Characterization of Population Structure in Northern Bottlenose Whales (Hyperoodon ampullatus) in the North Atlantic from an Analysis of Microsatellite Data

Anthony Einfeldt, Laura Joan Feyrer, Ian Paterson, Steven Ferguson, Patrick Miller, Evelien de Greef and Paul Bentzen

Science Branch
Fisheries and Oceans Canada
Maritimes Region
Dartmouth, Nova Scotia, B2Y 4A2, Canada

2022

Canadian Technical Report of
Fisheries and Aquatic Sciences 3467

Fisheries and Oceans Canada | Pêches et Océans Canada

Canada

Canadian Technical Report of Fisheries and Aquatic Sciences

Technical reports contain scientific and technical information that contributes to existing knowledge but which is not normally appropriate for primary literature. Technical reports are directed primarily toward a worldwide audience and have an international distribution. No restriction is placed on subject matter and the series reflects the broad interests and policies of Fisheries and Oceans Canada, namely, fisheries and aquatic sciences.

Technical reports may be cited as full publications. The correct citation appears above the abstract of each report. Each report is abstracted in the data base *Aquatic Sciences and Fisheries Abstracts*.

Technical reports are produced regionally but are numbered nationally. Requests for individual reports will be filled by the issuing establishment listed on the front cover and title page.

Numbers 1-456 in this series were issued as Technical Reports of the Fisheries Research Board of Canada. Numbers 457-714 were issued as Department of the Environment, Fisheries and Marine Service, Research and Development Directorate Technical Reports. Numbers 715-924 were issued as Department of Fisheries and Environment, Fisheries and Marine Service Technical Reports. The current series name was changed with report number 925.

Rapport technique canadien des sciences halieutiques et aquatiques

Les rapports techniques contiennent des renseignements scientifiques et techniques qui constituent une contribution aux connaissances actuelles, mais qui ne sont pas normalement appropriés pour la publication dans un journal scientifique. Les rapports techniques sont destinés essentiellement à un public international et ils sont distribués à cet échelon. II n'y a aucune restriction quant au sujet; de fait, la série reflète la vaste gamme des intérêts et des politiques de Pêches et Océans Canada, c'est-à-dire les sciences halieutiques et aquatiques.

Les rapports techniques peuvent être cités comme des publications à part entière. Le titre exact figure au-dessus du résumé de chaque rapport. Les rapports techniques sont résumés dans la base de données *Résumés des sciences aquatiques et halieutiques.*

Les rapports techniques sont produits à l'échelon régional, mais numérotés à l'échelon national. Les demandes de rapports seront satisfaites par l'établissement auteur dont le nom figure sur la couverture et la page du titre.

Les numéros 1 à 456 de cette série ont été publiés à titre de Rapports techniques de l'Office des recherches sur les pêcheries du Canada. Les numéros 457 à 714 sont parus à titre de Rapports techniques de la Direction générale de la recherche et du développement, Service des pêches et de la mer, ministère de l'Environnement. Les numéros 715 à 924 ont été publiés à titre de Rapports techniques du Service des pêches et de la mer, ministère des Pêches et de l'Environnement. Le nom actuel de la série a été établi lors de la parution du numéro 925.

Canadian Technical Report of

Fisheries and Aquatic Sciences 3467

2022

Characterization of Population Structure in Northern Bottlenose Whales (*Hyperoodon ampullatus*) in the North Atlantic from an Analysis of Microsatellite Data

by

Anthony Einfeldt[1,2], Laura Joan Feyrer[2,3], Ian Paterson[2], Steven H. Ferguson[4], Patrick Miller[5], Evelien de Greef[2], and Paul Bentzen[2]

Science Branch

Fisheries and Oceans Canada

Maritimes Region

Dartmouth, Nova Scotia, B2Y 4A2, Canada

[1] Balaena Institute of Cetacean Conservation Studies, Halifax, NS, Canada
[2] Department of Biology, Dalhousie University, Halifax, NS, Canada
[3] Bedford Institute of Oceanography, Fisheries and Oceans Canada, Dartmouth, NS, Canada
[4] Arctic Aquatic Research Division, Fisheries and Oceans Canada, Winnipeg, MB, Canada
[5] Sea Mammal Research Unit, Scottish Oceans Institute, University of St Andrews, St Andrews, Fife, UK

© Her Majesty the Queen in Right of Canada, 2022.
Cat. No. Fs97-6/3467E-PDF ISBN 978-0-660-42379-1 ISSN 1488-5379

Correct Citation for this publication:

Einfeldt, A., Feyrer, L., Paterson, I., Ferguson, S. H., Miller, P., de Greef, E. and Bentzen, P. 2022. Characterization of Population Structure in Northern Bottlenose Whales (*Hyperoodon ampullatus*) in the North Atlantic from an Analysis of Microsatellite Data. Can. Tech. Rep. Fish. Aquat. Sci. 3467: vii + 17 p.

Table of Contents

List of Tables

List of Figures

Abstract

Einfeldt, A., Feyrer, L., Paterson, I., Ferguson, S. H., Miller, P., de Greef, E. and Bentzen, P. 2022. Characterization of Population Structure in Northern Bottlenose Whales (*Hyperoodon ampullatus*) in the North Atlantic from an Analysis of Microsatellite Data. Can. Tech. Rep. Fish. Aquat. Sci. 3467: vii + 17 p.

The objective of this genetic analysis was to better understand northern bottlenose whale (NBW, *Hyperoodon ampullatus*) population structure and viability across the North Atlantic using microsatellite data from 60 contemporary specimens not previously available. Tissue samples were collected by biopsy from the Scotian Shelf, the Davis Strait – Labrador Sea, Jan Mayen near Iceland, and from a stranding in southern Newfoundland between 2014-2019. A total of 55 unique individuals were identified, expanding the current genetic dataset of NBW specimens to 218 individuals (1967-2019). Population structure analyses were consistent with site fidelity, as individuals resampled at different time periods were from the same geographic region. We detected a low number of unique microsatellite alleles in contemporary samples (2017-2019, n = 29) from the Davis Strait – Labrador Sea relative to historical samples (1971, n = 78), consistent with population reductions from harvesting and genetic drift. Population structure analyses supported three distinct genetic groups: the Scotian Shelf, the western North Atlantic (Davis Strait, Labrador Sea, West Iceland), and Jan Mayen (East Iceland). The differentiation of contemporary samples from Jan Mayen to other regions has not yet been reported, highlighting the importance of adequate size and distribution of samples for detecting subtle population structure in *NBW*.

