SC/68C/SD & DNA/02

Sub-committees/working group name: SD & DNA

Neutral microsatellite loci reveal the presence of distinct genetic units of North Atlantic humpback whales (Megaptera novaeangliae) off the Icelandic coast feeding area

Marc-Alexander Gose, Maximilian Schweinsberg, Christophe Pampoulie, Sverrir Daníel Halldórsson, Marianne H. Rasmussen, Gísli Víkingsson, Ralph Tollrian

Neutral microsatellite loci reveal the presence of distinct genetic units of North Atlantic humpback whales (*Megaptera novaeangliae*) off the Icelandic coast feeding area.

Marc-Alexander Gose[1, 4] & Maximilian Schweinsberg[1], Christophe Pampoulie[2], Sverrir Daníel Halldórsson[2], Marianne H. Rasmussen[3], Gísli Víkingsson[2], Ralph Tollrian[1]

1 Department of Animal Ecology, Evolution and Biodiversity, Ruhr-University Bochum, Bochum, Germany

2 Marine and Freshwater Research Institute, Hafnarfjörður, Iceland

3 Húsavík Research Center, University of Iceland, Húsavík, Iceland

4 Royal (Dick) School of Veterinary Studies and the Roslin Institute, University of Edinburgh, Edinburgh, UK

Correspondence

Marc-Alexander Gose, Department of Animal Ecology, Evolution and Biodiversity, Ruhr-University Bochum, 44801 Bochum, Germany

E-Mail: marc-alexander.gose@rub.de

Abstract

North Atlantic humpback whales (*Megaptera novaeangliae*; NAHW) have recovered to a population density close to estimated pre-exploitation levels. Large-scale genetic studies on NAHW, as one important indicator about the fitness of a population, have not been carried out for decades and are therefore much needed to explore genetic properties of a growing population. Capture-recapture studies suggest that Iceland represents an important feeding area for NAHW from different breeding stocks and is therefore a suitable area to assess/confirm the potential presence of several breeding stocks within a single feeding area. In this study, 101 individual humpback whales sampled off the Icelandic coast were genotyped using sequences of the mitochondrial control region and 17 microsatellite markers. Genetic diversity was relatively high (π=0.0096; H_O=0.690) and we detected no sign of inbreeding (F_{IS}=0.0587). Haplotype analysis revealed two separate clades and analysis of nuclear markers revealed a differentiation into three distinct genetic clusters with a moderate genetic distance (F_{ST}=0.044-0.055). Kinship analysis detected a high density of closely related pairings in our sample set, indicating inflicted site-fidelity as a possible driver of genetic differentiation. Our study highlights Iceland as an important habitat for genetically diverse humpback whales.

KEYWORDS: feeding area, humpback whale, Iceland, North Atlantic, Megaptera novaeangliae, population genetics, relatedness

1. Introduction

Baleen whales are large marine predators with a complex life history which includes high seasonal variability in energy intake and have been characterized as capital breeders (Christiansen et al. 2014; Lockyer 1986; Víkingsson 1995). Species like the humpback whale (*Megaptera novaeangliae*) are dependent on seasonal fluctuations in nutrient-availability and have adapted to these circumstances by displaying some of the longest known animal migrations. In general, during summer months most baleen whale species forage on large aggregations of prey such as swarming euphausiids, small schooling pelagic fish or copepods in temperate and polar waters followed by a migration to warmer waters near the equator where calving and mating behavior is prioritized (Clapham 1996; Corkeron and Connor 1999). Multi-institutional collaborations across the North Atlantic have increased our knowledge about the migratory behavior of the North Atlantic humpback whale (NAHW) during the last decades. Nevertheless, our understanding of the movements of NAHW is being altered frequently as new migration patterns of individuals are detected, revealing that many details of their migration remain unknown (Berrow et al. 2021; Horton et al. 2011; Jann et al. 2003; Rinaldi et al. 2009). Although recognized as a generalist species, there are behavioral factors that could remain widely stable throughout an individual's lifespan, suggesting that biological patterns, such as maternally inflicted site-fidelity, are influencing migratory behavior (Barendse et al. 2013; Richard et al. 2018; Weinrich et al. 2006). Breeding ground origin of an individual seems to play an important role in the development of its migratory behavior and could remain constant throughout an individual's life span. Long-term observations identified a large number of individuals at the same breeding and calving areas for decades with inter-breeding-area exchange occurring only at an extremely low frequency (Palsbøll et al. 1995; Richard et al. 2018; Stevick et al. 2003; 2016; Wenzel et al. 2009). Similarly, occupancy of broader feeding areas as well as route fidelity during migration periods also seems to be constant in some individuals (Horton et al. 2017; 2020; Stevick et al. 2006). This site-fidelity seems to be especially well developed in females, potentially guiding their offspring to well-known areas with high probability of foraging success (Craig and Herman 1997; Stevick et al. 2016).

Low levels of admixture between breeding areas could lead to a separation into more than two (i.e., eastern, and western) breeding stock subpopulations, especially in the western North Atlantic where known breeding habitats can be far apart from each other. Earlier studies have suggested the presence of at least two genetic units utilizing the Icelandic feeding grounds (Larsen et al. 1996; Palsbøll et al. 1995). Large-scale capture-recapture studies have used photographs of the ventral side of the fluke for identification of individuals and linking them to their feeding and breeding area. In Icelandic waters, a complex migratory pattern and four separate migratory units potentially representing four different breeding ground destinations were suggested: Cape Verde Archipelago, the Lesser Antilles, the Greater Antilles, and the Dominican Republic (Chosson-P et al. 2015). These findings indicate that NAHW coexisting in productive feeding grounds like Iceland seem to split into several breeding stock units, which use different areas for their breeding and mating behavior. However, it is yet unclear if this pattern is consistent enough throughout an individual's lifespan or throughout generations to influence the genetic structure of the NAHW population.

Although the assessment of effective population sizes of highly mobile cetaceans remains difficult, a variety of studies identified extensive population growth in NAHW during the last decades (Heide-Jørgensen et al. 2012; Jackson et al. 2014; Pike et al. 2009). Conservation measures, such as the protection from commercial whaling in Icelandic waters in 1955 and in the North Atlantic 1956 (and worldwide since 1966) combined with its generalist behavior regarding prey and habitat choices seem to have promoted rapid growth of humpback whale populations not only in the North Atlantic, but worldwide (Chero et al. 2020; Ruegg et al. 2013; Víkingsson et al. 2015; Zerbini et al. 2010). Nevertheless, multiple anthropogenic activities such as fishing interactions, vessel noise, ship strikes, commercial tourism, seismic activities (oil and gas drilling), military activities and intake of pollutants already put increasingly high pressure on cetacean populations (Basran et al. 2019; Christiansen et al. 2015; Cooke 2018; Hoyt 2018; Ryan et al. 2013). If not managed properly, these stressors may have severe negative effects on NAHW in the near future (Noad et al. 2019; Santora et al. 2020). Therefore, detecting population structure and recognizing areas with high genetic diversity may help to inform conservation entities about the value of these areas.

Altogether, humpback whales seem to be characterized by a higher adaptability than closely related species of baleen whales, perhaps being the advantageous species during changing environmental circumstances. This adaptability is mostly shown in their occupancy of foraging areas, where humpback whales were not particularly present a few decades ago (Brown et al. 2019; Ramm et al. 2020). However, a certain stability of their behavior could be resulting off their site-fidelity, potentially restricting geneflow and facilitating population differentiation. Understanding underlying patterns and processes influencing and shaping migratory routes is important for assessing the status of a population. Especially in recent times where human activities may put pressure on the entire population(s) of NAWHs, potentially endangering their future survival in a changing environment. Here, we aim to investigate if previously suggested differentiated breeding stock units in Icelandic humpback whales (Chosson-P et al. 2015) can be detected via classical population genetic approaches. We analyzed genetic diversity and population structure in order to detect potentially differentiated units and their level of relatedness. The results of this study increase our understanding of admixture and gene flow within a highly mobile cetacean population and give implications about the value of Icelandic feeding grounds for their long-term survival.

2. Materials and Methods

2.1 Sample collection, DNA extraction and amplification

A total of 101 tissue samples were collected between 1998 and 2018. Samples were taken from stranded animals (*n*=21) and actively taken from live animals during research expeditions (*n*=80). Of the live biopsy samples, 20 individuals were taken before 2010, 15 were taken in 2011, 22 were taken between 2012 and 2013, 8 were taken between 2014 and 2015 and 15 were taken between 2016 and 2018. Most biopsy samples were taken in two bays/fjords (Skjálfandalói and Eyjafjörður) located in northern Iceland while stranded animals occurred all around Iceland (Figure 3). Biopsy sampling was conducted using a modified rifle with biopsy darts attached to a 7 x 20 mm sanitized stainless-steel cutting tip. Individuals were photographed and catalogued to avoid resampling of individuals.

111 All samples were stored in 70% ethanol and -20 °C until DNA extraction was carried out. Qiagen
112 DNeasy© Blood and Tissue kit was used to extract total nuclear and mitochondrial DNA following the
113 manufacturer's protocol. DNA concentration was measured using a BioPhotometer from Eppendorf.

114 The mitochondrial control region was amplified using previously published primer sequences
115 light-strand Dlp-10 and heavy-strand Dlp-10 (Baker et al. 1993). Additionally, two sex specific primers
116 (SRY and ZFX/ZFY; Berube and Palsbøll 1996) were amplified for sex determination. PCR (12.5 µl)
117 contained 1.0 µl total DNA, 6.25 µl GOTaq® G2 Colorless Master Mix (x10), 0.25 µl of each primer
118 (reverse and forward; 10 pmol/µl) and 4.75 µl nuclease-free water. PCR settings differed between
119 sex-specific primers and the mitochondrial primer. SRY and ZFX/ZFY were amplified with an initial
120 denaturation at 94 °C of 1:00 min, 35 cycles of denaturation at 94 °C with 0:45 min, annealing at 58 °C
121 with 0:45 min, elongation at 72 °C with 1:00 min and finished with a final elongation step at 72 °C with
122 5:00 min. The PCR for the mitochondrial region was carried out with an initial denaturation at 95 °C of
123 2:00 min, 35 cycles of denaturation at 95 °C with 1:00 min, annealing at 60 °C with 0:30 min, elongation
124 at 72 °C with 1:30 min and a final elongation step at 72 °C with 3:00 min.

125 PCR amplification success was estimated running 1 µl of amplification products in an agarose
126 electrophoresis. Subsequently, the unincorporated primers and nucleotides were removed with ExoSAP
127 digestion (GE Healthcare, Illustra) and the samples were sequenced at the department of Receptor
128 Biochemistry, Faculty of Chemistry, Ruhr University Bochum. Mitochondrial control region sequences
129 were deposited at GenBank (Acc. #).

