
<https://doi.org/10.1111/j.1600-0633.2006.00201.x>.
- Morin, Phillip A., Gordon Luikart, and Robert K. Wayne. 2004. 'SNPs in Ecology, Evolution and Conservation'. *Trends in Ecology and Evolution* 19 (4): 208–16.  
<https://doi.org/10.1016/j.tree.2004.01.009>.
- Nielsen, R., D.K. Mattila, P. J. Clapham, and P. J. Palsbøll. 2001. 'Statistical Approaches to Paternity Analysis in Natural Populations and Applications to the North Atlantic Humpback Whale.' *Genetics* 157 (4): 1673–82.
- Paetkau, D., and C. Strobeck. 1994. 'Microsatellite Analysis of Genetic Variation in Black Bear Populations.' *Molecular Ecology* 3: 489–95.  
<https://doi.org/10.1111/j.1365-294X.1994.tb00127.x>.
- . 1995. 'The Molecular Basis and Evolutionary History of a Microsatellite Null Allele in Bears'. *Molecular Ecology* 4 (4): 519–20.  
<https://doi.org/10.1111/j.1365-294X.1995.tb00248.x>.
- Palsbøll, Per J. 1999. 'Genetic Tagging: Contemporary Molecular Ecology'. *Biological Journal of the Linnean Society* 68 (1–2): 3–22.  
<https://doi.org/10.1111/j.1095-8312.1999.tb01155.x>.
- Palsbøll, Per, F. Larsen, and E. Sigurd-Hansen. 1991. 'Sampling of Skin Biopsies from Free-Ranging Large Cetaceans at West Greenland: Development of Biopsy Tips and New Designs of Bolts'. *Reports of the International Whaling Commission Special Issue* 13: 71–79.
- Peakall, R., and P.E. Smouse. 2012. 'GenAlEx 6: Genetic Analysis in Excel. Population Genetic Software for Teaching and Research - an Update'. *Bioinformatics Advance Access*.
- Peakall, R., and Peter E. Smouse. 2006. 'Genalex 6: Genetic Analysis in Excel. Population Genetic Software for Teaching and Research'. *Molecular Ecology Notes* 6 (1): 288–95. <https://doi.org/10.1111/j.1471-8286.2005.01155.x>.
- Peterson, Brant K, Jesse N Weber, Emily H Kay, Heidi S Fisher, and Hopi E Hoekstra. 2012. 'Double Digest RADseq: An Inexpensive Method for de Novo SNP Discovery and Genotyping in Model and Non-Model Species.' *Plos One* 7 (5): e37135.  
<https://doi.org/10.1371/journal.pone.0037135>.
- Primmer, Craig R., Hans Ellegren, Nicola Saino, and Anders Pape Møller. 1996. 'Directional Evolution in Germline Microsatellite Mutations'. *Nature Genetics* 13 (4): 391–93.  
<https://doi.org/10.1038/ng0896-391>.
- Saiki, R K, T L Bugawan, G T Horn, K B Mullis, and H A Erlich. 1986. 'Analysis of Enzymatically Amplified SS-Globin and HLA-DQaDNA with Allele-Specific Oligonucleotide Probes'. *Nature* 324: 163–66.
- Sambrook, J, and D W Russell. 2001. *Molecular Cloning. A Laboratory Manual*. Third. Vol. 1. New York: Cold Spring Harbor Laboratory Press.
- Satopaa, Ville, Jeannie Albrecht, David Irwin, and Barath Raghavan. 2011. 'Finding a "Kneedle" in a Haystack: Detecting Knee Points in System Behavior'. In *2011 31st*

- International Conference on Distributed Computing Systems Workshops*, 166–71. Minneapolis, MN, USA: IEEE. <https://doi.org/10.1109/ICDCSW.2011.20>.
- Savitzky, Abraham., and M. J. E. Golay. 1964. ‘Smoothing and Differentiation of Data by Simplified Least Squares Procedures.’ *Analytical Chemistry* 36 (8): 1627–39. <https://doi.org/10.1021/ac60214a047>.
- Schuelke, M. 2000. ‘An Economic Method for the Fluorescent Labeling of PCR Fragments’. *Nature Biotechnology* 18: 1–2. <https://doi.org/DOI:10.1038/71855>.
- Selvin, Steve. 1980. ‘Probability of Nonpaternity Determined by Multiple Allele Codominant Systems’. *American Journal of Human Genetics* 32 (2): 276–78.
- Shimozuru, Michito, Mina Jimbo, Keisuke Adachi, Kei Kawamura, Yuri Shirane, Yoshihiro Umemura, Tsuyoshi Ishinazaka, et al. 2022. ‘Estimation of Breeding Population Size Using DNA-Based Pedigree Reconstruction in Brown Bears’. *Ecology and Evolution* 12 (9): e9246. <https://doi.org/10.1002/ece3.9246>.
- Skaug, H J. 2001. ‘Allele-Sharing Methods for Estimation of Population Size’. *Biometrics* 57 (3): 750–56.
- Skaug, H. J., M. Bérubé, and P. J. Palsbøll. 2010. ‘Detecting Dyads of Related Individuals in Large Collections of DNA-Profiles by Controlling the False Discovery Rate’. *Molecular Ecology Resources* 10 (4): 693–700. <https://doi.org/10.1111/j.1755-0998.2010.02833.x>.
- Tautz, Diethard. 1989. ‘Hypervariability of Simple Sequences as a General Source for Polymorphic DNA Markers’. *Nucleic Acids Research* 17 (16): 6463–71. <https://doi.org/10.1093/nar/17.16.6463>.
- Tollis, Marc, Jooke Robbins, Andrew E Webb, Lukas F K Kuderna, Aleah F Caulin, Jacinda D Garcia, Martine Berube, et al. 2019. ‘Return to the Sea, Get Huge, Beat Cancer: An Analysis of Cetacean Genomes Including an Assembly for the Humpback Whale (*Megaptera Novaeangliae*)’. *Molecular Biology and Evolution* 36 (8): 1746–63. <https://doi.org/10.1093/molbev/msz099>.
- Untergasser, A., I. Cutcutache, T. Koressaar, J. Ye, B.C. Faircloth, M. Remm, and S.G. Rozen. 2012. ‘Primer3--New Capabilities and Interfaces.’ *Nucleic Acids Research* 40 (15).
- Valière, Nathaniel. 2002. ‘Gimlet: A Computer Program for Analysing Genetic Individual Identification Data’. *Molecular Ecology Notes* 2 (3): 377–79. <https://doi.org/10.1046/j.1471-8286.2002.00228.x-i2>.
- Wang, Jinliang. 2006. ‘Informativeness of Genetic Markers for Pairwise Relationship and Relatedness Inference’. *Theoretical Population Biology* 70 (3): 300–321. <https://doi.org/10.1016/j.tpb.2005.11.003>.
- Weng, Zhuoying, Yang Yang, Xi Wang, Lina Wu, Sijie Hua, Hanfei Zhang, and Zining Meng. 2021. ‘Parentage Analysis in Giant Grouper (*Epinephelus lanceolatus*) Using Microsatellite and SNP Markers from Genotyping-by-Sequencing Data’. *Genes* 12 (7): 1042. <https://doi.org/10.3390/genes12071042>.
- Zane, Nelson, Jones, and Avise. 1999. ‘Microsatellite Assessment of Multiple Paternity in Natural Populations of a Live-bearing Fish, *Gambusia Holbrooki*’. *Journal of Evolutionary Biology* 12 (1): 61–69. <https://doi.org/10.1046/j.1420-9101.1999.00006.x>.

