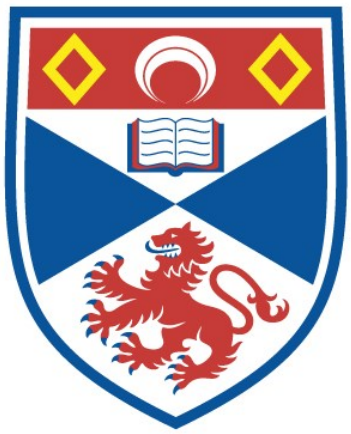
DNA IN THE DEEP: COMPARATIVE MOLECULAR ECOLOGY FOR THE CONSERVATION OF BEAKED WHALES

Aubrie Booth Onoufriou

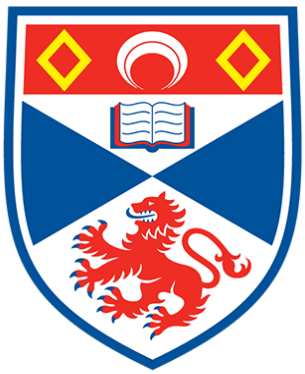
A Thesis Submitted for the Degree of PhD
at the
University of St Andrews



2023

DNA in the deep: Comparative molecular ecology for the conservation of beaked whales

Aubrie Booth Onoufriou



University of
St Andrews

This thesis is submitted in partial fulfilment for the degree of

Doctor of Philosophy (PhD)

at the University of St Andrews

April 2023

Abstract

Beaked whales (Ziphiidae) are an incredibly diverse and elusive family, and behavioural responses linked to certain anthropogenic sounds have resulted in mass stranding events. The substantial knowledge gaps regarding abundance and population structure for most ziphiids highlight that more information is required for their effective management and conservation. To this end, next-generation DNA sequencing was used to investigate beaked whale ecology and evolution, showcasing the utility of reduced representation sequencing in non-model organisms. Single nucleotide polymorphisms (SNPs) were derived using double-digest restriction site associated DNA sequencing (ddRAD) from a newly established international beaked whale tissue and DNA archive.

Globally, hierarchical genetic structure and diversity of cosmopolitan Cuvier's and Blainville's beaked whales (*Ziphius cavirostris* and *Mesoplodon densirostris*, "Cuvier's" and "Blainville's") were investigated. Biogeographic barriers and differing life and evolutionary histories have contributed to the observed patterns, and the findings are evaluated in the context of management units for conservation. Regionally, population structure and demographic history were characterised for four North Atlantic beaked whales: Cuvier's, Blainville's, Sowerby's beaked whales (*M. bidens*) and northern bottlenose whales (*Hyperoodon ampullatus).* Fluctuations in effective population size (N_e) were likely responses to climatic change. With these findings, potential responses of beaked whale populations to future climate change are discussed. Locally, genetic Essential Biodiversity Variables (EBVs) were calculated for paired 'disturbed' and 'semi-pristine' populations of Cuvier's and Blainville's in three well studied populations: the Bahamas, Canary Islands and Mediterranean Sea (Cuvier's only). At least one 'disturbed' site was found in each region with reduced genetic variation and at risk of genetic erosion.

In summary, genetic population structure has been identified in many ziphiids. These populations have different levels of genetic diversity, different demographic responses to historic climatic change and it is likely that they have different abilities to adapt to future anthropogenic and climate impacts and should be managed on a population level.

Keywords: Beaked whales, Molecular ecology, Genomics, ddRAD, Mitogenome, *Ziphius cavirostris*, *Mesoplodon densirostris*, *Mesoplodon bidens*, *Hyperoodon ampullatus*, Population structure, Genetic diversity, Inbreeding, Demographic history, Anthropogenic noise

Candidate's declaration

I, Aubrie Booth Onoufriou, do hereby certify that this thesis, submitted for the degree of PhD, which is approximately 54,716 words in length, has been written by me, and that it is the record of work carried out by me, or principally by myself in collaboration with others as acknowledged, and that it has not been submitted in any previous application for any degree. I confirm that any appendices included in my thesis contain only material permitted by the 'Assessment of Postgraduate Research Students' policy.

I was admitted as a research student at the University of St Andrews in September 2017.

I received funding from an organisation or institution and have acknowledged the funder(s) in the full text of my thesis.

Date

10 April 2023

Signature of candidate

[redacted]

Supervisor's declaration

I hereby certify that the candidate has fulfilled the conditions of the Resolution and Regulations appropriate for the degree of PhD in the University of St Andrews and that the candidate is qualified to submit this thesis in application for that degree. I confirm that any appendices included in the thesis contain only material permitted by the 'Assessment of Postgraduate Research Students' policy.

Date

Signature of supervisor

Permission for publication

In submitting this thesis to the University of St Andrews we understand that we are giving permission for it to be made available for use in accordance with the regulations of the University Library for the time being in force, subject to any copyright vested in the work not being affected thereby. We also understand, unless exempt by an award of an embargo as requested below, that the title and the abstract will be published, and that a copy of the work may be made and supplied to any bona fide library or research worker, that this thesis will be electronically accessible for personal or research use and that the library has the right to migrate this thesis into new electronic forms as required to ensure continued access to the thesis.

I, Aubrie Booth Onoufriou, confirm that my thesis does not contain any third-party material that requires copyright clearance.

The following is an agreed request by candidate and supervisor regarding the publication of this thesis:

Printed copy

No embargo on print copy.

Electronic copy

No embargo on electronic copy.

Date	Signature of candidate
10 April 2023	
Date	Signature of supervisor

Underpinning Research Data or Digital Outputs

Candidate's declaration

I, Aubrie Booth Onoufriou, understand that by declaring that I have original research data or digital outputs, I should make every effort in meeting the University's and research funders' requirements on the deposit and sharing of research data or research digital outputs.

Date 10 April 2023 Signature of candidate

Permission for publication of underpinning research data or digital outputs

We understand that for any original research data or digital outputs which are deposited, we are giving permission for them to be made available for use in accordance with the requirements of the University and research funders, for the time being in force.

We also understand that the title and the description will be published, and that the underpinning research data or digital outputs will be electronically accessible for use in accordance with the license specified at the point of deposit, unless exempt by award of an embargo as requested below.

The following is an agreed request by candidate and supervisor regarding the publication of underpinning research data or digital outputs:

No embargo on underpinning research data or digital outputs.

Date 10 April 2023 Signature of candidate

Date Signature of supervisor

Ethical Approval

University of St Andrews

School of Biology Ethics Committee

26th July 2018

Project Title:	Population genetics and molecular ecology of cetaceans
Researchers Name(s):	Oscar Gaggiotti, Emma Carroll, Marie Louis and Aubrie Booth
Supervisor(s):	Prof Oscar Gaggiotti
Biology SEC Ref:	SEC18020

Thank you for submitting your application which was considered by the Biology School Ethics Committee on the 26th July. The following documents were reviewed:

1. Animal Ethics Form 6/07/2018

The School of Biology Ethics Committee approves this study from an ethical point of view.

Approval is given for five years. Projects, which have not commenced within two years of original approval, must be re-submitted to the School Ethics Committee.

You must inform the School Ethics Committee when the research has been completed. If you are unable to complete your research within the five year validation period, you will be required to write to the School Ethics Committee to request an extension or you will need to re-apply.

Any serious adverse events or significant change which occurs in connection with this study and/or which may alter its ethical consideration, must be reported immediately to the School Ethics Committee, and an Ethical Amendment Form submitted where appropriate.

Approval is given on the condition that local permits are obtained prior to the project starting, and also on the understanding that the ASAB Guidelines for the Treatment of Animals in Behavioural Research and Teaching published in Animal Behaviour, 2003, 65, 249-255, are adhered to.

Yours sincerely,

Convenor of the School Ethics Committee

Ccs School Ethics Committee
Dr Tamara Lawson (Home Office Liaison Officer)

SEC Convenor, CBD, Dyers Brae House, Greenside Place, St Andrews, Fife KY16 9TH, Scotland
Email: bioethics@st-andrews.ac.uk Tel: 01334 462054

Table of Contents

List of Figures

List of Tables

Acknowledgements

When I started this PhD in 2017, I'd never have predicted what the next five years would bring. Three house moves, two dog surgeries, one "Life in the UK" test, one global pandemic, hundreds of hours working at home, hundreds of days of "social distancing", one baby, and perhaps most surprising of all considering everything that happened: this thesis. From the beginning, I was instructed to think of a PhD as if I was told to eat an elephant. This is, of course, an impossible task to do in one sitting, but if you just take one bite at a time, you will eventually finish (thanks Emma). This has proven to be true, but what I didn't realise at the start was how many forks I would need to pass around to get the help I needed to finish this massive endeavour.

My first and most heartfelt thanks go to my supervisors: Natacha Aguilar de Soto, Oscar Gaggiotti and Emma Carroll. Thanks Natacha, for your never-ending optimism and enthusiasm, and trusting me to continue this project that was started long before I came along. I am now well and truly a member of the "Beaked Whales are Amazing!" club.

Oscar, you have been the perfect supervisor for me from the beginning: a pillar of knowledge, expertise, encouragement, and support. I will never forget that you always had my back and the ability to pacify even my most dramatic meltdowns. I'm honoured to have been one of your students and appreciate every opportunity that you gave me to learn and share my work.

To Emma, you've been an inspiration since we met and a stalwart example of how to be an intelligent and capable scientist. We were friends before you were my supervisor and I appreciate how strange of a transition it must have been for you. While I could wield a pipette and read a protocol, you taught me everything else I needed to know to complete the practical side of this thesis, skills that I will take with me throughout my career. You encouraged me to reach further, pushing me to look for inspiration outside the world of marine mammal genomics and constantly reminded me that my work was interesting and important. Thank you for helping me navigate the interpersonal skills required for collaboration and demonstrating the importance of elevating the voices and stories from people and places around the world that need it the most. Let's keep making cool beaked whale science!

This work was supported by the Office of Naval research (N000141613017) and he University of St Andrews, School of Biology with a partial studentship.

There's a long list of people who have been involved in the laboratory and analytical aspects of this project and helped make this thesis as comprehensive as it is. Morten Olsen, thank you for the years of advice, supervising the mitogenome work, assisting with the administrative side of sequencing, and for your general enthusiasm and encouragement. I'd still love to come to Copenhagen someday! For their help with the mitogenome work, thank you also to Sascha Dreyer and Sabine Hansen. Special thanks to Morgan McCarthy for all the mitogenome analysis and expertise, writing help, and general friendliness- I promise never to message you on Slack again!

Here in St. Andrews, many people were involved in the lab work. Sincere thanks go to Devi Veytia, Cristel Reyes Suárez and Emma Carroll for the months of DNA extractions and library preparations that were completed before I started on the project. Many thanks to Valeria Montano, Gabby Genty, Lauren Arkoosh and Emma Carroll again, for helping with the lab work after I started. Your company and assistance in the lab were always welcome, and I'd never have been able to get it all done without you. Thanks especially to Kim Andrews for providing the ddRAD protocol.

Many of the analyses conducted for this thesis were done using the high-performance cluster managed by the University of St Andrews Bioinformatics Unit and special thanks go to Ramon Fallon, Joe Ward and Peter Thorpe with their help using it. Later analyses used the UK Crop Diversity Bioinformatics high performance cluster managed by the Research/Scientific computing teams at the James Hutton Institute and NIAB. Many thanks to Iain Milne and Peter Thorpe for their assistance with this cluster.

Many other people provided technical assistance with the processing, storage and shipment of samples including David Ferrier for providing lab space; Tanya Sneddon for storing samples and allowing me to use lab equipment; Pete Baxter and Ian Johnston for help with lab equipment, sample storage and shipment; Franca Eichenberger for your incredible help to catalogue, organise and ship the beaked whale tissue and DNA archive; and Ellen Garland for organising and shipping the beaked whale tissue and DNA archive. At the University of St Andrews, I thank Nathan Bailey and Neil Hazon for their pastoral support and encouraging discussions during our review meetings. To Simon Moss, Ryan Milne, Matt Bivins and Steve Balfour, thank you for the DOZENS of times you let me into the basement, for helping me with some random bit of equipment I needed for whatever random task I was doing, and for looking after all of my samples in the freezers! To Holly Armstrong, thank you for saving me boxes, collecting my mail, sending my samples, and lighting up every room with your kindness. Thank you also to everyone at the GeoGenetics Sequencing Core at the University of Copenhagen. I also thank Agus Schiavi and Chloé Yzoard at ULL for their administrative help in processing invoices.

It goes without saying that this project would not have been possible without the dozens of people who collected and contributed the samples from all of the world. I especially want to thank the field teams at the University of La Laguna (Natacha Aguilar de Soto, Chloé Yzoard, Agustina Schiavi, Jesúz Alcázar-Treviño, Daniel Miranda González, Jacobo Marrero, Patricia Arranz, A. Martin and Raquel Viñé) and CIMA (Massimiliano Rosso, Aurelie Moulins and Paola Tepsich) for collecting samples specifically for this project.

While all contributions were absolutely crucial, there were a few people that I need to thank specifically. Thanks to Phil Morin (SWFSC) for always lending an ear, offering advice, quickly reading, and commenting every draft, and coordinating the transfer of samples from NOAA's Southwest Fisheries Science Centre for use in this thesis. To Robin Baird (Cascadia Research Collective), thank you for your ever enthusiastic interest in this project, for providing samples and for providing me with the individual stories of the animals that the samples were collected from. Diane Claridge (BMMRO), thank you for the samples and individual stories as well. Thanks to Andrew Kitchener (National Museums Scotland) for not only providing samples, but also the taxonomic advice and coming up with the proposed subspecies names. Thank you to Andrew Brownlow and Nick Davison at SMASS. We tried so hard to work the 2018 Cuvier's UME into this

thesis and although we weren't successful, at least we have the next grant! Thanks for providing so many samples from Scotland, especially all of the UME samples and Sowerby's samples. An especially huge "thank you" goes to Merel Dalebout, for not only providing a substantial proportion of the samples, but also paving the way for my thesis with the library of beaked whale genetics work you have already done. I stood on your shoulders to complete this thesis!

For the hundreds of samples that were sent to us for this project, I wholeheartedly thank (in no order): Janette Norman and Wayne Longmore (Museum of Victoria), Megan Anderson and Daniel Burns (Southern Cross University), Charlotte Dunn and Diane Claridge (BMMRO), Leigh Hickmott (Open Ocean Consulting), Robin Baird (Victoria Whale Stranding Network, Cascadia Research Collective), John Calambokidis (Cascadia Research Collective), Zoe Luca (Sable Island), Carlos Olavarría (CEAZA), Jose Luis Brito Montero (Museo de Historia Natural e Histórico de San Antonio y Centro Rescate de Fauna), Nan Hauser (Center for Cetacean Research and Conservation), Ana Galov and Martina Duras (University of Zagreb), Daniel Palacios (Charles Darwin Research Station), Paul Jepson (ZSL and CSIP), Fabien Demaret and Florence Cuarant (CRMM), Alexandros Frantzis (Pelagos Cetacean Institute), Andy Foote (University of Copenhagen), Emer Rogan (UCC), Conor Ryan (IWDG, GMIT, Song of the Whale), Simon Berrow (IWDG, Shannon Dolphin and Wildlife Foundation), Dani Kerem (IMMRAC and University of Haifa), Marco Ballardini (BluWest and Woods Hole), Massi Rosso (CIMA), Jorge Urban (UABCS), Scott Baker (Oregon State University, University of Auckland), Tim Collins (Oman Whale and Dolphin Research Group), the late R. Natalie P. Goodall (RNP Goodall Foundation), Louella Dolar (Silliman University, Philippines), Robert Pitman (SWFSC), the late William Perrin (SWFSC), Jay Barlow (SWFSC), Kathryn Medlock (Tasmanian Museum and Art Gallery), Denise Hamerton and the late Peter Best (IZIKO Museums of South Africa), Luis Freitas (Madeira Whale Museum), Marisa Ferreira (Portugal Marine Animal Tissue Bank), Antonio Mignucci (Caribbean Stranding Network), Andrew Hebda (Nova Scotia Museum), Pablo Covelo (CEMMA), Natacha Aguilar de Soto (ULL), Bob Reid (Scottish Agricultural College), Nick Davison (SMASS), Vidal Martin (SECAC), Karen Evans (University of Tasmania), Andrew Dizon (SWFSC), John Durban (BMMRO and Southall Environmental Associates), Charley Potter and James Mead (National Museum of Natural History, Smithsonian Institution), Nelio Barros (Mote Marine Lab), Dan Odell and Judy St. Leger (Hubbs-SeaWorld Research Institute), Jim Dines, John Heyning and Dave Janiger (Los Angeles County Museum or Natural History), Rochelle Constantine (University of Auckland), Marc Oremus (WWF and South Pacific Whale Research Consortium), M. Michael Poole (Marine Mammal Research Program, French Polynesia), Kristi West (Hawaii Institute of Marine Biology), Emma Carroll (University of Auckland), Port Ellen Museum, National Marine Mammal Lab (Alaska), National Marine Fisheries Service (Alaska), NOAA Yakutat Salmon BD (Alaska), California Wildlife Centre, SeaWorld Florida, and Pacific Islands Regional Office (NMFS, Hawai'i).

Five years is a long time to work on a PhD, and there are so many friends and colleagues that kept me happy, sane, and inspired during that time. Firstly, thank you so much to everyone in the Gaggiotti Lab: Oscar Gaggiotti, Marie Louis, Éadin O'Mahony, Xinghu Qin, and Valeria Montano. We were a small little family, but I really enjoyed the time we spent together! Thanks also go to my good buddy Luke. Thank you for the endless chats about how amazing genetics is, for the visits to Cellardyke and for introducing us to Mollusc Mondays!

The luckiest people get to work with their friends, and I can safely say that my colleagues are the best in the world. Thank you to a VERY long list of people that brightened up my time in SMRU with innumerable chats over coffee breaks, pub nights, dinner parties, sleepovers and generally the best craic I could have asked for! I'll miss bumping into all of you: Debbie Russell and Matt Bivins, Matt Carter and Sarah Nelms, Chris McKnight, Steve Balfour and Saana Isojunno, Ryan Milne, Chris Morris, Simon Moss and Kimberley Bennet, Heather Vance, Katey Whyte, Laura Palmer, Janneke Ransijn, Alina Loth, Clint Blight, Rene Swift, and Kelly Robinson. I also want to acknowledge two great men who passed during this time: Bernie McConnell and Jeff Graves. Bernie, your enthusiasm, intelligence, and wit were something to behold and I am thankful for having known you. Dear Jeff, you were my first supervisor in St Andrews and helped me through the biggest academic failure of my life. Thank you for your warmth, kindness, and generosity with your time. I think I will forever see you and Ellen walking East Sands together in the mornings.

To my "Stich and Bitch" girls, thank you for inspiring me with your creativity, the hours of laughter, and for being just a wonderful community of strong women. I miss our meetups to craft and talk about life, love, and jobs. Thank you, Ailsa Hall, Monica Arso Civil, Nora Hanson, Kimberley Bennett, Alina Loth, Magda Chudzinska, Claire Lacey, Nadya Ramirez-Martinez, Tania Mendo. Lauren Arkoosh and Amy Milne.

To "Ma Girlz", I will never know a more supportive or loving group of strong, female scientists and you will never know how inspiring you all have been to me. Lots of love to Jo Kershaw, Holly Armstrong, Izzy Langley, Emily Hague, Heather Vance, Katey Whyte and Laura Palmer. Who run the world? Girls!

A most loving thank you goes to those who made our short tenure as west coasters in Oban full of fun and friendship. Tanya Houston and Jean Graham, you were the best neighbours we could have ever hoped to have while in lockdown. To me, you will always be the maidens of Glenshellach Farm, with your noble steeds: Jeff and Mac. To Ben and Mara Wilson, thank you for welcoming us into your bubble, showing us the beauty of Argyll, and for hours of fun meals and walks!

My son was born during this PhD and the sleepless nights and tired days were made a little bit better by the network of mom friends that were also navigating the world of postpartum and science. Thank you, Sarah Nelms, Monica Arso Civil, Heather Vance, Debbie Russel, Nora Hanson, Theoni Photopoulou, Marie Smedley, Amelia Saxby, Lauren Bierman, Chloe Malinka, Valentina Islas Villanueva, Anja Wittich and Valeria Montano. You are all superheroes!

To my dear friend Amelia Parish, thank you for being a wonderful and supportive friend for the past 11 years since we started our master's degree at the University of St Andrews. While our friendship has grown in distance (New Zealand is too far away!), it has also grown in love. For every 3 am text, hilarious meme, long-armed hug, and awkward dance, thank you for being you!

Dear Izzy Langley and Holly Armstrong, thank you for being such dedicated and loving friends to all of us Onions over the years. Your smiles and hugs make the world better, and your intelligence is inspiring. Your friendship and support during this thesis have meant so much!

To a work friend turned best friend and personal cheerleader, Jo Kershaw, I will always be thankful that I met you! You're the first person I want to share all of my accomplishments with because of the scream of delight that I know I will get in response. We never realised how good we had it living across the road from each other in Cellardyke and it brings me the most joy knowing you are finally just down the road again. Thank you for sharing your life and friends with me, and for showing me that its totally fine to be a bad ass scientist and a smoking hot woman. Here's to years of collaboration, great food and drink, fun times and "Despacito" dance parties!

Thank you to my two best friends and sisters I never had: Kari and Carley. We have been through everything together, and as sad as I am that we all live so far apart, there's not a second that goes by that I don't feel your love. From high school, to university, to new jobs, new loves and creating new lives, thank you for being there with me through it all.

I was extremely lucky to inherit an incredible family through marriage that has loved and supported me through this endeavour. Thank you, Caro, for all the trips to visit us in Scotland, the hours of brilliant conversation, the snuggly couch time with Zero and our whiskies to learn about British history, and for letting me school you that one (and only!) time when we visited the whale exhibit at Natural History Museum, I will never forget it! To my sister-in-law Chloe, thank you for not only being a loving and supporting sister, but also a beacon of knowledge, an inspiration in feminism, and a wonderful friend to share great experiences with. To my amazing in-laws, Mike and Alilee, thank you for loving and supporting me while I took on this project. You've both set a wonderful example of how to succeed in your careers individually and together as parents, and I hope to be able to conquer this challenge at least half as well has you have!

To my own family, thank you for loving and supporting me so well even while I have been so far away. Thank you Sam and Megan, for the love and encouragement these past 5 years and for making the long journey over to visit and see a bit of our life here! To my parents Neal and Kathy, there will never be the right words to express how much I love and respect you both. From the 3rd grade when I decided to be a marine biologist, you took me to every aquarium, signed me up for every summer camp, encouraged me through every internship, supported me through every degree, and cried at every airport when you dropped me off to fly far away home. The time apart will never get easier, but I hope you know how much I appreciate getting to have the adventure of a lifetime over here.

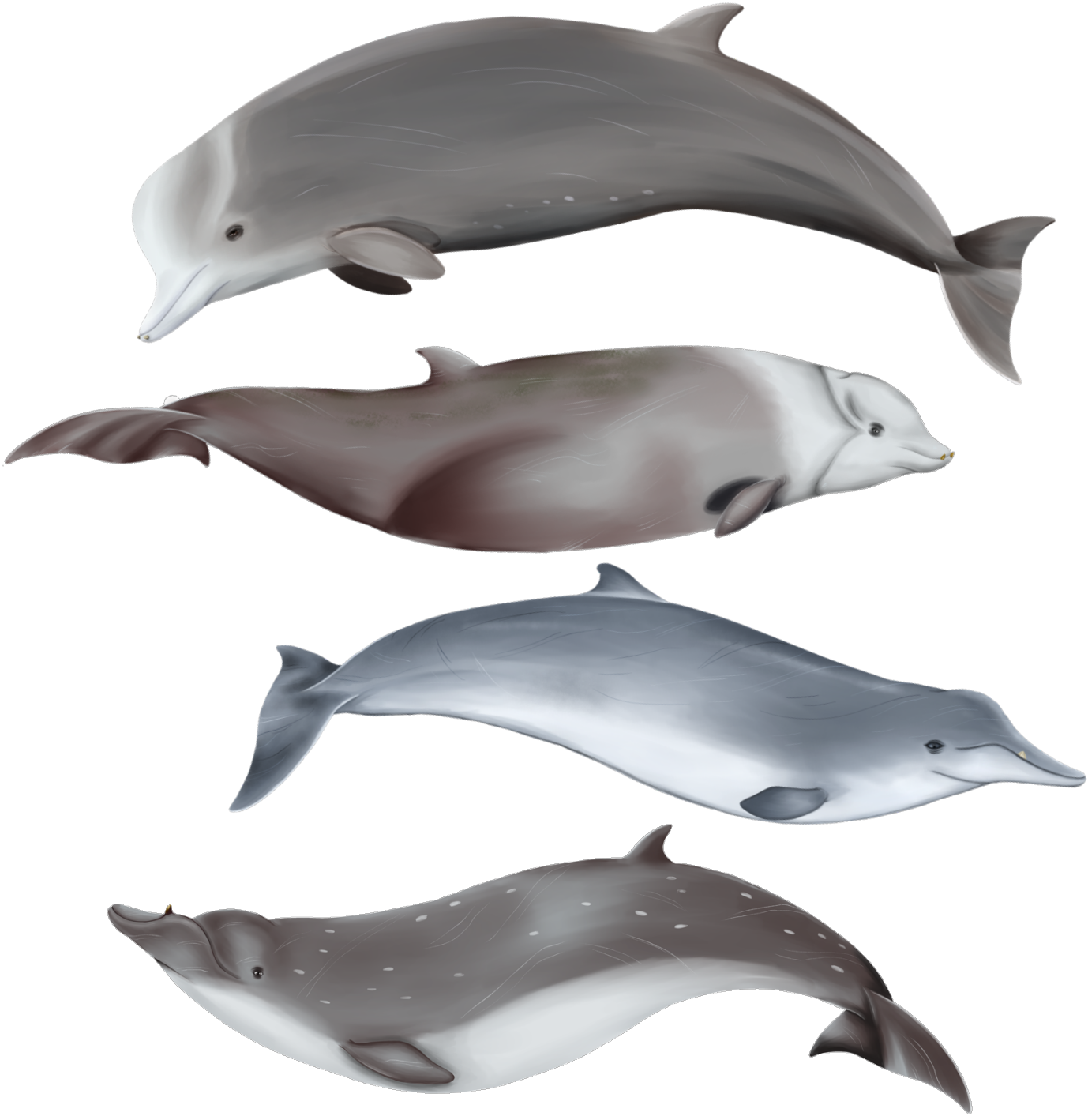
To my Joe: thank you. You have been the backbone of this family since the beginning, and not a day goes by that I don't wonder in disbelief how you are mine. Thank you for supporting me in every decision to get where I am today. You have provided for me in every way possible: with a home to come to when I moved to Scotland, by sharing your family with me when mine was another world away, and by always providing the stability so that I could take the risks I needed to take (like when I gambled and quit my permanent job to find a PhD!). You have been the best man, the best friend, the best partner and now the best father too, and I finally feel like we get to have a fresh start with this new post-PhD life. For everything you have done these past 6 months to keep me going and alive, thank you. Thank you for every meal, every cup of tea, every shoulder rub, every dog walk, every swim lesson, every errand, every bill, every hug, every pizza, every glass of wine, every load of laundry, every nursery run. You have held this family together and there is no way that I could have finished this thesis without you. Let's get our life started,

shall we? You should have plenty of time now to pursue a career as a professional whistler ☺

To Otis, you changed my life in every way possible and I will do everything I can to give you the love and support that I have been given. You make everyone around you smile and I am so proud that you are mine. Thank you for giving me a reason to stay afloat, and for bringing so much laughter to this final era of my PhD!

To my dear friend Gabriel Melo-Santos, thank you for sharing the gift of your beautiful art in this thesis.

I dedicate this thesis to my littlest Onion, Zero, and to the sweetest man I was blessed to call my father-in-law, Mike.



1 GENERAL INTRODUCTION

1.1 FAMILY ZIPHIIDAE

1.1.1 Beaked Whale Phylogenetics

The family Ziphiidae ("ziphiids") encompasses a monophyletic group of cetaceans called "beaked whales" that are found in all oceans around the world (MacLeod 2018). There are 24 recognized species, with three new species described in the last 20 years (Commitee on Taxonomy 2022). Although there have been important recent advances in our understanding of the diving, foraging and acoustic behaviour of some beaked whales, very little is understood about most species due to their often elusive behaviour at sea, offshore distributions, and deep-diving capabilities (MacLeod 2018). Most of the data available on beaked whale biology comes from individuals who have stranded, and some have no confirmed live sightings (Dalebout et al. 2008).

Ziphiids are one of the oldest families of odontocetes, appearing in the fossil record in the early Miocene, after the sperm whales and about the same time as the diversification of early hominids in Africa (Nikaido et al. 2001; Bianucci et al. 2016; Harrison 2017; Fordyce 2018). Using whole mitochondrial genomes (mitogenomes), the estimated origin for all beaked whale haplotypes was approximately 20.06 million years ago (mya, Morin et al. 2012). More recently, target sequence capture data suggested that beaked whale species began to diversify 15.6 million years ago (mya, 95% CI: 13.65-17.79 mya; (McGowen et al. 2019). Bianucci *et al.* (2016) conducted a global analysis of all fossil and extant ziphiids and found that they were well represented in the fossil record and most diverse through the middle Miocene (15.97-11.61 mya; Mead 1989; Mead 2008).

All extant and extinct ziphiids can be organised into two clades based on paleo-biogeographical analysis and phylogenetic relationships: Messapicetus (MC) and crown Ziphiidae (CZ). Of the 23 described genera of ziphiids, six remain and are all members of the CZ (*Mesoplodon, Ziphius, Tasmacetus, Indopacetus, Hyperoodon* and *Berardius*). Fossil studies have shown that four main convergent traits evolved in both the CZ and MC: reduction of dentition, vertex elevation, ossification of the rostrum in males, and an increase in body size (Bianucci et al. 2016).

In lieu of extensive ecological knowledge, the fossil record can help to examine the evolutionary patterns within the family (Bianucci et al. 2016). The reduction in dentition seen across ziphiids is likely an adaptation to suction feeding and the changing morphology of the face was most likely attributed to the optimisation of echolocation (Bianucci et al. 2016). Opposite to their current, non-polar distributions, fossil records indicate that the origin of *Mesoplodon* and *Ziphius* species was likely in the high latitudes of the Southern Ocean (Bianucci et al. 2016). Ziphiidae is the second most speciose family of extant cetaceans (after the delphinids), however in terms of the total number of described genera and species (extant and extinct), the ziphiids are the richest family of cetaceans (Bianucci et al. 2016).

The earliest molecular work involving beaked whale samples were surveys of meat market products from Japan and South Korea. Fragments of the mtDNA control region (CR, 550bp) and cytochrome B gene (CytB, 500bp) from whale meat were compared to a database of 33 baleen whale sequences and 18 odontocete sequences sourced from the literature (Baker et al. 1996). At least 8 different species were identified, one of which was identified as a beaked whale, but couldn't be identified further.

Table 1-1. Full list of recognized beaked whale species and overall length in meters based on Dalebout 2002; MacLeod 2005; MacLeod et al. 2006; Commitee on Taxonomy 2022).

Genus	Species	Common name	Authority	Range	Length (m)
Berardius	*arnuxii*	Arnoux's beaked whale	Duvernoy, 1851	Cold waters of Southern Hemisphere	6-9.3
	bairdii	Baird's beaked whale	Stejneger, 1883	North Pacific	10-11
	minimus	Sato's beaked whale	Yamada *et al.*, 2019	North Pacific	6.9
Hyperoodon	*ampullatus*	Northern bottlenose whale	Forster, 1770	North Atlantic	6.5-10
	planifrons	Southern bottlenose whale	Flower, 1882	Cold waters of Southern Hemisphere	6.5-7.5
Indopacetus	*pacificus*	Longman's beaked whale	Longman, 1926	Tropical Indo-Pacific	6-6.5
Mesoplodon	*bidens*	Sowerby's beaked whale	Sowerby, 1804	North Atlantic	5.5
	bowdoini	Andrew's beaked whale	Andrews, 1908	Cold temperate waters of Southern Hemisphere	4.4
	carlhubbsi	Hubb's beaked whale	Moore, 1963	North Pacific	5.3
	eueu	Ramari's beaked whale	Carroll *et al.* 2021	Temperate waters of Southern Hemisphere	5.3
	europaeus	Gervais' beaked whale	Gervais, 1855	Warm temperate & subtropical Atlantic	5.2
	ginkgodens	Ginkgo-toothed beaked whale	Nishiwaki & Kamiya, 1958	Warm temperate & tropical Indian and Pacific	5.3
	grayi	Gray's beaked whale	von Haast, 1876	Cold temperate waters of Southern Hemisphere	5.6
	hectori	Hector's beaked whale	Gray, 1871	Cold temperate waters of Southern Hemisphere	4.4
	hotaula	Deraniyagala's beaked whale	Deraniyagala, 1963	Tropical Indo-Pacific	4.8
	layardii	Strap-toothed beaked whale	Gray, 1865	Cold temperate waters of Southern Hemisphere	6.2
	mirus	True's beaked whale	True, 1913	Temperate North Atlantic	5.3
	periini	Perrin's beaked whale	Dalebout *et al.* 2002	North Pacific	4.4
	peruvianus	Pygmy beaked whale	Reyes *et al.*, 1991	Eastern tropical Pacific, western South Pacific	3.9
	stejnegeri	Stejneger's beaked whale	True, 1885	Cold temperate, subarctic North Pacific	5.7
	traversii	Spade-toothed beaked whale	Gray, 1874	Temperate South Pacific	5.3
	densirostris	Blainville's beaked whale	Blainville, 1817	Global, warm temperate & tropical	4.7
Tasmacetus	*sheperdi*	Shepherd's beaked whale	Oliver, 1937	Cold temperate waters of Southern Hemisphere	6-7
Ziphius	*cavirostris*	Cuvier's beaked whale	Cuvier, 1823	Global, temperate, subtropical, and tropical	7

In 1997, Henshaw *et al.* compiled the first database of specifically ziphiid mtDNA sequences for the use of identifying species taken in the California driftnet fishery. Sequences from 10 species were presented, with taxonomic confirmation based on skull morphology (Henshaw et al. 1997).

As molecular taxonomy has become a more widely accepted method of describing and discovering species, our understanding of beaked whale phylogeny has benefited greatly. Using molecular methods, four new species have been described (Reyes et al. 1991; Dalebout et al. 2002; Morin et al. 2017; Yamada et al. 2019; Carroll et al. 2021) and two species have been resurrected (VanHelden et al. 2002; Dalebout et al. 2014) in the past 30 years alone. Next-generation sequencing technologies are likely to contribute to this understanding even further (Cammen et al. 2016).

In 2004, a comprehensive validated beaked whale database was published, including a 437bp region of mtDNA CR, 384bp region of CytB, and complementary nuclear DNA actin intron sequences for 17 of the 21 then known species (Dalebout et al. 2004). Analysis of the sequences confirmed the monophyly of the family and accurate status of each species. The process of compiling this database and testing old samples from collections resulted in three interesting findings within the group that were published in the subsequent articles described below.

Three partial skull specimens collected in Chile and New Zealand were originally assigned to three different species, yet after mtDNA analysis all specimens were declared the same: spade-toothed whale (*Mesoplodon traversii*), a species that until 2010, was only known by these three partial skulls and considered the rarest of all living cetaceans (VanHelden et al. 2002; Thompson et al. 2012). Longman's beaked whale (*Indopacetus pacificus*) was described from two skull remains, but after two stranded individuals with photo records were linked by mtDNA sequences to these two specimens, new individuals were identified from across the tropical Indian and Pacific Oceans (Dalebout et al. 2003). Based on morphology, four individuals that stranded in California were identified as Hector's beaked whales (*Mesoplodon hectori*) and one Cuvier's beaked whale (*Ziphius cavirostris*), yet mtDNA analysis suggested they were all the same species, and different from all other beaked whale sequences (Dalebout et al. 2002). Later morphological studies showed subtle differences to Hector's beaked whales and these individuals were described as a new species, Perrin's beaked whale (*Mesoplodon perrini*).

In 2014, the total number of species within the genus *Mesoplodon* increased to 15 with the resurrection of Deraniyagala's beaked whale (*Mesoplodon hotauloa*) based on mitochondrial and nuclear DNA sequences (Dalebout et al. 2014). One beach-cast (El Hierro, Canary Islands) and one at-sea dead (the Azores) individual of the rarely observed True's beaked whale (*Mesoplodon mirus*) was confirmed using mtDNA fragments (Aguilar de Soto et al. 2017). Evidence from these individuals indicated a new southern limit of the species' disjointed, anti-tropical distribution and a new colouration pattern for the species. These new data allowed the species confirmation of other documented sightings and the first video recording of the species at sea. A later in-depth genomic investigation into the anti-tropical populations of True's beaked whales found that they actually represented two separate species, and the Southern Hemisphere population was raised to species level: Ramari's beaked whale (*Mesoplodon eueu*; Carroll *et al.*, 2021). A third species of *Berardius* was suggested based on morphology (smaller, black form) and mtDNA CR sequences (Morin et al. 2017). The 'black form' has been known by

Japanese whalers in the Northern Hemisphere (Omura et al. 1955) and genetic analysis revealed that these individuals were genetically more similar to the Southern Hemisphere Arnoux's beaked whale (*Berardius arnuxii*) than the sympatric Northern Hemisphere Baird's beaked whale (Morin et al. 2017). This 'black form' has now been raised to species level: Sato's beaked whale (*B. minimus;* Yamada *et al.*, 2019).

1.1.2 General Ecology

Beaked whales are so named due to the pronounced rostrum that is present in all species (Mead 2008). Diagnostically, all beaked whales have a spindle-shaped body, coming to a point at their diverse range of rostrum morphologies. Beaked whales are tooth-bearing odontocetes and display a variety of unusual dentition patterns, setting this family apart from all other cetaceans. This provides the primary distinguishing feature used to differentiate similarly sized species (particularly males) both at-sea and when stranded (Mead 1989; MacLeod 2018). In most cases, males' teeth are dramatically reduced to a single pair of "tusks" erupting from the lower jaw (with a few exceptions; Mead, 2008). This pattern occurs in all species of *Ziphius, Indopacetus, Hyperoodon* and *Mesoplodon.* Both sexes of *Berardius* have two pairs of teeth erupting from the lower jaw, and *Tasmacetus* has a full set of functional teeth in both sexes with two tusks erupting in adult males (Oliver 1937; Kirino 1956). Males and females of Gray's beaked whales (*Mesoplodon grayi)* commonly have small maxillary teeth (Thompson et al. 2014). Tusk morphologies vary extensively among species of *Mesoplodon* and this has been proposed to correlate with the amount of male-male social aggression, but sexual selection and species recognition are also likely contributors (Heyning 1984; Dalebout et al. 2008; Mead 2008). Females and sub-adults can be very difficult to distinguish to species-level due to a lack of characteristic dentition, however stranded, bycaught or biopsied individuals can be identified using DNA sequences (Dalebout et al. 2004).

Beaked whales have larger than average tail-flukes which typically lack a central notch (Heyning, 1989; Mead, 1989; MacLeod, 2018). Adults range from 4m (pygmy beaked whale, *Mesoplodon peruvianus*) to 12m (Baird's beaked whale, *Berardius bairdii*) in length (Table 1-1). The only species to exhibit sexual dimorphism in overall body size are Gervais' beaked whale (females are slightly larger, *M. europaeus*) and the northern bottlenose whale (a small proportion of the males are larger than the largest females, *Hyperoodon ampullatus;* MacLeod, 2005). The dorsal fin is typically small and located about 2/3 down the back. Pectoral fins are proportionally smaller than other species, tucking into depressions on either side of the body termed "flipper pockets" to increase streamlining while diving (True 1910).

Until recently, most of what is known about beaked whale foraging and diving behaviour came from stranded individuals and very few at-sea sightings (Cox et al. 2006). In recent years however, tagging technology has advanced and allowed fine-scale analyses of beaked whale diving behaviour to occur (Johnson and Tyack 2003; Johnson et al. 2004; Baird et al. 2006; Schorr et al. 2014). In 2003, Johnson and Tyack introduced the DTAG (www.soundtags.org) as a method to record environmental sound and animal vocalisations, as well as measure movement and orientation in space (Johnson et al. 2004). Using the data generated by DTAGs and other biologging devices, it has been found that all studied beaked whales have commonalities in their dive cycle (Aguilar de Soto et al. 2006; Tyack et al. 2006; Tyack et al. 2011; Allen et al. 2014). Typical dive profiles include a foraging dive to depths between 400 m and 3000 m for between 30 min to 120 min during both the day and night. Foraging dives are defined by a steep descent

to the deepest depth with echolocation starting at around 400-500 m (Johnson et al. 2004; Tyack et al. 2006). At the maximum depth, individuals spend 25-35 minutes actively foraging, using ultrasonic clicks (inter-click-interval of 0.2-0.4 seconds) to locate individual prey (Johnson et al. 2004). Echoes from targeted prey can be heard in DTAG acoustic recordings immediately prior to an increase in acceleration with click train ICI increasing to "buzzes" (250 clicks/second), both of which are behaviours assumed to be prey capture attempts (Johnson et al. 2004). The ascent rate is slower than descent rate, and deep dives are followed by a 1-2 hours long series of shallower dives ("bounce dives") to 100-400 m and 10-30 min duration each (Hooker and Baird 1999a; Cox et al. 2006; Tyack et al. 2006; R.W. Baird et al. 2008). Ascents are made at a low pitch angle and about half of the speed of descents, a strategy that is not yet understood (Tyack et al. 2006). The shallower "bounce" dives are silent, primarily likely to serve an anti-predator function and may or may not serve a physiological function (as they are not always seen in night dives; (Heithaus and Frid 2003; Tyack et al. 2006; Zimmer and Tyack 2007; Baird et al. 2008; Aguilar de Soto et al. 2012).

All beaked whale species are thought to be deep divers, preferring to live in water greater than 200 m (MacLeod et al. 2006; MacLeod 2018). Beaked whales use echolocation to find their prey (Johnson et al. 2004; Mead 2008) and the echoes off prey that have been recorded by DTAGs suggests that individuals are foraging for small, singular prey items (Johnson et al. 2004). Acoustic analysis of the frequency of ultrasonic "buzzes" indicate that approximately 30 prey capture attempts are made per dive (Johnson et al. 2004; Tyack et al. 2006). A lack of recorded diel differences in foraging dives suggests that beaked whales are feeding throughout the day and night on prey that does not migrate (Hooker and Baird 1999a; Baird et al. 2006; R.W. Baird et al. 2008; Schorr et al. 2014). However, Arranz *et al.* (2011) observed that the distribution of echolocation activity of Blainville's beaked whales during day and night dives was different, suggesting they may be following the migration of the deep scattering layer in some locations.

At-sea sightings indicate that beaked whales live in areas with underwater features that create upwelling and aggregate prey such as steep slopes, underwater canyons, and sea stacks, though they can also be found in the waters above the abyssal plane (MacLeod and D'Amico 2006; Baird 2018; Pitman 2018). At depth, beaked whales feed on meso- and benthopelagic prey using suction generated by a single pair of throat grooves extending from the lower jaw (Heyning and Mead 1996). The primary prey type for all beaked whales are schooling and solitary squid species, though some are known to eat fish and crustaceans (MacLeod et al. 2003; MacLeod and D'Amico 2006; West et al. 2017). Analyses of stomach contents and stable isotopes from stranded, bycaught and harvested whales have revealed some general trends in beaked whale diet (MacLeod et al. 2003; Whitehead et al. 2003; Riccialdelli et al. 2017; Smith, Trueman, et al. 2021). Differences in prey size found in sympatric species indicate that individuals likely occupy different ecological niches or segregate themselves temporally and/or geographically. For example, members of the genus *Mesoplodon* will co-exist with *Hyperoodon* and *Ziphius* whales by predating on generally smaller prey items, and *Ziphius* and *Hyperoodon* will eat the same prey type/size but segregate geographically due to competitive exclusion (MacLeod et al. 2003; Whitehead et al. 2003). Species within the *Mesoplodon* genus will segregate from each other spatially (Riccialdelli et al. 2017).

1.1.3 Beaked Whale Population Genetics

Since the early 2000's molecular methods have been used to investigate the population genetics of beaked whales. Evidence of genetic structure has primarily been identified from mtDNA and refined as sequencing technologies have improved or from the inclusion of nuclear DNA markers. Below, I will provide an overview of the findings from beaked whales that have been studied the most using molecular methods: northern bottlenose whales, Gray's beaked whales and Cuvier's beaked whale.

Using a 434bp fragment of the mtDNA CR, three variable sites (four haplotypes) were discovered between 45 northern bottlenose whales sampled in the Gully (near Nova Scotia, Canada), Labrador and Iceland (Dalebout et al. 2001). Within each region, there was very little genetic diversity detected, although there were significant regional differences (AMOVA, $p<0.05$ for F_{ST} and Φ_{ST}) between the Gully and Labrador stocks (not enough samples from Iceland to analyse). In 2006, this study was updated with more samples and 10 nuclear microsatellites (Dalebout et al. 2006). Genetic diversity based on microsatellite heterozygosity was similar among all three regions and most alleles appeared in all three populations. Pairwise comparisons using microsatellites and mtDNA indicated significant differences between the Gully and both Iceland and Labrador but no difference between Labrador and Iceland. Using next generation sequencing, whole mitochondrial genomes (mitogenomes) and nuclear genomes revealed the genetic distinctiveness of three populations (Scotian Shelf, Jan Mayen and western North Atlantic) and a strong pattern of isolation-by-distance (Feyrer et al. 2019; de Greef et al. 2022). Adding more molecular markers to these studies re-affirmed the need to protect the small population of animals present in the Gully, as they are likely to be a historically distinct population.

Population structure has been examined in Gray's beaked whale throughout most of their proposed Southern Ocean distribution. Thompson, *et al.* (2016a) examined a 530bp fragment of the CR and 12 microsatellites to test whether population structure exists using samples from stranded individuals collected from Australia and New Zealand. Thompson et al., (2014) found differences in morphology between Gray's beaked whales measured on the east and west sides of New Zealand, suggesting some form of reproductive isolation. Although there were high levels of genetic diversity (and no indication of a bottleneck), there was no measurable structure in either marker throughout the 6,000 km range sampled. Using complete mitogenomes and partial nuclear genomes, an absence of population structure was still found in 16 individuals from Australia, New Zealand and South Africa (Westbury et al. 2021). It is suggested that after diverging from their most recent common ancestor, Gray's beaked whales underwent a population expansion with abundant habitat, limited geographic barriers, and no directed exploitation by humans (Thompson, Patel, Baker, et al. 2016; Westbury et al. 2021). Gray's beaked whales are unique among ziphiids in that when they strand, it is often in large groups (the holotype was one of 28 in a mass stranding; von Haast, 1876). Based on this unusual pattern, Patel *et al.* (2017) used microsatellites and mtDNA haplotypes from 113 stranded animals to determine if groups of animals that stranded together were related. Although there were some cases of mother-calf pairs stranding together (n=6), there was no evidence that adults were related (average relatedness within groups was not significantly higher than between groups, $p<0.414$). The evidence presented in this paper suggests that both male and female dispersal occurs in this species.

Cuvier's are the widest ranging beaked whale for which a study of genetic structure and diversity has been done (Dalebout et al. 2005; Morin et al. 2012). Using a 290bp fragment of the CR, 87 samples of Cuvier's beaked whales collected around the world (Figure 1-1) clustered together in a monophyletic group within the ziphiid family (Dalebout et al. 2005). Four haplotype groups were found (derived from 33 haplotypes) and although there was sharing between ocean basins, an AMOVA (Analysis of Molecular Variance) could detect population differentiation, showing that 42% of the variance was due to ocean basin. All ocean basins were found to be significantly different from each other at the p=0.05 level, and within the North Atlantic, the Mediterranean was different from the other two regions.

To improve the resolution of Cuvier's population structure, and provide the first population genetics analysis of Blainville's beaked whales (*Mesoplodon densirostris*), a study of beaked whale phylogeography was conducted using whole mitogenomes (Morin et al. 2012). Three beaked whale species were investigated: Cuvier's (*n*=22), Blainville's (*n*=19) and Gervais' beaked whale (*n*=8) (Figure 1-2). Mitogenomes were analysed in their entirety, and by separating out three different partitions (12s + 16s, all coding sites, and CR; Morin et al. 2012). This study found that unlike most cetaceans where the putatively neutrally evolving CR is the most polymorphic, beaked whale mitogenomes were most polymorphic in the coding sites (Morin et al. 2012). Like Dalebout *et al.*, (2004), species-level relationships were highly supported by bootstrap. Within species, Blainville's assigned to two highly supported clades that were reciprocally monophyletic to ocean basin and likely diverged approximately 2 mya. Like Dalebout *et al.* (2005), Cuvier's genetic structure generally correlated with geographic distance, however there was no reciprocal monophyly of haplotypes to ocean basin. Three clades were identified in the Cuvier's samples, two of which contained individuals from both the Atlantic and Pacific Oceans. Although phylogenetic patterns in Cuvier's weren't as strong as Blainville's, like Dalebout *et al.* (2005), similar haplotypes did cluster geographically. Morin *et al.* (2012) suggest that mitogenomes from more geographic regions and the inclusion of nuclear loci are needed to better understand the population structure within Cuvier's beaked whales. The Gervais' samples were not analysed in this manner as all samples were from the same geographic region (Morin et al. 2012).

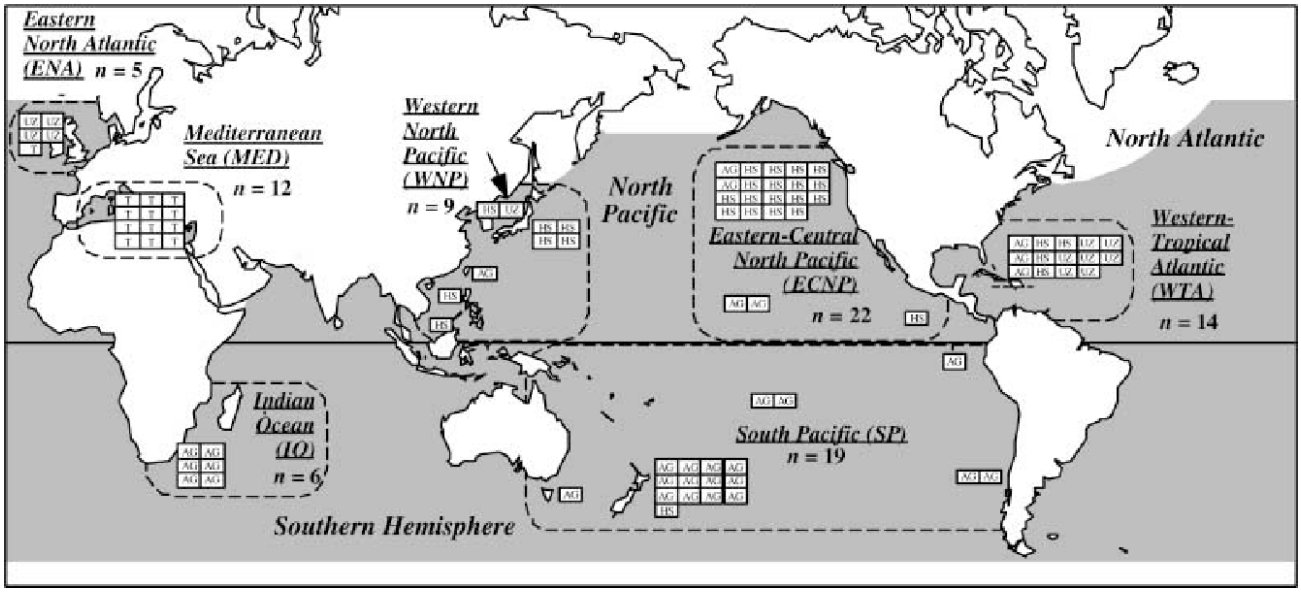


Figure 1-1. Map of samples and haplotypes from Dalebout et al., 2005.

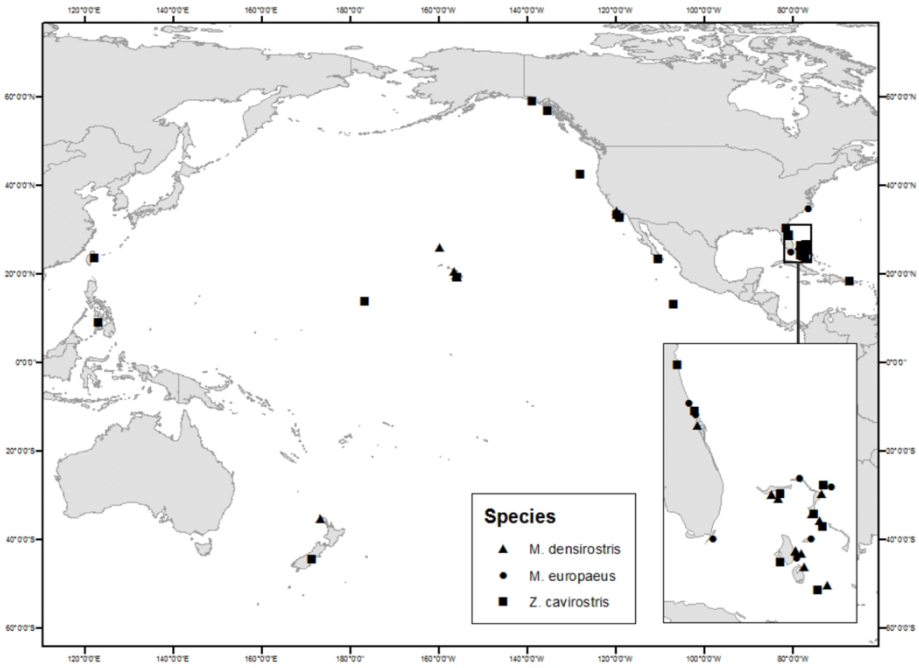


Figure 1-2. Samples from Morin et al., 2012.

1.1.4 Beaked Whale Social Structure

Due to difficulty in obtaining data on ziphiids, few long-term studies have been conducted at the population level (Baird 2019; Hooker et al. 2019). Those photo-identification studies that have been conducted have found small, persistent and resident populations, with very little connectivity between nearby locations (McSweeney et al. 2007; Claridge 2013; Aguilar de Soto et al. 2015; Reyes 2018). Though little is understood about beaked whale social structure, different species are thought to fall into two group size categories (MacLeod and D'Amico 2006). The first category consists of *Hyperoodon, Ziphius and* some species of *Mesoplodon*, where groups contain a maximum of 20 individuals and an average of 2 – 4. For example, northern bottlenose whales in north east Canada were identified in groups averaging 3 individuals, often with mixed sex (Gowans et al. 2001). The second category has a maximum group size of 100 individuals and includes *Berardius* (average group size of 8; Fedutin et al. 2015) and *Indopacetus* (average group size of 19 individuals; Pitman et al. 1999). Similar data are not available for *Tasmacetus.*

In areas where longitudinal photo-identification studies are taking place, more details about the social structure of these groups can be assessed (Baird 2019). In the Bahamas, Blainville's beaked whales (*Mesoplodon densirostris)* are thought to live in a polygynous harem society (average group size 4) where one adult male will associate with a group of adult females for up to a year (Claridge 2013). Females form associations for up to three years with preferred individuals, often of the same reproductive state (Dunn 2014). These females may or may not have calves, and males will join the group at times when the females are not ready to mate, suggesting that these harems are important for more than just mating. These types of interactions with animals moving in and out of groups is termed "fission-fusion". Without genetic studies of paternity however, the nature of this "social philopatry" will remain unknown. Blainville's and Cuvier's in Hawaii have been photo-identified in groups averaging 3.53 and 2.57 individuals, respectively (though not

significantly different, McSweeney et al. 2007). As with the Bahamas, the results of this study support a polygynous structure in Blainville's, however, too few data were collected from Cuvier's individuals to infer their breeding behaviour and social structure. In El Hierro (Canary Islands), a similar fission-fusion structure to what is seen in Hawaii and the Bahamas has been observed in Blainville's beaked whales (Reyes 2018). Cuvier's beaked whales in El Hierro also live in a fission-fusion society, however associations are shorter in duration and can change between dives rather than years (Reyes 2018). In each of these locations, females were re-sighted more frequently than males, suggesting some sort of difference in distribution and/or stronger site fidelity.

1.1.5 Anthropogenic Threats to Beaked Whale Populations

Major threats from anthropogenic activities have been identified across marine mammals, most of which can impact beaked whales too (Hooker et al. 2019; Nelms et al. 2021). Specific threats that have been highlighted for beaked whales include climate change, entanglement in fisheries gear, exposure to contaminants and the ingestion of plastics (Hooker et al. 2019; Nelms et al. 2021). The threat that has been most studied for beaked whales however is the impact of anthropogenic sound, such as military sonar (Parsons 2017; Hooker et al. 2019).

The northern bottlenose whale and members of the genus *Berardius* are the only ziphiids to be targeted in large-scale, commercial whaling industries, though other species have been taken in small fisheries or opportunistically (Heyning and Mead 2008; Mead 2008; Pitman 2018). Mead (1989) briefly describes that Japan, Taiwan and Newfoundland conducted small-scale harpoon fisheries that would take small cetaceans, including beaked whales, opportunistically. However due to the sporadic sightings and lack of knowledge of at-sea behaviour, beaked whales were primarily overlooked as a sustainable or commercially valuable resource.

There are numerous accounts of stranded beaked whales that have ingested both macro- and microplastics with varying levels of impact (Abreo et al. 2016; Alexiadou et al. 2019). In Brazil, a female Blainville's beaked whale was found with a stomach predominantly full of plastic and no evidence of recent feeding, suggesting the plastic in her stomach could have led to false satiation (Secchi and Zarzur 1999). When three True's beaked whales stranded in Ireland, both macro and microplastics were found throughout their digestive tracts, suggesting that these particles can pass through an individual. In this case, there was no sign of malnutrition as the small amount of plastic found was probably not enough to cause the false satiation hypothesised to have occurred in the Brazilian Blainville's (Lusher et al. 2015). The ingestion of plastic is known to be a problem in Cuvier's beaked whales with several incidents of impaction in the stomach and necrosis of the stomach lining (Gomerčić et al. 2006; Brownlow et al. 2015).

Sound is extremely important to all cetacean species (Nowacek et al. 2007) and like other odontocetes, beaked whales rely on their sensitive hearing underwater to echolocate for prey (Johnson et al. 2004). The frequency range of many anthropogenic sounds overlaps with the range emitted and perceived by beaked whales, potentially impacting them in a number of ways (Aguilar de Soto et al. 2006; Cox et al. 2006; Nowacek et al. 2007). The four primary concerns of animals exposed to loud sounds underwater are PTS (permanent threshold shift), TTS (temporary threshold shift), acoustic masking and behavioural disturbance (Nowacek et al. 2007). PTS and TTS are physiological changes that would be difficult to measure in free-ranging beaked whales. Acoustic masking and behavioural

disturbance however, have both been measured in beaked whales in response to various sources of noise (shipping traffic, predator calls, and sonar playbacks) using DTAGs (Aguilar de Soto et al. 2006; Tyack et al. 2011; DeRuiter et al. 2013; Allen et al. 2014).

As early as the 1970's, a connection has been suggested between certain military exercises and mass strandings of cetaceans, particularly beaked whales (van Bree and Kristensen 1974; Frantzis 1998; Balcomb III and Claridge 2001; Evans et al. 2001; Jepson et al. 2003; Cox et al. 2006; Nowacek et al. 2007). van Bree and Kristensen (1974) described four Cuvier's beaked whales that washed ashore in Bonaire, Lesser Antilles. Two days before the stranding, locals described hearing heavy explosions at sea, coinciding with a navy vessel dumping ammunition and causing underwater explosions. This study was the first to suggest a link between naval exercises and beaked whale mass strandings. Simmonds and Lopez-Jurado (1991) describe a series of multi-species mass strandings occurring in the Canary Islands between 1985 and 1989, all of which coincided temporally and geographically with at-sea military manoeuvres. The authors here suggest the naval activity may have driven the animals to shore. The first time active-sonar was suspected to be involved in beaked whale strandings was by Frantzis (1998). He correlated an atypical mass stranding of 12 Cuvier's beaked whales over 2 days (12-13 May 1996) and 38.2 km of Greek coastline with low-frequency active sonar tests being carried out by the NATO vessel *Alliance* from the 11-15 May 1996. The United States Government determined that a multi-species stranding event involving the deaths of at least 4 Cuvier's, 1 Blainville's and 1 spotted dolphin was linked to mid-frequency sonar use in the Bahamas in March 2000 (Balcomb III and Claridge 2001; Evans et al. 2001). Reviews of atypical strandings in conjunction with military activity have found linked events as far back as the 1960s (Jasny et al. 2005; Podestà et al. 2016).

Although the correlation between sonar and beaked whales is now recognized, the mechanism that causes the animals to strand is still not completely understood. The animals from the 2000 stranding in the Bahamas (Balcomb III and Claridge 2001) that were fresh enough to be necropsied were in good body condition with no sign of disease or contact trauma, yet all showed some type of auditory structural damage (Evans et al. 2001). One Blainville's beaked whale was the best preserved and showed bleeding in the kidney, lung and larynx, lesions in the heart, and bleeding in the inner ear and brain, symptoms providing strong evidence for acoustic trauma. It was concluded that some sort of acoustic insult (likely mid-frequency sonar) had caused a behavioural response leading to stranding, and the injuries from stranding were the cause of death. After a 2002 atypical stranding event coinciding with the *Neo Tapon* naval exercise in the Canary Islands, necropsy results from 8 Cuvier's, 1 Blainville's and 1 Gervais' beaked whale showed lesions consistent with *in vivo* bubble formation, similar to those seen in human decompression sickness (i.e. wide-spread microvascular haemorrhages, fat emboli in organs and intravascular bubbles; Jepson *et al.*, 2003; Fernández *et al.*, 2005). Jepson *et al.* (2003) suggest that such symptoms of decompression sickness could result from behavioural changes while diving (ascending quicker than usual) in response to sonar exposure.

Since the detection of decompression-like symptoms in stranded beaked whales, controlled play-back studies have attempted to explain the connection between sonar and strandings (Tyack et al. 2006; Zimmer and Tyack 2007; Tyack et al. 2011; DeRuiter et al. 2013; Allen et al. 2014). Although Baird *et al.* (2006) and Jepson *et al.* (2003) suggest ascending too quickly as the likely mechanism for mass strandings, beaked whale

responses to sonar playbacks show that both Blainville's and Cuvier's make longer and slower ascents away from the sound source during exposure (Tyack et al. 2011; DeRuiter et al. 2013). The response seen to sonar playbacks is similar, though less prolonged, than that seen to killer whale vocalisations, suggesting beaked whales are eliciting an anti-predator response (Tyack et al. 2011; Allen et al. 2014). Zimmer and Tyack (2007) developed a model to predict nitrogen saturation and to examine the effect on nitrogen tension and bubble growth from different diving behaviours. Based on the results, the authors suggest that the sustained anti-predator response consisting of repeated bounce dives (to below the limit of alveolar collapse) and directed swimming away from the source for the duration of the sound could result in decompression sickness. Repeated dives to this shallow depth (25 m-72 m) while moving away from the sound source may increase the horizontal distance travelled while underwater and keep the individual out of the water depths where their primary predators are found. The longer an individual continues to make these shallow bounce dives, the greater the predicted risk of decompression sickness. It has therefore been suggested that reducing the length of sonar transmissions may shorten this anti-predator response and reduce the risk of injury, yet it is still unknown if an animal discontinues the response as soon as the sound is gone (Zimmer and Tyack 2007).

Although beaked whale mass strandings are not particularly common (except Gray's beaked whales in New Zealand; Patel et al. 2017), they do raise a lot of public attention when they occur (Parsons 2017). During these events, a discreet number of individuals are in danger or are being removed, yet without understanding the underlying population structure and level of connectivity, it is impossible to know whether these strandings are causing long-term, population-level impacts.

1.2 THESIS OVERVIEW

1.2.1 Justification

As the first part of this review has highlighted, beaked whales are an incredibly diverse and elusive family, and little is known about population structure, connectivity, and genetic diversity for any ziphiid, on local or global scales. In recent years, beaked whales have come into the public eye due to their exceptional diving behaviour and dramatic behavioural responses linking certain anthropogenic sounds to mass stranding events (Cox et al. 2006; D'Amico et al. 2009). The recognized susceptibility to noise pollution and substantial knowledge gaps regarding abundance and population structure, strongly highlight that more information is required for effective management and conservation of these species. Highly publicised mass-stranding events clearly impact a discrete number of individuals but without data on population structure it is unknown whether they have long-term, population-level effects. Cuvier's are one of the most studied beaked whales in terms of publications, and although mounting evidence suggests population structure, the insufficient amount of data means very little has been officially recognized. Due to the recognized susceptibility and genetic distinctiveness of Mediterranean Cuvier's beaked whales, a "data-deficient" Mediterranean subpopulation has now been listed by the International Union for the Conservation of Nature (IUCN; Cañadas, 2012).