130 Furthermore, all samples were genotyped using 20 published microsatellites: GATA28, GATA53,
131 GATA98, GATA417, GGAA520, TAA31 (Palsbøll et al. 1997), EV1, EV14, EV21, EV37, EV94,
132 EV96, EV104 (Valsecchi and Amos 1996), RW-18, RW-48 (Waldick et al. 2002), GT023, GT211
133 (Bérubé et al. 2000), 199/200, 417/418, 464/465 (Schlötterer et al. 1999). Each forward primer was
134 attached to a fluorescent-dye labeled M13-tail (5'-CACGACGTTGTAAAACGA-3'). Amplification
135 was carried out in 12.5 µl reaction volumes containing 1.0 µl template DNA, 6.25 µl GOTaq® G2
136 Colorless Master Mix (x10), 0.25 µl labeled forward primer, 0.25 µl unlabeled reverse primer
137 (10 pmol/µl each), 0.25 µl M13-IRDye® and 4.5 µl nuclease-free water. PCR settings differed between

the loci and are summarized in Table S1 of the supplementary material. Allelic size was determined with a high-resolution polyacrylamide gel electrophoresis and the software Saga 2GT (LI-COR, Biosciences, Bad Homburg, Germany). Deviating samples were re-amplified and rescored two times independently analyzed by a second person in order to confirm the deviating genotype. All peak profiles that were either faint or ambiguous were omitted from the analyses. Moreover, two samples were established as standard samples to ensure cohesive fragment size evaluation.

2.2 Molecular analyses

Alignment and cropping of sequences of the mitochondrial control region was carried out in GENEIOUS 11.1.5 (Kearse et al. 2012). A sample of every detected haplotype was tested for matching genotypes of the target species using NCBI BLAST to check for species specificity of the sequenced region. All resulting sequences were aligned using MAFFT (multiple alignment using fast Fourier transform) alignment. Using POPART version 1.7 (Leigh and Bryant 2015) a minimum spanning network according to Bandelt et al.(1999) was created. The software was also used to calculate nucleotide diversity, number of polymorphic sites and Tajima's D value. POPART was also used to perform an analysis of molecular variance (AMOVA) between male and female haplotypes to test for sex-biased dispersal.

Each microsatellite locus was tested for the occurrence of null alleles, large allele dropout or stutter bands using CERVUS (Kalinowski et al. 2007). Additionally, deviations from Hardy-Weinberg equilibrium (HWE) and linkage was tested with ARLEQUIN 2.3.4 (Excoffier and Lischer 2010). The microsatellite genotypes were further analyzed in R 3.6.0 and R STUDIO (Ihaka and Gentleman 1996; RStudio Team 2015). A discriminant analysis of principle components (DAPC) was carried out with groups defined by genetic clusters using functions of *adegenet* package (Jombart 2008; Jombart et al. 2010). Genetic clustering was performed with the amount of clusters *K* determined by the Bayesian information criterion (BIC) which has proven to be the most suitable in *K*-means clustering (Schwarz 1978). Deviations from Hardy-Weinberg equilibrium were again tested using *pegas* package (Paradis 2010). The inbreeding coefficient F_{IS} for the overall population and individual F_{IS} for each cluster were calculated using *hierfstat*. The overall fixation index F_{ST} (Nei and Chesser 1983) and pairwise F_{ST} (Weir

and Cockerham 1984) was calculated in a modelled situation, where individuals were assigned to a population *a priori* based on the cluster assignment inferred by BIC using *hierfstat* package (Goudet 2005). We further examined the number of discrete populations (*K*) with a second Bayesian clustering program STRUCTURE 2.3.4 (Pritchard et al. 2000) for a more detailed population differentiation. Initial STRUCTURE runs used the default settings with a burn-in period of 10,000 and 10,000 MCMC repeats with 20 iterations per *K* (1-10). The results were uploaded to STRUCTURE HARVESTER (Earl and von Holdt 2012). The cluster ($K = 3$) determined most likely was used to perform a second run with more iteration (100) and higher values for the burn-in period (100,000) and MCMC repeats (100,000), to increase the accuracy of clustering. The resulting merged data file was then processed using the program CLUMPP 1.1.2 (Jakobsson and Rosenberg 2007) and visualized by DISTRUCT 1.1 (Rosenberg 2004). The Bayesian cluster analyses and DAPCs were additionally run with datasets representing only males and only females and different time periods to check temporal or sex-specific differences in population structure.

We further estimated the statistical power of our microsatellite dataset to detect population differentiation through the program POWSIM 4.1 (Ryman and Palm 2006). We modelled different scenarios with a fixed population size (n=100 to reflect this study) and varying levels of population differentiation (F_{ST} = 0.005, 0.01, 0.025, 0.05, 0.1) as well as a model for the estimation of α-error (t=0). Number of burn-ins, batches and iterations were set to 1000, 100 and 1000 and for each scenario, 1000 runs were performed.

2.3 Genealogy

The full-pedigree likelihood program COLONY 2.0.5.6 (Jones and Wang 2010) was used to determine full and half siblingships. Settings were set for a dioecious, diploid organism with a polygamic mating system. Genotypes of male and female individuals were put into separate files and loaded into COLONY to infer sex of the individuals. The analysis method was set to FPLS (full likelihood and pairwise likelihood combined) and the run length was set to medium. The run was carried out twice with the same settings to increase precision. Using the pairwise likelihood approach, the program was able to calculate

191 possible parent-offspring pairs with a confidence interval of 95% which are represented as full siblings
192 by the full likelihood approach.

193 A second pairwise likelihood approach was conducted using the program ML-RELATE (Kalinowski et
194 al. 2006) to compare the results of COLONY with the results of another program. Results of both
195 programs were merged, and uncertain pairings were excluded to reach higher precision in genealogy
196 analysis.

197 All pairings were then compared with cluster assignments inferred by the Bayesian cluster programs
198 and their individual haplotype at the mitochondrial control region. We further included the sex and date
199 of sampling for each individual that has been assigned to a pairing.

200 # 3. Results

201 ## 3.1 Marker quality

202 DNA extraction was successful for all 101 samples. The analyzed loci were tested for their quality using
203 Hardy-Weinberg Equilibrium (HWE). For each locus, the expected heterozygosity (H_E) was compared
204 with the observed heterozygosity (H_O) that resulted out of the allele frequencies gained by genotyping
205 (Table 1). A total of 17 markers showed no deviation from HWE, while three markers show significant
206 deviation and potential presence of null alleles. Three other markers show a possibility of an error
207 because of stuttering or shadow bands. Both sources of error were checked again and could be excluded
208 due to accurate and independent scoring, which was conducted twice with the same results for all loci.

209 ## 3.2 Genotypic and allelic diversity

210 All analyzed loci were polymorphic and the numbers of alleles per locus in microsatellites analyzed
211 ranged from 4-14 alleles per locus. The mean number of alleles per locus was 6.35 which is slightly
212 lower than observed allelic diversities in previous studies (Cypriano-Souza et al. 2010; Larsen et al.
213 1996).

214 **Table 1: Marker quality of 20 microsatellite loci. Given is the specific annealing temperature for amplification (T_A), the**
215 **core sequence, alleles per locus (N_A), the expected heterozygosity (H_E) and the observed heterozygosity (H_O). Significant**

deviations (p=0.05 (*), p=0.001 (***)) in heterozygosity from Hardy-Weinberg Equilibrium are represented by bold typeface. The overall values were calculated without the three omitted markers.

Locus	T_A	Core Sequence	N_A	H_E	H_O
EV21[a]	50	$(GATA)_n$	7	0.793	0.703
EV37[a]	54	$(TAA)_n$	14	0.879	0.792
GATA28[b]	50	$(GATA)_n$	6	0.445	0.465
EV14[a]	54	$(GATA)_n$	6	0.592	0.515
GATA53[b]	50	$(GATA)_n$	10	0.803	0.772
GATA417[b]	54	$(GGAA)_n$	11	0.871	0.851
RW18[d]	50	$(AC)_n(TC)_n$	2	0.496	**0.670 (*)**
EV1[a]	50	$(GT)_n$	4	0.443	0.436
EV94[a]	50	$(AC)_n$	7	0.699	0.604
EV96[a]	50	$(AC)_n$	11	0.818	0.733
RW48[d]	50	$(TC)_n$ […] $(AC)_n$	5	0.490	0.485
EV104[a]	50	$(AC)_n$	4	0.367	0.307
GGAA520[b]	50	$(AC)_n(GCAC)_n$	22	0.831	**0.510 (***)**
GT211[c]	52	(TG)TA(TG)	9	0.827	0.663
GT023[c]	57	$(TG)_n$	8	0.831	0.703
GATA98[b]	50	$(GT)_n$	7	0.717	**0.446 (*)**
TAA31[b]	50	$(GT)_n$	8	0.796	0.789
199/200[e]	45	$(CT)_n$	7	0.762	0.760
464/465[e]	45	$(GT)_n$	5	0.656	0.542

417/418[e]	45	$(GT)_n$	11	0.817	0.670
Overall	-	-	6.35	0.690	0.640

a Valsecchi and Amos 1996
b Palsbøll et al. 1997
c Bérubé et al. 2000
d Waldick et al. 2002
e Schlötterer et al. 1999

Two microsatellites showed allele numbers which were not in the postulated range of the source studies. However, both markers also showed possibility of null alleles. Due to the deviation from HWE and possibility of null alleles, the markers RW18, GGAA520 and GATA98 were omitted from analyses. Thus, all subsequent analyses were performed with 17 instead of 20 microsatellite markers.

There was no consistent pattern of linkage observed in all loci. The significance level for this test was p=0.005.

3.3 Haplotype analysis

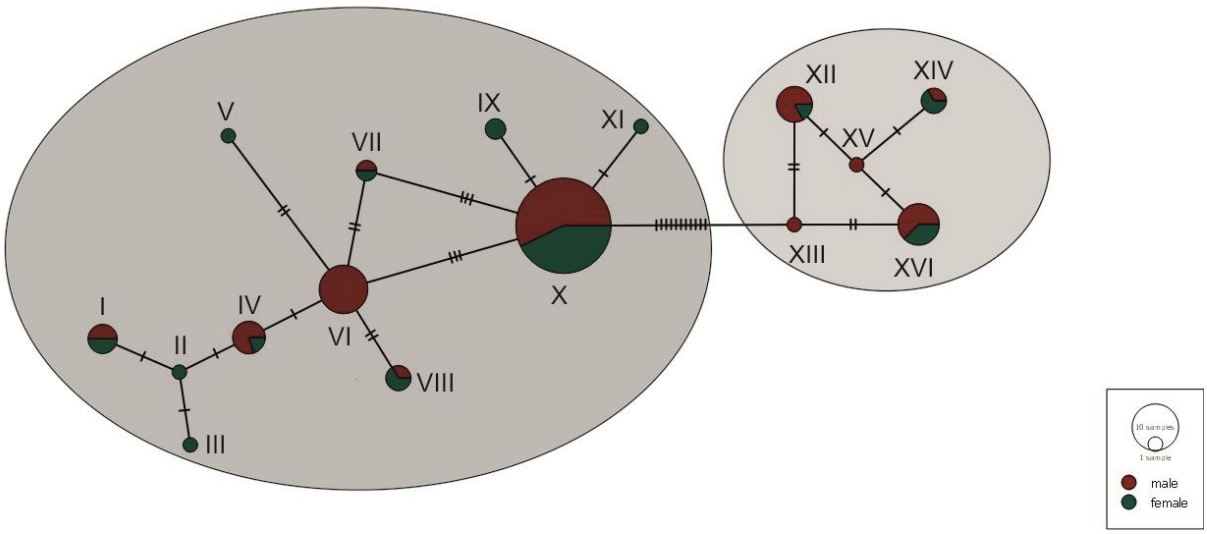


Figure 1: Minimum spanning network of the mitochondrial control region sequence for 94 sequenced individuals of *Megaptera novaeangliae*. Roman numbers indicate the haplotype, colours represent sex (red=male, green=female) and hatch mark represent mutations in a polymorphic site.