## Tables

Table 1. Primer sequences and genomic coordinates of the 28 new STR loci used in the study.

Locus	PCR fragment size	Primer designation	Scaffold <sup>&amp;</sup>	Position <sup>&amp;</sup>	Oligo-nucleotide sequence (5' - 3')
GATA19406	295	GATA19406F <sup>s</sup> GATA19406R	CM020949.2	56,264,761	AGGTTTAGTGAGGATCCTTCC TGCACATCTTGTAGCTGTGTT
GATA25072	407	GATA25072F GATA25072R	CM020942.2	158,399,714	TGGACACATTTAAGGGGATAA AACTTGATTTCGCCTTACTTTG
GATA29055	159	GATA29055F GATA29055R	CM020942.2	130,013,628	GACTGGTGTTTCTCTGGAGAA GACTTACCAGCCCCTACAAAT
GATA36068	353	GATA36068F GATA36068R	CM020947.2	92,217,069	CCAAATTGCTCTCAAGAAAGA CTTTGGAGATCACCGTTTAGA
GATA3635	310	GATA3635F GATA3635R	CM020946.2	34,123,516	TCAAATATGGGGAGAAAAACA TATTTATGCTTTTTTGCCCATC
GATA38314	331	GATA38314F GATA38314R	CM020946.2	42,779,213	AGGAGACAGAAAACACGACTG TACACAGGAACCTGGAGGAAG
GATA43950	368	GATA43950F GATA43950R	CM020943.2	2,789,293	TGTGGAGAAGATGGGAAATAA CCTAAACATTTACCCACAAC
GATA52422	321		CM020957.2	12,277,462	

		GATA52422F			TGGGAATCTGCTCTAGAAAAA
		GATA52422R			GTGGACTTGCTGAGGACTTAA
GATA5890064	280		CM020957.2	52,807,970	
		GATA5890064F			ATTACCAGAACTTGGGTCTCC
		GATA5890064R			AGTGGAGTGTCATCTGAAAGC
GATA5890240	273		CM020947.2	14,589,449	
		GATA5890240F			GCACTTTGGACAGAGAACAGT
		GATA5890240R			TAAAAAGGTGACTCGATGAGC
GATA5892687	261		CM020958.2	68,848,906	
		GATA5892687F			ACTTCCTAGCCAAACTGGAAT
		GATA5892687R			ACAGATAATTGGGCCTTAGCT
GATA5943219	252		CM020944.2	122,602,677	
		GATA5943219F			CACCATGAGAGGACTTAAGGA
		GATA5943219R			ATCAAATTAAGTGTGGGCAA
GATA5946992	283		CM020944.2	110,695,007	
		GATA5946992F			ATCGTATCAGCCACACATTTT
		GATA5946992R			TTTAGAGCACCTCTTTTCAGA
GATA5947654	250		CM020941.2	72,374,148	
		GATA5947654F			CAAAGCATAAAACCAGCAACT
		GATA5947654R			TTATCAGGAATTGGCTTATGC
GATA5984139	280		CM020949.2	40,228,786	
		GATA5984139F			TAGGACACGATGCTTTCACTT
		GATA5984139R			AACAGGGCTGGACTTAGAGAT
GATA6013633	287		CM020944.2	127,372,707	
		GATA6013633F			ACCAGAGATGTGGAACCTGTA
		GATA6013633R			TAAGGTGTTGCCTACAAGAGG
GATA6057581	231		CM020956.2	41,459,017	
		GATA6057581F			CCTAACTATACTGGAGCCCTGA

		GATA6057581R			ATTTCCAGGTCTCTGACACAG
GATA6058119	308		CM020961.2	27,069,836	
		GATA6058119F			GACCAGCTTCTCTTCTCCTCT
		GATA6058119R			TAAGTCAACGATGAGAGGGAG
GATA6058394	335		CM020941.2	15,139,983	
		GATA6058394F			AGCAGTACCCCTCACTAGCTT
		GATA6058394R			AACACTTATCAAGCCCCCTAC
GATA6059012	352		CM020942.2	38,429,421	
		GATA6059012F			CAGGAATCTCAGGGGATTTA
		GATA6059012R			AAAATGAAATGTTGCCTGAAG
GATA6059993	312		CM020950.2	91,833,689	
		GATA6059993F			AATGATCAGCCTCTCATCCTA
		GATA6059993R			GCAAACGAGGACTTTGAAATA
GATA6063318	319		CM020959.2	6,562,237	
		GATA6063318F			CCCTAAGTCCTTCTTCAGGAC
		GATA6063318R			GCACTTAGGCATCTGGAAGT
GATA6063862	251		CM020962.2	4,337,731	
		GATA6063862F			GGTCAAGCACAGAAAGACTGT
		GATA6063862R			CTGCTTCATAAGATGGCAGAT
GATA6064765	344		CM020956.2	54,087,188	
		GATA6064765F			CTTTTCTGCTTCTGTAGTGGG
		GATA6064765R			GTTTTGGGGATGAACCTAGAC
GATA6065910	340		CM020943.2	12,718,001	
		GATA6065910F			CAGAACGCTCATCTGAAAAAT
		GATA6065910R			TATGTTAGGCACCCAATAAGC
GATA6237777	116		CM020942.2	169,313,978	
		GATA6237777F			CCCATTCCACTAGATGACAGA
		GATA6237777R			TGTACCCATATCTGCCCATA

GATA91083	183		CM020959.2	54,208,483	
		GATA91083F			CCAAATTGAGACAGCAACTCT
		GATA91083R			ATTGGAAAGGAGAAGGATCAC
GATA97408	179		CM020942.2	166,113,966	
		GATA97408F			GTTGTGTTCCATTGGTTCATT
		GATA97408R			CATGTCGGTCTTTAATCCATC

Notes: <sup>s</sup>F and R denotes forward and reverse orientation, respectively. <sup>&</sup>Genome coordinates (GenBank assembly accession: GCA\_009873245)

Table 2. Multiplex oligo-nucleotide panels for mysticete species.