1.2.2 Study Species

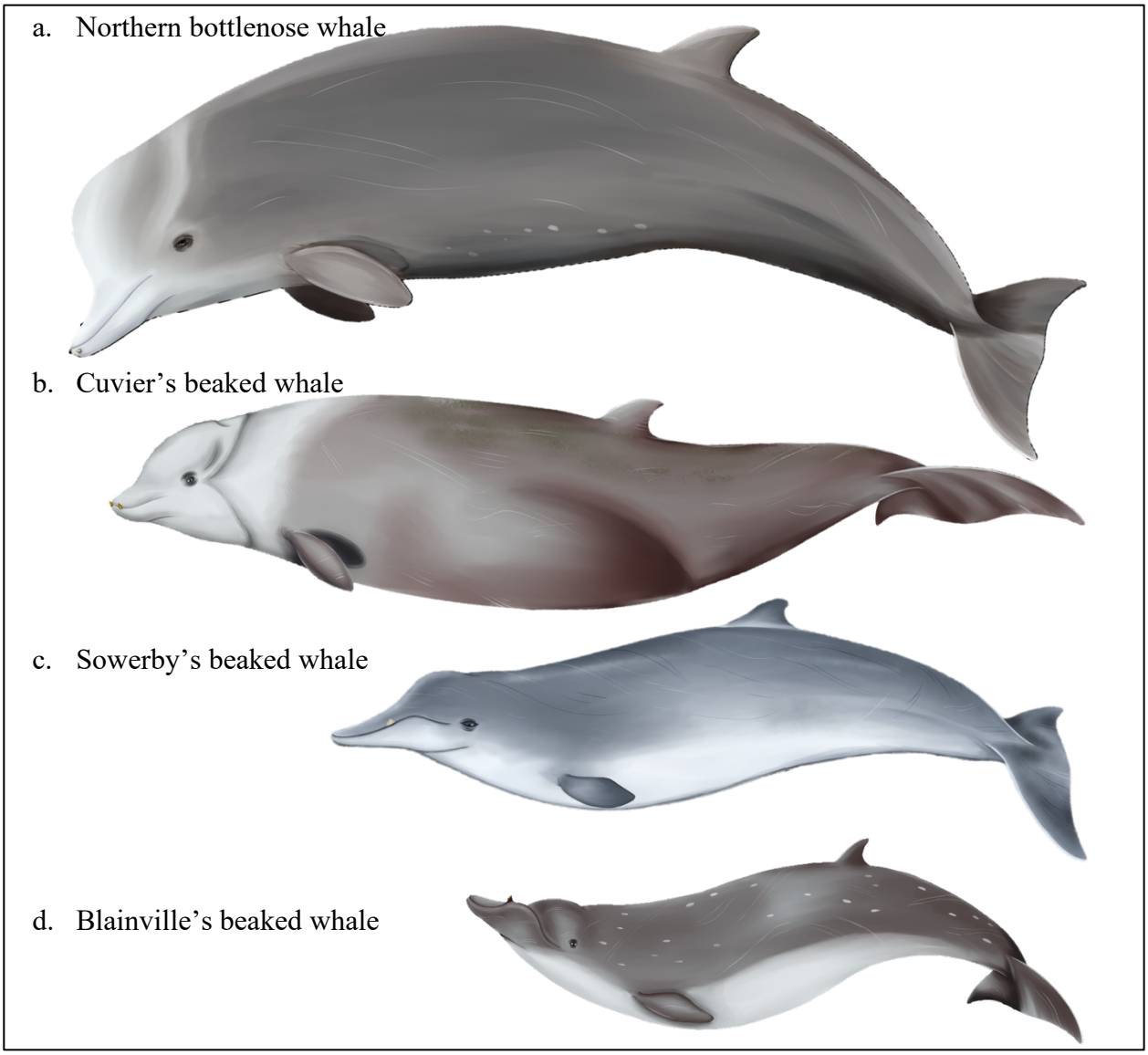


Figure 1-3. The four beaked whale species investigated this thesis and their relative sizes: a) Northern bottlenose whale (*Hyperoodon ampullatus*; 7-9 m), b) Cuvier's beaked whale (*Ziphius cavirostris*, 7 m), c) Sowerby's beaked whale (*Mesoplodon bidens*, 5.5 m), and d) Blainville's beaked whale (*Mesoplodon densirostris*, 4.7 m). Drawings by Gabriel Melo-Santos.

*1.2.2.1 Cuvier's Beaked Whale (*Ziphius cavirostris*)*

Cuvier's beaked whale (or "goose-beaked" whale) is the only species in its genus and was described by Cuvier in 1823 from a specimen that was mistakenly identified as a fossil due to its heavily ossified rostrum (Heyning 1989). Cuvier's reach lengths of 7m (Figure 1-3b) and the relatively blunt profile of their heads terminates at a small and poorly defined rostrum (Heyning 1989; Baird 2018). Like all ziphiid males, two tusks erupt from the lower jaw outside the gape of the mouth, however unlike some members of the *Mesoplodon* genus (mesoplodonts), they are located on the terminal end of the mandibles.

As with other ziphiids, Cuvier's are rarely sighted at sea due to the short amount of time that they spend on the surface (Heyning and Mead 2008). Although easier to identify to species level than mesoplodonts, sex is difficult to determine as the colour patterns are not consistently sexually dimorphic (Coomber et al. 2016). Adult males vary in colour

from nearly white to very dark and have a highly contrasted white cape extending from the melon to the dorsal fin. Most adult females are brown with a much less contrasting cape extending a third of the way down their back however, up to 1/3 of the females identified in some locations have displayed the same pigment pattern as adult males (Coomber et al. 2016). Adults of both sexes usually display a large number of circular scars caused by cookie-cutter sharks, however this varies among locations with different abundances of the sharks. Males will display linear scars as a result of male-male aggression using their tusks (Heyning 1989; Coomber et al. 2016; Baird 2018). Such scarring and pigmentation patterns facilitate the use of photo-identification studies to identify individuals within populations.

Cuvier's beaked whales have the most cosmopolitan distribution of all ziphiids based on stranding records and at-sea sightings (Figure 1-4). Individuals have been sighted in all oceans and seas apart from high polar areas, including resident populations in the Mediterranean Sea (Allen, Mead, et al. 2011; Coomber et al. 2016; Podestà et al. 2016). As with other ziphiids, at-sea sightings are usually in deep water or near continental slopes and individuals are usually sighted in small groups or singularly (McSweeney et al. 2007; Schorr et al. 2014; Baird 2018). Similar to Blainville's beaked whales, resident populations of Cuvier's in Hawaii, the Bahamas, the Canary Islands and also the Mediterranean have been identified and individual Cuvier's have been re-sighted over many years using photo-identification (McSweeney et al. 2007; Rosso et al. 2011; Reyes et al. 2012; Claridge et al. 2015).

There is no single estimate of population size or abundance trend in Cuvier's beaked whales, and many localized counts have had high levels of uncertainty (McSweeney et al. 2007; Taylor et al. 2008a; Cañadas and Vázquez 2014; Rogan et al. 2017; Baird 2018; Reyes 2018). Cuvier's are the most common beaked whales to strand suggesting they are probably not as rare as originally believed (Kenyon 1961; Heyning 1989; Allen, Brownell Jr., et al. 2011). The IUCN estimates that there are over 100,000 individuals (Baird et al. 2020) and classifies Cuvier's beaked whales globally as "Least Concern" (Baird et al. 2020). A sub-population has been assigned to Cuvier's found in the Mediterranean, and since most of the required data on abundance, trends, and structure are still unknown, these animals are listed as "Data-Deficient" (Cañadas 2012). Although there are known threats that could cause localized declines, the species has such a large range that it is unlikely that a 30% reduction of population size over three generations will occur (Baird et al. 2020). Never the target of large-scale fisheries, Cuvier's beaked whales were taken in both the Japanese *Berardius* fishery and the small cetacean fishery in Lesser Antilles (Allen, Brownell Jr., et al. 2011). Like Blainville's beaked whales, Cuvier's are also predated upon by killer whales (Heyning 1989) and have been discovered as bycatch in long-line and gillnet fisheries (Allen, Brownell Jr., et al. 2011).

Long-term photo-identification studies have revealed that this species lives in small, discrete populations with some degree of site fidelity (Hooker et al. 2019) and it is likely that the current species-wide conservation status is thus inappropriate (Allen, Brownell Jr., et al. 2011).

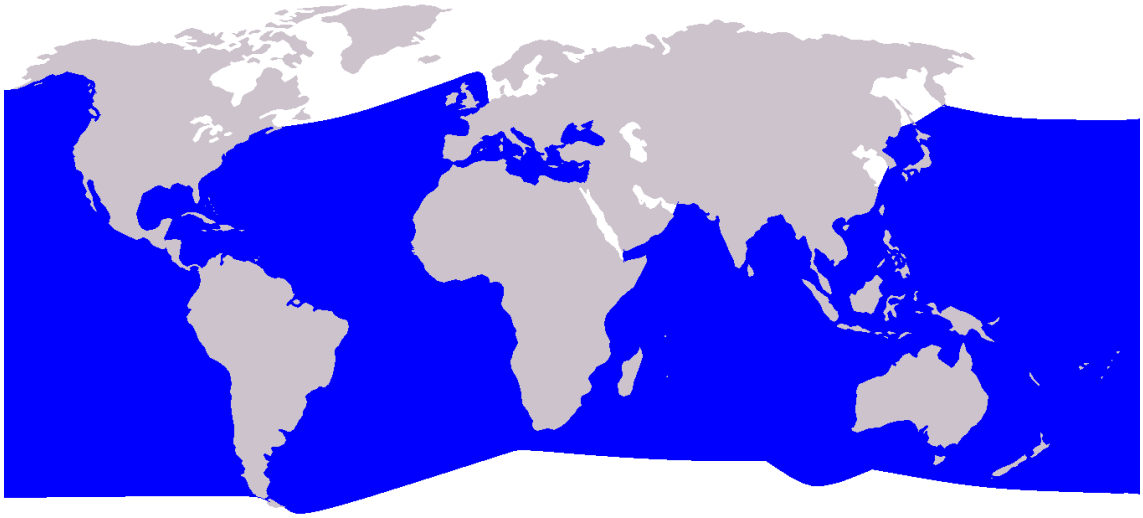


Figure 1-4. Known distribution of Cuvier's beaked whale (*Ziphius cavirostris*).

*1.2.2.2 Blainville's Beaked Whale (***Mesoplodon densirostris***)*

Blainville's beaked whales (or "dense-beaked" whale*)* were described in 1817 by Blainville and are a member of the most speciose genera of ziphiids (Mead 1989). In Blainville´s, males exhibit erupted teeth raised up on the bony arch of the lower jaw (Mead 1989). Sharp denticles form on the tip of the teeth, which may wear down over time and are most often encrusted with barnacles. Based on their tooth morphology, it is suggested that Blainville's are the most derived of the mesoplodont whales (Heyning 1984). The erupted tusks and heavily ossified rostrum are expected to be used in male-male combat (Heyning 1984).

Blainville's reach 4.7m in length (Figure 1-3d) and there is no evidence for sexual size dimorphism (MacLeod 2005). Females and sub-adults are difficult to differentiate as they are usually a non-descript grey/brown colour and lack distinctive erupted tusks (Mead 1989). Dominance battles lead to linear body scarring in males, and in conjunction with circular scars on both sexes from cookie-cutter sharks (Mead 1989), these marks provide an opportunity to identify individuals using photo-identification (McSweeney et al. 2007; Allen, Mead, et al. 2011; Reyes et al. 2012; Claridge 2013).

Blainville's also have a global distribution (Figure 1-5) and they are the most widely distributed of all mesoplodonts (Mead 1989; MacLeod et al. 2006; Pitman 2018). Mesoplodonts are difficult to distinguish to a species level and much of what is known about individual distribution comes from stranding records (Mead 1989; Allen, Mead, et al. 2011). Individuals are thought to associate in small groups (Mead 1989; Claridge 2013; Reyes 2018; Marques et al. 2019) and be distributed throughout all warm-temperature waters in the world and some colder waters, although they are notably absent from the Mediterranean Sea (MacLeod et al. 2006; Allen, Mead, et al. 2011). In some locations, e.g. Hawaii, the Bahamas and the Canary Islands, resident populations have been identified and individual whales have been re-sighted over many years (Aguilar de Soto 2006; McSweeney et al. 2007; Allen, Mead, et al. 2011; Reyes et al. 2012; Claridge 2013; Reyes 2018).

As with all other ziphiid species, there are no published global population size estimates of Blainville's beaked whales (Pitman 2018). Only a handful of populations have been well studied and overall estimates of abundance and population trends are unknown. The IUCN recently updated their listing from "Data Deficient" (Taylor, Baird, J. P. Barlow, *et al.*, 2008) to "Least Concern" (Pitman and Brownell Jr. 2020a). Although there are no documented observations, killer whales are likely predators based on tooth-scars and Blainville's behavioural response to killer whale vocalization play-backs (Mead 1989; Allen, Mead, et al. 2011; Tyack et al. 2011; Allen et al. 2014). There has never been a directed *Mesoplodon* fishery however, Blainville's have been taken opportunistically in Taiwanese small-scale, harpoon fisheries (Kasuya and Nishiwaki 1971; Mead 1989). Blainville's are also taken incidentally in gill-net and long line fisheries (Mead 1989; Pitman 2018).

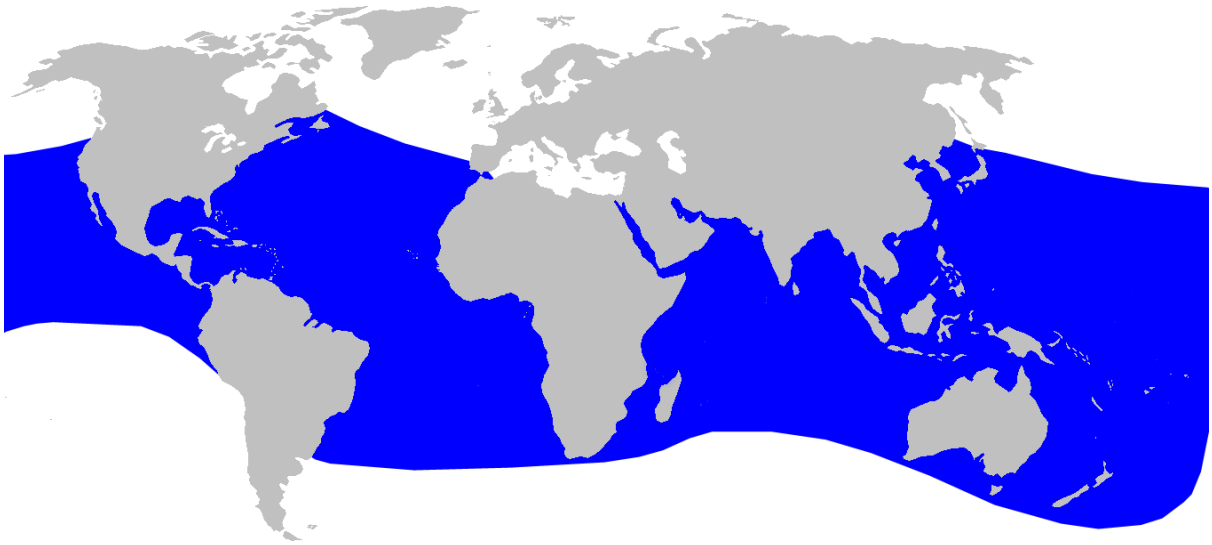


Figure 1-5. Known distribution of Blainville's beaked whale (Mesoplodon densirostris).

*1.2.2.3 Sowerby's beaked whale (***Mesoplodon bidens***)*

Sowerby's beaked whales were the first mesoplodont to be officially recognized. Originally classified in 1804 as *Physeter bidens* ("two-toothed cachalot"; Sowerby, 1804) based on the skull of an individual stranded in Scotland, Sowerby's share many of the same physical characteristics as other mesoplodont beaked whales (Ellis and Mead 2017a). Notably, Sowerby's reach approximately 5.5m in length (Figure 1-3c), have a longer beak length than many other mesoplodonts, and are typically dark grey in colour with some scarring (COSEWIC 2006; Ellis and Mead 2017a; Pitman 2018).

Occupying a medium-sized range, Sowerby's are the most northerly mesoplodont in the Atlantic with a distribution spanning from Massachusetts, USA to Labrador, Canada and across to Madeira and Northern Norway with some extralimital strandings and sightings in the Mediterranean, Gulf of Mexico and Brazil (Figure 1-5; MacLeod et al. 2006; Bittau et al. 2018; Pitman and Brownell Jr. 2020b). There's no information about migrations or site fidelity, though Sowerby's likely prefer deep water and/or continental shelves like other beaked whale species (COSEWIC 2006; MacLeod et al. 2006). With few external features to identify Sowerby's at sea, there are not many sightings of live individuals

(Ellis and Mead 2017a; Pitman and Brownell Jr. 2020b). In the handful of cases where live individuals have been sighted at sea, groups have included up to 8-10 individuals (Hooker and Baird 1999b) and been of mixed composition with males, females and calves together (Hooker and Baird 1999b; Berrow et al. 2018).

Few studies have specifically aimed to investigate Sowerby's due to the difficulty sampling or observing individuals at sea. The ontogeny of tusk and jaw formation was evaluated across the skulls of Sowerby's in Scotland, showing sexual and age-related dimorphism likely linked to intraspecific combat (Macleod and Herman 2004). Movement ecology and spatial structure were investigated in Sowerby's museum specimens using stable isotopes and morphological variation, suggesting that a metapopulation may structure individuals across their range (Smith, Trueman, et al. 2021; Smith, Mead, et al. 2021). Recently, two individuals outfitted with DTAGs showed for the first time that the foraging behaviour of Sowerby's differed greatly from similar-sized Blainville's, suggesting their ability to exploit different ecological niches and prey types using faster movements and echolocation clicks (Visser et al. 2022).

Like other beaked whale species, there are no estimates of abundance across the entire Sowerby's range. The species is classified as "Least Concern" by the IUCN Red List (Pitman and Brownell Jr. 2020b), where the justification for this category suggests that the species may encompass demographically independent populations (due to the wide geographic range), which would buffer potential risks to the species as a whole (Pitman and Brownell Jr. 2020b). In Canada, Sowerby's are classified as "Special Concern" due to the observed risk of anthropogenic sounds to closely related species (Cox et al. 2006; Filadelfo et al. 2009), and an overlap of these potentially harmful activities (specifically sonar) with the known range of Sowerby's (COSEWIC 2006). In the eastern North Atlantic, an estimate of n=3518 (95% CI: 1570-7883) Sowerby's was calculated based on only six observations during ship-based and aerial surveys in 2005 and 2007 (Rogan et al. 2017).

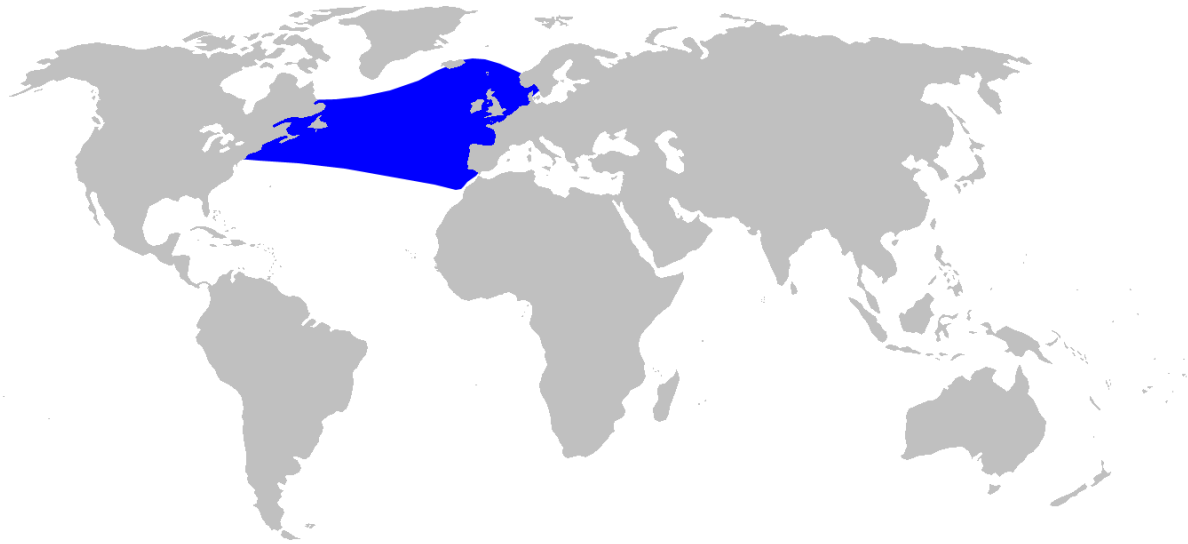


Figure 1-6. Known distribution of Sowerby's beaked whale (*Mesoplodon bidens*).

*1.2.2.4 Northern bottlenose whale (*Hyperoodon ampullatus*)*

Northern bottlenose whales were the first named extant beaked whale species (Forster, 1770) and are one of the best studied beaked whale species due to their heavy exploitation in the 19th and 20th centuries (Whitehead and Hooker 2012; Moors-Murphy 2018). Striking sexual dimorphism resulted in the proposal of two separate species for males and females by Gray (1860). Males are larger (9.15m vs 7.3m; MacLeod, 2005; Ellis and Mead, 2017; Figure 1-3a) and have more bulbous and dense foreheads likely for head-butting in male-male combat (Gowans and Rendell 1999). Colour patterns also vary between sexes with darker females and mottled brown to yellow males (Ellis and Mead 2017b; Moors-Murphy 2018).

Like Sowerby's beaked whales, northern bottlenose whales are exclusively found in the North Atlantic and range from the northeast USA up to Davis Strait in Canada, across to Iceland and Norway, and down to the Azores (Figure 1-6; MacLeod et al. 2006). Extralimital records include the Canary Islands and Mediterranean, as well as in the shallow North and Baltic Seas (MacLeod et al. 2006). Based on whaling data, northern bottlenose whales were concentrated in six areas including north east Canada, Greenland, Iceland and north Norway (Whitehead and Hooker 2012), with some genetic differences between them. Northern bottlenose have a fission-fusion social structure, where females form loose networks and males can have some long-term associations (Gowans et al. 2001). Group sizes tend to be smaller (<10) but larger groups have been seen (>20 individuals) (Ellis and Mead 2017b; Moors-Murphy 2018).

There is no single updated estimate for northern bottlenose whale abundance across their range, and the species is listed as "Near Threatened" by the IUCN Red List (Whitehead et al. 2021). Recent local estimates include *n*=143 individuals (95% CI: 129 - 156) in the Scotian Shelf (Canada) population (O'Brien and Whitehead 2013) and *n*=19539 (95% CI: 9921 – 38482) in the Faroe Islands and Europe (Rogan et al. 2017). The Scotian Shelf ("Gully") population has been heavily studied using photo-identification and genetics (Dalebout et al. 2006; O'Brien and Whitehead 2013) and is listed as "Endangered" by the Canadian government (COSEWIC 2011). The Davis Strait-Baffin Bay-Labrador Sea population is listed as "Special Concern" (COSEWIC 2011).

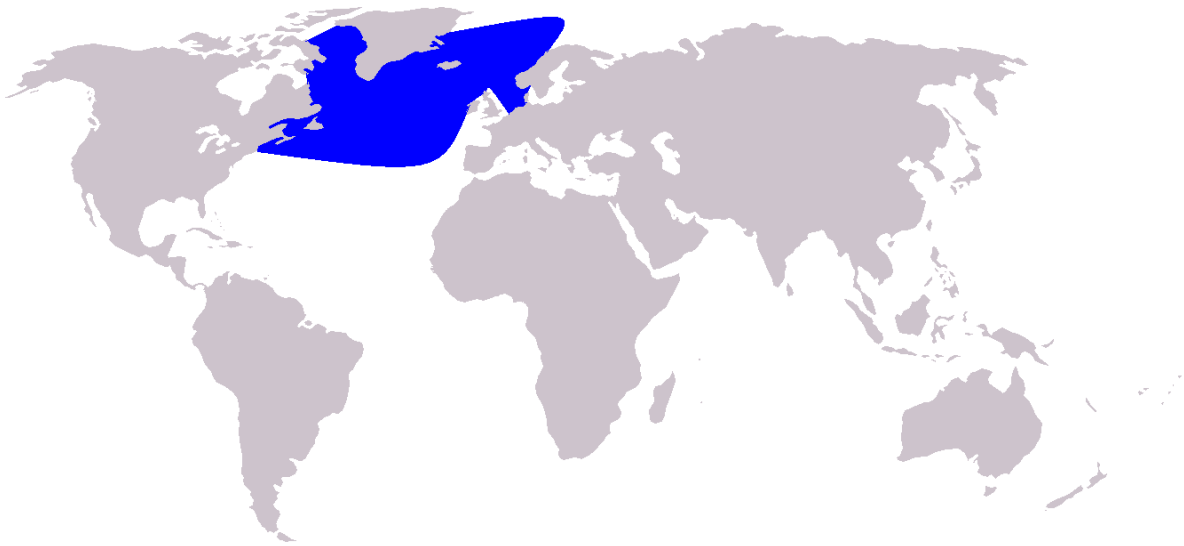


Figure 1-7. Known distribution of northern bottlenose whale (*Hyperoodon ampullatus*)

1.2.3 Molecular Ecology

> "Species are the result of descent with modification, and molecular characters are the direct archive of this history" (Dalebout et al. 2004, p. 471)

Improvements in DNA sequencing technology have opened up new avenues of research that incorporate molecular methods with ecology, advancing the field of "molecular ecology" (Monsen-Collar and Dolcemascolo 2010; Andrews and Luikart 2014; Hoelzel 2018). Within this field, molecular data are used to investigate questions across spatial and temporal scales, from individuals to species and communities (DeYoung and Honeycutt 2005; Monsen-Collar and Dolcemascolo 2010). Marine mammals are well suited for this field of study since many of their behaviours are hard to observe in the wild and incorporating genetic tools can provide a more comprehensive overview of the life history and population dynamics of species and individuals (Carroll and Garland 2022). Some examples of the use of molecular ecological methods in marine mammal science have been to inform management by defining species and conservation units, estimate life history parameters, investigate evolutionary hypotheses regarding adaptation to the aquatic environment, assess individual and population health and diet, and for forensic identification of market products (Baker et al. 1996; Cammen et al. 2016; Leroy et al. 2017; Waples et al. 2018; Zhou et al. 2018; Carroll and Gaggiotti 2019; Carroll and Garland 2022).

1.2.3.1 Population genetics theory

While the theories used to explain evolution have advanced greatly since Darwin first proposed it in 1859, the findings from some of the earliest theoretical publications still underpin the assumptions and analytic methods of population genetics (Darwin 1859; Charlesworth 2010). Namely, that genetic variation is driven by four evolutionary forces: mutation, natural selection, migration, and genetic drift. Mutations are the source of all new genetic variation in a closed gene pool. Natural selection concerns generational changes in phenotypes due to different survival or reproductive rates amongst individuals within a gene pool, and may lead to changes in gene frequencies. Natural selection can be directional (where one gene or genotype is consistently favoured), disruptive (where homozygotes are favoured), balancing (where heterozygotes are favoured), positive (selection for a particular genotype that increases fitness) or negative (when deleterious alleles/genotypes are removed from a gene pool). Migration, in this context, relates to the movement of individuals between populations and the exchange of genes between them (gene flow). Genetic drift is defined as the random changes in allele frequencies that occur from one generation to the next.

Early theoretical models provide a framework for understanding the complexities of evolution. One of the most used is the Wright-Fisher model, which is an idealised population that is not evolving, with discrete generations, random mating, and equal reproductive success (Fisher 1922; Wright 1931). The theoretical size of a population that is subject to the same strength of genetic drift as an idealised Wright-Fisher population is termed the effective population size (N_e) (Wright 1931; Wright 1938). When the ideal population and Wright-Fisher population share the same strength of drift, the Ne will be the same as the census population size (N_c). A Wright-Fisher population is expected to be in Hardy-Weinberg equilibrium (HWE), which means that allele and genotype frequencies will stay the same over time (Hardy 1908; Weinberg 1908). The equation that models the frequencies of alleles p and q at a single locus is $p^2 + 2pq + q^2 = 1$. Several assumptions are made about these theoretical models: the species is hermaphroditic, no

natural selection, no mutation, no migration, infinite population size (therefore no genetic drift), non-overlapping generations and random mating. While many of the assumptions are violated in wild populations, these theoretical models provide the framework to estimate genetic diversity, structure, and demographic changes over time.

The neutral theory of molecular evolution states that all mutations are random (Kimura 1991). The changes in allele frequency that occur by chance is termed genetic drift. Under neutral theory, most genetic variation on a molecular level does not convey a fitness advantage, and the fixation rate of these mutations differs from those that can cause amino acid changes (Kimura 1991).This theory provides the basis for most population genetic analyses, with the assumption that the molecular marker being used is under no selective pressure. In the case of SNPs used for demographic analyses, selection on a few loci when thousands of SNPs are being analysed is thought to not impact the inferences as long as any physical linkage among SNPs is negligible (Waples et al. 2018). Neutral loci located in autosomal chromosomes are suitable for most applications of population structure, diversity and demographic reconstructions, since they are distributed across the genome and have been similarly impacted by the evolutionary history of populations (Luikart et al. 2003; Johnson and Lachance 2012).

1.2.3.2 Population genetics applications: Beaked whales

Based on aspects of the population genetics theories described above, this PhD aims to characterize the genetic structure, diversity, relatedness, and demographic history of beaked whale species across a wide temporal and geographic range using next-generation DNA sequencing. Bi-parentally inherited nuclear DNA markers (single nucleotide polymorphisms, SNPs) were derived from double-digest restriction-site associated DNA (ddRAD) sequences (Peterson et al. 2012). Whole mitochondrial genomes were sequenced from a representative number of individuals (Duchêne et al., 2011; Morin et al., 2012), allowing for investigations of broader patterns of population structure.

1.2.3.2.1 Phylogenetics and population structure

The current management of most ziphiids is done on a species-wide level, though mounting evidence suggests that population structure within beaked whale species may be more common than panmixia (Dalebout et al. 2001; Dalebout et al. 2004; Dalebout et al. 2006; Feyrer et al. 2019; de Greef et al. 2022). Although beaked whales appear to be a monophyletic group (Dalebout et al. 2004; McGowen et al. 2019), advancing genomic technologies reveal the presence of new species more frequently than any other cetacean family (e.g., Morin et al. 2017; Carroll et al. 2021). Delimiting higher order groups such as species and subspecies can be done using phylogenetic trees, which may be based on sequence/genotype similarities or optimality methods (maximum parsimony or likelihood) (Waples et al. 2018).

Species are made up of populations with varying levels of isolation, and conservation and management of marine mammals on a species-level is considered to be insufficient as so many have wide-ranging distributions (Morin and Dizon 2018). Although there is not one single definition of a population (Waples and Gaggiotti 2006), it has been suggested that it would be more relevant to conserve the diversity within a species to a population or "stock" level (Reeves 2018). All definitions of population assume that some process unites individuals, and typically fall into either an ecological or evolutionary paradigm (Waples and Gaggiotti 2006). Within these two paradigms, the forces uniting individuals relate more to demography (individuals co-occur and can interact with each other; such

as “Demographically Independent Populations” (DIPs), Martien, Lang, et al. 2019) or genetics (individuals co-occur and can breed with each other; such as “Evolutionarily Significant Units” (ESUs), (Waples 1995), respectively (Waples and Gaggiotti 2006).

Several methods are available to investigate genetic structure in cetaceans (see Waples et al. 2018), and for this thesis, population structure was characterised using clustering and without assigning populations *a priori* using the R package *tess3r* (Caye et al. 2018; R Core Team 2019). Analyses using *tess3r* are fast and allow you to incorporate the geographic coordinates of each sample to help guide the formation of clusters (see 2.7.1 for a more in-depth review). Once biologically meaningful genetic clusters were identified, the commonly used fixation index (F_{ST}) was calculated between each population pair (Wright 1931; Wright 1951; Wright 1965; Weir and Cockerham 1984). This index is calculated based on allele frequencies and provides a measure of divergence between two populations ranging from 0 (no differentiation between populations) and 1 (each population is fixed for a different allele). The statistical significance of F_{ST} values can be addressed by bootstrapping and calculating the 95% confidence intervals around the F_{ST} point estimates and the p-value. There are some limitations to F_{ST} (Jost 2008; Jost et al. 2018; Waples et al. 2018), but it is still recommended to infer population structure in cetacean populations due to the validity of its theoretical foundation and the wide use of it across wildlife studies (Waples et al. 2018).

1.2.3.2.2 Genetic diversity

In wild populations, there is a positive correlation between population size, genetic diversity and fitness (Hansson and Westerberg 2002; Bouzat 2010; Hoelzel 2018). Genetic diversity is an important measure of adaptive potential and is also linked to an individual’s or population’s resilience to environmental change (Reed and Frankham 2003; Bouzat 2010; Leroy et al. 2017). Longitudinal photo-identification studies of several beaked whale species show they form resident populations (Gowans et al. 2001; McSweeney et al. 2007; Reyes et al. 2012; Claridge 2013; Aguilar de Soto et al. 2015). Living in small and socially-complex groups may indicate that beaked whales could naturally exhibit low levels of genetic diversity, thus increasing the risk of genetic erosion by the removal of individuals due to human impacts (Leroy et al. 2017).

Many metrics are available to investigate the genetic diversity of populations and species, which are typically calculated from neutral DNA markers (nuclear microsatellites or SNPs) and mtDNA haplotypes (Waples et al. 2018). In this thesis, observed heterozygosity (H_o; frequency of heterozygotes per locus) and gene diversity (H_s; the expected probability that an individual will be heterozygous at that locus) were calculated in Chapters 3-5 for each ddRAD SNP locus and averaged within populations (Nei 1987a; Goudet 2005). Both measures estimate the evenness of genetic diversity within a population (Hoban et al. 2022). Allelic richness (AR, mean number of alleles per locus calculated through rarefaction) was also averaged within populations in Chapter 5 using ddRAD SNPs, as it can provide an indication of adaptive potential (Hurlbert 1971; El Mousadik and Petit 1996; Goudet 2005; Hoban et al. 2022). To estimate diversity of the mtDNA in Chapter 3, the number of segregating sites between haplotypes (S) and the number of unique haplotypes (h) were calculated per population, as well as the haplotype diversity (Hd, an indication of the uniqueness of a haplotype in a population) and nucleotide diversity (π, the average number of differences between sequences) per population (Rozas et al. 2017).

1.2.3.2.3 Inbreeding and Relatedness

To maintain genetic diversity within a population, it is useful to understand the diversity within individuals. When inbreeding occurs, genetic diversity is lost due to individuals sharing genetic material that is derived from a common ancestor. The negative impacts of inbreeding on fitness (inbreeding depression) are due to the accumulation of deleterious recessive alleles, and evidence of this has been seen across taxa (Keller and Waller 2002; DeWoody et al. 2021; Hohenlohe et al. 2021).

Inbreeding can be quantified using metrics such as the inbreeding coefficient (F_{IS}), which was calculated for each population according to the method by Weir and Cockerham (1984) in Chapters 3-5. Individual-level metrics of inbreeding are also available, including Internal Relatedness (IR, Amos et al. 2001), an estimate of parental relatedness, which was calculated in Chapter 5.

1.2.3.2.4 Demographic history

Contemporary molecular data can tell us about the evolutionary history of past populations using analyses based on coalescent theory (Salmona et al. 2017). The coalescent method is a framework to reconstruct and trace the alleles in a population back through time to a common, ancestral allele (Charlesworth 2009). Based on this theory, several analytical techniques have been developed to estimate changes in effective population size backwards through time. Many such programs require that an underlying demographic model be specifically tested (i.e. N_e changes, population divergence, migration, etc), however some allow for a relaxed estimation of N_e changes overtime without a pre-defined demographic model (Salmona et al. 2017).

In this thesis, demographic histories were investigated using two methods. Firstly, Tajima's D was calculated directly from the SNP VCF files in Chapters 3 and 4 (Tajima 1989; Danecek et al. 2011). Tajima's D is a neutrality test statistic that can detect if a population is under any selective pressure. In practice, a population is in mutation-drift equilibrium when the Tajima's D value does not differ significantly zero. When the Tajima's D value is significantly positive, this indicates that there is a lack of rare alleles, a potential indicator of a recent population bottleneck. A positive Tajima's D could also be an artefact of mixing individuals from different populations, which would look like balancing selection. When the Tajima's D value is negative, there is an excess of rare alleles, which can indicate that a population expansion is occurring after a bottleneck or a recent selective sweep.

Demographic histories were reconstructed for each population from unfolded site frequency spectrum (ancestral allele unknown, SFS) files using Stairway Plot 2 in Chapter 4 (Liu and Fu 2020). The resulting plots from this analysis show changes in N_e overtime, which are calculated from theta (θ, the output of the Stairway Plot 2 calculations) and species-specific mutation rate per generation (μ) using the following equation: $\theta=4N_e\mu$.

1.2.3.3 Population genomics and wildlife conservation

Improved sequencing technologies mean that genome-wide data are available for potentially any species, even in the absence of a reference genome (Hohenlohe et al. 2021). Access to these genomic resources has provided conservation biologists with additional tools to assess the biodiversity of this planet (Hohenlohe et al. 2021). One of

the foundations of biodiversity is genetic diversity, as it is correlated with the structure, function and resilience of ecosystems (DeWoody et al. 2021; Hoban et al. 2022). Genetic markers are a particularly powerful tool to investigate biodiversity, as the timescale of the potential inferences can go back thousands of generations and be based on data from a single individual (Hoban et al. 2022).

There are still many knowledge gaps and challenges to overcome to fully implement genomic methods in wildlife conservation (Hohenlohe et al. 2021; Hoban et al. 2022). For example, while specific metrics have been proposed to investigate genetic variation of populations and species (Hoban et al. 2022), some are difficult to calculate with insufficient sample sizes and some are best to evaluate when there is data from more than one time-point. Another great challenge will be to standardise the methods for sampling, generating and analysing the data (Hoban et al. 2022).

Beaked whales are one of the most speciose families of cetaceans and some of largest predators in the deep sea, and therefore the viability and persistence of these populations will be important for maintaining the overall biodiversity and ecosystem function of this planet. Global collaborations and the use of archival specimens will be necessary for calculating conservation metrics to assess the health of populations and species across their sometimes-wide distributions.

1.3 THESIS OUTLINE

This thesis is made up of six chapters, to provide a comprehensive investigation of the ecology and evolution of beaked whales across time and space, with an aim to provide updated data for conservation and management applications.

1.3.1 Chapter 2: General Methods

Here I provide an in-depth explanation and justification for the laboratory, bioinformatic, and analytical methods used to conduct the analyses across the three investigative chapters. I also provide the methods and results from four optimisation studies done.

1.3.2 Chapter 3: “Biogeography in the deep: Hierarchical population genomic structure of two beaked whale species”

The primary objective of this chapter was to investigate the global population genetic structure of Cuvier’s and Blainville’s beaked whales and estimate whether current management units for conservation are appropriate. Using SNPs derived from ddRAD sequencing data and whole mitogenomes, I characterised population structure across the wide ranges of two cosmopolitan species and investigated structure at a finer spatial scale than in previous studies. Using this data, I proposed new management units to conserve in the form of ESUs and DIPs. This chapter was peer-reviewed and published in the journal *Global Ecology and Conservation* in October 2022 and is presented in its peer-reviewed, pre-print format. I was the lead author of this work, and the list of co-authors and their contributions are described in more detail in Chapter 3.

1.3.3 Chapter 4: “Demography in the deep: Reconstructed demographic histories of North Atlantic beaked whales

In this chapter, I investigated the population structure and demographic histories of North Atlantic beaked whale species to infer their response to future climate change. Specifically, I used SNPs generated from four species found in the North Atlantic:

Cuvier's, Blainville's and Sowerby's beaked whales (using ddRAD sequencing) and northern bottlenose whales (genotyping-by-sequencing, GBS) to calculate and compare demographic histories in the context of past glacial events. As both Sowerby's and northern bottlenose whales are only found in the North Atlantic, I hypothesised that periods of extended historic climate change would reduce their populations more than the globally distributed Cuvier's and Blainville's, who would have access to lower latitude areas of refuge.

1.3.4 Chapter 5: "Conservation in the deep: Genetic Essential Biodiversity Variables in 'disturbed' and 'semi-pristine' beaked whale populations"

In this chapter I investigated whether population-level impacts of repeated anthropogenic disturbance are evident in small beaked whale populations based on genomic diversity metrics. To do this, I calculate four genetic Essential Biodiversity Variables (EBVs) for paired 'disturbed' and 'semi-pristine' beaked whale populations in three regions to estimate genetic variation. In this chapter, I predicted that the 'disturbed' populations would have stronger genetic drift (smaller N_e), lower genetic diversity and higher inbreeding. I also generated kin networks, for visualising and comparing the relatedness of individuals within beaked whale populations.

1.3.5 Chapter 6: General Discussion

In the final chapter, I synthesized the findings from the data chapters and highlighted the key contributions of my methods to population genomics. I finished the chapter with a discussion on the future directions that beaked whale genomics should take.

2 GENERAL METHODS

2.1 INTRODUCTION

To meet the primary objectives described in Chapter 1 of investigating the ecology and evolution of an understudied group of non-model organisms, this PhD combined several laboratory and analytical methods to process and analyse high-throughput sequencing data. In the following chapter, I will present the foundational components of this thesis (the available tissue, DNA, and genomic sequences) as well as a thorough explanation of the laboratory and analytical methods used, including their justification when appropriate. In the final section, I provide details of the steps that I took to optimise parameters and test the suitability of various methods prior to the final analyses in the accompanying data chapters.

2.2 AVAILABLE GENOMIC RESOURCES

2.2.1 Genomic Databases

Several international databases store and maintain repositories of genomic data for open access use by researchers. Currently, there are two which house genomic data for beaked whales. Genbank® is maintained and distributed by the National Centre for Biotechnology Information (NCBI) and houses all publicly available sequence data that has been generated across the widest range of species and sequence types in the world (Benson et al. 2005). The DNA Zoo database (https://www.dnazoo.org/assemblies) stores high quality reference genomes that have been generated using the Hi-C sequencing approach (Dudchenko et al. 2017) for as many species as possible to aid in wildlife conservation.

2.2.2 Whole Nuclear Genomes

As of 13 August 2022, the following whole nuclear genomes were available in open-access databases for beaked whales:

Genbank:
Sowerby's beaked whale (*Mesoplodon bidens*, GCA_004027085.1)
Cuvier's beaked whale (*Ziphius cavirostris*, GCA_004364475.1)
Northern bottlenose whale (*Hyperoodon ampullatus*, GCA_024363105.1)

DNA Zoo:
Blainville's beaked whale (*Mesoplodon densirostris*, https://www.dnazoo.org/assemblies/Mesoplodon_densirostris)
Gervais' beaked whale (*Mesoplodon europaeus*, https://www.dnazoo.org/assemblies/Mesoplodon_europaeus)
Stejneger's beaked whale (*Mesoplodon stejnegeri*, https://www.dnazoo.org/assemblies/Mesoplodon_stejnegeri)

Draft whole genomes of Cuvier's and Sowerby's beaked whales were provided by Jeremy Johnson of the Broad Institute before they were made available on Genbank. These genomes were sequenced as part of the NIH funded project, "The 200 Mammals Project: sequencing genomes by a novel cost-effective method, yielding a higher resolution annotation of the human genome", a large-scale collaborative study comparing whole genomes from >200 species of mammal. The genomes were assembled with large contigs and little scaffolding but due to varying quality of DNA from the voucher specimens, the genome qualities vary accordingly. The Sowerby's genome was considered good quality

(N50=41908) however the Cuvier's genome is heavily fragmented (N50=7140). At the time the genomes were being used to align the sequencing data in Chapter 3 of this thesis, they were still undergoing the process of being uploaded to NCBI (https://www.ncbi.nlm.nih.gov/genome) and it was advised that any edits to the genome would have been extremely localised and not affect the alignments. These genomes have since been uploaded to Genbank and the accession numbers are provided above.

2.2.3 Whole Mitochondrial Genomes

Whole mitochondrial genomes (mitogenomes) were available from 236 individuals of 12 ziphiid species (Table 2-1) on the NCBI Genbank database as of 13 August 2022. While I did not conduct any analyses using mitogenomes for this thesis, collaborators at the University of Copenhagen used mitogenomes from Cuvier's and Blainville's to provide complementary genetic structure and diversity estimates in Chapter 3.

Table 2-1. Available whole mitochondrial genomes of family Ziphiidae on NCBI Genbank database.

Genus	**Species**	***n***
Berardius	*bairdii*	2
Hyperoodon	*ampullatus*	138
Indopacetus	*pacificus*	2
Mesoplodon	*bidens*	2
	densirostris	20
	eueu	2
	europaeus	9
	ginkgodens	2
	grayi	23
	mirus	10
	stejnegeri	2
Ziphius	*cavirostris*	24
Total		236

2.2.4 Collaborator Data

Phylogenetic trees were generated for Cuvier's and Blainville's in Chapter 3 based on SNP genotypes generated from double digest restriction-site associated DNA (ddRAD) sequencing. To root these trees, ddRAD sequences generated using the same laboratory protocol were provided by Dr. Emma Carroll from Southern right whales (*Eubalaena australis*).

The DNA sequence data for northern bottlenose whales were generated using genotyping-by-sequencing" (GBS) by the project collaborator Dr. Morten Tange Olsen at the University of Copenhagen. DNA was extracted from these samples at the Centre for Geogenetics, Natural History Museum of Denmark using either a Qiagen DNeasy® Blood and Tissue kit or Thermo Scientific™ KingFisher™ Cell and Tissue DNA kit. DNA was sent to the Cornell University Institute of Biotechnology for GBS library preparation and sequencing following the methods of Elshire *et al.* (2011). The raw sequencing reads were incorporated in the comparative analysis of North Atlantic beaked whale species presented in Chapter 4.

2.3 SAMPLE COLLECTION AND ARCHIVE

2.3.1 Sample Collection Methods

Tissue samples were either collected specifically for this project or provided on loan from archives maintained by contributors. A large set of samples was provided by Dr. Merel Dalebout, who used these samples for both her PhD thesis and subsequent publications (Dalebout et al. 2002; Dalebout 2002; VanHelden et al. 2002; Dalebout et al. 2003; Dalebout et al. 2004; Dalebout et al. 2005; Dalebout et al. 2006; Gomerčić et al. 2006; Dalebout et al. 2007; Dalebout et al. 2008; Dalebout et al. 2014). All samples were either skin biopsies collected directly from free-swimming animals, tissue collected from dead beach-cast or ship-strike individuals, or bone collected from dead stranded or museum specimens. Already-extracted DNA samples were provided by NOAA's Southwest Fisheries Science Centre Marine Mammal and Turtle Molecular Research Sample Collection (https://www.fisheries.noaa.gov/west-coast/science-data/marine-mammal-and-sea-turtle-research-tissue-collection). This DNA was extracted from biopsies collected from live whales, or tissue from dead beach-cast, ship-strike or entangled individuals. Freshly collected tissue samples were typically stored in DMSO or 70-99% ethanol and stored at -20°C. A full list of the samples that were sequenced for this thesis and their associated metadata are available in Appendix A. The organisation and/or individual that provided the sample is presented, as well as any animal welfare or ethical considerations required to collect tissue from deceased animals.

2.3.1.1 Tissue collection via skin biopsy sampling

Individual research institutions maintain their own protocols and regulations with regards to the collection of biopsy samples from live animals based on their own experiences, animal ethics and welfare guidelines, and government regulations. Typically, biopsy samples were collected using a small dart deployed by cross bow, adjustable-pressure modified air-gun, black powder gun or a pole (methods reviewed in Noren and Mocklin, 2012). In some instances, sloughed skin was collected after the deployment of telemetry devices such as the suction-cup mounted DTAG on free-ranging individuals (Miller et al. 2004). The specific procedures for close boat approaches and skin biopsy sampling are determined by the animal ethics approved protocols that each research group operates under, which may determine the number of individuals, their sex, age and breeding status, and the methods for targeting and approaching individuals for biopsy sampling. Typically, groups are approached slowly in small boats and sampled when they are within 5-30 meters away. The reaction of the sampled individual, and sometimes the whole group, is recorded on a standardised scale (Noren and Mocklin 2012), following the approved animal ethics protocol the fieldwork is conducted under (Appendix A).

2.3.1.2 Tissue collection from dead whales

Most samples in this thesis were collected opportunistically from already deceased animals that were either beach-cast, dead at sea, or bycaught. Most of the beach-cast animals were sampled by members of organised stranding networks with the aim of collecting data on the health of individuals and wild populations, while also allowing for investigations into the biology and life history of often hard to study marine mammals (Appendix A).

2.3.2 Ethical Considerations and Approval

The use of samples for this project was approved by the University of St Andrews School of Biology Ethics committee and the Animal Welfare and Ethics Review Body (AWERB). All research using vertebrate animals is strictly regulated in the UK under the Animals (Scientific Procedures) Act 1986 (ASPA) and the University of St Andrews adheres to the principles of the NC3Rs initiative (https://www.nc3rs.org.uk/). This PhD project involved the use of samples collected from live and dead animals overseas and tissue collected from dead-stranded specimens in the UK and is thus considered a non-ASPA study which did not require a license from the UK Home Office. Ethics approval for this study was provided for the umbrella project supervised by Prof. Oscar Gaggiotti titled "Population genetics and molecular ecology of cetaceans" (see Ethics Approval letter, page v). All samples collected or donated by project collaborators, including skin biopsy samples and tissue collected from dead beach-cast or ship-strike individuals, were done so under appropriate local permits and if applicable, local animal ethics approvals as described in Appendix A.

2.3.3 International Tissue Archive for Beaked Whales (ITABW)

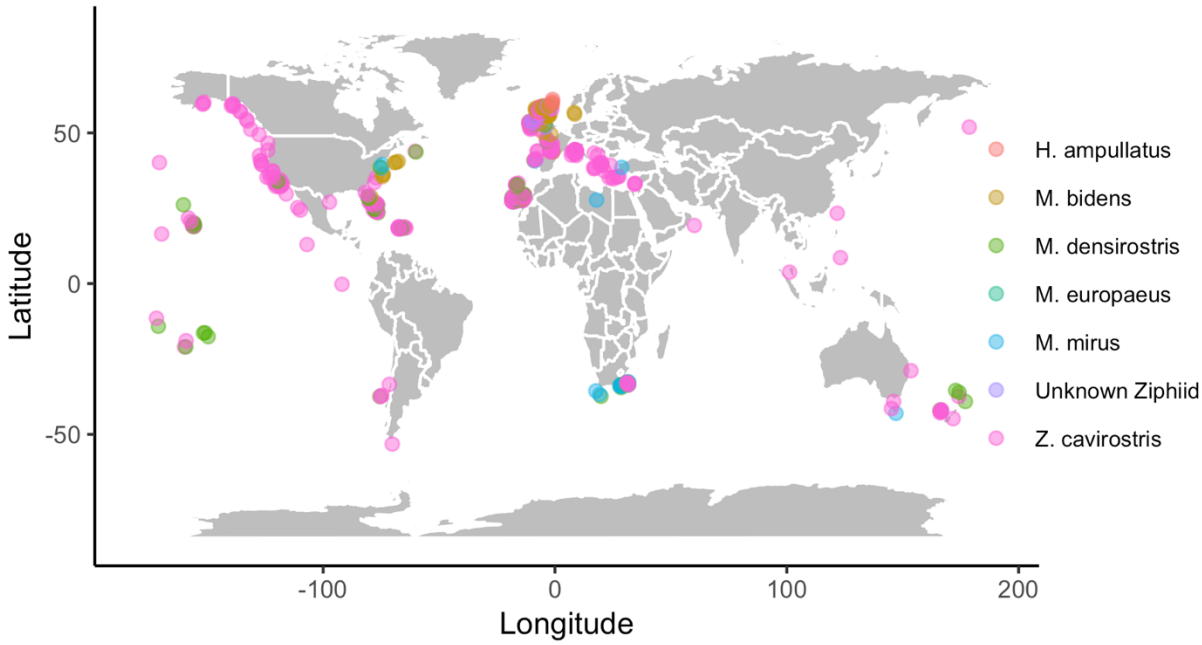


Figure 2-1. Map of all tissue (and DNA from NOAA) samples submitted to the International Tissue Archive for Beaked Whales (ITABW) for northern bottlenose whales (*H. ampullatus*), Sowerby's (*M. bidens*), Blainville's (*M. densirostris*), Gervais' (*M. europaeus*), True's/Ramari's (*M. mirus/eueu*), and Cuvier's beaked whale (*Z. cavirostris*). Some samples were also provided without species identified and are presented as "Unknown Ziphiid".

Between samples that were provided directly by collaborators on the Office of Naval Research (ONR) project that funded this PhD and from other contacted research, government and stranding organisations, this project has amassed an extremely large, international, beaked whale tissue sample collection. Due to the uniqueness of this dataset and rarity of beaked whale samples, an international repository for beaked whale samples (tissue and DNA) was created to be maintained between three institutions: the University of Auckland Waipapa Taumata Rau (maintained by Dr. Emma Carroll), the University of Copenhagen (maintained by Dr. Morten Olsen) and the University of La Laguna (maintained by Dr. Natacha Aguilar de Soto). The newly named "International Tissue

Archive for Beaked Whales", or ITABW, will facilitate access to these samples by future researchers and ensure that none of these rare samples are ever lost or forgotten. This global initiative aims to increase the knowledge of the molecular ecology of these poorly understood species. In the future, it will be easy to submit any tissue collected from a beaked whale to one of the three archives, where they will be subsampled (if enough tissue is provided), archived in the appropriate medium and temperature, and any future use will require evaluation and approval by a small committee. A map with all samples in the ITABW database is presented in Figure 2-1.

2.4 DNA EXTRACTION

2.4.1 Method

The Phenol:Chloroform:Isoamyl Alcohol method of DNA extraction as described by Sambrook, Fritsch and Maniatis (1989) was used in this thesis, with modifications allowing for greater recovery of DNA from small and difficult to digest samples (Baker et al. 1994), such as skin collected from free-ranging whales. This well-established method was suitable for the current study as it yields the greatest quantity and quality DNA, which is required for downstream next-generation sequencing applications.

DNA was extracted from approximately 30-50mg of tissue (preferably from the skin-blubber interface). Tissue samples were manually scored with sterile razor blades and digested overnight at 55°C in a mixture of 450µl SET buffer (53.9mL H2O, 0.1mL 0.5M EDTA, 5mL 1M Tris pH 8, 1mL 5M NaCl), 25µl SDS and 20-40µl Proteinase K. Following digestion, 500µl Phenol was added to each tube and rocked for 10 minutes before being spun down at 10,000rpm for 10 minutes. In this step, the phenol removes and absorbs the proteins, leaving a biphasic mixture with DNA, salts, and other contaminants in the top layer. After centrifugation, the top layer was removed and added to 500µl of Phenol:Chloroform:Isoamyl Alcohol (P:CI) at a 25:24:1 ratio. Samples were again rocked and centrifuged at 10,000rpm for 10 minutes each. The upper layer containing the DNA was removed and added to 500µl CI (24:1 ratio), rocked and centrifuged at 10,000rpm for 10 minutes each. The resulting supernatant with all proteins, salts and contaminants removed was moved to a new tube with 50µl of 3M sodium acetate and 1.2mL of isopropanol or pure ethanol (if samples were low quality/quantity) and inverted gently. If DNA precipitated immediately, the samples proceeded directly into the ethanol washes. If no visible precipitate occurred, 0.5µl linear polyacrylamide was added to each sample and then stored overnight at -4°C. After precipitation or the overnight storage, samples were spun at 13,000rpm for 10 minutes, pelleting the DNA at the bottom of the tube. All liquid (isopropanol or ethanol) was removed, 1mL of 70% absolute ethanol was added and samples were spun again at 13,000rpm for 10 minutes. This step was repeated one more time. With most of the ethanol removed, samples were left to dry completely on the bench before elution in 50-100µl of Qiagen EB buffer. At this stage, the DNA samples were checked for quality/quantity before laboratory analyses or archival at -4°C.

2.4.2 Sample Quality Control and Scoring

The extracted DNA was checked for quality and quantity using a NanoDrop™ (Thermo Scientific ™) spectrophotometer and gel electrophoresis. The NanoDrop™ measures sample purity based on the ratio of absorbances at 260nm/230nm (~2.0-2.2) and 260nm/280nm (~1.8). Excessive deviations from these numbers can indicate carryover of contaminants such as phenol and ethanol from the extraction. Although fluorometric

methods such as the Qubit™ (Invitrogen™) are more accurate at quantifying precise amounts of double stranded DNA (dsDNA), spectrophotometry measurements by the NanoDrop™ (Thermo Scientific™) are sufficient for estimating the approximate quantity. Extracted DNA quality with regards to its integrity (state of degradation) can be checked using agarose gel (Figure 2-2). High molecular weight DNA will show a thick and well-defined band, which indicates that the DNA has not fragmented into smaller pieces. Degraded DNA will show up as a smear.

As the protocol for ddRAD requires both high molecular weight and high concentration DNA (>20ng/ul), a scoring system was developed to rank samples (Table 2-2) prior to preparing libraries for pooling and sequencing. DNA was run on 1.2% agarose gels to assess the overall quality of the sample and the concentration was measured using a NanoDrop and Qubit. Samples selected based on their DNA score were pooled into libraries with individuals that shared the same score whenever possible.

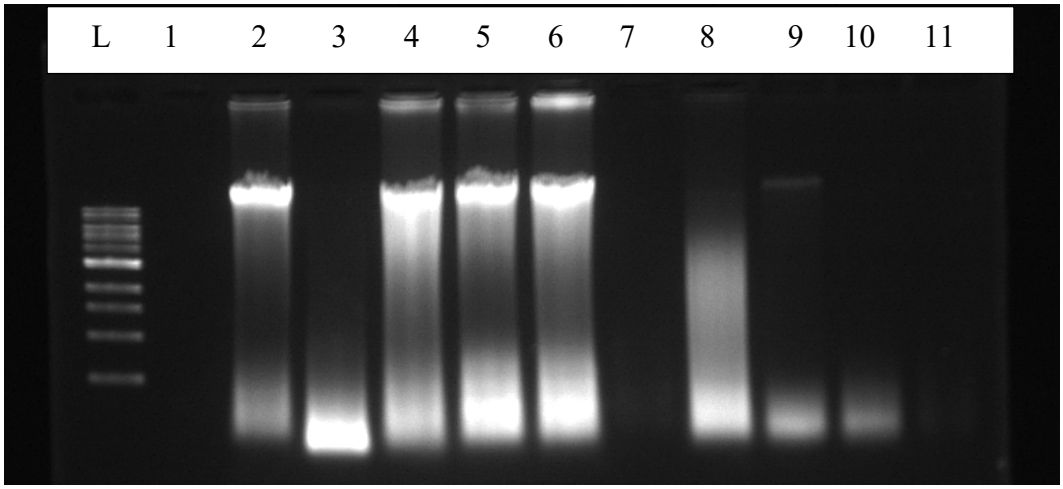


Figure 2-2. Example gel image of recently extracted DNA. In the 12 lanes above, the first has the 100bp ladder and the following lanes contain extracted DNA. Lanes 1, 7 and 11 have no visible DNA; lanes 3, 8, and 10 have only very fragmented DNA; and lanes 2, 4, 5, 6 and 9 show a band of high molecular weight (HMW) DNA with some fragmented DNA.

Table 2-2. The scoring system developed to rank DNA samples before ddRAD library preparation based on the amount of DNA in the sample (measured using a Qubit fluorometer) and the molecular weight according to 1.2% agarose gels (HMW =High Molecular Weight, Smear=degraded DNA of varying sizes, LMW= Low Molecular Weight).

Score	DNA Concentration (Qubit)	DNA Gel Result
Great	>20ng/ul	HMW
Good	>20ng/ul	HMW + smear
Good	>20ng/ul	Smear
Good	>20ng/ul	No visible DNA
Good	15-20ng/ul	HMW
OK	>20ng/ul	LMW
OK	15-20ng/ul	Smear
OK	<15ng/ul	HMW
Some	<15ng/ul	Faint HMW
Some	<15ng/ul	Smear
Some	<15ng/ul	No visible DNA

2.5 GENOMIC LABORATORY METHODS

2.5.1 Double-digest Restriction Associated Digest Sequencing

Restriction site-associated DNA sequencing, or RADseq, was first described by Miller *et al.* (2007) out of a need to identify and genotype polymorphisms of non-model organisms in a high-throughput, yet cost-effective manner. The original use of the method generated species-specific microarrays and Baird *et al.* (2008) went on to adapt the method for Illumina sequencing, where one run could do the equivalent of hundreds of RAD arrays. This method is massively parallel due to the multiplexing of samples with individual barcodes, and allows for the discovery and genotyping of single nucleotide polymorphisms (SNPS) to occur in the same step (Baird *et al.*, 2008).

In the original RADseq protocol (Baird *et al.*, 2008), one restriction enzyme (RE) was used to cut the genome at a rarely occurring motif. P1 adaptors were ligated to the RE cut site that contain primers for Illumina sequencing as well as a 4-5bp nucleotide barcode (short DNA sequence that can be used downstream to de-multiplex pooled samples). Once the adaptors were ligated, the DNA fragments were pooled together, mechanically sheared with a sonicator and broadly size-selected using gel electrophoresis. The final step before sequencing ligated a Y-shaped adaptor (P2) to the other end of the fragment, ensuring that fragments are only sequenced from the P1 end.

ddRAD (**D**ouble-**D**igest RAD) sequencing builds upon this method by adding a second RE to the digest and uses explicit size selection, allowing researchers to have more control over the fraction of the genome that is sequenced (Peterson et al. 2012). In this protocol, samples were digested with two REs, one that targets a commonly occurring motif (4bp) and a rarely occurring motif (6bp). Unique P1 adaptors containing individual barcodes and PCR primers, and universal P2 adaptors with PCR primers, were ligated to both ends of the digested DNA. The samples were then cleaned, pooled and size selected using a Pippin Prep. A PCR step added a secondary identifier (index) to the P2 end and Illumina flow cell annealing sequences on both ends. After this step, samples can be cleaned, pooled, and sequenced.

These methods have revolutionized the field of genomics for non-model organisms by dramatically increasing the ability to quickly and cost-effectively discover and genotype thousands of polymorphisms in the absence of a reference genome (Andrews et al. 2016; Cammen et al. 2016; Lowry et al. 2017). By massively increasing the number of markers per individual, the resolution and statistical power is great enough to conduct studies in ecology and evolution of species or populations that are difficult to sample or show little genetic diversity using traditional genotyping approaches (Cammen et al. 2016). The method described by Peterson *et al.* (2012) has many advantages over the single digest RADseq protocol by Baird *et al.* (2008). By adding a second RE digest, and eliminating random mechanical shearing and broad size selection, individual studies can be much more reproducible and precise. Limiting the DNA window that is sequenced and precisely size-selecting, means that the sequenced fragments from different individuals are more likely to be recovered from the same region of the genome (Peterson et al. 2012).

To assess the usefulness of ddRAD for beaked whales, Carroll *et al.* (2016) conducted a pilot study using four samples each of Cuvier's and Blainville's beaked whales. Following the ddRAD protocol of Peterson *et al.* (2012) with modifications, it was found that approximately 10,000 variable loci (likely to be a sufficient number to detect

structure in beaked whales) at 30x coverage could be obtained with >4 million paired end reads. All four Blainville's samples in this pilot study were collected from El Hierro, Canary Islands and showed no genetic differentiation using STRUCTURE (Pritchard et al. 2000; Carroll et al. 2016). The four Cuvier's samples had a wider geographic origin (*n*=1 Scotland, *n*=1 Ligurian Sea, *n*=2 Canary Islands) and did show differentiation using STRUCTURE (Carroll et al. 2016). The initial analysis indicated that the Scottish sample belonged to a different population and when removed, the two Canary Island samples clustered together separate from the Ligurian Sea sample. The samples in this pilot study were sequenced using an Illumina MiSeq, and scaling up to a HiSeq2500, it was expected that 45-50 samples could be multiplexed in a single lane to obtain this level of coverage (Carroll et al. 2016).

In the three data chapters that follow, laboratory methods for ddRAD library preparation followed the method described in Peterson et al. (2012) as modified in Carroll *et al.* (2016, 2021). In order to combine many individuals into a single pool for sequencing, a combinatorial labelling system was used (Peterson *et al.*, 2012). Each sequencing library contained DNA from up to 10 individuals labelled with 10 unique 5-8 bp long barcodes. Each sequencing library was labelled with a unique 6 bp long reverse index. This combinatorial approach using barcodes and indices allowed for multiplexing up to 50 individuals in a single sequencing lane on a HiSeq2500, increasing the cost efficiency significantly.

The specific steps to generate the ddRAD libraries were as follows. Samples selected for sequencing were normalised to 20ng/µl and underwent an overnight digestion at 37°C with two restriction enzymes: MspI and HindIII. After a 20-minute heat kill step at 65°C, adaptors were ligated with one of 10 forward barcodes per sample. This mixture was kept at 22°C for 2 hours followed by 65°C for 20 minutes. With unique barcodes now ligated, all samples within a library were pooled and cleaned using 3 PureLink PCR Micro Kit columns (Invitrogen) per library. Following the final elution step, 30µl of cleaned ligate underwent size selection to a 300-400bp range using a Pippin Prep (Sage Science). The resulting size-selected ligate was divided into 8 wells and library-specific reverse indices were annealed during PCR using a Phusion High-Fidelity PCR kit (ThermoFisher). PCR products were pooled and cleaned using AMPURE-XP (Beckman-Coulter) beads and eluted to a final volume of 10µl in EB buffer (Qiagen).