Amplification of the mitochondrial control region sequence was successful in 94 samples. The sequences were trimmed and aligned, resulting in a 544 bp long sequence. A sample from every detected haplotype was tested in silico using NCBI BLAST. All tested samples showed at least 99.98% similarity with the target species *Megaptera novaeangliae*. A total of 16 haplotypes were detected. Nucleotide

diversity was π=0.0096 and a total of 25 polymorphic sites have been identified. Haplotype X is the most common haplotype, which is shared by 43 individuals. Six haplotypes were only found in a single individual. The minimum spanning network shows few permutations between most haplotypes (Figure 1). However, a higher number of permutations was detectable between two larger groups with a minimum of 11 permutations between two haplotypes of each group. Haplotype diversity is higher in females with 13 haplotypes detected in female individuals in contrast to 11 haplotypes in male individuals despite the higher proportion of males in our sample set. Five sequences were exclusively found in females while only three were exclusive for males. AMOVA between defined groups by sex showed no significance in molecular variance (Φ_{ST}=20.949) between males and females.

3.4 Population structure

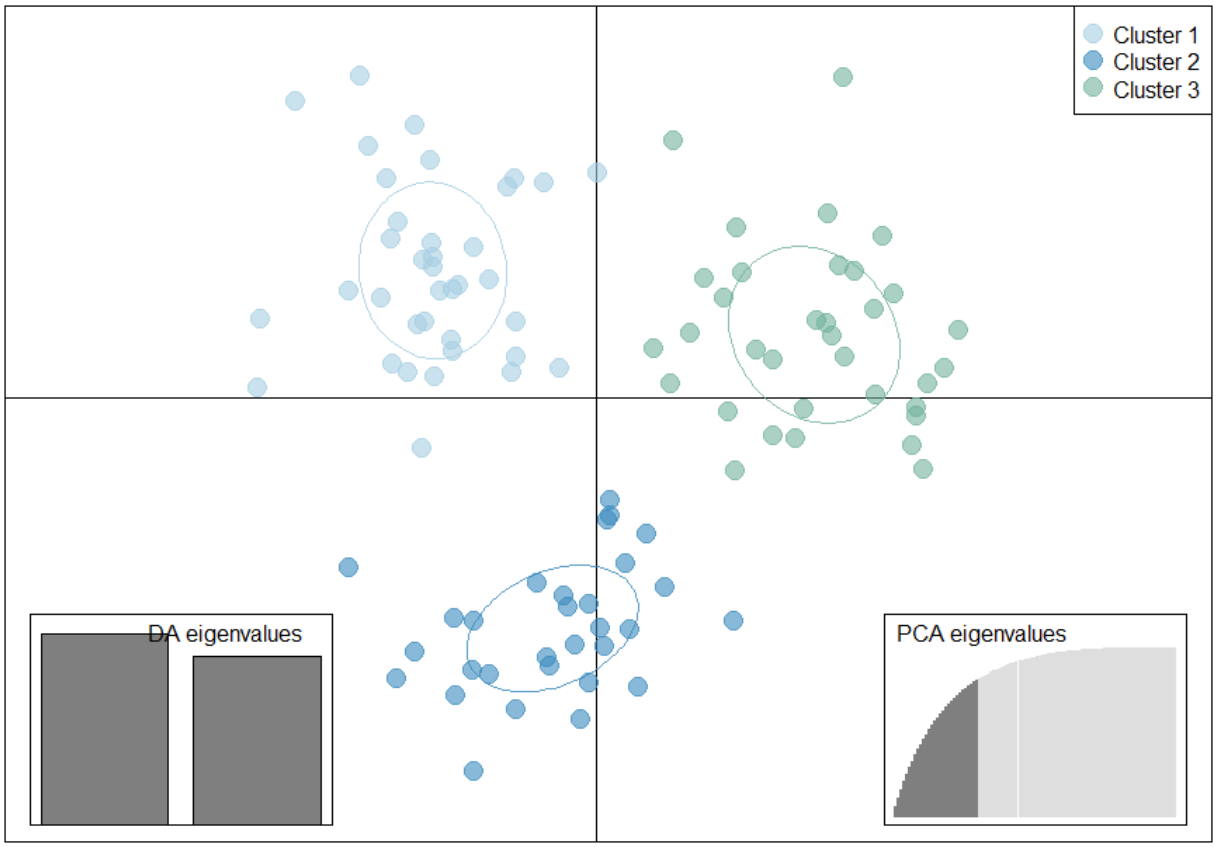


Figure 2: Discriminant analysis of principal components (DAPC) based on multilocus genotypes from 17 microsatellite markers within 101 individual humpback whales. 30 principal components and 2 discriminant functions were retained as indicated by the graphs. Group membership is based on the *K*-mean clustering as inferred by the Bayesian information criterion (BIC).

All 17 microsatellite loci were successfully amplified in 101 individual samples with only very few missing data occurring (<2%). All samples show unique genotypes confirming that no duplicates were included in the analysis. The final sample set comprises of 60 males, 37 females and 4 samples of unknown sex due to insufficient amplification of the sex-specific regions. Male to female ratio therefore is 1.62:1. The overall fixation index for the whole population is F_{ST}=0.0336.

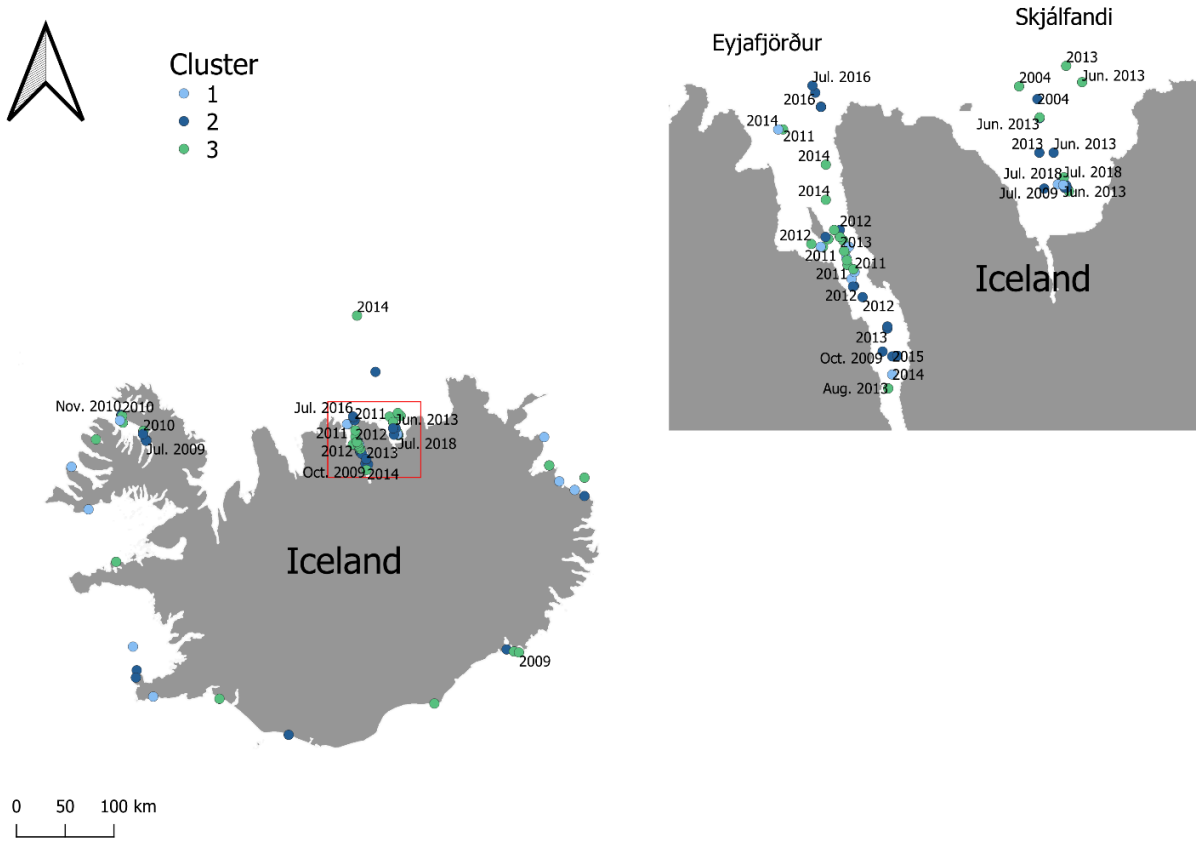


Figure 3: Geographical distribution of samples with group membership (represented by colours) and rough date of sampling, if available.

Genetic clustering using *adegenet* functions and BIC predictions revealed *K*=3 to be the most suitable number of genetic units within the sample set (Figure 2). STRUCTURE software also estimated *K*=3 to be the most suitable model (Supplementary material Figure S1). Clusters were similarly distributed through sex (Supplementary material Figures S2&S3) and there is no pattern of clustering by geographic region (Figure 3). In contrast, Bayesian clustering in five defined different sampling periods yielded optimal *K* between 2 and 5 clusters. Pairwise F_{ST} using three putative populations regarding to *K* detected values between F_{ST}=0.044-0.055 and therefore a moderate differentiation between the clusters. All F_{ST} values were within the 95% confidence interval as calculated by bootstrapping and are therefore

252 significantly different from 0. Overall inbreeding factor F_{IS} is 0.0587 and indicates no extensive
253 inbreeding within the sample size. Individual F_{IS} for each cluster revealed a much lower value for cluster
254 1 (F_{IS}=0.024) than in cluster 2 and 3 (F_{IS}=0.078 and 0.074 respectively). Pairwise F_{ST} between the
255 clusters separated by sex show slight but insignificantly higher values in males than in females. The
256 modelling of statistical power using POWSIM confirms that our dataset provides sufficient power for
257 detecting genetic structure within our sampled population as it is capable to detect population
258 differentiation as little as F_{ST}=0.005 with an accuracy of 89.2% and with a certainty of 100% for all
259 other tested levels of differentiation (Table 2). The α-error for this test as inferred from a model sampling
260 directly from the base population without any generation of genetic drift (t=0) is calculated as 3.5%. All
261 estimates of statistical power are therefore significant ($p<0.05$).