<b>Panel A</b>					
Locus	Primer Reverse	Primer Forward	Dye	Range	
GATA6237777	R	F	HEX	108	152
GATA6064765*	R2	F2	HEX	182	230
GATA6063318	R	F	HEX	300	350
GATA91083	R	F	FAM	149	220
GATA38314	R	F	FAM	303	367
<b>Panel B</b>					
Locus	Primer Reverse	Primer Forward	Dye	Range	
GATA5947654*	R3	F3	HEX	130	160
GATA5984139	R	F	HEX	204	280
GATA52422RF	R	F	HEX	320	479
GATA25072*	R2	F2	FAM	76	148
GATA97408	R2	F2	FAM	164	215
SRY	R	F	FAM	332	332

\*New primers were re-designed in order to get different size fragments to fit the multiplex: GATA25072F2 (5'-CACCTGCTTTAAACTGTGTATAGT-3'), GATA25072R2 (5'-GATCTAGCAACTCTTTCTAGGC-3'), GATA6064765F2 (5'-GTACAAATGCACTTTCTCCCG-3') and GATA6064765R2 (5'-AGGCACTTATCAGTTCCAAGT-3'), SRYF, (5'-TGTGAACGGTGAGGATTA-3') and SRYR, (5'-GTGCATGGCTCGTAGTCT-3') GATA5947654R3 (5'-TCAGCCTCCATAATTGCATAAG-3') and GATA5947654F3 (5'-GTTATTAGATAGGGTTCTCTGCAG-3'), GATA97408F2 (5'-CTCCCACCACTGATTTGTAATA-3') and GATA97408R2 (5'-AGCCTAGTTTTATGGTACCTCT-3').

Table 3. Summary statistics for 10 microsatellite loci in 48 individuals in a selection of mysticete species.

Species	Locus	Na	H <sub>O</sub>	H <sub>E</sub>	Pr( <i>I</i> )	HW Signif
<i>B. borealis</i>	GATA25072	6	0.83	0.75	9.55E-02	ns
	GATA38314	7	0.63	0.73	1.04E-01	***
	GATA52422	4	0.65	0.64	1.57E-01	ns
	GATA5947654	7	0.83	0.77	7.75E-02	ns
	GATA5984139	8	0.79	0.72	8.91E-02	ns
	GATA6063318	9	0.88	0.83	4.04E-02	ns
	GATA6064765	10	0.94	0.86	2.71E-02	ns
	GATA6237777	2	0.13	0.12	7.76E-01	ns
	GATA91083	4	0.81	0.67	1.64E-01	*
	GATA97408	7	0.69	0.76	7.81E-02	ns
<i>E. australis</i>	GATA25072	3	0.42	0.41	4.06E-01	***
	GATA38314	20	0.63	0.83	3.74E-02	**
	GATA52422	11	0.25	0.76	6.81E-02	***
	GATA5947654	1	0	0	1.00E+00	Monomorphic
	GATA5984139	5	0.21	0.35	4.31E-01	***
	GATA6063318	8	0.5	0.79	6.82E-02	***
	GATA6064765	11	0.54	0.88	1.99E-02	ns
	GATA6237777	7	0.81	0.8	5.57E-02	ns
	GATA91083	3	0.46	0.43	3.68E-01	ns
	GATA97408	8	0.79	0.79	6.56E-02	ns
<i>B. acutorostrata</i>	GATA25072	4	0.58	0.59	2.04E-01	ns
	GATA38314	7	0.77	0.79	6.80E-02	ns
	GATA52422	6	0.65	0.66	1.57E-01	ns
	GATA5947654	4	0.23	0.28	5.18E-01	ns
	GATA5984139	8	0.48	0.53	2.43E-01	***
	GATA6063318	6	0.75	0.7	1.36E-01	ns
	GATA6064765	7	0.77	0.76	8.68E-02	ns
	GATA6237777	5	0.63	0.61	2.03E-01	ns



<i>M. novaeangliae</i>	GATA91083	6	0.63	0.61	2.18E-01	ns
	GATA97408	4	0.6	0.67	1.74E-01	ns
	GATA25072	9	0.85	0.84	3.65E-02	ns
	GATA38314	6	0.77	0.67	1.43E-01	ns
	GATA52422	6	0.75	0.72	1.19E-01	ns
	GATA5947654	5	0.73	0.67	1.38E-01	ns
	GATA5984139	11	0.81	0.82	4.27E-02	***
	GATA6063318	9	0.9	0.81	5.17E-02	ns
	GATA6064765	6	1	0.77	8.21E-02	**
	GATA6237777	8	0.9	0.85	3.29E-02	ns
<i>B. physalus</i>	GATA91083	7	0.77	0.77	7.30E-02	ns
	GATA97408	7	0.73	0.75	7.96E-02	ns
	GATA25072	12	0.88	0.85	2.94E-02	***
	GATA38314	10	0.75	0.75	8.93E-02	***
	GATA52422	15	0.81	0.76	6.12E-02	**
	GATA5947654	8	0.67	0.69	1.36E-01	***
	GATA5984139	13	0.85	0.86	2.76E-02	***
	GATA6063318	12	0.73	0.73	9.80E-02	***
	GATA6064765	10	0.88	0.85	3.20E-02	ns
	GATA6237777	4	0.31	0.47	3.62E-01	***
<i>B. mysticetus</i>	GATA91083	5	0.69	0.69	1.30E-01	ns
	GATA97408	8	0.85	0.8	5.72E-02	ns
	GATA25072	1	0	0	1.00E+00	Monomorphic
	GATA38314	12	0.79	0.81	4.18E-02	ns
	GATA52422	6	0.75	0.73	1.01E-01	ns
	GATA5947654	7	0.4	0.71	1.21E-01	***
	GATA5984139	3	0.44	0.5	3.49E-01	ns
	GATA6063318	7	0.67	0.61	1.90E-01	ns
	GATA6064765	12	0.79	0.78	5.51E-02	***
	GATA6237777	6	0.81	0.74	9.40E-02	ns

<i>B. musculus</i>	GATA91083	5	0.46	0.75	9.70E-02	***
	GATA97408	4	0.08	0.12	7.66E-01	***
	GATA25072	8	0.83	0.78	6.49E-02	ns
	GATA38314	15	0.77	0.88	1.55E-02	***
	GATA52422	3	0.29	0.5	2.82E-01	ns
	GATA5947654	3	0	0.1	8.10E-01	***
	GATA5984139	13	0.56	0.88	1.91E-02	***
	GATA6063318	14	0.65	0.86	2.45E-02	*
	GATA6064765	28	0.73	0.93	4.17E-03	***
	GATA6237777	1	0	0	1.00E+00	Monomorphic
	GATA91083	11	0.81	0.78	6.84E-02	ns
	GATA97408	9	0.6	0.75	8.55E-02	ns