Résumé

Einfeldt, A., Feyrer, L., Paterson, I., Ferguson, S. H., Miller, P., de Greef, E. and Bentzen, P. 2022. Characterization of Population Structure in Northern Bottlenose Whales (*Hyperoodon ampullatus*) in the North Atlantic from an Analysis of Microsatellite Data. Can. Tech. Rep. Fish. Aquat. Sci. 3467: vii + 17 p.

L'objectif de cette analyse génétique était de mieux comprendre la structure et la viabilité de la population de baleines à bec communes (*Hyperoodon ampullatus*) dans l'Atlantique Nord en utilisant des données microsatellitaires de 60 spécimens contemporains non disponibles auparavant. On a prélevé des échantillons de tissus par biopsie dans le plateau néo-écossais, dans le détroit de Davis, dans la mer du Labrador, au large de l'île Jan Mayen près de l'Islande et sur un échouage dans le sud de Terre-Neuve entre 2014 et 2019. On a identifié un total de 55 individus, ce qui étend l'ensemble actuel de données génétiques des spécimens de baleine à bec commune à 218 individus (de 1967 à 2019). Les analyses de la structure de la population étaient cohérentes avec la fidélité aux sites, car les individus qui ont fait l'objet d'un nouvel échantillonnage à différentes périodes provenaient de la même région géographique. Nous avons détecté un faible nombre d'allèles microsatellitaires uniques dans les échantillons contemporains (de 2017 à 2019, $n = 29$) du détroit de Davis et de la mer du Labrador par rapport aux anciens échantillons (1971, $n = 78$), ce qui correspond à des réductions de populations découlant des récoltes et de la dérive génétique. Les analyses de la structure de la population ont soutenu trois groupes génétiques distincts : le plateau néo-écossais, l'ouest de l'Atlantique Nord (détroit de Davis, mer du Labrador) et l'île Jan Mayen (est de l'Islande). La différenciation des échantillons contemporains de l'île Jan Mayen par rapport aux autres régions n'a pas encore été déclarée, ce qui souligne l'importance d'une taille et d'une répartition adéquates des échantillons aux fins de la détection d'une structure subtile de la population de baleines à bec communes.

Key Highlights

1. New samples of northern bottlenose whales were analyzed for microsatellites and combined with data from Feyrer et al. 2019 to identify genetic sex and population structure across the North Atlantic. The current dataset represents a total of 218 unique individuals, including additions from the Scotian Shelf, Iceland and the Davis Strait.
2. Population structure analysis supported the genetic distinctiveness of the Scotian Shelf subpopulation and with additional samples identified a previously undocumented population structure between northern bottlenose whales found east of Iceland (Jan Mayen) and the western North Atlantic.
3. Historical sampling and sample size may affect population structure results for sparsely sampled intermediary areas. Results grouping individuals from Southern Labrador and Newfoundland with the Davis Strait region, as well as western and eastern Iceland, should be considered putative until further study.
4. Future work should include increasing effort in under-sampled habitat areas across the species range and continuing to explore the broader application of genomic approaches.

Introduction

The northern bottlenose whale (NBW), *Hyperoodon ampullatus*, is a species beaked whale that inhabits deep (500- 2500m) shelf edge waters of the northern North Atlantic (Feyrer et al. 2019). The Scotian Shelf population, found primarily in the waters off Nova Scotia and Newfoundland, Canada, is listed as Endangered under the Canadian *Species at Risk Act* (SARA). Fisheries and Oceans Canada (DFO) prepared an action plan for this species (DFO 2017), which includes measures to better understand population structure and viability including:

Recovery measure 4. Research linkages between the Scotian Shelf population and the Davis Strait – Baffin Bay – Labrador Sea and other populations

Recovery measure 9. Update the Scotian Shelf population viability analysis when new information becomes available

Biopsies collected from animals on the Scotian Shelf and Davis Strait in 2019, and Iceland in 2014 and 2016, were added to an existing dataset of $n = 164$ sequenced for the same microsatellite markers (Feyrer et al. 2019). These additional samples (biopsy $n = 60$, unique $n = 55$) were included in a new analysis of the entire dataset ($n = 218$) to understand whether sample size, temporal sampling period or geography would influence Feyrer et al.'s (2019) assessment of linkages between populations. These analyses help to validate that previously identified patterns of population

subdivision can be replicated across sampling periods. Data from these samples also provide important demographic information (sex and unique individual identification) and are required for other ongoing research using biopsy tissues (fatty acid, stable isotope, contaminants) and monitoring of the population(s).

Specifically, the work described in this report: (1) contributes to ongoing genetic monitoring of the endangered Scotian Shelf population; (2) provides additional support for the conclusions of contemporary population structure, as published results (Feyrer et al. 2019) were based on a large proportion of samples from whaled animals and animals from the Scotian Shelf; and (3) helps address questions on movement and linkages between and within populations through the identification of population structure and repeat sampling events. These analyses contribute to ongoing efforts to improve our understanding of genetic connectivity of NBW populations within Canada and across the North Atlantic.

Materials and Methods

Sample collection

Biopsy samples were collected from northern bottlenose whales in the Gully submarine canyon in 2019 (biopsy n = 8, unique n = *7*) and in the Davis Strait-Labrador Sea in 2019 (biopsy n = 25, unique n = 21) and Iceland in 2014 and 2016 (biopsy n = 26, unique n = 26). Some biopsy samples were subsampled and sequenced independently for quality control (see Table 1). A female NBW stranded in 2019 in Southern Newfoundland was also analyzed (Figure 1). Biopsy sampling details generally followed those provided by Feyrer et al. (2019) and targeted only healthy adult animals (i.e. good body condition). Biopsy material (skin and blubber) was flash frozen within an hour of collection and stored at -20^{o}C until analysis.