262 **Table 2: Estimation of statistical power of the 17 microsatellite markers used in this study in POWSIM 4.1. Population**
263 **size (n) was fixed at 100 while different levels of predefined genetic differentiation were tested with 1000 runs per model.**
264 **Percentage of runs detecting the F_{ST} represents the statistical power in each scenario.**

n	F_{ST}	runs	n = Chi^2<0.05	% runs detecting F_{ST} (p<0.05)
100	0.005	1000	892	89.2
100	0.010	1000	1000	100
100	0.025	1000	1000	100
100	0.050	1000	1000	100
100	0.100	1000	1000	100

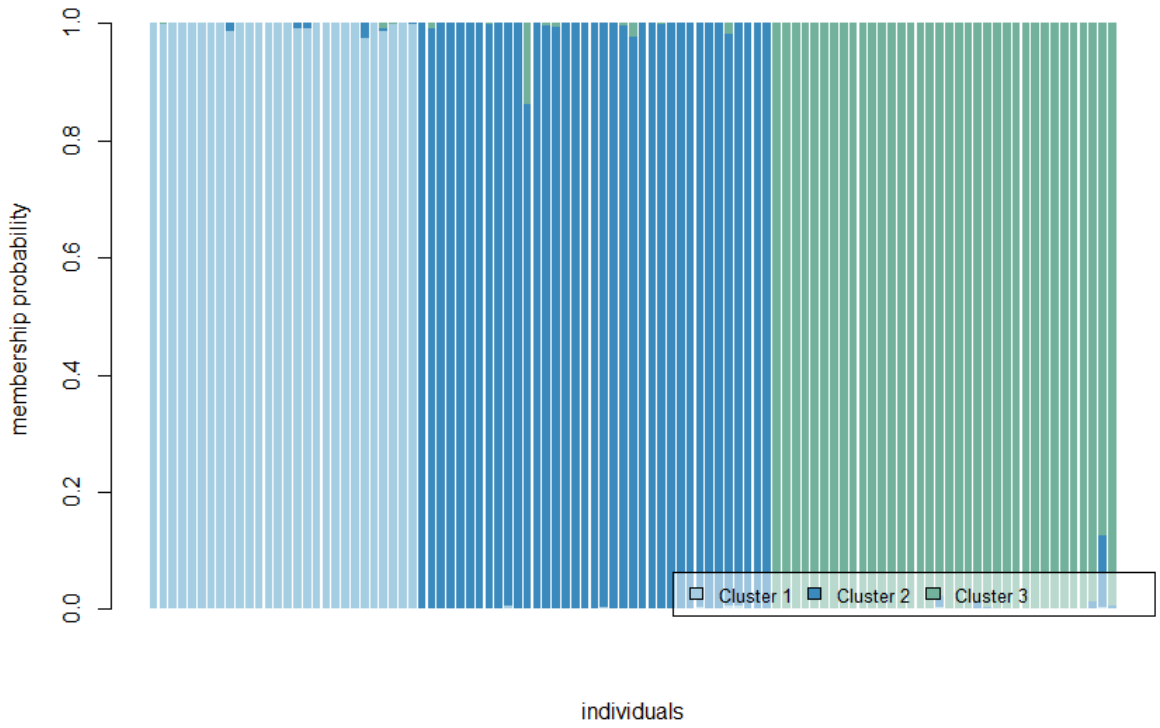


Figure 4: Assignment plot for all multilocus genotypes (*K* = 3). Each bar on the x-axis represents one individual and their individual group membership is represented by the color assigned to one cluster on the y-axis.

3.5 Genealogy

The pairwise and full-likelihood approaches detected 18 full-siblingship pairings with probabilities between 100 and 90%. Four of these pairings were identified as parent-offspring by the pairwise likelihood approach in COLONY. Since there is no long-term observational data for these individuals, it is uncertain what the true relationship of these pairs is. Each of these pairings were compared with their cluster assignment based on the Bayesian clustering of *adegenet* and STRUCTURE and their haplotype at the mitochondrial control region and their sex (Supplementary material Table S2). All except for one pairing shared the same cluster assignment and the same haplotype (except for four samples, where sequencing of the control region was unsuccessful). Interestingly, eight of the pairings were between individuals which have been sampled in 2011. Only four pairings between individuals sampled in two different years could be detected. Out of all pairings, only two were detected in the genetic cluster 2. All remaining pairings were between individuals who were assigned to cluster 1 (8 pairings) or 3

(7 pairings). Nine pairings shared the most common of the detected haplotypes (X) and three pairings shared the second most common haplotype (VI). One pairing occurred between individuals of the haplotype VIII and four pairings had one individual without a known haplotype and could therefore only be reassured by Bayesian cluster assignment. Lastly, it is notable that all detected pairings were exclusively between members of the same sex except for one pairing, where the sex-specific regions could not be sufficiently amplified in one sample.

4. Discussion

Detecting genetic variability and assessing population structure is essential in understanding the complexity of a population and revealing invisible patterns that shape the genetic composition of a species (Razgour et al. 2019). In marine mammals, population differentiation is often a cryptic process. The lack of physical barriers within one ocean basin calls for high resolution markers and a deep understanding of a species' ecology to explain genetic differentiation. In many highly social odontocetes, cultural differences between subpopulation are a major force that contributes to population differentiation (Genoves et al. 2020; Wiszniewski et al. 2010). However, in species with more loose social behavior such as the humpback whale (Clapham 1996), the role of cultural transmission and its effect on genetic structure is poorly understood. Areas with high densities of humpback whales such as Iceland represent excellent study regions for the assessment of genetic variability and population structure. Detecting genetic diversity in these hot spots can help to acknowledge their importance for a population and thus, inform conservation entities about the value of these areas for their long-term viability on regional and global scales.

Thus, the present study aims to investigate genetic structure within Icelandic NAHW using Bayesian clustering models and discriminant analysis of principal components within 101 individual NAHW. We further investigated the genetic relatedness between the individuals to get insights about the potential effect of cultural transmission on genetic structure in NAHW.

The results from this study should inform conservation management entities about the complexity of this species' population structure in the North Atlantic but most importantly, should encourage

307 researchers to further characterize these units and investigate them separately in terms of diet, habitat
308 preference, contaminant load etc. to increase our understanding of this population and detect potential
309 differences between the units.

310 4.1 Fine-scale population structure

311 Genetic variability at our multilocus genotypes was calculated as an observed heterozygosity of H_O=0.64
312 and the number of alleles N_A=6.35 over all 17 loci. Similar studies on humpback whales in the South
313 Atlantic and North Pacific generally found higher values for N_A and slightly higher values for H_O.
314 However, these studies had larger sample sizes or included individuals from different regions (Baker et
315 al. 2013; Cypriano-Souza et al. 2010). Nucleotide diversity (π=0.0096) is similar to previous studies on
316 mitochondrial control regions of NAHW (Palsbøll et al. 1995). We therefore conclude that the genetic
317 diversity in Icelandic humpback whales is similar to populations in other ocean basins.

318 Bayesian information criterion (BIC) and Evanno clustering for all sampled individuals based on 17
319 microsatellites estimated three clusters to be the most suitable model for our sample set (Evanno et al.
320 2005; Schwarz 1978). The level of genetic differentiation was relatively moderate suggesting
321 differentiated units rather than fully isolated populations (pairwise F_{ST} = 0.044-0.055). Admixture events
322 between breeding stocks, although thought to occur in lower frequencies, further support these results.
323 Inter-breeding stock migration is rarely but regularly detected within west Atlantic breeding stocks as
324 well as between the eastern and western breeding areas (Mackay et al. 2019; Stevick et al. 2016). The
325 later event recently demonstrated that all four detected migration events were performed by male
326 individuals, indicating the possibility of a sex-biased dispersal, which is a common driver of genetic
327 structure in mammals (Lawson et al. 2007). This effect was also previously detected via mtDNA studies
328 on NAHW, suggesting maternally inflicted philopatry and male-biased gene flow (Baker et al. 1993;
329 Palsbøll et al. 1995; Rosenbaum et al. 2002). However, we were not able to detect significant differences
330 in nuclear or mitochondrial population structure within the sexes and our sample set might be somewhat
331 biased due to the higher proportion of males. Baker et al. (2013) found similar results in a study on
332 North Pacific humpback whales and discusses a limited male dispersal in humpback whales due to
333 strong site-fidelity in both sexes as a potential explanation. Thus, site-fidelity may be a strong driver for

334 population structure in NAHW and therefore, likely contributes to the structure we find in Icelandic
335 specimens as previously described for Russian and Australian populations (Barendse et al. 2013;
336 Richard et al. 2018). Altogether, our results generally confirm the findings of Chosson-P et al. (2015)
337 on a genetic basis by detecting fine-scale genetic structure in Icelandic humpback whales. However, our
338 data suggests three differentiated units rather than four groups as suggested by photo-identification data.
339 A possible reason could be short distances and/or high admixture between two breeding areas, most
340 likely in the West Atlantic.

341 Nevertheless, the reasons for fine-scale structure, especially within geographically close breeding areas,
342 may be multifactorial. Beside maternally inflicted site-fidelity, another factor potentially restricting gene
343 flow between breeding stocks is spatiotemporal variation in migratory behavior, that is differences in
344 time of departure and/or arrival between breeding stocks, that use different areas for foraging as
345 postulated by Kershaw et al. (2017). These variations of timing lead to some individuals arriving later
346 than others in their breeding and/or feeding areas, restricting possibilities of encounters between
347 breeding stocks that could facilitate random mating behavior (Elwen et al. 2014; Rosenbaum et al. 2014).
348 In the North Atlantic, this effect is particularly detectable in two breeding stocks of the West Atlantic.
349 Here, individuals who prefer to migrate to West Atlantic feeding areas such as the Gulf of Maine or the
350 Gulf of St. Lawrence (mainly the Greater Antilles breeding stock) seem to initiate their northward
351 migration significantly earlier than the Guadeloupe breeding stock, which preferably feeds towards
352 Iceland and Norway. Their southward migration is similarly shifted, decreasing the possibility of mating
353 encounters (Stevick et al. 2003). In Icelandic waters, the onset of the autumn migration seems to stretch
354 over several months, and considerable numbers of humpback whales are regularly observed in every
355 month of the year (Magnúsdóttir and Lim 2019; Pike et al. 2009; Víkingsson et al. 2015)

356 Lastly, although widely acknowledged as a generalist species regarding its diet, differences in prey
357 preference could be another factor influencing population structure. Foote et al. (2016) used genome
358 re-sequencing in killer whales to reveal evolutionary trends that lead to the adaptation to several different
359 ecotypes, which coexist in the same ocean basin. The authors postulated cultural transmission of
360 behavior (e.g., specialization on certain prey items) to be a main driver of genetic structure. While this

effect is by far not as well developed in humpback whales, it highlights the importance of inheriting behavior from closely related individuals that might play a role in many cetacean species. Additionally, reports of cultural transmission of specific hunting techniques such as lobtail feeding highlights that humpback whales indeed display cultural inheritance (Allen et al. 2013). In NAHW, it was reported several times that the diet of humpbacks feeding in the eastern feeding grounds as well as Newfoundland consists of krill and capelin, while Gulf of Maine humpback whales mainly consume sand lance and herring (Johnson and Davoren 2020; Sigurjónsson and Víkingsson 1997). Given the exceptional inter-annual resighting rate of over 70% in the Gulf of Maine, prey preferences could indeed favor population structure in humpback whales to some extent (Clapham et al. 2011). It is uncertain if and in what dimension the dietary choices may have an impact on population structure and prey preferences could well be solely a result of site-fidelity. Future analyses of stable isotopes combined with genetic data could detect correlations and give insights on its potential effect.