Two separate rounds of sequencing were done in 2018 and 2020. The aim of the 2018 sequencing round was to develop an understanding of the baseline patterns of genetic diversity and structure of Blainville's and Cuvier's beaked whales across their global ranges, providing context for investigating the impacts of anthropogenic activities on resident populations in future studies. It was decided that five lanes of sequencing on a HiSeq 2500 (Illumina) would be sufficient to balance the coverage of samples across their respective distributions and budgetary constraints. Finished libraries were sent to the National High-throughput DNA Sequencing Centre at the University of Copenhagen where the final library quantities and quality were checked with qPCR and Bioanalyzer (Agilent Genomics), and the libraries were normalised, pooled into sequencing lanes, and sequenced on a HiSeq2500 (Illumina) using paired-end (PE, 150bp) chemistry.

The sequencing round in 2020 focused on individuals that were sampled in the three focal regions in Chapter 5 which are the focus of the ONR funded project: the Bahamas, Canary Islands and Mediterranean Sea. Following ddRAD library prep, final library quantities

were checked using the qPCR NGS Library Quantification Kit (Agilent Genomics) following manufacturer instructions. The final library quality and size was checked using a TapeStation 4200 (Agilent Genomics) and D1000 ScreenTape. Since the University of Copenhagen retired their HiSeq2500, the 2020 sequencing round was upgraded to the NextSeq (Illumina) platform, and ddRAD libraries for 150 individuals (three pools, five libraries per pool, 10 individuals per library) were sent to NovogeneAIT Genomics Singapore Pte. Ltd for PE (150bp) sequencing.

2.5.2 Whole Mitogenome Sequencing

Sequencing regions of mitochondrial DNA (mtDNA) has been a common method of studying genetic variation and divergence between species/populations for several decades (Harrison 1989). MtDNA is small, circular and conserved DNA that is found in the mitochondria, thus only inherited maternally and without any recombination (Harrison 1989). Traditionally, it was only possible to sequence fragments of the mtDNA genome due to cost and time, such as the non-coding control region (CR), cytochrome b (CytB) gene or the cytochrome c oxidase subunit 1 gene (CO1 or Cox1) (Duchêne et al. 2011). Studies have shown that cetaceans have particularly slow molecular clocks and low genetic diversity, as such traditional short fragment mtDNA methods have been unable to resolve some phylogenies on species and/or population levels (Duchêne et al. 2011; Morin et al. 2012).

As the cost of DNA sequencing decreases, the ability to sequence whole mitogenomes of cetaceans has allowed for more accurate phylogenies and estimates of divergence time to be generated (Morin *et al.*, 2008, 2012; Duchêne *et al.*, 2011; Hancock-Hanser *et al.*, 2013; Thompson, *et al.*, 2016). In a comparative analysis, Duchêne *et al.* (2011) compared the resolution of cetacean phylogenies using whole mitogenomes, subsets of informative genes, and single genes. Whole mitogenomes were determined to be the most reliable based on topologies that were highly supported and exhibited clock-like behaviour. This study also found that informative subsets of genes could summarize the phylogenies generated by the whole mitogenome, however they were species-specific and could only be determined by first sequencing entire mitogenomes. In longer time scales (>15 mya), single genes would also be sufficient for phylogenetic analyses.

To explore major phylogeographical patterns and supplement the ddRAD analyses for Cuvier's and Blainville's in Chapter 3, complete mitogenomes were generated by Dr. Morten Tange Olsen and colleagues at the University of Copenhagen from a subset of the ddRAD individuals, including some low-quality museum specimens. "Shotgun" sequencing was used, whereby DNA was fragmented to approximately 350bp, using the M220 Focused-Ultrasonicator™ (Covaris), according to the manufacturer's protocol. After fragmentation, samples were quantified on the Agilent 2200 TapeStation according to the protocol for genomic DNA, to verify fragmentation success. DNA libraries were built, using the blunt-end single-tube protocol described by Carøe et al. (2018) with a few modifications. To each library, 2 µl of 10 µM Illumina® adapters were added to the fragmented DNA, followed by a MiniElute (Qiagen) clean-up step before indexing with P5 and P7 indices. Libraries were sequenced using paired-end (PE) 150 bp chemistry on two lanes of Illumina® HiSeq 4000 at the National High-Throughput Sequencing Centre at University of Copenhagen, Denmark. In addition, 16 libraries characterized by average fragment lengths <300bp were sequenced on a single lane of HiSeq4000 using single-end 80 bp chemistry.

2.6 BIOINFORMATIC PIPELINES

2.6.1 Assessing Sequencing Read Quality

All sequencing reads regardless of method (ddRAD, GBS and mitogenome) underwent quality control using the program FastQC (Andrews et al. 2010). FASTQ formatted files provided by the sequencing facilities were imported directly into FastQC to obtain a detailed report summarising the quality of the reads according to several analysis modules. Quality reports included per sequence quality scores, per sequence GC content, per base N content, sequence length distribution, etc. One of the most useful modules, per base sequence quality, gave an overview of how the quality of each base changed according to its position in the total read. Using these plots, you can decide if and at what length you would like to trim your sequencing read, as quality usually decreases as the read increases. An example plot of one sequencing library (Mde_L3, reverse) is in Figure 2-3. As the sequencing read approaches 90 base pairs, the read quality decreases.

There are some special considerations for interpreting FastQC results of sequences generated using restriction enzymes in the library preparation. Each sequencing read will have a restriction enzyme overhang which is visible in the slightly reduced quality score for the first 5 bases in Figure 2-3. This is more obvious in the "Per base sequence content" of a FastQC report, which will show extremely skewed per base composition for the first five bases which are the restriction enzyme overhangs.

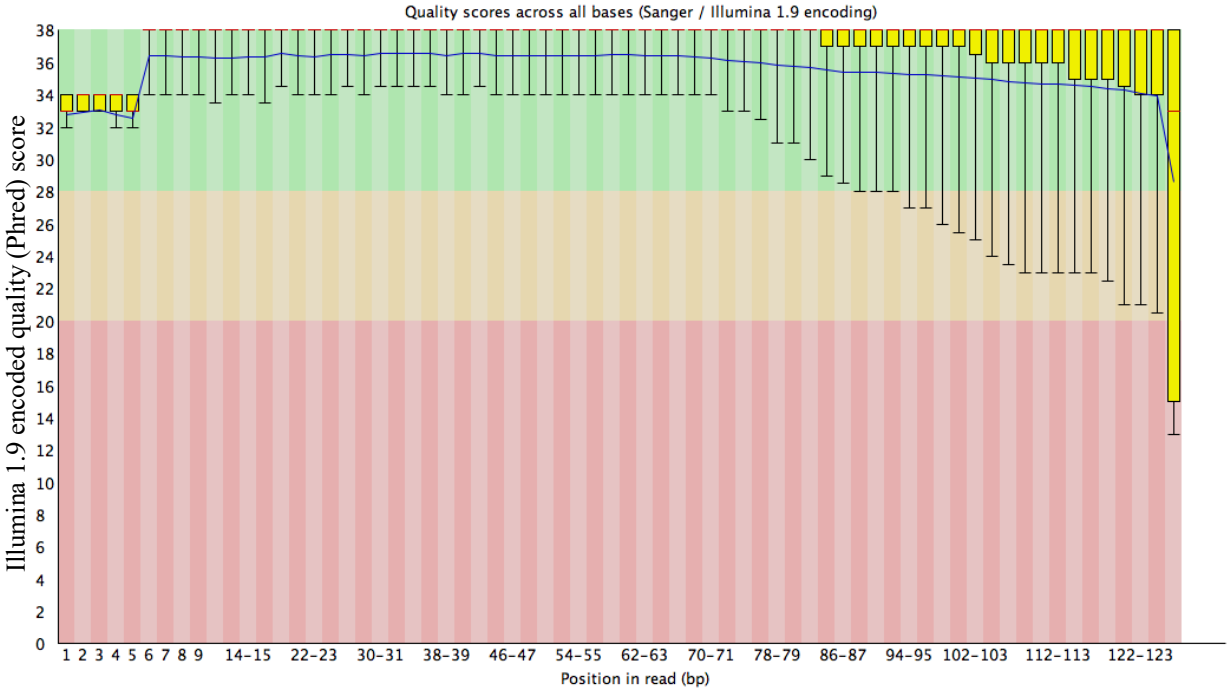


Figure 2-3. Per base sequence quality report for one sequencing library (Mde_L3) generated in FastQC. The quality score is based on the Illumina 1.9 encoded Phred score. In each box and whisker plot the red line indicates the median, the yellow box is the inter-quartile range, the whiskers represent 10% and 90% and the blue line indicates the mean quality. The shaded green region indicates good calls, the orange indicates reasonable quality, and the red area indicates poor quality.

2.6.2 SNP Discovery and Genotyping

2.6.2.1 Bioinformatic pipeline selection

Nuclear SNP genotypes were derived from ddRAD sequencing data using the program Stacks (v2, Catchen *et al.*, 2011, 2013; Rochette, Rivera-Colón and Catchen, 2019). Unlike other genotyping programs, Stacks is particularly useful for non-model organisms as it can discover and genotype SNPs with or without a reference genome, and version 2 is especially well suited to deal with paired-end sequencing data. As it was unknown whether reference genomes would be available in time for the analysis, Stacks was selected for its utility with *de novo* assembly. To further justify the use of Stacks, I compared the number of peer-reviewed papers citing the five most used genotyping programs or pipelines on Web of Science since 2013: dDocent (Puritz et al. 2014), UNEAK (Lu et al. 2013), Pyrad (Eaton 2014), ANGSD (Korneliussen et al. 2014) and Stacks (Catchen et al. 2011; Catchen et al. 2013; Rochette et al. 2019) (Figure 2-4). Within the Stacks pipeline, there are a number of parameters that can be modified which impact the genotyping stringency (Rivera-Colón and Catchen 2022) and rather than apply an arbitrary cut-off across all species, I undertook an optimisation study to determine the best set of parameters for using Stacks with the beaked whale ddRAD data (2.8.1).

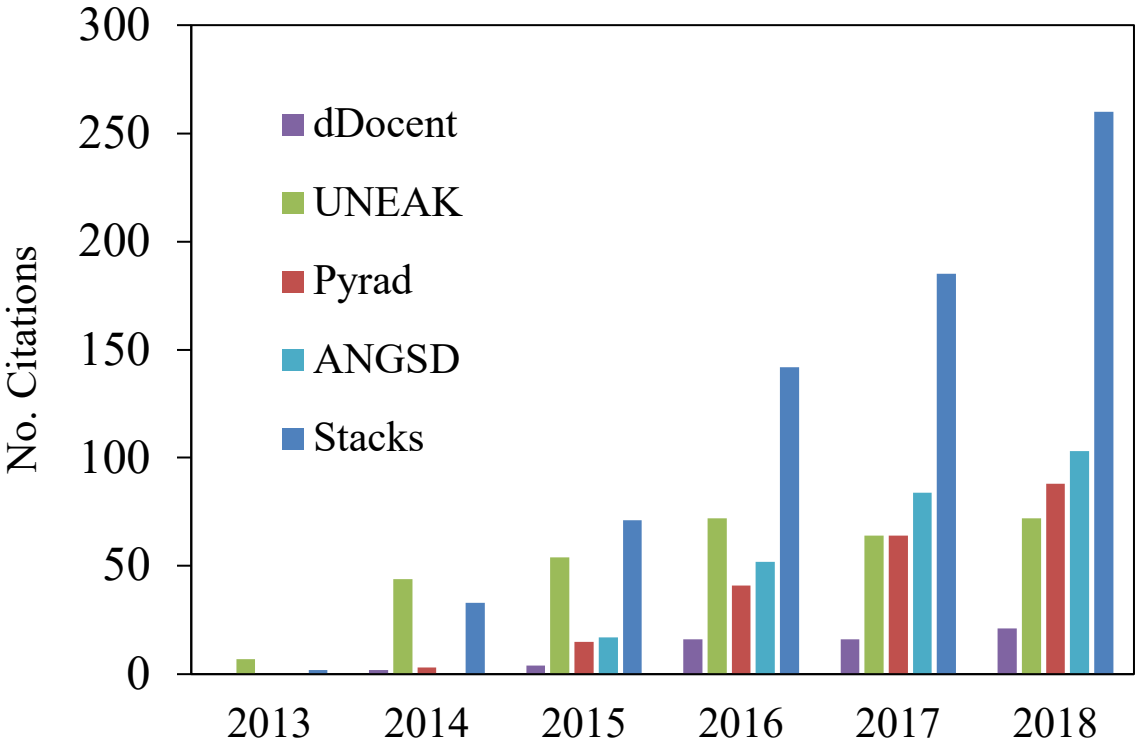


Figure 2-4. The number of journal citations for each of the five most used programs and pipelines (dDocent, UNEAK, PYRAD, ANGSD and Stacks) to discover and genotype SNPs between 2013 and 2019, when the analysis was started.

As described in section 2.2.2, reference genomes became available in April 2018 from Jeremy Johnson at the Broad Institute, and the reference map pipeline was followed in Stacks. The multi-step process of going from raw sequencing data to genotypes usable for various analyses in this pipeline involves five main steps: 1. cleaning the data, 2.

aligning the data, 3. building the loci catalogue, 4. applying a population-level framework and 5. filtering the dataset. Steps 1, 3, and 4 were conducted in Stacks while steps 2 and 5 were done externally. Each step is described in detail below.

2.6.2.2 Cleaning the data

During library preparation, samples were pooled together based on the combinatorial labelling system of forward barcodes and reverse indices. The sequences were demultiplexed based on the library-specific reverse index by the sequencing centres prior to returning the data. The program 'process_radtags' in Stacks was then used to demultiplex individuals within libraries using their individual-specific forward barcodes. For each individual, this program allows you to check and correct the enzyme cut sites, check the quality score of a read in a sliding window and discard those below a determined threshold, and trim the ends off reads if the base quality deteriorates as the sequence length increases (see section 2.6.1).

2.6.2.3 Aligning and sorting the data

Once the sequencing reads were cleaned, trimmed, and demultiplexed, they were aligned to the reference genome using the BWA-MEM algorithm in the Burrows-Wheeler Alignment Tool implemented in the software BWA (Li and Durbin 2009; Li 2013). This algorithm is particularly well suited for Illumina reads between 70-100bp long and is faster and more accurate than the other alignment options (Li and Durbin 2009; Li 2013). Prior to alignment, the reference genome must be indexed, condensing the file, and therefore decreasing computation time. The alignments are output as human-readable SAM (Sequence Alignment/Map) files that are converted using SAMtools (Li et al. 2009) to a compressed, binary version called a BAM (Binary Alignment Map) file for more efficient downstream applications. Once the BAM files are sorted by their genomic coordinates in SAMtools, they were used to discover and genotype SNPs in Stacks.

2.6.2.4 Building the loci

SNP loci were discovered and genotyped from the sorted BAM files in the 'gstacks' program within Stacks. Stacks has incorporated the Bayesian genotype caller (BGC) algorithm of Maruki and Lynch (2015, 2017; Rochette, Rivera-Colón and Catchen, 2019) which uses a Bayesian genotype-frequency prior that takes into account population-level allele frequencies, does not assume Hardy-Weinberg equilibrium and estimates error rates directly from the sequence data (not from read quality scores) when calling genotypes (Maruki and Lynch 2015; Maruki and Lynch 2017). 'gstacks' first estimated significant polymorphic loci from the BAM alignments with confidence set by the parameter "var-alpha" in 'gstacks' (Maruki and Lynch 2015; Rochette et al. 2019). Genotypes for each individual were called at these loci, using a method that takes into account allele balance and read depth (Maruki and Lynch 2017). Confidence in the genotype calling was done using a likelihood ratio test which compared the likelihood of the two most likely genotypes (Maruki and Lynch 2015).

Four parameters must be defined within the 'gstacks' command and can be optimised in a preliminary analysis with a subset of the data: "min-mapq" (minimum mapping quality score to consider a read; 10,20), "max-clipped" (maximum soft-clipping level as a fraction of the read length; 0.1, 0.2), "var-alpha" (SNP discovery threshold; 0.05, 0.01) and "gt-alpha" (genotype calling threshold; 0.05, 0.01). The best combination of parameters was selected based on resulting datasets with the highest number of SNP loci and the lowest amount of missing data (see 2.8.1).

2.6.2.5 Applying a population framework

The resulting loci from 'gstacks' were passed through the 'populations' program in Stacks. In 'populations', individuals and their genotyped loci can be analysed in a framework incorporating some sort of group assignment (such as geographical origin or sex) and then filtered according to minor allele frequency (MAF) or locus frequency within the entire population. To reduce bias from potentially arbitrary population designations, no such population framework was provided for Chapters 3 and 4, however *a priori* populations were assigned in Chapter 5 to maximise the number of loci retained for each population. MAF filters were applied only to ensure that there were no monomorphic loci.

2.6.2.6 Filtering and Quality Control

Massively parallel sequencing can lead to high error rates and genotypic uncertainties that can be introduced at any step throughout the analysis (O'Leary et al. 2018). Fortunately, many of these can be overcome by employing rigorous filtering to identify and reduce errors before analysing the final dataset (O'Leary et al. 2018). In this study, I implemented a tiered approach to filtering, starting with low cut-off values for missing data (applied separately per locus and individual) and finalizing the dataset with higher thresholds. This alternative and iterative filtering method whereby you increase the cut-off threshold has been shown to retain more loci and individuals as poor-quality individuals can deflate genotype call rates in otherwise acceptable loci while poor-quality loci can increase the amount of missing data in otherwise acceptable individuals (O'Leary et al. 2018). Therefore, a rigorous quality control (QC) process followed Stacks 'populations' (Nielsen et al. 2011; O'Leary et al. 2018), using R v.3.6.0 (R Core Team 2019) and VCFtools v.0.1.12a (Danecek et al. 2011) to filter individuals and loci based on the amount of missing data, read depth, and quality score, following the recommendations of O'Leary et al. (2018; see Table 2-3).

Table 2-3. List of filtering commands and steps used in the program VCFtools to filter loci and individuals based on locus depth, genotype quality, minor allele frequency (MAF) and missingness.

Threshold Level	VCFtools Command	Description
Low	--minDP 5 –minGQ20	Recode genotypes with quality <20 and depth <5 to zero
	--maf 0.001	Remove the sites made monomorphic by previous step.
	--max-missing 0.5	Remove sites with >50% missing data
	--missing-indv	Calculate missingness per individual, write a list of individuals with >50% missing data
	--remove	Remove individuals on list with >50% missing data
High	--site-depth	Calculate site depth, write a list of loci with mean site depth >3x the overall mean
	--exclude-positions	Remove sites with site depth >3x the overall mean
	--max-missing 0.75	Remove sites with more than 75% missing data
	--missing-indv	Calculate missingness per individual, write a list of individuals with >25% missing data
	--remove	Remove individuals on list with >25% missing data

The list of remaining loci ("whitelist") was passed back into 'populations', retaining loci that appeared in a predetermined proportion of individuals and had predetermined minor allele frequencies (depending on the analysis). The resulting data were exported as a VCF file using Stacks with a single random SNP written per locus, and any individuals with >25% missing data were removed using VCFtools.

2.6.2.7 Assessing final datasets

The resulting bi-allelic genotypes were plotted using the command "glPlot" in the R package, *adegenet* v. 2.1.1 (Jombart 2008), to visually inspect the frequency of the minor allele for each locus. Such plots allow you to see where individuals or loci may have excessive amounts of missing data and in this study, individuals to remove were identified using "glPlot" due to sequencing errors from library-specific artefacts (Jombart 2008; Jombart and Collins 2017). To assess the quality of the final dataset for each species, the total number of SNP loci per species, per individual and SNP depth per individual were derived using VCFtools. The total proportion of missing data per species was calculated using *adegenet*. Ascertainment bias was not explicitly accounted for, as GBS-type approaches are not

2.6.3 Mitogenome Assembly

The following work involving mitogenomes was done by the collaborators at the University of Copenhagen. In summary, de-multiplexing of mitogenome sequencies and removal of indices with mismatches and initial QC were performed with bcl2fastq Conversion Software (Illumina®). All PE sequences were paired, merged and trimmed with the "BBDuk" plugin in Geneious v.9.1.5 (Kearse et al. 2012), and subsequently mapped to the reference mitogenomes for Cuvier's (NCBI GenBank: LN997430) and Blainville's beaked whales (NCBI GenBank: NC_021974.2), using Geneious mapper. Coverage was assessed for the newly generated mitogenomes using Geneious.

2.7 ANALYSIS OF NUCLEAR SNP DATA

2.7.1 Investigating and Measuring Population Structure

The R package *tess3r* (Caye et al. 2018) was selected to carry out investigations into population structure for the ddRAD data for Chapters 3 and 4. This package incorporates genotypic and geographical information (latitude and longitude coordinates for each sample) in a spatially explicit, least-squares optimization approach to estimate ancestry. Results are presented in bar plots, where each individual is represented by a bar and the ancestry coefficients reflect the probability of population membership and estimates of admixture. Prior to running the program, a predefined range of ancestral population clusters (K) is assigned, and the appropriate value of K is selected based on the cross-entropy scores and parsimony of resulting bar plots. Cross-entropy scores are plotted by the package against the range of K values to select the optimal value. In these plots, smaller cross-entropy scores indicate better fit, and often the best estimate of K is found when the curve reaches a plateau or starts to increase. In this study, there was rarely a clear minimum or plateau, and in these cases the optimal value of K was that which led to the most parsimonious assignment of individuals (least amount of admixture) to populations in the bar plots. This version of *tess3r* doesn't assume an underlying biological model but does assume that individuals sampled close together will share more ancestry than those sampled from further away.

Population structure was also visualised using Discriminant Analysis of Principle Components (DAPC) in the R package, *adegenet* (Jombart 2008; Jombart et al. 2010). DAPC summarises the amount of genetic differentiation between groups (either determined *a priori* or using K-means clustering) while minimising the amount of variation within groups. The DAPC method comprises two steps. In the first, data are transformed using principal component analysis (PCA) based on a genotype matrix. In the second step, the uncorrelated variables generated by PCA in the first step are input in the Discriminant Analysis (DA) (Jombart et al. 2010; Miller et al. 2020). To optimise the number of principle components (PCs) to retain from the first step in the DAPC analysis, *adegenet* offers a cross-validation tool with "xvalDapc" (Jombart et al. 2010; Jombart and Collins 2015). This command subsets the data to use as a training set, runs the analysis over a pre-determined number of repeats (n=30), and determines the best number of PCs to retain based on whichever yields the highest predictive success of the training data with the lowest root mean squared error (RMSE). The resulting DAPC scatter plots visualise structure across two dimensions. In the following data chapters, DAPC optimisation was done using genetic populations identified *a priori* in *tess3r.*

The amount of genetic differentiation between the defined clusters was measured by calculating pairwise F_{ST} values (Weir and Cockerham 1984) and their 95% confidence intervals (*dartr* v1.3.3, based on 100 bootstraps; Gruber, Unmack, Berry, & Georges, 2018). Statistical significance of the fixation indices was determined by permutation test using the *strataG* R package or *dartr* (Archer et al. 2017; Gruber et al. 2018). Calculations of F_{ST} give a measurement of similarity of allele frequencies between populations, with high F_{ST} values indicative of greater genetic differentiation (Waples et al. 2018). F_{ST} is widely used to measure population divergence since its value and significance are easily calculated using different types of data and it is straightforward to interpretate (Waples et al. 2018).

2.8 OPTIMISATION AND SENSITIVITY STUDIES

Summarised here are the four analyses I undertook to ensure sound methodology: optimise the genotype calling algorithm in Stacks for the beaked whale dataset (2.8.1), compare the relative success of different cetacean genomes to retain reads and enable loci to be called across study species (2.8.2), compare the utility of two different sequencing platforms and estimate the reproducibility of ddRAD sequencing (2.8.3), and conduct a sensitivity analysis to determine the number of SNPs required to produce accurate estimates of genetic structure and diversity (2.8.4).

2.8.1 Optimising 'gstacks' Parameters

The Stacks SNP discovery pipeline (Rochette et al. 2019) implements the Bayesian genotype caller (BGC) algorithm of Maruki & Lynch (2015, 2017) in the command 'gstacks' (2.6.2.4). Four parameters are defined within the 'gstacks' command and were optimised in a preliminary analysis with a subset of the data for each species analysed. The parameters are: "min-mapq" (minimum mapping quality score to consider a read), "max-clipped" (maximum soft-clipping level as a fraction of the read length), "var-alpha" (SNP discovery threshold) and "gt-alpha" (genotype calling threshold).

To optimise the best 'gstacks' parameters for the Cuvier's and Blainville's analysis in Chapter 3, a subset of demultiplexed and QC reads from the 2018 sequencing run were selected. Cuvier's (n=40) and Blainville's *(*n=56) individuals were chosen as they were

considered high quality (>98% retained reads and >1,000,000 retained reads) and covered the widest geographical ranges. The full bioinformatic pipeline described above was run for each combination of parameters, and the best parameters were those that maximised the number of SNP loci and minimised the amount of missing genotypes(Appendix B).

In summary, the following parameters were modified to optimise the 'gstacks' command for each dataset using the values in brackets: "min-mapq" (10,20), "max-clipped" (0.1, 0.2), "var-alpha" (0.05, 0.01) and "gt-alpha" (0.05, 0.01). The same parameters were used for Cuvier's and Blainville's in Chapters 3 and 5 but were re-calculated using just the samples used in Chapter 4. The final optimised 'gstacks' parameters used for each chapter are provided below in Table 2-4.

Table 2-4. The final optimised parameters for 'gstacks' used to generate the SNP datasets in each chapter for the four study species: "min-mapq" (minimum mapping quality score to consider a read; 10,20), "max-clipped" (maximum soft-clipping level as a fraction of the read length; 0.1, 0.2), "var-alpha" (SNP discovery threshold; 0.05, 0.01) and "gt-alpha" (genotype calling threshold; 0.05, 0.01).

Species	Chapter	min-mapq	max-clipped	var_apha	gt_alpha
Cuvier's beaked whale	3, 4, 5	10	0.2	0.05	0.05
Blainville's beaked whale	3, 5	10	0.2	0.05	0.05
	4	20	0.2	0.05	0.05
Sowerby's beaked whale	4	20	0.2	0.01	0.05
Northern bottlenose whale	4	10	0.2	0.05	0.05

2.8.2 Comparing Suitability of Different Reference Genomes

In Chapter 4, I investigated the population structure and demographic history of four beaked whales found in the North Atlantic: Northern bottlenose whale ("N. bottlenose"), Sowerby's beaked whale ("Sowerby's"), Cuvier's beaked whale ("Cuvier's") and Blainville's beaked whale ("Blainville's"). Sequence data were generated for each species using ddRAD (Cuvier's, Blainville's and Sowerby's, 2.5.1) or GBS (N. bottlenose by the University of Copenhagen, 2.2.4). In this chapter, I originally planned to use a reference genome basal to all four species to align the sequences to. I chose the sperm whale genome, as it was of very high quality and fully annotated (*Physeter macrocephalus*; Genbank Accession: GCA_002837175.2). A bioinformatic analysis was done to determine the differences in datasets when aligning sequences to the Cuvier's genome or the sperm whale genome for SNP discovery and future analyses. In doing this, I was also able to assess the suitability of the Stacks and VCFtools pipeline for the N. bottlenose GBS data.

DNA sequences from individuals analysed in Chapter 4 (n=58 N. bottlenose, n=89 Cuvier's, n=40 Sowerby's and n=35 Blainville's from the North Atlantic only) were demultiplexed using 'process_radtags' in Stacks. BWA was used to align the cleaned reads to either the Cuvier's or sperm whale genomes, and Table 2-5 shows the cleaning and alignment summaries for each species/alignment. The percentage of retained reads was high for each species and the percentage of kept alignments varied widely. The percentage of kept alignments was consistently higher when the Cuvier's reference genome was used, however.

The 'gstacks' component of Stacks identified and genotyped a catalogue of SNPs in each set of alignments using the same parameter values in Chapter 3 (--max-clipped 0.2 --min-

mapq 10 --var-alpha 0.05 --gt-alpha 0.05). The 'gstacks' calls were run through the 'populations' program in Stacks without any *a priori* population designation in this step. The resulting VCF files underwent the stringent filtering steps using VCFtools and R based on missingness of loci and individuals as well as locus depth (2.6.2.6). Finally, the remaining individuals and SNPs were passed back through 'populations' for a second time, resulting in a final VCF file. Dataset summaries for each species/alignment combination are in Table 2-6. For the datasets based on ddRAD data, the final number of SNPs was high (>25k), and the N. bottlenose data had many fewer SNPs (>1400). Across all species, more SNPs were retained when the Cuvier's dataset was used.

Table 2-5. Summaries from demultiplexing and aligning DNA sequences of the four species analysed in Chapter 4 using Cuvier's and sperm whale reference genomes.

Species	***n***	**Mean Total Reads**	**Mean Retained Reads**	**Mean % Retained**	**Reference Genome**	**Mean Total Alignments**	**Mean Alignments Kept**	**Mean % Kept**
Cuvier's	89	6174056	6108788	94%	sperm whale	5601180	3154768	54%
					Cuvier's	5488349	3698910	65%
Blainville's	35	5059604	5010074	96%	sperm whale	5241059	3256010	61%
					Cuvier's	5197128	3649521	69%
Sowerby's	40	7553143	7484999	89%	sperm whale	7810003	4519592	51%
					Cuvier's	7712545	5366259	62%
N. bottlenose	58	1401190	1384101	98%	sperm whale	1387796	360727	25%
					Cuvier's	1388155	385256	27%

Table 2-6. Summary of SNP loci and individuals retained across each of the Stacks 'populations' runs and the final filtering steps in VCFtools and R.

Species	**Reference Genome**	***n***	**Pop1 Polymorphic Sites**	**Pop2 Polymorphic Sites**	**Final *n***	**Final No SNPs**	**Mean Coverage**	**Mean % Missing**
Cuvier's	Sperm whale	89	1473308	29533	55	22987	53.5	1.99
	Cuvier's	89	1752665	32899	55	32899	54.4	2.45
Blainville's	Sperm whale	35	588541	26698	29	26698	46.0	3.07
	Cuvier's	35	689592	29612	29	29612	46.8	3.22
Sowerby's	Sperm whale	40	293965	36847	29	33525	65.8	2.02
	Cuvier's	40	370378	36847	29	36847	68.1	2.04
N. bottlenose	Sperm whale	58	27007	1145	28	1145	10.8	1.69
	Cuvier's	58	30348	1308	28	1308	10.2	1.56

The final VCF files for each combination of species and reference genome were loaded into R to inspect visually. The function "glPlot" in the R package *adegenet* was used to display the VCF files as a heatmap, with each row corresponding to an individual, each column to a SNP locus, and each colour representing a different genotype. These plots were used to highlight individuals that had seemingly excessive rare alleles, which were inferred to be suffering from some batch effect such as contamination. These samples were removed, and the pipeline rerun to call loci.

In conclusion, ddRAD and GBS sequences aligned to the Cuvier's reference genome consistently yielded more SNPs across all four species (Table 2-7) and so these data were selected for the analyses in Chapter 4. These data were also deemed most appropriate for the demographic reconstructions, because sequences aligned to the sperm whale genome were likely biased towards highly conserved regions, which is not desirable for demographic reconstructions. Regardless of the reference genome used, the N. bottlenose GBS samples yielded far fewer SNPs compared to the ddRAD datasets, however it was determined that this pipeline was adequate to include these samples in the chapter.

Table 2-7. Final SNP datasets for each species and reference genome combination.

Species	Reference Genome	*n*	No SNPs
Cuvier's	Sperm whale	55	22987
	Cuvier's	55	32899
Blainville's	Sperm whale	25	20466
	Cuvier's	25	24120
Sowerby's	Sperm whale	27	25074
	Cuvier's	27	29660
N. bottlenose	Sperm whale	28	1145
	Cuvier's	28	1308

2.8.3 Concordance of SNP loci between different sequencing runs and platforms

Little has been published about the reproducibility between sequencing platforms when the same library preparation method and samples have been used. To address this, a subset of Blainville's (n=21) and Cuvier's (n=25) that were sequenced in both the 2018 and 2020 sequencing runs (HiSeq2500 and NextSeq platforms, respectively) underwent the same bioinformatic pipeline to determine how consistently SNP loci were discovered and genotyped.

Individuals were selected if they had more than 1,000,000 reads after demultiplexing and quality control in the Stacks 'process_radtags' program. 'Process_radtags' was used to remove barcodes, remove bad quality reads (phred score =10) and trim sequences to 90bp (the point at which quality readings began to decline according to FastQC). The trimmed and demultiplexed sequences were aligned to reference genomes (Cuvier's and Blainville's) using the BWA-MEM algorithm implemented in BWA (Li and Durbin 2009; Li 2013). The bioinformatic pipeline outlined above was run twice. First, the duplicate samples were analysed separately by year. Then, the 2018 and 2020 alignment files were merged for each individual. Alignments were run through the Stacks program 'gstacks' using the optimised parameters found for both Cuvier's and Blainville's in Chapter 3. The 'gstacks' catalogue was passed through the 'populations' module with no *a priori* population designation and including a minor allele frequency filter (MAF) to ensure no monomorphic loci were present. The resulting VCF files were filtered using VCFtools and R as described above (2.6.2.6) with the final filtered VCF file used as input for a second run in Stacks 'populations'.

VCFtools provides several options for comparing two VCF files. First, I compared the final 2018 and 2020 VCF files (--diff-indv) to see if the same individuals passed the QC and filtering steps. I calculated the shared loci between the two files (--diff-site) and estimated the number of discordant genotypes between the 2018 and 2020 VCF files for each individual and locus (--diff-indv-discordance and --diff-site-discordance,

respectively). To use the comparison functions in VCFtools, the two different VCF files must share all the same chromosomes and since the reference genomes used were scaffolds and not chromosome level, many (>1000) weren't shared between the files and needed to be excluded from the discordance calculations.

The resulting datasets from the optimisation subset are displayed in Table 8, including the 2018, 2020 and the 2018/2020 merged dataset. Many SNPs were discovered and genotyped per individual, with little missing data, regardless of the sequencing platform used. Most individuals were retained through the bioinformatic pipeline in both species, regardless of the sequencing platform used (Table 2-8).

Table 2-8. Sample sizes and final number of loci in the optimisation datasets following the full bioinformatic pipeline. "Merged" indicates that the FASTA files generated in 2018 and 2020 were merged for each duplicated individual prior to the bioinformatic pipeline.

	Blainville's			**Cuvier's**		
	2018	**2020**	**Merged**	**2018**	**2020**	**Merged**
Starting_N	21	21	21	25	25	25
Final_N	17	17	19	24	22	24
Final_SNP_Loci	34,443	32,707	36,441	81,211	80,664	75,375
Final_Missing_%	3.03	5.2	4	4.52	5.64	5.33

Comparison of the 2018 and 2020 datasets showed a high percentage of SNP loci were common to both datasets, compared to the total number of SNP loci found across both datasets (Blainville's: 76%, Cuvier's: 81%, Table 2-9). Of the loci genotyped in common between 2018 and 2020, individual site discordance ranged from 0.12-0.90% (0.33% mean discordance) in Blainville's and 0.13-37.04% (3.47% mean discordance) in the Cuvier's (Table 2-9). Overall, site discordance was low per site in both datasets (Blainville's: 0.35%, Cuvier's: 3.60%, Table 2-9). The findings are further explored in the discordance matrices presented in Table 2-10 and Table 2-11. Each matrix gives the counts of loci that shared the same genotype across both files, shared 1 allele across both files, shared no alleles across both files or were missing in either file. There are many more discordant loci in the Cuvier's dataset (also seen in the maximum individual discordance in Table 2-9), which is due to two likely contaminated individuals with >30% discordance across all sites.

Table 2-9. Sample sizes, final number of loci and discordance between the datasets generated in 2018 and 2020.

	Blainville's	**Cuvier's**
Starting *n*	21	25
Final *n*	17	22
Total loci in both files	34443	81211
Common loci between files	26182	65874
2018 only loci	6765	7098
2020 only loci	5249	7133
Minimum individual discordance	0.12%	0.13%
Maximum individual discordance	0.90%	37.04%
Mean individual discordance	0.33%	3.47%
Discordant alleles	1219	48081
Discordant loci	856	36277
Mean locus discordance	0.35%	3.60%

Table 2-10. Discordance matrix for the 2018 and 2020 Blainville's SNP datasets. The row and column heading indicate the genotype and the dot indicates missing data.

Blainville's Discordance Matrix	2018: 0/0	2018: 0/1	2018: 1/1	2018: ./.
2020: 0/0	243723	466	1	2958
2020: 0/1	321	105806	142	1967
2020: 1/1	0	217	23532	578
2020: ./.	5634	3943	867	2377

Table 2-11. Discordance matrix for the 2018 and 2020 Cuvier's SNP datasets. The row and column heading indicate the genotype and the dot indicates missing data.

Cuvier's Discordance Matrix	2018: 0/0	2018: 0/1	2018: 1/1	2018: ./.
2020: 0/0	985370	14513	5399	22004
2020: 0/1	17699	231966	2896	6690
2020: 1/1	4642	2598	82927	3070
2020: ./.	32523	11599	4369	19445

Overall, the two sequencing platforms yielded consistent loci and genotypes between years. This is reassuring given that reproducibility is one of the main advantages advertised by the ddRAD methodology, however, there has been little published evidence to show this.

2.8.4 SNP Dataset Sensitivity Study

In Chapter 4, GBS sequence data for N. bottlenose was provided by project collaborators at the University of Copenhagen (2.2.4). The bioinformatic pipeline for ddRAD sequences was able to identify and genotype SNP loci from this data, however the number of loci generated from the GBS data (~1000 SNPs) was far lower than the ddRAD data (~30 k-100 k SNPs). Using the Cuvier's dataset (which had the most SNPs in the final dataset, ~130 k), I conducted a sensitivity analysis to determine whether the number of SNPs in the N. bottlenose dataset was sufficient to produce accurate estimates of genetic structure and diversity.

The final Cuvier's SNP dataset was randomly subset using the package *dartr* (Gruber et al. 2018) down to 100 k, 50 k, 10 k, 1 k and 200 SNPs. A DAPC analysis was done for each SNP subset to see if the structure that had been characterised using ~130 k SNPs would be as clear using fewer loci. The four genetic clusters identified in the full analysis (Canary Islands, Northeast, Spain, West) were included in the DAPC as *a priori* populations. The results of these DAPCs are presented in Figure 2-5. As in Chapter 3, the population of 2 individuals from Spain were highly divergent, and to visualise the clusters more easily, the 2nd and 3rd discriminant functions are presented. Clear differentiation between the clusters is present down to 1k SNPs, with some overlap appearing in the 200 SNP plot Figure 2-5.

Each of the SNP datasets were run through *tess3r* to see if there were any major differences between values of K depending on the number of SNP loci used. Figure 2-6 shows the cross-entropy plots for each of the five datasets and the resulting co-ancestry plots are presented in Figure 2-7. As with the DAPC, the same patterns of population structure are retained until 200 SNPs, when increased admixture at *K*=4 starts to muddle the structure.

Estimates of genetic differentiation were calculated using *dartr* as described above in section 2.7.1 between each genetic cluster using each of the five datasets, and the point estimates of F_{ST}, the 95% confidence intervals and p-values are presented in Table 2-12. F_{ST} estimates are relatively stable, though with widening confidence intervals down to 1k SNPs. Using only 200 SNPs, two pairs are no longer significantly differentiated (Table 2-12). The Weir and Cockerham inbreeding coefficient, F_{IS} was calculated for each population using the five datasets (Table 2-13). Like the F_{ST} calculations, confidence intervals widen as the number of SNPs is reduced, and the estimates based on 200 SNPs are significantly different from the other datasets.

Using DAPC, *tess3r* and calculations of F_{ST} and F_{IS}, I conclude that using this subset of 1000 random SNPs (the approximate number of SNPs generated from the N. bottlenose GBS data), nearly identical results are found compared to those calculated using 130k SNPs in Chapter 4.

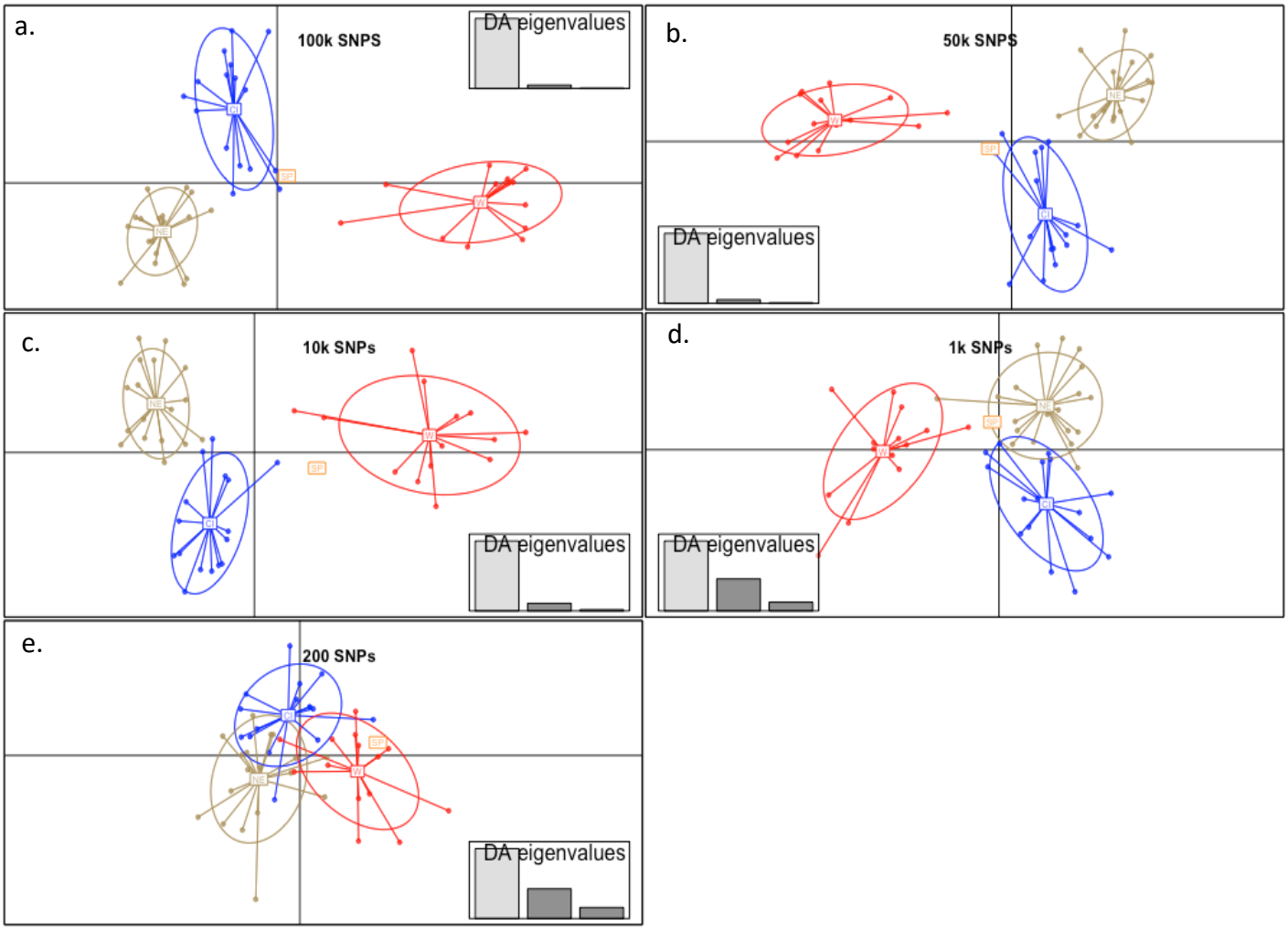


Figure 2-5. DAPC scatter plots for each of the Cuvier's SNP datasets (a. 100k SNPs, b. 50k SNPs, c. 10k SNPs, d. 1000 SNPs, e. 200 SNPs) with the original genetic clusters as priors. The second discriminant function is shown along the x-axis and the third is along the y-axis.

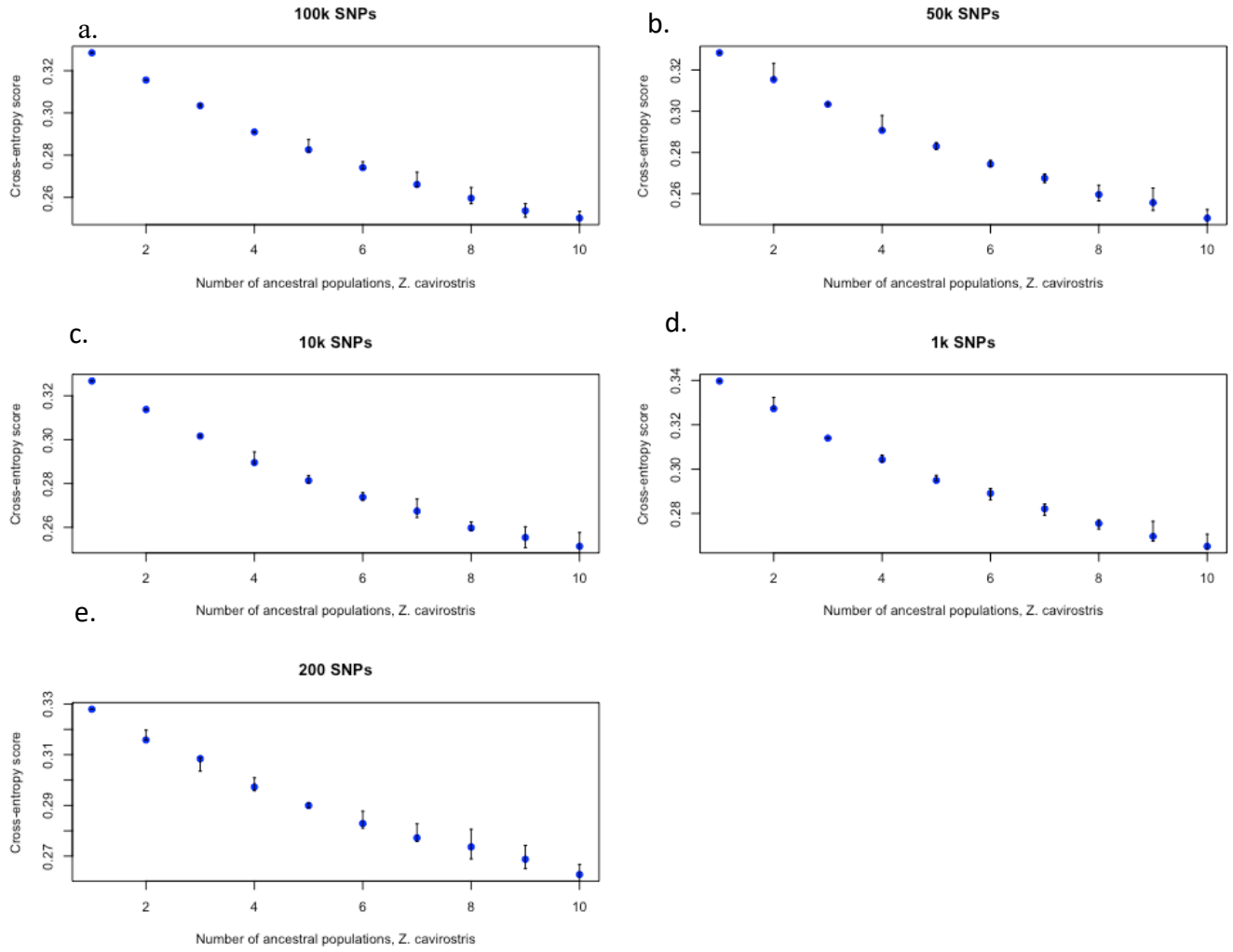


Figure 2-6. Cross entropy plots from the *tess3r* analysis of each of the 5 SNP datasets using a. 100 k SNPs, b. 50 k SNPs, c. 10 k SNPs, d. 1 k SNPs, and e. 200 SNPs.

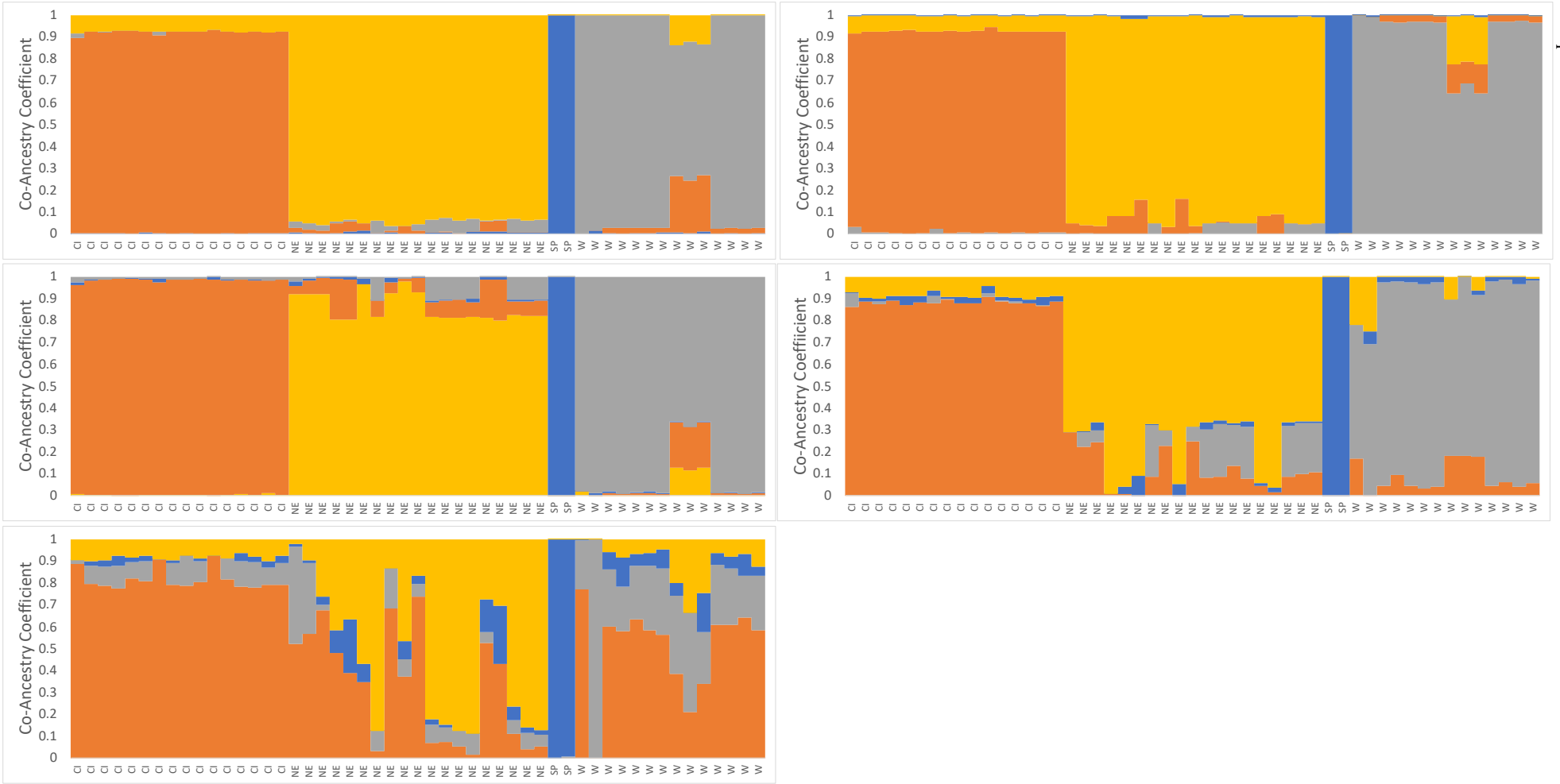


Figure 2-7. Bar plots of coancestry coefficients using *K*=4 clusters for each SNP dataset generated using a. 100k SNPs, b. 50k SNPs, c. 10k SNPs, d. 1k SNPs, and e. 200 SNPs. Each bar in the plots represent one individual whose North Atlantic population is listed along the x-axis (CI: Canary Islands, NE: Northeast, SP: Spain, W: West). The colours correspond to the proportion of ancestry shared amongst the four genetic clusters.

Table 2-12. Estimates of genetic differentiation (F_{ST}), 95% confidence intervals and associated p-values are presented for each population pair, calculated using each SNP dataset. The values in bold are significant at $p<0.05$.

Pop 1	Pop 2	100k SNPs	50k SNPs	10k SNPs	1k SNPs	200 SNPs
CI	**NE**	**0.007 (0.006, 0.007)**	**0.007 (0.006, 0.008)**	**0.006 (0.005, 0.008)**	**0.007 (0.003, 0.011)**	-0.003 (-0.013, 0.005)
CI	**W**	**0.012 (0.011, 0.013)**	**0.012 (0.011, 0.013)**	**0.012 (0.01, 0.014)**	**0.018 (0.013, 0.024)**	0.002 (-0.009, 0.017)
CI	**SP**	**0.153 (0.15, 0.156)**	**0.156 (0.15, 0.161)**	**0.156 (0.145, 0.164)**	**0.127 (0.101, 0.154)**	**0.165 (0.086, 0.22)**
NE	**W**	**0.011 (0.01, 0.012)**	**0.011 (0.01, 0.012)**	**0.011 (0.009, 0.014)**	**0.015 (0.009, 0.02)**	**0.01 (0, 0.024)**
NE	**SP**	**0.141 (0.138, 0.144)**	**0.143 (0.139, 0.147)**	**0.145 (0.134, 0.153)**	**0.123 (0.094, 0.148)**	**0.153 (0.078, 0.2)**
W	**SP**	**0.153 (0.15, 0.156)**	**0.156 (0.151, 0.16)**	**0.155 (0.143, 0.162)**	**0.123 (0.099, 0.149)**	**0.169 (0.079, 0.221)**

Table 2-13. Estimates of Weir and Cockerham's F_{IS} 95% confidence intervals are presented for each population pair, calculated using each SNP dataset.

Pop	100k SNPs	50k SNPs	10k SNPs	1k SNPs	200 SNPs
CI	(0.142, 0.146)	(0.141, 0.147)	(0.142, 0.158)	(0.118, 0.168)	(0.068, 0.208)
NE	(0.145, 0.148)	(0.142, 0.149)	(0.133, 0.147)	(0.115, 0.156)	(0.089, 0.193)
SP	(-1.043, -1.032)	(-1.036, -1.023)	(-1.053, -1.019)	(-1.039, -0.957)	(-1.291, -1.000)
W	(0.152, 0.157)	(0.146, 0.153)	(0.149, 0.168)	(0.142, 0.193)	(0.077, 0.194)

3 Biogeography in the deep: Hierarchical population structure of two beaked whale species

This chapter was published in *Global Ecology and Conservation* as "Biogeography in the deep: Hierarchical population structure of two beaked whale species" (https://doi.org/10.1016/j.gecco.2022.e02308). Here I present this publication in the pre-proof version that was submitted to the journal following peer-review and acceptance for publication. All electronic supplemental materials and tables that are referred to throughout the text are in Appendix C. I was the lead author on this work, and the full list of authors, affiliations and contributions are as follows:

Aubrie B. Onoufriou[1,2], Oscar E. Gaggiotti[3], Natacha Aguilar de Soto[2], Morgan L. McCarthy[4], Phillip A. Morin[5], Massimiliano Rosso[6], Merel Dalebout[7], Nicholas Davison[8], Robin W. Baird[9], C. Scott Baker[10], Simon Berrow[11,12], Andrew Brownlow[8], Daniel Burns[13,14], Florence Caurant[15], Diane Claridge[16], Rochelle Constantine[17], Fabien Demaret[15], Sascha Dreyer[4], Martina Đuras[18], John Durban[16,19], Alexandros Frantzis[20], Luis Freitas[21], Gabrielle Genty[22], Ana Galov[23], Sabine S. Hansen[4], Andrew C. Kitchener[24,25], Vidal Martin[26], Antonio A. Mignucci-Giannoni[27,28], Valeria Montano[1], Aurelie Moulins[6], Carlos Olavarría[29], M. Michael Poole[30], Cristel Reyes Suárez[2], Emer Rogan[31], Conor Ryan[11,32], Agustina Schiavi[2], Paola Tepsich[6], Jorge Urban[33], Kristi West[34], Morten Tange Olsen[4] & Emma L. Carroll[1,17,22]

[1]School of Biology, University of St Andrews, St Andrews, UK
[2]BIOECOMAC Department of Animal Biology, Edaphology and Geology, University of La Laguna, San Cristóbal de La Laguna, Tenerife, Canary Islands, Spain
[3]Centre for Biological Diversity, School of Biology, University of St Andrews, St Andrews, UK
[4]Section for Molecular Ecology and Evolution, Globe Institute, University of Copenhagen, Copenhagen, Denmark
[5]Southwest Fisheries Science Centre, National Marine Fisheries Service, National Oceanic and Atmospheric Administration, La Jolla, California, USA
[6]CIMA Research Foundation, Savona, Italy
[7]School of Biological, Earth, and Environmental Sciences, University of New South Wales, Sydney, New South Wales, Australia
[8]Scottish Marine Animal Stranding Scheme, School of Biodiversity, One Health and Veterinary Medicine, College of Medical, Veterinary and Life Sciences, University of Glasgow, Glasgow, UK
[9]Cascadia Research Collective, Olympia, Washington, USA
[10]Marine Mammal Institute, Hatfield Marine Science Centre, Oregon State University, Newport, Oregon, USA
[11]Marine and Freshwater Research Centre, Galway-Mayo Institute of Technology, Galway, Ireland
[12]Irish Whale and Dolphin Group, Merchant's Quay, Kilrush, County Clare, Ireland
[13]Southern Cross University, Marine Ecology Research Centre, Lismore, Australia
[14]South Pacific Whale Research Consortium, Avarua, Cook Islands
[15]Observatoire Pelagis, UMS3462 La Rochelle University-CNRS, La Rochelle, France
[16]Bahamas Marine Mammal Research Organisation, Abaco, Bahamas
[17]School of Biological Sciences Te Kura Mātauranga Koiora, University of Auckland Waipapa Taumata Rau, Auckland 1010, Aotearoa New Zealand
[18]Department of Anatomy, Histology and Embryology, Faculty of Veterinary Medicine, University of Zagreb, Zagreb, Croatia
[19]Southall Environmental Associates, Inc., Aptos, CA 95003, USA
[20]Pelagos Cetacean Research Institute, Vouliagmeni, Greece

[21]Museu da Baleia da Madeira, Caniçal, Madeira, Portugal
[22]Sea Mammal Research Unit, Scottish Oceans Institute, University of St Andrews, St Andrews, UK
[23]Department of Biology, Faculty of Science, University of Zagreb, Zagreb, Croatia
[24]Department of Natural Sciences, National Museums Scotland, Edinburgh, UK
[25]School of Geosciences, University of Edinburgh Drummond Street, Edinburgh, UK
[26]Society for the Study of the Cetaceans in the Canary Archipelago (SECAC), Lanzarote, Canary Islands, Spain
[27]Centro de Conservación de Manatíes del Caribe, Universidad Interamericana, Bayamón, Puerto Rico
[28]Center for Conservation Medicine and Ecosystem Health, Ross University School of Veterinary Medicine, Basseterre, St. Kitts, West Indies
[29]Centro de Estudios Avanzados en Zonas Áridas (CEAZA), La Serena, Chile
[30]Marine Mammal Research Program, Maharepa, Moorea, French Polynesia
[31]School of Biological, Earth & Environmental Sciences, University College Cork, Cork, Ireland
[32]Song of the Whale Research Team, Marine Conservation Research International, Kelvedon, Essex, UK
[33]Departamento de Ciencias Marinas y Costeras, Universidad Autonoma de Baja California Sur, La Paz, B.C.S., Mexico
[34]Hawaii Insitute of Marine Biology, University of Hawaii at Manoa, Kaneohe, HI, USA

Author contributions (my name bold and underlined):
Research and sampling design: **ABO**, NAS, ELC, PAM, MTO, OEG, MR and VM
Laboratory work: **ABO**, ELC, SSH, SDN, MTO, GG, and VM2
Analysis and writing: **ABO**, ELC, MLM and MTO,
Manuscript editing: **ABO**, MLM, GG, MMP, LF, AB, AF, AK, CR, RC, ND, RWB, SB, VM, CSB, ER, PAM, ELC, OEG and MTO.
Sample contribution: NAS, MTO, MR, MD, ND, RWB, CSB, SB, AB, DB, FC, DC, RC, FD, JD, MD2, JD2, AF, LF, AG, ACK, VM, AAM-G, AM, CO, MMP, CRS, ER, CR, AS, PT, JU and KW.

3.0 ABSTRACT

The deep sea is the largest ecosystem on Earth, yet little is known about the processes driving patterns of genetic diversity in its inhabitants. Here, we investigated the macro- and microevolutionary processes shaping genomic population structure and diversity in two poorly understood, globally distributed, deep-sea predators: Cuvier's beaked whale (*Ziphius cavirostris*) and Blainville's beaked whale (*Mesoplodon densirostris*). We used double-digest restriction associated DNA (ddRAD) and whole mitochondrial genome (mitogenome) sequencing to characterise genetic patterns using phylogenetic trees, cluster analysis, isolation-by-distance, genetic diversity, and differentiation statistics. Single nucleotide polymorphisms (SNPs; Blainville's *n*=43 samples, SNPs=13988; Cuvier's *n*=123, SNPs= 30479) and mitogenomes (Blainville's *n*=27; Cuvier's *n*=35) revealed substantial hierarchical structure at a global scale. Both species display significant genetic structure between the Atlantic, Indo-Pacific and in Cuvier's, the Mediterranean Sea. Within major ocean basins, clear differentiation is found between genetic clusters on the east and west sides of the North Atlantic, and some distinct patterns of structure in the Indo-Pacific and Southern Hemisphere. We infer that macroevolutionary processes shaping patterns of genetic diversity include biogeographical barriers, highlighting the importance of such barriers even to highly mobile, deep-diving taxa. The barriers likely differ between the species due to their thermal tolerances and evolutionary histories. On a microevolutionary scale, it seems likely that the balance between resident populations displaying site fidelity, and transient individuals facilitating gene flow, shapes patterns of connectivity and genetic drift in beaked whales. Based on these results, we propose management units to facilitate improved conservation measures for these elusive species.

3.1 INTRODUCTION

The deep sea is the largest habitat on Earth (Sutton and Milligan 2019), yet less is understood about the species that live here, their distributions, population structure and connectivity than any other ecosystem. In the deep sea, underwater features, such as steep continental slopes, canyons, and oceanic islands, can create areas of upwelling, aggregating prey species on which predators forage (Worm et al. 2003). Biodiversity hotspots are often contained within well-defined biogeographical provinces (Costello et al., 2017) where barriers may be fully impassable (e.g., Isthmus of Panama), or permeable (e.g., shallow rises; Bianchi & Morri, 2000), providing the basis for hierarchical levels of population structure in the deep sea.

Beaked whales (Ziphiidae) comprise the most speciose group of deep-sea marine mammals, yet their often elusive, offshore, and deep-diving lifestyle means that most information on these taxa comes from dead, beach-stranded specimens and several species are listed as data deficient on the IUCN Red List. Blainville's and Cuvier's beaked whales (*Mesoplodon densirostris,* Blainville 1817, and *Ziphius cavirostris,* G. Cuvier 1823*,* respectively; 'Blainville's' and 'Cuvier's' hereafter) are among the most well-studied beaked whale species, based on the number of publications. Both have cosmopolitan distributions and occupy the deep seas of all major ocean basins, although only Cuvier's is regularly found in the Mediterranean and at higher latitudes (MacLeod et al., 2006). Both species show site fidelity to some nearshore areas with steep bathymetry (Baird, 2019), potentially leading to population genetic structuring, especially in the presence of biogeographic barriers.

While many natural populations are at risk of genetic erosion due to declining abundance from human impacts (Leroy et al. 2017), some marine mammals (and particularly cetaceans) may be amongst the most susceptible taxa (Schipper et al. 2008; Cammen et al. 2016). Beaked whales are known to strand in unusual mortality events (UMEs) driven by underwater noise from human activities (Cox et al. 2006; Bernaldo de Quirós et al. 2019; Simonis et al. 2020). The majority of UMEs involving beaked whales have involved Cuvier's and/or members of the *Mesoplodon* genus (D'Amico et al. 2009). However, the extent to which these mortalities might cause long-term, population-level impacts of these poorly understood deep divers, is currently unknown. Identifying the baseline patterns of population genetic diversity and structure of these species, and understanding the underlying environmental drivers, would enable mortality to be linked with population units, thus facilitating future management efforts in the face of anthropogenic impacts.

To date, only mitochondrial DNA (mtDNA) has been analysed to investigate population structure and genetic diversity in Cuvier's and Blainville's (Dalebout et al. 2005; Morin et al. 2012). Cuvier's mtDNA haplotypes were shared between ocean basins with significant differences in haplotype frequencies, suggesting limited gene flow (Dalebout et al. 2005; Morin et al. 2012). Blainville's mitogenomes formed two reciprocally monophyletic clades according to ocean basin (Morin et al. 2012). Although these studies suggest ocean basin-level genetic structure, the paucity of beaked whale data makes it difficult to confirm these hypotheses and extend analyses to a finer geographical scale. Indeed, other deep-sea cetaceans can exist in populations that are socially-driven (e.g., sperm whales, *Physeter macrocephalus;* Alexander et al., 2016), island-associated (e.g., rough-toothed dolphins, *Steno bredanensis;* Jefferson 2018) or panmictic (e.g., Gray's beaked whale, *Mesoplodon grayi;* Westbury et al., 2021).