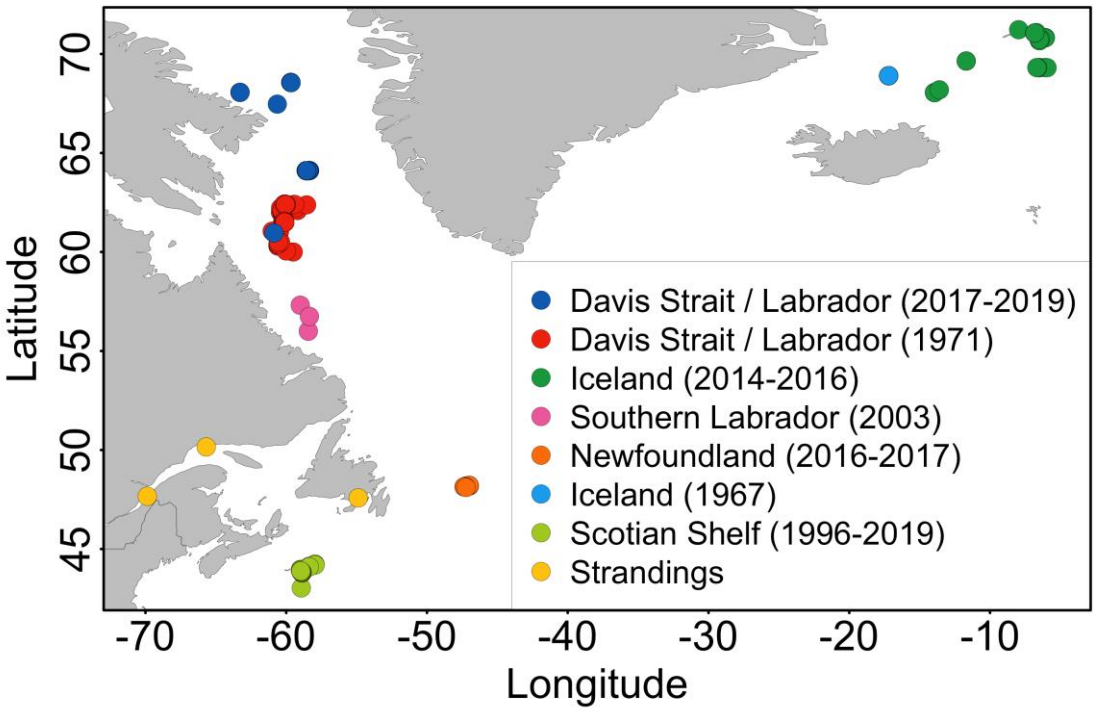


Figure 1. Map of samples available to this study. Groups are defined based on geography and sampling year for consistency with previous reports and publications.

DNA Extraction

Biopsy skin tissues arrived frozen and were thawed and preserved with ethanol when received by the Marine Gene Probe Laboratory (MGPL). DNA extraction was performed with a phenol-chloroform protocol (Sambrook & Russell., 2006) on frozen tissues of recent collection.

Sex Identification

Sex of each sampled individual was inferred from haplotype read depth of the zinc finger gene locus ZFX/ZFY, which is known to be linked to the sex chromosome. We identified samples with only ZFX haplotypes as females, samples with equal ratios of ZFX/ZFY ratios as males, and samples with abnormally high (~2x) ratios of ZFX/ZFY ratios as intersex. This method of identifying intersex genotypes has been validated by morphological examination of reproductive tract phenotypes (Einfeldt et al. 2019).

Microsatellite Sequencing

We amplified and sequenced nine NBW samples for 47 microsatellite loci developed by the MGPL. PCR amplifications and library preparation followed Zahn et al. (2017). Of these loci, five were dropped due to missing data (maximum missing data per locus threshold = 30%) in the combined dataset across two samples, and 42 were retained for analyses presented in this report and combined

with data from Feyrer et al. (2019). While Feyrer et al. (2019) only report on analyses of 37 loci, 5 additional loci were included in this report due to new sequencing recovering polymorphisms in some previously monomorphic loci and other loci passing the 30% threshold for missing data. DNA libraries created from pooled PCR amplicons were sequenced on an Illumina MiSeq DNA sequencer using a 150 cycle v3 reagent kit in a single read. Microsatellite DNA sequences were demultiplexed and scored using MEGASAT 1.0 (Zahn et al., 2017). The resulting data were curated to remove poorly amplifying loci, monomorphic loci, repeat samples, and technical duplicates.

To infer the number of genetic clusters in *NBW* and their spatio-temporal distributions, we performed Bayesian clustering of microsatellite genotypes using Structure v2.3.4 (Pritchard et al., 2000; Falush et al., 2007). Where individuals were repeatedly sampled or subsampled, only one of the sequenced samples was included in subsequent analyses to avoid biasing results. We analyzed results using models with and without admixture, which allows for mixed ancestry of individuals between genetic clusters. To account for differences in sample sizes and the expectation that both sampling location and year may be informative about ancestry (Hubisz et al., 2009; Wang, 2017) we used sampling region and year as location priors in all analyses. We averaged model log-likelihoods and individual assignment coefficients over 10 runs of 100,000 steps following a burn-in of 100,000 steps for each value of k from 1 to 5, and determined the best value of k using the ΔK method (Evanno et al., 2005).

Mitogenome Sequencing and Assembly

Although mitogenomes were not used in the final analysis due to a laboratory error, we provide a description of the method for recovering whole mtDNA. The analysis involved four steps: (i) Isolate DNA; (ii) Create a genomic library for each individual; (iii) Sequence the DNA libraries; and (iv) Use bioinformatic methods to recover the mtDNA genomes from each individual.

Genomic libraries were prepared in the MGPL at Dalhousie University for each sample by following a modified Illumina Nextera protocol (see Baym et al. 2015). Sequencing was performed on an Illumina MiSeq in the MGPL or Illumina NovaSeq at GeneWiz (Boston, MA).

Custom pipeline bioinformatic methods developed by the MGPL were used to isolate mtDNA fragments for each individual and assemble them into a complete mtDNA genome. Assembly performance was optimized by simultaneously trimming all reads for Illumina adapter sequences and applying stringent thresholds for sequence quality at leading and trailing bases (score > 25) and over a sliding window of 4 bases (average score > 23) using Trimmomatic (Bolger et al., 2018).

Reads from each sample were mapped to an *NBW* reference sequence (GenBank Accession: NC_005273_1) using the Bowtie2 (Langmead & Salzberg, 2012) software. We then performed an iterative assembly process with MIRA (Chevreux et al., 1999). First, we created a draft guided assembly for each sample using the *NBW* reference mitogenome and calculated intermediate statistics of assembly performance. Because guided assembly can lead to miscalled insertion and deletion variants, we then performed de novo draft assembly with MIRA for samples that passed an initial completeness threshold of 95%. For samples passing guided assembly but failing de novo assembly,

the guided draft assembly was passed to the next step in the pipeline and manually inspected for errors at the end of the assembly process. To account for overhanging genome ends that result from the MIRA assembler treating circular mitochondrial genomes as linear (a common issue for all assemblers), we split draft assemblies at the beginning of the mitochondrial control region and merged these sequences based on their overlap, creating draft mitochondrial genomes of consistent length. To identify errors in each assembly, we re-mapped reads from each sample to the corresponding assembly with Bowtie2 and then used Pilon (Walker et al., 2014) to correct miscalled bases, fill gaps, and identify ambiguous bases using read-based evidence. All code used in the custom pipeline is publicly available on GitHub (https://github.com/einfeldt/Hyperoodon).

