Matching of humpback, *Megaptera novaeangliae,* and fin whale*, Balaenoptera physalus*, biopsies collected by HDR.

Authors: Martine Bérubé, and Per J. Palsbøll

Marine Evolution and Conservation Group, Groningen Institute of Evolutionary Life Sciences, University of Groningen, Groningen, The Netherlands

Contact information: p.j.palsboll@rug.nl (phone +31 6 57779495), m.berube@rug.nl

Introduction

This report outlines the molecular genotyping, sexing and DNA sequencing of skin biopsies collected from free-ranging humpback (*Megaptera novaeangliae*) and fin (*Balaenoptera physalus*) whales by HDR. The purpose of the contracted work was to ascertain if any whales sampled by HDR had been sampled elsewhere by means of so-called "genetic tagging", which consists of identifying samples with identical genetic "fingerprints" (Palsbøll, Allen, et al., 1997). Two aspects are key to reliable genetic tagging; (a) employing sufficient genetic markers that each unique genetic fingerprint is distinct for each individual, and (b) low laboratory error rates, which can result in otherwise identical genetic fingerprints becoming non-identical due to random errors. Aspect (a) can be evaluated statistically by estimating and employing a sufficiently low probability of identity (i.e., the probability of two different individuals having identical genetic fingerprints). Aspect (b) is assessed by (i) employing more genetic markers than the bare minimum, and (ii) by conducting regular, internal assessments of the genotype and laboratory error rates. In our laboratory past assessments have yielded an error rate of one incorrect genotype for every 1,000 genotypes.