4.2 Genealogy

As described above, cultural transmission may play a significant role in restricting gene flow between breeding stocks through conservation of migratory destinations, timing and perhaps prey preferences. Many of those factors are most likely maternally inherited or transmitted through closely related individuals (Barendse et al. 2013). Therefore, we performed genealogy analysis based on our multilocus genotypes to estimate level of relatedness within our sample set. The high level of relatedness (18 full-siblings including four possible parent-offspring relationships) detected in our dataset give a suggestion that cultural transmission may play a significant role in humpback whale population structure. A possible explanation for the relatively high number of related pairings could be, that most of our samples were taken in two restricted areas which are also close to one another (Skjálfandalói and Eyjafjörður). Although, this still supports the significance of cultural transmission of habitat and diet preferences in humpback whales as closely related individuals may use the same foraging areas over large temporal scales. At the Gulf of Maine feeding ground, high numbers of stable associations between two or three individuals were recorded, with about one third occurring in consecutive years. The pairings in their study mainly consisted of females, further supporting differences in male behavior beyond their

388 breeding season (Weinrich 1991). A follow up study investigating mtDNA of the pairings revealed a
389 significant trend towards stable social association between individuals sharing the same haplotype. This
390 indicates stronger social association within one maternal lineage, potentially displaying cultural
391 transmission e.g. regarding prey type or hunting techniques, enhancing their overall metabolic fitness
392 (Weinrich et al. 2006). Here, we find that most detected pairings were sampled in the same year,
393 suggesting that close related individuals simultaneously use the same foraging areas. Interestingly, all
394 but one pairing (due to unknown sex of one individual) consisted exclusively of individuals of the same
395 sex but there is no observational data available to resolve social behavior among those individuals.
396 Further, our analysis detected a much lower number of pairings in one genetic cluster compared to the
397 other two. This could be an indication on the overall population size of each genetic cluster as the chance
398 of sampling close related individuals within one genetic unit decreases with a higher population size and
399 vice versa. Thus, cluster 2 could potentially represent a unit with a high number of individuals while the
400 effective size of cluster 1 and 3 might be smaller.

401 Altogether, the relatively high number of related individuals in our samples implies that close related
402 individuals (full-siblings and parent-offspring pairs) use the same area for foraging. This effect is
403 especially well detected within one year. This further supports the hypothesis of high site-fidelity in
404 NAHW and strengthens the implication of the restrictive effect this may have on gene flow between
405 genetic clusters.

406 4.3 Implications for future research

407 Humpback whales experienced an extraordinary comeback from their near depletion during 19th and
408 20th century in almost all ocean basins (Noad et al. 2019; Pallin et al. 2019; Ruegg et al. 2013; Zerbini,
409 et al. 2010). Its flexibility and adaptability regarding prey and habitat use potentially give humpback
410 whales advantages compared to some other baleen whales in adapting to a changing environment
411 (Fleming et al. 2016; Ramp et al. 2015). Nevertheless, all cetaceans face multiple increasingly severe
412 environmental pressures, almost all of which are directly or indirectly anthropogenically induced.
413 Several of those stressors have been shown to impact individual humpback whales such as entanglement,
414 ship strikes, noise pollution, contaminant intake or warming sea temperatures (Basran et al. 2019;

Fournet et al. 2018; Redfern et al. 2013; Ryan et al. 2013; Simmonds and Eliott 2009). Therefore, a regular monitoring of the populations and their potential threats need to be maintained. Conservation management of highly migratory species such as the humpback whale faces many organizational challenges since the species crosses multiple national borders with varying degrees of wildlife protection measures. It is therefore essential to identify the genetic structure of a population and locate diversity hot spots as priority protection areas. In the field of conservation genetics, ensuring the maintenance of genetic variability strongly increases adaptive potential to fast changing environments (Razgour et al. 2019). Here, we show that the Icelandic feeding ground represents an important habitat for three differentiated genetic units of NAHW. We therefore encourage further characterization of these groups (e.g., their individual breeding ground origins) as an important contribution to understanding the population structure of NAHW.

4.4 Conclusion

In this study, we present fine scale population structure of Icelandic North Atlantic humpback whales using classical population genetic approaches and analysis of relatedness as an indicator for culturally inherited site-fidelity. Our results suggest a complex population structure, showing a minimum of three differentiated units of humpback whales that coexist in the highly productive waters around Iceland during their feeding period. These results are generally consistent with photo-identification data suggesting a separation of NAHW into several breeding stock units (Chosson-P et al. 2015). High levels of relatedness among sampled individuals suggest high site-fidelity regarding their feeding areas. We therefore postulate that site-fidelity to both breeding and feeding grounds could be a major driver of fine-scale population differentiation but may be facilitated by other factors as well. Our findings characterize Icelandic waters as important habitats for genetically diverse groups of humpback whales. Future studies should focus on retaining reduced representation next generation sequencing approaches to verify our results and get further, more detailed insights into the ecological characteristics of those units.

Acknowledgements

We gratefully acknowledge the staff, students, and numerous volunteers at the University of Iceland's Húsavík Research Center and Marine and Freshwater Research Institute that have been involved in stranding response, sampling effort and storage of samples throughout the last three decades. This study was partly funded by the Swiss Whale Society's SWG Discovery Scholarship 2020.

Authorship contributions

Marc-Alexander Gose: Conceptualization; Data curation; Formal analysis; Investigation; Project administration; Validation; Visualization; Writing – original draft (lead); Writing – review & editing **Maximilian Schweinsberg**: Conceptualization; Data curation; Formal analysis; Project administration; Validation; Investigation; Writing – original draft (lead); Writing – review & editing **Christophe Pampoulie**: Conceptualization; Supervision; Resources; Validation; Writing -original draft (supporting), Writing – review & editing **Sverrir Daníel Halldórsson**: Resources; Writing – review & editing **Marianne H. Rasmussen**: Funding acquisition; Project administration, Investigation Resources; Writing – review & editing **Gísli Víkingsson**: Project administration; Resources; Supervision; Writing – review & editing **Ralph Tollrian**: Conceptualization; Funding acquisition; Project administration; Investigation; Supervision

6. References

Allen, J., M. Weinrich, W. Hoppitt, and L. Rendell. (2013). Network-Based Diffusion Analysis Reveals Cultural Transmission of Lobtail Feeding in Humpback Whales. *Science* 340 (6131): 485–88. https://doi.org/10.1126/science.1231976.

Baker, C. S., A. Perry, J. L. Bannister, M. T. Weinrich, R. B. Abernethy, J. Calambokidis, J. Lien, R. H. Lambertsen, J. U. Ramírez, and O. Vasquez. (1993). Abundant Mitochondrial DNA Variation and World-Wide Population Structure in Humpback Whales. *Proceedings of the National Academy of Sciences* 90 (17): 8239–43. https://doi.org/10.1073/pnas.90.17.8239.

Baker, C. S., D. Steel, J. Calambokidis, E. Falcone, U. González-Peral, J. Barlow, A. M. Burdin, et al. (2013). Strong Maternal Fidelity and Natal Philopatry Shape Genetic Structure in North Pacific Humpback Whales. *Marine Ecology Progress Series* 494 (December): 291–306. https://doi.org/10.3354/meps10508.

Bandelt, H. J., P. Forster, and A. Röhl. (1999). Median-Joining Networks for Inferring Intraspecific Phylogenies. *Molecular Biology and Evolution* 16 (1): 37–48. https://doi.org/10.1093/oxfordjournals.molbev.a026036.

Barendse, J., P. B. Best, I. Carvalho, and C. Pomilla. (2013). Mother Knows Best: Occurrence and Associations of Resighted Humpback Whales Suggest Maternally Derived Fidelity to a Southern Hemisphere Coastal Feeding Ground. *PLOS ONE* 8 (12): e81238. https://doi.org/10.1371/journal.pone.0081238.

Basran, C. J., C. G. Bertulli, A. Cecchetti, M. H. Rasmussen, M. Whittaker, and J. Robbins. (2019). First Estimates of Entanglement Rate of Humpback Whales Megaptera Novaeangliae Observed in Coastal Icelandic Waters. *Endangered Species Research* 38 (February): 67–77. https://doi.org/10.3354/esr00936.

Berrow, S. D., N. Massett, P. Whooley, B. V. M. Jann, P. Lopez-Suarez, P. T. Stevick, and F. W. Wenzel. (2021). Resightings of Humpback Whales (Megaptera Novaeangliae) from Ireland to a Known Breeding Ground: Cabo Verde, West Africa. *Aquatic Mammals* 47 (1): 63–71. https://doi.org/10.1578/AM.47.1.2021.63.

Berube, M., and P. Palsbøll. (1996). Identification of Sex in Cetaceans by Multiplexing with Three ZFX and ZFY Specific Primers. *Molecular Ecology* 5 (2): 283–87. https://doi.org/10.1046/j.1365-294X.1996.00072.x.

Bérubé, M., H. JØrgensen, R. McEwing, and P. J. Palsbøll. (2000). Polymorphic Di-Nucleotide Microsatellite Loci Isolated from the Humpback Whale, Megaptera Novaeangliae. *Molecular Ecology* 9 (12): 2181–83. https://doi.org/10.1046/j.1365-294X.2000.105315.x.

Brown, J. L. (2019). Abundance and Distribution Shifts of Humpback Whales (Megaptera Novaeangliae) in Ísafjarðardjúp. Thesis. https://skemman.is/handle/1946/34020.

Chero, G., R. Pradel, S. Derville, C. Bonneville, O. Gimenez, and C. Garrigue. (2020). Reproductive Capacity of an Endangered and Recovering Population of Humpback Whales in the Southern Hemisphere. *Marine Ecology Progress Series* 643 (June): 219–27. https://doi.org/10.3354/meps13329.

Chosson-P V., O. Shears, M. H. Rasmussen, C. G. Bertulli, P. Stevick, S. Rickert, C. Pampoulie, J. Sigurjónsson, G. A. Víkinsson (2015). Long-Distance Migration Pattern of Humpback Whale (Megaptera Novaeangliae) from Icelandic Waters. Society for Marine Mammalogy's 21st Biennial Conference. 2015.