Results

A total of 60 NBW tissue samples (including subsamples) were sequenced for microsatellite loci and genetic identification of sex: n = 8 from the Scotian Shelf, n = 25 from the Davis Strait – Labrador Sea, n = 26 from Iceland, and n = 1 stranding (Figure 1). Of these samples, five had identical microsatellite profiles reflecting repeated sampling of the same individual during field data collection occurring within the same region and year (Table 1). Thus, 55 unique individuals were added to an existing database of 164 NBW samples genotyped at 42 microsatellite loci (Appendix 1, Table S1). One individual had been sampled on the Scotian Shelf in different years, first in 2016 and again in 2019, consistent with site fidelity (Table 1). After removing all duplicate sequences, a total of 218 unique NBW individuals were used for downstream population structure analyses (Appendix 1, Table S2).

Table 1. Information for new samples sequenced in this report. Region abbreviations: Scotian Shelf (SS), Davis Strait – Labrador Sea (DS/L). Sex is determined by genetic ZFX/ZFY assay: M = male, F = female, I = intersex.

Code	Alternate code	Year collected	Region	Sex	Latitude	Longitude	Resample information
BKW0955	NBW19-01	2019	SS	F	43.8642	-58.9274	
BKW0956	NBW19-02	2019	SS	F	43.8306	-58.8833	
BKW0957	NBW19-03	2019	SS	F	43.8581	-58.9346	Sampled later: BKW0959
BKW0958	NBW19 04	2019	SS	F	43.8614	-58.9346	
BKW0959	NBW19 05	2019	SS	F	43.8304	-58.8948	Resample of BKW0957 (from SS)
BKW0960	NBW19 06	2019	SS	F	43.8384	-58.9054	
BKW0961	NBW19 07	2019	SS	F	43.8680	-58.9372	
BKW0962	NBW 19 08	2019	SS	F	43.8895	-58.9614	Resample of NBW-2016-12 (from SS)
BKW1278	NF20190812	2019	Stranding	F	47.5801	-54.8844	

BKW1303	NBW_2019_B1	2019	DS/L	M	64.1134	-58.3410	
BKW1304	NBW_2019_B2	2019	DS/L	M	64.1124	-58.3442	
BKW1305	NBW_2019_B3	2019	DS/L	M	64.1106	-58.3844	Sampled later: BKW1313
BKW1306	NBW_2019_B4	2019	DS/L	M	64.1102	-58.3867	
BKW1307	NBW_2019_B5	2019	DS/L	M	64.1097	-58.3896	
BKW1308	NBW_2019_B6	2019	DS/L	M	64.1109	-58.3913	
BKW1309	NBW_2019_B7	2019	DS/L	M	64.1106	-58.3921	
BKW1310	NBW_2019_B8	2019	DS/L	M	64.1105	-58.3957	Sampled later: BKW1312, BKW1314
BKW1311	NBW_2019_B9	2019	DS/L	M	64.1102	-58.3974	
BKW1312	NBW_2019_B10	2019	DS/L	M	64.1109	-58.3992	Resample of BKW1310
BKW1313	NBW_2019_B11	2019	DS/L	M	64.1071	-58.5764	Resample of BKW1305
BKW1314	NBW_2019_B12	2019	DS/L	M	64.1071	-58.5764	Resample of BKW1310
BKW1315	NBW_2019_B13	2019	DS/L	M	64.1071	-58.5764	
BKW1316	NBW_2019_B14	2019	DS/L	M	64.1071	-58.5764	
BKW1317	NBW_2019_B15	2019	DS/L	M	64.1071	-58.5764	
BKW1318	NBW_2019_B16	2019	DS/L	F	64.1075	-58.5448	
BKW1319	NBW_2019_B17	2019	DS/L	M	64.1075	-58.5448	
BKW1320	NBW_2019_B18	2019	DS/L	M	64.1075	-58.5448	
BKW1321	NBW_2019_B19	2019	DS/L	M	64.1075	-58.5448	
BKW1322	NBW_2019_B20	2019	DS/L	Intersex	64.1075	-58.5448	
BKW1323	NBW_2019_B21	2019	DS/L	F	60.9588	-60.8922	Sampled later: BKW1324
BKW1324	NBW_2019_B22	2019	DS/L	F	60.9595	-60.8680	Resample of BKW1323
BKW1325	NBW_B23	2019	DS/L	F	68.0618	-63.2835	
BKW1326	NBW_B24	2019	DS/L	F	68.0601	-63.2806	
BKW1327	NBW_B25	2019	DS/L	F	68.0606	-63.2794	
BKW1328	Ha14-01- RED	2014	Iceland	M	69.6444	-11.6887	Lab tissue subsampled (BKW1334)
BKW1329	Ha14-02- RED	2014	Iceland	M	70.9734	-6.6691	Lab tissue subsample of BKW1335
BKW1330	Ha14-03- RED	2014	Iceland	F	69.3114	-6.4826	
BKW1331	Ha14-04- RED	2014	Iceland	F	71.2254	-8.9289	Lab tissue subsample of BKW1342
BKW1332	Ha14-05- RED	2014	Iceland	F	70.9710	-6.6779	
BKW1333	Ha14-06- RED	2014	Iceland	F	70.8231	-6.0703	