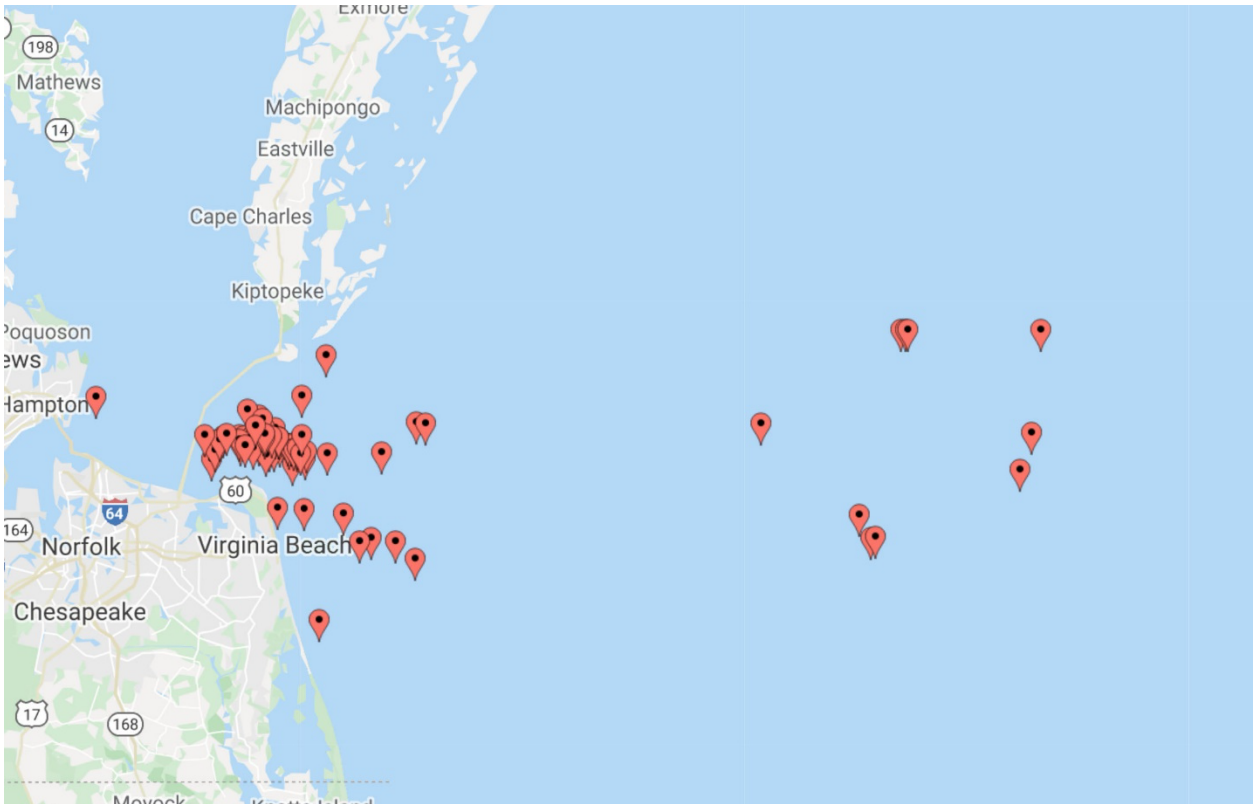
Materials and methods

HDR biopsy samples

With the exception of a single sample collected with a satellite dart, the remainder 71 samples were collected as skin biopsies from free-ranging whales, using either a compound crossbow or a

Paxarms™ biopsy gun. In total eight samples were collected from fin whales and 63 samples from humpback whales. Most humpback whale samples were collected along the coast line, except for two samples; VA180001 (HDR sample no. 2018Feb09_DTE_Mn_001) and VA190001 (HDR sample no. 2019Jan04_DTE_Mn_001). These two latter humpback whale samples were collected in the same offshore area where the fin whale samples were obtained (Figure1).

Figure 1. Location of biopsy samples.



Laboratory methods

Total-cell DNA was extracted from the supplied tissue samples using Puregene Gentra™ DNA extraction columns (Qiagen Inc.) according to the manufacturer's instructions.

The sex of each sample was determined by the presence of a Y-chromosome specific fragment (a small region of the SRY gene on the Y chromosome), which was co-amplified during the microsatellite multiplex genotyping.

The first 400 nucleotides of the mitochondrial control region (the most variable part) were PCR-amplified (polymerase chain reaction, Mullis & Faloona, 1987), using the primers MT4F

(Arnason, Gullberg, & Widegren, 1993) and BP16071R (Drouot et al., 2004). The initial PCR amplifications were performed in a 10 µL volume comprising 0.2 µM of each dNTP, 67mM Tris-HCl (pH 8.8), 2mM $MgCl_2$, 17mM NH_3SO_4, 10mM β-mercaptoethanol, 0.1µM of each primer, 0.4 units of *Taq* DNA polymerase and approximately 10 - 20 ng of DNA extraction. The thermo-cycling conditions were: 2 min at 94° C, followed by 25 cycles each consisting of 15 sec. at 94°C, 30 sec. at 54°C and 120 sec. at 72°C. After amplification, unincorporated nucleotides and excess primes were enzymatically removed using *shrimp alkaline phosphatase* and *exonuclease* I as described by Werle *et al.* (1994). The cleaned PCR amplification products were sequenced using fluorescently labelled ddNTPs according to the manufacturer's instructions (Big Dye™ v3.1 Terminator Ready Reaction Mix, Life Technologies Inc.), using either the primers MT4F or BP16071R. Excess dideoxy-terminator nucleotides were removed by ethanol/EDTA precipitation and re-suspended in 10 µL deionized formamide (Calbiochem Inc.). The order of sequencing products was resolved by capillary electrophoresis using an ABI Prism™ 3730 sequencer (Applied Biosystems Inc.).

Humpback whale samples were genotyped at 20 microsatellite loci: AC087 (Bérubé et al., 2005), EV001, EV037, EV094, EV096 (Valsecchi & Amos, 1996), 1996, GATA028, GATA098, GATA053, GATA417, TAA031 (Palsbøll, Bérubé, Larsen, & Jørgensen, 1997), GT011 (Bérubé et al., 1998), GT015, GT023, GT101, GT195, GT211, GT271, GT575 (Bérubé, Jørgensen, McEwing, & Palsbøll, 2000), GATA43950, GATA97408 and the Y-chromosome specific marker SRY (Bérubé, unpublished, and Palsbøll, Vader, & Bakke, 1992).