Christiansen, F., C. G. Bertulli, M. H. Rasmussen, and D. Lusseau. (2015). Estimating Cumulative Exposure of Wildlife to Non-Lethal Disturbance Using Spatially Explicit Capture–Recapture Models. *The Journal of Wildlife Management* 79 (2): 311–24. https://doi.org/10.1002/jwmg.836.

Christiansen, F., G. A. Víkingsson, M. H. Rasmussen, and D. Lusseau. (2014). Female Body Condition Affects Foetal Growth in a Capital Breeding Mysticete. *Functional Ecology* 28 (3): 579–88. https://doi.org/10.1111/1365-2435.12200.

Clapham, P. J. (1996). The Social and Reproductive Biology of Humpback Whales: An Ecological Perspective. *Mammal Review* 26 (1): 27–49. https://doi.org/10.1111/j.1365-2907.1996.tb00145.x.

Clapham, P. J., L. S. Baraff, C. A. Carlson, M. A. Christian, D. K. Mattila, C. A. Mayo, M. A. Murphy, and S. Pittman. (2011). Seasonal Occurrence and Annual Return of Humpback Whales, Megaptera Novaeangliae, in the Southern Gulf of Maine. *Canadian Journal of Zoology*, February. https://doi.org/10.1139/z93-063.

Cooke, J. (2018). IUCN Red List of Threatened Species: Megaptera Novaeangliae. *IUCN Red List of Threatened Species*, March. https://www.iucnredlist.org/en.

Corkeron, P. J., and R. C. Connor. (1999). Why Do Baleen Whales Migrate?. *Marine Mammal Science* 15 (4): 1228–45. https://doi.org/10.1111/j.1748-7692.1999.tb00887.x.

Craig, A. S., and L. M. Herman. (1997). Sex Differences in Site Fidelity and Migration of Humpback Whales (Megaptera Novaeangliae) to the Hawaiian Islands. *Canadian Journal of Zoology* 75 (11): 1923–33. https://doi.org/10.1139/z97-822.

Cypriano-Souza, A. L., G. P. Fernández, C. A. V. Lima-Rosa, M. H. Engel, and S. L. Bonatto. (2010). Microsatellite Genetic Characterization of the Humpback Whale (Megaptera Novaeangliae) Breeding Ground off Brazil (Breeding Stock A). *The Journal of Heredity* 101 (2): 189–200. https://doi.org/10.1093/jhered/esp097.

Earl, D. A., and B. M. vonHoldt. (2012). STRUCTURE HARVESTER: A Website and Program for Visualizing STRUCTURE Output and Implementing the Evanno Method. *Conservation Genetics Resources* 4 (2): 359–61. https://doi.org/10.1007/s12686-011-9548-7.

Elwen, S. H., N. Tonachella, J. Barendse, T. Collins, P. B. Best, H. C. Rosenbaum, R. H. Leeney, and T. Gridley. (2014). Humpback Whales off Namibia: Occurrence, Seasonality, and a Regional Comparison of Photographic Catalogs and Scarring. *Journal of Mammalogy* 95 (5): 1064–76. https://doi.org/10.1644/14-MAMM-A-108.

Evanno, G., S. Regnaut, and J. Goudet. (2005). Detecting the Number of Clusters of Individuals Using the Software Structure: A Simulation Study. *Molecular Ecology* 14 (8): 2611–20. https://doi.org/10.1111/j.1365-294X.2005.02553.x.

Excoffier, L., and H. E. L. Lischer. (2010). Arlequin Suite Ver 3.5: A New Series of Programs to Perform Population Genetics Analyses under Linux and Windows. *Molecular Ecology Resources* 10 (3): 564–67. https://doi.org/10.1111/j.1755-0998.2010.02847.x.

Fleming, A. H., C. T. Clark, J. Calambokidis, and J. Barlow. (2016). Humpback Whale Diets Respond to Variance in Ocean Climate and Ecosystem Conditions in the California Current. *Global Change Biology* 22 (3): 1214–24. https://doi.org/10.1111/gcb.13171.

Foote, A. D., N. Vijay, M. C. Ávila-Arcos, R. W. Baird, J. W. Durban, M. Fumagalli, R. A. Gibbs, et al. (2016). Genome-Culture Coevolution Promotes Rapid Divergence of Killer Whale Ecotypes. *Nature Communications* 7 (1): 11693. https://doi.org/10.1038/ncomms11693.

Fournet, M. E. H., L. P. Matthews, C. M. Gabriele, S. Haver, D. K. Mellinger, and H. Klinck. (2018). Humpback Whales Megaptera Novaeangliae Alter Calling Behavior in Response to Natural Sounds and Vessel Noise. *Marine Ecology Progress Series* 607 (December): 251–68. https://doi.org/10.3354/meps12784.

Genoves, R. C., P. F. Fruet, S. Botta, L. B. Beheregaray, L. M. Möller, and E. R. Secchi. (2020). Fine-Scale Genetic Structure in Lahille's Bottlenose Dolphins (Tursiops Truncatus Gephyreus) Is Associated with Social Structure and Feeding Ecology. *Marine Biology* 167 (3): 34. https://doi.org/10.1007/s00227-019-3638-6.

Goudet, J. (2005). Hierfstat, a Package for r to Compute and Test Hierarchical F-Statistics. *Molecular Ecology Notes* 5 (1): 184–86. https://doi.org/10.1111/j.1471-8286.2004.00828.x.

Heide-Jørgensen, M. P., K. L. Laidre, R. G. Hansen, M. L. Burt, M. Simon, D. L. Borchers, J. Hansen, et al. (2012). Rate of Increase and Current Abundance of Humpback Whales in West Greenland. *Journal of Cetacean Research and Management* 12(1): 1–14, 2012.

Horton, T. W., N. Hauser, A. N. Zerbini, M. P. Francis, M. L. Domeier, A. Andriolo, D. P. Costa, et al. (2017). Route Fidelity during Marine Megafauna Migration. *Frontiers in Marine Science* 4:422. https://doi.org/10.3389/fmars.2017.00422.

Horton T. W., Holdaway R. N., Zerbini A. N., Hauser N., Garrigue C., Andriolo A., and Clapham P. J. (2011). Straight as an Arrow: Humpback Whales Swim Constant Course Tracks during Long-Distance Migration. *Biology Letters* 7 (5): 674–79. https://doi.org/10.1098/rsbl.2011.0279.

Horton, T. W., A. N. Zerbini, A. Andriolo, D. Danilewicz, and F. Sucunza. (2020). Multi-Decadal Humpback Whale Migratory Route Fidelity Despite Oceanographic and Geomagnetic Change. *Frontiers in Marine Science* 7:414. https://doi.org/10.3389/fmars.2020.00414.

Hoyt, E. (2018). Tourism. In *Encyclopedia of Marine Mammals (Third Edition)*, edited by B. Würsig, J. G. M. Thewissen, and K. M. Kovacs, 1010–14. Academic Press. https://doi.org/10.1016/B978-0-12-804327-1.00262-4.

Ihaka, R., and R. Gentleman. (1996). R: A Language for Data Analysis and Graphics. *Journal of Computational and Graphical Statistics* 5 (3): 299–314. https://doi.org/10.1080/10618600.1996.10474713.

Jackson, J. A., D. J. Steel, P. Beerli, B. C. Congdon, C. Olavarría, M. S. Leslie, C. Pomilla, H. Rosenbaum, and C. S. Baker. (2014). Global Diversity and Oceanic Divergence of Humpback Whales (Megaptera Novaeangliae). *Proceedings. Biological Sciences* 281 (1786). https://doi.org/10.1098/rspb.2013.3222.

Jakobsson, M., and N. A. Rosenberg. (2007). CLUMPP: A Cluster Matching and Permutation Program for Dealing with Label Switching and Multimodality in Analysis of Population Structure. *Bioinformatics* 23 (14): 1801–6. https://doi.org/10.1093/bioinformatics/btm233.

Jann, B, J. Allen, M. Carrillo, S. Hanquet, S. K. Katona, A. R. Martin, R. R. Reeves, R. Seton, P. T. Stevick, and F. W. Wenzel. (2003). Migration of a Humpback Whale (Megaptera Novaeangliae) between the Cape Verde Islands and Iceland, *Journal of Cetacean Research and Management* 5(2):125–129.

Johnson, K. F., and G. K. Davoren. (2020). Distributional Patterns of Humpback Whales (Megaptera Novaeangliae) along the Newfoundland East Coast Reflect Their Main Prey, Capelin (Mallotus Villosus). *Marine Mammal Science* 2021;37:80–97. https://doi.org/10.1111/mms.12730.

Jombart, T. (2008). Adegenet: A R Package for the Multivariate Analysis of Genetic Markers. *Bioinformatics* 24 (11): 1403–5. https://doi.org/10.1093/bioinformatics/btn129.

Jombart, T., S. Devillard, and F. Balloux. (2010). Discriminant Analysis of Principal Components: A New Method for the Analysis of Genetically Structured Populations. *BMC Genetics* 11 (1): 94. https://doi.org/10.1186/1471-2156-11-94.

Jones, O. R., and J. Wang. (2010). COLONY: A Program for Parentage and Sibship Inference from Multilocus Genotype Data. *Molecular Ecology Resources* 10 (3): 551–55. https://doi.org/10.1111/j.1755-0998.2009.02787.x.

Kalinowski, S. T., M. L. Taper, and T. 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.

Kalinowski, S. T., A. P. Wagner, and M. L. Taper. (2006). Ml-Relate: A Computer Program for Maximum Likelihood Estimation of Relatedness and Relationship. *Molecular Ecology Notes* 6 (2): 576–79. https://doi.org/10.1111/j.1471-8286.2006.01256.x.

Kearse, M., R. Moir, A. Wilson, St. Stones-Havas, M. Cheung, S. Sturrock, S. Buxton, et al. (2012). Geneious Basic: An Integrated and Extendable Desktop Software Platform for the Organization and Analysis of Sequence Data. *Bioinformatics* 28 (12): 1647–49. https://doi.org/10.1093/bioinformatics/bts199.

Kershaw, F., I. Carvalho, J. Loo, C. Pomilla, P. B. Best, K. P. Findlay, S. Cerchio, et al. (2017). Multiple Processes Drive Genetic Structure of Humpback Whale (Megaptera Novaeangliae) Populations across Spatial Scales. *Molecular Ecology* 26 (4): 977–94. https://doi.org/10.1111/mec.13943.

Larsen, A. H., Sigurjónsson J., Øien N., Vikingsson G., and Palsbøll P. (1996). Populations Genetic Analysis of Nuclear and Mitochondrial Loci in Skin Biopsies Collected from Central and Northeastern North Atlantic Humpback Whales (Megaptera Novaeangliae): Population

Identity and Migratory Destinations. *Proceedings of the Royal Society of London. Series B: Biological Sciences* 263 (1376): 1611–18. https://doi.org/10.1098/rspb.1996.0236.