Here, we aim to shed light on the population genetic structure, phylogeography and diversity of Cuvier´s and Blainville´s beaked whale species across their cosmopolitan ranges to inform conservation efforts and understand the micro- and macroevolutionary processes driving genetic structure in deep-sea species. We improve the spatial resolution of previous genetic studies by Dalebout et al. (2005) and Morin et al. (2012) by identifying finer scale structure using larger sample sizes, and analysing whole mitogenomes and reduced-representation nuclear genomic data. We propose macro- and microevolutionary drivers for the substantial hierarchical genetic structure identified, which also allows us to recommend evolutionarily significant units (ESUs; Waples, 1995) and demographically independent populations (DIPs; Martien et al., 2019; Palsbøll, Bérubé, & Allendorf, 2007) for conservation and management.

3.2 METHODS

3.2.1 Sample Collection and Lab Protocols

Tissue samples were selected from the newly established International Tissue Archive of Beaked Whales (ITABW) to cover the full geographical ranges of Cuvier's and Blainville's (Electronic Supplementary Material (ESM) 1, Supplementary Spreadsheet Table (SST) 1). All sampling was done under appropriate local permits and, if applicable, local animal ethics approvals. DNA was extracted from tissue samples (ESM1) and sequencing libraries were prepared following the method described by Peterson et al. (2012) and optimised for beaked whales (Carroll et al. 2016; Carroll et al. 2021) (ESM2). To explore major phylogeographical patterns, and supplement the nuclear analyses, we

generated and analysed complete mitogenomes from a subset of the samples used for ddRAD as well as additional low-quality museum samples by "shotgun" sequencing (Carøe et al., 2018; ESM2 for protocol), and augmented with publicly available sequences (Morin et al. 2012; SST1).

3.2.2 Bioinformatics

DdRAD SNP discovery and filtering followed the pipeline of Carroll et al. (2021) and the full protocol is available at https://github.com/aono87/bw_reference_mapped. In summary, sequence quality was inspected using FastQC v.0.11.8a (Andrews et al. 2010) and 'process_radtags' in Stacks (v. 2.0Beta9; Rochette, Rivera-Colón, & Catchen, 2019) was used to demultiplex, quality check and trim reads. Sequences were aligned to a reference genome using BWA-MEM v. 0.7.15 (Li 2013), converted and sorted using SAMtools v. 1.6 (Li et al. 2009), and assigned read groups using PICARD TOOLKIT v. 2.14.1 (Broad Institute, 2019). Cuvier's sequences were aligned to the Cuvier's reference genome (GenBank accession: GCA_004364475.1), and Blainville's sequences were aligned to the Sowerby's beaked whale reference genome (*Mesoplodon bidens*, GenBank accession: GCA_004027085.1), the closest relative available at the time of analysis. DdRAD parameter optimisation, SNP genotyping methods and data quality control are described in detail in ESM3 and ESM4, following previously described methods (Carroll et al. 2016; O'Leary et al. 2018; Carroll et al. 2021).

Mitochondrial sequencing reads were inspected with FastQC. Adapters and sequences were trimmed (ktrim=r, k=23, mink=8, hdist=1, qtrim=rl, trimq=15, maq=20, and minlen=40) using 'bbduk.sh' (https://jgi.doe.gov/data-and-tools/bbtools/bb-tools-user-guide/bbduk-guide/). Reads were mapped to a reference mitogenome for each species (Cuvier's GenBank accession: NC_021435.1; Blainville's: NC_021974.2) using BWA-MEM and PCR duplicates were removed with SAMtools v.1.9. The consensus mitogenome was extracted using ANGSD v.0.931 (Korneliussen et al. 2014) with the filters -doFasta 2 -doCounts 1 -minQ 30 -minMapQ 30 -setMinDepth 3. Depth of coverage was assessed using SAMtools.

3.2.3 Phylogeography

Global structure based on ddRAD SNPs was visualized by generating rooted neighbour-joining trees (BIONJ, Gascuel, 1997) for each species, using sequences from southern right whales (SRW, *Eubalaena australis*) as the outgroup (ESM5). Full mitogenomic sequences were analysed in a Bayesian phylogenetic framework implemented in BEAST v.2.5.2 (Bouckaert et al. 2019). The 12srRNA, 16srRNA and 13 protein-coding genes from each unique mitogenome were individually aligned, and the stop codons for the coding genes were removed in Geneious Prime v2020.1.2 (https://www.geneious.com/). The final alignments from the two rRNA and 13 protein-coding genes were subsequently concatenated into a single alignment, and substitution models were inferred for each codon position for each gene with PartitionFinder v.2.1.1 (Lanfear et al. 2017). Control region (CR) data were analysed separately (see below). The resultant partitions were incorporated into BEAUTI, the graphical user interface for generating BEAST XML files (SST2).

For the Cuvier's mitogenome phylogeny, a strict clock with a Yule tree model was applied, as closely related species are not expected to have deviating clocks. Calibration dates followed most recent common ancestor (MRCA) estimates inferred from McGowen et al. (2019). Baird's beaked whale (*Berardius bairdii;* GenBank accession:

NC_005274.1) was used as an outgroup and the MRCA between all Cuvier's and the Baird's beaked whale outgroup was set to a mean age of 15 million years ago (mya; 95% probability range 12.3-17.6). The model was set to a 15 million chain length, sampled every 1,500 steps, and replicated three times to assess model convergence. Log and tree files were combined with LogCombiner v.2.5.2 (Drummond and Rambaut 2007) after a 10% burn-in for each run. All parameters in the combined log file had ESS values greater than 200. The Maximum Clade Credibility (MCC) tree was prepared in TreeAnnotator v.2.5.2 (Bouckaert et al. 2019) and visualised in FigTree v.1.4.4 (http://tree.bio.ed.ac.uk/software/figtree/).

For the Blainville's mitogenome phylogeny, a strict clock with a Yule tree model was also applied. Two outgroups were used as they are both closely related to Blainville's (Gray's beaked whale, GenBank accession: NC_023830.1 ; Stejneger's beaked whale, *Mesoplodon stejnegeri,* GenBank accession: NC_036997.1) and their nodes for the MRCA were also calibrated using estimates from McGowen et al. (2019). The calibration node for the Blainville's alignments and the two outgroups was set to the mean age 5.19 mya (95% probability range 4.12-6.26 mya). The calibration node for the Blainville's alignments with the Stejneger's beaked whale outgroup was set to the mean age 4.59 mya (95% probability range 3.6-5.58 mya). The Blainville's model needed fewer iterations to converge than Cuvier's and was set to a 10 million chain length, sampled every 1000 steps, and replicated three times to check for model convergence. The tree and log files were combined, and the output visualized as described for Cuvier's.

3.2.4 Genetic Population Structure

Isolation by distance (IBD) was calculated per species using a Mantel test in *ade4* v1.7-16 in R (Dray and Dufour 2007) to compare Euclidean genetic distance for the ddRAD SNPs and geographical distances calculated via the least cost (LC) path distance over seawater in *marmap* v1.0.5 (Pante and Simon-Bouhet 2013) (ESM6).

We conducted a hierarchical clustering analysis incorporating ddRAD SNPs and geographical information in a spatially explicit, least-squares optimization approach to estimate ancestry in the R package *tess3r* (v1.1.0, Caye, Jay, Michel, & Francois, 2018). The most likely number of clusters was estimated from a range (K=2-10) with n=20 repetitions and n=200 maximum iterations of the optimisation algorithm. Cross-entropy scores were plotted against K values to infer the most likely number (ESM7). When hierarchical spatial structure is likely, genetic clustering algorithms tend to find structure at the highest level of divergence (Evanno et al. 2005). Previous mtDNA work (Dalebout et al. 2005; Morin et al. 2012) suggests ocean basin-level differentiation, so we carried out analyses both between and within ocean basins to allow for the detection of finer-scale structure. The genetic clusters determined using *tess3r* were further explored using Discriminant Analysis of Principal Components (DAPC; ESM8; Jombart, Devillard, & Balloux, 2010). A simulation study suggests that an *a priori* approach to DAPC successfully resolves the composition of the genetic clusters and the genetic distance between clusters correlates well with the value of F_{ST} used to generate the simulated data (Miller et al. 2020). On the other hand, *de novo* DAPC analyses perform poorly in situations that are common to marine species; low genetic differentiation and IBD (Jombart et al. 2010; Miller et al. 2020). Therefore, while we explored both *de novo* and *a priori* DAPC analyses, we only report on the *a priori* results.

We measured the amount of genetic differentiation in the ddRAD data between the defined clusters by calculating pairwise F_{ST} and net nucleotide divergence (d_A, Nei, 1987, equation 10.21; Weir & Cockerham, 1984) and F_{ST} 95% confidence intervals (*dartr* v1.3.3, based on 100 bootstraps; Gruber, Unmack, Berry, & Georges, 2018). Statistical significance of F_{ST} and d_A was determined by permutation test (n=1000 permutations; Archer et al., 2017).

In addition, for future comparisons with other studies and criteria recently suggested for species, subspecies and population delineations in cetaceans (Taylor et al., 2017), we estimated F_{ST} (and associated p-value) and d_A between ocean basins for full mitogenomes and just the CR with the software package DNAsp v6.12.03 (Rozas et al. 2017).

3.2.5 Genetic Diversity and Demographic History

To investigate differences in nuclear genetic diversity across defined populations, we estimated observed heterozygosity (H_o) and gene diversity (H_s) (Nei 1987b), the inbreeding coefficient (F_{IS}) (Weir and Cockerham 1984) using *hierfstat* v.0.4.22 (Goudet 2005), and Tajima's D (Tajima 1989) using VCFtools v 0.1.12a (Danecek et al. 2011) for the SNP dataset. Diversity and inbreeding statistics were calculated per locus and Tajima's D was calculated based on a 100 kb sliding window for each genetic cluster. Confidence intervals (95%) were calculated around mean H_o, H_s and Tajima's D using *rcompanion* (v2.3.25; n=1000 boostraps; Mangiafico, 2020) and "boot.ppfis" in *hierfstat* (n=100 bootstraps) for the F_{IS} values. Mitochondrial DNA genetic diversity (segregating sites (S), number of haplotypes (h), haplotype diversity (*Hd*) and nucleotide diversity (π)) were calculated separately for the mitogenome and CR data, for each species and ocean basin, using DNAsp.

3.3 RESULTS

The final Cuvier's dataset included 30479 ddRAD SNPs (mean=29697 SNPs per individual, SE=72) from 123 Cuvier's individuals (Figure 3-1, ESM9) as well as mitogenomes (15291 bp) and CR (860 bp) sequences from 35 individuals (n=22 downloaded from GenBank, SST3; n=13 generated for this study, SST4). The final Blainville's dataset included 13988 ddRAD SNPs (mean=13760 SNPs per individual, SE=64) from 43 individuals (Figure 3-2, ESM9), as well as mitogenomes (14147 bp) and CR (852 bp) sequences from 27 individuals (n=15 downloaded from GenBank, SST3; n=12 generated for this study, SST4). Overall, both ddRAD datasets had low levels of missing data (2.6% in Cuvier's and 1.6% in Blainville's across the entire datasets) and high depth of coverage per ddRAD SNP (Cuvier's: mean=59x, SE=4; Blainville's: mean=60x, SE=4, ESM9) and mitogenome (Cuvier's: mean=156x, SE=23; Blainville's: mean=158x, SE=21) (SST4). The final datasets used to generate the ddRAD phylogenetic trees were n=118 Cuvier's (n=33137 SNPs) and n=42 Blainville's (n=29904 SNPs), with the six Southern right whales as the outgroup (ESM5).

3.3.1 Cuvier's beaked whales

3.3.1.1 Phylogeography and Genetic Population Structure

The Cuvier´s exhibited substantial phylogeographic and population genetic structure in both the nuclear ddRAD SNPs and the mitogenomes. The mitogenome phylogeny of 35 Cuvier's indicated a separation into three major clades with high levels of support (Figure 3-3a); Clade 1 contained lineages sampled exclusively in the North Atlantic (including

the Mediterranean) ,whereas Clades 2 and 3 contained individuals from the western North Atlantic, and divided Indo-Pacific lineages (and one from Argentina) broadly along the equator (Figure 3-3a, see SST1 for individual metadata and clade assignments). Individuals from the North Atlantic (Clade 1) appear to have diverged from the mostly Indo-Pacific Clade 2 approximately 1.5 mya, while the individuals in the Mediterranean diverged from the rest of the North Atlantic approximately 0.5 mya (Figure 3-3a). Unlike the paraphyletic mitogenome phylogeny, the BIONJ phylogeny based on the final ddRAD SNP data consisting of 33137 SNPs genotyped in 118 Cuvier's (ESM5) and 6 SRW outgroup showed clearer monophyly of Atlantic and Pacific Ocean clades (Figure 3-3b). Like the mitogenome phylogeny, the Mediterranean forms a separate monophyletic clade in the ddRAD BIONJ phylogeny (Figure 3-3b and S5.1).

In the *tess3r* analysis based on ddRAD SNPs, the most likely number of genetic clusters for the 123 Cuvier's was K=3 (Figure 3-1a and S7.1), clearly dividing samples from the Atlantic, Mediterranean and Indo-Pacific (Figure 3-1a). This was supported by statistically significant levels of pairwise differentiation between the Atlantic and the Indo-Pacific in both mtDNA (mitogenome F_{ST} =0.178, CR F_{ST} =0.241, p<0.01, SST5) and nuclear markers (ddRAD F_{ST} =0.018, p<0.01, SST6). Substantial differentiation was also seen between the Mediterranean and the Atlantic and Indo-Pacific in both the mtDNA (mitogenome F_{ST} =0.623 and 0.375, CR F_{ST} =0.531 and 0.347, respectively, SST5) and nuclear (ddRAD F_{ST} =0.184 and F_{ST} =0.197, p<0.01, respectively; SST6) markers. Net nucleotide differences showed the same pattern as the CR (SST5).

In addition to ocean-level genetic structure, the ddRAD SNPs revealed finer scale structure within ocean basins. The most likely number of genetic clusters for Cuvier's within the North Atlantic was K=5 clusters, K=3 in the Indo-Pacific and K=2 in the Mediterranean (Figure 3-1b-e and S7.1; ESM7). In the Mediterranean (Figure 3-1b-c), distinct structure was present between the west and east. Within the North Atlantic, there was little admixture and individuals grouped into five discrete clusters, reflecting their general sampling location (Figure 3-1d). Four individuals from Puerto Rico and the Virgin Islands shared low levels of admixed ancestry (<30%), suggesting connectivity with the other clusters. One cluster consisted only of two individuals from Spain and was not included in further analyses.

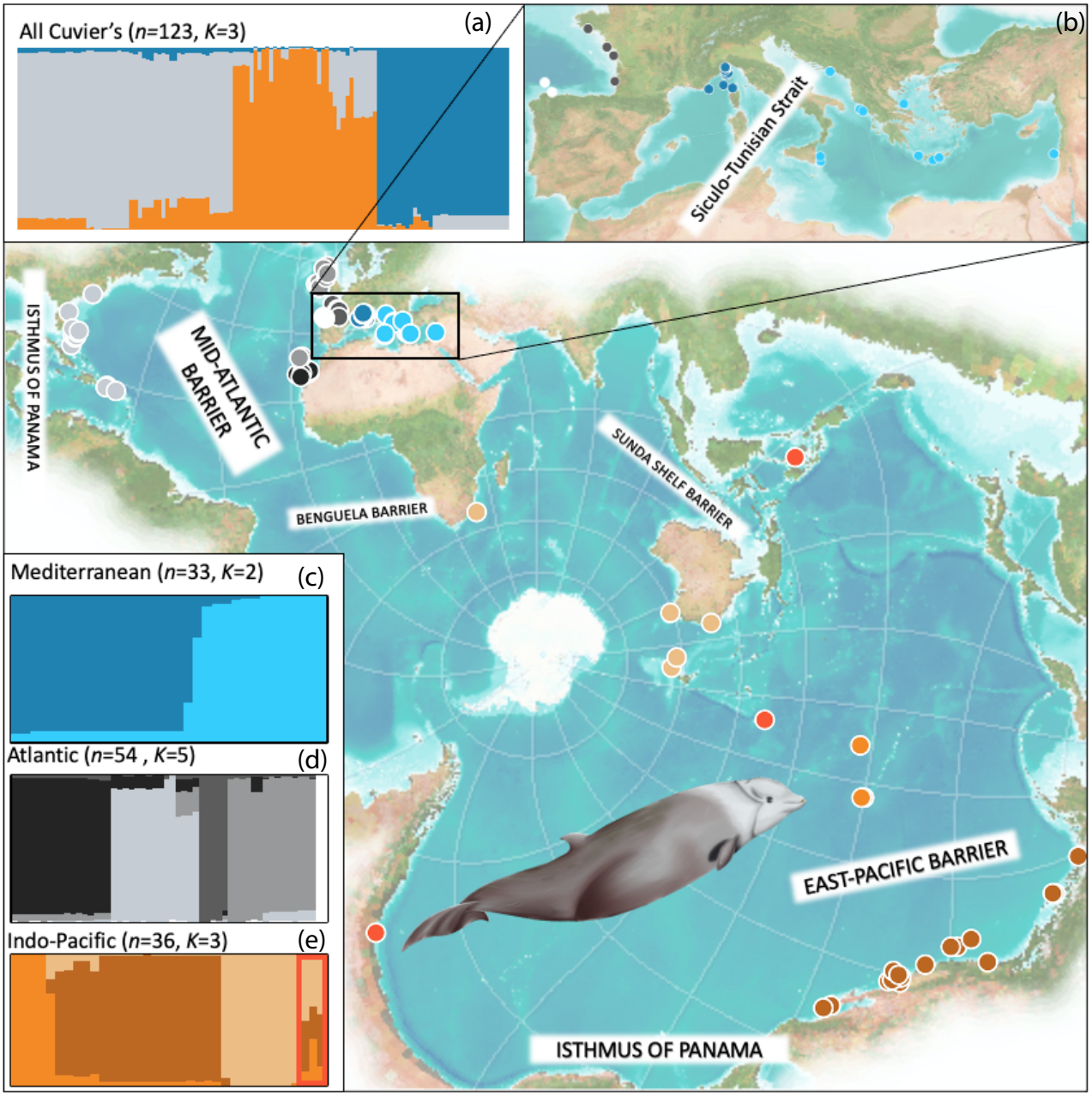


Figure 3-1. Sampling locations and genetic clustering of Cuvier's beaked whales (*Ziphius cavirostris,* *n*=123 individuals, *n*=30479 SNPs). Main map shows sample locations plotted on the Spilhaus projection, to highlight the connectivity and continuity of the world's oceans for a globally distributed species, with names and locations of relevant major biogeographical barriers. Bar plots of co-ancestry coefficients generated in *tess3r* are presented for (a) all samples, (c) Mediterranean (location shown more detail in (b) map inset), (d) Atlantic and (e) Indo-Pacific. The three Indo-Pacific samples enclosed in a red box represent the admixed individuals and are plotted on the map as red points.

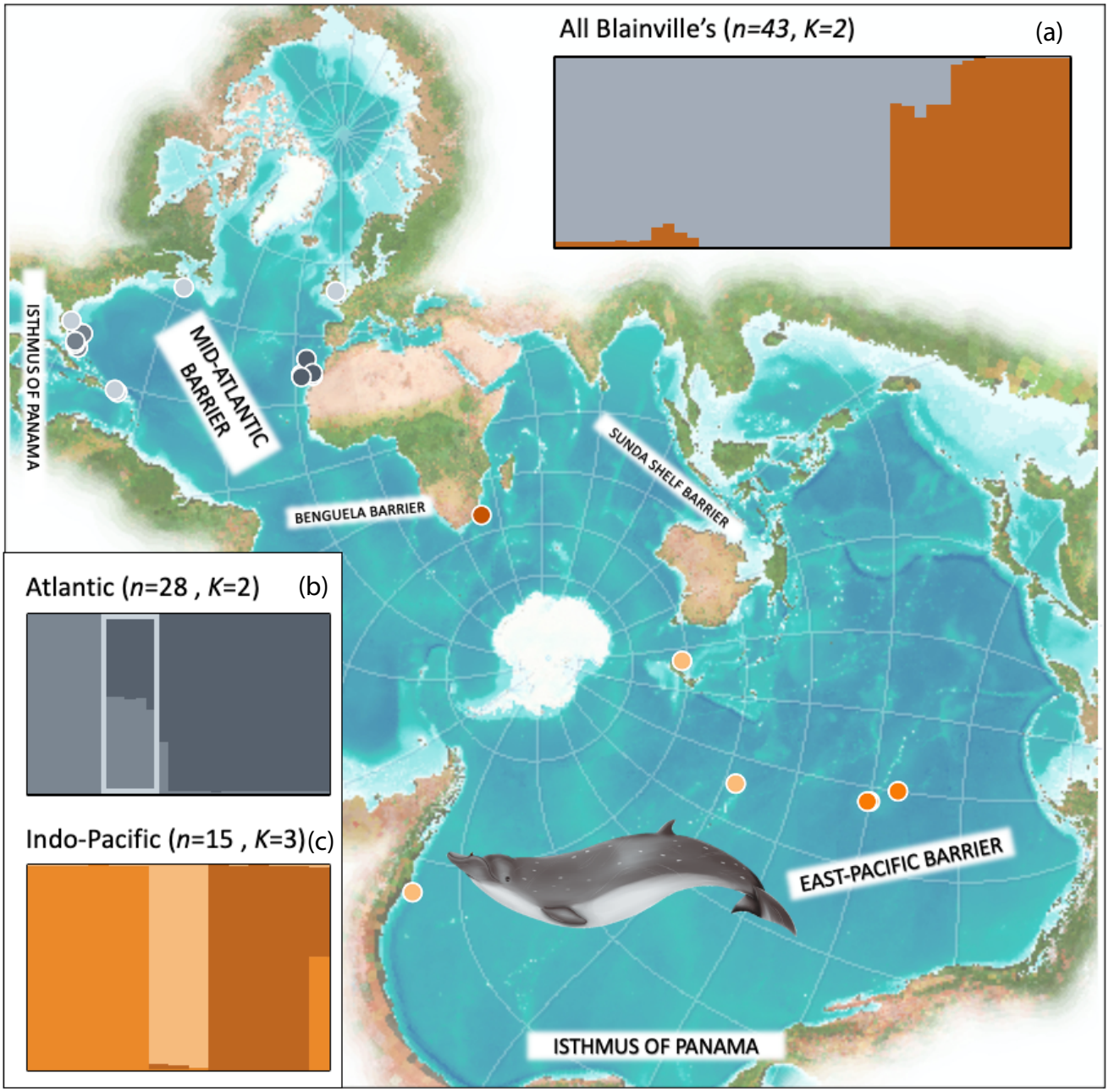


Figure 3-2. Sampling locations and genetic clustering of Blainville's beaked whales (*Mesoplodon densirostris, n*=43 individuals, *n*=13988 SNPs). Main map shows sample locations plotted on the Spilhaus projection, to highlight the connectivity and continuity of the world's oceans for a globally distributed species, with names and locations of relevant major biogeographical barriers. Bar plots of co-ancestry coefficients generated in *tess3r* are presented for (a) all samples, (b) Atlantic and (c) Indo-Pacific.

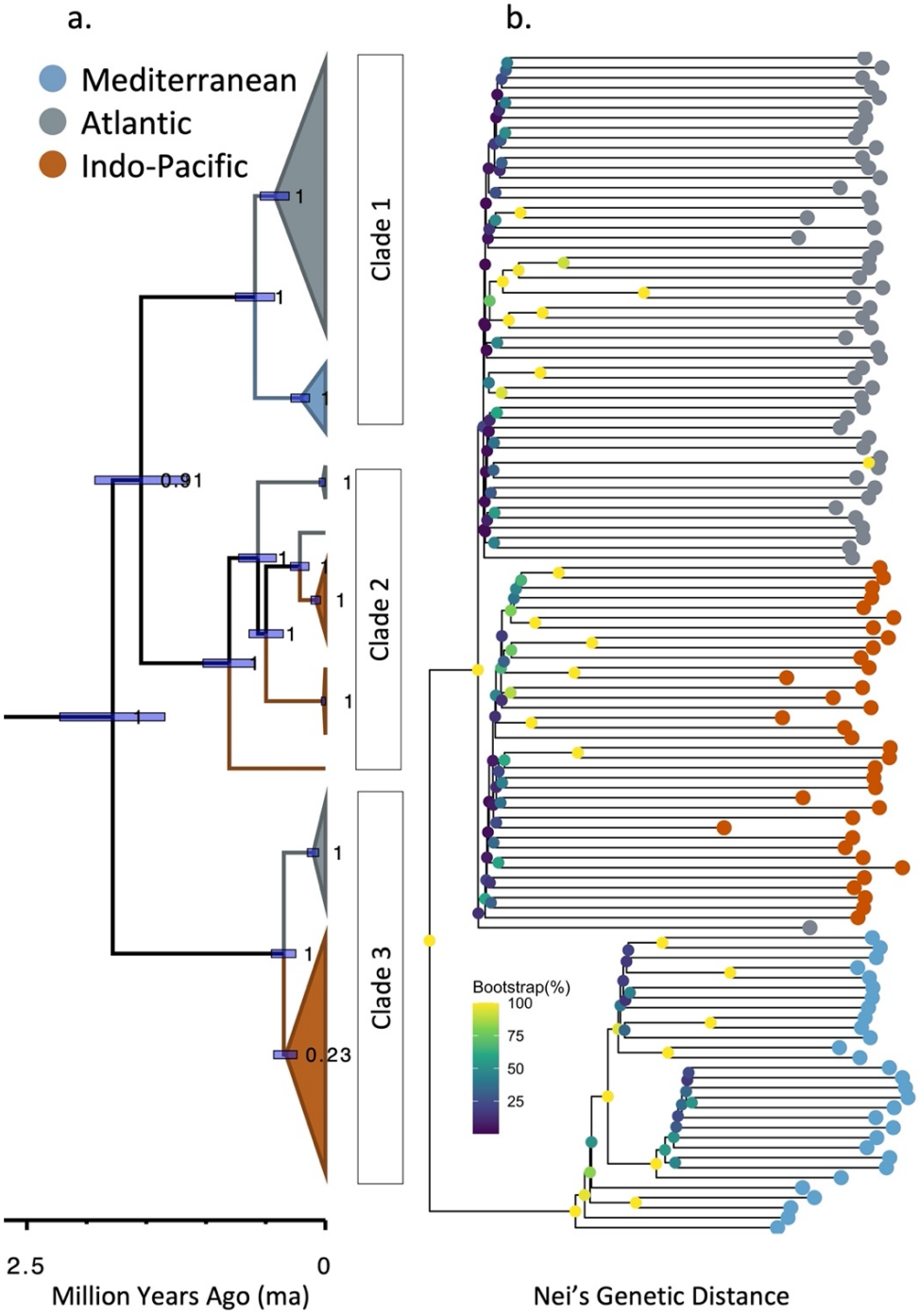


Figure 3-3. Phylogenetic relationship among globally distributed and sampled Cuvier's beaked whales (Ziphius cavirostris). (a) Bayesian mitogenome phylogeny generated from 32 unique Cuvier's mitogenome haplotypes with posterior estimates (as a proportion) with divergence shown in million years ago and the purple bars represent the 95% probability range of divergence dates (HPD). (b) BIO neighbour-joining (BIONJ) phylogeny of 118 Cuvier's beaked whales with bootstrapped support for each node >50% (n=1000 bootstraps) generated using 33137 nuclear SNPs. The six Southern right whales (*Eubalaena* australis) used to root the tree were removed prior to plotting (see Figure S5.1 for tree with outgroup).

All pairwise ddRAD SNP F_{ST} values were significant ($p<0.05$), except for between the France and Canary Island clusters in the Atlantic (SST6).

Within the Indo-Pacific, Cuvier's subclusters also corresponded with the geographical origin of samples (Figure 3-1e). The broadest distribution of samples was found in the South Indo-Pacific (Indo_Sou). Three admixed individuals from the Philippines, Chile and Samoa shared most of their ancestry coefficient with the remaining South Indo-Pacific individuals (South Africa, Australia, and Aotearoa New Zealand), and were not included in the remaining analyses.

As DAPC and *tess3r* analyses produced similar results, and because *tess3r* is expected to perform better in the presence of admixture and IBD (reviewed in Francois and Waits 2016), we present the DAPC analyses primarily in the Supplementary Material (ESM8). The first axis in the DAPC separates the two Mediterranean clusters from the Atlantic and Indo-Pacific ones (Figure S8.1). The second axis separates the Indo-Pacific from the Atlantic, as well as the east and west Mediterranean. Finer scale differentiation between clusters within ocean basins is observed when the 2nd vs 3rd (Figure S8.3) and 3rd vs 4th (Figure S8.4) axes are plotted. Samples were reassigned to the a-priori clusters with high probability, except for a few samples in the east Atlantic and Indo-Pacific clusters (S8.1).

Isolation-by-distance was present and significant when all Cuvier's were combined and within ocean basins (ESM6, Table S6.1).

3.3.1.2 Genetic Diversity and Demography

Genetic diversity summary statistics (H_o, H_s, F_{IS}, and Tajima's D) and their 95% confidence intervals are presented for the 118 Cuvier's that were assigned a genetic cluster (i.e., not admixed, Table 3-1). Both the overall Mediterranean dataset and populations therein had positive values of Tajima's D (indicative of population contraction), the lowest levels of diversity and inbreeding (H_o, H_s, and F_{IS}; Table 3-1), and the lowest mtDNA diversity (π and S, Table 3-1, SST5). Slightly greater overall diversity values were seen in the North Atlantic, compared to the Indo-Pacific in both the ddRAD and mtDNA data (Table 3-1). All sample partitions outside of the Mediterranean had negative values of Tajima's D, indicative of population expansion (Table 3-1).

F_{is} ranged widely amongst Cuvier's clusters within ocean basins (F_{IS}=0.077-0.141; Table 3-1). Mitogenome haplotype diversity was very high, with only one haplotype shared between two individuals from the Bahamas, and one shared between two Canary Islands and one Irish individual (Table SST7).

Table 3-1. Nuclear diversity statistics calculated from ddRAD SNPs (Cuvier's: n=30479, Blainville's: n=13988): Mean and 95% confidence intervals (in brackets) are provided for each ocean basin and genetic cluster of observed heterozygosity (H_o; Nei 1987) gene diversity (H_s; Nei 1987), inbreeding coefficient (F_{IS}; Weir & Cockerham, 1984) and Tajima's D (100kb sliding window). Mitogenomic diversity statistics are provided for mitogenomes and mtDNA CR (in brackets) for each ocean basin: no. segregating sites (S), no. of haplotypes (h), haplotype diversity (Hd) and nucleotide diversity (π).

	Population	ddRAD					mtDNA				
		n	H_o	H_s	F_{IS}	Tajima's D	n	S	h	Hd	π
Cuvier's	**Atlantic**	**54**	**0.112 (0.111, 0.113)**	**0.128 (0.126, 0.129)**	**0.121 (0.117, 0.125)**	**-0.279 (-0.289, -0.270)**	**18**	**462 (14)**	**17 (9)**	**0.993 (0.882)**	**0.0097 (0.0048)**
	Atl_Carib	15	0.113 (0.111, 0.114)	0.127 (0.125, 0.128)	0.115 (0.110, 0.121)	-0.33 (-0.342, -0.318)					
	Atl_CanIs	17	0.111 (0.110, 0.113)	0.126 (0.124, 0.128)	0.121 (0.115, 0.127)	-0.304 (-0.316, -0.291)					
	Atl_France	5	0.112 (0.110, 0.114)	0.127 (0.124, 0.129)	0.116 (0.108, 0.125)	-0.25 (-0.267, -0.233)					
	Atl_NE	15	0.112 (0.111, 0.114)	0.126 (0.124, 0.127)	0.110 (0.104, 0.115)	-0.374 (-0.387, -0.361)					
	Indo-Pacific	**36**	**0.109 (0.107, 0.110)**	**0.123 (0.122, 0.125)**	**0.118 (0.114, 0.122)**	**-0.306 (-.0316, -0.295)**	**14**	**344 (9)**	**13 (6)**	**0.989 (0.747)**	**0.0079 (0.0041)**
	Indo_Central	5	0.103 (0.101, 0.105)	0.119 (0.117, 0.121)	0.141 (0.131, 0.151)	-0.177 (-0.195, -0.159)					
	Indo_NE	19	0.111 (0.109, 0.112)	0.122 (0.121, 0.124)	0.097 (0.092, 0.102)	-0.302 (-0.314, -0.290)					
	Indo_South	9	0.110 (0.108, 0.112)	0.124 (0.122, 0.125)	0.110 (0.104, 0.118)	-0.337 (-0.351, -0.322)					
	Mediterranean	**33**	**0.078 (0.077, 0.080)**	**0.089 (0.088, 0.091)**	**0.122 (0.116, 0.128)**	**0.443 (0.424, 0.462)**	**3**	**47 (5)**	**3 (3)**	**1.000 (1.000)**	**0.0021 (0.0039)**
	Med_East	14	0.072 (0.070, 0.074)	0.078 (0.076, 0.080)	0.080 (0.071, 0.088)	0.386 (0.363, 0.408)					
	Med_West	19	0.083 (0.082, 0.085)	0.090 (0.089, 0.092)	0.077 (0.070, 0.084)	0.407 (0.388, 0.427)					
Blainville's	**Atlantic**	**28**	**0.105 (0.103, 0.108)**	**0.114 (0.112, 0.117)**	**0.078 (0.070, 0.085)**	**-0.073 (-0.097, -0.050)**	**16**	**164 (9)**	**13 (7)**	**0.975 (0.800)**	**0.0028 (0.0030)**
	Atl-Bah	7	0.102 (0.099, 0.105)	0.110 (0.107, 0.113)	0.064 (0.048, 0.080)	0.038 (0.007, 0.065)					
	Atl-East	16	0.106 (0.104, 0.109)	0.113 (0.110, 0.115)	0.073 (0.059, 0.087)	0.015 (-0.012, 0.041)					
	Atl-Oth	5	0.106 (0.103, 0.110)	0.114 (0.110, 0.117)	0.057 (0.048, 0.066)	-0.017 (-0.046, 0.014)					
	Indo-Pacific	14	**0.127 (0.125, 0.129)**	**0.141 (0.138, 0.143)**	**0.098 (0.090, 0.105)**	**-0.450 (-0.468, -0.431)**	**11**	**312 (11)**	**10 (9)**	**0.982 (0.964)**	**0.0050 (0.0040)**
	Indo-Haw	6	0.125 (0.121, 0.128)	0.142 (0.139, 0.145)	0.124 (0.111, 0.137)	-0.285 (-0.309, -0.260)					
	Indo-Afr	5	0.128 (0.125, 0.131)	0.137 (0.133, 0.140)	0.062 (0.052, 0.076)	-0.263 (-0.287, -0.238)					
	Indo-Sou	3	0.127 (0.124, 0.131)	0.138 (0.135, 0.142)	0.079 (0.060, 0.095)	-0.217 (-0.245, -0.191)					

3.3.2 Blainville's beaked whales

3.3.2.1 Phylogeography and Genetic Population Structure

The mitogenome phylogeny of 23 unique Blainville's lineages showed clear division between the Indo-Pacific and North Atlantic which occurred ~0.75 mya (
Figure 3-4a). One individual from Car Nicobar, Bay of Bengal, was nested within the North Atlantic clade, near the base of the group (Figure 3-4a, SST1). No ddRAD data were available for this individual to compare locations in the phylogenies. Significant clustering according to ocean basin (100% bootstrap support) was also found in the BIONJ phylogeny of 42 ddRAD Blainville's and the 6 SRW used as an outgroup (Figure 3-4b and S5.2). One individual sampled in South Africa was nested within the Atlantic individuals. No mitogenome data were available to compare phylogenies for this individual.

In the *tess3r* analyses based on ddRAD SNPs, the most likely number of genetic clusters of the n=43 Blainville's was K=2, corresponding with North Atlantic and Pacific Oceans (Figure 3-2 and S7.2). Blainville's samples from South Africa also clustered with individuals from the Pacific, forming an Indo-Pacific cluster. The pairwise difference between the North Atlantic and Indo-Pacific clusters in Blainville's was substantial and statistically significant (ddRAD F_{ST} = 0.119, p<0.001, SST6), and greater than the parallel differentiation between North Atlantic and Indo-Pacific Cuvier's clusters (ddRAD F_{ST}=0.018, p<0.01, SST6). Mitogenomic differentiation between the North Atlantic and Indo-Pacific was also statistically significant (p<0.01 for all comparisons) and extremely high (mitogenome F_{ST}=0.711, CR F_{ST}=0.328, SST5). Net nucleotide divergence between ocean basins was d_A =0.0097 (mitogenome) and d_A =0.0017 (CR) (SST5).

Like the Cuvier's, the Blainville's ddRAD SNP data revealed considerable within-ocean-basin structure, with individuals grouping best into K=2 Atlantic and K=3 Indo-Pacific genetic clusters (Figure 3-2a-c). Five individuals from across the North Atlantic (Atl_Oth, Figure 3-2b), clustered together with equal amounts of admixture from the Bahamas and eastern North Atlantic clusters. Blainville's from South Africa (Indo_Afr) formed a cluster separate from the large South Pacific (Indo_Sou) cluster (Figure 3-2c). One sample from New Zealand was equally admixed between the South African and Hawaiian clusters and was therefore removed from the remaining analyses. All pairwise F_{ST} comparisons of clusters between ocean basins were highly differentiated and significant (SST6).

In support of the *tess3r* analyses, the first axis in the DAPC broadly separates the Atlantic from the Indo-Pacific clusters, with the South African cluster spanning both sides (Figure S8.2). The second axis separates the east from the west Atlantic clusters while in the Indo-Pacific, it divides the Hawaiian and southern Indo-Pacific individuals. Individuals were reassigned to the prior clusters with high probability and few misassignments (Figure S8.1).

Finally, some IBD was found for all Blainville's together and within ocean basins (ESM6, Table S6.1), with two clearly separated patches of points in the Atlantic Blainville's suggesting genetic sub structuring (ESM6, Figure S6.2).

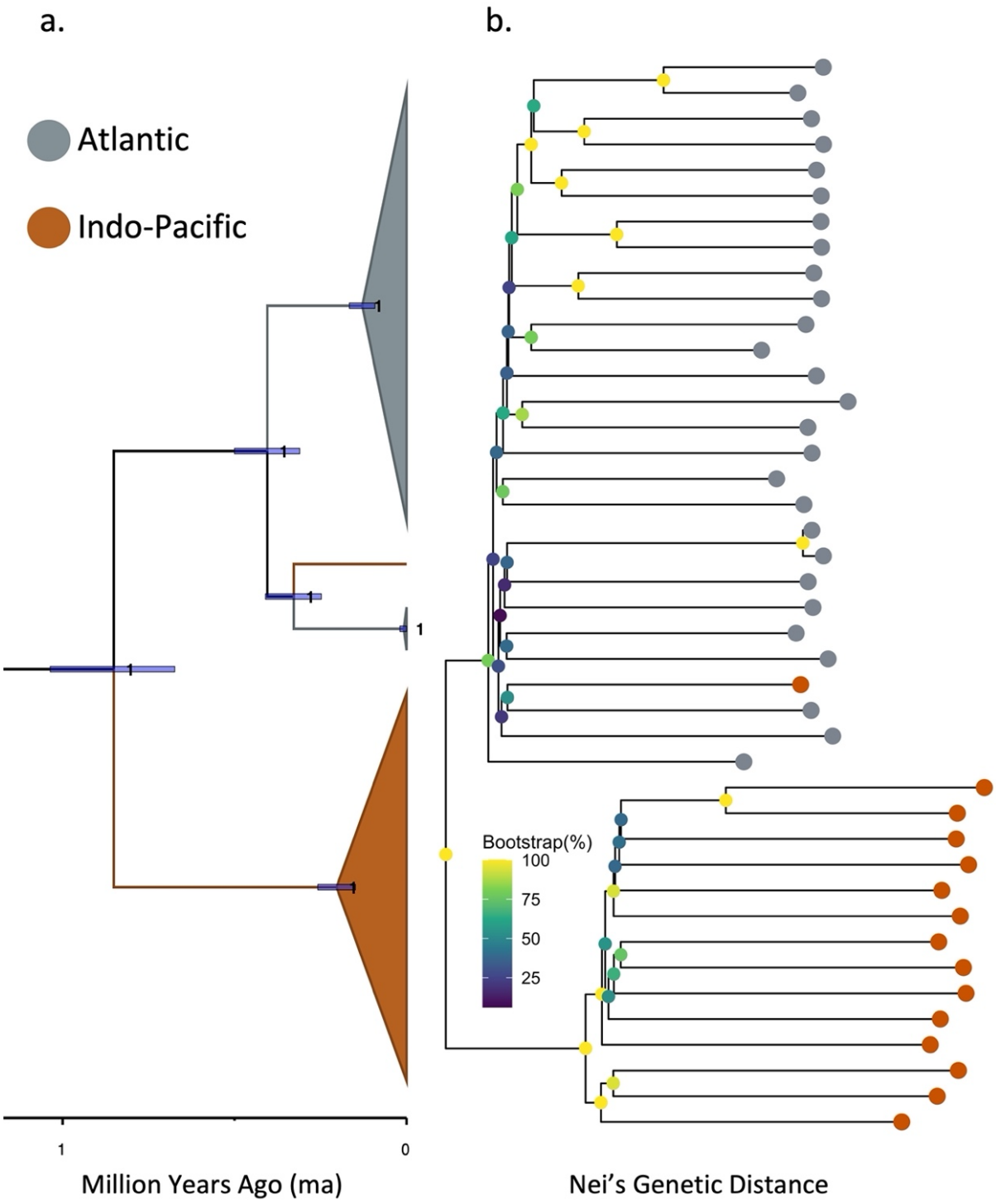


Figure 3-4. Phylogenetic relationship among globally distributed and sampled Blainville's beaked whales (*Mesoplodon densirostris*). (a) Bayesian mitogenome phylogeny generated from 23 unique Blainville's mitogenome haplotypes with posterior estimates (as a proportion) with divergence shown in million years ago (MYA) and the purple bars represent the 95% probability range of divergence dates (HPD). (b) BIO neighbour-joining (BIONJ) phylogeny of 42 Blainville's beaked whales with bootstrapped support for each node >50% (*n*=1000 bootstraps) generated using 29904 SNPs. The six Southern right whales (*Eubalaena australis*) used to root the tree were removed prior to plotting (for tree with outgroup see Figure S5.2).

3.3.2.2 Genetic Diversity and Demography

Nuclear diversity summary statistics and their 95% confidence intervals are presented for 42 Blainville's that assigned to a cluster and mtDNA statistics are provided for whole mitogenomes and mtDNA CR data in Table 3-1. The overall Indo-Pacific dataset and clusters therein had higher levels of diversity (H_o and H_s) and inbreeding (F_{IS}) compared with the North Atlantic cluster, a pattern also reflected in the mitochondrial data (Table 3-1). One haplotype was shared between two individuals sampled in Hawai'i and three pairs from the Bahamas shared unique haplotypes (SST7). Both ocean-basin clusters had negative Tajima's D values, though the North Atlantic value was very close to 0 (D=-0.073) indicating long term population size stability. Within ocean basins, all Indo-Pacific clusters and only the admixed North Atlantic (Atl-Other) cluster had negative Tajima's D values (Table 3-1), indicative of increasing population sizes. The remaining North Atlantic clusters showed positive Tajima's D values, signalling a population decline (Table 3-1).

3.4 DISCUSSION

Here we show that two deep-sea predators, the Cuvier's and Blainville's beaked whales, exhibit substantial hierarchical population structure on a global scale and substantial genetic differentiation at within ocean basin scales. At a macroevolutionary scale, this suggests that patterns of spatial genetic structure can vary between deep-sea species depending on their evolutionary history, e.g., as other similar predators demonstrate panmixia (Winkelmann et al. 2013; Westbury et al. 2021) or isolation-by-distance (Amaral et al. 2012; Gonçalves da Silva et al. 2020). At a microevolutionary scale, we propose that patterns of gene flow and genetic drift are shaped by life history traits and site fidelity in Blainville's and Cuvier's. These microevolutionary processes may be similar to those thought to influence other deep-sea associated marine mammals, such as island preference (Albertson et al. 2017; VanCise et al. 2017), social structure and prey specialisation (Foote et al. 2016; Martien, Taylor, et al. 2019).

Some level of structure has been identified in many of the recently published studies on beaked whale population genetics (though is notably absent from Gray's beaked whale, Westbury et al., 2021). At a global scale, recent work indicated that True's beaked whales (*Mesoplodon mirus*) in the Northern and Southern Hemispheres had been separated for 0.5-2 mya, resulting in the designation of a new species in the south (*M. eueu,* Carroll et al., 2021). At an ocean-basin-level, population structure or IBD was detected in Baird's beaked whale (*Berardius bairdii,* Morin et al., 2017) across the North Pacific, and the genetic divergence measured between the "black" and "gray" stocks using the mtDNA CR was found to represent separate species within the genus (Morin et al. 2017; Yamada et al. 2019). The beaked whale species with the most genetic studies is the northern bottlenose whale (*Hyperoodon ampullatus*), which is found exclusively in the North Atlantic (Dalebout et al. 2001; Dalebout et al. 2006; Feyrer et al. 2019; de Greef et al. 2022). The culmination of these publications is the understanding that northern bottlenose whales show genetic structure and IBD across their range, likely through philopatry linked to bathymetric features like the "Gully" canyon (Dalebout et al. 2006), with resulting conservation management implications (COSEWIC 2011). Adding to the sparse knowledge about beaked whale population structure, we talk about drivers of Cuvier's and Blainville's genetic structure at macro- and microevolutionary scales and discuss the conservation management implications of this work.

3.4.1 Macroevolutionary drivers of genetic population structure in Cuvier's and Blainville's

Biogeographic barriers on both global and regional scales (Toonen et al. 2016), have been known to influence patterns of genetic structure in widely-distributed marine species (e.g., Clarke et al., 2015; Gaither, Bowen, Rocha, & Briggs, 2016; Leslie, Archer, & Morin, 2018). Such barriers can be fully impassable (if only recently in geological history; such as the Isthmus of Panama), whereas others are physically passable features, such as shallow rises (e.g., the Siculo-Tunisian Strait; Bianchi & Morri, 2000; Mejri, Lo Brutto, Hassine, & Arculeo, 2009), and expanses of deep sea (e.g., East Pacific and Mid-Atlantic Ridge Barriers; Toonen, Bowen, Iacchei, & Briggs, 2016). At a global scale, the genetic structuring found here in Cuvier's and Blainville's appears to be shaped by such biogeographic barriers previously highlighted in the marine realm, including the Isthmus of Panama, East-Pacific Barrier, Mid-Atlantic Ridge, Benguela and Sunda Shelf. These barriers vary in permeability between our focal species, primarily due to differences in the timing and geography of each species' origin and spread across the globe (e.g., before or after the formation of the Isthmus of Panama).

Blainville's demonstrate a significant genetic distinction between Atlantic and Pacific clusters, supporting the previous mitogenome phylogeny showing reciprocal monophyly to ocean basin (Morin et al. 2012). In contrast, Cuvier's mtDNA CR previously showed no monophyly, but significant haplotype frequency differences between basins indicating high and significant levels of differentiation (Dalebout et al. 2005; Morin et al. 2012). With additional mitogenomes and nuclear SNPs, we were able to extend these results for Cuvier's and show ocean-specific clades, greatly refining our understanding of genetic structuring in this species. This is similar to what was found in northern bottlenose whales, where additional population structure was identified when nuclear genomes were added to the analysis (de Greef et al. 2022).

The estimated divergence dates of the deepest splits that separate primarily North Atlantic and Pacific clades (Cuvier's: ~1.5-2 mya Figure 3-3b) support a role of the formation of the Isthmus of Panama (O'Dea et al. 2016) in driving genetic structure in Cuvier's, as it has in other odontocete species (Amaral et al., 2012; da Silva et al., 2015; Van Cise et al., 2019). The later divergence date for Blainville's (~0.75 mya, Figure 3-4b) suggests a link to the Mid-Pleistocene Transition (MPT), which is noted for its overall cooling of sea-surface temperatures, elongation and cooling of glaciation cycles, and northward shift of Antarctic polar fronts (Clark et al. 2006; Kemp et al. 2010). The resulting cooling of tropical waters is hypothesised to have driven the speciation between *M. mirus* and *M. eueu* as well (Carroll et al. 2021). Thermal constraints linked to the smaller body size of Blainville's compared to Cuvier's may determine the former's contemporary southerly limit, which coincides with the Southern Ocean Sub-Tropical Front (STF; Graham & De Boer, 2013; Pitman & Brownell Jr., 2020). A northward shift of the STF during the MPT would have further limited migration between the Atlantic and Pacific via the southern tip of South America and potentially South Africa. Therefore, we posit that oceanographic conditions and thermal constraints may have reduced gene flow between Atlantic and Indo-Pacific Blainville's populations compared with Cuvier's. Given the well-described relationship in mammals between body size and lifespan (Blueweiss et al., 1978), we also hypothesise that genetic drift and subsequent differentiation could have occurred more quickly in Blainville's compared with Cuvier's, although we acknowledge that many factors can contribute to genetic drift (Leroy et al 2017).

The East Pacific and Mid-Atlantic Ridge Barriers are large stretches of deep and open water that fall within the oligotrophic centres of the ocean and may act as "nutritional" barriers. Beaked whales prefer productive habitats with complex bathymetry, including underwater canyons and deep slopes (MacLeod and D'Amico 2006), and no sightings records of Cuvier's nor Blainville's exist within the open areas of the East Pacific basin and the Mid-Atlantic (OBIS, 2021: although such surveys are difficult to conduct effectively; Barlow et al 2006). The Benguela and Sunda Shelf Barriers may also influence Cuvier's and Blainville's population structure. However, small sample sizes from the Southern Ocean mean further sampling will be needed to clarify the relevance of these barriers as the Southern basins are connected via the Antarctic Circumpolar Current, and the resulting productive convergence zones could promote connectivity of species across wide geographical expanses, limiting genetic differentiation (e.g. Gray's beaked whales; Thompson et al., 2016; Westbury et al., 2021).

Genetic discontinuities are well documented across the Strait of Gibraltar that separates the North Atlantic and Mediterranean Sea in many marine taxa (Patarnello et al. 2007), including a number of odontocetes (e.g. Gaspari, Airoldi, & Hoelzel, 2007; Kraft, Pérez-Álvarez, Olavarría, & Poulin, 2020). Here, we find genetic structuring not only between the North Atlantic and Mediterranean, but also within the Mediterranean, a pattern found in other odontocetes (Gaspari et al. 2007; Gkafas et al. 2017; Gaspari et al. 2019). A small amount of gene flow from individuals in the Atlantic entering the Mediterranean may be driving higher genetic diversity in the western populations or sequential founder effects driving lower diversity in the east. Further investigation requires samples from the Alborán Sea population (Cañadas and Vázquez 2014).

3.4.2 Microevolutionary drivers of genetic population structure in Cuvier's and Blainville's

Physiology, life history, behaviour, and dispersal capabilities have also likely contributed to the population structure and genetic diversity patterns of Cuvier's and Blainville's observed in our work. Cuvier's are considerably larger than Blainville's (MacLeod, 2005), make deeper dives (Baird 2019) and occupy a greater geographical range (including the circumpolar Southern Ocean; (MacLeod et al. 2006). Such factors increase the potential connectivity of Cuvier's populations in the Southern Hemisphere, though the substantial genetic structure between the North Atlantic and Indo-Pacific clusters indicates the low likelihood of individuals travelling great distances between ocean basins.

Blainville's and Cuvier's often live in small, resident populations around island groups (Baird 2019; Hooker et al. 2019), where such philopatry could enhance inter-oceanic distinctiveness, regardless of the presence of biogeographical barriers. Such site fidelity is also seen in northern bottlenose whales from the Scotian Shelf, where their preference for the submarine canyons of the "Gully" has led to genetic distinctiveness from other populations (de Greef et al. 2022). Some genetic clusters from areas of known site fidelity (Cuvier's from the Canary Islands (Reyes, 2018) and Blainville's from the Bahamas (Claridge 2006; Claridge 2013)) have slightly lower diversity compared to the remaining Atlantic clusters. In contrast, Blainville's from Hawai'i belong to either a resident or oceanic population (Baird et al. 2011), and had greater genetic diversity and lower inbreeding coefficients than their island-associated Cuvier's counterparts. Improving abundance and residency estimates in areas without dedicated photo-identification studies

will help to determine the role of site fidelity and ongoing gene flow in shaping genetic diversity in these species.

3.4.3 Proposed management units for conservation strategies

Genetic structure has now been found in several beaked whale species (see discussion above), suggesting that genetically structured populations may be a common feature in the beaked whale species that are yet to be studied. Many of the beaked whale species surveyed to date are known to occur at low abundances in small, localized populations, which given their substantial site-fidelity and population structure, may render them particularly vulnerable to human disturbances (Hooker et al., 2019).

Maintenance of genetic diversity within populations is a priority for conservation and management measures, and small populations, such as those observed in many beaked whales (Hooker et al. 2019), are more susceptible to genetic erosion (Leroy et al. 2017). Consequences of genetic erosion include inbreeding depression, accumulation of deleterious alleles, maladaptation, and reduced adaptive potential to climate change, and hence monitoring of genetic diversity is imperative for the management and conservation of wild populations (Leroy et al. 2017). In particular, multiple populations of Blainville's and Cuvier's have been subject to repeated UMEs, including in the Ligurian Sea where more than 35 Cuvier's mortalities have occurred as a result of UMEs since the 1960's, corresponding to a loss of approximately 1/3 of the current estimated population size (Podestà et al. 2006; Podestà et al. 2016). Similar incidences have impacted Cuvier's in Puerto Rico and the Virgin Islands, coinciding with US Navy activity between the 1960s and 2000 (Mignucci-Giannoni 1996; Mignucci-Giannoni and Rosario-Delestre 1998). In the eastern North Atlantic, UMEs between 2008 and 2018 involved nearly 200 deep-diving whales, around half of which were Cuvier's (Dolman et al. 2010; Brownlow et al. 2014; Brownlow et al. 2018). Given that only a small proportion of dead animals reach land during UMEs, and there are estimated to be less than 2300 Cuvier's in the eastern North Atlantic (Rogan et al. 2017), these events may have had substantial impacts on the species. The classification of taxonomic and management units below species is therefore necessary to determine how these human activities are impacting populations with respect to conservation objectives, and investigate the potential evolutionary consequences such activities may have (Taylor, Perrin, et al. 2017).

In Table 3-2, we propose both ESUs; important evolutionary components of a species which are necessary for maintaining genetic variability and successfully responding to future environmental changes (Waples 1995; Taylor et al. 2010), and DIPs; population units that depend on internal demographic dynamics rather than immigration (Waples and Gaggiotti 2006; Palsbøll et al. 2007; Taylor et al. 2010; Martien, Lang, et al. 2019).

Nearly reciprocal monophyly to ocean basin in both the mitogenomic and ddRAD phylogenies and minimal amounts of admixture between ocean basins in the cluster analysis provide clear and strong evidence that populations are on separate evolutionary trajectories, consistent with distinct ESUs. To quantify and further demonstrate the distinctiveness of ocean-basin-level populations, we used mean F_{ST} and d_A as evidence of population isolation (Taylor, Archer, et al., 2017; Waples et al., 2018).

For both species, nuclear SNP based F_{ST} (Cuvier's F_{ST} =0.018-0.197, Blainville's F_{ST}=0.119) values fall within the range of F_{ST} values calculated from nuclear SNPs in recognized odontocete subspecies (spinner dolphins: F_{ST}=0.0045-0.012; pantropical

spotted dolphins: F_{ST}=0.055; harbour porpoise (*Phocoena phocoena*): F_{ST}=0.187−0.207; and short-finned pilot whale: F_{ST} =0.1−0.4; Lah et al., 2016; Leslie & Morin, 2018; Van Cise et al., 2019). The F_{ST} of ocean-basin level populations using full mitogenomes (Cuvier's F_{ST}=0.178-0.623, Blainville's F_{ST} =0.711) are much higher than mitogenomic F_{ST} values calculated in recognized odontocete subspecies (spinner dolphins: F_{ST} =0.013, spotted dolphins: F_{ST} =0.013, fin whales (*Balaenoptera physalus*): F_{ST} =0.005−0.018; Archer et al., 2019; Leslie, Archer, & Morin, 2018). The F_{ST} of the ocean-basin populations calculated using the mtDNA CR (Cuvier's F_{ST} = 0.241-0.531, Blainville's F_{ST}=0.628) is also greater than the range of F_{ST} values in recognized cetacean subspecies (F_{ST}=0.013-0.209; Rosel et al., 2017). In contrast, our estimates of d_A did not reach the previously described subspecies threshold that ranges from 0.004-0.02 (Rosel et al., 2017). The exception to this was the Indo-Pacific and Mediterranean Cuvier's comparison (Cuvier's d_A=0.0045, F_{ST} = 0.531). Beaked whales show a contrasting pattern to other cetaceans, where the CR is less diverse compared to other regions of the mitochondrial genome (Dalebout et al., 2004; Morin et al., 2012). It is possible that d_A measurements using only the CR are not as appropriate for beaked whales as other cetaceans, and additional evaluation of mitogenomic divergence across known taxonomic groups would be useful to further resolve divergence levels and taxonomy within and among beaked whale species.

The proposed ESUs show different patterns of expansion and contraction and may therefore respond differently to UMEs and other human activities (Taylor, Perrin, et al. 2017). For example, both proposed Mediterranean ESUs have low genetic diversity and historical demography indicative of population contraction, in contrast to Cuvier's in other regions. Furthermore, in both species, individuals from the Southern Hemisphere showed genetic divergence from the Northern Hemisphere and while sample sizes in the former are small, Southern Hemisphere Cuvier's and Blainville's should be managed as distinct ESUs until confirmation of genetic distinctiveness with further sampling.

The designation of DIPs stems from multiple lines of evidence (Martien et al., 2017). Based both on the statistically significant genetic differentiation shown here and the results of long-term studies on resident populations (recently reviewed by Hooker et al., 2019), we suggest the following DIPs be identified for future conservation and management initiatives: Bahamas, Canary Islands and Hawai'i in both Cuvier's and Blainville's, and separate Mediterranean Cuvier's populations in the Ligurian and Ionian seas (Table 3-2). We acknowledge the limitations of our ability to identify DIPs given the available dataset, as increasing the number and distribution of samples will likely increase the number of DIPs identified. However, Martien et al., (2019) advises, that for rare or elusive species such as beaked whales, the sample size used to infer management units should not only be assessed with respect to statistical power and representativeness, but the length of time it could take to increase the sample size. Given that the samples in the current study span 140 years, we recommend application of the precautionary principle by incorporating the management units suggested here into conservation and management activities in light of the anthropogenic pressures that these species face (Moore and Barlow 2013; Parsons 2016).

Table 3-2 Sampling locations, sample size and *tess3r* population ID across and within ocean basins for Cuvier's (Ziphius cavirostris, *n*=123 individuals, *n*=30479 SNPs) and Blainville's beaked whales (Mesoplodon densirostris, *n*=43 individuals, *n*=13988 SNPs). Proposed ESUs formed distinct genetic clusters in *tess3r* that corresponded to distinct geographic regions and significant ddRAD F_{ST}>0.01. Proposed DIPs were based on resident populations investigated by long-term photo-identification and telemetry studies and differentiated by significant F_{ST}. Abbreviations include Med for Mediterranean and S. Hem for Southern Hemisphere, and cardinal compass points abbreviated to their first letter (e.g., west is W).

Species	Ocean Basin	Population ID	*n*	Sampling Region	Sampling Locality	Proposed ESUs	Proposed DIPs
Cuvier's	Atlantic	Atl_Carib	15	Caribbean	Bahamas, East USA, Puerto Rico, Virgin Islands	NW Atlantic	
		Atl_CanIs	17	Canary Islands	Canary Islands	NE Atlantic	Canary Islands
		Atl_NE	15	NE. Atlantic	Madeira, France, Ireland, Scotland		
		Atl_Sp	2	Spain	Spain		
		Atl_France	5	France	France		
	Indo-Pacific	Indo_Cent	5	Central Pacific	Hawaii, Johnston Atoll	Central Pacific	Hawai'i
		Indo_Sou	9	S. Indo-Pacific	Australia, New Zealand, South Africa	S Hem	
		Indo_NE	19	NE Indo-Pacific	Mexico, West Canada, West USA	NE Pacific	
		Indo_Mix	3	S. Indo-Pacific	Chile, Samoa, Philippines		
	Med	Med_West	19	W. Mediterranean	Corsica, Italy-Ligurian Sea	W Med	W Med
		Med_East	14	E. Mediterranean	Croatia, Italy-Ionian Sea, Greece, Israel	E Med	E Med
Blainville's	Atlantic	Atl_Bah	7	Bahamas	Bahamas	NW Atlantic	Bahamas
		Atl_East	16	E. Atlantic	Madeira, Canary Islands	NE Atlantic	Canary Islands
		Atl_Oth	5	Admixed Atlantic	East USA, Puerto Rico, Canada, UK		
	Indo-Pacific	Indo_Haw	6	Hawaii	Hawaii	Central Pacific	Hawai'i
		Indo_Afr	5	S. Africa	South Africa	S Hem	
		Indo_Sou	3	S. Pacific	French Polynesia, Chile		

4 Demography in the deep: Reconstructed demographic histories of North Atlantic beaked whales

4.1 INTRODUCTION

The North Atlantic is a dynamic ocean basin with complex circulation and variable weather patterns that are inherently linked to both the global climate and the trajectory of climate change (Buckley and Marshall 2016). On centennial scales, the Atlantic Meridional Overturning Circulation (AMOC) transfers cold and salty water from the surface of the subpolar North Atlantic to the deep sea (>2000m) via convection, and together with the Antarctic bottom water (AABW) generated in the Southern Hemisphere, helps to fuel the conveyor belt of water, heat and carbon transport around the globe (Buckley and Marshall 2016). On decadal scales, the oceanic conditions of the North Atlantic Ocean are impacted by physical features such as gyres and currents, continental run-off, and cyclic climate oscillations (Rossby 1996; Atkinson and Grosch 1999; Reid et al. 2003; Hurrell and Deser 2010). On a spatial scale, landscapes within this basin such as undersea mountain ranges, oceanic islands, continental shelf slopes, and shallow seas interact with currents to upwell nutrients from the deep sea, and form a number of different ecosystems (Reid et al. 2003).

Wave-driven vertical mixing increases the nutrient content in the upper layer of North Atlantic and pairing this with a seasonal abundance of light, makes it one of the most biologically productive basins in the world (Mueter et al. 2009; Osman et al. 2019). Though productivity can vary spatially and temporally, it is more stable around certain topographic features, such as the landscapes mentioned above, encouraging upwelling and productivity (Mueter et al. 2009), and aggregating higher trophic level species such as squid, fish, sea birds and marine mammals (Reid et al. 2003). The highly productive North Atlantic supports the presence of 51 marine mammal species across all trophic levels (*n*=8 mysticetes, *n*=32 odontocetes, *n*=8 pinnipeds, *n*=2 sirenians and the polar bear; Bowen, 1997; Waring *et al.*, 2006).

Climate variability can greatly influence the lower trophic levels that drive productivity and support the diversity of marine megafauna in the North Atlantic. Interannual changes in sea surface temperature (SST) and variable factors like wind, affect developmental rates of both phyto- and zooplankton and power the turbulence and mixing of plankton and nutrients in the water column (Drinkwater et al. 2003). The most significant source of climate variability in the North Atlantic is the North Atlantic Oscillation (NAO), which directly impacts a number of variables (SST, wind direction and intensity, water column mixing, precipitation, etc.) that can affect productivity across trophic levels (Drinkwater et al. 2003; Sundby et al. 2016). For example, the NAO index and endangered North Atlantic right whale (*Eubalaena glacialis*) reproduction are linked through changes in copepod (*Calanus finmarchicus)* abundance in the western North Atlantic (Meyer-Gutbrod et al. 2015); a large drop in NAO index in 1996 caused copepod populations to crash, resulting in reduced calf production of North Atlantic right whales. Warmer SST associated with positive NAO phases in the eastern North Atlantic is linked to earlier squid migrations (*Loligo forbesi,* Sims *et al.*, 2001), which may result in a mismatch of abundant prey and the predators who rely on them.

It is unequivocally understood that climate change is impacting the marine environment by increasing temperature and sea level, reducing sea-ice cover and therefore salinity through increased freshwater run-off, and ultimately changing patterns of circulation and overall climate (Simmonds and Eliott 2009). In the North Atlantic specifically, the current and projected trend is a pattern of warming around the periphery of the basin, with cooling

in the centre (Caesar et al. 2018; Pershing and Stamieszkin 2020). As predators in the marine system, cetaceans have been proposed as sentinels for monitoring climate change and ecosystem health (Hazen et al. 2019; Williamson et al. 2021). Cetaceans are responding to global climate change through a number of mechanisms (reviewed in van Weelden *et al.*, 2021) including range shifts (Storrie et al. 2018; Stafford 2019; van den Berg et al. 2021), abundance changes (Víkingsson et al. 2015), changes in timing/location of migrations (Rugh et al. 2001; Ramp et al. 2015; Hauser et al. 2017), and changes in community and trophic structure in relation to changes in prey abundance (Simmonds and Eliott 2009; Fleming et al. 2016). It is suggested that cold water species and those with discontinuities, or those that live within a narrow temperature range or near the limit of their range, will see range contractions and potentially local extirpation or extinctions (MacLeod 2009; van Weelden et al. 2021).