Fin whale samples were genotyped at 21 microsatellite loci: EV037 and EV094 (Valsecchi & Amos, 1996), GATA028, GATA098, GATA417, and TAA023 (Palsbøll, Bérubé, et al., 1997), GT011 (Bérubé et al., 1998), GT023, GT211, GT271, GT310, and GT575 (Bérubé et al., 2000), AC087 and CA234 (Bérubé et al., 2005), GATA25072, GATA43950, GATA5947654, GATA6063318, GATA91083 (Bérubé, unpublished), and EV001 (Valsecchi & Amos, 1996).

Samples were genotyped in multiplex PCR reactions (between six to eight microsatellite loci per amplification), using the MM2X™ Multiplex kit Plus (Qiagen Inc.) in 5µL reaction volumes. The thermocycling conditions were: 2 min. at 94°C, followed by 35 cycles each of 30 sec. at 94°C, 90 sec. at 57°C and 30 sec. at 72°C followed by a final cycle of 10 min. at 68°C. The PCR

products were separated by capillary electrophoresis using an ABI Prism™ 3730 (Applied Biosystems Inc.). The size of the amplification products was estimated against a Genescan™ ROX-500 size standard (Applied Biosystems Inc.) in the software GENEMAPPER™ (version 4.0; Applied Biosystems Inc.).

Data analyses

The goodness of fit of the observed sex ratio with parity was assessed using a log-likelihood ratio test.

The multi-locus genotypes obtained from all HDR samples were matched against the collection of individual multi-locus genotypes from North Atlantic humpback and fin whales curated by our group, in the following manner.

The minimum number of identical microsatellite locus genotypes necessary to rigorously identify samples from the same individual was determined from the probability of identity (denoted *I*, Paetkau & Strobeck, 1994) for unrelated individuals, estimated from the entire sample. The parameter *I* denotes the probability that two different individuals have an identical multi-locus genotype, and is simply the product of each locus' *I*. While full-siblings and parent-offspring pairs have the highest probabilities of genotype identity, they constitute a very small fraction of the total number of possible pairs (see Rew, Robbins, Mattila, Palsbøll, & Bérubé, 2011). Therefore, we employed *I* for unrelated individuals as guide for determining the minimum required genotypes to discern among individuals by adding loci (starting with the locus with the highest estimate of *I*) until the total number of expected chance matches (assuming all samples were unrelated) was below 0.001.

All samples genotyped at less than the minimum loci required to discern among different individuals (above) were removed from the data set prior to matching.

Results

The sex ratio among the seven fin whale samples were heavily male biased (6 males:1 female). Such male-bias among fin whales on foraging grounds was reported earlier by Berube and co-workers and may represent pre-mating behavior (Bérubé, Berchok, & Sears, 2001). We detected

32 males and 31 females among the 63 individual humpback whales (Table 1), a ratio that did not differ significantly from parity (χ^2_{1df} = 0.016, P = 0.90).

The North Atlantic genetic humpback whale catalogue comprised a total of 9,265 genotyped at 20+ microsatellite loci. Among the 9,265 samples were 4,121, and 5,049 females and males, respectively. Sex was undetermined in 95 samples. The total number of expected single-locus genotypes (i.e., 9,265 times 20 loci) was 185,300 among which were 4,680 missing single-locus genotypes yielding an overall genotype rate at 0.975. Locus GATA43950 was only typed in 72 % of all samples. The reason for the lower genotyping rate was the identification of a so-called null-allele (and allele that is undetected), which implied that we needed to redesign the PCR amplification oligos and retype all homozygote individuals genotyped with the original PCR amplification oligos. This work update is yet to be completed and consequently, locus GATA43950 was removed from all samples resulting in a genotype rate at 0.988 for 19 loci.

The expected number of pairs of different individuals matching at all 19 loci among a total of 579,726 possible pairs (63 HDR samples compared to 9,202 non-HDR samples) were estimated at 4.2 10^{-3}, 5.52 10^{-7} and 1.82 10^{-15} for full-siblings, parent-offspring pairs and unrelated individuals, respectively (Table 1). Since a chance match at 11 loci between unrelated individuals (the vast majority of comparisons) was estimated at less than 0.001 pairs (of 579,726); a match at 11 loci was set as the threshold for identifying unique individuals (Table 1). Accordingly, all specimens with identical microsatellite genotypes at minimum 11 loci were inferred as duplicate specimens collected from the same individual.