---

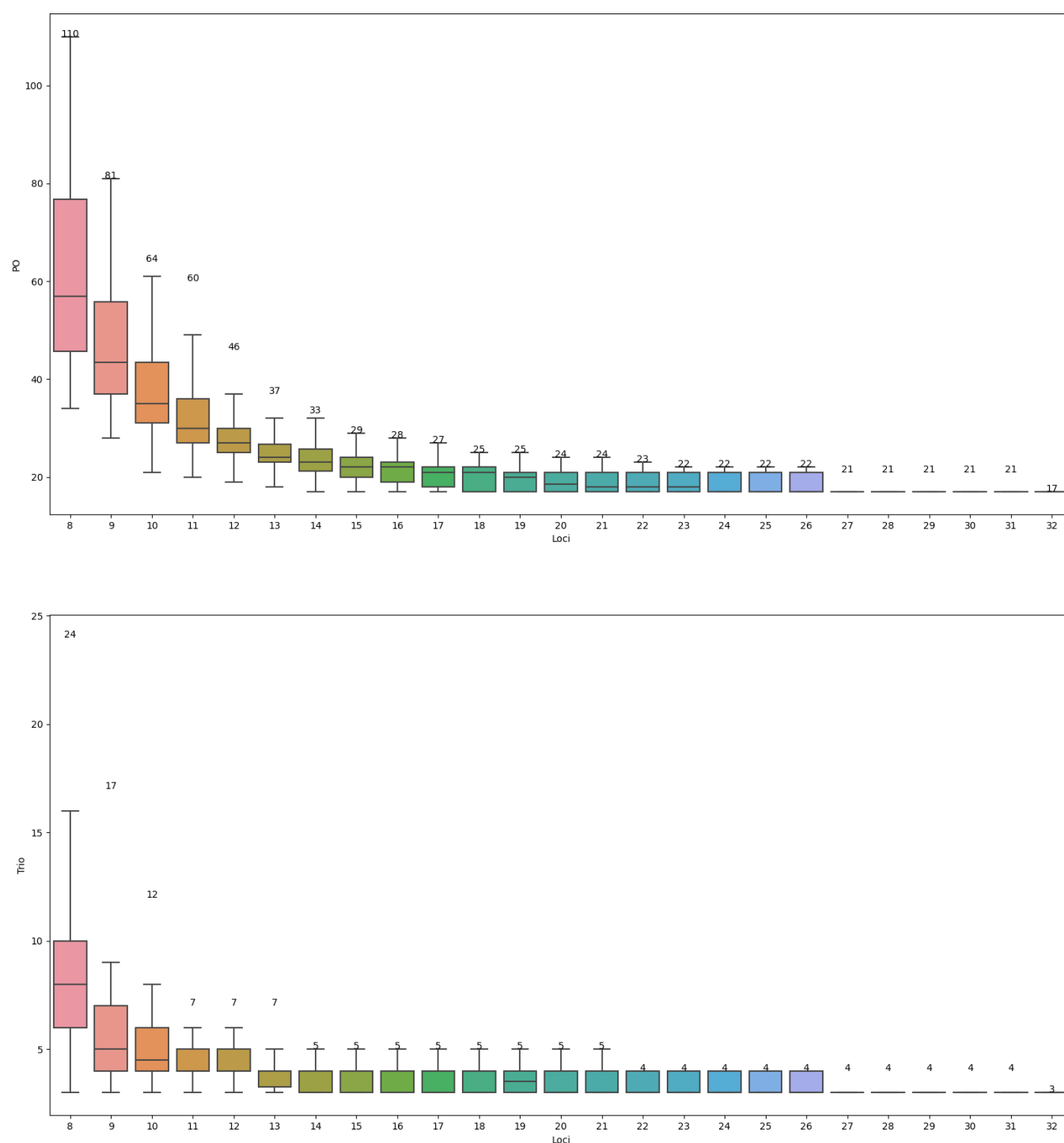
Notes. Na denotes the number of alleles at each locus and  $H_O$  and  $H_E$ , observed and expected heterozygosity, respectively. HW signif, denotes the level of significance for Hardy-Weinberg equilibrium. The probability of identity,  $P(I)$  (for a population where individuals randomly mate with correction for small samples of individuals), are estimated as described in Paetkau et Strobeck (1994) and Kendall et Stewart (1977). Finally, ns=not significant, \*  $P<0.05$ , \*\*  $P<0.01$ , \*\*\*  $P<0.001$ .

Table 4. Humpback whale combined exclusion probabilities of first parents estimated for ranked STR loci based on the informativeness of relationship criterion (IR).

Loci	Highest	Lowest
8	0.002638	0.15063484
9	0.00151085	0.10668238
10	0.00087211	0.07395704
11	0.00050162	0.05130777
12	0.00028957	0.03557103
13	0.00017144	0.02442526
14	0.0001043	0.01612181
15	0.00006355	0.01068883
16	0.00003802	0.00676013
17	0.00002404	0.00404405
18	0.00001594	0.00246402
19	0.00001052	0.00149905
20	0.00000722	0.00088752
21	0.00000501	0.00051233
22	0.00000347	0.00029468
23	0.00000241	0.0001701
24	0.00000171	0.00009742
25	0.00000128	0.00005367
26	0.00000094	0.00002792
27	0.0000007	0.000014
28	0.00000053	0.00000685
29	0.00000042	0.00000338
30	0.00000034	0.00000154
31	0.00000029	0.00000065
32	0.00000026	0.00000026