One way to predict how populations may respond to future climate change is to look at how they responded to historic change. Earth's climate has fluctuated between periods of glaciation and warm interglacials for the past 2.5 mya, with the most recent glacial maximum (LGM) occurring between 26.5-19 kya (Clark *et al.*, 2009). Within the North Atlantic, shifts between these periods resulted in wide swings in temperature, ice cover and sea level, which are thought to have influenced population structure and demographic history of a number of marine megafauna. For example, during periods of glaciation and lower sea levels, biogeographic barriers are proposed to have restricted gene flow between smooth hammerhead shark (*Sphyrna zygaena)* populations, with demographic expansions occurring when the glacial period terminated (Lopes da Silva Ferrette et al. 2021). Loggerhead sea turtles (*Caretta caretta*) underwent a substantial contraction in population size due to changing climate during the late Pliocene, with populations likely maintained in tropical refugia during periods of extended cooling (Reece et al. 2005). Following the LGM, populations of baleen whales and their prey in the North Atlantic experienced a mix of responses to global warming, due to the widely variable oceanographic and climatic features of the North Atlantic impacting environmental conditions and prey resources (Cabrera et al. 2022).

One of the most specious and least understood cetacean families is the beaked whales (Ziphiidae; Hooker *et al.*, 2019). Six ziphiid species are found in the North Atlantic, three of which have an "Unfavourable" predicted conservation implication with regards to global climate change (MacLeod 2009). This indicates that the geographic area that a species inhabits will likely decrease as SST increases (MacLeod 2009). In the case of higher latitude species, this contraction in their range will occur as species shift towards the poles. Here, I investigate the responses of beaked whale populations in the North Atlantic to past climate change using genomic methods, to better understand how populations may fare in the face of predicted global climate change. The demographic histories of top predators like ziphiids were likely impacted by fluctuations of lower trophic level species due to historic climate oscillations which impacted the environment and resulting primary productivity (Cabrera et al. 2022).

Reconstructions of demographic history using genetic methods have been used to estimate the effect of glacial/interglacial conditions on past populations and have found that many populations of Northern Hemisphere species declined during the LGM (Moura *et al.*, 2014; Skovrind *et al.*, 2021). This, combined with the rapid population expansions that were measured in other populations following the termination of the LGM (Jenkins et al. 2018; Feyrer et al. 2019; Cabrera et al. 2022), suggest that periods of ice cover were

disadvantageous. It has also been suggested that periods of rapid environmental change or extended periods of cold weather are proposed to be the most likely time for population declines and extinction events in cetaceans (Moura et al. 2014).

Actual sea ice cover and resulting sea level decline during glaciations is unlikely to have impacted the deepest-diving cetaceans of family Ziphiidae (ziphiids) directly, as they prefer the deep waters offshore (Moulins et al. 2007; Rogan et al. 2017; Hooker et al. 2019). It is hypothesised that deep-sea cetaceans may be able to readily adapt to climate-driven changes in the ocean by changing their often-wide-ranging distributions (Whitehead et al. 2008) and MacLeod (2009) does state that the "Unfavourable" conservation implication for Northern Hemisphere beaked whales could change to "Favourable" if they are able to colonise the Arctic Ocean. However, many ziphiids form distinct genetic clusters around specific island archipelagos and seascape features in the absence of physical barriers (Feyrer *et al.*, 2019; Hooker *et al.*, 2019, Chapter 3). Such philopatry and habitat preferences indicates that in the face of climate change, species change in spatiotemporal distributions may be constrained, reducing the availability of suitable habitat.

Here I investigate four ziphiids in the North Atlantic: Northern bottlenose whale (*Hyperoodon ampullatus,* "N. bottlenose"), Sowerby's beaked whale (*Mesoplodon bidens,* "Sowerby's"), Blainville's beaked whale (*M. densirostris,* "Blainville's") and Cuvier's beaked whale (*Ziphius cavirostris,* "Cuvier's"). While all four are found in the North Atlantic, two have cosmopolitan distributions (Cuvier's and Blainville's) and two are found exclusively in this basin (N. bottlenose and Sowerby's; Figure 4-1; MacLeod *et al.*, 2006). During periods of extended sea ice cover, available deep sea and continental slope habitat would have decreased across the North Atlantic, and it is possible that Cuvier's and Blainville's could have shifted their ranges to more equatorial areas of their distributions. With exclusively North Atlantic contemporary distributions, it is unclear how populations of Sowerby's and N. bottlenose would have responded or fared during glaciations. Possible responses include range shifts southwards, the separation of populations into refugia, or local extirpation and recolonisation.

Changes in effective population (N_e) size are estimated over the past million years for individual populations using methods based on coalescent theory, whereby past population dynamics can be inferred from modern day DNA sequences (Drummond et al. 2005; Liu and Fu 2015; Liu and Fu 2020). A number of software packages are available to estimate past population changes, and here I use "Stairway Plot 2" as it does not require the specification of an underlying demographic scenario, is well suited for reduced representation sequencing data such as ddRAD, and does not require any knowledge of ancestral alleles (Liu and Fu 2020).

Finally, I consider how contemporary structure and historical demographic responses may influence the response of these species in the face of future climate change. Based on their contemporary ranges and the predicted conservation implications from MacLeod (2009), I hypothesize that greater population declines will have occurred in the two species limited to the North Atlantic (Sowerby's and N. bottlenose) in response to glaciation and that populations in higher latitudes will see a greater decline in N_e during periods of glaciation.

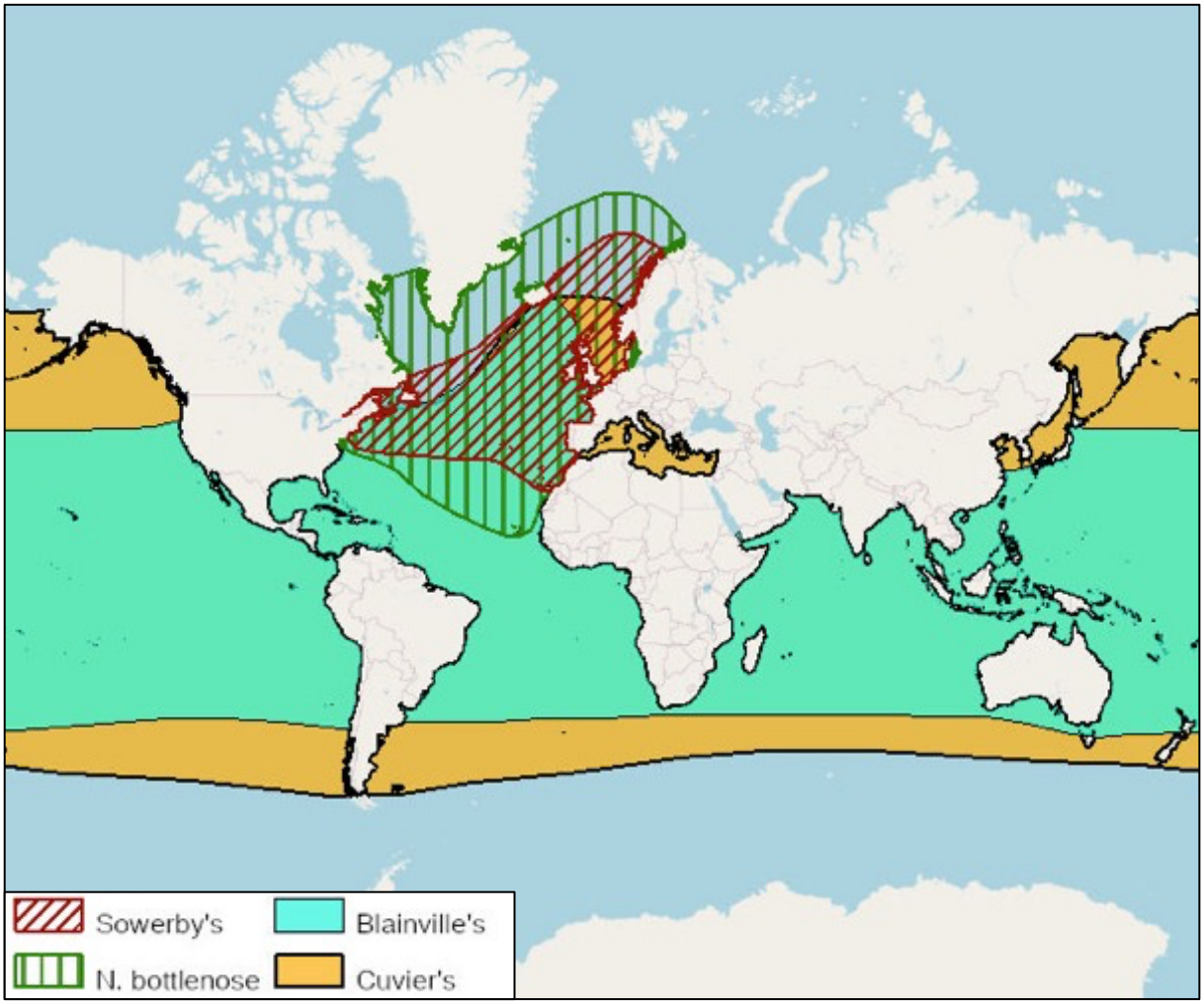


Figure 4-1. Map of the inferred distributions of each species based on shapefiles provided by the IUCN Red List: Cuvier's beaked whale ("Cuvier's"; IUCN (International Union for Conservation of Nature), 2012), Blainville's beaked whale ("Blainville's"; IUCN (International Union for Conservation of Nature), 2012b), northern bottlenose whale ("N. bottlenose"; IUCN (International Union for Conservation of Nature), 2008), and Sowerby's beaked whale ("Sowerby's"; IUCN (International Union for Conservation of Nature), 2008b). Note that the Cuvier's range encompasses all the Blainville's range.

4.2 METHODS:

4.2.1 Bioinformatics

Double-digest restriction-site associated DNA (ddRAD) sequencing was used to generate data for Cuvier's, Blainville's and Sowerby's. Sample selection, DNA extraction and ddRAD library preparation followed the protocols in Chapters 2 and 3 using samples selected from the International Tissue Archive of Beaked Whales (ITABW; 2.3.3). GBS data for N. bottlenose were provided by the University of Copenhagen as described in Chapter 2.

Following a preliminary analysis to test the suitability of the bioinformatic pipeline for the GBS data provided for N. bottlenose (Chapter 2), the Stacks (v. 2; Rochette et al. 2019) pipeline and filtering steps were run on samples from all four species to optimise the best species-specific 'gstacks' parameters (Chapter 2). The remaining bioinformatic

pipeline was conducted, and two final datasets were generated for each species to 1) characterise the genetic structure and diversity and 2) estimate demographic history. The datasets for population structure were generated using the pipeline described in Chapters 2 and 3. For demographic histories, recent mutations may be at a very low frequency and can therefore be lost when filtering by minor allele frequency (MAF). Therefore, all steps involving MAF filters were removed for the dataset used for demographic reconstructions. It is also necessary to include invariant sites when investigating demographic history, as using mutation rate to calibrate results requires all the sites characterised by ddRAD sequencing, not just the variable ones. As Stacks only outputs the variable sites into the VCF file, an R script was used to modify it to include the monomorphic sites before proceeding with the demographic analyses (https://github.com/laninsky/bats_and_rats/tree/master/fastsimcoal2_inputs).

4.2.2 Population Structure and Genetic Diversity

Demographic analyses assume that a population is panmictic, so *tess3r* (Caye et al. 2016; Caye et al. 2018) was used to determine how many genetic clusters were present within each species. *tess3r* was run with K=2-10 genetic clusters for each of the four species, using the datasets generated for population structure (with MAF filters). Significance of the genetic clustering determined by *tess3r* was tested by calculating the pair-wise F_{ST} and 95% confidence intervals between clusters in each species (*dartr*; Gruber *et al.*, 2018). A discriminant analysis of principle components (DAPC) was conducted using the genetic clusters identified from *tess3r* as *a priori* populations in the R package *adegenet* (Jombart 2008; Jombart et al. 2010) to further investigate genetic structure within these four species in the North Atlantic. The cross-validation tool "xvalDapc"' was used to optimise the number of principle components (PCs) to retain from the analysis, using the genetic clusters determined from *tess3r* as priors. The resulting DAPCs were plotted to observe the spatial structure of SNP genotypes across each species and the "assignplot" function was used to determine how well the assignment to DAPC clusters corresponded with the *a priori* population assignments.

Summary statistics of genetic diversity were calculated for each population using the R package *hierfstat* (Goudet 2005). Mean observed heterozygosity (H_o) and gene diversity (H_s) were calculated within populations according to Nei (1987) and the inbreeding coefficient (F_{IS}) was calculated according to Weir and Cockerham (1984). The 95% confidence intervals were calculated around H_o and H_s using *rcompanion* (Mangiafico 2020) and the F_{IS} 95% confidence intervals were calculated using *hierfstat* ("boot.ppfis"; Goudet, 2005).

4.2.3 Reconstructing Demographic History

Once the underlying population structure was investigated and the appropriate genetic clusters were identified, site frequency spectra (SFS) files were generated for each population within species. Folded SFS files (ancestral alleles of SNPs are unknown) were generated from the datasets without MAF filters (and with the monomorphic sites added) using the EasySFS python script (https://github.com/isaacovercast/easySFS), down projecting to the population size that resulted in the greatest number of segregating sites per population.

Demographic histories were reconstructed for each population with Stairway Plot 2 (Liu and Fu 2020) using the population specific folded SFS files, and species-specific mutation rates (sites/generation) and generation times (in years, Table 4-1). Stairway Plot 2 is a

nonparametric method of inferring demographic histories using folded SFSs. As SFSs can summarise the genotypes of the thousands of samples and SNPs produced with genotyping-by-sequencing technologies, it is particularly suitable for studies of non-model organisms. Species-specific mutation rates were calculated based on the estimated mean rate of nucleotide evolution for nuclear DNA (UCLN model) at 9.10e-4 substitutions/site/million years (Dornburg et al. 2012). The equation for species-specific rates based on Dornburg *et al.* (2012) is as follows:

$$\frac{No.\,substitutions\ per\ nucleotide\ per\ million\ years * generation\ time}{1{,}000{,}000\ years}$$

To run Stairway Plot 2, "blueprint" files were generated for each genetic cluster within species using the folded SFS, species-specific mutation rate and generation times. Generation times were summarised from the literature and these with the mutation rates are found in Table 4-1. N_e was estimated based on 200 bootstraps and the final output provided the median N_e, 75% and 95% confidence intervals.

An additional metric of demographic history, Tajima's D, was calculated for each population using a 100 kb sliding window in the program VCFtools (Danecek et al. 2011). The mean value and 95% confidence intervals were calculated using the R package *rcompanion* (Mangiafico 2020).

Table 4-1. Species-specific mutation rates (substitutions/nucleotide/generation) based on the calculation in Dornburg *et al.* (2012), the mean rate of nucleotide evolution for nuclear DNA (9.10e-4 substitutions/site/million years, Dornberg *et al.* 2012), and generation time (years) from the literature.

Species	Generation time	Mutation rate	Citation
Cuvier's	15	1.365e-8	(ACCOBAMS; Dalebout et al. 2005)
Blainville's	10	9.1e-9	(Mead 1984; Aguilar de Soto et al. 2015)
Sowerby's	7	6.37e-9	(NOAA Fisheries 2022)
N. bottlenose	10	9.1e-9	(Mead 1984; COSEWIC 2011)

4.3 RESULTS:

4.3.1 Final Dataset

DdRAD sequences were available from *n*=89 Cuvier's, *n*=40 Sowerby's and *n*=35 Blainville's (2.6) and GBS data were available from *n*=58 N. bottlenose (2.2.4). The sampling locations of the final individuals that passed quality control are mapped in Figure 4-2. The final datasets for each species with MAF filters (wMAF) and without MAF filters and with monomorphic loci added back in (noMAF+MM) are presented in Table 4-2. Within species-specific datasets, the amount of missing data was consistently low (mean missing SNPs per species: 1.19% - 2.65%). For the demographic reconstructions, summaries of the folded SFS files are presented in Table 4-2 with the down projected population size and the number of segregating sites.

Table 4-2. Summary of sample metadata and number of loci used in analyses presented in this Chapter, as well as diversity summary statistics for each population of Cuvier's beaked whale (ZC), Blainville's beaked whale (MD), Sowerby's beaked whale (MB) and N. bottlenose whale (HA). SNP data sets either had minor allele frequency filters for structure and diversity analyses (wMAF) or without MAF filters plus the monomorphic loci added back in (noMAF+wMM) for demographic history analyses. Cluster abbreviations: CI (Canary Islands), NE (Northeast), W (West), SP (Spain), E (East), OTH (Other), D (Denmark), SF (Scotland+France), FI (Faroe Islands), ICE (Iceland), GBI (UK+Ireland), CAN (Canada). Sample origin abbreviations: CI (Canary Islands), FR (France), MD (Madeira), SP (Spain), FL (Florida, USA), NC (North Carolina, USA), NJ (New Jersey, USA), PR (Puerto Rico), UK (United Kingdom), IR (Ireland), BH (Bahamas), DK (Denmark), CA (Canada), PT (Portugal), SW (Sweden), NO (Norway), VI (Virgin Islands), FI (Faroe Islands), IC (Iceland). Summary statistic abbreviations: H_o (observed heterozygosity), H_s (gene diversity), F_{IS} (inbreeding coefficient), SFS *n* (down projected population size used to generate site frequency spectra, SFS), seg. sites (number of segregating sites used to generate the SFS).

Species	No. Loci (wMAF)	Cluster	*n*	Sample Origin	H_o (95% CI)	H_s (95% CI)	F_{IS} (95% CI)	No. Loci (noMAF + wMM)	SFS *n* (seg. sites)	Tajima's D
Cuvier's	132795	ZC_CI	16	CI	0.162 (0.161,0.162)	0.187 (0.187,0.188)	0.145 (0.143, 0.147)	8445897	28 (130842)	-0.589 (-0.599, -0.579)
		ZC_NE	19	UK, FR, MD	0.162 (0.161,0.163)	0.188 (0.187,0.189)	0.146. (0.144, 0.148)		32 (146779)	-0.698 (-0.708, -0.689)
		ZC_SP	2	SP	0.172 (0.170,0.174)	0.085 (0.084,0.086)	NA		NA	NA
		ZC_W	14	FL, NC, BH, PR, VI	0.162 (0.161,0.162)	0.189 (0.188,0.190)	0.154 (0.152, 0.156)		24 (121861)	-0.524 (-0.534, -0.514)
Blainville's	25541	MD_E	16	CI, UK, MD	0.150 (0.148,0.153)	0.159 (0.156,0.161)	0.052 (0.046, 0.059)	5465244	28 (38486)	0.055 (0.042, 0.067)
		MD_W	9	CA, FL, NJ, PR, BH	0.143 (0.141,0.146)	0.155 (0.153,0.157)	0.077 (0.071, 0.086)		16 (31984)	0.026 (0.013, 0.039)

Species	No. Loci (wMAF)	Cluster	*n*	Sample Origin	H_o (95% CI)	H_s (95% CI)	F_{IS} (95% CI)	No. Loci (noMAF + wMM)	SFS *n* (seg. sites)	Tajima's D
Sowerby's	24961	MB_D	2	DK	0.243 (0.238,0.248)	0.121 (0.118,0.124)	NA	6628357	NA	NA
		MB_OTH	8	UK, IR, CI, PT, USA	0.233 (0.230,0.235)	0.242 (0.240,0.244)	0.054 (0.048, 0.061)		14 (36565)	-0.163 (-0.175, -0.151)
		MB_SF	17	UK, FR	0.235 (0.232,0.237)	0.243 (0.241,0.245)	0.048 (0.044, 0.053)		30 (47927)	-0.208 (-0.220, -0.196)
N. bottlenose	935	HA_CAN	3	CA	0.326 (0.303,0.348)	0.263 (0.247,0.278)	-0.241 (-0.305, -0.191)	156119	6 (501)	0.167 (0.082, 0.261)
		HA_FI	7	FI	0.329 (0.310,0.349)	0.275 (0.264,0.286)	-0.193 (-0.237, -0.144)		12 (758)	-0.146 (-0.226, -0.073)
		HA_GBI	9	UK, IR	0.323 (0.305,0.340)	0.273 (0.261,0.284)	-0.182 (-0.218, -0.135)		16 (886)	-0.417 (-0.480, -0.344)
		HA_ICE	5	IC	0.328 (0.308,0.349)	0.283 (0.270,0.296)	-0.157 (-0.207, -0.113)		10 (681)	-0.038 (-0.119, 0.050)
		HA_OTH	4	UK, FR, SW, NO	0.345 (0.323,0.368)	0.272 (0.258,0.287)	-0.265 (-0.318, -0.224)		8 (574)	0.129 (0.0419, 0.2170)

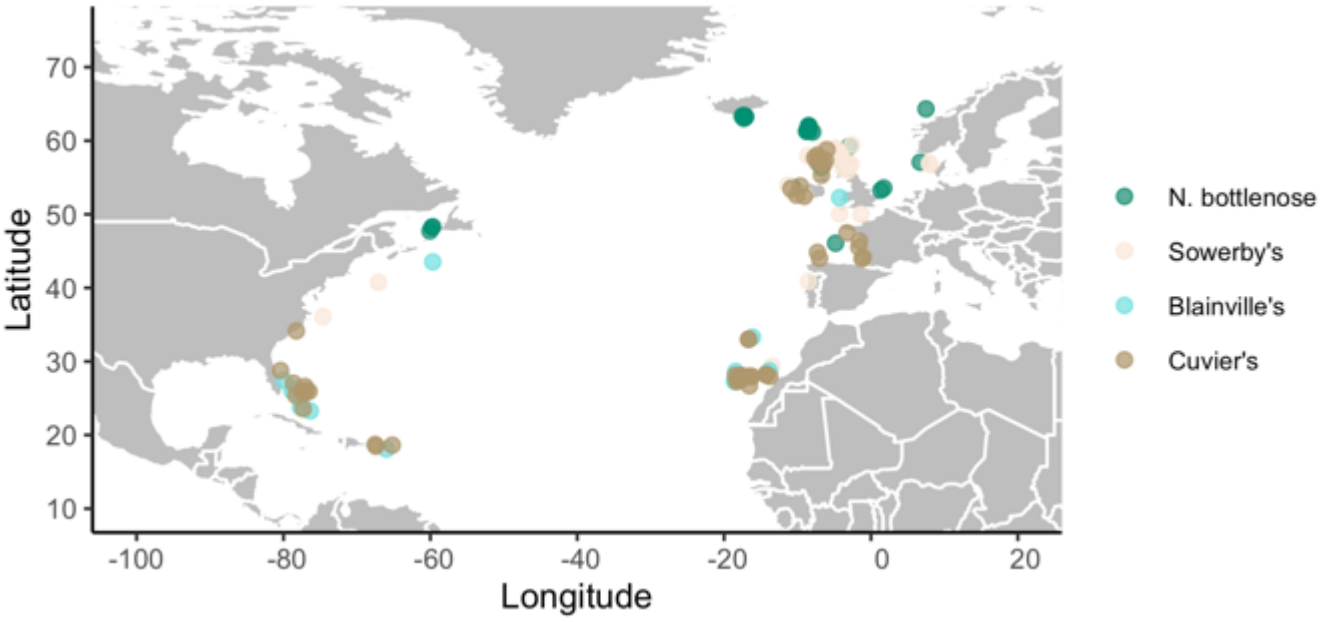


Figure 4-2. Distribution of samples from northern bottlenose whales ("N. bottlenose"), Sowerby's beaked whale ("Sowerby's"), Blainville's beaked whale ("Blainville's"), and Cuvier's beaked whale ("Cuvier's") that passed quality control and were included in the analyses.

Due to differences in the sequencing methodology, the N. bottlenose dataset resulted in less data than the Cuvier's (935 SNPs vs ~132k SNPs). With SNP counts spanning two orders of magnitude, I conducted a sensitivity study using the Cuvier's dataset to ensure that the number of SNPs retained in the N. bottlenose dataset would yield sufficient genetic structure resolution to conduct the remaining analyses and make comparisons with the other species (2.8.4).

4.3.2 Population Structure and Genetic Diversity

In the *tess3r* analysis, the most likely number of genetic clusters was selected as in Chapter 3, based on the most parsimonious clustering of individuals (Figure 4-3 to Figure 4-6). The best number of clusters was defined as K=4 for Cuvier's, K=2 for Blainville's, K=3 for Sowerby's, and K=3 for N. bottlenose (Figure 4-3 to Figure 4-6).

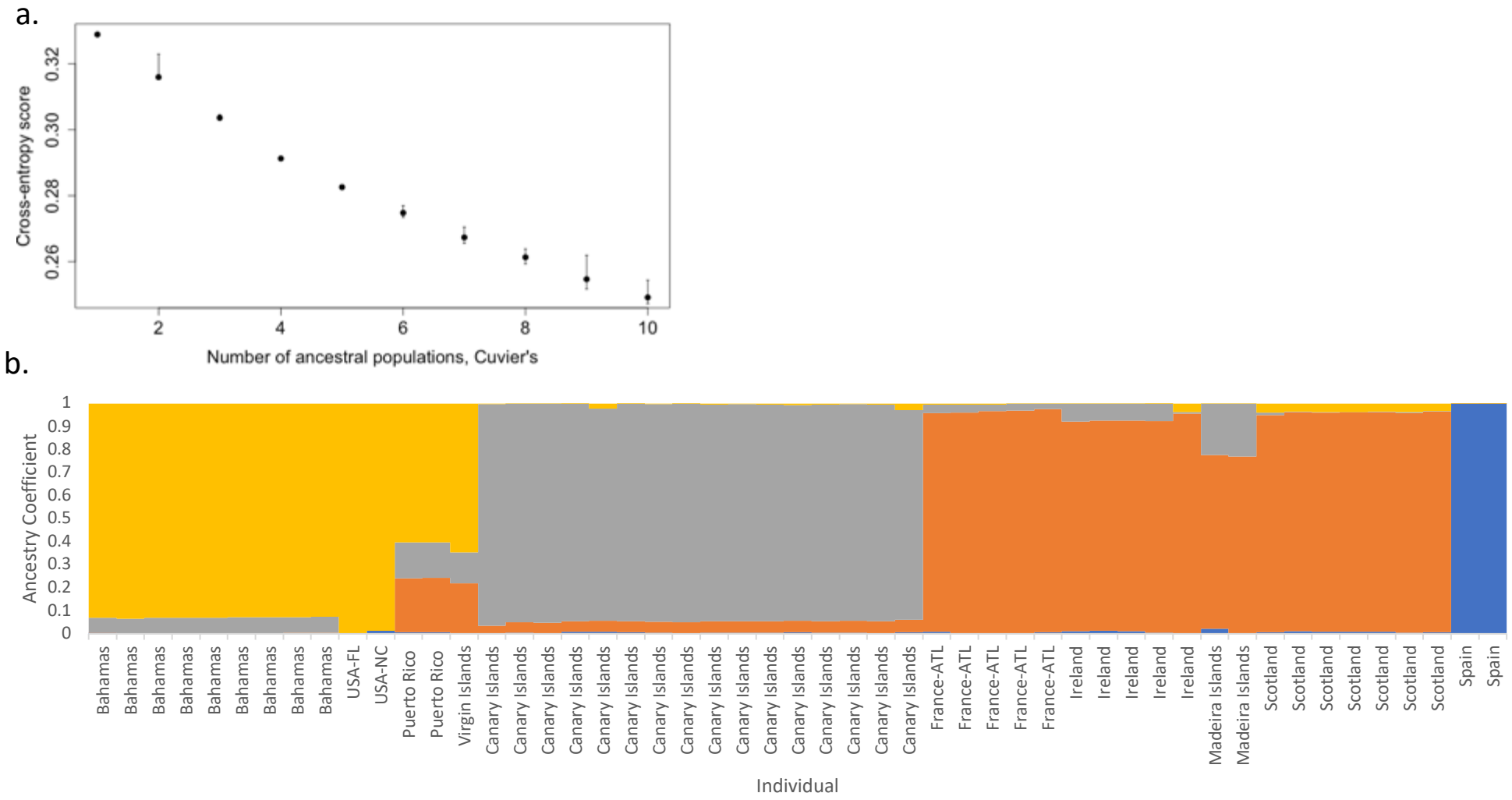


Figure 4-3. *tess3r* plots of cross entropy score (a) and coancestry coefficients (b) for Cuvier's (K=4)

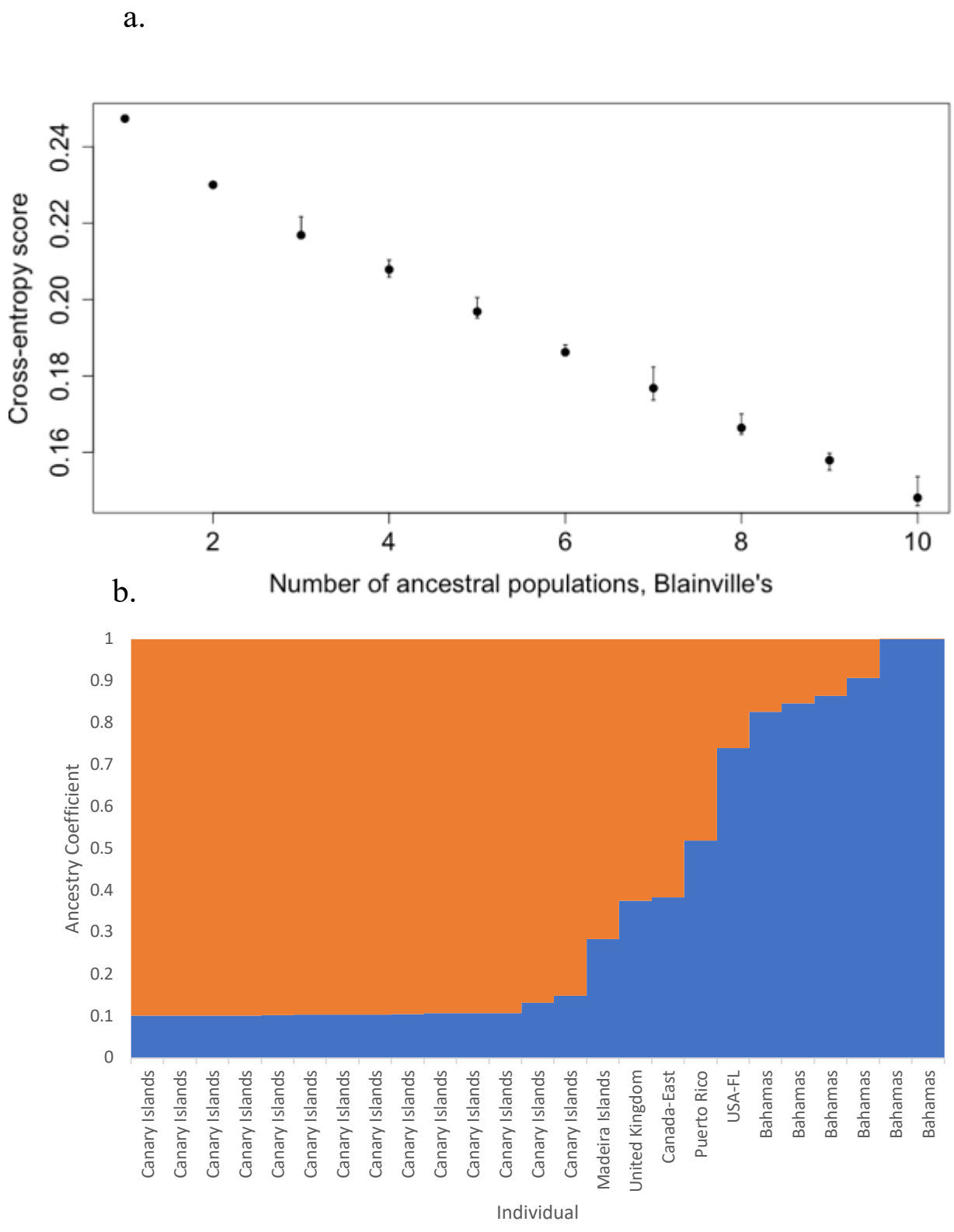


Figure 4-4. *tess3r* plots of cross entropy score (a) and coancestry coefficients (b) for Blainville's (K=2).

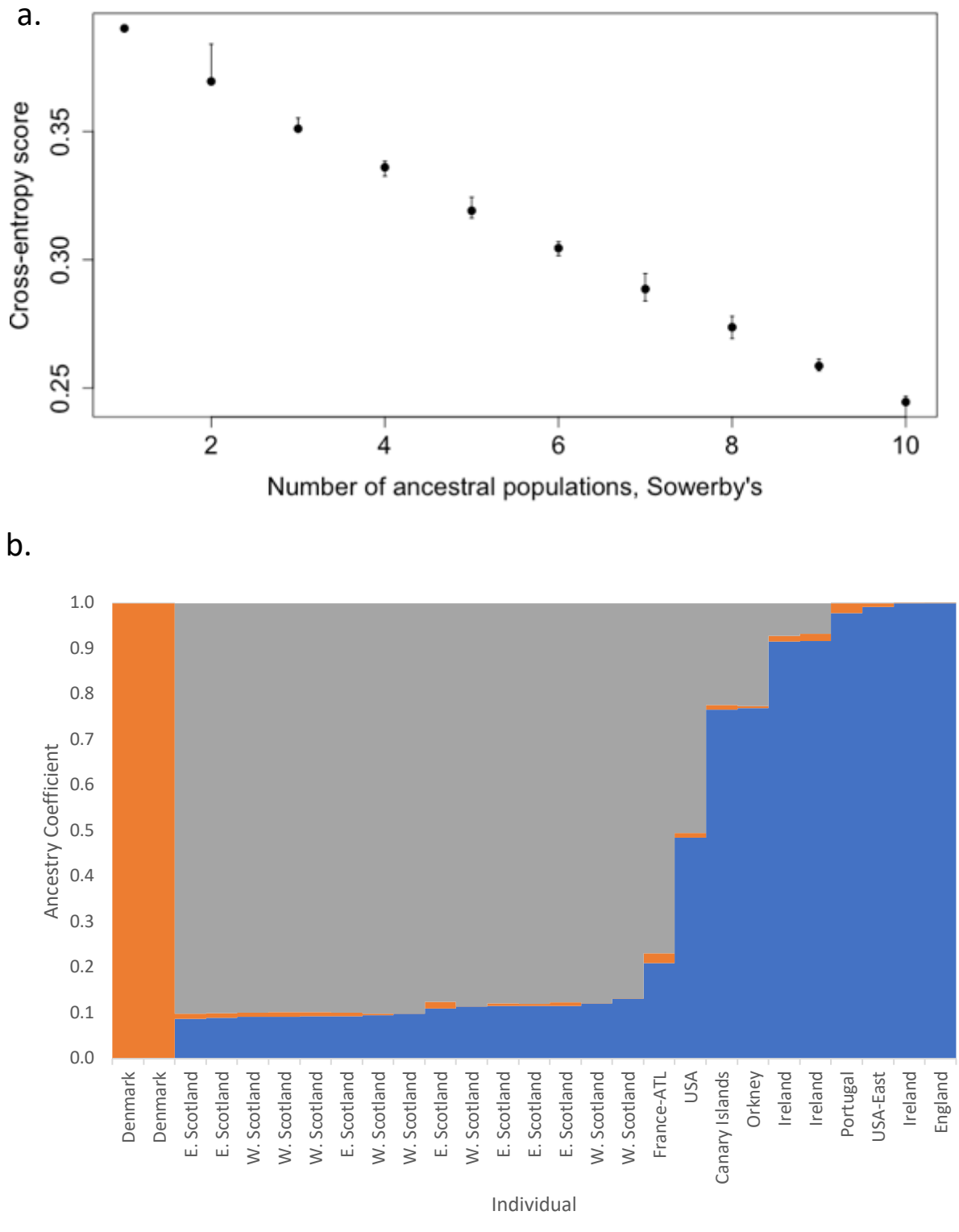


Figure 4-5. *tess3r* plots of cross entropy score (a) and coancestry coefficients (b) for Sowerby's (K=3).

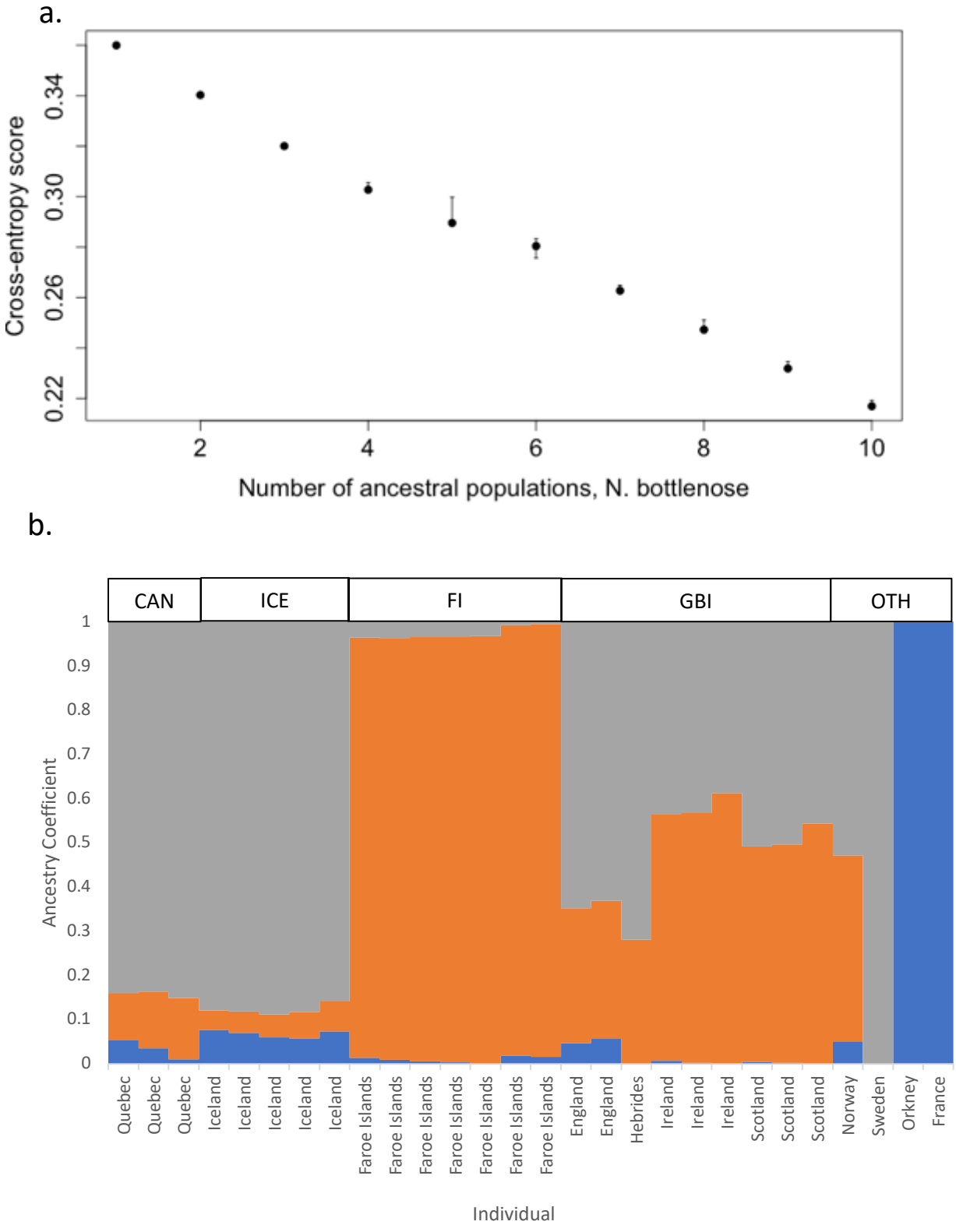


Figure 4-6. *tess3r* plots of cross entropy score (a) and coancestry coefficients (b) for N. bottlenose (K=3). The names above correspond to the final cluster names in Table 4-2 which were used for the genetic diversity and demographic history analyses.

As in Chapter 3, distinct genetic clusters corresponded to the geographic origin of the Cuvier's samples in the Caribbean, Canary Islands, Spain and eastern North Atlantic (Figure 4-3). All Blainville's clustered according to their area of origin (eastern or western North Atlantic), except for one individual from the east coast of Canada that shared more coancestry with the eastern North Atlantic individuals (Figure 4-4). Sowerby's formed three distinct clusters (Figure 4-5), one made of two Danish individuals, one made of primarily of Scottish individuals (and one from France), and the final cluster with individuals from the east coast of the USA, Canary Islands, Ireland, Portugal, and England.

N. bottlenose individuals from the Faroe Islands formed a distinct cluster as did those from Quebec and Iceland together (Figure 4-6). One individual from Sweden shared coancestry with the Quebec/Iceland group. The third cluster contained one individual from France and Orkney and the final cluster included samples from the UK, Norway and Ireland and shared ancestry with both the Faroese and Quebec/Icelandic clusters (Figure 4-6). Not all N. bottlenose samples clustered using 'tess3r' according to geographic origin (for example on individual from Sweden shared coancestry with the Quebec/Iceland group and one individual from Orkney shared it's coancestry with one individual from France instead of the other individuals from Scotland and the rest of the British Isles, Figure 4-6).

Using the genetic clusters identified in 'tess3r' as priors, the Cuvier's cluster consisting of 2 individuals from Spain was highly differentiated from the remaining populations and dominated the visible structure in Cuvier's when the 1st and 2nd principal components were plotted (Figure 4-7a). When the 2nd and 3rd principal components were plotted, the separation of the remaining three clusters were clearer (Figure 4-8a) and all individuals assigned to their geographic origin (Figure 4-7b and Figure 4-8b). The Blainville's DAPC (Figure 4-9a) showed clear division between eastern and western North Atlantic clusters with only one individual assigned to a location different from where it was sampled in the assignment plot (Figure 4-9b).

Clear differentiation is seen between the Danish and remaining Sowerby's clusters (Figure 4-10a), but little distinction is visible between the two remaining clusters. When the Danish cluster is removed, the two remaining clusters still overlap (Figure 4-11a). This unclear distinction between clusters is evident in the assignment plots, where several individuals do not assign to their population of origin regardless of whether the Danish cluster is included or not (Figure 4-10b and Figure 4-11b). In the N. bottlenose dataset, the "Other" cluster is clearly differentiated from the remaining clusters (Figure 4-12a) and when removed, the remaining 3 clusters are more clearly differentiated from each other (Figure 4-13a), and most samples assign to the cluster from their geographic origin (Figure 4-13b).

As it is not possible to know whether the unclear patterns of genetic clustering of N. bottlenose according to geographic origin is an artefact of the sample sizes or of the underlying population structure itself, the remaining analyses were conducted on N. bottlenose population assignments that factored in the geographic origin of the sample as well as the genetic clustering results from 'tess3r'. For the analysis, the

Quebec/Iceland group was split into separate clusters (HA_CAN and HA_ICE, respectively), the UK (except Orkney) and Ireland samples were clustered together (HA_GBI), and an "Other" group was used (HA_OTH), consisting of four individuals from Orkney, France, Norway, and Sweden. The final assigned genetic cluster, abbreviation, sample size, and sample origin are provided for all species in Table 4-2.

The amount of genetic differentiation (F_{ST}, p-value, and 95% confidence interval) between each genetic clusters described in Table 4-2 is presented in Table 4-3. All pairs of Cuvier's and Blainville's populations were genetically distinct (p-value < 0.05, 95% confidence intervals didn't span zero). In Sowerby's, the only significant differentiation was seen pairs involving the 2 individuals from Denmark (Table 4-3). All pairs of N. bottlenose populations that included the "Other" group showed significant differentiation, as well as the Faroe Islands and Iceland pair (Table 4-3).

Estimates of genetic diversity and heterozygote deficiency are presented for each population in Table 4-2. Within Cuvier's, H_o was significantly higher (no overlap of 95% confidence intervals) and H_s was significantly lowest in Spain, with little difference between the remaining clusters. F_{IS} was significantly highest in the West; with the sample size being too small in the Spanish cluster (n=2) to calculate. In Blainville's H_o and H_s were highest, and F_{IS} was lowest in the east. The small Danish cluster (n=2) of Sowerby's had significantly higher H_o and lower H_s and the remaining two clusters had overlapping values. The sample size was too small to calculate F_{IS} in the Sowerby's Danish cluster, and values from the other two clusters overlapped. The 95% confidence intervals of H_o and H_s overlapped across N. bottlenose populations, and all F_{IS} estimates were significantly negative.

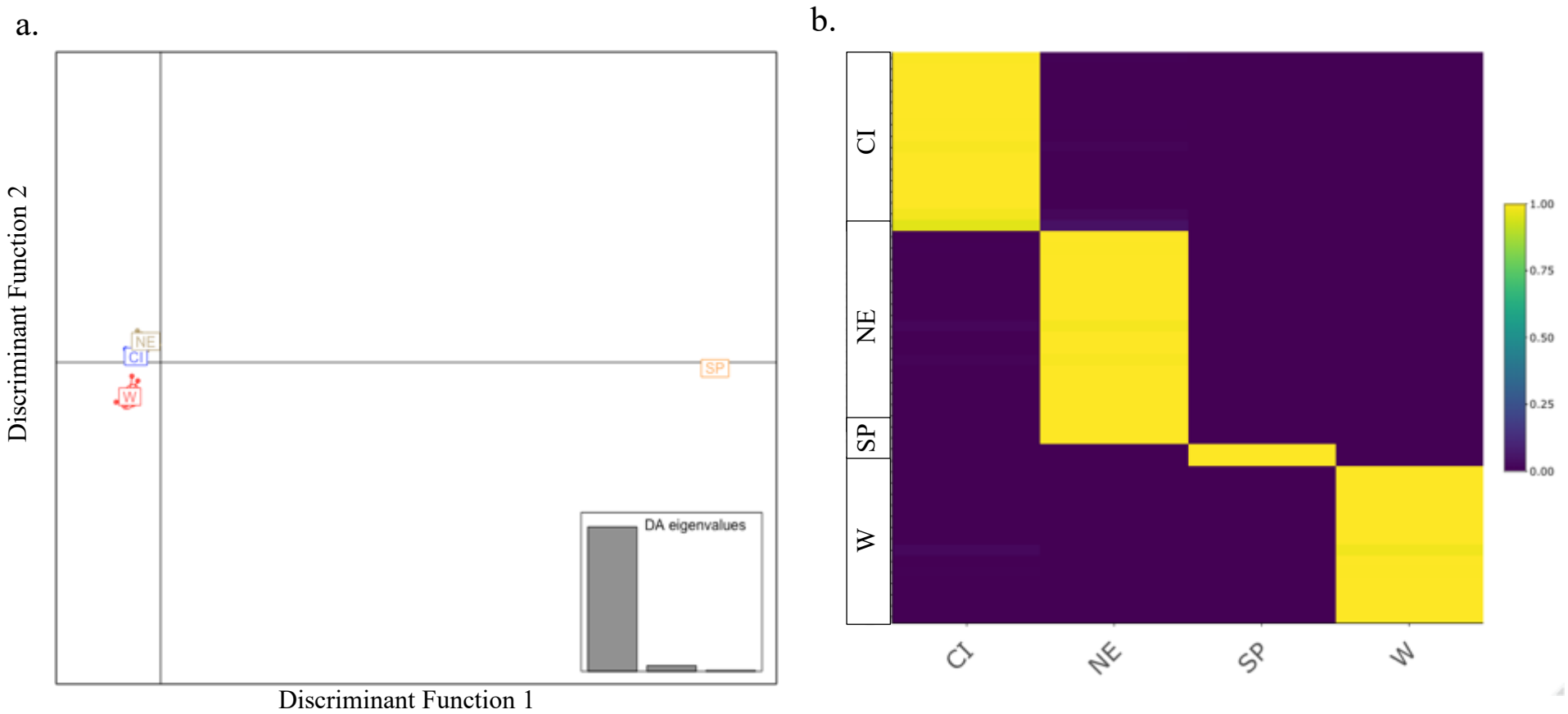


Figure 4-7. DAPC scatter of the 1st and 2nd discriminant functions (a) and assignment plot (b) for the four Cuvier's genetic clusters: Canary Islands (CI), Northeast (NE), Spanish (SP) and western North Atlantic (W). In the assignment plot, each row represents an individual and is labelled with the genetic cluster assigned *a priori*. Each column represents the DAPC cluster classification, and the shade corresponds to the membership probability.

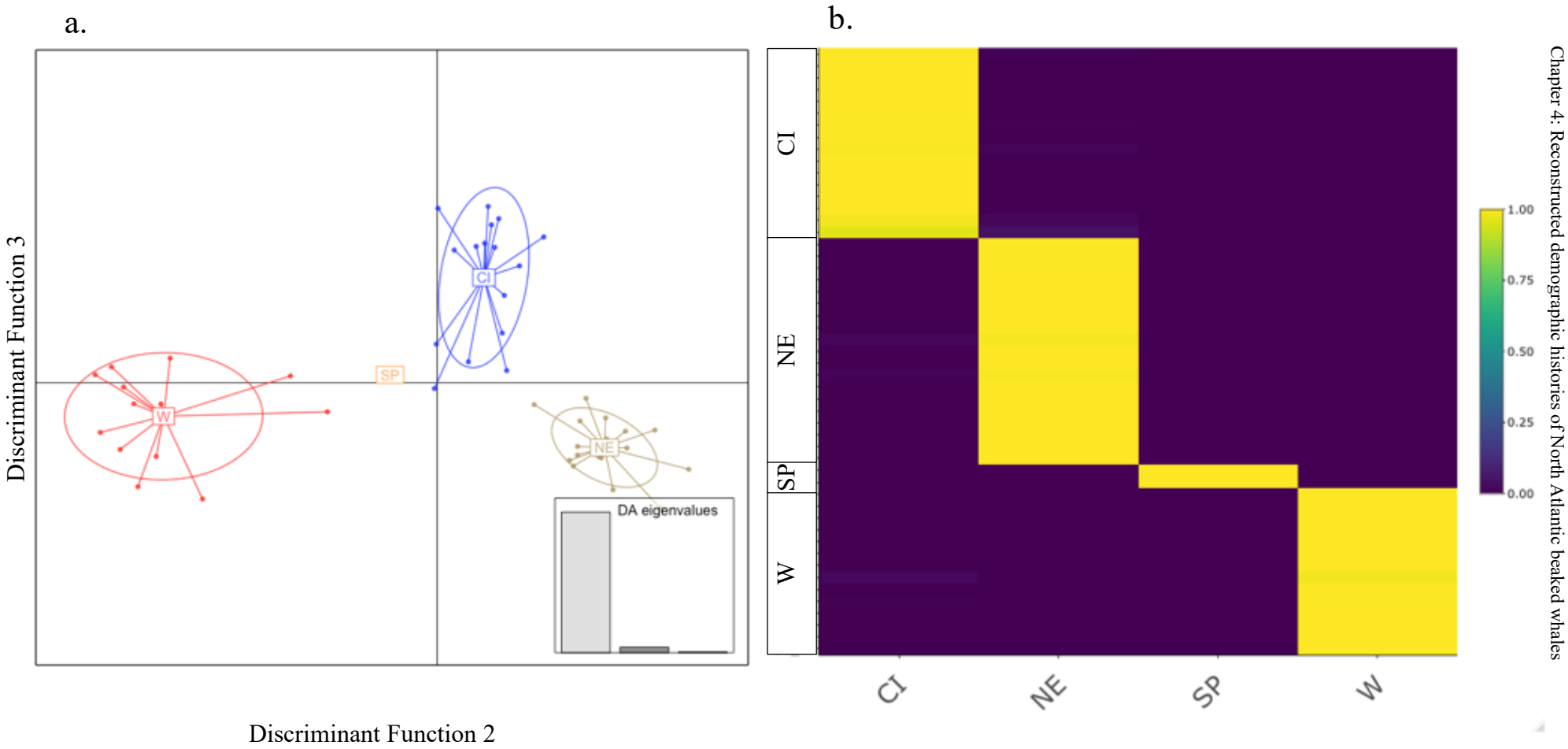


Figure 4-8. DAPC scatter of the 2nd and 3rd discriminant functions (a) and assignment plot (b) for the four Cuvier's genetic clusters: Canary Islands (CI), Northeast (NE), Spanish (SP) and western North Atlantic (W). In the assignment plot, each row represents an individual and is labelled with the genetic cluster assigned *a priori*. Each column represents the DAPC cluster classification, and the shade corresponds to the membership probability.

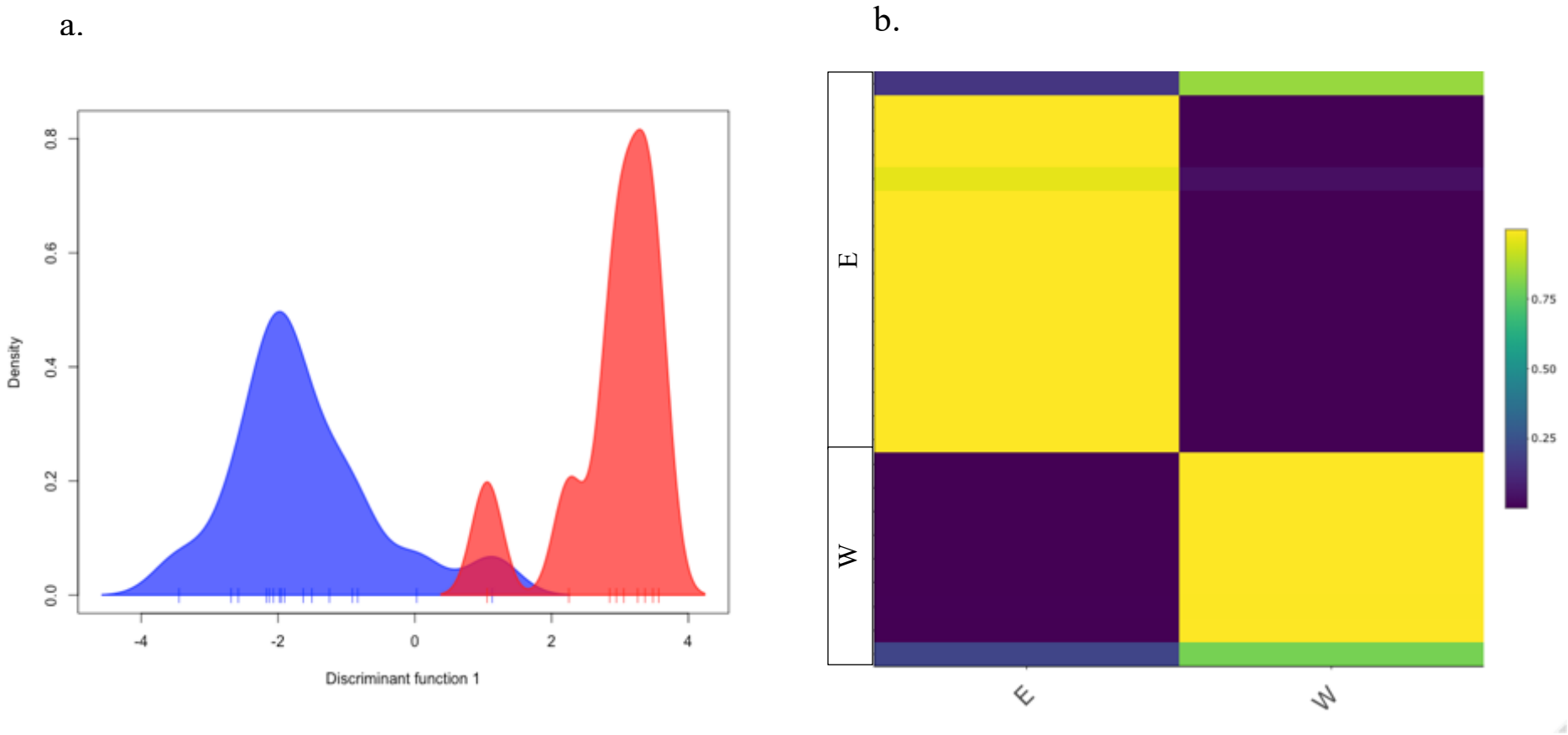


Figure 4-9. Density plot for the first discriminant function of a DAPC (a) and assignment plot (b) for the two Blainville's genetic clusters: East (E) and West (W). In the assignment plot, each row represents an individual and the label along the y-axis indicates the genetic cluster assigned *a priori*. Each column represents the DAPC cluster classification, and the shade corresponds to the membership probability.

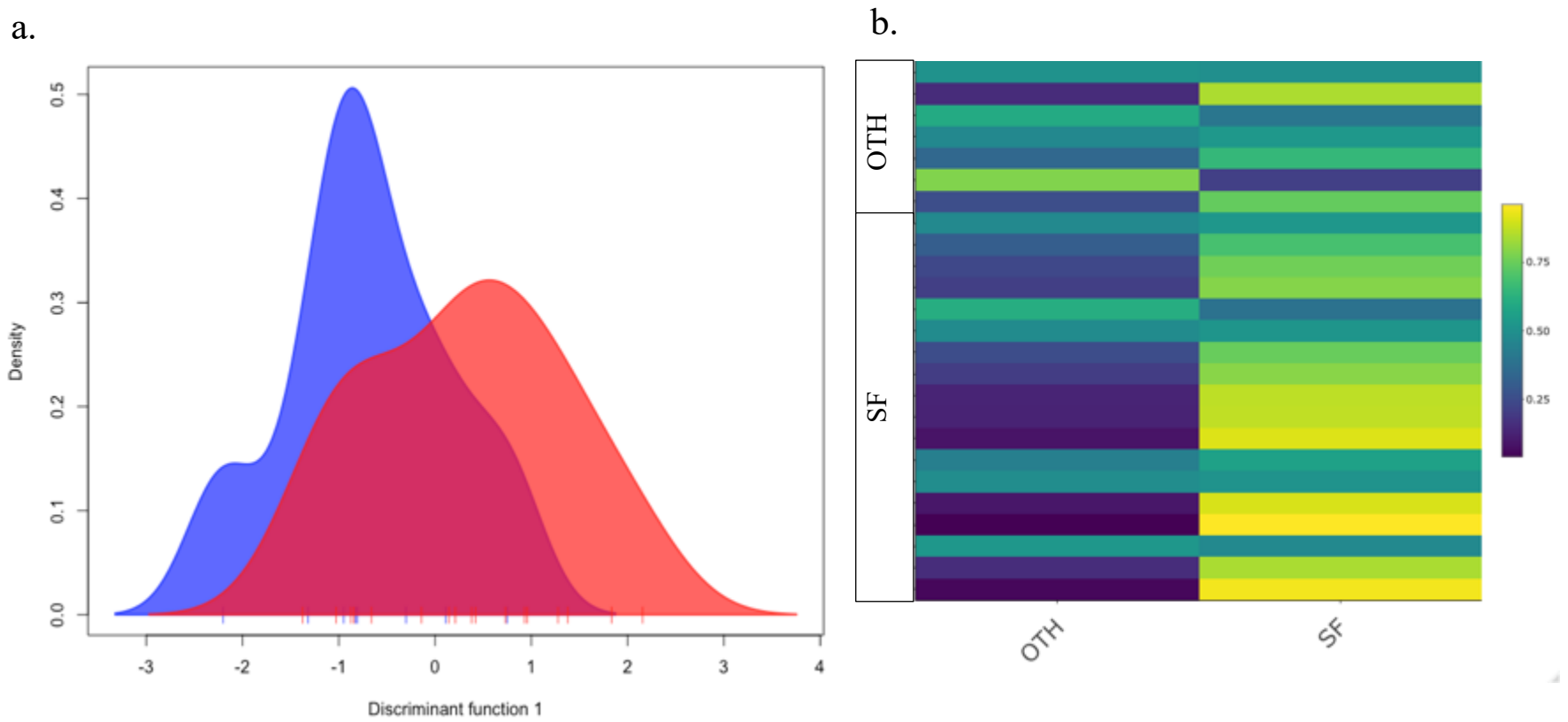


Figure 4-11. Density plot for the first discriminant function of a DAPC (a) and assignment plot (b) for two Sowerby's clusters: Scotland and France (SF) and Other (OTH). In the assignment plot, each row represents an individual and is labelled with the genetic cluster assigned *a priori*. Each column represents the DAPC cluster classification, and the shade corresponds to the membership probability.

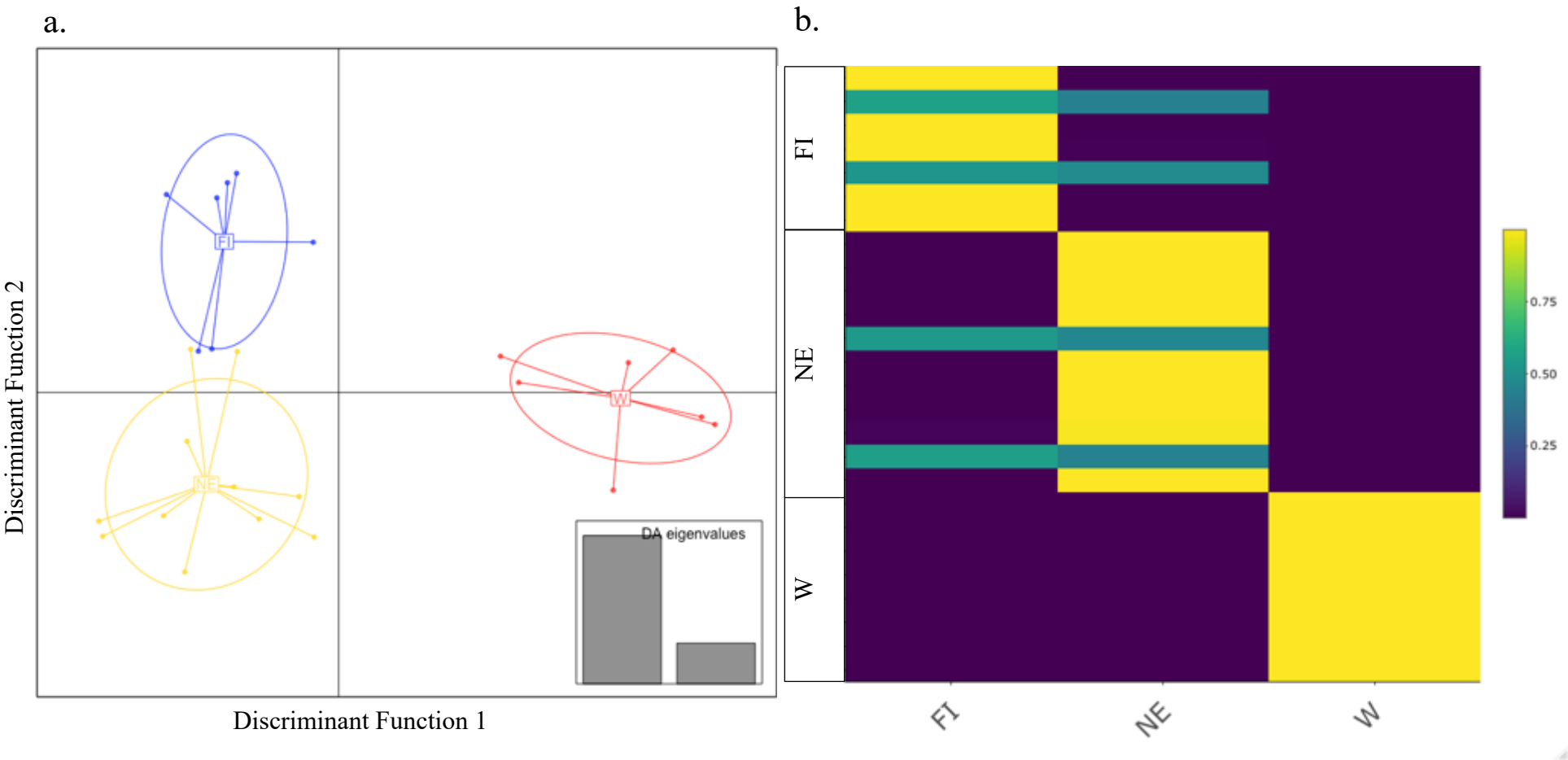


Figure 4-13. DAPC scatter of the 1st and 2nd discriminant functions (a) and assignment plot (b) for three N. bottlenose clusters: Faroe Island (FI), Northeast Atlantic (NE), and western North Atlantic (W). In the assignment plot, each row represents an individual and is labelled with the genetic cluster assigned *a priori*. Each column represents the DAPC cluster classification, and the shade corresponds to the membership probability

Table 4-3. Estimates of differentiation (F_{ST}, p-value and 95% confidence intervals) for each population pair of Cuvier's beaked whale (ZC), Blainville's beaked whale (MD), Sowerby's beaked whale (MB) and N. bottlenose whale (HA) calculated using the wMAF dataset (Table 4-2). Rows in bold italics indicate population pairs without significant (p>0.05) differentiation. Cluster definitions: CI (Canary Islands), NE (Northeast), W (West), SP (Spain), E (East), OTH (Other), D (Denmark), SF (Scotland+France), FI (Faroe Islands), ICE (Iceland), GBI (UK+Ireland), CAN (Canada).