Table 1. Probability of identical multi-locus genotypes between different individuals

# Loci	Locus	*Full siblings*			*Parent-offspring*			*Unrelated*		
		I_{locus}	$I_{cumulative}$	$N_{expected\ matches}$	I_{locus}	$I_{cumulative}$	$N_{expected\ matches}$	I_{locus}	$I_{cumulative}$	$N_{expected\ matches}$
1	GATA028	0.516	5.16E-01	2.99E+05	0.531	5.32E-01	3.08E+05	0.304	3.05E-01	1.77E+05
2	EV001	0.462	2.38E-01	1.38E+05	0.424	2.26E-01	1.31E+05	0.23	7.01E-02	4.06E+04
3	GT271	0.462	1.10E-01	6.39E+04	0.425	9.59E-02	5.56E+04	0.209	1.46E-02	8.49E+03
4	GT195	0.427	4.70E-02	2.73E+04	0.353	3.39E-02	1.96E+04	0.197	2.88E-03	1.67E+03
5	EV094	0.395	1.86E-02	1.08E+04	0.291	9.84E-03	5.70E+03	0.134	3.86E-04	2.24E+02
6	GATA098	0.411	7.65E-03	4.43E+03	0.322	3.17E-03	1.84E+03	0.131	5.04E-05	2.92E+01
7	GT575	0.398	3.04E-03	1.76E+03	0.295	9.37E-04	5.43E+02	0.108	5.47E-06	3.17E+00
8	GT101	0.372	1.13E-03	6.55E+02	0.243	2.28E-04	1.32E+02	0.094	5.16E-07	2.99E-01
9	GATA97408	0.362	4.09E-04	2.37E+02	0.223	5.09E-05	2.95E+01	0.077	3.96E-08	2.30E-02
10	**GT015**	**0.35**	**1.43E-04**	**8.29E+01**	**0.2**	**1.02E-05**	**5.91E+00**	**0.064**	**2.53E-09**	**1.46E-03**
11	GT211	0.342	4.90E-05	2.84E+01	0.185	1.88E-06	1.09E+00	0.06	1.52E-10	8.81E-05
12	GT023	0.341	1.67E-05	9.68E+00	0.182	3.43E-07	1.99E-01	0.057	8.70E-12	5.04E-06
13	TAA031	0.341	5.70E-06	3.30E+00	0.182	6.26E-08	3.63E-02	0.056	4.90E-13	2.84E-07
14	EV096	0.341	1.94E-06	1.13E+00	0.181	1.13E-08	6.57E-03	0.056	2.74E-14	1.59E-08
15	GT011	0.337	6.54E-07	3.79E-01	0.174	1.97E-09	1.14E-03	0.053	1.46E-15	8.45E-10
16	GATA053	0.334	2.19E-07	1.27E-01	0.169	3.33E-10	1.93E-04	0.05	7.24E-17	4.20E-11
17	AC087	0.334	7.30E-08	4.23E-02	0.167	5.58E-11	3.23E-05	0.047	3.41E-18	1.98E-12
18	GATA417	0.321	2.34E-08	1.36E-02	0.142	7.90E-12	4.58E-06	0.035	1.19E-19	6.92E-14
19	EV037	0.31	7.27E-09	4.21E-03	0.12	9.52E-13	5.52E-07	0.026	3.14E-21	1.82E-15

Removing all samples typed at fewer than 11 loci (a total of 51 samples) resulted in a final data set at 9,214 samples (incl. the 63 HDR samples) typed at 11 - 19 loci and an overall genotyping rate at 0.992 (i.e., the vast majority of samples were genotyped at all 19loci). The samples sizes per general area and range of sampling years are summarized in Table 2. It is important to realize that apart from a few areas then the overall sampling effort has been highly heterogenous across space and time.