BKW1334	Ha14-07- RED	2014	Iceland	M	69.6444	-11.6887	Lab tissue subsample of BKW1328
BKW1335	Ha14-08- RED	2014	Iceland	M	70.9734	-6.6691	Lab tissue subsampled (BKW1329)
BKW1336	Ha14-09- RED	2014	Iceland	M	NA	NA	
BKW1337	Ha14-10- RED	2014	Iceland	M	69.3137	-6.0733	
BKW1338	Ha14-11- RED	2014	Iceland	F	69.3114	-6.4826	
BKW1339	Ha14-12- RED	2014	Iceland	F	70.8231	-6.0703	
BKW1340	Ha14-13- RED	2014	Iceland	M	70.8147	-6.0789	
BKW1341	Ha14-14- RED	2014	Iceland	F	69.3094	-5.9469	
BKW1342	Ha14-15- RED	2014	Iceland	F	70.9721	-6.6788	Lab tissue subsampled (BKW1331)
BKW1343	Ha14-16- RED	2014	Iceland	F	70.9710	-6.6779	
BKW1344	Ha14-17- RED	2014	Iceland	F	69.3093	-6.6218	
BKW1345	Ha14-18- RED	2014	Iceland	F	69.3127	-6.6761	
BKW1346	Ha16-01- RED	2016	Iceland	F	68.0518	-13.9446	
BKW1347	Ha16-02- RED	2016	Iceland	F	68.0503	-13.9416	
BKW1348	Ha16-03- RED	2016	Iceland	M	68.0488	-13.9409	
BKW1349	Ha16-04- RED	2016	Iceland	F	68.1906	-13.5989	
BKW1350	Ha16-05- RED	2016	Iceland	F	70.7623	-6.4379	
BKW1351	Ha16-06- RED	2016	Iceland	F	70.6711	-6.4865	
BKW1352	Ha16-07- RED	2016	Iceland	F	70.6711	-6.4865	
BKW1353	Ha16-08 - RED	2016	Iceland	F	71.2263	-7.9653	
BKW1354	Ha16-09 - RED	2016	Iceland	F	71.0570	-6.8242	
BKW1355	Ha16-10 - RED	2016	Iceland	F	71.0593	-6.8190	
BKW1356	Ha16-11 - RED	2016	Iceland	F	71.0807	-6.7076	

Mitogenomes

A total of 26 samples sequenced for microsatellites and ZFX/ZFY were chosen for whole genome and mitogenome sequencing, based on quality and quantity of DNA extracted from tissue during work. After sequencing was completed, disagreement between targeted ZFX/ZFY and whole-genome approaches to assaying sex for a number of individuals revealed that a laboratory error had occurred related to sequencing indexes not matching their recorded sample identification numbers. While the exact source of error could not be identified, the affected samples were limited to a single plate of sequencing. Due to the small quantity of NBW tissue available, additional DNA extraction and sequencing was not possible for most of these samples.

In total, two mitogenomes were assembled that were not affected by the indexing error: BKW1350 and BKW1351. Coupled with the low mitogenomic diversity previously reported in NBW, this sample size is insufficient for meaningful analysis of regional genetic diversity and differentiation. From previous reports, the main application of mitogenomes in this species has been in reconstructing historical demographic trends. As these trends can now be investigated using genome resequencing data and mitogenomes are a product of full genome resequencing, we anticipate reconstructing demographic history from genomic data will supersede analyses of mitogenomic markers. Aside from comparative analyses, targeted resequencing of mitogenomes is unlikely to have future applications for NBW.

Genetic Diversity

Genetic diversity in NBW from 42 microsatellite loci across 218 unique individuals was low overall, with very few alleles detected for each locus (mean = 3.6). 15 loci had 2 alleles, 12 loci had 3 alleles, 4 loci had 4 alleles, 4 loci had 5 alleles, 4 loci had 6 alleles, and only 1 locus each was found with 7, 8, and 10 alleles (Table 2).

Table 2. Diversity indices for microsatellite loci retained in final analyses.

Locus	n Alleles	H_o	H_e
Hyam-002	8	0.7619	0.8045
Hyam-007	3	0.1357	0.1357
Hyam-008	4	0.3484	0.3659
Hyam-012	5	0.6036	0.6457
Hyam-015	6	0.7083	0.7339
Hyam-016	3	0.7136	0.6647
Hyam-020	2	0.0090	0.0090
Hyam-024	3	0.5928	0.5451
Hyam-026	10	0.2349	0.3994
Hyam-027	4	0.5762	0.6106
Hyam-031	5	0.5667	0.5839
Hyam-032	4	0.0955	0.0924
Hyam-034	4	0.3784	0.3685
Hyam-040	5	0.4434	0.4260
MK6	6	0.4949	0.4885
New-Hyam-094	2	0.5336	0.4951
New-Hyam-095	2	0.5202	0.4903
New-Hyam-098	3	0.3004	0.2975
New-Hyam-099	2	0.4574	0.4575
New-Hyam-102	3	0.2773	0.2741
New-Hyam-106	2	0.4118	0.4744
New-Hyam-108	2	0.3604	0.4896
New-Hyam-109	2	0.1121	0.1137
New-Hyam-113	2	0.0717	0.0692

New-Hyam-114	2	0.2242	0.4814
New-Hyam-116	2	0.0500	0.0488
New-Hyam-117	3	0.1121	0.1224
New-Hyam-118	2	0.0583	0.0566
New-Hyam-119	3	0.4234	0.4805
New-Hyam-120	3	0.1442	0.1577
New-Hyam-121	6	0.7027	0.6997
New-Hyam-122	6	0.3604	0.3633
New-Hyam-124	3	0.3516	0.3581
New-Hyam-126	5	0.7149	0.6856
New-Hyam-127	3	0.5450	0.4932
New-Hyam-128	2	0.0045	0.0045
New-Hyam-129	2	0.4820	0.4622
New-Hyam-130	2	0.9507	0.4988
New-Hyam-131	3	0.5112	0.5335
New-Hyam-133	2	0.0045	0.0045
New-Hyam-135	7	0.3964	0.4120
New-Hyam-136	3	0.0987	0.0940

Measures of genetic diversity including observed and expected heterozygosity, allelic diversity, and the number of alleles per locus were similar across sampling regions (Table 3). The number of private alleles – those detected in one sampling region but not others – differed among groups. Samples from the Scotian Shelf had 3 microsatellite alleles not detected elsewhere, indicating unique genetic diversity endemic to this region. The greatest number of private alleles detected in a sampling group was 7 in Davis Strait (n = 78), which consists of samples collected in 1971. These alleles were not detected in more recent sampling between 2017-2019 (n = 29). Only 1 private allele was detected in the contemporary Davis Strait – Labrador Sea sampling, which is consistent with a loss of genetic variation due to genetic drift or removal by harvesting (Table 3). In Iceland, 6 private alleles were detected in contemporary samples (2014-2016, n = 26) while only 1 was detected in historical samples collected in 1967 (n = 7), but this is likely due to the large difference in sample sizes, and particularly the low sample size in 1967.