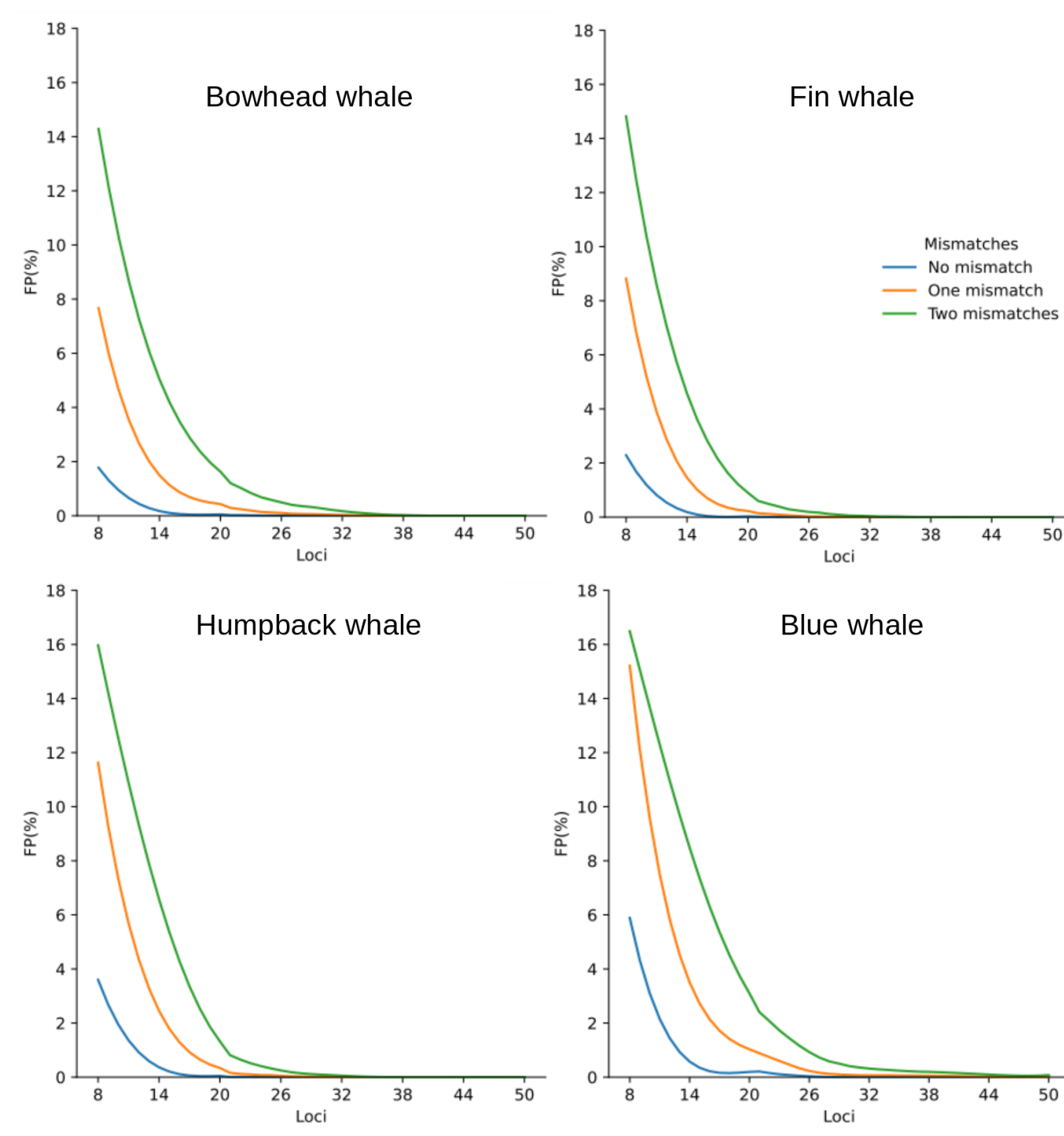
Notes. Starting with the eight highest/lowest ranked loci.

Figure 1. Effects of the number and informativeness (IR) of STR loci on parentage offspring and trio assignment.



Notes. Values on top of each box indicate the maximum number of PO assignments in any dataset for that specific set of loci.

Figure 2. Percentage of false positives of paternity assignments per loci within each species.



# Supplementary Material

Table S1. Allele number, size range (bp), observed and expected heterozygosity of 28 novel microsatellite loci in seven *Mysticetes* species.

Locus	Species	N	Na	Range*		Ho	He
				Lower	Higher		
<b>GATA19406</b>	<b>Mysticetes</b>						
	Bp	7	4	301	311	0.71	0.75
	Ba	8	4	291	307	0.50	0.56
	Ea	8	2	291	295	0.12	0.12
	Mn	8	6	291	319	0.75	0.76
	Bm	8	1	276		0.00	0.00
	Bb	8	1	276		0.00	0.00
<b>GATA25072</b>	Bmyst	8	4	287	299	0.75	0.70
	<b>Mysticetes</b>						
	Bp	8	8	375	414	1.00	0.86
	Ba	8	4	367	379	0.75	0.54
	Ea	6	2	356	360	0.16	0.38
	Mn	8	7	391	420	0.62	0.74
	Bm	8	6	387	411	0.87	0.77
<b>GATA29055</b>	Bb	8	4	387	399	0.87	0.73
	Bmyst	8	1	358		0.00	0.00
	<b>Mysticetes</b>						
	Bp	8	1	146		0.00	0.00
	Ba	8	1	150		0.00	0.00
	Ea	6	2	119	146	0.17	0.15
	Mn	7	6	146	204	0.71	0.70
<b>GATA36068</b>	Bm	8	2	134	151	0.13	0.12
	Bb	8	2	146	151	0.38	0.49
	Bmyst	8	3	143	151	0.13	0.54
	<b>Mysticetes</b>						
	Bp	8	5	379	399	0.75	0.73
	Ba	8	1	344		0.00	0.00
	Ea	7	2	355	363	0.14	0.13
<b>GATA3635</b>	Mn	8	5	360	380	0.75	0.75
	Bm	8	1	337		0.00	0.00
	Bb	8	3	352	372	0.38	0.40
	Bmyst	8	2	343	353	0.63	0.43
	<b>Mysticetes</b>						
	Bp	8	4	317	333	0.50	0.65
	Ba	8	5	317	337	0.50	0.73
<b>GATA38314</b>	Ea	8	6	301	329	0.50	0.76
	Mn	7	6	307	335	0.71	0.81
	Bm	8	1	297		0.00	0.00
	Bb	8	4	305	321	0.38	0.62
	Bmyst	8	7	300	322	0.88	0.79
	<b>Mysticetes</b>						
	Bp	8	7	350	378	0.63	0.74
<b>GATA43950</b>	Ba	8	6	329	349	1.00	0.80
	Ea	2	1	364		0.00	0.00
	Mn	6	5	328	352	0.67	0.72
	Bm	8	5	345	373	0.63	0.68
	Bb	7	3	353	361	0.57	0.61
	Bmyst	8	5	343	361	0.50	0.57
	<b>Mysticetes</b>						
<b>GATA43950</b>	Bp	7	3	425	433	0.71	0.62
	Ba	4	2	359	363	0.00	0.50
	Ea	2	2	366	416	0.00	0.50