Table 2. Sample sizes per general area and sampling years

General area	sampling years	samples
eastern US sea border	1990-2019	2,546
eastern Canadian sea border	1990-2019	1,157
Central Atlantic	1991-2017	419
eastern North Atlantic sea border	1988-2019	473
western Caribbean	1989-2005	4,538
eastern Caribbean and Cape Verde	1995-2015	78
Miscellaneous		3

No duplicate samples were identified among the HDR humpback whale samples. A total of 18 HDR samples matched to samples collected elsewhere along the US eastern sea border (Table 3). All samples matched 100% on all loci genotyped in both samples in each pair (i.e., no mismatching genotypes were detected). In addition, as an additional affirmation, the sex and mitochondrial control region DNA sequences agreed for all matching pairs. The sample identification numbers in Table 3 are reference numbers for the matching samples. Additional information regarding each matching individuals can be obtained by contacting the institution that provided the sample. These institutes have also been informed about matches and provided the HDR reference number. Contact information to the relevant person is provided in Table 3. Samples collected by WCS are mainly from the New York coastal area. Although the majority of samples provided by the Center for Coastal Studies (CCS) originate from the Gulf of Maine, CCS also receives samples from their network of collaborators from other areas along the US eastern sea border.

Table 3. HDR humpback whale samples matching to samples collected by other institutions

Sample ID	Matching reference	Sex	Matching sample reference	# Matching loci
VA150001	HDR001	female	WCS001	18
VA150006	HDR002	female	WCS002	19
VA150007	HDR003	male	CCS001	18
VA150011	HDR004	female	CCS002	14
VA160009	HDR005	male	CCS003	19
VA160011	HDR006	female	CCS004	19
VA160011	HDR007	female	CCS005	19
VA160014	HDR008	female	CCS006	18
VA170005	HDR009	male	CCS007	19
VA170006	HDR010	male	CCS008	18
VA170009	HDR011	female	CCS009	19
VA170009	HDR012	female	CCS010	15
VA170010	HDR013	male	CCS011	19
VA180001	HDR014	male	CCS012	19
VA190003	HDR015	female	CCS013	18
VA190004	HDR016	female	CCS014	14
VA190005	HDR017	female	CCS015	19
VA190006	HDR018	male	CCS016	19

Notes: WCS reference samples contact: Dr. Howard Rosenbaum, Wildlife Conservation Society, email: hrosenbaum@wcs.org, CCS reference samples contact: Dr. Jooke Robbins, Center for Coastal Studies, email: jrobbins@coastalstudies.org

Following the above procedures nine loci were deemed sufficient to discern amongst individuals. A single pair of duplicate samples was detected between two fin whale samples (the two HDR fin whale biopsy samples collected in 2015). None of the HDR fin whale samples matched to the 1,789 (incl. HDR samples) samples contained in the North Atlantic genetic fin whale archive genotyped at 21 microsatellite loci at a genotyping rate of 0.9909. The samples in the data base are summarized in Table 4.

Table 4. Sample areas, years and sample sizes for fin whales

General area	sampling years	total sample size
eastern US sea border	1991-2019	193
eastern Canadian sea border	1990-2019	582
Central North Atlantic	1985-2018	479
eastern North Atlantic	1982-2020	535

Data availability

All genetic data generated from the HDR samples were transmitted by in attachments along with this report. The sex and microsatellite genotypes are provided in a so-called one-line STRUCTURE format (Pritchard, Stephens, & Donnelly, 2000), where the first column after sample identification number denotes the sex, and a **-9** denotes missing data. The humpback file is named `hdr_mn_genotpes_w_sex.str.` The fin whale genotype data are in file `hdr_bp_genotypes_w_sex.str`. Mitochondrial control region DNA sequences are also provided as an attachment in fasta format. The file for the humpback DNA sequence data is named `hdr_mn_mt_sequences.fas`, and the fin whale DNA sequence data file `hdr_bp_mt_sequences.fas`.