Species	**Cluster1**	**Cluster 2**	**F_{ST}**	**Lower CI**	**Upper CI**	**p-value**
Cuvier's	ZC_CI	ZC_NE	0.007	0.007	0.008	<0.01
	ZC_CI	ZC_W	0.013	0.012	0.013	<0.01
	ZC_CI	ZC_SP	0.155	0.153	0.158	<0.01
	ZC_NE	ZC_W	0.012	0.011	0.012	<0.01
	ZC_NE	ZC_SP	0.143	0.141	0.146	<0.01
	ZC_W	ZC_SP	0.155	0.153	0.158	<0.01
Blainville's	MD_E	MD_W	0.03	0.028	0.032	<0.01
Sowerby's	MB_OTH	MB_D	0.145	0.139	0.151	<0.01
	MB_OTH	***MB_SF***	***0.001***	***0.000***	***0.002***	***0.06***
	MB_D	MB_SF	0.131	0.125	0.136	<0.01
N. bottlenose	HA_OTH	HA_FI	0.045	0.031	0.055	<0.01
	HA_OTH	HA_ICE	0.031	0.018	0.043	<0.01
	HA_OTH	HA_GBI	0.039	0.027	0.049	<0.01
	HA_OTH	HA_CAN	0.032	0.017	0.049	<0.01
	HA_FI	HA_ICE	0.008	0.001	0.015	0.02
	HA_FI	***HA_GBI***	***0.006***	***-0.012***	***-0.002***	***1***
	HA_FI	***HA_CAN***	***0.011***	***-0.002***	***0.024***	***0.06***
	HA_ICE	***HA_GBI***	***<0.001***	***-0.011***	***0.004***	***0.69***
	HA_ICE	***HA_CAN***	***<0.001***	***-0.013***	***0.009***	***0.56***
	HA_GBI	***HA_CAN***	***<0.001***	***-0.013***	***0.008***	***0.59***

4.3.3 Reconstructing Demographic History

The reconstructed demographic histories from Stairway Plot 2 are plotted for each population within species (>2 individuals, excluding "Other" clusters) in Figure 4-14a-d. In each plot, the median N_e is presented using the thick line, and the shaded area represents the 95% confidence interval. The vertical shaded bars represent major climatic events that are known to have impacted the demographic history of other marine species: Last Glacial Period (LGP; 110-15kya), Penultimate Glacial Period (PGP; 194-135 kya) and the Mid-Pleistocene Transition (MPT; 1250-700kya).

Tajima's D in Cuvier's populations was negative (Table 4-2), indicative of population expansion, contrasting with the reconstructed demographic histories (Figure 4-14a). In each population, there is a rise in N_e from the start and a gradual decline to present day. Additionally, both the eastern North Atlantic populations have bottlenecks and expansions that began ~800 kya. While all three populations began with approximately the same N_e (~25 k), the present value is substantially higher in the West population (25k, 95% CI: 1.5-11.5 k) compared to the two east populations (~5-10 k, Figure 4-14a), though the 95% confidence intervals overlap (Figure 4-14a).

The positive Tajima's D values for both Blainville's populations indicate that the populations are contracting, and the demographic history reconstructions shows that both populations go through bottlenecks (Figure 4-14b) approximately 100 kya, with the West population plateauing at 60 k (95% CI: 30-80 k) individuals 40 kya (Figure 4-14b). In the East, the population goes through another bottleneck 4 kya and the population plateaus at 15 k (95% CI: 0-50 k) individuals. The West population starts with fewer individuals compared to the East (<6 k vs 25 k) and its bottleneck drops to fewer individuals (9 k vs 30 k). However, the most recent N_e for the West is higher than the East, although the 95% confidence intervals overlap.

The analysed Sowerby's population (Scotland+France) had negative Tajima's D, indicative of population expansion (Table 4-2), also seen in the demographic history (Figure 4-14c). This population expanded until ~600 kya and reached a plateau at ~100 k individuals until modern times (95% CI: 20 k-100 k; Figure 4-14c).

Reconstructed demographic histories of the four N. bottlenose populations are found in Figure 4-14d. The two populations with negative Tajima's D (Table 4-2, therefore expanding, Faroe Islands and UK+Ireland) do show a pattern where the modern-day N_e is greater than the historic N_e, though the Faroe Islands population appears to experience a bottleneck followed by an expansion between ~400 k-50 kya. A similar, though less dramatic, bottleneck is seen in the UK+Ireland population, with an expansion following around the same time as the Faroe Islands population. The positive Tajima's D value in the Canadian population suggests a contraction (Table 4-2), which is seen in the bottleneck that takes place around 400 kya. As with the other 3 populations, the Icelandic population appears to go through a bottleneck/expansion similar in timing and magnitude, with the final N_e higher than the starting N_e. The final N_e for Faroe Islands, UK+Ireland and Iceland populations are similar in magnitude (175 k-250 k) and much higher than the final N_e for the Canadian population (~30 k) though the 95% confidence intervals overlap (Figure 4-14d).

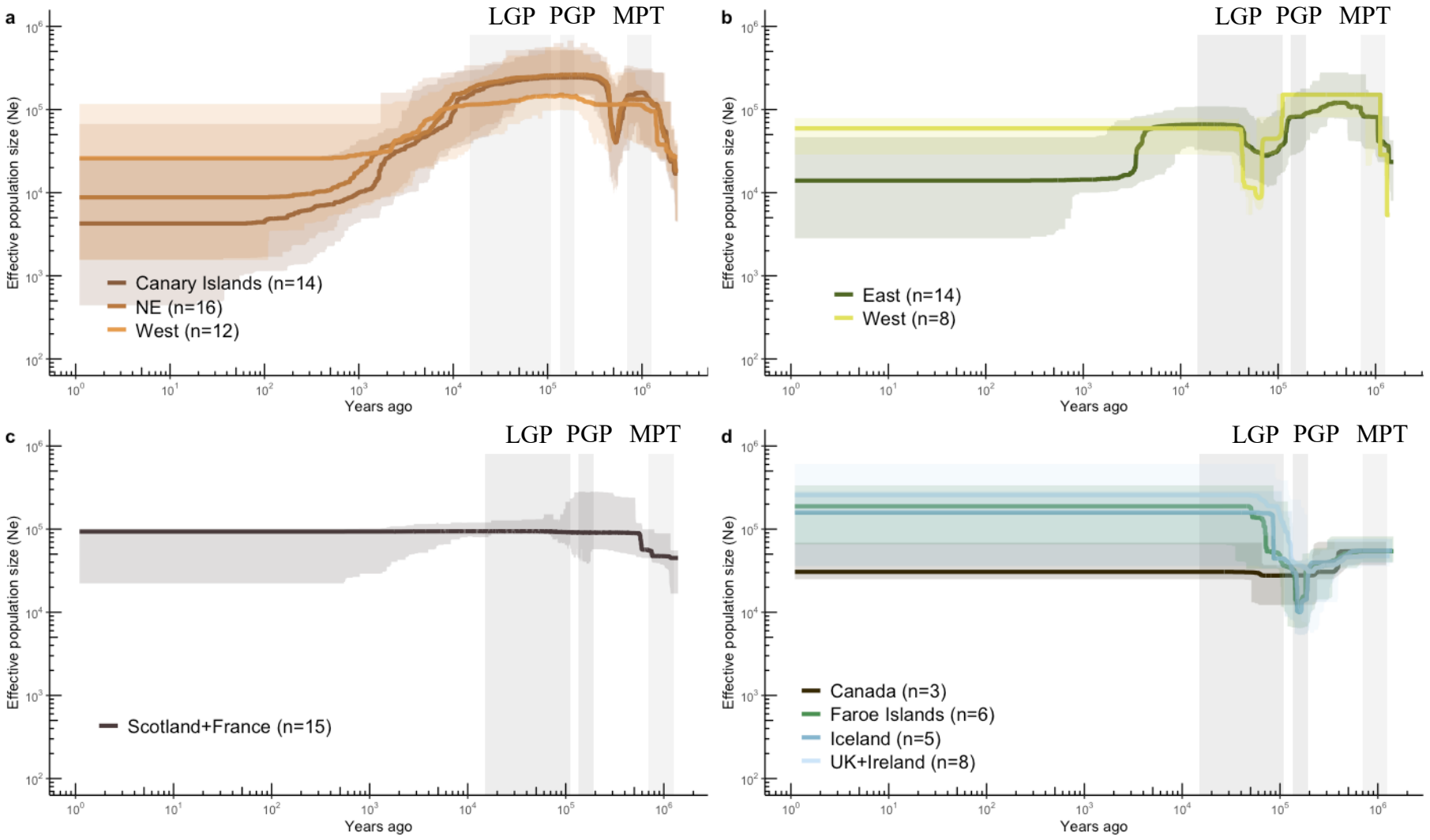


Figure 4-14 (a-d). Reconstructed demographic histories of a) Cuvier's beaked whale populations, b) Blainville's beaked whale, c) Sowerby's beaked whale, and d) northern bottlenose whale. Solid lines are the median effective population size (N_e), and shaded areas are 95% confidence intervals. Both axes are on a log scale and are in the thousands. The vertical shaded bars represent major climatic events: Last Glacial Period (LGP; 110-15 k years ago (kya), Penultimate Glacial Period (PGP; 194-135 kya) and the Mid-Pleistocene Transition (MPT; 1250-700 kya).

4.4 DISCUSSION

The demographic histories of North Atlantic beaked whale populations are presented here in comparison to periods of notable historic climate change, and notable changes in population trajectories may be linked to wide-scale cooling and warming of the ocean basin, with fluctuations in salinity, ice cover, trophic dynamics, and currents. While the reconstructed demographic histories presented here did not show straightforward responses to past climate change according to species, body size or species range, I hypothesize that the changes in N_e were most likely influenced by a complex relationship between physiology, behaviour, and evolutionary history.

4.4.1 Population Structure and Genetic Diversity

In the absence of physical barriers to movement and dispersal, genetic population structure is found in many large and mobile marine species (Bowen and Karl 2007; Castro et al. 2007; Hoelzel 2009; Clarke et al. 2015; Gonçalves da Silva et al. 2020). While comprising a complex system of oceanographic features, the North Atlantic Ocean does not present any physical barriers to large mammals such as beaked whales, yet structure on both large and fine scales has been detected (results here, Chapter 3, Dalebout *et al.*, 2001, 2006; Feyrer *et al.*, 2019; de Greef *et al.*, 2022). The patterns of genetic structure in Cuvier's and Blainville's are discussed at length in Chapter 3, so here I will focus on the Sowerby's and N. bottlenose findings.

Based on differences in skull morphology and staple isotopes, separate Sowerby's populations in the eastern and western North Atlantic have been proposed (Smith, Mead, et al. 2021; Smith, Trueman, et al. 2021). In this first genomic study of Sowerby's population structure, genetic clusters detected using *tess3r* did not correspond to geographic origin of the samples. Several factors could have contributed to the lack of genetic structure found here. Sample sizes were sparse across the species' entire range, potentially missing populations. Samples were also biased towards the eastern North Atlantic (only two individuals from the west), similar to the studies that indicated east and west differences in Sowerby's skull morphology (west: n=45, east: n=129; Smith, Mead, et al. 2021) and stable isotopes (west: n=38, east: n=64; Smith, Trueman, et al. 2021). Additionally, all Sowerby's samples came from stranded animals, making it impossible to pinpoint the fine scale origin of those individuals. Sowerby's populations may also be structured by something other than geography, such as preference for different environmental conditions or ecological niches (like bottlenose dolphins, Louis *et al.*, 2014). Finally, it was presumed that genetic structure would be found as it has been found in most other studied ziphiids, however it may be that Sowerby's in the North Atlantic are a panmictic species like Gray's beaked whale across the Southern Hemisphere (*Mesoplodon grayii*, Westbury *et al.*, 2021).

N. bottlenose have shown lower levels of genetic diversity than other beaked whale species occupying similar niches due to their more limited geographic range (Feyrer et al. 2019). Investigations using mitochondrial DNA, nuclear microsatellites and nuclear genomes have not been able to paint a clear picture about N. bottlenose structure and connectivity, though it is understood that the Scotian Shelf population (including the well-studied "Gully" population, Dalebout *et al.*, 2006) seems to be genetically distinct from N. bottlenose sampled elsewhere (Dalebout et al. 2001; Dalebout et al. 2006; Feyrer et al. 2019; de Greef et al. 2022). Evidence also shows that the population in Iceland is genetically distinct (de Greef et al. 2022). In the current study however, we did not see

differentiation between the individuals sampled from eastern Canada compared to Iceland using *tess3r*. The lack of differentiation between the Canadian and Icelandic individuals may have been due to a smaller sample size or a lack of resolution in the SNP dataset. The "Other" cluster showed significant differentiation from the remaining clusters, however this cluster included individuals from a range of geographic locations who all shared similar levels of admixture between the Quebec/Iceland and Faroe Islands clusters and it is very unlikely that this group represents a meaningful demographic or evolutionary unit. The Faroe Islands were genetically distinct from the Icelandic cluster, an interesting result considering their close geographic proximity and that these were considered to be part of the same historic whaling stock (Whitehead and Hooker 2012; Whitehead et al. 2021). North-south and inshore-offshore migrations have been suggested for N. bottlenose in the eastern North Atlantic, which may have resulted in the lack of clear genetic similarity amongst individuals sampled in the same region, as they could have originated, or reproduced, elsewhere.

4.4.2 Demographic History

Population bottlenecks and expansions were detected in three species (Figure 4-14) spanning four discrete time periods. Starting the furthest back in time, Cuvier's populations (Canary Islands and Northeast) experienced a bottleneck ~800 kya with an expansion following ~500 kya (Figure 4-14a). Between ~150-50 kya, a bottleneck and later expansion occurred in N. bottlenose population (Faroe Islands, UK+Ireland and Iceland, Figure 4-14d) and Blainville's populations (East and West, Figure 4-14b). Three Cuvier's populations (Canary Islands, Northeast, and West) also experienced a gradual contraction in N_e from ~10 kya and Blainville's (East) underwent a substantial decline in N_e from 60 k to 30 k over the course of ~1 k years. Each of these time periods coincides with glacial and interglacial events that have impacted the demography of other species and will be discussed here.

The bottleneck and following expansion ~800 kya in Cuvier's may have coincided with the culmination of the Mid-Pleistocene Transition (MPT). This period saw a shift in the global climate from less severe glaciations on a 40 k year scale to high amplitude swings between extreme glacial and interglacial periods on 100 k year scales (McClymont et al. 2013). This transition has been suggested as the driver behind differentiation in rockhopper penguin (*Eudyptes chrysocome sensu lato*) populations (De Dinechin et al. 2009) and an expansion of green sea turtle (*Chelonia mydas*) N_e (Fitak and Johnsen 2018).

The temporal range of the bottlenecks experienced by N. bottlenose and Blainville's encompasses the penultimate glacial period (PGP, 194 kya – 135 kya), the Eemian Interglacial (130 kya-115 kya) and the start of the last glacial period (LGP, 115 kya), suggesting these declines could have been the result of periods of warming, cooling, or the transition between the two. During the PGP, glacial ice covered its most expansive range in the last 400 k years, corresponding with an extensive drop in global sea level and shifting temperature zones (Colleoni et al. 2016). During this time, there was a possible increase in killer whale N_e (*Orcinus orca*; Moura *et al.*, 2014), divergences in both beluga and walrus populations (Shafer et al. 2015; Skovrind et al. 2021), and decreases in Antarctic fur seal (*Arctocephalus gazella)* and California sea lion (*Zalophus californianus)* populations (Peart et al. 2020). The Eemian interglacial was a period of similar air and sea surface temperatures as today, and saw an increase in sperm whale N_e (Morin et al. 2018) and a decline in killer whale N_e (Moura et al. 2014). The start of the

LGP brought on the start of a marked decline in beluga N_e (*Delphinapterus leucas;* Skovrind *et al.*, 2021).

The later contraction shared by Cuvier's populations could have coincided with the end of the LGP or the "8.2 kiloyear event" (Alley et al. 1997; Törnqvist and Hijma 2012), where rapid warming resulted in a massive discharge of glacial melt water into the North Atlantic, disrupting and reducing primary productivity and potentially limiting prey. Environmental changes during this time are proposed to have impacted the population trajectories of a number of baleen whales and their prey (Cabrera et al. 2022). The most recent decline in the east Blainville's population may have coincided with the "4.2 ka" event (Ran and Chen 2019), however population declines occurring near the present may be an artefact of underlying population structure in skyline plots (Heller et al. 2013).

Any event with a lasting impact on primary productivity in the ocean can cause downstream effects on the population trajectories of higher trophic level species. Extended periods of global warming (such as the Eemian interglacial) can impact pelagic species by disrupting upwelling and reducing overall primary productivity and prey availability (Schmittner 2005). However, productivity may be maintained in areas with complex topography that facilitates upwelling (such as the areas where persistent beaked whale populations are found; MacLeod and Zuur, 2005; Ferguson *et al.*, 2006). Reducing prey availability can impact demographic history by reducing the overall population size and may lead to reproductive isolation and resulting genetic structure if the ranges of prey species also become fragmented. This disruption of upwelling and reduction of prey is thought to have led to speciation in minke whales (*Balaenoptera acutorostrata*) during the warm Pliocene (Pastene et al. 2007). Alternatively, periods of warming were beneficial for some species with limited ranges during glacial periods, as evidenced by the beluga population expansions during interglacials (Skovrind et al. 2021). Populations of N. bottlenose in this study also appear to rebound during the Eemian interglacial, before reaching their peak N_e during the LGP.

Prolonged cool periods can shift distributions due to reduced sea level, reduced SST, and extended ice sheets (Banguera-Hinestroza et al. 2010). These extended cool periods and their terminations are proposed to be the most likely events associated with population declines and/or extinctions, such as that seen in killer whales (Moura *et al.*, 2014). Redistributions of species ranges into glacial refugia may also increase reproductive isolation, resulting in genetic structure of populations and/or speciation over time (Taguchi et al. 2010). Extended periods of cooling have also resulted in peaks of primary productivity, with predators expanding (and potentially diverging) to exploit new ranges of prey distributions (Harlin-Cognato et al. 2007).

A potentially synchronous bottleneck was detected in populations of N. bottlenose and Blainville's ~150-50 kya, but not in Cuvier's or Sowerby's, whose populations were at their highest N_e during these years. Populations that did and did not experience the bottleneck spatially overlap, ruling out a direct latitudinal link to the climatic event. Differences in body size may also be ruled out, as N. bottlenose and Cuvier's are the two biggest species, however the available habitat to N. bottlenose would have been reduced in glacial periods. Although Cuvier's and Blainville's both have global distributions, this bottleneck does not appear to have impacted the Cuvier's populations. In chapter 3, I propose that the smaller Blainville's may have a higher thermal limit than Cuvier's, suggesting they may respond to extended cool periods in a different way. Perhaps the

reduction in available habitat resulted in less gene flow between populations, with declines in N_e representing population declines as well as increasing genetic structure. Sowerby's are not much bigger than Blainville's (5.5m vs 4.7m; Pitman, 2018), however as their normal distribution extends into much colder waters than Blainville's suggesting a lower thermal limit. Site fidelity can also be ruled out, as Cuvier's, Blainville's and N. bottlenose show some level of philopatry, but differing demographic histories (Reyes 2018; Baird 2019; Feyrer et al. 2019; de Greef et al. 2022). A straightforward cause and effect relationship does not seem to explain the presence/absence of this bottleneck in North Atlantic beaked whale species, and a more complex relationship involving niche partitioning may be more appropriate.

Beaked whales display resource partitioning, with different species preferring certain types and/or sizes of prey (MacLeod et al. 2003; Spitz et al. 2011; Baird 2019). In general, *Mesoplodon* species eat consistently smaller prey than Cuvier's and N. bottlenose, and tend to forage in shallower depths (Baird et al. 2006; Rogan et al. 2017; Baird 2019). Cuvier's and N. bottlenose occupy a similar dietary niche, though the niche breadth of N. bottlenose is significantly smaller (Whitehead et al. 2003). Populations of Cuvier's and Blainville's are often geographically segregated; for example in the Bahamas and Hawai'i, Blainville's are more commonly observed in shallower water compared to Cuvier's and dive to shallower depths to forage (MacLeod et al. 2003; Baird 2019). In the eastern North Atlantic, Sowerby's diet includes significant amounts of fish and crab, and they forage on the shallower continental shelf compared to the more offshore Cuvier's (Spitz et al. 2011). Although Sowerby's and Blainville's are of similar size and are thought to share the same ecological niche, their foraging behaviour differs greatly (Visser et al. 2022). While most ziphiids use a relatively slow and energy conserving style of diving, Sowerby's swim, find and catch their prey faster, and make shorter duration dives to a broad depth range, expanding the breadth of their niche to include larger and more energetic prey types (Visser et al. 2022). Perhaps differences in ecological niche and foraging behaviour drove the different demographic responses of Blainville's and Sowerby's to historic climate change.

While not impacted by the bottleneck discussed previously, Cuvier's in the eastern North Atlantic did experience an earlier bottleneck that the western North Atlantic population did not. Perhaps productivity changes occurred more significantly during the wider amplitude glaciations following the MPT in the eastern North Atlantic compared to the more tropical west. It is interesting to note that this MPT related decline in eastern Cuvier's populations was not seen in the Blainville's data, as these species are often found in sympatric resident populations (Baird et al. 2009; Claridge et al. 2015; Reyes 2018; Baird 2019; Hooker et al. 2019).

While the sharp declines in N_e detected here in beaked whale populations were assumed to be the result of a bottleneck, the formation of new populations from a small subset of individuals ("founder effect") and underlying population structure could also have resulted in the patterns seen here (Heller et al. 2013; Mazet et al. 2015). For example, reductions in N_e as a result of the foundation of new populations has been seen in European roe deer (*Capreolus capreolus;* de Jong *et al.*, 2020), the brown rat (*Rattus norvegicus;* Puckett *et al.*, 2020), and the Eurasian lynx (*Lynx lynx;* Lucena-Perez *et al.*, 2020).

4.4.3 Beaked Whales and Future Climate Change

The geographic ranges of cetaceans are primarily influenced by water temperature (MacLeod 2009), and with climate change expected to increase global SST by up to 3° C by 2100 (Hoegh-Guldberg et al. 2014), scientists have tried to predict what this affect may be on cetacean communities (van Weelden et al. 2021). As a result of the complex interactions of currents and climate systems, notably the AMOC, warming of the North Atlantic is occurring around the periphery of the basin and cooling in the centre (Caesar et al. 2018; Pershing and Stamieszkin 2020). Trophic dynamics in the North Atlantic are based on the seasonal blooms of phytoplankton, which feed the seasonally abundant zooplankton (genus *Calanus*). The predictable booms in *Calanus* provide the basis for all higher trophic level species in the North Atlantic, and the current warming trend is disrupting this timing (Sundby et al. 2016; Pershing and Stamieszkin 2020). Furthermore, a shift is occurring whereby the range of *C. finmarchicus*, the smallest and lowest lipid content of the three *Calanus species,* is expanding northward and changing the distribution of biomass in the Arctic away from the more nutritious *Calanus* species with likely impacts on higher trophic levels (Freer et al. 2022).

Warming of the eastern North Atlantic has already resulted in shifting dynamics of both prey and cetacean species (MacLeod et al. 2005; Golikov et al. 2013), and the two species in this study with North Atlantic distributions (N. bottlenose and Sowerby's) are predicted to have an unfavourable response to climate change (MacLeod 2009; van Weelden et al. 2021). Warming seas will likely mean poleward shifts in distribution, and for species in the Northern Hemisphere, this could result in the mixing of previously genetically isolated populations and/or competition with similar niche species in the ice-free Arctic Ocean (MacLeod 2009; van Weelden et al. 2021). Cuvier's and Blainville's have nearly cosmopolitan distributions, with unchanged or favourable predicted responses to climate change, respectively (MacLeod 2009; van Weelden et al. 2021). Both species may also expand their ranges poleward and encounter species of similar niches to compete for resources with, which could also increase the likelihood of potential mixing between genetically distinct units (MacLeod 2009). Site fidelity has been identified on some scale in each of these four species (Dalebout et al. 2006; Hooker et al. 2019; Smith, Trueman, et al. 2021), and preference for local bathymetry and habitat preference may trap and ultimately extirpate geographically isolated populations, keeping them from following suitable habitat as it shifts poleward with rising temperatures (like belugas MacLeod, 2009; Skovrind *et al.*, 2021). If populations do ending up shifting their ranges to follow suitable habitat, this may impede conservation efforts if mismatches occur between the population of interest and any protected areas in place to protect them (i.e. the proposed MPA for Cuvier's in the Alborán Sea or the Gully MPA for N. bottlenose whales; O'Brien and Whitehead, 2013; Cañadas and Vázquez, 2014).

4.4.4 Caveats and Future Work

The results of the demographic analyses presented here are only as good as the data used to generate them, and several parameters are subject to caveats. Methods used to calculate N_e have many of the same assumptions as a Wright-Fisher population: random mating, non-overlapping generations, constant population size, panmixia and no forces such as selection, recombination, mutation or drift (Charlesworth 2009; Salmona et al. 2017). Many of these assumptions are violated in most natural populations, impacting estimations of effective population size (reviewed in Charlesworth, 2009), however some can be accounted for in the sampling scheme and analysis.

Stairway Plot 2 was specifically developed to estimate changes in population size over time without needing to assign an underlying population model, based on the site frequency spectrum (SFS). This fundamental summary statistic collapses genome-wide SNP data into a single vector of the allele frequency distribution of a (or several) population (Salmona et al. 2017). SFS files and the resulting analyses of demographic history are biased by the bioinformatic pipeline chosen to identify and call SNP genotypes (Shafer et al. 2017). Shafer *et al.* (2017) make several recommendations for studies involving RADseq data, one of which is to use a closely related reference genome for calling SNPs. This was done in the current study, as the Cuvier's reference genome was used for each of our four species which are within the same family. Other recommendations that were not considered here but could be incorporated into future analyses include using multiple bioinformatic pipelines to assess the robustness of the findings and to generate SFS files directly from genotype likelihoods instead of explicit calls, which is an option in 'ANGSD', (Korneliussen et al. 2014; Rochette et al. 2019; Warmuth and Ellegren 2019). This genotype likelihood method is advantageous over genotype calls in many situations, namely when the data are low or medium coverage (Korneliussen et al. 2014).

Scaling the demographic history trajectories to discrete years is dependent on the estimates of species-specific generation times. Multiple definitions of "generation time" exist, including the most commonly used average parental age of offspring in a cohort (Bienvenu 2019), however this is not available for most beaked whale species and is therefore a potential caveat in the demographic histories presented here. This definition of generation time was available for N. bottlenose, and ranged from 15.5 years (COSEWIC 2011) to 17.8 years (in an assumed stable population; Taylor *et al.*, 2007). In the remaining three species, generation times have been assumed based on those of other beaked whale species, but species-specific ages at sexual maturity are available. For Cuvier's, the estimated age at sexual maturity is 10 to 15 years (ACCOBAMS) and a generation time of 15 years has previously been used (based on N. bottlenose estimates; Dalebout *et al.*, 2005). Estimates for Sowerby's range from reaching sexual maturity at 7 years (NOAA Fisheries 2022) to a generation time of 15-30 (based on other beaked whales; COSEWIC, 2006). The estimated age of sexual maturity for Blainville's ranges from 9 to 11 years (Mead 1984; Aguilar de Soto et al. 2015). In the absence of sufficient species-specific generation times, 10 years was used as an estimate for all ziphiids in Dalebout (2002). In this study, I used the generation times presented in Table 4-1, balancing the various estimates of generation time and age at sexual maturity. While the pattern of demographic changes (in terms of N_e) would not change with different generation times, the timing will (e.g.,
Figure 4-15). It must be stated that differences in the timings of trajectory changes could therefore be a result of incorrect generation times, and not differing response to climate and ecosystem changes.

Demographic reconstructions are unreliable if they do not account for underlying population structure. This can appear as a decline in N_e seen close to the present, a phenomenon called the "structure effect" (Heller et al. 2013). Underlying structure could be the reason for the recent decline in Blainville's from the eastern North Atlantic (primarily Canary Island population). Photo-identification studies have found little exchange of individuals between the West and East Canary Islands and Blainville's are thought to exhibit higher site fidelity than their Cuvier's counterparts in the archipelago (Reyes 2018). As all individuals from this region were analysed together, recent

separation and genetic isolation of these two populations may be influencing the population trajectory seen here. Genetic structure associated with geographic location of origin was not clear for Sowerby's, with the "Other" population including individuals from both the eastern and western North Atlantic. Sample sizes were sparse, and since structure has been proposed based on skull morphology and stable isotopes (Smith, Trueman, et al. 2021; Smith, Mead, et al. 2021), further sampling across the North Atlantic for Sowerby's is recommended to determine whether the lack of genetic structure we have detected is real, or an artefact of sampling.

Stairway Plot 2 is a useful tool to model and generate hypotheses about the demographic histories of populations. Moving forward, simulation-based approaches such as those employed using Approximate Bayesian Computation (ABC), can be used to test for more complex scenarios that may include changes in population size as well as population divergence. A useful analysis would be to test the synchronicity of the bottleneck seen in N. bottlenose and Blainville's populations using the framework implemented by 'Multi-Dice' (Xue and Hickerson 2017). Finally, divergence dating between species and populations would provide complementary evidence of impacts of past climate change if these events coincided with known climatic events.

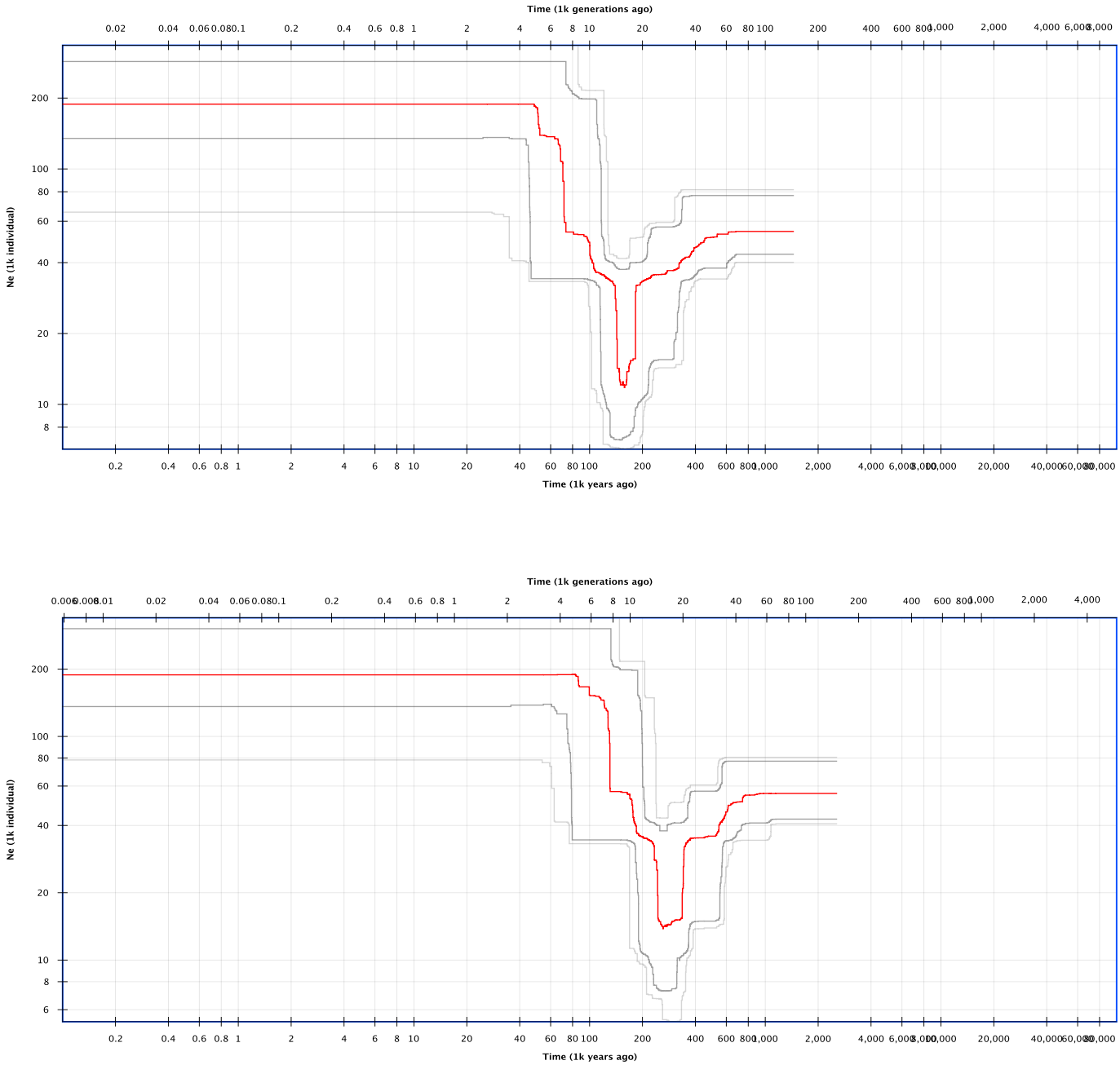


Figure 4-15. A comparison of reconstructed demographic histories for N. bottlenose sampled in the Faroe Islands assuming a generation time of 10 years (upper) and 17.8 years (lower).

5 Conservation in the deep: Genetic essential biodiversity variables in ‘disturbed’ and ‘semi-pristine’ beaked whale populations

5.1 INTRODUCTION

Anthropogenic noise pollution has been classified as one of the greatest threats to ecosystem health in both terrestrial and marine environments (Shannon et al. 2016). Marine environments, especially the deep sea, are particularly sensitive to these threats, as sound propagates a greater distance in the deep sea and many of the diverse inhabitants here use acoustic cues due to the absence of light (Nowacek et al. 2007; Hildebrand 2009; Sutton and Milligan 2019). A group of major deep-sea predators particularly sensitive to anthropogenic noise are marine mammals (Nowacek et al. 2007; Hildebrand 2009; Nelms et al. 2021). One specific intersect between noise pollution and deep-sea research has been the recent understanding that the beaked whale family (Ziphiidae), and likely other cetaceans, are negatively impacted by naval sonar use (Evans et al. 2001; Nowacek et al. 2007; Tyack et al. 2011; Parsons 2017).

Since the early 1970's a connection was suggested between military exercises at sea and strandings of beaked whales (van Bree and Kristensen 1974; Frantzis 1998; Balcomb III and Claridge 2001; Jepson et al. 2003). The development of biologging devices that can collect sound recordings (DTAGs; Johnson and Tyack, 2003) resulted in the measurement of acoustic "masking" (obscuring or interfering with natural sounds; Weilgart, 2007) and behavioural responses to noise (shipping traffic, predator calls, sonar playbacks, etc) in beaked whales (Aguilar de Soto et al. 2006; Tyack et al. 2011; Aguilar de Soto et al. 2012; DeRuiter et al. 2013; Allen et al. 2014). While this correlation between anthropogenic noise and beaked whale strandings is widely accepted among the scientific community, the mechanism causing animals to strand is not completely understood (Cox et al. 2006).

Beaked whale mass strandings are not particularly common (Filadelfo et al. 2009), but they raise a lot of public attention when they occur (Parsons 2017). During these events, a discrete number of individuals may be removed from the population through fatal strandings, but it is not understood how this may cause long-term, population-level impacts. This uncertainty has led to a dramatic increase in studies relating to beaked whale biology and ecology (Hooker et al. 2019). In particular, much work has been done investigating the two widest ranging ziphiids (MacLeod et al. 2006; Baird 2019), Cuvier's and Blainville's beaked whales (*Ziphius cavirostris* and *Mesoplodon densirostris*, respectively and henceforth "Cuvier's" and "Blainville's"), to provide baselines from which impacts can be measured.

Cuvier's and Blainville's have nearly cosmopolitan distributions (Macleod 2000; MacLeod et al. 2006) and hierarchical genetic structure has been found in both species (Chapter 3, Dalebout *et al.*, 2005; Morin *et al.*, 2012). Biogeographic barriers, site fidelity, evolutionary and life history drive the distributions of Blainville's and Cuvier's (Chapter 3). Resident populations of both species, where individuals are re-sighted over many years in the same region using photo-identification, have been identified across the globe (Baird 2019; Hooker et al. 2019). Three regions in which Blainville's and Cuvier's are regularly found are the Canary Islands, the Bahamas, and the Mediterranean Sea (latter, Cuvier's only; Figure 5-1). In each of these regions, "key" areas for beaked whales have been identified (MacLeod and Mitchell 2006) and resident populations have been investigated in long-term studies (Rosso et al. 2011; Claridge 2013; Podestà et al. 2016; Reyes 2018; Hooker et al. 2019). Populations in each of these regions were identified as genetically distinct Demographically Independent Populations (DIPs) (Chapter 3) and

have experienced mass strandings which correlate with naval sonar activity (D'Amico et al. 2009; Hooker et al. 2019).

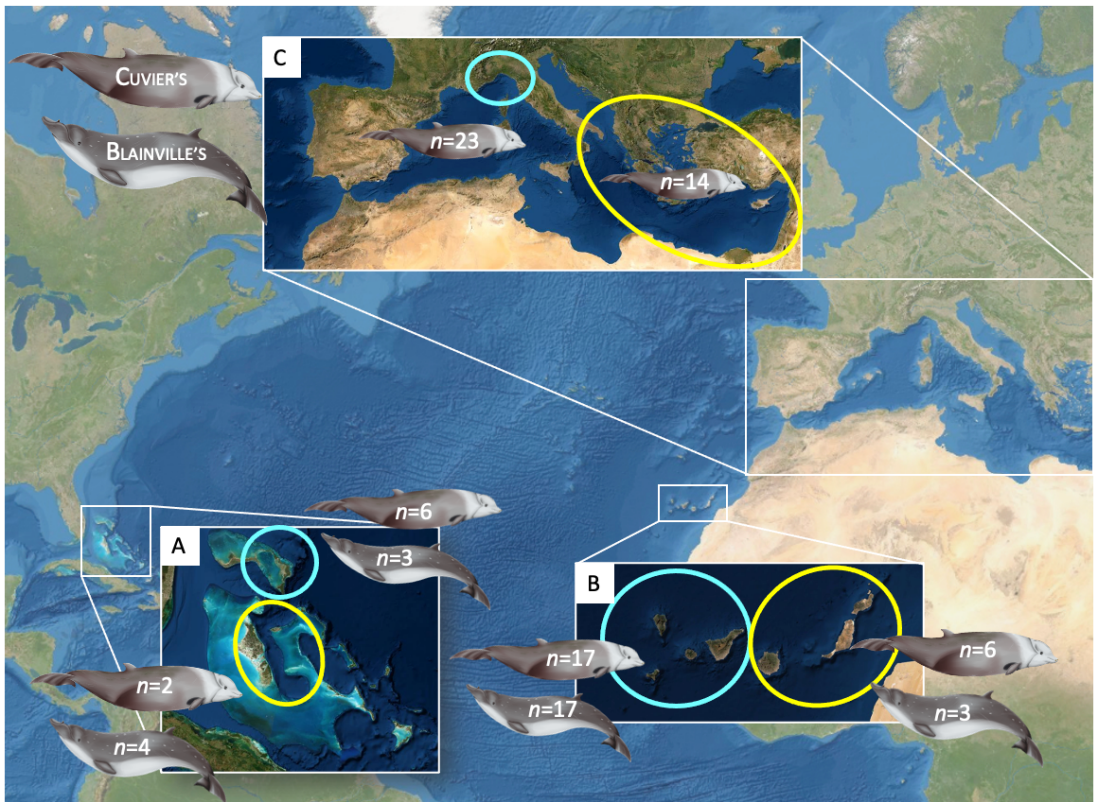


Figure 5-1. Map of the three focal regions included in this study (A. Bahamas, B. Canary Islands, and C. Mediterranean Sea), including species-specific sample sizes for each of the "disturbed" populations (circled in yellow) and "semi-pristine" populations (circled in blue).

Cuvier's and Blainville's are regularly sighted in the Canary Islands, where long-term photo-identification studies have estimated residency patterns and life-history parameters (Aguilar de Soto 2006; Tejedor et al. 2010; Reyes 2018). The population sizes of Blainville's and Cuvier's off El Hierro (western Canary Islands) are estimated to be *n*=103 and *n*=87 individuals based on mark recapture, respectively, with respective estimates of the percentage of resident individuals being 35% and 53% (Reyes 2018). Both species are regularly sighted in the Bahamas, and long-term photo-identification studies of Blainville's and Cuvier's are ongoing in the Atlantic Undersea Test and Evaluation Centre (AUTEC; a naval testing range) and around Abaco Island (Claridge 2013; Claridge et al. 2015; Hooker et al. 2019). In Abaco, *n*=48 Blainville's have been individually identified and 34% of these have been sighted more than once (Claridge 2013). In AUTEC, *n*=30 Blainville's have been individually identified, 35% of which were resighted at least once (Claridge 2013). There have not been any photographic matches between AUTEC and Abaco (Claridge 2013). A thorough analysis of Cuvier's photo-identification data in the Bahamas is underway (Claridge 2013), and so far photo-identification, telemetry, dietary and contaminant analyses show that Cuvier's in the Bahamas exhibit site-fidelity to the geographically distinct regions that were investigated, including Tongue of the Ocean and Abaco Island, with limited movements between them (Claridge et al. 2015). Across the Mediterranean Sea, there is an estimated ~5700 Cuvier's (Cañadas et al. 2018), with the three highest density areas being Genoa Canyon in the Ligurian Sea: 100 individuals (95% CI: 79-116; estimated from photo-identification data; Podestà *et al.*, 2016), the Hellenic Trench: 512 individuals (95% CI:

387-755; estimated from visual surveys between 2018-2021; Frantzis *et al.*, 2022), and the Alborán Sea: 429 individuals (95% CI: 334-557; estimated from line transect surveys; Cañadas and Vázquez, 2014).

Beaked whale populations living in each of the areas described above have also been subjected to historic or current anthropogenic disturbance which resulted in mass strandings (D'Amico et al. 2009; Filadelfo et al. 2009). In the Canary Islands, nine mass strandings (2 or more individuals) occurred between 1985 and 2004 around Fuerteventura and Lanzarote in the east involving 63 individuals (from five species including *n*=49 Cuvier's and *n*=5 Blainville's), some of which were correlated in space and time with naval sonar activity (Simmonds and Lopez-Jurado 1991; Jepson et al. 2003; Martín et al. 2003; Fernández et al. 2005; D'Amico et al. 2009; Bernaldo de Quirós et al. 2019). The last of these strandings, which took place in 2004, resulted in the deaths of four Cuvier's, and occurred 6-12 days after international naval exercises were carried out north of Lanzarote and Fuerteventura (Fernández et al. 2012). Following this, the Spanish government implemented a sonar moratorium around the Canary Islands in 2004, and no mass strandings have taken place since (Fernández et al. 2012; Fernández and Martín 2013). The AUTEC naval range in the Tongue of the Ocean (TOTO), Bahamas is the site of regular mid-frequency active sonar (MFAS) used during training exercises (Claridge 2013). In 2000, a mass stranding involving at least 14 beaked whales (including *n*=9 Cuvier's, *n*=3 Blainville's, and *n*=2 unidentified ziphiids) followed antisubmarine warfare exercises occurring nearby in the Northern Bahamas (Balcomb III and Claridge 2001; Evans et al. 2001). This is considered the first major stranding event known to be associated with a U.S. Navy sonar exercise (Filadelfo et al. 2009).

In the Mediterranean, seven mass strandings of Cuvier's occurred between 1961 and 1996 in the Ligurian Sea (Podestà et al. 2006; D'Amico et al. 2009) involving 44 individuals (>1/3 of the current estimated abundance; Podestà *et al.*, 2006, 2016; D'Amico *et al.*, 2009). Three of these events are thought to have been correlated with naval activity (D'Amico et al. 2009). Mass strandings of Cuvier's in Greece have been noted since 1996, however the first to be correlated with naval exercises occurred in 1996, when 12 Cuvier's died after stranding in geographic and temporal proximity to low frequency active sonar (LFAS) testing by a NATO research vessel (Frantzis 1998; D'Amico et al. 2009). The most recent mass stranding of 6-10 Cuvier's in Greece occurred in 2014 in Crete, coincident with multinational naval exercises involving anti-submarine warfare and military sonar (Frantzis 2015).

The events described above suggest that beaked whale responses to disturbance can result in fatal strandings. Studies are ongoing to understand individual-level responses to disturbance (Tyack et al. 2011; Aguilar de Soto et al. 2012; DeRuiter et al. 2013; Allen et al. 2014; Aguilar de Soto et al. 2020; Joyce et al. 2020) and why those responses may result in strandings (Fahlman et al. 2014; Quick et al. 2020); however, the potential long-term impacts of these events on beaked whale populations and species are yet to be studied in detail (Harwood et al. 2016). It is thought that the mortalities resulting from mass stranding events may have significant negative impacts on beaked whale populations through mechanisms including energetic costs to direct mortality, though this is difficult to assess without population data that pre-dates sonar use (Bernaldo de Quirós et al. 2019). There is also some evidence of long-term impacts in Blainville's from the AUTEC naval range, where reproductive rates and recruitment through births are lower compared to the nearby Abaco population (Claridge 2013). Energetic models suggest that

the disturbances which may displace individuals from high quality habitat, could impact the lifetime production of females by increasing the interval between calves (New et al. 2013).

One means by which population health can be assessed in response to mass strandings is through the use of Essential Biodiversity Variables (EBVs, Pereira et al. 2013; Hoban et al. 2022). This series of metrics has been proposed to assess impacts of natural and anthropogenic drivers, while also tracking changes in biodiversity (Pereira et al. 2013). Within this framework, measurements of genetic variation are considered to be an integral component in the assessment of biodiversity change by providing species with resilience and adaptability to perturbations (Hoban et al. 2022). Four EBVs are proposed to measure genetic variation within a species. These encompass the components required to assess the health and viability of a population (Hoban et al. 2022): genetic differentiation, genetic diversity, inbreeding and effective population size (N_e).

Genetic variation is shaped by the four forces of evolution (mutation, genetic drift, gene flow and selection), which in turn are influenced by species-specific life history traits, and external drivers such as anthropogenic and environmental variation (Hoban et al. 2022). In wild populations, there is a positive correlation between population size, genetic diversity, and fitness, with smaller populations being subject to stronger genetic drift as measured by N_e (Hansson and Westerberg, 2002; Bouzat, 2010; Hoelzel, 2018). Life history traits found in many marine mammal species can also influence the rate of genetic drift, including mating systems that create reproductive skews (i.e., the harem structure in Blainville's and Cuvier's; Baird, 2019) and overlapping generations (Waples et al. 2014). Estimating N_e using genomic data is potentially a useful conservation tool to monitor the rate at which genetic variation declines in a population, and N_e is also reduced when inbreeding and underlying genetic structure are present in a population (Charlesworth 2009).

Longitudinal photo-identification studies show that both Cuvier's and Blainville's live in small resident populations and socially-complex groups (Baird 2019; Hooker et al. 2019) and genetic studies indicate that these populations are genetically distinct (Chapter 3 and 4). The negative anthropogenic impacts summarised above in conjunction with the presence of small, genetically isolated population suggests these species may be at an increased risk of declining genetic variation (Leroy *et al.*, 2017; Hooker *et al.*, 2019).

In this chapter, I use the Genetic EBV framework (Hoban et al. 2022) to quantify and compare the genetic composition of well-studied beaked whale populations and discuss their viability to continued and future perturbations. In each of the three regions described above (Canary Islands, Bahamas and Mediterranean), beaked whales in neighbouring locations have been impacted by naval activities at different intensities, and here I use a paired-site approach to evaluate potential population-level effects in the form of Genetic EBV metrics. The focal sites that have been less disturbed by anthropogenic noise (termed "semi-pristine") for this study are El Hierro, west Canary Islands ("west Canary Islands"); the island of Abaco, Bahamas ("Abaco"); and the Ligurian Sea ("west Mediterranean"). The focal sites that have been more subjected to disturbance by naval sonar are termed "disturbed" and include the east Canary Islands ("east Canary Islands"); Tongue of the Ocean, Bahamas ("TOTO"); and Ionian Sea ("east Mediterranean") (Figure 5-1). I first assess EBV1 (population structure), to confirm the distinct genetic units at which to measure subsequent genetic diversity metrics. I then calculate metrics for EBVs 2-4

(genetic diversity, inbreeding and N_e), and predict that there will be reduced genetic variation in the "disturbed" compared with "semi-pristine" sites, specifically:

1. stronger genetic drift as measured by smaller N_e values
2. lower genetic diversity, as measured by heterozygosity and allelic richness
3. higher inbreeding, as measured by the inbreeding coefficient (F_{IS}), internal relatedness (IR), and an estimate of the potential for future inbreeding though an estimation of kin networks.

5.2 METHODS

5.2.1 Laboratory Methods

Samples were selected from the International Tissue Archive for Beaked Whales (ITABW; 2.3.3) for inclusion in this study if they originated in one of the regions of interest: Canary Islands, Mediterranean and Bahamas. DNA was extracted using the Phenol:Chloroform:Isoamyl Alcohol protocol (Sambrook *et al.*, 1989; see Chapter 2 for details) and ddRAD libraries were prepared following the methods of (Peterson et al. 2012) and modified for beaked whales (Carroll *et al.*, 2016, 2021; see Chapter 2 for details). A total of 16 sequencing libraries were prepared containing DNA from 10 individuals in each. One additional library contained four individuals. Libraries were prepared with a mix of Cuvier's and Blainville's samples, pooling together samples of similar quality (Chapter 2, Table 2-2). Prepared libraries were checked for DNA quantity and quality using a Tape Station by Xelect Ltd. (https://xelect-genetics.com/), with 15 libraries successfully passing quality control. The final libraries were evenly divided into three sequencing pools, containing DNA from *n*=42 Blainville's and *n*=108 unique Cuvier's. DNA libraries were sequenced on an Illumina NovaSeq, using paired-end 150 bp chemistry at NovogeneAIT Genomics Singapore. Some of the samples that were sequenced in 2018 for Chapters 3 and 4 underwent ddRAD library preparation and sequencing again to improve the quality. Finally, high quality ddRAD sequences were available from some Cuvier's and Blainville's that were generated in 2018 for Chapters 3 and 4. A summary of the samples from just the three focal regions and their sequencing date is presented in Table 5-1.

5.2.2 Assessing reproducibility of ddRAD sequencing

A total of *n*=21 Blainville's and *n*=25 Cuvier's that were sequenced in both 2018 and 2020 were selected to investigate the reproducibility of ddRAD sequencing across multiple sequencing platforms and laboratories. This also provided an opportunity to determine the effectiveness of the bioinformatic protocol developed for the Illumina HiSeq2500 using a new sequencing platform (Illumina NovaSeq). A summary of this investigation is presented in Chapter 2 (2.8.3).

5.2.3 SNP discovery and genotyping

DNA sequences were analysed using the bioinformatic pipeline described in Chapter 2 with some modifications. Individuals that were sequenced in 2020 with the new platform (NovaSeq) were returned as multiple FASTA files which had to be merged using SAMtools v1.11 and the "-merge" flag (Li et al. 2009). Each merged FASTA file was inspected using FastQC v.0.11.8a (Andrews et al. 2010) and sequences were demultiplexed, cleaned and trimmed to 90 base pairs using 'process_radtags' in Stacks (Catchen et al. 2011; Rochette et al. 2019). Sequences were aligned in BWA v.0.7.15 (Li

and Durbin 2009) using the BWA-MEM algorithm (Li and Durbin 2009; Li 2013). Cuvier's sequences were aligned to the published Cuvier's reference genome (NCBI Genbank database accession: PRJNA399469) and Blainville's sequences were aligned to the Blainville's assembly generated by DNAZoo (https://www.dnazoo.org/assemblies/Mesoplodon_densirostris; Dudchenko *et al.*, 2017, 2018). Alignment files were converted to BAM files and sorted using SAMtools. Where individuals had been sequenced multiple times, the BAM files were also merged using SAMtools "–merge" (Li et al. 2009).

All BAM files were run through the 'gstacks' module in Stacks v.2.53 (Catchen et al. 2013; Rochette et al. 2019) to discover and genotype SNPs using the Bayesian genotype caller (BCG) algorithm of Maruki and Lynch (Maruki and Lynch 2015; Maruki and Lynch 2017; Rochette et al. 2019). The 'gstacks' parameters found to provide the greatest number of SNPs for Cuvier's and Blainville's in Chapter 3 (2.8.1) were used (mapq=10, sclip=0.2, var_alpha=0.05, gt_alpha=0.05) and population info was assigned *a priori* (Canary Islands, Bahamas or Mediterranean). The resulting genotype catalogue was analysed using the 'populations' program in Stacks with the same *a priori* population designations and a minor allele frequency (MAF) representing one allele per individual. This program applies a predetermined population framework to summarise and filter the dataset and convert it to the widely used VCF file format. The VCF files were filtered as described in Chapter 2 and following the recommendations of O'Leary *et al.* (2018) using VCFtools and R (Danecek et al. 2011; R Core Team 2019). The filtered and QC VCF files and a whitelist of good SNPs were input back into Stacks 'populations', with the *a priori* populations defined, a MAF filter representing one allele per individual, and only selecting one random SNP per locus. SNPs were kept if they were only found in one population, however only if the SNP was present in 80% of that population's individuals. The results from Chapter 3 indicate that each of the focal regions show significant genetic differentiation and for finer scale analyses it is better to allow for private alleles in regional populations. The final filtered genotypes were uploaded into R (R Core Team 2019) and converted to the widely used "genlight" format using the vcfR package (Knaus and Grünwald 2017). The "glPlot" function in *adegenet* was used to visually inspect the MAF of each locus and individual for excessive degrees of missing or inconsistent data, ensuring there were no library-specific artefacts.

5.2.4 Final Datasets and Geographic Sites

Samples were assigned to one of three focal regions: Bahamas, Canaries or Mediterranean and were subdivided within these regions into focal sites based on historic and contemporary mass strandings correlated with naval activities. Multiple SNP datasets were generated depending on the geographic scale of the analyses: species-specific (when comparing between focal regions) and region-specific (when comparing focal sites within focal regions). Region-specific datasets were generated by subsetting the species-specific datasets by sample region and removing any loci with missing data using the R package *dartr* (Gruber et al. 2018). East and West Mediterranean Cuvier's were split into two datasets prior to the calculations of relatedness in *Demerelate* since underlying population structure may lead to overestimation of relatedness (Kraemer and Gerlach 2017).

5.2.5 Calculating Genetic Composition using the four Genetic EBVs

5.2.5.1 EBV 1: Genetic Differentiation

To quantify the number of distinct genetic units to include in the analysis, a Discriminant Analysis of Principle Components (DAPC) was conducted in the R package, *adegenet* (Jombart 2008; Jombart et al. 2010). DAPC analysis proceeded as described in Chapter 2, incorporating the sampling regions informed by the findings of Chapter 3 (Bahamas, Canary Islands, East Mediterranean, or West Mediterranean) *a priori*. This function subset a training dataset, ran the analysis over a pre-determined number of repeats (n=30), and determined the best number of PCs to retain based on whichever yielded the highest predictive success of the training data with the lowest root mean squared error (RMSE). The resulting DAPCs were plotted to observe the spatial structure of SNP genotypes across both Cuvier's and Blainville's and the "assignplot" function was used to determine how well the assignment to DAPC clusters corresponded with the *a priori* population assignments.

The fact that there is no indication of movement between the east and west Canary Islands based on photo-identification (Reyes 2018) suggests that there could be underlying population structure which would inflate the relatedness estimates in *Demerelate* (Ayres 2000; Kraemer and Gerlach 2017). To investigate this, I ran an additional DAPC analysis using just individuals from the Canary Islands with the east or west focal site as population identifier.

5.2.5.2 EBV 2: Genetic Diversity

Measures of the two components of genetic diversity (richness and evenness) proposed by Hoban *et al.* (2022) were calculated for each focal region and site. To summarise richness, allelic richness was calculated in the R package *hierfstat* (Goudet 2005) using a rarefaction method that accounts for sample size (El Mousadik and Petit 1996). Evenness of genetic diversity was calculated using observed heterozygosity (H_o) and expected heterozygosity (H_s) in the R package *hierfstat*. The 95% confidence intervals were calculated using n=1000 bootstraps around the mean of each variable using the *rcompanion* R package (Mangiafico 2020).

5.2.5.3 EBV 3: Inbreeding and Relatedness Between Individuals

Inbreeding was estimated directly using inbreeding coefficients and the risk of inbreeding (Hoban *et al.*, 2022) was estimated by calculating relatedness and kinship within populations. The Weir and Cockerham inbreeding coefficient (F_{IS}, Weir and Cockerham, 1984) and the 95% confidence intervals around the mean were calculated for each focal region and site using *hierfstat* and n=1000 bootstraps in R (Goudet 2005). An additional measure of inbreeding, Internal Relatedness (IR), was calculated for each individual using the R package *Rhh* (Alho et al. 2010). This multilocus heterozygosity measure has been applied to other marine species, and is negatively correlated with reproductive success (Amos et al. 2001).

The R package *Demerelate* (Kraemer and Gerlach 2017) was used to calculate the pairwise relatedness of individuals within sites and estimate the likelihood of first and second order kin relationships (hereby referred to as full-sibling and half-sibling) within focal regions. The M_{xy} estimator of relatedness (Blouin et al. 1996) was used, which estimates the number of shared alleles between individuals with no previous knowledge of population allele frequencies required. This metric was selected for its suitability with

bi-allelic data and cetacean data, as shown by its use in other published cetacean studies (Brüniche-Olsen et al. 2018; O'Corry-Crowe et al. 2020; Morin et al. 2021). To estimate M_{xy}, *Demerelate* simulated a pre-determined number of pairs of unrelated individuals, half-siblings and full siblings based on the allele frequencies in the provided genotype data (n=1000 pairs for each category). Based on these, a threshold was calculated by *Demerelate* for each relatedness category based on the distribution of M_{xy} values. A chi-squared (χ^2) test was performed by *Demerelate* to determine whether the number of observed sibling pairs was significantly different from the number of expected sibling pairs in a group made up of randomized individuals of the same size and based on allele frequencies from the empirical data.

Due to large differences in allele frequencies between focal regions and the fact that *Demerelate* will not run if there are any missing data or monomorphic sites, individuals from the focal regions were analysed separately. The developers of *Demerelate* also note that any underlying population structure may lead to overestimation of relatedness and suggest that populations be split prior to analysis if any underlying structure is suspected. For this reason, I split the east and west Mediterranean regions as Chapter 3 showed significant structure between these two populations. Region specific files for the Canary Islands, Bahamas, east Mediterranean and west Mediterranean were generated by removing all missing and monomorphic loci, and were then converted to *Demerelate* format using the R-package *dartr* (Gruber et al. 2018). *Demerelate* was run without assigning a reference population and using 1000 simulated pairs.

Since there is no evidence of movement between the east and west Canary Islands (Reyes 2018), I recalculated the estimates of M_{xy} within the west Canary populations of Cuvier's and Blainville's (sample size was too low to calculate the east populations) in case underlying population structure between the two sites could be inflating M_{xy} relatedness estimates.

The proportion of related individuals (combined half and full siblings) within each focal region and based on region-specific M_{xy} thresholds, was compared using a 'comparison of proportions' test implemented in 'radiant' (Nijs 2022). A two-tailed test with Bonferroni correction was used to test the null hypothesis that the differences in proportion of related individuals between each region was zero. Pairwise-relatedness (M_{xy}) and patterns of sibship within focal regions were visualised using the network analysis R package, *igraph* (Csárdi and Nepusz 2006). In each graph, individuals were plotted as dots ("nodes") with lines drawn between individuals ("edges") that represent kinship. As M_{xy} relatedness was calculated for all pairs of individuals, only lines that connected proposed half and full siblings were kept. Within these plots, individual nodes were coloured according to the finer-scale focal site to look for any connections between areas that were previously considered to be separate based on photo-identification.

5.2.5.4 EBV 4: Effective Population Size (N_e)

Contemporary effective population size (N_e) was calculated for each focal site and region with NeEstimator V2.1 (Do et al. 2014). N_e was calculated using the linkage disequilibrium method with bias correction and using a random subset of 10,000 SNPs from each of the region-specific datasets (Waples 2006). The 95% confidence intervals were calculated using a non-parametric method that jack-knives over individuals and is recommended when the number of loci used is >100 (Jones et al. 2016). A minor allele threshold of P_{crit}=0.02 was applied.

5.3 RESULTS

5.3.1 Sequencing and Sample selection

The 2020 sequencing run generated 128.7gb of raw sequencing data. Combining data from the 2018 and 2020 sequencing runs, ddRAD sequences were available from *n*=44 Blainville's and *n*=113 Cuvier's (Table 5-1) with the highest representation of samples from the Canary Islands and Bahamas (Table 5-1).

5.3.2 Optimisation

The results of the optimisation protocol are presented in 2.8.3. In summary, there was good consistency with very little site discordance between the SNPs genotyped in duplicate individuals sequenced across different platforms, in different labs and in different years. Of the SNP loci that were based on sequences generated in both sequencing runs from 2018 and 2020, individual site discordance ranged from 0.12-0.90% (0.33% mean discordance) in Blainville's and 0.13-37.04% (3.47% mean discordance) in the Cuvier's. Of note, two individuals in the Cuvier's dataset had ~35% discordance, and without these individuals the mean discordance was 0.33%. Overall, site discordance was low per site in both datasets (Blainville's: 0.35%, Cuvier's: 3.60%).

5.3.3 Final SNP Datasets

SNP loci were recovered from both Cuvier's and Blainville's datasets, with little missing data and high site read depth (Table 5-2). A breakdown of the number of sites and private alleles for each focal region are also presented in Table 5-2. The proportion of missing data was higher in this analysis compared to Chapters 3 and 4, as the population of origin was incorporated *a priori* in both the Stacks 'gstacks' and 'populations' runs, which allowed for SNPs to be genotyped even if they are only included in one population (if >80% individuals within that population have the locus). Substantial genetic structure found in Chapters 3 and 4 indicated that private alleles within populations are likely to occur and there were regional differences in allele frequencies with many private alleles. The final species-specific datasets consisted of *n*=27 Blainville's (*n*=47127 SNPs, 5.6% missing) and *n*=68 Cuvier's (*n*=140088 SNPs, 11.2% missing). The datasets used to calculate relatedness and kinship in *Demerelate* are in Table 5-3.

Table 5-1. Summary of sequence data available from the three focal regions (Canary Islands, Bahamas and Mediterranean) and the year(s) in which they were sequenced (2018, 2020) or whether they were sequenced in both 2018 and 2020 (2018 + 2020). Individuals sequenced from nearby areas that were not included in the final analyses but are included in the counts below: individuals from Madeira were included in the "Canary Islands" region and individuals from Florida, Puerto Rico and the Virgin Islands were included in the "Bahamas" region.

	Sequencing Run	Canary Islands	Bahamas	Mediterranean	Total
Blainville's	Total	30	14	-	44
	2018	1	1	-	2
	2020	12	1	-	13
	2018 + 2020	17	12	-	29
Cuvier's	Total	44	25	44	113
	2018	4	3	5	12
	2020	18	4	8	30
	2018 + 2020	22	18	31	71
Total		74	39	44	157

Table 5-2. Final sample size and SNP loci for Blainville's and Cuvier's that passed QC and filtering steps in each of the three focal regions. SNPs were kept if they passed filtering, appeared in at least one region, and were present in >80% of individuals within that region. MAF was filtered to allow for at least one individual per population to have a variable site. Sampling region was provided in the Stacks 'populations' program, so region-specific sites and private alleles are also provided.

	Blainville's	**Cuvier's**
Final_N	27	68
Final_SNP_Loci	47127	140088
Mean_SNP_Loci	44511 (SD=1690)	124360 (SD=4794)
Mean_Missing_Loci	5.6% (SD=3.6%)	11.2% (SD=3.4%)
Mean_Site_Depth	118x (SD=68x)	119x (SD=68x)
Bahamas_Polymorphic_Sites (private alleles)	32784 (2668)	84755 (9676)
Canaries_Polymorphic_Sites (private alleles)	43874 (11313)	111435 (32509)
Mediterranean_Polymorphic_Sites (private alleles)	NA	57498 (10826)

5.3.4 Genetic EBV Results

5.3.4.1 EBV 1: Genetic Differentiation

Using the full Blainville's SNP dataset (*n*=47127 SNPs), the first 12 PCs and one discriminant function were used to resolve two distinct Blainville's clusters consistent with the structure patterns found in Chapters 3 and 4 (Figure 5-2a). Individuals from each sampling region were also placed into the correct genetic cluster as defined by the DAPC (Canary Islands and Bahamas, Figure 5-2b). Using the full Cuvier's SNP dataset (*n*=140088), the first 10 PCs and three discriminant functions were used to resolve four distinct Cuvier's clusters consistent with previous analyses: Canary Islands, Bahamas, West, and East Mediterranean (Figure 5-3a). All individuals from each sampling region were placed into the correct cluster (Figure 5-3b). Axis 1 discriminates the two Mediterranean sites from the rest of the focal regions. Axis 2 discriminates between the two Mediterranean sites (Figure 5-3b).

Results from the Canary Island focal site DAPC analysis is presented in Figure 5-4. Two Cuvier's that were sampled in the west clustered within the east individuals and one east individual clustered with the west (Figure 5-4a). The three Blainville's that were sampled in the east clustered with the west samples (Figure 5-4b).

5.3.4.2 EBV 2: Genetic Diversity

Diversity statistics are provided for each focal region (Table 5-4) and focal site (Table 5-5). Using the full SNP dataset (*n*=47127 SNPs), genetic diversity metrics (H_o, H_s and AR) of Blainville's were significantly higher in the Canary Islands (no overlap of 95% confidence intervals, Table 5-4) compared to the Bahamas. Within the Bahamas, significantly lower diversity (H_s and AR) was found in the "disturbed" TOTO population (Table 5-5).

In Cuvier's, diversity was lowest in the Mediterranean (H_o, H_s, AR), and H_o and H_s differed significantly between the Bahamas and Canary Islands (Table 5-4). The more "disturbed" east Canary Islands had significantly lower diversity (H_o and H_s) than the "semi-pristine" site. In the Bahamas, AR was significantly lower in the "disturbed" TOTO population. All measures of diversity (H_o, H_s and AR) were significantly lower in the "disturbed" east Mediterranean population compared to the west, regardless of the SNP dataset used (Table 5-4 and Table 5-5).