	Mn	8	7	293	413	0.88	0.84
	Bm	8	6	385	413	1.00	0.80
	Bb	0	0			0.00	0.00
	Bmyst	0	0			0.00	1.00
	<b>Mysticetes</b>						
<b>GATA52422</b>	Bp	6	5	436	468	0.83	0.68
	Ba	7	4	375	391	0.71	0.66
	Ea	0	0			0.00	0.00
	Mn	5	4	333	345	0.60	0.70
	Bm	8	2	349	353	0.13	0.12
	Bb	7	4	317	329	0.57	0.60
	Bmyst	8	5	347	359	0.63	0.75
	<b>Mysticetes</b>						
<b>GATA5890064</b>	Bp	8	5	295	311	0.75	0.74
	Ba	8	4	283	295	0.75	0.60
	Ea	8	3	253	265	0.50	0.60
	Mn	8	6	283	303	0.75	0.77
	Bm	8	6	306	330	0.88	0.64
	Bb	8	2	279	283	0.38	0.43
	Bmyst	8	1	261		0.00	0.00
	<b>Mysticetes</b>						
<b>GATA5890240</b>	Bp	0	0			0.00	0.00
	Ba	8	2	288	292	0.25	0.22
	Ea	5	5	290	306	0.80	0.74
	Mn	8	7	288	320	0.88	0.78
	Bm	5	1	292		0.00	0.00
	Bb	8	5	318	335	0.63	0.72
	Bmyst	8	4	276	302	0.63	0.66
	<b>Mysticetes</b>						
<b>GATA5892687</b>	Bp	0	0			0.00	0.00
	Ba	8	5	255	285	0.88	0.67
	Ea	5	1	241		0.00	0.00
	Mn	7	4	272	288	0.57	0.61
	Bm	7	5	277	293	0.71	0.72
	Bb	7	1	269		0.00	0.00
	Bmyst	8	1	237		0.00	0.00
	<b>Mysticetes</b>						
<b>GATA5943219</b>	Bp	3	4	220	230	0.67	0.67
	Ba	6	2	277	281	0.67	0.44
	Ea	7	4	229	259	0.29	0.37
	Mn	7	5	260	276	0.71	0.72
	Bm	6	4	265	283	0.67	0.69
	Bb	7	4	264	276	0.57	0.54
	Bmyst	8	6	219	261	0.88	0.77
	<b>Mysticetes</b>						
<b>GATA5946992</b>	Bp	6	3	290	298	0.83	0.61
	Ba	8	6	261	281	0.75	0.78
	Ea	0	0			0.00	0.00
	Mn	8	6	279	303	0.50	0.77
	Bm	8	5	290	306	0.75	0.74
	Bb	8	6	268	292	1.00	0.79
	Bmyst	0	0			0.00	1.00
	<b>Mysticetes</b>						
<b>GATA5947654</b>	Bp	8	6	287	311	1.00	0.77
	Ba	8	3	256	264	0.25	0.23
	Ea	8	1	228		0.00	0.00
	Mn	8	4	259	271	0.63	0.67
	Bm	8	1	222		0.00	0.00
	Bb	8	5	269	285	0.88	0.73

	Bmyst	8	4	227	261	0.38	0.51
	<b>Mysticetes</b>						
<b>GATA5984139</b>	Bp	5	7	260	288	1.00	0.84
	Ba	8	4	262	284	0.50	0.49
	Ea	7	2	226	230	0.29	0.25
	Mn	6	7	257	285	1.00	0.83
	Bm	7	6	242	284	0.86	0.77
	Bb	8	6	261	281	0.88	0.80
	Bmyst	8	2	221	225	0.25	0.47
	<b>Mysticetes</b>						
<b>GATA6013633</b>	Bp	8	6	287	315	1.00	0.80
	Ba	8	2	295	299	0.25	0.38
	Ea	6	3	299	307	0.33	0.57
	Mn	8	5	299	315	0.88	0.72
	Bm	8	2	299	311	0.50	0.47
	Bb	8	4	340	352	0.13	0.68
	Bmyst	8	3	272	298	0.38	0.32
	<b>Mysticetes</b>						
<b>GATA6057581</b>	Bp	0	0			0.00	0.00
	Ba	8	6	186	214	0.88	0.75
	Ea	8	3	199	207	0.38	0.46
	Mn	8	7	237	261	1.00	0.83
	Bm	8	4	189	207	0.75	0.65
	Bb	8	1	199		0.00	0.00
	Bmyst	8	3	206	214	0.25	0.26
	<b>Mysticetes</b>						
<b>GATA6058119</b>	Bp	0	0			0.00	0.00
	Ba	8	3	317	325	0.75	0.54
	Ea	8	2	289	305	0.25	0.22
	Mn	8	5	321	341	0.63	0.70
	Bm	6	2	206	242	0.50	0.49
	Bb	8	6	305	354	0.88	0.79
	Bmyst	8	3	289	305	0.63	0.48
	<b>Mysticetes</b>						
<b>GATA6058394</b>	Bp	7	5	343	359	0.86	0.76
	Ba	8	4	366	378	0.63	0.60
	Ea	6	4	362	374	0.83	0.67
	Mn	7	5	355	399	1.00	0.75
	Bm	7	6	344	367	0.57	0.74
	Bb	8	5	340	356	0.88	0.66
	Bmyst	8	5	358	374	0.88	0.73
	<b>Mysticetes</b>						
<b>GATA6059012</b>	Bp	6	4	368	418	0.67	0.65
	Ba	8	4	365	377	0.50	0.49
	Ea	6	1	342		0.00	0.00
	Mn	8	6	357	377	0.88	0.81
	Bm	8	8	372	400	0.75	0.79
	Bb	8	1	352		0.00	0.00
	Bmyst	0	0			0.00	1.00
	<b>Mysticetes</b>						
<b>GATA6059993</b>	Bp	5	1	287		0.00	0.00
	Ba	4	3	307	315	0.75	0.53
	Ea	0	0			0.00	0.00
	Mn	2	2	335	339	0.50	0.38
	Bm	0	0			0.00	0.00
	Bb	3	1	288		0.00	0.00
	Bmyst	8	4	296	336	0.50	0.74
	<b>Mysticetes</b>						
<b>GATA6063318</b>	Bp	7	5	321	338	0.43	0.65



	Ba	6	3	330	338	0.67	0.57
	Ea	4	5	322	336	1.00	0.75
	Mn	6	6	339	355	0.83	0.74
	Bm	7	4	327	343	0.43	0.46
	Bb	5	4	322	338	0.60	0.64
	Bmyst	8	5	308	330	0.63	0.66
	<b>Mysticetes</b>						
<b>GATA6063862</b>	Bp	8	7	245	269	0.88	0.80
	Ba	8	4	272	284	1.00	0.71
	Ea	8	1	245		0.00	0.00
	Mn	8	8	245	285	0.75	0.83
	Bm	8	6	264	296	1.00	0.77
	Bb	8	3	261	269	0.75	0.65
	Bmyst	8	4	249	265	0.38	0.63
	<b>Mysticetes</b>						
<b>GATA6064765</b>	Bp	5	5	353	381	1.00	0.74
	Ba	7	5	368	407	0.86	0.74
	Ea	7	6	420	444	0.86	0.79
	Mn	8	3	355	367	0.63	0.65
	Bm	7	8	341	385	0.71	0.87
	Bb	8	5	360	391	0.88	0.66
	Bmyst	8	6	311	347	0.63	0.74
	<b>Mysticetes</b>						
<b>GATA6065910</b>	Bp	4	4	377	389	1.00	0.72
	Ba	8	5	347	363	0.63	0.56
	Ea	0	0			0.00	0.00
	Mn	8	6	350	370	0.63	0.58
	Bm	4	5	354	406	0.50	0.78
	Bb	0	0			0.00	0.00
	Bmyst	8	5	346	366	0.38	0.76
	<b>Mysticetes</b>						
<b>GATA6237777</b>	Bp	8	5	293	309	0.75	0.63
	Ba	8	4	273	285	0.50	0.65
	Ea	7	4	270	286	0.57	0.61
	Mn	8	5	270	294	0.75	0.73
	Bm	7	5	278	294	0.86	0.77
	Bb	8	5	278	294	0.75	0.78
	Bmyst	8	1	233	233	0.00	0.00
	<b>Mysticetes</b>						
<b>GATA91083</b>	Bp	8	3	199	207	0.75	0.57
	Ba	8	4	179	195	0.63	0.63
	Ea	8	3	199	207	0.38	0.46
	Mn	6	5	179	207	1.00	0.68
	Bm	7	6	179	235	1.00	0.76
	Bb	8	2	171	195	0.50	0.47
	Bmyst	8	2	184	196	0.38	0.49
	<b>Mysticetes</b>						
<b>GATA97408</b>	Bp	8	5	293	309	0.75	0.63
	Ba	8	4	273	285	0.50	0.65
	Ea	7	4	270	286	0.57	0.61
	Mn	8	5	270	294	0.75	0.73
	Bm	7	5	278	294	0.86	0.77
	Bb	8	5	278	294	0.75	0.78
	Bmyst	8	1	233	233	0.00	0.00