A brief cautionary note on interpretations of matches

While matches identified in this analysis can be viewed as "true" and provide insights into connections between the whales sampled by HDR and other areas, great caution should be applied with regards to absences of matches as well as relative matching rates. As mentioned above the sampling effort varies substantially across time and space. The lack of any matches to the fin whale HDR biopsies in our database is likely due to the observation that the overall abundance of fin whales in the North Atlantic is ~70,000 individuals, which implies that ~ 1,800 samples represent a very low sampling proportion and hence a low chance of a match overall.

Acknowledgements

The work presented here would have been impossible but for our many collaborators who have tirelessly collected and contributed samples over three decades

Appendix I. HDR and corresponding laboratory sample identification numbers and molecular sex determinations

Original Animal Number This is HDR's ID number	Current Animal Number This is SAMPLE number	Species	MarECon Ids	Sex
HDRVAMn009	20150102_DTE_Mn_001	*M. novaeangliae*	VA150001	female
HDRVAMn010	20150106_DTE_Mn_001	*M. novaeangliae*	VA150002	male
HDRVAMn011	20150106_DTE_Mn_002	*M. novaeangliae*	VA150003	female
HDRVAMn013	20150111_DTE_Mn_001	*M. novaeangliae*	VA150004	female
HDRVAMn014	20150111_DTE_Mn_002	*M. novaeangliae*	VA150005	female
HDRVAMn015	20150111_DTE_Mn_003	*M. novaeangliae*	VA150006	female
HDRVAMn023	20150120_DTE_Mn_001	*M. novaeangliae*	VA150007	male
HDRVAMn024	20150122_DTE_Mn_001	*M. novaeangliae*	VA150008	male
HDRVAMn025	20150122_DTE_Mn_002	*M. novaeangliae*	VA150009	male
HDRVAMn005	20150129_DTE_Mn_001	*M. novaeangliae*	VA150010	female
HDRVAMn027	20150129_DTE_Mn_002	*M. novaeangliae*	VA150011	female
HDRVAMn029	20150209_DTE_Mn_001	*M. novaeangliae*	VA150012	male
HDRVABp005	20150429_DTE_Bp_001	*B. physalus*	VA150013	male
HDRVABp005	20150429_DTE_Bp_002	*B. physalus*	VA150014	male
HDRVAMn039	20151207_DTE_Mn_001	*M. novaeangliae*	VA150015	male
HDRVAMn041	2015Dec09_DTE_Mn_001	*M. novaeangliae*	VA150016	male
HDRVAMn044	2015Dec10_DTE_Mn_001	*M. novaeangliae*	VA150017	female
HDRVAMn045	2015Dec20_DTE_Mn_001	*M. novaeangliae*	VA150018	male
HDRVAMn048	2015Dec20_DTE_Mn_002	*M. novaeangliae*	VA150019	male
HDRVAMn052	2016Jan15_DTE_Mn_001	*M. novaeangliae*	VA160001	female
HDRVAMn050	2016Jan15_DTE_Mn_002	*M. novaeangliae*	VA160002	male
HDRVAMn051	2016Jan15_DTE_Mn_003	*M. novaeangliae*	VA160003	male
HDRVAMn054	2016Jan15_DTE_Mn_004	*M. novaeangliae*	VA160004	male
HDRVAMn063	2016Feb09_DTE_Mn_001	*M. novaeangliae*	VA160005	male
HDRVAMn061	2016Feb17_DTE_Mn_001	*M. novaeangliae*	VA160006	male
HDRVAMn069	2016Nov01_DTE_Mn_001	*M. novaeangliae*	VA160007	female
HDRVAMn071	2016Nov01_DTE_Mn_002	*M. novaeangliae*	VA160008	male
HDRVAMn031	2016Nov03_DTE_Mn_001	*M. novaeangliae*	VA160009	male
HDRVAMn059	2016Nov03_DTE_Mn_002	*M. novaeangliae*	VA160010	male
HDRVAMn049	2016Nov18_DTE_Mn_001	*M. novaeangliae*	VA160011	female