Lawson Handley, L. J., and N. Perrin. (2007). Advances in Our Understanding of Mammalian Sex-Biased Dispersal. *Molecular Ecology* 16 (8): 1559–78. https://doi.org/10.1111/j.1365-294X.2006.03152.x.

Leigh, J. W., and D. Bryant. (2015). Popart: Full-Feature Software for Haplotype Network Construction. *Methods in Ecology and Evolution* 6 (9): 1110–16. https://doi.org/10.1111/2041-210X.12410.

Lockyer, C. (1986). Body Fat Condition in Northeast Atlantic Fin Whales, Balaenoptera Physalus, and Its Relationship with Reproduction and Food Resource. *Canadian Journal of Fisheries and Aquatic Sciences* 43 (1): 142–47. https://doi.org/10.1139/f86-015.

Mackay, M., C. Bacon, L. Bouveret, S. Fossette, and P. Stevick. (2019). Humpback Whale (Megaptera Novaeangliae) Intra/Inter- Seasonal Exchanges Between Puerto Rico and the Southeastern Caribbean. *Animal Behavior and Cognition* 6 (April): 98–104. https://doi.org/10.26451/abc.06.02.02.2019.

Magnúsdóttir, E. E., and R. Lim. (2019). Subarctic Singers: Humpback Whale (Megaptera Novaeangliae) Song Structure and Progression from an Icelandic Feeding Ground during Winter. *PLOS ONE* 14 (1): e0210057. https://doi.org/10.1371/journal.pone.0210057.

Nei, M., and R. K. Chesser. (1983). Estimation of Fixation Indices and Gene Diversities. *Annals of Human Genetics* 47 (3): 253–59. https://doi.org/10.1111/j.1469-1809.1983.tb00993.x.

Noad, M. J., E. Kniest, and R. A. Dunlop. (2019). Boom to Bust? Implications for the Continued Rapid Growth of the Eastern Australian Humpback Whale Population despite Recovery. *Population Ecology* 61 (2): 198–209. https://doi.org/10.1002/1438-390X.1014.

Pallin L. J., Baker C. S., Steel D., Kellar N. M., Robbins J., Johnston D. W., Nowacek D. P., Read A. J., and Friedlaender A. S. (2019). High Pregnancy Rates in Humpback Whales (Megaptera Novaeangliae) around the Western Antarctic Peninsula, Evidence of a Rapidly Growing Population. *Royal Society Open Science* 5 (5): 180017. https://doi.org/10.1098/rsos.180017.

Palsbøll, P. J., M. Bérubé, A. H. Larsen, and H. Jørgensen. (1997). Primers for the Amplification of Tri- and Tetramer Microsatellite Loci in Baleen Whales. *Molecular Ecology* 6 (9): 893–95. https://doi.org/10.1046/j.1365-294X.1997.d01-214.x.

Palsbøll, P. J., P. J. Clapham, D. K. Mattila, F. Larsen, R. Sears, H. R. Siegismund, J. Sigurjónsson, O. Vasquez, and P. Arctander. (1995). Distribution of MtDNA Haplotypes in North Atlantic Humpback Whales: The Influence of Behaviour on Population Structure. *Marine Ecology Progress Series* 116 (1/3): 1–10.

Paradis, E. 2010. Pegas: An R Package for Population Genetics with an Integrated–Modular Approach. *Bioinformatics* 26 (3): 419–20. https://doi.org/10.1093/bioinformatics/btp696.

Pike, D. G., C. G. M. Paxton, T. Gunnlaugsson, and G. A. Víkingsson. (2009). Trends in the Distribution and Abundance of Cetaceans from Aerial Surveys in Icelandic Coastal Waters,

1986-2001. *NAMMCO Scientific Publications* 7 (September): 117–42. https://doi.org/10.7557/3.2710.

Pritchard, J. K., M. Stephens, and P. Donnelly. (2000). Inference of Population Structure Using Multilocus Genotype Data. *Genetics* 155 (2): 945–59.

Ramm, T. (2020). Hungry during Migration? Humpback Whale Movement from the Barents Sea to a Feeding Stopover in Northern Norway Revealed by Photo-ID Analysis, June. https://munin.uit.no/handle/10037/19109.

Ramp, C., J. Delarue, P. J. Palsbøll, R. Sears, and P. S. Hammond. (2015). Adapting to a Warmer Ocean—Seasonal Shift of Baleen Whale Movements over Three Decades. *PLOS ONE* 10 (3): e0121374. https://doi.org/10.1371/journal.pone.0121374.

Razgour, O., B. Forester, J. B. Taggart, M. Bekaert, J. Juste, C. Ibáñez, S. J. Puechmaille, R. Novella-Fernandez, A. Alberdi, and S. Manel. (2019). Considering Adaptive Genetic Variation in Climate Change Vulnerability Assessment Reduces Species Range Loss Projections. *Proceedings of the National Academy of Sciences* 116 (21): 10418–23. https://doi.org/10.1073/pnas.1820663116.

Redfern, J. V., M. F. Mckenna, T. J. Moore, J. Calambokidis, M. L. Deangelis, E. A. Becker, J. Barlow, K. A. Forney, P. C. Fiedler, and S. J. Chivers. (2013). Assessing the Risk of Ships Striking Large Whales in Marine Spatial Planning. *Conservation Biology* 27 (2): 292–302. https://doi.org/10.1111/cobi.12029.

Richard, G., O. V. Titova, I. D. Fedutin, D. Steel, I. G. Meschersky, M. Hautin, A. M. Burdin, E. Hoyt, O. A. Filatova, and J.-L. Jung. (2018). Cultural Transmission of Fine-Scale Fidelity to Feeding Sites May Shape Humpback Whale Genetic Diversity in Russian Pacific Waters. *Journal of Heredity* 109 (7): 724–34. https://doi.org/10.1093/jhered/esy033.

Rinaldi, C., R. Sears, P. T. Stevick, and C. Carlson. (2009). First Re-Sighting of a Humpback Whale between the French Lesser Antilles and the North Atlantic Feeding Grounds off Canada, Paper IWC/SC61/013 presented to the Scientific Committee of the International Whaling Commission, June 2009. 4 pp.

Rosenbaum, H. C., M. T. Weinrich, S. A. Stoleson, J. P. Gibbs, C. S. Baker, and R. DeSalle. (2002). The Effect of Differential Reproductive Success on Population Genetic Structure: Correlations of Life History With Matrilines in Humpback Whales of the Gulf of Maine. *Journal of Heredity* 93 (6): 389–99. https://doi.org/10.1093/jhered/93.6.389.

Rosenbaum, H. C., S. M. Maxwell, F. Kershaw, and B. Mate. (2014). Long-Range Movement of Humpback Whales and Their Overlap with Anthropogenic Activity in the South Atlantic Ocean. *Conservation Biology* 28 (2): 604–15. https://doi.org/10.1111/cobi.12225.

Rosenberg, N. A. (2004). Distruct: A Program for the Graphical Display of Population Structure. *Molecular Ecology Notes* 4 (1): 137–38. https://doi.org/10.1046/j.1471-8286.2003.00566.x.

RStudio Team, R. (2015). RStudio: Integrated Development for R. *RStudio, Inc., Boston, MA URL Http://Www. Rstudio. Com* 42: 14.

Ruegg, K., H. C. Rosenbaum, E. C. Anderson, M. Engel, A. Rothschild, C. S. Baker, and S. R. Palumbi. (2013). Long-Term Population Size of the North Atlantic Humpback Whale

within the Context of Worldwide Population Structure. *Conservation Genetics* 14 (1): 103–14. https://doi.org/10.1007/s10592-012-0432-0.

Ryan, C., B. McHugh, B. Boyle, E. McGovern, M. Bérubé, P. Lopez-Suárez, C. Elfes, et al. (2013). Levels of Persistent Organic Pollutants in Eastern North Atlantic Humpback Whales. *Endangered Species Research* 22 (3): 213–23. https://doi.org/10.3354/esr00545.

Ryman, N., and S. Palm. (2006). POWSIM: A Computer Program for Assessing Statistical Power When Testing for Genetic Differentiation. *Molecular Ecology Notes* 6 (3): 600–602. https://doi.org/10.1111/j.1471-8286.2006.01378.x.

Santora, J. A., N. J. Mantua, I. D. Schroeder, J. C. Field, E. L. Hazen, S. J. Bograd, W. J. Sydeman, et al. (2020). Habitat Compression and Ecosystem Shifts as Potential Links between Marine Heatwave and Record Whale Entanglements. *Nature Communications* 11 (1): 536. https://doi.org/10.1038/s41467-019-14215-w.

Schlötterer, C., and T. Wiehe. (1999). Microsatellites, a Neutral Marker to Infer Selective Sweeps. *Microsatellites*, 238–48.

Schwarz, G. (1978). Estimating the Dimension of a Model. *The Annals of Statistics* 6 (2): 461–64. https://doi.org/10.1214/aos/1176344136.

Sigurjónsson, J., and G. Víkingsson. (1997). Seasonal Abundance of and Estimated Food Consumption by Cetaceans in Icelandic and Adjacent Waters. *Journal of Northwest Atlantic Fishery Science* 22 (December): 271–87. https://doi.org/10.2960/J.v22.a20.

Simmonds, M. P., and W. J. Eliott. (2009). Climate Change and Cetaceans: Concerns and Recent Developments. *Journal of the Marine Biological Association of the United Kingdom* 89 (1): 203–10. https://doi.org/10.1017/S0025315408003196

Stevick, P. T., J. Allen, P. J. Clapham, S. K. Katona, F. Larsen, J. Lien, D. K. Mattila, et al. (2006). Population Spatial Structuring on the Feeding Grounds in North Atlantic Humpback Whales (Megaptera Novaeangliae). *Journal of Zoology* 270 (2): 244–55. https://doi.org/10.1111/j.1469-7998.2006.00128.x.

Stevick, P. T., J. Allen, M. Bérubé, P. J. Clapham, S. K. Katona, F. Larsen, J. Lien, et al. (2003). Segregation of Migration by Feeding Ground Origin in North Atlantic Humpback Whales (Megaptera Novaeangliae). *Journal of Zoology* 259 (3): 231–37. https://doi.org/10.1017/S0952836902003151.

Stevick, P. T., S. D. Berrow, M. Bérubé, L. Bouveret, F. Broms, B. Jann, A. Kennedy, et al. (2016). There and Back Again: Multiple and Return Exchange of Humpback Whales between Breeding Habitats Separated by an Ocean Basin. *Journal of the Marine Biological Association of the United Kingdom* 96 (4): 885–90. https://doi.org/10.1017/S0025315416000321.

Valsecchi, E., and W. Amos. (1996). Microsatellite Markers for the Study of Cetacean Populations. *Molecular Ecology* 5 (1): 151–56. https://doi.org/10.1111/j.1365-294X.1996.tb00301.x.

Víkingsson, G. A. (1995). Body Condition of Fin Whales during Summer off Iceland. In *Developments in Marine Biology*, edited by A. S. Blix, L. Walløe, and Ø. Ulltang, 4:361–69.