Notes: \* Indicates that the length of the M13 (5'- TGT AAA ACG ACG GCC AGT-3') or the T7 ( 5'-TAA TAC GAC TCA CTA TAG GGC-3') are included in the length of the microsatellite fragment. Species' abbreviations: Bp (*Balaenoptera physalus*), Ba (*Balaenoptera acutorostrata*), Ea (*Eubalaena australis*), Mn (*Megaptera novaeangliae*), Bm (*Balaenoptera musculus*), Bb (*Balaenoptera borealis*), Bmyst (*Balaena mysticetus*).

Table S2. STR loci used for parentage assignment.

STR loci	Species	$H_E$				$P_{\text{NON-EXCL}}$			
	isolated from	Bp	Mn	Bm	Bmyst	Bp	Mn	Bm	Bmyst
AC045 <sup>&amp;</sup>	Mn	n/a	0.74	n/a	n/a	n/a	0.52	n/a	n/a
AC082 <sup>&amp;</sup>	Mn	n/a	n/a	0.60	0.68	n/a	n/a	0.35	0.40
AC087 <sup>&amp;</sup>	Mn	0.69	0.80	0.57	0.74	0.74	0.60	0.35	0.52
AC137 <sup>&amp;</sup>	Mn	n/a	n/a	0.67	n/a	n/a	n/a	0.45	n/a
Bmy16 <sup>†</sup>	Bmyst	n/a	n/a	n/a	0.72	n/a	n/a	n/a	0.50
Bmy19 <sup>†</sup>	Bmyst	n/a	n/a	n/a	0.84	n/a	n/a	n/a	0.69
Bmy26 <sup>†</sup>	Bmyst	n/a	n/a	n/a	0.91	n/a	n/a	n/a	0.80
Bmy33 <sup>†</sup>	Bmyst	n/a	n/a	n/a	0.80	n/a	n/a	n/a	0.61
Bmy41 <sup>†</sup>	Bmyst	n/a	n/a	n/a	0.93	n/a	n/a	n/a	0.84
Bmy42 <sup>†</sup>	Bmyst	n/a	n/a	n/a	0.80	n/a	n/a	n/a	0.63
Bmy58 <sup>†</sup>	Bmyst	n/a	n/a	n/a	0.92	n/a	n/a	n/a	0.83
CA128 <sup>&amp;</sup>	Bp	n/a	n/a	0.67	n/a	n/a	n/a	0.38	n/a
CA141 <sup>&amp;</sup>	Bp	0.69	n/a	0.53	0.61	0.73	n/a	0.32	0.33
CA232 <sup>&amp;</sup>	Mn	0.57	n/a	n/a	0.76	0.83	n/a	n/a	0.52
CA234 <sup>&amp;</sup>	Mn	0.81	0.73	0.71	n/a	0.54	0.48	0.48	n/a
EV001 <sup>°</sup>	Pm	0.84	0.63	n/a	n/a	0.49	0.38	n/a	n/a
EV037 <sup>°</sup>	Mn	0.85	0.87	0.53	n/a	0.46	0.74	0.33	n/a
EV094 <sup>°</sup>	Mn	0.91	0.68	n/a	n/a	0.31	0.43	n/a	n/a
EV096 <sup>°</sup>	Mn	n/a	0.80	n/a	n/a	n/a	0.61	n/a	n/a
GATA028 <sup>¥</sup>	Mn	0.87	0.46	0.82	0.84	0.41	0.28	0.63	0.68
GATA053 <sup>¥</sup>	Mn	n/a	0.85	n/a	n/a	n/a	0.68	n/a	n/a
GATA098 <sup>¥</sup>	Mn	0.74	0.65	0.67	0.63	0.66	0.43	0.44	0.41
GATA19406 <sup>α</sup>	Mn	0.83	0.72	n/a	0.64	0.51	0.50	n/a	0.35

GATA25072 <sup>α</sup>	Mn	0.87	0.84	n/a	n/a	0.41	0.68	n/a	n/a
GATA38314 <sup>α</sup>	Mn	0.82	0.71	n/a	0.85	0.52	0.46	n/a	0.69
GATA417 <sup>¥</sup>	Mn	0.86	0.86	0.87	n/a	0.42	0.71	0.73	n/a
GATA43950 <sup>α</sup>	Mn	0.60	0.88	0.81	n/a	0.79	0.75	0.63	n/a
GATA52422 <sup>α</sup>	Mn	0.85	0.70	n/a	0.69	0.44	0.44	n/a	0.44
GATA5947654 <sup>α</sup>	Mn	0.78	0.67	n/a	n/a	0.59	0.41	n/a	n/a
GATA5984139 <sup>α</sup>	Mn	0.83	0.84	n/a	0.49	0.50	0.67	n/a	0.19
GATA6059012 <sup>α</sup>	Mn	0.75	n/a	n/a	n/a	0.65	n/a	n/a	n/a
GATA6063318 <sup>α</sup>	Mn	0.77	0.81	n/a	0.66	0.61	0.63	n/a	0.42
GATA6063862 <sup>α</sup>	Mn	n/a	0.72	n/a	n/a	n/a	0.49	n/a	n/a
GATA6064765 <sup>α</sup>	Mn	0.83	n/a	n/a	0.77	0.50	n/a	n/a	0.58
GATA6237777 <sup>α</sup>	Mn	0.43	0.83	n/a	n/a	0.91	0.65	n/a	n/a
GATA91083 <sup>α</sup>	Mn	0.61	0.79	0.78	n/a	0.80	0.59	0.58	n/a
GATA97408 <sup>α</sup>	Mn	0.77	0.76	0.79	n/a	0.62	0.55	0.58	n/a
GT011 <sup>□</sup>	Mn	0.83	0.80	n/a	0.60	0.50	0.60	n/a	0.33
GT015 <sup>α</sup>	Mn	n/a	0.78	n/a	n/a	n/a	0.57	n/a	n/a
GT023 <sup>#</sup>	Mn	0.78	0.80	0.83	0.74	0.60	0.60	0.64	0.51
GT101 <sup>#</sup>	Mn	n/a	0.71	0.69	n/a	n/a	0.49	0.46	n/a
GT129 <sup>&amp;</sup>	Bp	n/a	n/a	n/a	0.62	n/a	n/a	n/a	0.34
GT195 <sup>#</sup>	Mn	n/a	0.62	n/a	n/a	n/a	0.32	n/a	n/a
GT211 <sup>#</sup>	Mn	0.80	0.79	n/a	n/a	0.56	0.58	n/a	n/a
GT271 <sup>#</sup>	Mn	0.64	0.51	0.53	n/a	0.75	0.31	0.28	n/a
GT310 <sup>#</sup>	Mn	0.71	n/a	0.51	n/a	0.67	n/a	0.23	n/a
GT541 <sup>&amp;</sup>	Mn	n/a	n/a	0.73	n/a	n/a	n/a	0.50	n/a
GT575 <sup>#</sup>	Mn	0.69	0.73	0.63	n/a	0.72	0.52	0.35	n/a
TAA031 <sup>¥</sup>	Mn	n/a	0.78	n/a	n/a	n/a	0.57	n/a	n/a
TAA023 <sup>¥</sup>	Mn	0.76	n/a	n/a	n/a	0.62	n/a	n/a	n/a