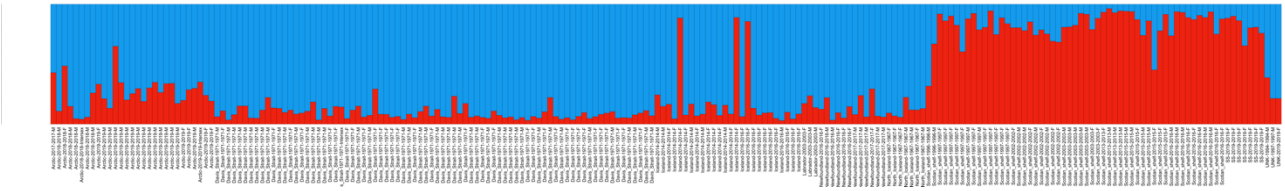
Table 3. Genetic diversity of microsatellites in NBW, including observed heterozygosity (H_o), expected (H_e) heterozygosity, and fixation index for individuals within groups (F_{is}). M:F:I refers to Male, Female, Intersex. Groups are defined *a priori* based on geography and sampling year for consistency with previous reports and publications.

Region	*n*	M	F	I	Percent Missing	Alleles/ locus	Allelic diversity	Mean H_o	Mean H_e	Private alleles	F_{is}
West Iceland (1967)	7	3	4	0	3.40	2.33	0.34	0.33	0.34	1	0.09
Jan Mayen (2014-2016)	26	6	20	0	0.73	2.79	0.37	0.38	0.37	5	-0.02
Davis Strait – Labrador Sea (1971)	78	44	33	1	4.21	3.24	0.37	0.37	0.37	7	0.01
Davis Strait – Labrador Sea (2017-2019)	29	21	6	2	1.97	2.90	0.37	0.38	0.37	1	0.01
Southern Labrador (2003)	3	2	1	0	1.59	2.0	0.33	0.40	0.33	1	-0.03
Newfoundland (2016-2017)	12	7	5	0	0.59	2.64	0.38	0.40	0.38	1	-0.01
Scotian Shelf (1996-2019)	60	26	34	0	0.63	2.95	0.37	0.38	0.37	3	-0.01
Stranded	3	2	1	0	0	2.07	0.34	0.42	0.35	0	-0.04
All	218	111	104	3	2.20	3.60	0.38	0.38	0.38	NA	-0.01

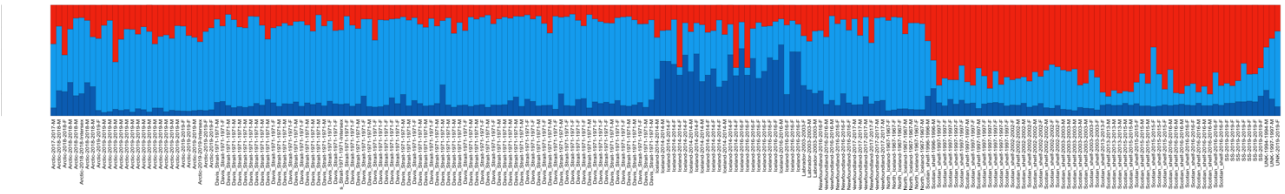
Population Analyses

Analyses of population structure supported similar patterns of genetic structure overall, though with different values of *k* genetic clusters for models run with or without admixture (Figure 2). In interpreting these results, it is important to consider the limitations of Structure and the nature of the underlying dataset. Given there are likely differing degrees of population subdivision in NBW across the entire North Atlantic, it is known that the highest level of hierarchical structure can swamp more fine-scale subdivisions, leading to an underestimation of the true number of clusters (Janes et al. 2017). In addition, the strong similarity of closely related family groups in a dataset can bias analysis and lead to an overestimation of the true number of genetic clusters (Porras-Hurtado et al. 2013). Using the Evanno method, both the admixture and no admixture models supported $k = 2$ genetic clusters. However, given the above noted issues with Structure, and probable biases in the underlying dataset (e.g., probable relatedness across the Scotian Shelf samples), we used $k = 3$ as the most biologically relevant number of clusters supported by both analyses.