5.3.4.3 EBV 3: Inbreeding and Relatedness Between Individuals

Using the full Blainville's SNP dataset (n=47127 SNPs), F_{IS} was significantly higher in the Canary Islands (no overlap of 95% confidence intervals, Table 5-4) compared to the Bahamas. Within these regions, both measures of inbreeding were near zero and overlapped between the Canary Islands and Bahamas (Table 5-5). The only significant difference within regions was in the Bahamas (lower F_{IS} in TOTO, Table 5-5).

Using the full Cuvier's SNP dataset (n=140088 SNPs), F_{IS} was significantly lower in the Mediterranean compared to the Canary Islands and Bahamas and IR was significantly higher (Table 5-4). F_{IS} and IR were significantly lower in the Canary Islands (Table 5-4) compared to the Bahamas. The more disturbed east Canary Islands and east Mediterranean sites had significantly higher levels of inbreeding (F_{IS} and IR) compared with their semi-pristine comparators. In comparison, there was no significant difference between the two Bahamas populations (Table 5-5).

Across all focal regions in Cuvier's and Blainville's, mean M_{xy} pairwise relatedness was higher than the simulated threshold for unrelated individuals (Figure 5-5 and Figure 5-6, Table 5-3). Pairwise M_{xy} in Blainville's was lower in the Bahamas compared to the Canary Islands (Figure 5-5, Table 5-3). M_{xy} was also highest in the Canary Islands Cuvier's, with similarly lower M_{xy} values for the Bahamas, East and West Mediterranean regions (Figure 5-6, Table 5-3).

Networks generated using the M_{xy} relatedness estimates calculated for each focal region are presented in Figure 5-7 and Figure 5-8. A summary of the detected sibling pairs is presented in Table 5-3.

In the Blainville's, one full-sibling pair was found in each of the Bahamas and Canary Island regions, and half-sibling pairs were detected in the Canary Islands (Figure 5-7, Table 5-3). Neither of the full-sibling pairs spanned focal sites in either species, however there were some half-sibling pairs between the east and west Canary Islands (Figure 5-7). When M_{xy} was recalculated for the individual sites, one new sibling pair was estimated and three of the original sibling pairs were not.

In the Cuvier's, no sibling pairs were detected in the Bahamas (Figure 5-8). Four full sibling pairs were found in the Canary Islands, two of which were in the west and two which included one east and two west Canary Island individuals. A further 21 half-sibling pairs were found in the Canary Islands, 7 of which linked the east and west through the same east Canary Island individual as above (Figure 5-8, Table 5-3). One full sibling pair was detected between two Ligurian Sea individuals in the West Mediterranean and 15 half-sibling pairs were also detected (Figure 5-8, Table 5-3). No full siblings were detected in the east Mediterranean, though 17 half-sibling pairs were found spanning the entire east Mediterranean basin (Figure 5-8, Table 5-3). The recalculated M_{xy} values in the west Canary Islands Cuvier's resulted in the same sibling pairs as the original analysis.

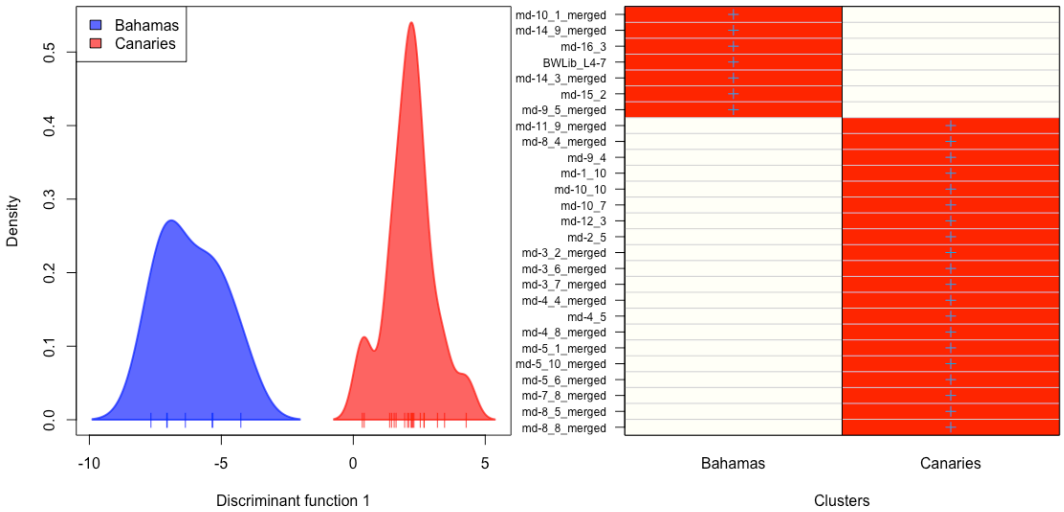


Figure 5-2. Density plot for the first discriminant function of two Blainville's regions (left; Canary Islands and Bahamas) generated using *n*=53862 SNPs (a). Assignment plot showing the accuracy of the DAPC genetic cluster assignment compared to the sampling region (b). The "+" symbol indicates the *a priori* population and the red colour indicates the population assigned based on the DAPC analysis.

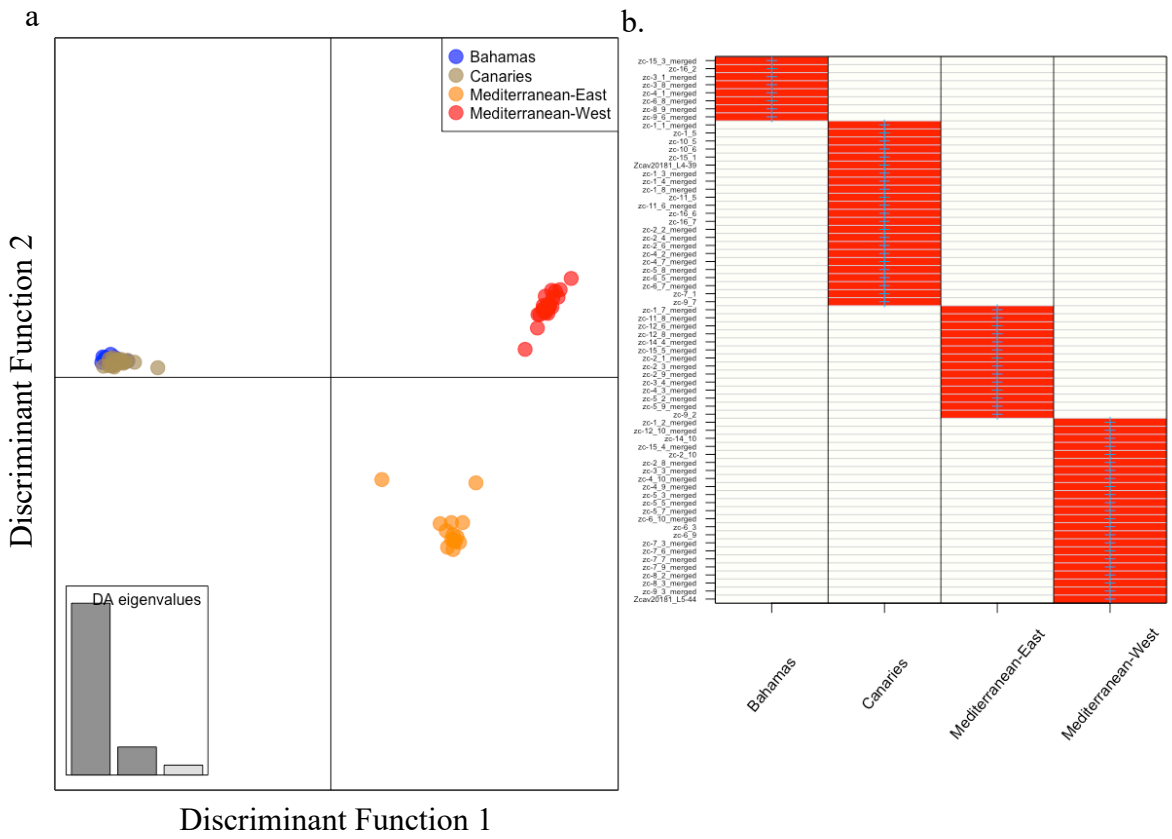
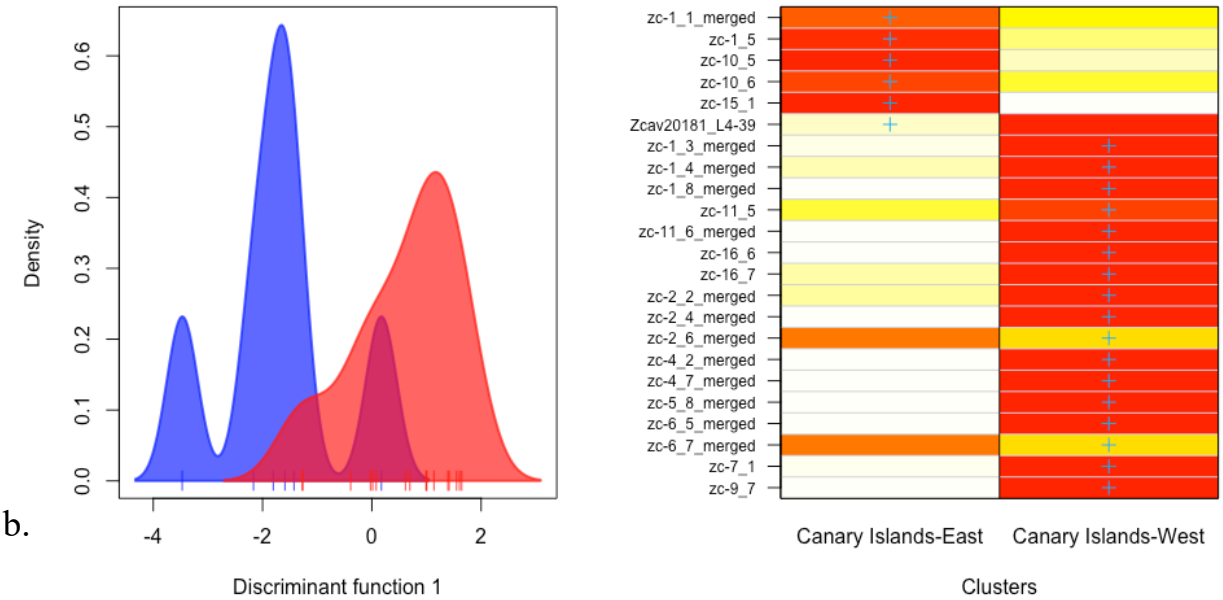


Figure 5-3. First and second discriminant functions of the DAPC scatter plot for Cuvier's regions (left; Bahamas, Canary Islands, West Mediterranean, and East Mediterranean) generated using *n*=149616 SNPs (a). Assignment plot showing the accuracy of the DAPCC genetic cluster assignment compared to the sampling region (b). The "+" symbol indicates the a priori population given to the analysis and the red colour indicates the population assigned based on the DAPC analysis.

a.



b.

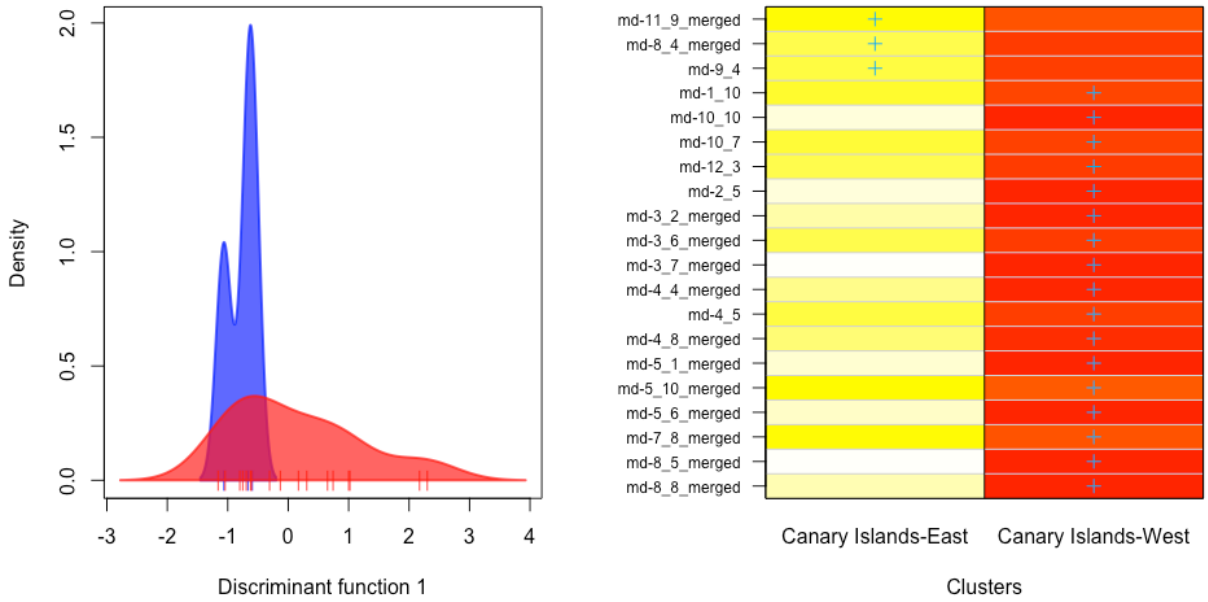


Figure 5-4. Density plots for the first discriminant function and assignment plots for Cuvier's (a) and Blainville's (b) in the two Canary Island focal sites. In both plots, the east individuals are presented in blue and the west individuals are presented in red. Assignment plots show the accuracy of the DAPC genetic cluster assignment compared to the sampling region (b). The "+" symbol indicates the *a priori* population and the red colour indicates the population assigned based on the DAPC analysis.

Table 5-3. Inputs and results of *Demerelate* analysis of relatedness using the M_{xy} estimator in each focal region. Inputs shown include the sample size and loci used for region-specific calculations. Outputs shown include the relatedness thresholds for sibship, number of estimated sibling pairs and results of the Chi-square analysis to determine if more siblings (combined full and half sibling pairs) were detected compared to the expected number based on 1000 simulated pairs.

	Cuvier's				**Blainville's**	
Focal Site	Canary Islands	Bahamas	E. Mediterranean	W. Mediterranean	Canary Islands	Bahamas
***n*, Individuals**	23	8	14	23	20	7
***n*, Loci**	56246	61643	26269	34741	28911	26890
***n*, Simulated pairs**	1000	1000	1000	1000	1000	1000
mean M_{xy}	0.840	0.756	0.752	0.755	0.776	0.718
M_{xy} threshold, Half/Full Siblings	0.852/0.899	0.787/0.856	0.764/0.841	0.759/0.839	0.784/0.855	0.744/0.829
***n*, Half/Full/Unrelated pairs**	21/4/228	0/0/28	17/0/77	15/1/237	17/1/172	0/1/20
Observed/Expected Freq. Siblings	0.099/0	0/0	0.154/0	0.063/0	0.095/0	0.048/0
Chi2 statistic (*p<0.05)	24.238*	NA	13.077*	14.522*	16.854*	0

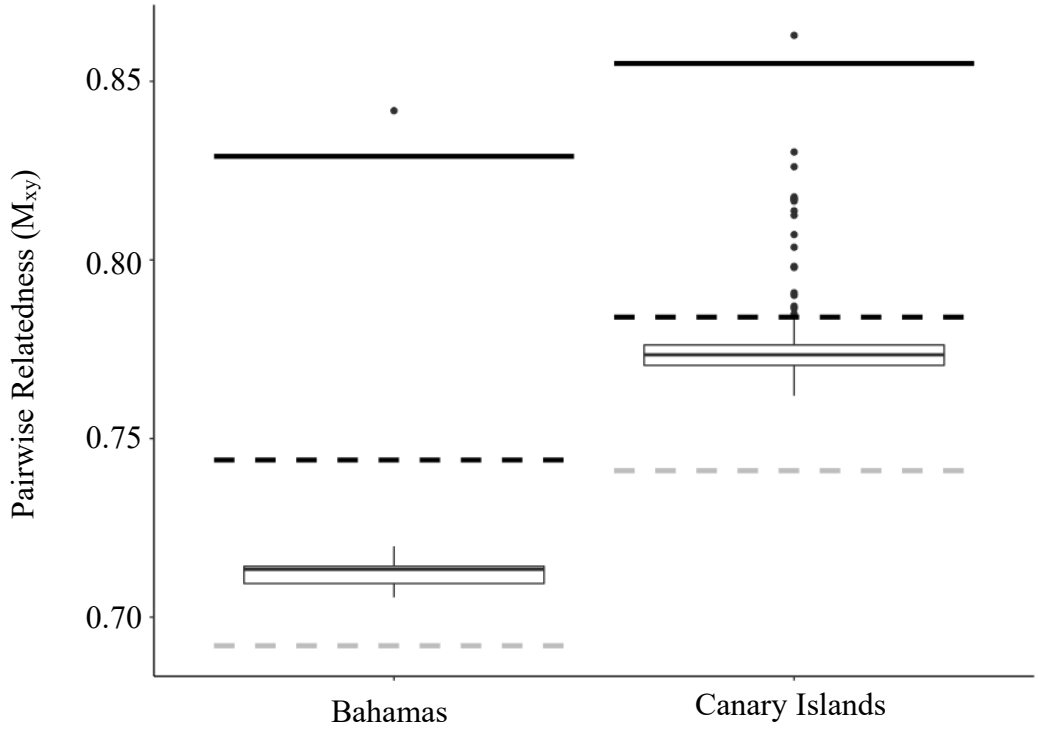


Figure 5-5. Pairwise relatedness (M_{xy}) and sibship thresholds (unrelated: light-grey dashed line, half-sibling: black dashed line, full sibling: black line) for Blainville's pairs using *n*=26890 loci (Bahamas) and *n*=28911 loci (Canary Islands). Relatedness thresholds were calculated based on the allele frequencies of *n*=1000 simulated pairs. The mean relatedness presented in Table 5-3.

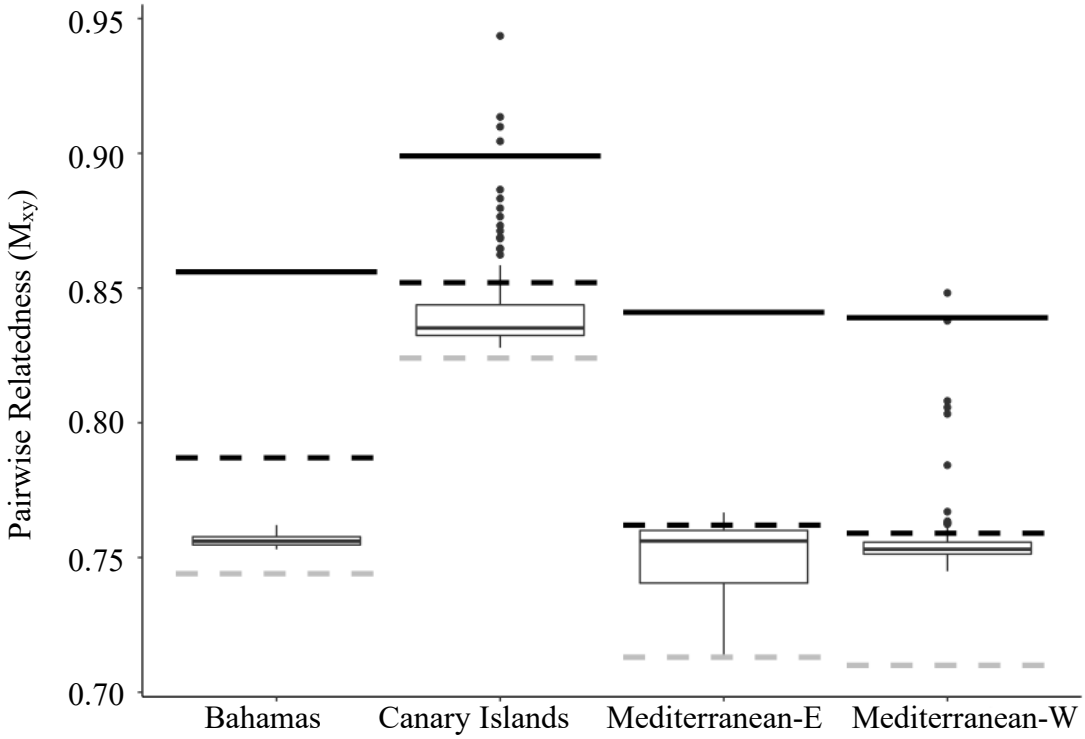


Figure 5-6. Pairwise relatedness (M_{xy}) and sibship thresholds (unrelated: light-grey dashed line, half-sibling: black dashed line, full sibling: black line) for Cuvier's pairs using *n*=61643 loci (Bahamas), *n*=56246 loci (Canary Islands), *n*=26269 loci (East Mediterranean) and *n*=34741 loci (West Mediterranean). Relatedness thresholds were calculated based on the allele frequencies of *n*=1000 simulated pairs. The mean relatedness is presented in Table 5-3.

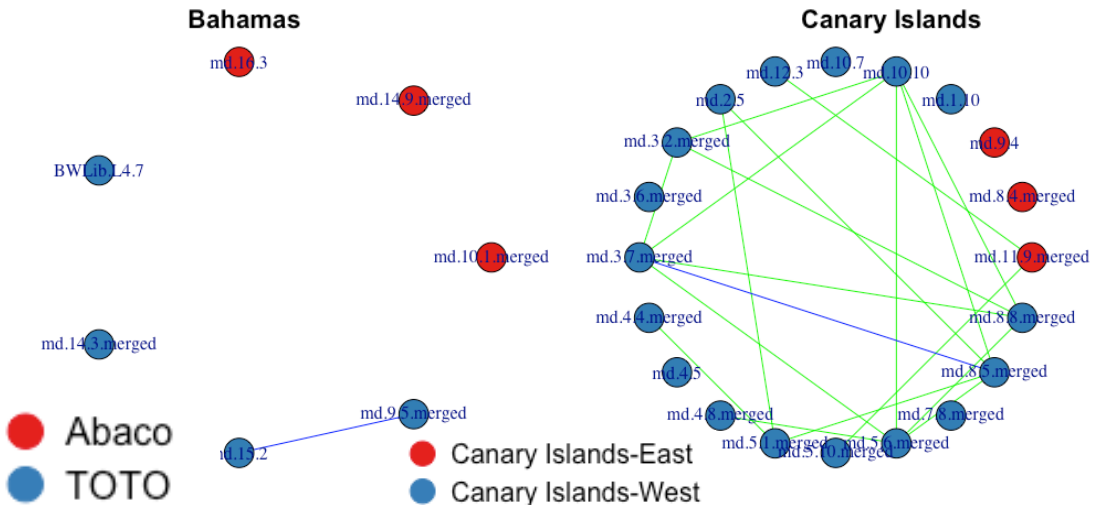
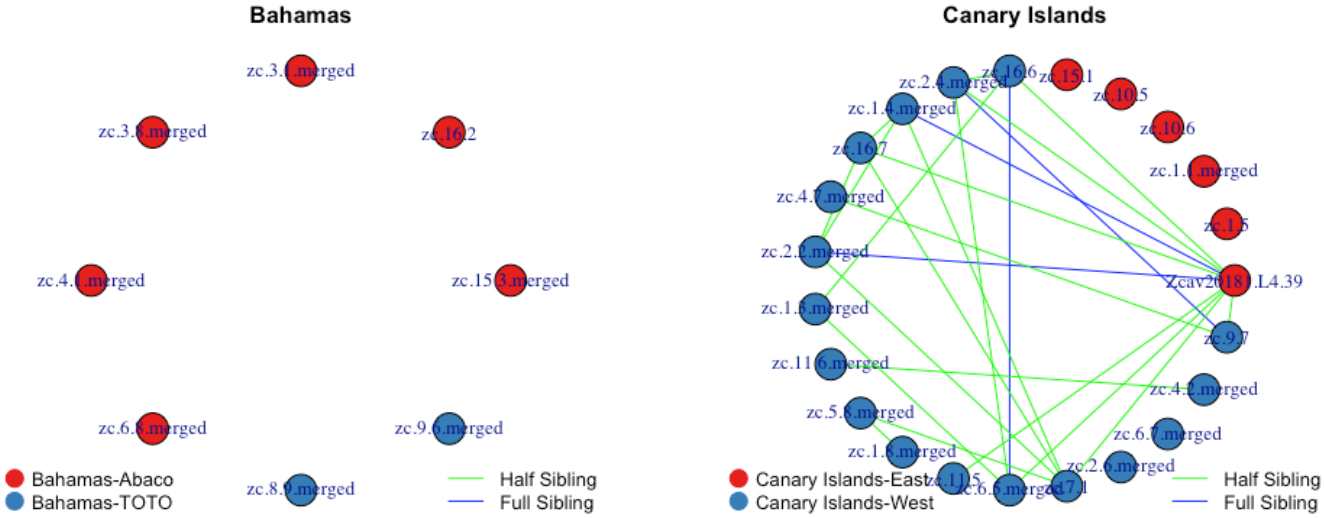


Figure 5-7. Sibling networks for Blainville's in the Bahamas (*n*=26890 loci) and Canary Islands (*n*=28911 loci). The dot colour represents the sampling location and line colour represents the estimated relationship based on the M_{xy} estimate calculated in *Demerelate*.



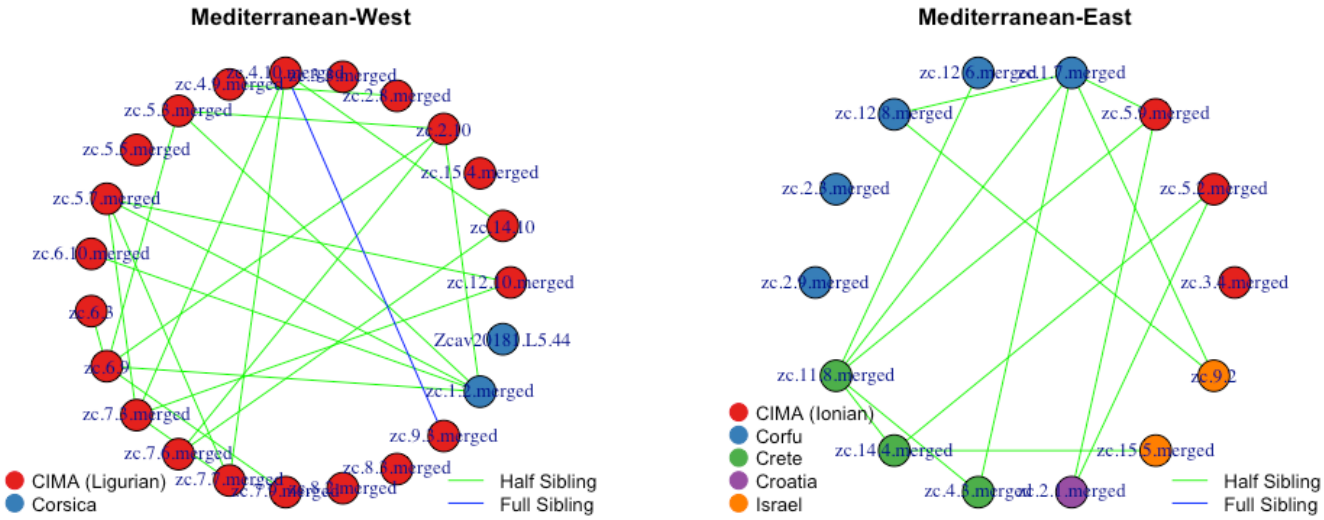


Figure 5-8. Sibling networks for Cuvier's in the Bahamas (*n*=61643 loci), Canary Islands (*n*=56246 loci) West Mediterranean (*n*=34741 loci) and East Mediterranean (*n*=26269 loci). The dot colour represents the sampling location and colour of the line represents the estimated relationship based on the M_{xy} estimate calculated in *Demerelate*.

Table 5-4. Diversity and inbreeding statistics and their 95% confidence intervals calculated for each focal region using species specific datasets. Observed heterozygosity (H_o), expected heterozygosity (H_s), Weir and Cockerham's inbreeding coefficient (F_{IS}) and Allelic Richness (AR) were calculated using the *hierfstat* package in R. Internal Relatedness (IR) was calculated using the *Rhh* package in R and effective population size (N_e) was calculated using NeEstimator V2.1 using a random subset 10,000 SNPs.

Species	Focal Region	n	Loci (% missing)	H_o (95% CI)	H_s (95% CI)	AR (95% CI)	F_{IS} (95% CI)	IR (95% CI)	N_e (95%)
Blainville's	Bahamas	7	47127 (5.6%)	0.232 (0.230,0.234)	0.251 (0.249,0.253)	1.68 (1.68,1.69)	0.069 (0.066,0.072)	0.040 (0.025,0.059)	33 (7,Inf)
	Canary Islands	20	47127 (5.6%)	0.237 (0.235,0.239)	0.254 (0.253,0.256)	1.70 (1.70,1.71)	0.075 (0.070,0.080)	0.012 (-0.003,0.029)	86 (44.4, 533.2)
Cuvier's	Bahamas	8	140088 (11.2%)	0.152 (0.151,0.153)	0.180 (0.178,0.180)	1.48 (1.48,1.48)	0.155 (0.152,0.158)	0.144 (0.130,0.159)	Inf (Inf,Inf)
	Canary Islands	23	140088 (11.2%)	0.155 (0.155,0.156)	0.176 (0.175,0.177)	1.48 (1.48,1.48)	0.118 (0.116,0.120)	0.107 (0.083,0.127)	62.8 (33.2,266.3)
	E. Mediterranean	14	140088 (11.2%)	0.106 (0.105,0.107)	0.114 (0.113,0.115)	1.28 (1.28,1.28)	0.071 (0.068,0.074)	0.206 (0.198,0.214)	108.3 (51.8,Inf)
	W. Mediterranean	23	140088 (11.2%)	0.118 (0.117,0.119)	0.128 (0.127,0.129)	1.31 (1.31,1.31)	0.074 (0.071,0.077)	0.279 (0.248,0.302)	132.9 (61.9,Inf)

Table 5-5. Essential Biodiversity Variables (EBVs) 2 (genetic diversity), EBV3 (inbreeding) and EBV4 (effective population size) and their 95% confidence intervals calculated for each fine-scale focal site using region-specific datasets with no missing data. Sites that are categorised as "semi-pristine" are presented in bold; others are "disturbed". Observed heterozygosity (H_o), expected heterozygosity (H_s), Weir and Cockerham's inbreeding coefficient (F_{IS}) and Allelic Richness (AR) were calculated using the *hierfstat* package in R. Internal Relatedness (IR) was calculated using the *Rhh* package in R and effective population size (N_e) was calculated using NeEstimator V2.1 and using a random subset of 10,000 SNPs. Site abbreviations: Tongue of the Ocean (TOTO).

Species	**Focal Site**	***n***	**Loci**	**EBV2**			**EBV3**		**EBV4**
				H_o (95% CI)	H_s (95% CI)	AR (95%)	F_{IS} (95% CI)	IR (95% CI)	N_e (95%)
Blainville's	**Canary Islands-West**	17	28911	0.255 (0.253,0.257)	0.265 (0.263,0.267)	1.60 (1.60, 1.60)	0.039 (0.035,0.043)	-0.037 (-0.056,-0.015)	68.3 (33.4,779.4)
	Canary Islands-East	3	28911	0.256 (0.253,0.259)	0.269 (0.266,0.272)	1.60 (1.60,1.61)	0.049 (0.040,0.057)	-0.032 (-0.038,-0.023)	Inf (Inf,Inf)
	Bahamas-Abaco	3	26890	0.313 (0.310,0.316)	0.340 (0.337,0.343)	1.75 (1.75,1.76)	0.079 (0.071,0.088)	-0.113 (-0.148,-0.060)	Inf (Inf,Inf)
	Bahamas-TOTO	4	26890	0.312 (0.309,0.315)	0.321 (0.318,0.323)	1.72 (1.71,1.72)	0.029 (0.022,0.036)	-0.147 (-0.174,-0.119)	Inf (36.7,Inf)
Cuvier's	**Canary Islands-West**	17	56246	0.169 (0.168,0.170)	0.181 (0.180,0.182)	1.63 (1.62,1.63)	0.065 (0.062,0.068)	-0.046 (-0.083,-0.019)	57.1 (28.9,376.2)
	Canary Islands-East	6	56246	0.158 (0.157,0.160)	0.179 (0.177,0.180)	1.63 (1.62,1.63)	0.114 (0.109,0.119)	0.019 (-0.012,0.070)	-21.9 (Inf,Inf)
	Bahamas-Abaco	6	61643	0.241 (0.239,0.242)	0.275 (0.273,0.276)	1.50 (1.50,1.51)	0.123 (0.119,0.127)	-0.092 (-0.112,-0.072)	-23.1 (Inf,Inf)
	Bahamas-TOTO	2	61643	0.243 (0.241,0.246)	0.276 (0.273,0.278)	1.49 (1.48,1.49)	0.117 (0.110,0.125)	-0.099 (-0.103,-0.094)	-1 (Inf,Inf)
	E. Mediterranean	14	27087	0.222 (0.219,0.224)	0.229 (0.227,0.231)	1.77 (1.76,1.77)	0.032 (0.028,0.036)	0.097 (0.063,0.121)	108.3 (51.8,Inf)
	W. Mediterranean	23	27087	0.259 (0.257,0.261)	0.266 (0.264,0.268)	1.88 (1.87,1.88)	0.027 (0.024,0.031)	-0.004 (-0.012,0.004)	132.9 (61.9,Inf)

As explained above (5.2.5.3), there were too few sibling pairs to conduct the χ^2 and comparison of proportions tests within focal sites, so they were done between focal regions. Combining the counts of full and half siblings, the χ^2 test in *Demerelate* indicated that there were more siblings than expected in all regions except the Bahamas in both species (Table 5-3). A comparison of proportions test between regions (within species) determined that there was no significant ($p<0.05$) difference in the proportion of related individuals between each of the regions, though the difference in proportions between the East and West Mediterranean (15.4% and 8.7% related pairs, respectively) was nearly significant ($p=0.067$) (Table 5-6).

Table 5-6. Comparison of proportions undertaken in Radiant to test if the proportion of related individuals (combined half and full siblings) differed between focal regions. For each null hypothesis, the two pairs being tested are presented with the percent of related pairs, number of pairs, difference between the percent of related pairs, χ^2 value and associated p-value. As all p-values are >0.05, none of the null hypotheses can be rejected.

Species	Null Hypothesis (% related, *n* pairs)	% Difference	χ^2 Value	p value
Blainville's	Bahamas (4.8%, 21) = Canary Islands (10%, 190)	-0.050	0.605	0.710
Cuvier's	Bahamas (0%, 28) = Canary Islands (9.9%, 253)	-0.099	3.037	0.897
	Bahamas (0%, 28) = E. Med. (15.4%, 91)	-0.154	4.882	0.201
	Bahamas (0%, 28) = W. Med. (8.7%, 253)	-0.087	2.642	0.819
	Canary Islands (9.9%, 253) = E. Med. (15.4%, 91)	-0.055	2.016	0.934
	Canary Islands (9.9%, 253) = W. Med. (8.7%, 253)	0.012	0.211	1.000
	E. Med. (15.4%, 91) = W. Med. (8.7%, 253)	0.067	3.196	0.067

5.3.4.4 EBV 4: Effective Population Size (N_e)

The estimates of effective population size (N_e) and 95% confidence intervals are provided for each focal region (Table 5-4) and focal site (Table 5-5). In Cuvier's from the Bahamas, the sample size was insufficient to calculate N_e and the upper limit of the confidence intervals for both Mediterranean sites was infinity (Table 5-4). The point estimate of N_e and the lower confidence interval limit were lower in the east Mediterranean than the west. The Blainville's in the Bahamas also had infinity for the upper confidence interval value, but both the point estimate and lower confidence interval value were lower than the Canary Islands N_e (Table 5-4).

Within focal regions, there were sufficient sample sizes to estimate N_e and 95% confidence intervals in the West Canary Islands (both species). All remaining populations had infinity or negative numbers for the point estimate and/or confidence intervals (Table 5-5). Both the point estimate and confidence interval of N_e was greater in the west Canary Island Blainville's compared to the Cuvier's, and the N_e estimate of West Mediterranean Cuvier's was greater than both (Table 5-5).

5.3.5 Importance of the dataset used

Two SNP datasets were used to conduct the analyses above. For investigations that compared the focal regions (DAPC, region-specific diversity (H_o, H_s, AR), inbreeding (F_{IS}, IR) and N_e), the full dataset with n=140088 (11.2% missing) and n=47127 loci (5.6% missing data; Cuvier's and Blainville's, respectively) was used. For investigations within focal regions (sibling analyses (M_{xy}), focal-site specific diversity (H_o, H_s, AR), inbreeding (F_{IS}, IR) and N_e), region-specific datasets with no missing data were used (Table 5-5).

The results using these two datasets varied, showing how important it is to be clear which dataset was used when drawing conclusions as it will not always be appropriate to use them both. For example, when the full dataset was used, IR was significantly higher in the two Mediterranean Cuvier's populations compared to the Bahamas and Canary Islands (no overlap of 95% CI; Table 5-4), and IR was also substantially higher than zero. When the region-specific datasets were used, IR in these two regions was less than zero and overlapped with most other sites (Table 5-5). This is likely due to the underlying population structure within Cuvier's, which resulted in fewer polymorphic sites in the Mediterranean (Table 5-2) and likely more missing data.

5.4 DISCUSSION

This study provides five separate comparisons of "disturbed" and "semi-pristine" beaked whale populations (Cuvier's: Canary Islands, Bahamas, and Mediterranean; Blainville's: Canary Islands and Bahamas), with the results indicating that four of the "disturbed" sites have significantly reduced genetic variation (lower genetic diversity, higher inbreeding, more sibling relationships) in at least one EBV. Here I have presented the first evidence of potential population-level impacts of naval sonar exposure on the genetic variation of beaked whale populations. I have demonstrated that, while causality cannot be directly inferred, some populations existing in areas regularly exposed to naval sonar may be more susceptible to deleterious interactions that those in more pristine habitats based on the four genetic EBVs proposed to monitor and compare the genetic health and viability of wild populations (Hoban et al. 2012).

5.4.1 Genetic differentiation

Though present throughout all temperate and warm waters of the world (except the Mediterranean in Blainville's), populations of Cuvier's and Blainville's appear to form resident, local populations that are genetically distinct (Chapter 3). In this chapter, I reinforced the genetic distinctiveness of the focal regions from this chapter as originally measured in Chapter 3, especially highlighting the differentiation between the Mediterranean and North Atlantic Cuvier's, and between the Cuvier's populations within the Mediterranean. The DAPC analysis presented here does not merely reiterate the findings from Chapter 3, it also proves the advertised reproducibility of the ddRAD methodology (Peterson et al. 2012). The population structure analyses used here were based on SNP datasets from sequences generated in different years and using different sequencing platforms, and the resulting DAPC scatter plots showed clear differentiation between each of the regions, as seen in Chapter 3. This differentiation is also evident in the variable number of polymorphic loci between regions and presence of private alleles (Table 5-2).

5.4.2 Comparisons of diversity, inbreeding and N_e between sites

To provide baseline values for genetic EBVs, investigate the potential long-term impacts of repeated anthropogenic disturbance, and address the potential for future risk from perturbations, measures of genetic diversity, inbreeding, relatedness, and kinship were calculated and compared between areas that are considered more or less impacted by anthropogenic disturbance ("disturbed" or "semi-pristine").

5.4.2.1 Bahamas

The Bahamas had the smallest sample size in both species (*n*=7 Blainville's and *n*=8 Cuvier's) and there was only one sibling pair detected in either species (Cuvier's). This was a mother/calf pair that live stranded at the AUTEC naval base (TOTO) and was mis-identified as two adults. The mother died before rescue attempts by Navy staff could commence and the calf was eventually re-floated, though it was found dead the next day (Diane Claridge, 2022. Personal Communication, 5 August).

For Cuvier's in the Bahamas, diversity and inbreeding stats were not significantly different between the disturbed and semi-pristine sites, except AR which was slightly lower in TOTO compared to the Abaco population. In Blainville's, H_s, AR and F_{IS} were significantly lower in TOTO. Estimates of N_e in the Bahamas are difficult to interpret, as sample size limitations resulted in point values and confidence intervals that included infinity.

5.4.2.2 Canary Islands

Most inferred sibling pairs were between individuals in the west Canary Islands, however one Blainville's from the east Canary Islands was estimated to be half-siblings with two El Hierro individuals. In Cuvier's, two full-sibling relationships were found between an individual that stranded in Fuerteventura and two El Hierro individuals. This same Fuerteventura animal also formed several half-sibling pairs with El Hierro animals. The original sibling relationships were estimated for combined east and west Canary Island individuals to find out if there was any connectivity between the two sites. In both species, one individual from the east was connected via half or full sibling relationships to the west Canary Islands, however the Canary Island DAPC (Figure 5-7) shows that these individuals clustered amongst samples from the west Canary Islands. The east Cuvier's individual was found dead stranded and covered with wounds, and was thought to have been dead about a week (Europa Press 2013). The east Blainville's individual with sibling ties to the west Canaries was biopsy sampled while alive. These findings support the photo-identification evidence that there's little or no movement between the east and west Canary Islands. It is interesting to note, that there are no detected sibling pairs between any individuals in the east Canary Islands, of either species.

A detailed analysis of the Blainville's and Cuvier's photo-identification data collected off El Hierro was done by the University of La Laguna, including social structure analysis, temporal analysis, and calculations of an individual-individual gregarious index, and showed that there was some preference in associations by individuals (Raquel Viñé Afonso, 2022, Personal Communication, 2 August). A study combining the genetic and social associations is planned.

Diversity and inbreeding statistics show that the east Canary Cuvier's have lower diversity and higher inbreeding statistics compared to the west Canary Islands, despite showing fewer inferred sibling relationships. There is no difference between these values in the Blainville's, although the east Canary Islands sample size was very small (*n*=3). Estimates of N_e are difficult to compare between sites in either species as the point estimates and confidence intervals in eastern sites included infinity however estimates were greater in Blainville's than Cuvier's.

5.4.2.3 Mediterranean

No sibling pairs were found between the east and west Mediterranean which is consistent with the substantial genetic structure measured in this basin (Chapter 3). Compared to the relatively small geographic distribution of samples in the west Mediterranean, the samples from the East covered a much larger area from the Ionian Sea to Israel. Inferred sibling pairs were detected between individuals spanning this distance (Israel and Corfu and Crete) and on a much smaller scale (siblings detected within samples from the 2011 Corfu and 2014 Crete mass stranding events). In the Ligurian Sea, photo-identification studies are carried out by the CIMA Research Foundation and analyses of association between identified individuals are underway with an aim to compare the association rate amongst individuals that I have identified here as being related (Massimiliano Rosso, 2022. Personal Communication, 30 August). In the Ionian Sea, photos are regularly taken during encounters with live Cuvier's by the Pelagos Cetacean Research Institute, and analyses are planned to investigate association rates (Alexandros Frantzis, 2022. Personal Communication, 29 August).

Within the Mediterranean, genetic structure was substantial and the east and west populations needed to be split prior to conducting the relatedness analysis in *Demerelate*; when east and west Mediterranean individuals were pooled and analysed together, all east individuals were classified as half-siblings. This is likely because underlying population structure may lead to overestimation of relatedness by *Demerelate* (Kraemer and Gerlach 2017). Of note, when M_{xy} was calculated for all Mediterranean individuals together, no sibling pairs were detected between the east and west. Since the east and west populations were split prior to estimating M_{xy} relatedness, and there were sufficient sample sizes in both populations, statistical tests were run for each area to compare the frequency of sibling pairs. Both sites had significantly more sibling pairs than expected, and the east had a higher (though not statistically significant, p=0.067) proportion of sibling pairs compared to the west. Genetic diversity (H_o, H_s and AR) was lowest and both inbreeding statistics were highest in the east, the more disturbed focal site.

Population size estimates in the Ligurian Sea (*n*~100 individuals; Podestà *et al.*, 2016) are lower than those in the Hellenic Trench (*n*~250-950 individuals; Frantzis *et al.*, 2022), and the geographic range that the East Mediterranean samples covered here is much greater than just the Hellenic Trench. The estimated N_e (point estimate and lower confidence interval) in the East Mediterranean was lower than the west Mediterranean, which supports the lower genetic diversity and allelic richness, and higher levels of inbreeding measured here compared to the west. Estimates of N_e are sensitive to sample size and the age structure of populations, and the east Mediterranean samples used here were collected over 15 years, potentially spanning several generations. Based on the relatedness networks, we now know that individuals throughout the eastern Mediterranean are connected genetically, and a more thorough understanding of habitat use across this wide basin is needed to better understand the dynamics of Cuvier's populations here.

5.4.3 Are there signals of population-level impact from naval sonar in the "disturbed" populations?

Genetic EBVs presented here provide evidence of reduced genetic variation in the face of repeated anthropogenic disturbance in the "disturbed" TOTO populations of both species, the "disturbed" east Canary Islands Cuvier's and the "disturbed" east

Mediterranean Cuvier's. Sample size was low in the "disturbed" Bahamas and Canary Island sites, however some of the EBV metrics are robust to this. The calculations of expected heterozygosity and allelic richness account for small sample sizes (Nei 1978; El Mousadik and Petit 1996) and when sample sizes are too small, estimates of N_e are negative or infinity and therefore unusable.

Most of the focal regions in this study are also proposed Evolutionarily Significant Units (ESUs), evolutionary components of a species that are necessary for the maintenance of overall genetic variability and resilience, in Chapter 3. Point estimates of N_e are available for five of these regions (none for Cuvier's in the Bahamas due to sample size constraints) and they range from N_e=33 (Bahamas Blainville's) to N_e=132.9 (west Mediterranean Cuvier's), though some of the upper confidence interval limits are infinity. Thresholds have been proposed to understand the risk of biodiversity loss based on effective population size, with N_e >100 suggested to avoid inbreeding depression and N_e >1000 required to maintain evolutionary potential (Frankham et al. 2014). Only the west and east Mediterranean Cuvier's have N_e >100, suggesting that each of the focal regions presented here are at risk of losing evolutionary potential and in some cases, inbreeding depression. Few studies have estimated N_e in cetacean populations using the linkage disequilibrium method and nuclear SNPs, but the few reported estimates are similarly small. The critically endangered western gray whale (*Eschrichtius robustus*) population is estimated to have N_e =14.1 (95% CI: 12.1, 16.7; Dewoody *et al.*, 2017) and three genetically distinct populations of killer whales (*Orcinus orca*) around Australia and New Zealand have N_e values that range from N_e ;=12-77 (Reeves et al. 2022).

5.4.4 Conclusions

Some of the core regions of beaked whale habitat have been repeatedly exposed to anthropogenic noise. Though the focus here has been on three specific regions and the impact of naval sonar use, other sound sources are thought to impact beaked whale behaviour (Aguilar de Soto et al. 2006), with potential impacts on long-term fitness (New et al. 2013). There are also other resident populations of Blainville's and Cuvier's that persist in areas with regular sonar use, including the Pacific Missile Range Facility in Hawai'i (Henderson et al. 2016) and the Southern California Anti-Submarine Warfare Range (Falcone et al. 2009; Curtis et al. 2021), and future studies should be directed at characterising the genetic variation within these populations as well. In this chapter, I have provided the first estimates of the four genetic EBVs and laid a path for future research to begin monitoring the genetic variation of beaked whale populations in response to anthropogenic and natural perturbations.

Natural and anthropogenic disturbances play a key role in driving patterns of genetic diversity and structure, and populations can respond in different ways depending on the type of disturbance and demography of different species (Banks et al. 2013). Banks *et al.* (2013) provide some examples of the way disturbance can impact the genetic diversity of populations with differing population structure and demography. In one example, coastal tailed frogs (*Ascaphys truei*) from nearby populations recolonised the area around Mt St Helens following it's eruption in 1980, increasing genetic diversity and decreasing genetic differentiation between populations (Spear et al. 2012). Galapagos giant tortoises (*Geochelone nigra vandenburghi*) underwent a substantial bottleneck following the eruption of Alcedo. The population size was recovered only by survivors of the original population (no immigrants) and the resulting genetic diversity was reduced (Beheregaray et al. 2003). Southern right whales (*Eubalaena australis*) and bowhead whales (*Balaena*

mysticetus) have lost genetic diversity due to demographic and genetic bottlenecks due to whaling (Alter et al. 2012; Jackson et al. 2016).

Here I have hypothesised that in the case of beaked whales, disturbance (naval sonar) will reduce the genetic variation of a population like the Galapagos tortoises above, where no immigrants will repopulate the disturbed local population. A lack of photo-identification matches between focal sites (Claridge 2013; Claridge et al. 2015; Reyes 2018), suggesting that neighbouring areas are not connected, supports this hypothesis. The appearance of individuals in the east Canary Islands genetically assigning to, and having close kin in, the west Canary Islands suggests that some level of movement is occurring. A deeper understanding of the dynamics and connectivity of neighbouring beaked whale populations will be imperative to predict the population-level impacts of disturbances going forward.

In the absence of historic samples to sequence, it is not possible to determine whether the measures of genetic variation found in contemporary beaked whale populations is representative of those from before the onset of sonar use, however, Chapter 3 of this thesis and a few studies that have investigated beaked whale genetic diversity using nuclear SNPs (and allowing for some missing data) provide some context for the findings presented here. The range of genetic diversity values (observed heterozygosity) estimated for the Cuvier's focal regions (Table 5-4) were all significantly higher than each population from Chapter 3 (Table 3-1) but falls within the range of published values in the three other beaked whale species that have been investigated using nuclear SNPs: True's beaked whale (*Mesoplodon mirus;* H_o=0.134, Carroll *et al.*, 2021), Ramari's beaked whale (*M. eueu;* H_o=0.159, Carroll *et al.*, 2021) and northern bottlenose whale (*Hyperoodon ampullatus;* H_o=0.172-0.184, de Greef *et al.*, 2022). Notably, the measured diversity in Mediterranean Cuvier's was substantially lower than each of the other species and Blainville's heterozygosity was substantially higher than the Blainville's populations in Chapter 3 and from the other beaked whale estimates. While it is not possible to know with certainty if the lower genetic variation observed here is due to repeated strandings and exposure to naval sonar, the fact remains that populations with lower genetic variation are less resilient to future perturbations (Hoban et al. 2022), especially when genetic drift is stronger (smaller N_e).

Few legislative measures are in place that directly address noise in the marine environment though the application of the "precautionary principle" is widely suggested by conservation and management bodies (Markus and Sánchez 2018). The fact that beaked whales are susceptible to military sonar is accepted by the scientific community and some nations (Evans et al. 2001; U.S. Department of the Navy 2018), providing evidence that this principle should be adopted and management efforts should incorporate the EBV metrics calculated here, which suggest uneven risks to Cuvier's and Blainville's populations across their wide ranges. In the Canary Islands, conservation measures are already in place that seem to have reduced the number of beaked whale strandings in response to naval sonar. Following the mass strandings that occurred coincidentally with naval exercises between 1985 – 2004, the Spanish government banned sonar use around the Canary Islands. Since this straightforward directive was implemented, no mass strandings have occurred around the archipelago (Fernández et al. 2012). The "disturbed" TOTO subpopulations of Cuvier's and Blainville's in the Bahamas overlap geographically with the active AUTEC naval sonar testing range, implying that these

individuals will continue to experience repeated disturbances and see further declines in their genetic variation.

In the Ligurian Sea, marine mammals including Cuvier's are protected within the Pelagos Sanctuary, an area with high levels of human activity that pose actual and potential threats to local populations including shipping, whale-watching, and ferries (Notarbartolo-di-Sciara et al. 2008). "Areas of Special Concern for Beaked Whales", including the Ligurian Sea and Hellenic Trench, have been mapped out by the Agreement on the Conservation of Cetaceans of the Black Sea, Mediterranean Sea and contiguous Atlantic area (ACCOBAMS) as locations where naval sonar and underwater explosions should not take place (ACCOBAMS 2013). In response, the Italian government proclaimed to discontinue military sonar use in the sanctuary (https://www.sanctuaire-pelagos.org/en/threats/underwater-noise). Despite the suggestion by ACCOBAMS in 2013, a mass stranding corresponding with naval activities took place in Greece in 2014 (ACCOBAMS 2013; Frantzis 2015; IUCN Marine Mammal Protected Areas Task Force 2017). Though the Canary Islands moratorium only presents one data point on the effectiveness of a sonar ban in preventing beaked whale mass strandings, it suggests that legally binding conservation actions are needed to protect these potentially vulnerable populations.

For the first time, genetic EBVs have been calculated to evaluate the genetic variation within cetacean populations, providing the basis to assess potential population-level impacts of naval sonar. I have also provided the first step in comparing susceptibility of different populations to future perturbations in the form of continued anthropogenic impact and/or global climate change, and in doing so have shown that "disturbed" populations are at a greater risk of a reduction in genetic variation. Since genetic variation is an integral indicator of the persistence and adaptability of a population to perturbations, it is prudent to comprehensively evaluate the four genetic EBVs within populations for inclusion in future conservation and management actions. Estimates of these variables were made possible thanks to a global collaboration (2.3.3), using samples collected over 19 years. Despite this huge collection effort, sample sizes were still limited in certain locations, suggesting that the measurement of all EBVs may not be as feasible in all populations. For example, although N_e is considered to be a fundamental parameter to monitor genetic biodiversity (Hoban et al. 2022), even focal regions with the greatest number of samples in this study (n>20) had variances in the 95% confidence intervals which were too high to make many reasonable inferences. Since obtaining samples from beaked whales is notoriously difficult (Barlow et al. 2006), it will be important for researchers using the EBV framework to use metrics which are appropriate for, and account for, smaller sample sizes, such as allelic richness and expected heterozygosity.

6 GENERAL DISCUSSION

6.1 OVERVIEW

Broad ecological questions underpin the research effort towards the effective conservation and management of wildlife populations. For example, the IUCN Red List collates data regarding geographic range, population size, habitat, ecology, threats, and current conservation actions to assess the extinction risk of species and subpopulations (IUCN Standards and Petitions Committee 2019). The techniques required to obtain this information vary widely depending on the species and location of interest, demonstrating the difficulty in comparing the findings of different studies to reach a common understanding on the management and protections of species with wide distributions (Pereira et al. 2013; Thakur et al. 2018; Hoban et al. 2022). The incorporation of rapidly advancing genomic methods is beginning to bridge this gap and provide ubiquitous tools which can be applied to a wide range of sample types, species, and ecosystems (Benestan et al. 2016; Carroll et al. 2018; Garner et al. 2020; Hoban et al. 2022). The DNA extracted from tissue samples collected from animals alive or dead, adult or infant, male or female, can provide insights into the demographic and evolutionary histories of populations and their genetic resilience to future perturbations, all contributing to the understanding of how targeted conservation efforts can be aimed (Salmona et al. 2017; Carroll et al. 2018; Waples et al. 2018; Hoban et al. 2022).

In this PhD I have shown that next-generation sequencing can be used to answer questions about beaked whale ecology and evolution over large temporal and spatial scales. I have also provided some of the evidence required for effective conservation and management, including the identification of genetically distinct management units with varying levels of genetic diversity and relatedness, different demographic histories and potentially differing abilities to adapt to future climate change. Further I have demonstrated that there is an overlap of harmful anthropogenic noise (reviewed in Chapter 1) and some genetically distinct populations, suggesting that a population-specific approach to management is more appropriate for the conservation of Cuvier's and Blainville's beaked whales.

An underlying theme of this thesis was to showcase the utility of reduced representation sequencing (RRS) in non-model organisms such as cetaceans. Prior to the start of this project, limited genomic data were available for any beaked whale species and the genomic contributions of this thesis are presented in Appendix D. In the sections below, I synthesise the findings from this thesis by summarising the outcomes from the methodology applied across chapters (Chapter 2) and each of the three research objectives (Chapters 3-5). Finally, I discuss the future research that I think is needed to continue this important work.

6.2 METHODOLOGY OUTCOMES: THE POWER OF DDRAD SEQUENCING

Molecular ecology is a rapidly growing field thanks to the constant improvements of DNA sequencing technology and the advancements of available analytical methods (Andrews et al. 2016; Cammen et al. 2016; da Fonseca et al. 2016). Access to high quality beaked whale samples is notoriously difficult and DNA sequencing, while going down in cost (Wetterstrand 2021), is still expensive. Dozens of sequencing approaches are available depending on the budget and study aims (Andrews et al. 2016; Illumina 2017), and in Chapter 2 I present the benefits of the sequencing approach taken in this thesis, double digest restriction-associated digest (ddRAD) sequencing. Here, I show that by

following strict filtering and quality assurance protocols, just one method of DNA sequencing preparation (ddRAD sequencing; Peterson et al. 2012), could result in high enough quality data to answer a broad range of research questions about ecology and evolution across time and space.

6.2.1 Small bioinformatic modifications for answering different questions

Specific alterations to the bioinformatic pipelines used to discover, genotype and filter SNPs, can allow users to harness RRS to investigate questions of ecology and evolution across a wide range of scales: species to individuals, global distributions to fine-scale populations, and temporally from measuring demographic histories during the last million years to estimating potential long-term effects of active sonar that has only been in use for the past 70 years. Using only ddRAD data, I generated SNP datasets that were specific to the unique objectives of each analysis.

In Chapter 3, samples and SNPs were selected to maximise the number of individuals and loci retained across the global distributions of the focal species, while minimising missing data and region-specific loci. In Chapter 4, two SNP datasets were generated for each of the four species analysed. The first dataset was made following the protocol in Chapter 3, as they shared the intended use of estimating genetic structure and diversity without assigning populations *a priori.* The second dataset was used to investigate demographic history, and all low frequency mutations were kept. These recent mutations provide essential information for inferring recent demographic histories and are lost when filtering by minor allele frequency (MAF). Additionally, all sites were included (not just the variable ones). In Chapter 5, SNP datasets were generated for each species using the pipeline in Chapter 3, except that populations were assigned *a priori* during the genotyping step and region-specific sites and private alleles were allowed. This step was taken to maximise the number of SNPs present in each of the region-specific datasets.

The greatest number of polymorphic sites was found when populations were defined *a priori,* and loci were kept even if they were only found in one population (if it was present in >80% of individuals; Chapter 5). This method also resulted in the highest percentage of missing data, which in conjunction with the presence of private alleles, showcases the substantial genetic structure between the studied populations.

6.2.2 Key contributions of my thesis methods

6.2.2.1 ddRAD sequencing is robust

Samples were pooled from different species into the same sequencing runs if they shared the same DNA quality/quantity score (See Chapter 2 for scoring rubric; mixing poor- and high-quality samples in the same sequencing library may result in poor-quality samples dominating the sequencing results) and so libraries in the 2020 sequencing run contained pooled Blainville's and Cuvier's individuals if they shared the same DNA score. This approach seems to have produced good outcomes both in terms of the coverage at each locus and the fact that SNPs generated using the 2020 dataset (Chapter 5) yielded similar population structure DAPC results as those generated in 2018 (Chapter 3).

In 2020, a new provider was found since the original sequencing centre at the University of Copenhagen no longer used the Illumina HiSeq2500. This new centre was based in a different country and used a different platform (Illumina NextSeq). By including duplicate individuals that had been sequenced twice, I showed that SNP genotypes are

consistently recovered (0.13-37.04% discordance across *n*=22 Cuvier's and *n*=65874 shared loci, and 0.12%-0.90% discordance across *n*=17 Blainville's and *n*=26182 shared loci, sequenced in both 2018 and 2020), at least when the libraries are prepared in the same lab and using the same protocols.

6.2.2.2 SNP datasets must match the desired objective

Above I describe the various ways in which the bioinformatic pipeline was modified to generate SNP datasets appropriate for each of the analyses conducted in Chapters 3-5. Seemingly small changes in the pipeline (i.e., filtering by minor allele frequency, *a priori* population assignments, etc.) resulted in different results across chapters, demonstrating the need to carefully interpret reported metrics in the context of the bioinformatic steps taken to generate the dataset used (Shafer et al. 2017; O'Leary et al. 2018).

Tajima's D was calculated in Chapters 3 and 4 using the same protocol, however different filtering parameters were used to generate the final SNP datasets. In Chapter 3, the same SNP datasets were used for the structure analyses and the calculations of Tajima's D. In Chapter 4, the datasets used to calculate Tajima's D were not filtered for MAF and they included all sites (not just variable ones). In the two species where Tajima's D was calculated using both datasets (Cuvier's and Blainville's), the results are overall very similar. In both chapters, Tajima's D values were in the same range (>0, <0, ~0) however the dataset in Chapter 4 yielded more extreme values (more negative or more positive) than the Chapter 3 dataset.

Genetic diversity and inbreeding were measured for each chapter using SNP datasets that were especially tailored for the unique objectives. One of the regions that was analysed across multiple chapters using different SNP datasets was the Mediterranean, with values of diversity (H_o, H_s) and inbreeding (F_{IS}) varying widely (Table 6-1). Across the datasets where a common SNP dataset was used for all Cuvier's populations, genetic diversity was lowest and inbreeding was highest in the Mediterranean. Chapter 3 shows that there is substantial structure between the Mediterranean Sea Cuvier's and those in the North Atlantic and Indo-Pacific. Chapter 3 also showed that there was just one mitogenome cluster of Mediterranean individuals which split from the North Atlantic approximately 700 kya, while the remaining polyphyletic clades contain individuals from both the Atlantic and Indo-Pacific. It is possible that the lower genetic diversity seen in the Mediterranean when analyzed with populations in other ocean basins is the result of a small founding population that has remained isolated from the rest of the North Atlantic Cuvier's. Genetic drift likely acted on this reproductively isolated, small founding population, reducing the number of polymorphic sites in the Mediterranean population (Chapter 5, Table 5-2) and biasing the estimates of diversity and inbreeding. The SNP dataset that yielded the highest genetic diversity and lowest inbreeding was the one that had the fewest loci, but was generated using *a priori* populations, allowing for SNPs to be kept even if they were only found in one population, and had no missing or monomorphic loci. The differences in the reported values suggest that caution is needed when comparing results to those in the published literature, paying particular attention to the filtering steps (MAF filter, missing data allowed, accounting for underlying population structure, etc.) used to generate them.

6.2.2.3 A comprehensive genomic beaked whale dataset

The utility of dedicated tissue and DNA archives for wildlife conservation, management and forensics is well established (Thompson et al. 2013; Pérez-Espona and CryoArks

Consortium 2021). To meet the objectives of this thesis and the broader aims of the project that this thesis was funded by, the largest collection of beaked whale-specific tissue and DNA samples was assembled and named the International Tissue Archive for Beaked Whales (ITABW). Samples were provided from seven species, collected in seven ocean basins (including the Mediterranean), by 70 collaborators, in 30 countries, as part of a global initiative to increase our knowledge of the world's most data deficient cetacean family. Since it's assembly, ITABW has not only provided the material for this PhD thesis, but also resulted in two peer-reviewed publications: Chapter 3 from this thesis (Onoufriou *et al.*, 2022) and the identification of a new beaked whale species (Ramari's beaked whale, *Mesoplodon eueu,* Carroll *et al.*, 2021). In the future, ITABW will be maintained across three institutions (the University of Auckland Waipapa Taumata Rau, maintained by Dr. Emma Carroll; the University of Copenhagen, maintained by Dr. Morten Olsen; and the University of La Laguna, maintained by Dr. Natacha Aguilar de Soto) where a committee will evaluate proposals for use of these samples to support the continued scientific research into the molecular ecology of beaked whales.

Table 6-1. Diversity estimates (H_o, H_s and F_{IS}) calculated for Cuvier's populations in the Mediterranean Sea in Chapters 3 and 5 using different SNP datasets.

Site	Chapter	*n* Individuals	*n* SNPs (% Missing)	H_o (95% CI)	H_s (95% CI)	F_{IS} (95% CI)
All	3	33	30479 (2.6%)	0.078 (0.077,0.080)	0.089 (0.088, 0.091)	0.122 (0.116,0.128)
	5	37	125769 (3.11%)	0.114 (0.113,0.115)	0.128 (0.127, 0.129)	0.114 (0.112,0.116)
	5	37	27087 (0%)	0.245 (0.243,0.247)	0.264 (0.262, 0.266)	0.071 (0.068,0.073)
East	3	14	30479 (2.6%)	0.072 (0.070,0.074)	0.078 (0.076, 0.080)	0.080 (0.071,0.088)
	5	14	140088 (11.2%)	0.106 (0.105,0.107)	0.114 (0.113,0.115)	0.071 (0.068,0.074)
	5	14	27087 (0%)	0.222 (0.219,0.224)	0.229 (0.227,0.231)	0.032 (0.028,0.036)
West	3	19	30479 (2.6%)	0.083 (0.082,0.085)	0.090 (0.089, 0.092)	0.077 (0.070,0.084)
	5	23	140088 (11.2%)	0.118 (0.117,0.119)	0.128 (0.127,0.129)	0.074 (0.071,0.077)
	5	23	27087 (0%)	0.259 (0.257,0.261)	0.266 (0.264,0.268)	0.027 (0.024,0.031)

Using the data generated by two sequencing runs allowed me to investigate beaked whale populations across vast temporal and spatial scales using only ddRAD SNPs filtered according to the objectives of each chapter (in Chapter 3, complementary phylogenomic and diversity analyses were conducted by collaborators using mitogenomes from Cuvier's and Blainville's). A summary of the genomic resources and datasets that were generated for this thesis are presented in the Supplementary Materials (D.1), including details about ITABW and ddRAD sequences, as well as the new mitogenomes developed for this project by collaborators. Beaked whales are difficult to obtain samples from due to their deep-diving capability, presence in deep and often offshore areas, and elusive surface behaviour (Heyning and Mead 2008; MacLeod 2018). The vast sample set collated for this project is unlikely to be surpassed in the near future.