TR3G2*	Eg	n/a	n/a	0.76	n/a	n/a	n/a	0.55	n/a
--------	----	-----	-----	------	-----	-----	-----	------	-----

---

Notes: Mn denotes *Megaptera novaeangliae*, Bmyst denotes *Balaena mysticetus* and Bm, *Balaenoptera musculus*, Bp, *Balaenoptera physalus*, Eg, *Eubalaena glacialis*, Pm, *Physeter macrocephalus*. H<sub>E</sub> denotes expected heterozygosity and P<sub>NON-EXCL</sub>, the non-exclusion probability of the first parent. References; <sup>□</sup>(Bérubé et al. 1998); <sup>#</sup>(Bérubé et al. 2000); <sup>&</sup>(Bérubé et al. 2005); <sup>\*</sup>(Frasier et al. 2006) ; <sup>¶</sup>(Huebinger et al. 2008); <sup>¥</sup>(Palsbøll et al. 1997); <sup>α</sup>This study; <sup>°</sup>(Valsecchi and Amos 1996)

Table S3 Informativeness estimates and power to detect relationships in the humpback whale dataset

Locus	R_Info	RMSD (LL)	RMSD (QG)	RMSD (R)	RMSD (LR)	RMSD (W)	Threshold	Power	Simu Thrshld	SimuPow	r_Infor
GATA43950	0.121	5.220	5.220	1.420	5.100	5.240	1.820	0.093	1.820	0.093	0.163
EV037	0.115	5.070	5.050	1.420	4.970	5.060	1.760	0.090	1.760	0.090	0.158
GATA417	0.109	4.820	4.820	1.630	4.780	4.850	1.710	0.088	1.710	0.089	0.151
GATA25072	0.101	4.590	4.560	1.180	4.520	4.570	1.720	0.092	1.720	0.092	0.144
GATA053	0.100	4.650	4.590	1.270	4.550	4.530	1.740	0.088	1.740	0.088	0.143
GATA598413	0.099	4.530	4.490	1.670	4.440	4.500	1.760	0.091	1.760	0.090	0.142
GATA623777	0.094	4.440	4.370	1.490	4.350	4.320	1.660	0.088	1.660	0.088	0.137
GATA606331	0.090	4.160	4.120	0.919	4.100	4.200	1.720	0.090	1.720	0.091	0.133
EV096	0.085	4.020	3.980	1.430	3.980	4.040	1.680	0.089	1.680	0.090	0.129
AC087	0.085	3.960	3.930	1.490	3.930	4.030	1.770	0.091	1.770	0.091	0.129
GT023	0.085	4.020	3.990	2.030	4.020	4.030	1.720	0.087	1.720	0.087	0.128
GT011	0.084	4.020	3.940	0.980	3.940	3.980	1.720	0.089	1.730	0.089	0.127
GATA91083	0.081	3.920	3.860	0.881	3.880	3.890	1.670	0.086	1.670	0.086	0.124
GT015	0.080	3.710	3.650	0.941	3.650	3.810	1.820	0.093	1.820	0.092	0.124
TAA031	0.079	3.730	3.660	0.715	3.660	3.790	1.700	0.090	1.700	0.089	0.123
GT211	0.079	3.910	3.810	0.718	3.800	3.810	1.620	0.085	1.620	0.086	0.122
GATA97408	0.075	3.560	3.550	1.670	3.660	3.670	1.650	0.087	1.650	0.086	0.118
GT575	0.072	3.290	3.390	1.230	3.630	3.510	1.730	0.091	1.730	0.092	0.115
AC045	0.070	3.360	3.350	1.640	3.490	3.470	1.710	0.087	1.710	0.087	0.113
GATA19406	0.065	3.200	3.220	0.677	3.430	3.310	1.600	0.083	1.600	0.083	0.107
GT101	0.064	3.120	3.100	0.931	3.280	3.240	1.750	0.092	1.750	0.091	0.107
GATA606386	0.064	3.140	3.100	0.579	3.270	3.230	1.640	0.090	1.640	0.090	0.106
CA234	0.062	3.200	3.050	0.736	3.090	3.170	1.570	0.086	1.570	0.086	0.104
GATA38314	0.059	3.060	2.890	0.885	2.940	3.060	1.660	0.088	1.6-60	0.088	0.101
GATA098	0.056	2.660	2.760	1.050	3.110	2.880	1.670	0.089	1.670	0.090	0.097
GATA52422	0.054	2.940	2.700	0.448	2.720	2.880	1.590	0.083	1.590	0.083	0.096
EV094	0.053	2.820	2.650	1.780	2.780	2.820	1.550	0.079	1.550	0.080	0.094
GATA594765	0.050	2.700	2.560	1.670	2.770	2.730	1.500	0.079	1.500	0.079	0.091
EV001	0.045	2.440	2.320	1.420	2.630	2.510	1.550	0.081	1.550	0.080	0.085
GT195	0.038	2.210	1.710	1.360	1.760	2.190	1.490	0.076	1.490	0.076	0.075
GT271	0.037	1.870	1.970	0.599	2.370	2.100	1.780	0.091	1.780	0.091	0.072
GATA028	0.033	1.650	1.790	0.460	2.130	1.890	1.750	0.091	1.750	0.090	0.065

Notes: R\_Info denotes informativeness for relationship, RMSD: the mean squared deviations of relatedness estimates prior to combination of LL (Lynch 1988) and (Li, Weeks, and Chakravarti 1993); QG (Queller and Goodnight 1989); R (Ritland 1996); LR (Lynch and Ritland 1999) and W (Wang 2002). The Power refers to the power for relationship inference and the threshold is based on the distribution of the likelihood ratio values for pairs of genotypes through simulations as described in Queller and Goodnight (1989) for the exception that it allows for errors in the dataset. Finally, r\_Info denotes the informativeness for relatedness.