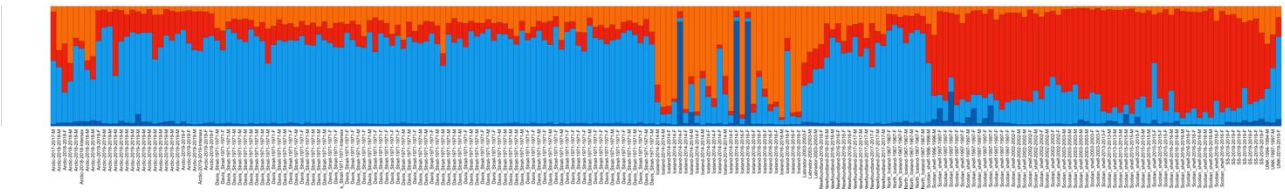
a) *k*=2, admixture model



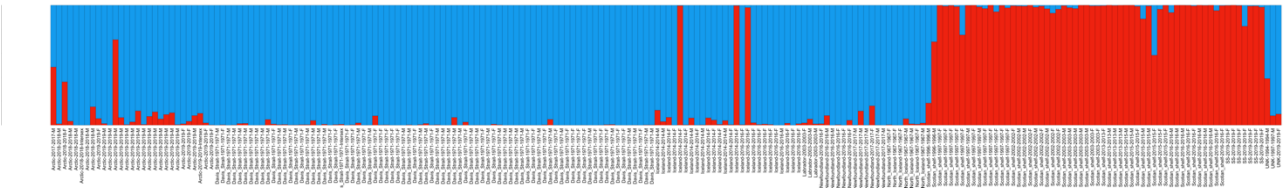
b) *k*=3, admixture model



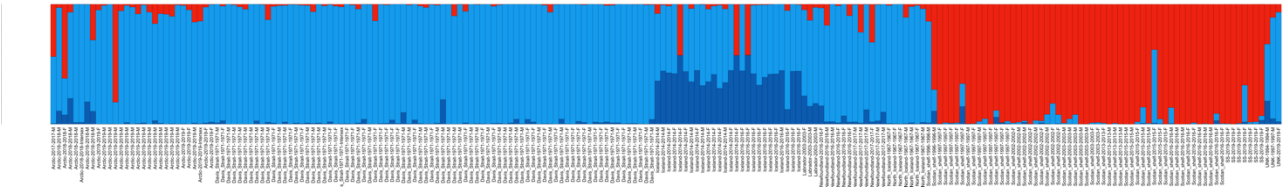
c) *k*=4, admixture model



d) *k*=2, no-admixture model



e) *k*=3, no-admixture model



f) *k*=4, no-admixture model

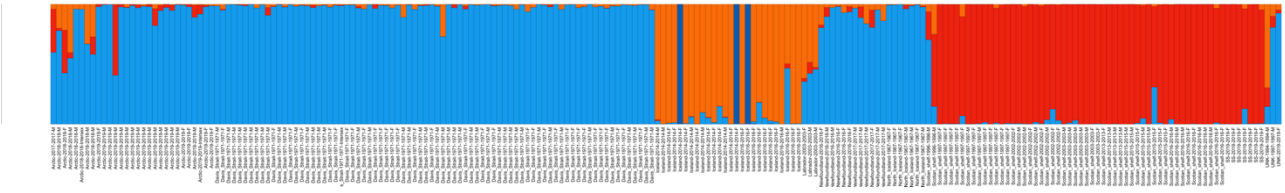


Figure 2. Assignment of individuals to k genetic clusters using models assuming admixture (a,b,c) or no admixture (d,e,f) and location priors in Structure. Bars indicate likelihood of each individual belonging to the genetic cluster for *k*=2 (a,d), *k*=3 (b,e), and *k*=4 (c,f) genetic clusters. From left to right samples are from Davis Strait, Iceland, Labrador-Newfoundland, Scotian Shelf

regions and strandings (UNK). Individual sample labels also include year and sex based on ZFX/ZFY.

In all analyses of population structure, individuals from the Scotian Shelf are assigned to a genetic cluster that is distinct from individuals sampled at all other locations. For analyses of k=3, recently sampled individuals from Iceland (i.e., biopsies from 2014 and 2016) are identified as a third distinct genetic cluster. Although k=2 was selected as the best fit under a model with no admixture, the clear separation of contemporary Iceland from other groups at k=3 indicates that varying scales of hierarchical structure is present in this dataset. The relatively stronger genetic subdivision between the Scotian Shelf individuals and all other regions likely masks weaker signals of subdivision using Evanno's method for determining the best value of k (i.e. the "k=2 conundrum" discussed by Janes et al. 2017). Overall, we suggest that the representation of three clusters is the most biologically meaningful result.

Genetic Subdivision of the Eastern North Atlantic

Individuals killed off of Iceland in 1967 clustered with samples from the western North Atlantic, while contemporary biopsy samples from 2014 and 2016 formed a unique cluster. The 1967 Iceland samples were taken by whalers further west than those collected by researchers off Jan Mayen in 2014 and 2016 (Figure 1). Geographic differences in sampling between western Iceland (1967) and Jan Mayen, eastern Iceland (2014-2016) leaves uncertainty as to whether the genetic separation detected reflects a barrier to gene flow or temporal change in genetic composition. Within contemporary samples collected in Iceland, three individuals (BKW1332, BKW1343, and BKW1351) stand out in analyses, grouping with different populations at different values of k (Figure 2). At k=2 and k=3 these individuals cluster more closely with the Scotian Shelf group, while at k=4 they form a unique cluster separate from all other individuals. These individuals likely represent a close family group with high genetic similarity, rather than representing a distinct sympatric subpopulation. Related kin-pairs would bias clustering analyses and lead to an overestimation of k in the model with no admixture (Porras-Hurtado et al. 2013). Unfortunately, close-kin relationships cannot be confidently estimated in this dataset due to low overall microsatellite diversity and high genetic similarity between individuals. However, due to their ancestry coefficients not supporting complete separation as a single distinct genetic group, close-kin relationships are the most likely explanation for the unusual clustering of these individuals.

Potential Ancestry and Migration

Across analyses, four individuals had ancestry coefficients that differed from other individuals sampled in the same geographic location as reported by Feyrer et al. (2019):

- NBW07-2015 was sampled in the Scotian Shelf and clustered more closely with individuals in the Davis Strait – Labrador Sea
- BNW12, NBW01ARC17, and BKW1309 were sampled in the Davis Strait – Labrador Sea but clustered more closely with the Scotian Shelf.

These results suggest that these individuals may be migrants, or the offspring of individuals that dispersed from their natal region and successfully reproduced. No new evidence of migration was detected from the additional samples sequenced in this report.

Summary

This report considers additional microsatellite genetic data from contemporary samples (n = 63) with the full dataset of NBW samples collected by Feyrer et al. 2019 (n = 164) across the Scotian Shelf, Davis Strait – Labrador Sea, and Iceland, representing a total of 218 unique individuals. A re-analysis of population structure continues to support the existence of a genetic subdivision of NBW in the Scotian Shelf from other regions as found by Feyrer et al. 2019. This indicates that this geographic population subdivision is unlikely to be a by-product of temporal biases in sampling periods. Individuals have been sampled multiple times across different years in the Scotian Shelf, consistent with other reports of site fidelity (e.g., long term resights in photo-identification reported in Feyrer et al. 2021).

The addition of contemporary samples from the eastern North Atlantic, off Jan Mayen, Iceland, revealed subtle population structure which was not previously detected by Feyrer et al. (2019) using seven samples collected by whalers off western Iceland in 1967. Due to uncertainty in the precise collection location of whaled animals, it is not clear whether subdivision between contemporary and historic samples in Iceland are due to fine-scale geographic subdivision within the eastern North Atlantic or population fragmentation over this period. The newly detected genetic subdivision between eastern and western Atlantic samples indicates more limited connectivity between NBW in this part of their range than previously understood. This finding suggests that fragmentation and barriers to connectivity may present additional challenges for demographic recovery across their range. The current NBW tissue database reflects concentrated periods of sampling effort that has emphasized core habitat areas. Ongoing monitoring with microsatellite markers can provide valuable information on individual movement, genetic subdivision, and the broad-scale providence of stranded individuals, but the quality of inferences ultimately depends on sampling effort. As NBW inhabit intermediary regions between the core areas where sampling effort has been concentrated so far, determining fine-scale geographic boundaries for poorly or unsampled areas still requires additional specimens to provide the power necessary to detect subtle population differentiation. Sampling outside the well documented Scotian Shelf canyons and Davis Strait regions (i.e., Newfoundland, Labrador and other areas) should be prioritized to identify potential mechanisms behind currently understood population structure.

An important application of population structure analyses for NBW is that the providence of stranded animals can now be reliably assigned to one of three major regions: the Scotian Shelf, the western North Atlantic (Davis Strait, Labrador, Newfoundland), or Iceland. This does not preclude the potential for detection of more subtle population structure within those three geographic regions sampled (e.g., within the western North Atlantic), as the small sample sizes from subregions (i.e., Newfoundland, Labrador) and low diversity of markers used here may restrict power over fine spatial scales.