HDRVAMn064	2016Dec13_DTE_Mn_001	*M. novaeangliae*	VA160012	female
HDRVAMn012	2016Dec21_DTE_Mn_001	*M. novaeangliae*	VA160013	female
HDRVAMn082	2016Dec21_DTE_Mn_002	*M. novaeangliae*	VA160014	female
HDRVAMn081	2016Dec28_DTE_Mn_001	*M. novaeangliae*	VA160015	female
HDRVAMn084	2017Jan01_DTE_Mn_001	*M. novaeangliae*	VA170001	female
HDRVAMn084?	2017Jan01_DTE_Mn_001	*M. novaeangliae*	VA170002	female
HDRVAMn066	2017Jan01_DTE_Mn_002	*M. novaeangliae*	VA170003	male
HDRVAMn092	2017Jan11_DTE_Mn_001			
HDRVAMn095	2017Jan16_DTE_Mn_001*Jan17 on tube	*M. novaeangliae*	VA170004	male
HDRVAMn092	2017Jan19_DTE_Mn_001	*M. novaeangliae*	VA170005	male
HDRVAMn090	2017Jan21_DTE_Mn_001	*M. novaeangliae*	VA170006	male
HDRVAMn101	2017Jan21_DTE_Mn_002	*M. novaeangliae*	VA170007	male
HDRVAMn102	2017Jan25_DTE_Mn_001	*M. novaeangliae*	VA170008	male
HDRVAMn007	2017Feb01_DTE_Mn_001	*M. novaeangliae*	VA170009	female
HDRVAMn091	2017Feb02_DTE_Mn_001	*M. novaeangliae*	VA170010	male
HDRVAMn099	2017Feb02_DTE_Mn_002	*M. novaeangliae*	VA170011	female
HDRVAMn088	2017Feb06_DTE_Mn_001	*M. novaeangliae*	VA170012	male
HDRVAMn093	2017Feb06_DTE_Mn_002	*M. novaeangliae*	VA170013	female
HDRVAMn098	2017Feb06_DTE_Mn_003	*M. novaeangliae*	VA170014	female
HDRVAMn104	2017Feb17_DTE_Mn_001	*M. novaeangliae*	VA170015	female
HDRVAMn105	2017Feb17_DTE_Mn_002	*M. novaeangliae*	VA170016	female
HDRVAMn111	2017Feb24_DTE_Mn_001	*M. novaeangliae*	VA170017	female
HDRVAMn096	2017Feb24_DTE_Mn_002	*M. novaeangliae*	VA170018	male
HDRVAMn112	2017Mar21_DTE_Mn_001	*M. novaeangliae*	VA170019	male
HDRVAMn109	2017Mar21_DTE_Mn_002	*M. novaeangliae*	VA170020	male
HDRVABp019	2017May10_DTE_Bp_001	*B. physalus*	VA170021	female
HDRVABp026	2017Aug17_DTE_Bp_001	*B. physalus*	VA170022	male
HDRVABp027	2017Aug17_DTE_Bp_002	*B. physalus*	VA170023	male
HDRVAMn120	2017Dec22_DTE_Mn_001	*M. novaeangliae*	VA170024	male
HDRVAMn122	2017Dec29_DTE_Mn_001	*M. novaeangliae*	VA170025	male
HDRVAMn126	2018Feb09_DTE_Mn_001	*M. novaeangliae*	VA180001	male
HDRVAMn132	2018Jul31_DTE_Mn_001	*M. novaeangliae*	VA180002	female
HDRVAMn146	2019Jan04_DTE_Mn_001	*M. novaeangliae*	VA190001	female
HDRVAMn153	2019Feb03_DTE_Mn_001	*M. novaeangliae*	VA190002	female
HDRVAMn154	2019Feb03_DTE_Mn_002	*M. novaeangliae*	VA190003	female
HDRVAMn151	2019Jan31_DTE_Mn_001	*M. novaeangliae*	VA190004	female
HDRVAMn152 changed to HDRVAMn156	2019Feb14_DTE_Mn_001	*M. novaeangliae*	VA190005	female
HDRVAMn154 changed to HDRVAMn152	2019Mar02_DTE_Mn_001	*M. novaeangliae*	VA190006	male
HDRVAMn163	2019May04_DTE_Mn_001	*M. novaeangliae*	VA190007	female
HDRVABp049	2018Apr22_DTE_Bp_001	*B. physalus*	VA180003	male
HDRVABp048	2018Apr22_DTE_Bp_002	*B. physalus*	VA180004	male
HDRVABp046	2018Apr22_DTE_Bp_003	*B. physalus*	VA180005	male

Cited literature

Arnason, U., Gullberg, A., & Widegren, B. (1993). Cetacean mitochondrial DNA control region: Sequences of all extant baleen whales and two sperm whale species. *Molecular Biology and Evolution*, *10*(5), 960–970.