737 Whales, Seals, Fish and Man. Elsevier Science. https://doi.org/10.1016/S0163-
738 6995(06)80037-5.

739 Víkingsson, G., D. Pike, A. Schleimer, H. Valdimarsson, T. Gunnlaugsson, T. Silva, B.
740 Elvarsson, et al. (2015). Distribution, Abundance, and Feeding Ecology of Baleen Whales in
741 Icelandic Waters: Have Recent Environmental Changes Had an Effect? *Frontiers in Ecology*
742 *and Evolution* 3 (February). https://doi.org/10.3389/fevo.2015.00006.

743 Waldick, R. C., S. Kraus, M. Brown, and B. N. White. (2002). Evaluating the Effects of
744 Historic Bottleneck Events: An Assessment of Microsatellite Variability in the Endangered,
745 North Atlantic Right Whale. *Molecular Ecology* 11 (11): 2241–49.
746 https://doi.org/10.1046/j.1365-294X.2002.01605.x.

747 Weinrich, M. T. (1991). Stable Social Associations among Humpback Whales (Megaptera
748 Novaeangliae) in the Southern Gulf of Maine. *Canadian Journal of Zoology* 69 (12): 3012–
749 19. https://doi.org/10.1139/z91-425.

750 Weinrich, M. T., H. Rosenbaum, C. S. Baker, A. L. Blackmer, and H. Whitehead. (2006). The
751 Influence of Maternal Lineages on Social Affiliations among Humpback Whales (Megaptera
752 Novaeangliae) on Their Feeding Grounds in the Southern Gulf of Maine. *Journal of Heredity*
753 97 (3): 226–34. https://doi.org/10.1093/jhered/esj018.

754 Weir, B. S., and C. C. Cockerham. (1984). Estimating F-Statistics for the Analysis of
755 Population Structure. *Evolution* 38 (6): 1358–70. https://doi.org/10.1111/j.1558-
756 5646.1984.tb05657.x.

757 Wenzel, F. W., J. Allen, S. Berrow, C. J. Hazevoet, B. Jann, R. E. Seton, L. Steiner, P.
758 Stevick, P. López Suárez, and P. Whooley. (2009). Current Knowledge on the Distribution
759 and Relative Abundance of Humpback Whales (Megaptera Novaeangliae) off the Cape Verde
760 Islands, Eastern North Atlantic. *Aquatic Mammals* 35 (4): 502–10.
761 https://doi.org/10.1578/AM.35.4.2009.502.

762 Wiszniewski, J., L. B. Beheregaray, S. J. Allen, and L. M. Möller. (2010). Environmental and
763 Social Influences on the Genetic Structure of Bottlenose Dolphins (Tursiops Aduncus) in
764 Southeastern Australia. *Conservation Genetics* 4 (11): 1405–19.
765 https://doi.org/10.1007/s10592-009-9968-z.

766 Zerbini, A. N., P. J. Clapham, and P. R. Wade. (2010). Assessing Plausible Rates of
767 Population Growth in Humpback Whales from Life-History Data. *Marine Biology* 157 (6):
768 1225–36. https://doi.org/10.1007/s00227-010-1403-y.

769

7. Supplementary Material

Table S1: PCR amplification programs for most of the microsatellite primers.

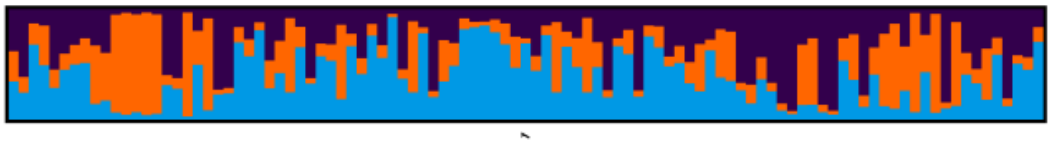
Step	Temperature [°C]	Duration [s]	Repeats
Initial denaturation	94	180	1
Denaturation	94	30	
Annealing	Specific (see Table 1 in text)	30	30 or 35
Elongation	72	30 to 45	
Final elongation	72	600 or 2100	1

Table S1: Details on 18 full-sibling or parent-offspring parings

ID	Pairing	Prob.	Cluster	sex	mt haplotype	Year of sampling
S045		1	1	w	X	2011
S049	PO/FS	1	1	w	X	2011
S042		1	3	m	VII	2011
S050	PO/FS	1	3	m	n.a.	2011
S065		1	1	w	VIII	2009
S066	PO/FS	1	1	w	VIII	2009
S077		1	3	w	n.a.	2010
S079	PO/FS	1	3	w	II	2010
S020		1	3	m	X	before 2010
S021	FS	1	3	m	X	before 2010

S038		1	1	m	X	2011
S041	FS	1	1	m	X	2011
S040		1	3	m	X	2011
S087	FS	1	3	m	X	2012
S041		1	1	m	X	2011
S051	FS	1	1	m	X	2011
S043		1	3	m	X	2011
S046	FS	1	3	m	X	2011
S044		1	1	m	VI	2011
S058	FS	1	1	m	VI	2011
S048		1	3	w	X	2011
S098	FS	1	3	w	X	2013
S021		0.9	3	m	X	before 2010
S028	FS	0.9	2	m	VI	before 2010
S038		0.9	1	m	X	2011
S051	FS	0.9	1	m	X	2011
S039		0.9	1	m	VI	2011
S044	FS	0.9	1	m	VI	2011
S039		0.9	1	m	VI	2011
S058	FS	0.9	1	m	VI	2014
S057		0.9	3	w	X	2014
S088	FS	0.9	3	w	X	2012
S063		0.9	2	m	n.a.	2008
S083	FS	0.9	2	m	IV	2012
S085		0.9	2	w	VII	2012
S086	FS	0.9	2	n.a.	n.a.	2012

775

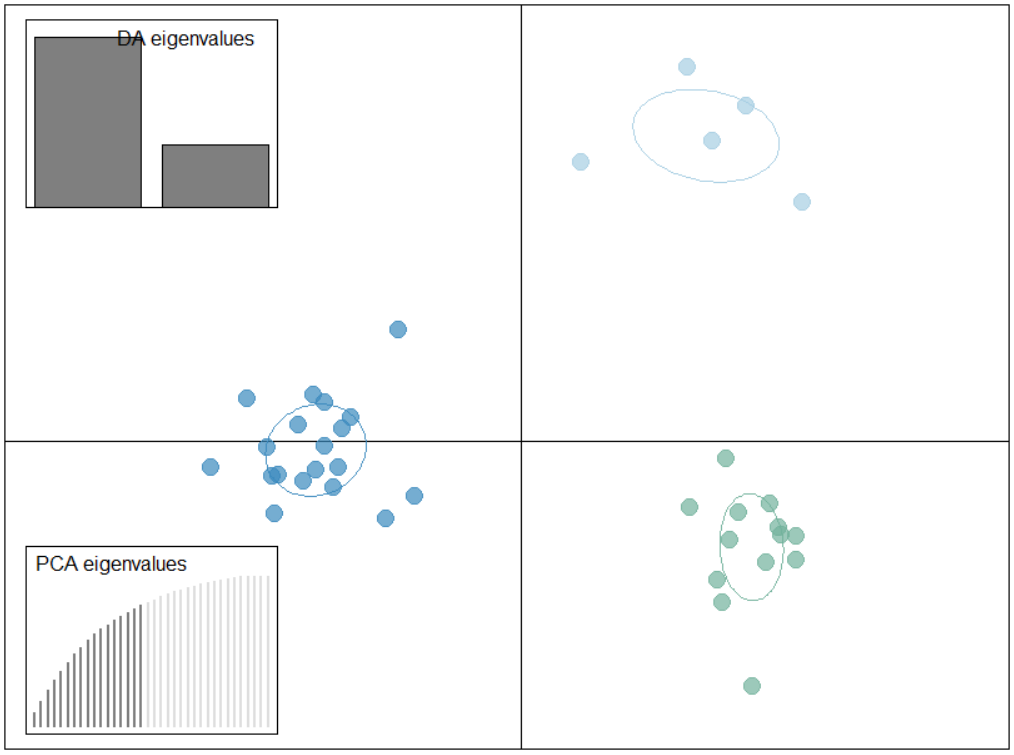


776

Figure S1: Membership assignment plot as inferred by cluster assignment in STRUCTURE.

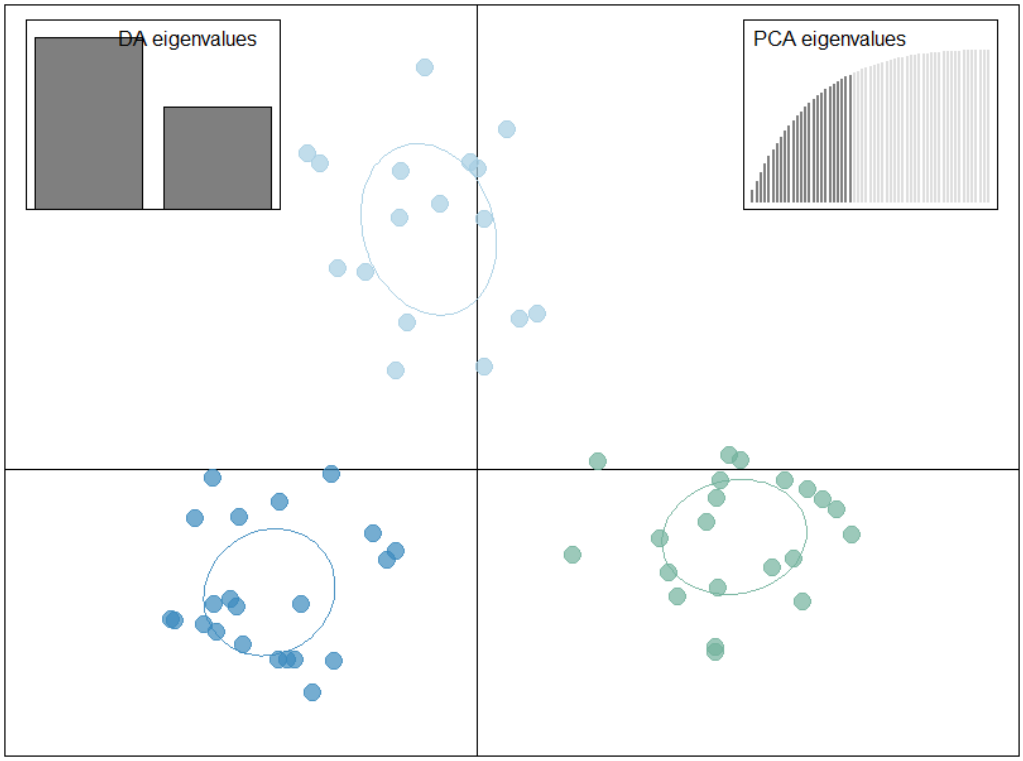
777

778



779

780 **Figure S2: DAPC for all female individuals**



781

782 **Figure S3: DAPC for all male individuals**