Recommendations for Future Work

1. *Additional sampling* of underrepresented subregions where NBW have been acoustically detected or observed, including southern Nova Scotia, Labrador and Newfoundland. A larger sample size has demonstrated utility in increasing the power to detect subtle population structure in this low diversity species.

2. *Genome resequencing* provides the resolution necessary to overcome some of the challenges of the low diversity of microsatellite markers and small sample sizes. Genome resequencing can be used to identify kin relationships amongst individuals and can also provide more information about population subdivision with fewer samples due to the increased number of polymorphisms detected when surveying the entire genome. The additional population structure identified in this report between Iceland and the western North Atlantic emphasizes the importance of sampling across a broad geographic area and genomic approaches can also test whether subpopulations harbor genetic diversity that is locally adapted, which cannot be addressed with microsatellite markers.

Acknowledgements

We thank Kelsey Johnson, Kyle Lefort, and Nigel Hussey for their help collecting and processing Arctic biopsy samples. Samples from the stranded animal were collected by Wayne Ledwell, Capelin Bay Stranding Network with support from Jack Lawson, DFO Newfoundland. Species at Risk, DFO Maritimes Region funded the genetic analyses, which was conducted by the Marine Gene Probe Lab at Dalhousie University. Fieldwork on the Scotian Shelf was conducted by LJF and researchers at Dalhousie University and funded by Fisheries and Oceans Canada, Marine Protected Areas group, Maritimes Region. Fieldwork off Jan Mayen was conducted by PM and members of the 3S collaborative research project, with funding from the US Office of Naval Research and SERDP grant RC-2337. This work was further supported by Fisheries and Oceans Canada (Arctic), Crown-Indigenous Relations and Northern Affairs Canada, Government of Nunavut, Nunavut Fisheries Association, Canadian Northern Economic Development, and Maritimes (SARP) funding to the Balaena Institute to support AE's work on this report.

Appendix 1

Data used in this analysis is stored in a publically available repository on Open Science at: https://doi.org/10.17605/OSF.IO/97AVS.

Citation for the Data:

Einfeldt, A., Feyrer, L., Paterson, I., Ferguson, S. H., Miller, P., de Greef, E. and P. Bentzen. 2022. Appendix 1: Characterization of Population Structure in Northern Bottlenose Whales (Hyperoodon ampullatus) in the North Atlantic from an Analysis of Microsatellite Data. Open Science. https://doi.org/10.17605/OSF.IO/97AVS

References

Chevreux B, Wetter T, and Suhai S. 1999. Genome Sequence Assembly Using Trace Signals and Additional Sequence Information. *Computer Science and Biology: Proceedings of the German Conference on Bioinformatics (GCB), 99*: 45-56.

DFO (Fisheries and Oceans Canada). 2017. Action Plan for the Northern Bottlenose Whale (Hyperoodon ampullatus), Scotian Shelf population, in Atlantic Canadian waters. Species at Risk Act Action Plan Series. Fisheries and Oceans Canada, Ottawa. iv + 37 pp.

Einfeldt, A., Feyrer, L., Paterson, I., Ferguson, S. H., Miller, P., de Greef, E. and P. Bentzen. 2022. Appendix 1: Characterization of Population Structure in Northern Bottlenose Whales (Hyperoodon ampullatus) in the North Atlantic from an Analysis of Microsatellite Data. Open Science. https://doi.org/10.17605/OSF.IO/97AVS

Einfeldt AL, Orbach DN, Feyrer LJ. 2019. A method for determining sex and chromosome copy number: sex-by-sequencing reveals the first two species of marine mammals with XXY chromosome condition. *Journal of Mammalogy*, *100*(5):1671-7.

Evanno G, Regnaut S, and Goudet J. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology*, *14*(8), 2611-2620.

Falush D, Stephens M, and Pritchard JK. 2007. Inference of population structure using multilocus genotype data: dominant markers and null alleles. *Molecular Ecology Notes*, *7*(4), 574-578.

Feyrer LJ, Bentzen P, Whitehead H, Paterson IG, Einfeldt A. 2019. Evolutionary impacts differ between two exploited populations of northern bottlenose whale (Hyperoodon ampullatus). *Ecology and evolution*, *9*(23):13567-84.

Feyrer LJ, Stewart M, Yeung J, Soulier C, Whitehead H. 2021. Origin and Persistence of Markings in a Long-Term Photo-Identification Dataset Reveal the Threat of Entanglement for Endangered Northern Bottlenose Whales (*Hyperoodon ampullatus*). *Frontiers in Marine Science, 8*:484.

Hubisz MJ, Falush D, Stephens M, and Pritchard JK. 2009. Inferring weak population structure with the assistance of sample group information. *Molecular Ecology Resources*, *9*(5), 1322-1332.

Janes JK, Miller JM, Dupuis JR, Malenfant RM, Gorrell JC, Cullingham CI, Andrew RL. 2017. The K= 2 conundrum. *Molecular Ecology, 26*(14):3594-602.

Langmead B and Salzberg S. 2012. Fast gapped-read alignment with Bowtie 2. *Nature Methods, 9*: 357-359.

Porras-Hurtado L, Ruiz Y, Santos C, Phillips C, Carracedo Á, Lareu M. 2013. An overview of STRUCTURE: applications, parameter settings, and supporting software. *Frontiers in genetics*, *29*(4):98.

Pritchard JK, Stephens M, and Donnelly P. 2000. Inference of population structure using multilocus genotype data. *Genetics, 155*(2), 945-959.

Sambrook J and Russell W. 2006. Purification of nucleic acids by extraction with phenol: chloroform. *Cold Spring Harbor Protocols,* 1:pp.pdb-prot4455.

Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q, Wortman J, Young SK, Earl AM. 2014. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PloS One, 9*(11):e112963.

Wang J. 2017. The computer program STRUCTURE for assigning individuals to populations: easy to use but easier to misuse. *Molecular Ecology Resources, 17*(5), 981-990.

Zhan L, Paterson IG, Fraser B, Watson B, Bradbury IR, Ravindran PN, Reznick D, Beiko RG, Bentzen P. 2017. MEGASAT: automated inference of microsatellite genotypes from sequence data. *Molecular Ecology Resources, 17*(2):247-256. doi: 10.1111/1755-0998.12561.