Bérubé, M., Aguilar, A., Dendanto, D., Larsen, F., Notarbartolo di Sciara, G., Sears, R., … Palsbøll, P. J. (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. *Molecular Ecology*, *7*(5), 585–599.

Bérubé, M., Berchok, C., & Sears, R. (2001). Observation of a male-biased sex ratio in the Gulf of St. Lawrence fin whales (\emphBalaenoptera physalus): Temporal, geographical or group segregation? *Marine Mammal Science*, *17*(2), 251–268.

Bérubé, M., Jørgensen, H., McEwing, R., & Palsbøll, P. J. (2000). Polymorphic di-nucleotide microsatellite loci isolated from the humpback whale, *Megaptera novaeangliae*. *Molecular Ecology*, *9*(12), 2181–2183. doi: 10.1046/j.1365-294X.2000.01053.x

Bérubé, M., Rew, M. B., Skaug, H., Jørgensen, H., Robbins, J., Best, P., … Palsbøll, P. J. (2005). Polymorphic microsatellite loci isolated from humpback whale, *Megaptera novaeangliae* and fin whale, *Balaenoptera physalus*. *Conservation Genetics*, *6*(4), 631–636. doi: 10.1007/s10592-005-9017-5

Drouot, V., Bérubé, M., Gannier, A., Gould, J. C., Reid, R. J., & Palsbøll, P. J. (2004). A note on genetic isolation of Mediterranean sperm whale, *Physeter macrocephalus*, suggested by mitochondrial DNA. *Journal of Cetacean Research and Management*, *6*, 29–32.

Mullis, K. B., & Faloona, F. (1987). Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction. *Methods in Enzymology*, *155*, 335–350.

Paetkau, D., & Strobeck, C. (1994). Microsatellite Analysis Of Genetic-Variation In Black Bear Populations. *Molecular Ecology*, *3*(5), 489–495.

Palsbøll, P. J., Allen, J. M., Bérubé, M., Clapham, P. J., Feddersen, T. P., Hammond, P. S., … Øien, N. (1997). Genetic tagging of humpback whales. *Nature*, *388*(6644), 767–769. doi: 10.1038/42005

Palsbøll, P. J., Bérubé, M., Larsen, A. H., & Jørgensen, H. (1997). Primers for the amplification of tri- and tetramer microsatellite loci in baleen whales. *Molecular Ecology*, *6*(9), 893–895. doi: 10.1111/j.1365-294X.1997.tb00146.x

Palsbøll, P. J., Vader, A., & Bakke, I. (1992). Determination of gender in cetaceans by the polymerase chain reaction. *Canadian Journal of ...*, *70*(11), 2166–2170. doi: 10.1139/z92-292

Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, *155*(2), 945–959.

Rew, M. B., Robbins, J., Mattila, D. K., Palsbøll, P. J., & Bérubé, M. (2011). How many genetic markers to tag an individual? An empirical assessment of false matching rates among close relatives. *Ecological Applications*, *21*(3), 877–887. doi: 10.1890/10-0348.1

Valsecchi, E., & Amos, W. (1996). Microsatellite markers for the study of cetacean populations. *Molecular Ecology*, *5*(1), 151–156. doi: 10.1111/j.1365-294X.1996.tb00301.x

Werle, E., Schneider, C., Renner, M., Volker, M., & Fiehn, W. (1994). Convenient Single-Step, One Tube Purification of Pcr Products for Direct Sequencing. *Nucleic Acids Research*, *22*(20), 4354–4355.