The tissue archive collated for this thesis (ITABW) and the DNA resources summarised above have also provided the basis for a new projected funded by the Office of Naval Research (ONR): "Developing a universal beaked whale genotyping panel for assessing population-level impacts of anthropogenic activities". In this project, a reproducible genotyping panel is being developed to provide species and sex identification, individual and population-level information from a range of sample types (degraded tissue, environmental DNA, faeces, and exhalations- or blow), across all beaked whale species. The aim is to generate a "rapid-response" to atypical stranding events and ultimately provide insight into elusive beaked whales from poor quality or low quantity DNA samples.

6.3 RESEARCH OUTCOMES

The three research objectives proposed in the introduction and presented in Chapters 3-5 are summarised below and showcase the utility of RRS data to investigate drivers of population structure, diversity, demographic history and relatedness on different temporal and spatial scales.

6.3.1 Objective 1: Investigate the global population genetic structure of Cuvier's and Blainville's beaked whales and estimate whether current management units for conservation are appropriate.

In 2005, Dalebout *et al.* (2005) identified substantial ocean-basin level structure in Cuvier's using the mtDNA control region. Morin *et al.* (2012) also found relative concordance between geographic distance and genetic distance using whole mitogenomes, though neither study found reciprocal monophyly in phylogenetic trees. On the other hand, Blainville's were divided into two, highly supported clades between the Atlantic and Pacific Oceans based on mitogenomes. Understanding that this ocean basin-scale structure was evident in mtDNA, I hypothesised that the same was likely to be found using nuclear ddRAD SNPs.

To investigate this hypothesis in more detail, ddRAD SNPs were genotyped for n=123 Cuvier's (n=30479 ddRAD SNPs) and n=43 Blainville's (n=13988 ddRAD SNPs). Complementary mitogenomes were generated and analysed by colleagues at the University of Copenhagen for n=35 Cuvier's and n=27 Blainville's. With these expansive datasets I investigated the presence of genetic structure and diversity on two different spatial scales (between and within ocean basins) and two different processes driving these patterns (macro- and micro-evolutionary drivers).

Confirming the large scale differences in population structure first described by Dalebout *et al.* (2005) and Morin *et al.* (2012), major ocean basins were highly genetically differentiated in both species. In Cuvier's, there was clear monophyly to ocean basin using ddRAD SNPs (Atlantic Ocean, Pacific Ocean, and Mediterranean Sea). The clear differentiation between Blainville's in the Atlantic and Pacific Oceans was also highly supported. This level of structure was also statistically supported in both species based on F_{ST} at the $p<0.01$ level. The F_{ST} between Atlantic and Indo-Pacific Cuvier's was low (F_{ST}=0.018), however still higher than the SNP-based estimates of F_{ST} between recognized odontocete subspecies such as spinner dolphins (*Stenella longirostris;* F_{ST}=0.0035-0.0119; Leslie and Morin, 2018). Genetic differentiation between Blainville's sampled in the Atlantic and Indo-Pacific was much higher (F_{ST} =0.119), and

within the F_{ST} range of other odontocete subspecies e.g. short-finned pilot whales (*Globicephala macrorhynchus*; F_{ST}=0.1-0.4; Van Cise *et al.*, 2019). The greatest genetic differentiation was seen between the Cuvier's in the Mediterranean and the Atlantic (F_{ST} =0.184) and Indo-Pacific (F_{ST} =0.197). While this level of structure was expected in the dataset (Dalebout et al. 2005; Morin et al. 2012), it is reassuring to see the trends found using mtDNA reflected in the ddRAD data which suggests that the structure patterns are likely shared between males and females.

Using a larger sample set and substantially more markers than previous studies using only mtDNA (Dalebout et al. 2005; Morin et al. 2012), structure was identified within ocean basins. Biopsy sampling effort was biased towards individuals in the focal regions discussed in Chapter 5, and genetic differentiation was found between Cuvier's and Blainville's in these areas with known resident populations (Canary Islands, Bahamas, Hawai'i). The greatest genetic difference found within an ocean basin in either species was the Cuvier's in the East and West Mediterranean (F_{ST} =0.092), though it would still be classified as only "moderate differentiation" according to Wright (1978).

Macroevolutionary processes driving differences between Cuvier's and Blainville's structure patterns include different evolutionary histories and abilities of each species to navigate the presence of well-known biogeographic barriers (Toonen et al. 2016). Variations in oceanographic conditions during periods of historic climate change may have limited the dispersal ability of the smaller Blainville's beaked whale, resulting in the stronger differentiation seen between Atlantic and Pacific populations. Smaller body size can be linked to life span, and the smaller Blainville's may also have experienced faster genetic drift and subsequent differentiation.

Microevolutionary processes likely drive the differing structure patterns found between species here. Like other deep-diving cetaceans (Alexander et al. 2016; VanCise et al. 2017; Vachon et al. 2022), resident populations of Cuvier's and Blainville's are typically found in close proximity to oceanic islands, shelf edges, sea mounts and undersea canyons. Some degree of site fidelity occurs in both species, contributing to genetic differentiation between populations. Cuvier's have larger bodies and geographic range, potentially increasing the dispersal potential for transient individuals, facilitating gene flow. Regardless of the drivers, establishing contemporary patterns of structure is necessary to understand how populations may respond to future disturbances.

As described throughout this thesis, studies of beaked whale population genomics are relatively few and many aspects of this chapter are novel. Previous studies have combined nuclear and mtDNA markers to study Gray's beaked whales (Thompson, Patel, Baker, et al. 2016; Westbury et al. 2021) and northern bottlenose whales (Dalebout et al. 2006; Feyrer et al. 2019; de Greef et al. 2022), however here I have presented the first study to do this in Cuvier's and Blainville's. I have also presented the first comparison of population structure and genetic diversity between beaked whale species using nuclear DNA (beaked whale mitogenomes were compared in Morin et al. 2012). Biogeographic barriers and their impact on population structure have been investigated in other deep-diving odontocetes (VanCise et al. 2019), but this study is the first to investigate them as drivers of macro- and micro-evolutionary beaked whale structure. Here I have also presented the first direct recommendations for designating management units in beaked whales (Evolutionarily Significant Units and Demographically Independent Populations)

and provided the first evidence of subspecies-level differentiation within beaked whales (discussed further in 6.4.1).

6.3.2 Objective 2: Investigate contemporary patterns of population structure and historical demographic histories in North Atlantic beaked whales with respect to future climate change.

In this chapter, I estimated changes in effective population size (N_e) through time of four beaked whale species found in the North Atlantic: Cuvier's, Blainville's and Sowerby's beaked whales and northern bottlenose whales. DNA sequences were prepared for Sowerby's in the same way (and at the same time) as those for Cuvier's and Blainville's in Chapter 3, but the northern bottlenose data were generated using a different method (GBS) by collaborators at the University of Copenhagen. Two SNP datasets were generated: (1) as described in Chapter 3 for structure and diversity measurements, and (2) suitable for demographic history reconstructions (i.e., with monomorphic loci and no MAF filter).

Reconstructions of demographic history assume panmixia within the population, so it was necessary to identify discrete units for these analyses by investigating population structure. Evidence for population structure that corresponds to geographic origin had already been established for Cuvier's and Blainville's in Chapter 3 and was confirmed using the SNP datasets generated for this study. In Sowerby's, no clear pattern of population structure was detected, with some evidence of structure in northern bottlenose.

Beaked whale populations responded to historic climate change via notable changes in population trajectories linked to wide-scale glacial and interglacial periods. During these periods, fluctuations in oceanographic conditions and prey abundance and/or composition would have played a key role in the observed bottlenecks, rapid expansions, and slow contractions within beaked whale populations. Individual demographic histories in the North Atlantic did not show clear responses to past climate change according to species, body size or species range, and instead are most likely influenced by a complex relationship between physiology (thermal constraints, dive capability), behaviour (site fidelity, foraging behaviour), and evolutionary history (population structure, geographic range). Declines in N_e can also be attributed to the presence of population structure, decline in gene flow or the establishment of new populations through the "founder effect", and it is possible that this historic climate change caused changes in population structure rather than population size (Chikhi et al. 2010; Heller et al. 2013; Mazet et al. 2015; Salmona et al. 2017; Loog 2021).

This chapter provides substantially novel contributions to beaked whale science, including the first dedicated investigation of Sowerby's population structure and genetic diversity across their North Atlantic range (though limited in sample size and discussed further in 6.4.2). I have also presented the first multispecies investigation of demographic history across any deep-diving cetacean, and reconstructed demographic histories from Cuvier's, Blainville's and Sowerby's for the first time (the demographic history of northern bottlenose was investigated in Feyrer et al. 2019 and de Greef et al. 2022).

6.3.3 Objective 3: Investigate whether population-level impacts of repeated anthropogenic disturbance are present in small beaked whale populations using genomic methods.

In Chapter 5, I compared the genetic variability of beaked whale populations that have been 'disturbed' repeatedly by naval sonar to those considered to live in a 'semi-pristine' environment using the framework of the genetic Essential Biodiversity Variables (EBVs, Hoban *et al.*, 2022). Genetic EBVs have been proposed as a tool to monitor and compare the genetic health and viability of wild populations to current and future threats (Hoban et al. 2022). Within this framework, genetic differentiation was estimated in Chapter 3 and confirmed here using discriminant analysis of principle components (DAPC).

This chapter took a comparative approach by investigating fine-scale populations of Cuvier's within three focal regions: Canary Islands, Bahamas, and the Mediterranean (Cuvier's only). Sequences were available from those generated for Chapters 3 and 4, and additional samples were sequenced from the three focal regions for this chapter. Using region-specific SNP datasets, the four genetic EBVs were calculated for each 'disturbed' and 'semi-pristine' site where the sample size was sufficient: genetic differentiation, genetic diversity, inbreeding and effective population size. At least one species showed signs of reduced genetic variation (lower genetic diversity, higher inbreeding, more sibling relationships) in each of the "disturbed" focal sites. While it is not possible to know with certainty if the lower genetic variation observed here is due to repeated strandings and exposure to naval sonar, the fact remains that populations with lower genetic variation are less resilient to future perturbations (Hoban et al. 2022), especially when genetic drift is stronger (smaller N_e).

Despite a global collaboration providing samples that had been collected over 19 years, sample sizes were still limited in certain locations, suggesting that the measurement of all EBVs may not be as feasible in all populations. Calculations of N_e were the most impacted by low sample sizes; even populations with the greatest number of samples in this study (n>20) had variances in the 95% confidence intervals which were too high to make many reasonable inferences. Some of the EBV metrics are more robust to lower samples sizes; expected heterozygosity (H_s) and allelic richness (AR) account for small sample sizes (Nei 1978; El Mousadik and Petit 1996).

In this chapter, I show that genetic EBVs can be used to provide baseline estimates of genetic variation, an integral indicator of the persistence and adaptability of a population to perturbations. While such calculations provide a framework for comparisons in the face of anthropogenic or natural disturbance, limitations in sample size must be considered by researchers hoping to calculate these metrics in the future. Future work would ideally measure genetic EBVs in historic samples from the same sites, providing information about whether the reductions of genetic variation measured here occurred since the offset of naval sonar use. Additional contemporary samples would help refine the estimates of N_e.

In this chapter, I applied the framework of genetic EBVs to cetacean populations for the first time (based on published literature). In doing so, I provided the first estimates of contemporary N_e based on nuclear markers in beaked whales (estimates based on mtDNA are provided in Dalebout 2002 and long-term N_e is estimated in Dalebout et al. 2006). This direct comparison of the susceptibility of beaked whale populations to future perturbations with regards to their genetic composition provided the first evidence that

continued and/or future disturbance may result in genetic erosion of beaked whale populations living in "disturbed" areas.

6.4 FUTURE DIRECTIONS

Investigating beaked whale genomics in this thesis has provided answers to many questions about their ecology and evolution. Notably, population structure is a common feature across most species and in the case of Cuvier's and Blainville's, it is present across all investigated geographic and temporal scales. Such findings naturally lead to more questions. What is driving the differences in genetic structure between Cuvier's and Blainville's compared to northern bottlenose and Sowerby's? Does the significant structure found in Cuvier's and Blainville's warrant new taxonomic classification? How does the reduced genetic variation in 'disturbed' populations translate to extinction risk?

In this section, I will discuss some of the ways in which these questions could be addressed in future studies. I specifically focus on the species for which we have the most data (Cuvier's, Blainville's and Sowerby's) as northern bottlenose populations are the focus of several past and ongoing population genetic studies (Dalebout et al. 2001; Dalebout et al. 2006; Feyrer et al. 2019; de Greef et al. 2022).

6.4.1 Describe Cuvier's and Blainville's Subspecies

In this thesis, I have found evidence to suggest the presence of separate subspecies in Cuvier's and Blainville's. Both Cuvier's and Blainville's showed nearly reciprocal monophyly to ocean basin using ddRAD SNPs and mitogenomes and there was minimal admixture between ocean basins using the *tess3r* cluster analysis. This provides evidence that Cuvier's in the Mediterranean, and the Atlantic and Indo-Pacific Blainville's are separately evolving lineages (Chapter 3). Two measurements of differentiation and divergence were estimated using the ddRAD and mitogenome data (F_{ST} and dA) to provide complementary evidence of population isolation.

The definition of a cetacean subspecies is a lineage that is separately evolving due to discontinuities that restrict gene flow, resulting in populations that are measurably distinct (Taylor, Perrin, et al. 2017). The 93 cetacean species are currently "under classified", with only 20% have recognized subspecies, none of which are identified in the beaked whales (Taylor, Perrin, et al. 2017; Commitee on Taxonomy 2022). Classification of cetacean taxonomic groups is important for management by providing definitions for the various units to conserve. Cetaceans are particularly under classified for a number of reasons: there is limited scope and expertise for morphological-based methods, specimens are legally protected and hard to gain access to, cetaceans live in the sea where access is often difficult, and many cetaceans have wide ranging habitats with little indication of barriers that could lead to regional adaptation or differentiation (Taylor, Perrin, et al. 2017). Molecular data has been proposed as a useful tool for the designation of cetacean subspecies due to the difficulties described above, recognizing that genetic discontinuity associated with geographical differentiation is a strong and suitable line of evidence to classify cetacean subspecies (Taylor, Perrin, et al. 2017).

The F_{ST} values for the proposed subspecies (Cuvier's in the Mediterranean, Blainville's in the Atlantic and Indo-Pacific) using nuclear SNPs (Cuvier's: Mediterranean vs Atlantic F_{ST}=0.184 and Mediterranean vs Indo-Pacific F_{ST}=0.197, Blainville's: Atlantic vs Indo-Pacific F_{ST}=0.119), fall within the range of F_{ST} values calculated from nuclear SNPs in

recognized odontocete subspecies of spinner dolphins (*Stenella longirostris;* F_{ST}=0.0035-0.012; Leslie and Morin 2018), pantropical spotted dolphins (*S. attenuata;* F_{ST}=0.055; Leslie and Morin 2018), harbour porpoise (*Phocoena phocoena*; F_{ST}=0.187−0.207; Lah et al. 2016), and short-finned pilot whale (*Globicephala macrorhynchus*; F_{ST}=0.1−0.4; Van Cise et al. 2019). The F_{ST} of proposed subspecies using full mitogenomes (Cuvier's F_{ST}=0.375 and F_{ST}=0.623, Blainville's F_{ST}=0.711) are much higher than mitogenomic F_{ST} values calculated in recognized odontocete subspecies of spinner dolphins (F_{ST}=0.013; Leslie and Morin 2018), pantropical spotted dolphins (F_{ST}=0.013; Leslie et al. 2018), fin whales (*Balaenoptera physalus*; F_{ST}=0.005−0.018; Archer et al. 2019). The F_{ST} of the proposed subspecies calculated using the mtDNA CR (Cuvier's F_{ST}= 0.347 and 0.531, Blainville's F_{ST}=0.328) is also greater than the range of F_{ST} values in recognized cetacean subspecies (F_{ST}=0.013-0.209; Rosel *et al.*, 2017).

Net nucleotide divergence, d_A (Nei, 1987; pg. 276) was found to effectively distinguish between cetacean populations using the mtDNA CR and the proposed range of d_A values between subspecies is 0.004-0.02 (Rosel, Taylor, et al. 2017; Rosel, Hancock-hanser, et al. 2017; Taylor, Archer, et al. 2017). Here, the d_A estimate between Indo-Pacific and Mediterranean Cuvier's falls within this range (Cuvier's d_A =0.0045). Unlike other cetaceans, the beaked whale control region is less diverse compared to other regions of the mitochondrial genome (Dalebout et al. 2004; Morin et al. 2012) and d_A measurements using only the CR may not be appropriate for beaked whales.

Guidelines to classify cetaceans using nuclear DNA markers like those that have been proposed using mtDNA (Taylor, Archer, et al. 2017) would be welcome in light of the nuclear data being generated for more cetacean species. Using the ddRAD and mitogenome sequences generated in this thesis, further evidence could be provided by calculating divergence times between putative subspecies. The sequencing of whole nuclear genomes could allow for comparisons of more ancient patterns of demographic history and divergence, and provide information about natural selection and local adaptation (Cammen et al. 2016; Fan et al. 2019). Further lines of evidence could also be examined to support the proposed subspecies. Differences in skull morphology have been used to identify potentially different Sowerby's beaked whale populations (Smith, Mead, et al. 2021) and provided evidence for speciation between True's and Ramari's beaked whales (*Mesoplodon mirus* and *M. eueu*; Carroll *et al.*, 2021), and this technique would be useful evidence for the description of Cuvier's and Blainville's beaked whales.

In the event that future lines of evidence become available in order to officially raise the proposed subspecies, the following names are suggested: The proposed Mediterranean subspecies of Cuvier's would become the nominate form, *Ziphius cavirostris cavirostris*, as the type specimen described by Cuvier (1824) was found at the mouth of the Rhone on the French Mediterranean coast. The next available name for Cuvier's found and described outside of the Mediterranean is *Ziphius cavirostris gervaisii* (Duvernoy 1851). The type specimen for Blainville's has no known locality (Blainville 1817), and would need genetic confirmation of subspecies. It is most likely that the type specimen was found in the Atlantic, which would be the nominate subspecies *Mesoplodon densirostris densirostris*. The first available name for the Indo-Pacific subspecies would be *Mesoplodon densirostris sechellensis* (Gray 1846). If the type specimen originated in the Indo-Pacific, a new name will need to be selected for the Atlantic subspecies.

6.4.2 Uncertain population structure in Sowerby's beaked whales

This thesis produced the first dedicated analysis of genomic population structure in Sowerby's beaked whales (Chapter 4). Like Gray's beaked whales (*Mesoplodon grayii*; Thompson *et al.*, 2016; Westbury *et al.*, 2021), I did not detect any clear patterns of population structure within their North Atlantic range that corresponded with geographic origin of samples.

Analyses based on skull morphology and stable isotopes have suggested that there is likely a Sowerby's metapopulation in the Atlantic (Smith, Trueman, et al. 2021; Smith, Mead, et al. 2021), potentially divided between the east and west. This pattern was not clear in the genomic analysis presented in this thesis, with the only cluster detected in *tess3r* that showed significant differentiation consisting of two highly divergent individuals from Denmark.

It is unclear from the samples included here if there is any population structure that corresponds with geographic location, and several other factors could contribute to the structuring of individuals (such as habitat or prey preference, social structure, etc.). The ddRAD sequencies generated here could be partitioned by sex and reanalysed to see if there are differing patterns of structure between males and females. This data could also be used to scan for outlier loci, which can be better at identifying genetic structure in some species than neutral SNPs (Freamo et al. 2011). The ddRAD data could also be incorporated into a seascape genetics approach to see if structure is linked to environmental characteristics such as temperature and/or productivity (Amaral et al. 2012). Future sequencing efforts could be used to resequence whole nuclear genomes to resolve structure in both neutral and non-neutral markers and/or investigate functional genes (Cammen et al. 2016). Greater sampling throughout the Sowerby's range (especially from the western extent of their distribution) is highly recommended to assist in refining the estimates of genetic structure and subdividing individuals into appropriate management units for conservation.

6.4.3 Population viability analyses

In Chapter 3, I have identified that the small, resident populations of Cuvier's and Blainville's found around many island archipelagos are genetically distinct and Chapter 5 shows that populations in areas more 'disturbed' by naval sonar show reduced genetic variation. Using this information, Population Viability Analysis (PVA) software, such as "Vortex" (Shaffer 1981; Lacy 1993; Lacy 2000; Lacy and Pollak 2017) could be used to simulate and predict population trends and calculate a risk of extinction for 'disturbed' beaked whale populations. PVA is a method used to simulate the trajectory of populations or species, allowing you to estimate the minimum viable population size required to ensure persistence and maintain adaptive potential (Shaffer 1981; Hoban et al. 2012). Such simulations are flexible and incorporate population-specific demographic parameters to apply various stochastic pressures to a system, forecasting the trajectory of the population over time (Shaffer 1981; Hoban et al. 2012). Small populations are more susceptible to extinction (Shaffer 1981), and as the size decreases, chance events are more likely to impact population persistence (Lacy et al. 2017). The minimum viable population size must then account for normal conditions as well as in the presence of extreme and unpredicted perturbations (Shaffer 1981).

The demographic parameters required as input for the simulations can be derived from the long-term datasets in areas with resident populations. Such parameters included in other PVA analyses include the minimum and maximum age of reproduction, longevity, percentage of reproductive individuals, sex ratio, breeding adults, adult and calf mortality, and census population size (Miller 2016). PVA using Vortex would incorporate the newly determined data on genetic diversity, population structure and connectivity and estimate the effects of stochastic events such as environmental change and anthropogenic impact. The outcome of the simulations can provide managers with the information they need to implement more specific conservation measures with measurable targets and timeframes for population recovery (Lindenmayer et al. 1993; Hoban et al. 2012). In the case of beaked whale populations, such measures could include sonar bans like the one successfully implemented in the Canary Islands (Fernández et al. 2012).

6.5 CONCLUSIONS

The revelations in my thesis about the elusive beaked whales would not have been possible without global collaborations and the provision of samples from dozens of sources. The maintenance and strengthening of this collaboration are of the utmost importance to uncover as much as we can about this potentially threatened family of whales. While dedicated biopsy sampling typically provides the highest quality DNA and can provide invaluable metadata thanks to the sampling of an individual in their known habitat and social group, the opportunistic collection of tissue from dead stranded and bycaught individuals continues to provide a wealth of information about the distribution of these difficult to study species in sometimes remote, geographic areas. Each new sample is a valuable contribution to beaked whale science, and I encourage the continued collaboration between researchers, conservation organisations, stranding networks, and governments to collect and archive tissue from as many beaked whales as possible to facilitate future studies like this one.

References

Abreo NAS, Macusi ED, Blatchley DD, Cuenca-Ocay GC. 2016. First evidence of plastic ingestion by the rare Deraniyagala's beaked whale (*Mesoplodon hotaula*). IAMURE Int J Ecol Conserv. 19. doi:ISSN 2244-1581.

ACCOBAMS. Cuvier's beaked whale. In: ACCOBAMS Cetacean Summary.

ACCOBAMS. 2013. Recommendation on the conservation of Cuvier's Beaked Whales in the Mediterranean: "Areas of Special Concern for Beaked Whales" (ASC-BW) and mitigation protocols for anthropogenic activities using intense underwater sound sources. In: Report of the Fifth Meeting of the Parties to ACCOBAMS. Tangier, Morocco.

Aguilar de Soto N. 2006. Acoustic and diving behaviour of the short finned pilot whale (*Globicephala macrorhynchus*) and Blainville´s beaked whale (*Mesoplodon densirostris*) in the Canary Islands. Implications on the effects of man-made noise and boat collisions. PhD Thesis. Universidad de La Laguna, Spain.

Aguilar de Soto N, Johnson MP, Madsen PT, Tyack PL, Bocconcelli A, Fabrizio Borsani J. 2006. Does intense ship noise disrupt foraging in deep-diving Cuvier's beaked whales (*Ziphius cavirostris*)? Mar Mammal Sci. 22(3):690–699. doi:10.1111/j.1748-7692.2006.00044.x.

Aguilar de Soto N, Madsen PT, Tyack PL, Arranz P, Marrero J, Fais A, Revelli E, Johnson MP. 2012. No shallow talk: Cryptic strategy in the vocal communication of Blainville's beaked whales. Mar Mammal Sci. 28(2):75–92. doi:10.1111/j.1748-7692.2011.00495.x.

Aguilar de Soto N, Martín V, Silva MA, Edler R, Reyes C, Carrillo M, Schiavi A, Morales T, García-Ovide B, Sanchez-Mora A, et al. 2017. True's beaked whale (*Mesoplodon mirus*) in Macaronesia. PeerJ. 5:e3059. doi:10.7717/peerj.3059.

Aguilar de Soto N, Reyes C, Schiavi A, Hammond PS. 2015. Population parameters of Blainville's and Cuvier´s beaked whales. Annual Report to Office of Naval Research.

Aguilar de Soto N, Visser F, Tyack PL, Alcazar J, Ruxton G, Arranz P, Madsen PT, Johnson MP. 2020. Fear of killer whales drives extreme synchrony in deep diving beaked whales. Sci Rep. 10(3):13. doi:10.1038/s41598-019-55911-3.

Albertson GR, Baird RW, Oremus M, Poole MM, Martien KK, Baker CS. 2017. Staying close to home? Genetic differentiation of rough-toothed dolphins near oceanic islands in the central Pacific Ocean. Conserv Genet. 18(1):33–51. doi:10.1007/s10592-016-0880-z.

Alexander A, Steel D, Hoekzema K, Mesnick SL, Engelhaupt D, Kerr I, Payne R, Baker CS. 2016. What influences the worldwide genetic structure of sperm whales (*Physeter macrocephalus*)? Mol Ecol. 25(12):2754–2772. doi:10.1111/mec.13638.

Alexiadou P, Foskolos I, Frantzis A. 2019. Ingestion of macroplastics by odontocetes of the Greek Seas, Eastern Mediterranean: Often deadly! Mar Pollut Bull. 146(May):67–75. doi:10.1016/j.marpolbul.2019.05.055.

Alho JS, Välimäki K, Merilä J. 2010. Rhh: An R extension for estimating multilocus heterozygosity and heterozygosity-heterozygosity correlation. Mol Ecol Resour. 10(4):720–722. doi:10.1111/j.1755-0998.2010.02830.x.

Allen AN, Schanze JJ, Solow AR, Tyack PL. 2014. Analysis of a Blainville's beaked whale's movement response to playback of killer whale vocalizations. Mar Mammal Sci. 30(1):154–168. doi:10.1111/mms.12028.

Allen BM, Brownell Jr. RL, Meade J. 2011. Species review of Cuvier's beaked whale,

Ziphius cavirostris. Int Whal Comm Sci Meet 63. SC/63/SM17:1–8.

Allen BM, Mead JG, Brownell Jr. RL. 2011. Species review of Blainville's beaked whale, *Mesoplodon densirostris*. Int Whal Comm Sci Meet 63. SC/63/SM16:1–18.

Alley RB, Mayewski PA, Sowers T, Stuiver M, Taylor KC, Clark PU. 1997. Holocene climatic instability: a prominent, widespread event 8200 yr ago. Geology. 25(6):483–486. doi:10.1130/0091-7613(1997)025<0483:HCIAPW>2.3.CO;2.

Alter SE, Rosenbaum HC, Postma LD, Whitridge P, Gaines C, Weber D, Egan MG, Lindsay M, Amato G, Dueck L, et al. 2012. Gene flow on ice: the role of sea ice and whaling in shaping Holarctic genetic diversity and population differentiation in bowhead whales (*Balaena mysticetus*). Ecol Evol. 2(11):2895–2911. doi:10.1002/ece3.397.

Amaral AR, Beheregaray LB, Bilgmann K, Boutov D, Freitas L, Robertson KM, Sequeira M, Stockin KA, Coelho MM, Möller LM. 2012. Seascape genetics of a globally distributed, highly mobile marine mammal: the short-beaked common dolphin (Genus *Delphinus*). PLoS One. 7(2). doi:10.1371/journal.pone.0031482.

Amos W, Worthington Wilmer J, Fullard K, Burg TM, Croxall JP, Bloch D, Coulson T. 2001. The influence of parental relatedness on reproductive success. Proc R Soc B Biol Sci. 268(1480):2021–2027. doi:10.1098/rspb.2001.1751.

Andrews KR, Good JM, Miller MR, Luikart G, Hohenlohe PA. 2016. Harnessing the power of RADseq for ecological and evolutionary genomics. Nat Rev Genet. 17(2):81–92. doi:10.1038/nrg.2015.28.

Andrews KR, Luikart G. 2014. Recent novel approaches for population genomics data analysis. Mol Ecol News Views. 23:1661–1667. doi:https://doi.org/10.1111/mec.12686.

Andrews S, Lindenbaum P, Howard B, Ewels P. 2010. FastQC High Throughput Sequence QC Report. www.bioinformatics.babraham.ac.uk/projects/.

Archer FI, Adams PE, Schneiders BB. 2017. stratag: An r package for manipulating, summarizing and analysing population genetic data. Mol Ecol Resour. 17(1):5–11. doi:10.1111/1755-0998.12559.

Archer FI, Brownell Jr. RL, Hancock-Hanser BL, Morin PA, Robertson KM, Sherman KK, Calambokidis J, Urbán Ramirez J, Rosel PE, Mizroch SA, et al. 2019. Revision of fin whale *Balaenoptera physalus* (Linnaeus, 1758) subspecies using genetics. J Mammal. 100(5):1653–1670. doi:10.1093/jmammal/gyz121.

Arranz P, Aguilar de Soto N, Madsen PT, Brito A, Bordes F, Johnson MP. 2011. Following a foraging fish-finder: Diel habitat use of Blainville's beaked whales revealed by echolocation. PLoS One. 6(12). doi:10.1371/journal.pone.0028353.

Atkinson LP, Grosch CE. 1999. Continental runoff and effects on the North Atlantic Ocean Subtropical Mode Water. Geophys Res Lett. 26(19):2977–2980. doi:10.1029/1998GL010377.

Ayres KL. 2000. Relatedness testing in subdivided populations. Forensic Sci Int. 114(2):107–115. doi:10.1016/S0379-0738(00)00289-9.

Baird NA, Etter PD, Atwood TS, Currey MC, Shiver AL, Lewis ZA, Selker EU, Cresko WA, Johnson EA. 2008. Rapid SNP discovery and genetic mapping using sequenced RAD markers. PLoS One. 3(10):1–7. doi:10.1371/journal.pone.0003376.

Baird RW. 2018. Cuvier's Beaked Whale. In: Wursig B, Thewissen JGM, Kovacs KM, editors. Encyclopedia of Marine Mammals, 3rd Edition. Academic Press. p. 234–237.

Baird RW. 2019. Behavior and ecology of not-so-social odontocetes: Cuvier's and Blainville's beaked whales. In: Würsig B, editor. Ethology and Behavioural Ecology of Odontocetes. Springer Nature. p. 1–47.

Baird RW, Brownell Jr. RL, Taylor BL. 2020. *Ziphius cavirostris* (Global), Cuvier's Beaked Whale. IUCN Red List Threat Species. e.T23211A5. doi:https://dx.doi.org/10.2305/IUCN.UK.2020- 3.RLTS.T23211A50379111.en.

Baird RW, McSweeney DJ, Schorr GS, Mahaffy SD, Webster DL, Barlow JP, Hanson MB, Turner JP, Andrews RD. 2009. Studies of beaked whales in Hawai'i: population size, movements, trophic ecology, social organisation and behaviour. In: Dolman SJ, MacLeod CD, Evans PGH, editors. ECS Special Publication Series. p. 23–25.

Baird RW, Schorr GS, Webster DL, Mahaffy SD, McSweeney DJ, Hanson MB, Andrews RD. 2011. Open-ocean movements of a satellite-tagged Blainville's beaked whale (*Mesoplodon densirostris*): Evidence for an offshore population in Hawai'i? Aquat Mamm. 37(4):506–511. doi:10.1578/AM.37.4.2011.506.

Baird RW, Webster DL, McSweeney DJ, Ligon AD, Schorr GS, Barlow JP. 2006. Diving behaviour of Cuvier's (*Ziphius cavirostris*) and Blainville's (*Mesoplodon densirostris*) beaked whales in Hawai'i. Can J Zool. 84(8):1120–1128. doi:10.1139/z06-095.

Baird RW, Webster DL, Schorr GS, McSweeney DJ, Barlow JP. 2008. Diel variation in beaked whale diving behavior. Mar Mammal Sci. 24(3):630–642. doi:10.1111/j.1748-7692.2008.00211.x.

Baker CS, Cipriano F, Palumbi SR. 1996. Molecular genetic identification of whale and dolphin products from commercial markets in Korea and Japan. Mol Ecol. 5:671–685. doi:10.1111/j.1365-294X.1996.tb00362.x.

Baker CS, Slade RW, Bannister JL, Abernethy RB, Weinrich MT, Lien J, Urbán Ramirez J, Corkeron P, Calambokidis J, Vasquez O, et al. 1994. Hierarchical structure of mitochondrial DNA gene flow among humpback whales *Megaptera novaeangliae*, world-wide. Mol Ecol. 3(4):313–27. doi:10.1111/j.1365-294X.1994.tb00071.x.

Balcomb III KC, Claridge DE. 2001. A mass stranding of cetaceans caused by naval sonar in the Bahamas. Bahamas J Sci. 5:2–12.

Banguera-Hinestroza E, Bjørge A, Reid RJ, Jepson PD, Hoelzel AR, Rus. 2010. The influence of glacial epochs and habitat dependence on the diversity and phylogeography of a coastal dolphin species: *Lagenorhynchus albirostris*. Conserv Genet. 11(5):1823–1836. doi:10.1007/s10592-010-0075-y.

Banks SC, Cary GJ, Smith AL, Davies ID, Driscoll DA, Gill AM, Lindenmayer DB, Peakall R. 2013. How does ecological disturbance influence genetic diversity? Trends Ecol Evol. 28(11):670–679. doi:10.1016/j.tree.2013.08.005.

Barlow JP, Ferguson M, Perrin WF, Gerrodette T, Joyce G, MacLeod CD, Mullin KD, Palka DL, Waring GT. 2006. Abundance and densities of beaked and bottlenose whales (family Ziphiidae). J Cetacean Res Manag. 7(3):263–270. doi:10.1111/j.1748-7692.1997.tb00656.x.

Beheregaray LB, Ciofi C, Geist D, Gibbs JP, Caccone A, Powell JR. 2003. Genes record a prehistoric volcano eruption in the Galápagos. Science. 302:75. doi:10.1126/science.1087486.

Benestan LM, Ferchaud A-L, Hohenlohe PA, Garner BA, Naylor GJP, Baums IB, Schwartz MK, Kelley JL, Luikart G. 2016. Conservation genomics of natural and

managed populations: Building a conceptual and practical framework. Mol Ecol. 25(13):2967–2977. doi:10.1111/mec.13647.

Benson DA, Karsch-Mizrachi, Lipman DJ, Ostell J, Wheeler DL. 2005. GenBank. Nucleic Acids Res. 33(Database issue):D34–D38. doi:10.1093/nar/gki063.

van den Berg GL, Vermeulen E, Valenzuela LO, Bérubé M, Ganswindt A, Gröcke DR, Hall G, Hulva P, Neveceralova P, Palsbøll PJ, et al. 2021. Decadal shift in foraging strategy of a migratory southern ocean predator. Glob Chang Biol. 27(5):1052–1067. doi:10.1111/gcb.15465.

Bernaldo de Quirós Y, Fernández A, Baird RW, Brownell Jr. RL, Aguilar de Soto N, Allen D, Arbelo M, Arregui M, Costidis AM, Fahlman A, et al. 2019. Advances in research on the impacts of anti-submarine sonar on beaked whales. Proc R Soc B Biol Sci. 286(1895):20182533. doi:10.1098/rspb.2018.2533.

Berrow S, Meade R, Marrinan M, McKeogh E, O'Brien J. 2018. First confirmed sighting of Sowerby's beaked whale (*Mesoplodon bidens* (Sowerby, 1804)) with calves in the Northeast Atlantic. Mar Biodivers Rec. 11(1):1–5. doi:10.1186/s41200-018-0154-1.

Bianchi CN, Morri C. 2000. Marine biodiversity of the Mediterranean Sea: Situation, problems and prospects for future research. Mar Pollut Bull. 40(5):367–376. doi:10.1016/S0025-326X(00)00027-8.

Bianucci G, Di Celma C, Urbina M, Lambert O. 2016. New beaked whales from the late Miocene of Peru and evidence for convergent evolution in stem and crown Ziphiidae (Cetacea, Odontoceti). PeerJ. 4:e2479. doi:10.7717/peerj.2479.

Bienvenu F. 2019. The equivocal mean age of parents in a cohort. Am Nat. 194(2):276–284. doi:10.1086/704110.

Bittau L, Leone M, Gannier Adrien, Gannier Alexandre, Manconi R. 2018. Two live sightings of Sowerby's beaked whale (*Mesoplodon bidens*) from the western Mediterranean (Tyrrhenian Sea). J Mar Biol Assoc United Kingdom. 98(5):1003–1009. doi:10.1017/S0025315416001892.

Blainville HMD. 1817. Delphinorhynques. Nouv Dict d'Histoire Nat. 9:151–152.

Blouin MS, Parsons M, Lacaille V, Lotz S. 1996. Use of microsatellite loci to classify individuals by relatedness. Mol Ecol. 5(3):393–401. doi:10.1111/j.1365-294X.1996.tb00329.x.

Bouckaert R, Vaughan TG, Barido-Sottani J, Duchêne S, Fourment M, Gavryushkina A, Heled J, Jones G, Kühnert D, De Maio N, et al. 2019. BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. PLoS Comput Biol. 15(4):1–28. doi:10.1371/journal.pcbi.1006650.

Bouzat JL. 2010. Conservation genetics of population bottlenecks: the role of chance, selection, and history. Conserv Genet. 11:463–478. doi:10.1007/s10592-010-0049-0.

Bowen BW, Karl SA. 2007. Population genetics and phylogeography of sea turtles. Mol Ecol. 16(23):4886–4907. doi:10.1111/j.1365-294X.2007.03542.x.

Bowen WD. 1997. Role of marine mammals in aquatic ecosystems. Mar Ecol Prog Ser. 158(1):267–274. doi:10.3354/meps158267.

van Bree PJH, Kristensen I. 1974. On the intriguing stranding of four Cuvier's Beaked Whales, *Ziphius cavirostris*, G. Cuvier, 1823, on the Lesser Antillean Island of Bonaire. Bijdr tot Dierkd. 44(2):235–238.

Broad Institute. 2019. Picard Toolkit.

Brownlow AC, Davison NJ, ten Doeschate M. 2014. Annual Report 2014, 1 January to 31 December 2014, for Marine Scotland, Scottish Government.

Brownlow AC, Davison NJ, ten Doeschate M. 2015. Annual Report 2015 - 1 January to 31 December 2015, for Marine Scotland, Scottish Government.

Brownlow AC, Davison NJ, ten Doeschate M. 2018. Annual Report 2018, 1 January to 31 December 2018, for Marine Scotland, Scottish Government.

Brüniche-Olsen A, Urban RJ, Vertyankin V V., Godard-Codding CAJ, Bickham JW, DeWoody JA. 2018. Genetic data reveal mixed-stock aggregations of gray whales in the North Pacific Ocean. Biol Lett. 14(10):1–4. doi:10.1098/rsbl.2018.0399.

Buckley MW, Marshall J. 2016. Observations, inferences, and mechanisms of the Atlantic Meridional Overturning Circulation: a review. Rev Geophys. 54(1):5–63. doi:10.1002/2015RG000493.

Cabrera AA, Schall E, Bérubé M, Anderwald P, Bachmann L, Berrow S, Best PB, Clapham PJ, Cunha HA, Dalla Rosa L, et al. 2022. Strong and lasting impacts of past global warming on baleen whales and their prey. Glob Chang Biol. 28(8):2657–2677. doi:10.1111/gcb.16085.

Caesar L, Rahmstorf S, Robinson A, Feulner G, Saba V. 2018. Observed fingerprint of a weakening Atlantic Ocean overturning circulation. Nature. 556(7700):191–196. doi:10.1038/s41586-018-0006-5.

Cammen KM, Andrews KR, Carroll EL, Foote AD, Humble E, Khudyakov JI, Louis M, McGowen MR, Olsen MT, Van Cise AM. 2016. Genomic methods take the plunge: Recent advances in high-throughput sequencing of marine mammals. J Hered. 107(6):481–495. doi:10.1093/jhered/esw044.

Cañadas A. 2012. *Ziphius cavirostris* (Mediterranean subpopulation), Cuvier's Beaked whale. IUCN Red List Threat Species 2012 eT16381144A16382769.

Cañadas A, Aguilar de Soto N, Aissi M, Arcangeli A, Azzolin M, B-Nagy A, Bearzi G, Campana I, Chicote C, Cotte C, et al. 2018. The challenge of habitat modelling for threatened low density species using heterogeneous data: the case of Cuvier's beaked whales in the Mediterranean. Ecol Indic. 85:128–136. doi:10.1016/j.ecolind.2017.10.021.

Cañadas A, Vázquez JA. 2014. Conserving Cuvier's beaked whales in the Alborán Sea (SW Mediterranean): Identification of high density areas to be avoided by intense man-made sound. Biol Conserv. 178:155–162. doi:10.1016/j.biocon.2014.07.018.

Carøe C, Gopalakrishnan S, Vinner L, Mak SST, Sinding M-HS, Samaniego JA, Wales N, Sicheritz-Pontén T, Gilbert MTP. 2018. Single-tube library preparation for degraded DNA. Methods Ecol Evol. 9(2):410–419. doi:10.1111/2041-210X.12871.

Carroll EL, Bruford MW, DeWoody JA, Leroy G, Strand A, Waits L, Wang J. 2018. Genetic and genomic monitoring with minimally invasive sampling methods. Evol Appl. 11(7):1094–1119. doi:10.1111/eva.12600.

Carroll EL, Gaggiotti OE. 2019. Making use of the social network in conservation genomics: Integrating kinship and network analyses to understand connectivity. Mol Ecol Resour. 19:307–309. doi:10.1111/1755-0998.12950.

Carroll EL, Garland EC. 2022. Viewing the Lives of Whales Through a Molecular Lens. In: Clark CW, Garland EC, editors. Ethology and Behavioural Ecology of Mysticetes.

Springer Nature. p. 125–146.

Carroll EL, McGowen MR, McCarthy ML, Marx FG, Aguilar de Soto N, Dalebout ML, Nielsen SD, Gaggiotti OE, Hansen SS, Van Helden AL, et al. 2021. Speciation in the deep: Genomics and morphology reveal a new species of beaked whale *Mesoplodon eueu*. Proc R Soc B Biol Sci. 288(1961). doi:10.1098/rspb.2021.1213.

Carroll EL, Reyes C, Gaggiotti OE, Olsen MT, Maaholm DJ, Rosso M, Davison NJ, Martín V, Schiavi A, Aguilar de Soto N. 2016. Pilot study to assess the utility of ddRAD sequencing in identifying species-specific and shared SNPs among Blainville's (*Mesoplodon densirostris*) and Cuvier's (*Ziphius cavirostris*) beaked whales. Int Whal Comm Sci Meet 66.(SC/66b/DNA/03):1–7. doi:10.13140/RG.2.1.2286.5527.

Castro ALF, Stewart BS, Wilson SG, Hueter RE, Meekan MG, Motta PJ, Bowen BW, Karl SA. 2007. Population genetic structure of Earth's largest fish, the whale shark (*Rhincodon typus*). Mol Ecol. 16(24):5183–5192. doi:10.1111/j.1365-294X.2007.03597.x.

Catchen JM, Amores A, Hohenlohe PA, Cresko WA, Postlethwait JH. 2011. Stacks: Building and genotyping loci de novo from short-read sequences. G3. 1:171–182. doi:10.1534/g3.111.000240.

Catchen JM, Hohenlohe PA, Bassham S, Amores A, Cresko WA. 2013. Stacks: An analysis tool set for population genomics. Mol Ecol. 22:3124–3140. doi:10.1111/mec.12354.

Caye K, Deist TM, Martins H, Michel O, François O. 2016. TESS3: Fast inference of spatial population structure and genome scans for selection. Mol Ecol Resour. 16(2):540–548. doi:10.1111/1755-0998.12471.

Caye K, Jay F, Michel O, Francois O. 2018. Fast inference of individual admixture coefficients using geographic data. Ann Appl Stat. 12(1):586–608. doi:https://doi.org/10.1214/17-AOAS1106.

Charlesworth B. 2009. Fundamental concepts in genetics: Effective population size and patterns of molecular evolution and variation. Nat Rev Genet. 10(3):195–205. doi:10.1038/nrg2526.

Charlesworth B. 2010. Molecular population genomics: A short history. Genet Res (Camb). 92:397–411. doi:10.1017/S0016672310000522.

Chen C, Durand E, Forbes F, François O. 2007. Bayesian clustering algorithms ascertaining spatial population structure: a new computer program and a comparison study. Mol Ecol Notes. 7(5):747–756. doi:10.1111/j.1471-8286.2007.01769.x.

Chikhi L, Sousa VC, Luisi P, Goossens B, Beaumont MA. 2010. The confounding effects of population structure, genetic diversity and the sampling scheme on the detection and quantification of population size changes. Genetics. 186(3):983–995. doi:10.1534/genetics.110.118661.

Claridge DE. 2006. Fine-scale distribution and habitat selection of beaked whales. MSc Thesis. University of Aberdeen, Scotland.

Claridge DE. 2013. Population ecology of Blainville's beaked whales (*Mesoplodon densirostris*). PhD thesis. University of St Andrews, Scotland.

Claridge DE, Dunn CA, Ylitalo G, Herman D, Parsons KM, Durban JW. 2015. Behavioral ecology of deep-diving odontocetes in the Bahamas. Final Report of Project

RC-2114. Department of Defense Strategic Environmental Research and Development Program.

Clark PU, Archer D, Pollard D, Blum JD, Rial JA, Brovkin V, Mix AC, Pisias NG, Roy M. 2006. The middle Pleistocene transition: characteristics, mechanisms, and implications for long-term changes in atmospheric pCO2. Quat Sci Rev. 25(23–24):3150–3184. doi:10.1016/j.quascirev.2006.07.008.

Clark PU, Dyke AS, Shakun JD, Carlson AE, Clark J, Wohlfarth B, Mitrovica JX, Hostetler SW, McCabe AM. 2009. The Last Glacial Maximum. Science. 325:710–714. doi:10.1126/science.1172873.

Clarke CR, Karl SA, Horn RL, Bernard AM, Lea JS, Hazin FH, Prodöhl PA, Shivji MS. 2015. Global mitochondrial DNA phylogeography and population structure of the silky shark, *Carcharhinus falciformis*. Mar Biol. 162(5):945–955. doi:10.1007/s00227-015-2636-6.

Colleoni F, Wekerle C, Näslund J-O, Brandefelt J, Masina S. 2016. Constraint on the penultimate glacial maximum Northern Hemisphere ice topography (≈140 kyrs BP). Quat Sci Rev. 137:97–112. doi:10.1016/j.quascirev.2016.01.024.

Commitee on Taxonomy. 2022. List of marine mammal species and subspecies. Society for Marine Mammalogy, www.marinemammalscience.org, consulted on 18 August, 2020.

Coomber F, Moulins A, Tepsich P, Rosso M. 2016. Sexing free-ranging adult Cuvier's beaked whales (*Ziphius cavirostris*) using natural marking thresholds and pigmentation patterns. J Mammal. 97(3):879–890. doi:10.1093/jmammal/gyw033.

COSEWIC. 2006. COSEWIC assessment and update status report on the Sowerby's beaked whale *Mesoplodon bidens* in Canada. Comm Status Endanger Wildl Canada Ottawa vi + 20 pp (www.sararegistry.gc.ca/status/status_e.cfm).

COSEWIC. 2011. COSEWIC assessment and status report on the Northern Bottlenose Whale *Hyperoodon ampullatus* in Canada. Comm Status Endanger Wildl Canada Ottawa xii + 31 pp.

Costello MJ, Tsai P, Wong PS, Cheung AKL, Basher Z, Chaudhary C. 2017. Marine biogeographic realms and species endemicity. Nat Commun. 8(1):1–9. doi:10.1038/s41467-017-01121-2.

Cox TM, Ragen TJ, Read AJ, Vos E, Baird RW, Balcomb III KC, Barlow JP, Caldwell J, Cranford TW, Crum L, et al. 2006. Understanding the impacts of anthropogenic sound on beaked whales. J Cetacean Res Manag. 7(3):177–187. doi:10.1109/LPT.2009.2020494.

Csárdi G, Nepusz T. 2006. The igraph software package for complex network research. InterJournal. Complex Sy:1695. doi:10.3724/sp.j.1087.2009.02191.

Curtis KA, Falcone EA, Schorr GS, Moore JE, Moretti DJ, Barlow JP, Keene E. 2021. Abundance, survival, and annual rate of change of Cuvier's beaked whales (*Ziphius cavirostris*) on a Navy sonar range. Mar Mammal Sci. 37(2):399–419. doi:10.1111/mms.12747.

Cuvier G. 1824. Recherches sur les ossements fossiles.

D'Amico A, Gisiner RC, Ketten DR, Hammock JA, Johnson C, Tyack PL, Mead JG. 2009. Beaked whale strandings and naval exercises. Aquat Mamm. 35(4):452–472.

doi:10.1578/AM.35.4.2009.452.

Dalebout ML. 2002. Species identity, genetic diversity and molecular systematic relationships among the Ziphiidae (beaked whales). PhD Thesis. University of Auckland, New Zealand.

Dalebout ML, Baker CS, Mead JG, Cockcroft VG, Yamada TK. 2004. A comprehensive and validated molecular taxonomy of beaked whales, family Ziphiidae. J Hered. 95(6):459–473. doi:10.1093/jhered/esh054.

Dalebout ML, Baker CS, Steel DJ, Robertson KM, Chivers SJ, Perrin WF, Mead JG, Grace R V., Schofield Jr. TD. 2007. A divergent mtDNA lineage among *Mesoplodon* beaked whales: Molecular evidence for a new species in the tropical pacific? Mar Mammal Sci. 23(4):954–966. doi:10.1111/j.1748-7692.2007.00143.x.

Dalebout ML, Baker CS, Steel DJ, Thompson KF, Robertson KM, Chivers SJ, Perrin WF, Goonatilake M, Anderson RC, Mead JG, et al. 2014. Resurrection of *Mesoplodon hotaula* Deraniyagala 1963: a new species of beaked whale in the tropical Indo-Pacific. Mar Mammal Sci. 30(3):1081–1108. doi:10.1111/mms.12113.

Dalebout ML, Hooker SK, Christensen I. 2001. Genetic diversity and population structure among northern bottlenose whales, *Hyperoodon ampullatus*, in the western North Atlantic Ocean. Can J Zool. 79(3):478–484. doi:10.1139/cjz-79-3-478.

Dalebout ML, Mead JG, Baker CS, Baker AN, van Helden AL. 2002. A new species of beaked whale *Mesoplodon perrini* sp. n. (Cetacea: Ziphiidae) discovered through phylogenetic analyses of mitochondrial DNA sequences. Mar Mammal Sci. 18(3):577–608. doi:10.1111/j.1748-7692.2002.tb01061.x.

Dalebout ML, Robertson KM, Frantzis A, Engelhaupt D, Mignucci-Giannoni AA, Rosario-Delestre RJ, Baker CS. 2005. Worldwide structure of mtDNA diversity among Cuvier's beaked whales (*Ziphius cavirostris*): Implications for threatened populations. Mol Ecol. 14(11):3353–3371. doi:10.1111/j.1365-294X.2005.02676.x.

Dalebout ML, Ross GJB, Baker CS, Anderson RC, Best PB, Cockcroft VG, Hinz HL, Peddemors V, Pitman RL. 2003. Appearance, distribution, and genetic distinctiveness of Longman's beaked whale, *Indopacetus pacificus*. Mar Mammal Sci. 19(July):421–461.

Dalebout ML, Ruzzante DE, Whitehead H, Øien NI. 2006. Nuclear and mitochondrial markers reveal distinctiveness of a small population of bottlenose whales (*Hyperoodon ampullatus*) in the western North Atlantic. Mol Ecol. 15(11):3115–3129. doi:10.1111/j.1365-294X.2006.03004.x.

Dalebout ML, Steel DJ, Baker CS. 2008. Phylogeny of the beaked whale genus *Mesoplodon* (Ziphiidae: Cetacea) revealed by nuclear introns: Implications for the evolution of male tusks. Syst Biol. 57(6):857–875. doi:10.1080/10635150802559257.

Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, et al. 2011. The variant call format and VCFtools. Bioinformatics. 27(15):2156–2158. doi:10.1093/bioinformatics/btr330.

Darwin C. 1859. The Origin of Species. Gramercy Books, 1979.

DeRuiter SL, Southall BL, Calambokidis J, Zimmer WMX, Sadykova D, Falcone EA, Friedlaender AS, Joseph JE, Moretti DJ, Schorr GS, et al. 2013. First direct measurements of behavioural responses by Cuvier's beaked whales to mid-frequency active sonar. Biol Lett. 9(4):20130223–20130223. doi:10.1098/rsbl.2013.0223.

Dewoody JA, Fernandez NB, Brüniche-Olsen A, Antonides JD, Doyle JM, Miguel PS, Westerman R, Vertyankin V V., Godard-Codding CAJ, Bickham JW. 2017. Characterization of the gray whale *Eschrichtius robustus* genome and a genotyping array based on single-nucleotide polymorphisms in candidate genes. Biol Bull. 232(3):186–197. doi:10.1086/693483.

DeWoody JA, Harder AM, Mathur S, Willoughby JR. 2021. The long-standing significance of genetic diversity in conservation. Mol Ecol. 30:4147–4154. doi:10.1111/mec.16051.

DeYoung R., Honeycutt R. 2005. The Molecular Toolbox: Genetic techniques in wildlife ecology and management. J Wildl Manag. 69(4):1362–1384.

De Dinechin M, Ottvall R, Quillfeldt P, Jouventin P. 2009. Speciation chronology of rockhopper penguins inferred from molecular, geological and palaeoceanographic data. J Biogeogr. 36(4):693–702. doi:10.1111/j.1365-2699.2008.02014.x.

Do C, Waples RS, Peel D, Macbeth GM, Tillett BJ, Ovenden JR. 2014. NeEstimator v2: Re-implementation of software for the estimation of contemporary effective population size (Ne) from genetic data. Mol Ecol Resour. 14(1):209–214. doi:10.1111/1755-0998.12157.

Dolman SJ, Pinn E, Reid RJ, Barley JP, Deaville R, Jepson PD, O'Connell M, Berrow S, Penrose RS, Stevick PT, et al. 2010. A note on the unprecedented strandings of 56 deep-diving whales along the UK and Irish coast. Mar Biodivers Rec. 3:1–8. doi:10.1017/s175526720999114x.

Dornburg A, Brandley MC, McGowen MR, Near TJ. 2012. Relaxed clocks and inferences of heterogeneous patterns of nucleotide substitution and divergence time estimates across whales and dolphins (Mammalia: Cetacea). Mol Biol Evol. 29(2):721–736. doi:10.1093/molbev/msr228.

Dray S, Dufour AB. 2007. The ade4 package: Implementing the duality diagram for ecologists. J Stat Softw. 22(4):1–20. doi:10.18637/jss.v022.i04.

Drinkwater KF, Belgrano A, Borja A, Conversi A, Edwards M, Greene CH, Ottersen G, Pershing AJ, Walker H. 2003. The response of marine ecosystems to climate variability associated with the North Atlantic oscillation. Geophys Monogr Ser. 134:263–279. doi:10.1029/134GM12.

Drummond AJ, Rambaut A. 2007. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol Biol. 7(1):1–8. doi:10.1186/1471-2148-7-214.

Drummond AJ, Rambaut A, Shapiro B, Pybus OG. 2005. Bayesian coalescent inference of past population dynamics from molecular sequences. Mol Biol Evol. 22(5):1185–1192. doi:10.1093/molbev/msi103.

Duchêne S, Archer FI, Vilstrup J, Caballero S, Morin PA. 2011. Mitogenome phylogenetics: the impact of using single regions and partitioning schemes on topology, substitution rate and divergence time estimation. PLoS One. 6(11). doi:10.1371/journal.pone.0027138.

Dudchenko O, Batra SS, Omer AD, Nyquist SK, Hoeger M, Durand NC, Shamim MS, Machol I, Lander ES, Aiden AP, et al. 2017. De novo assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds. Science. 356:92–95. doi:10.1126/science.aal3327.

Dudchenko O, Shamim MS, Batra SS, Durand NC, Musial NT, Mostofa R, Pham M,

Hilaire BGS, Yao W, Stamenova E, et al. 2018. The Juicebox Assembly Tools module facilitates de novo assembly of mammalian genomes with chromosome-length scaffolds for under $1000. bioRxiv.:254797. doi:http://dx.doi.org/10.1101/254797.

Dunn CA. 2014. Insights into Blainville's beaked whale (*Mesoplodon densirostris*) communication. PhD Thesis. University of St Andrews, Scotland.

Durand E, Jay F, Gaggiotti OE, François O. 2009. Spatial inference of admixture proportions and secondary contact zones. Mol Biol Evol. 26(9):1963–1973. doi:10.1093/molbev/msp106.

Duvernoy M. 1851. Memoire sure les characteres osteologiques des genres nouveaus ou des especes nouvelles de cetaces vivants out fossiles. Ann des Sci Nat. 15:1–71.

Eaton DAR. 2014. PyRAD: Assembly of *de novo* RADseq loci for phylogenetic analyses. Bioinformatics. 30(13):1844–1849. doi:10.1093/bioinformatics/btu121.

Ellis R, Mead JG. 2017a. Sowerby's Beaked Whale. In: Beaked Whales: A Complete Guide to Their Biology and Conservation. Johns Hopkins University Press.

Ellis R, Mead JG. 2017b. Northern Bottlenose Whale. In: Beaked Whales: A Complete Guide to Their Biology and Conservation. Johns Hopkins University Press.

Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, Mitchell SE. 2011. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS One. 6(5):1–10. doi:10.1371/journal.pone.0019379.

Europa Press. 2013. Encuentran el cadáver de un zifio común varado en la playa de La Caleta, en Fuerteventura. 20minutos. https://www.20minutos.es/noticia/1731237/0/.

Evanno G, Regnaut S, Goudet J. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. Mol Ecol. 14(8):2611–2620. doi:10.1111/j.1365-294X.2005.02553.x.

Evans DL, England GR, Lautenbacher Jr. CC, Livingstone SM, Hogarth WT, Johnson HT. 2001. Joint Interim Report, Bahamas Marine Mammal Stranding Event of 15-16 March 2000.

Fahlman A, Tyack PL, Miller PJO, Kvadsheim PH. 2014. How man-made interference might cause gas bubble emboli in deep diving whales. Front Physiol. 5(13):1–6. doi:10.3389/fphys.2014.00013.

Falcone EA, Schorr GS, Douglas AB, Calambokidis J, Henderson E, McKenna MF, Hildebrand J, Moretti D. 2009. Sighting characteristics and photo-identification of Cuvier's beaked whales (*Ziphius cavirostris*) near San Clemente Island, California: A key area for beaked whales and the military? Mar Biol. 156(12):2631–2640. doi:10.1007/s00227-009-1289-8.

Fan G, Zhang Y, Liu Xiaochuan, Wang Jiahao, Sun Z, Sun S, Zhang H, Chen J, Lv M, Han K, et al. 2019. The first chromosome-level genome for a marine mammal as a resource to study ecology and evolution. Mol Ecol Resour. 19:944–956. doi:10.1111/1755-0998.13003.

Fedutin ID, Filatova OA, Mamaev EG, Burdin AM, Hoyt E. 2015. Occurrence and social structure of Baird's beaked whales, *Berardius bairdii*, in the Commander Islands, Russia. Mar Mammal Sci. 31(3):853–865. doi:10.1111/mms.12204.

Ferguson MC, Barlow JP, Reilly SB, Gerrodette T. 2006. Predicting Cuvier's (*Ziphius cavirostris*) and *Mesoplodon* beaked whale population density from habitat

characteristics in the eastern tropical Pacific Ocean. J Cetacean Res Manag. 7(3):287–299.

Fernández A, Edwards JF, Rodríguez F, Espinosa de los Monteros A, Herráez P, Castro P, Jabert JR, Martín V, Arbelo M. 2005. 'Gas and fat embolic syndrome' involving a mass stranding of beaked whales (Family Ziphiidae) exposed to anthropogenic sonar signals. Vet Pathol. 42:446–457. doi:10.1354/vp.42-4-446.

Fernández A, Martín V. 2013. No mass strandings since sonar ban. Nature. 497:317. doi:https://doi.org/10.1038/497317d.

Fernández A, Sierra E, Martín V, Mendez M, Méndez M, Sacchini S, Bernaldo de Quirós Y, Andrada M, Rivero M, Quesada O, et al. 2012. Last 'atypical' beaked whales mass stranding in the Canary Islands (July, 2004). J Mar Sci Res Dev. 2(2). doi:10.4172/2155-9910.1000107.

Feyrer LJ, Bentzen P, Whitehead H, Paterson IG, Einfeldt AL. 2019. Evolutionary impacts differ between two exploited populations of northern bottlenose whale (*Hyperoodon ampullatus*). Ecol Evol. 9(23):13567–13584. doi:10.1002/ece3.5813.

Filadelfo R, Mintz J, Michlovich E, D'Amico A, Tyack PL, Ketten DR. 2009. Correlating military sonar use with beaked whale mass strandings: What do the historical data show? Aquat Mamm. 35(4):435–444. doi:10.1578/AM.35.4.2009.435.

Fisher RA. 1922. On the dominance ratio. Proc R Soc Edinburgh. 42:321–341.

Fitak RR, Johnsen S. 2018. Green sea turtle (*Chelonia mydas*) population history indicates important demographic changes near the mid-Pleistocene transition. Mar Biol. 165(7):1–7. doi:10.1007/s00227-018-3366-3.

Fleming AH, Clark CT, Calambokidis J, Barlow JP. 2016. Humpback whale diets respond to variance in ocean climate and ecosystem conditions in the California Current. Glob Chang Biol. 22(3):1214–1224. doi:10.1111/gcb.13171.

da Fonseca RR, Albrechtsen A, Themudo GE, Ramos-Madrigal J, Sibbesen JA, Maretty L, Zepeda-Mendoza ML, Campos PF, Heller R, Pereira RJ. 2016. Next-generation biology: Sequencing and data analysis approaches for non-model organisms. Mar Genomics. 30:3–13. doi:10.1016/j.margen.2016.04.012.

Foote AD, Vijay N, Ávila-Arcos MC, Baird RW, Durban JW, Fumagalli M, Gibbs RA, Hanson MB, Korneliussen TS, Martin MD, et al. 2016. Genome-culture coevolution promotes rapid divergence of killer whale ecotypes. Nat Commun. 7(May). doi:10.1038/ncomms11693.

Fordyce RE. 2018. Cetacean evolution. In: Wursig B, Thewissen JGM, Kovacs KM, editors. Encyclopedia of Marine Mammals, 3rd Edition. Academic Press. p. 180–185.

Frankham R, Bradshaw CJA, Brook BW. 2014. Genetics in conservation management: Revised recommendations for the 50/500 rules, Red List criteria and population viability analyses. Biol Conserv. 170:56–63. doi:10.1016/j.biocon.2013.12.036.

Frantzis A. 1998. Does acoustic testing strand whales? Nature. 392(6671):29. doi:10.1038/32068.

Frantzis A. 2015. Short report on the mass stranding of Cuvier's beaked whales that occurred on the 1st of April 2014 in South Crete, Greece euring naval exercises. FINS. 6(1):10–11.

Frantzis A, Koutouzi N, Gkikopoulou KC, Alexiadou P, Moschopoulos E. 2022. Results

synthesis. Deliverable C3 of project: 'Sampling, analysis and data synthesis for the assessment of Good Environmental Status (GES) of MSFD Descriptor 1 (biodiversity) for cetaceans', Green2Sustain-G. Tentes-A.Frantzis. Athens.

Freamo H, O'reilly P, Berg PR, Lien S, Boulding EG. 2011. Outlier SNPs show more genetic structure between two Bay of Fundy metapopulations of Atlantic salmon than do neutral SNPs. Mol Ecol Resour. 11(SUPPL. 1):254–267. doi:10.1111/j.1755-0998.2010.02952.x.

Freer JJ, Daase M, Tarling GA. 2022. Modelling the biogeographic boundary shift of *Calanus finmarchicus* reveals drivers of Arctic Atlantification by subarctic zooplankton. Glob Chang Biol. 28(2):429–440. doi:10.1111/gcb.15937.

Gaither MR, Bowen BW, Rocha LA, Briggs JC. 2016. Fishes that rule the world: circumtropical distributions revisited. Fish Fish. 17(3):664–679. doi:10.1111/faf.12136.

Garner BA, Hoban S, Luikart G. 2020. IUCN Red List and the value of integrating genetics. Conserv Genet. 21(5):795–801. doi:10.1007/s10592-020-01301-6.

Gascuel O. 1997. BIONJ: An improved version of the NJ algorithm based on a simple model of sequence data. Mol Biol Evol. 14(7):685–695. doi:10.1093/oxfordjournals.molbev.a025808.

Gaspari S, Airoldi S, Hoelzel AR. 2007. Risso's dolphins (*Grampus griseus*) in UK waters are differentiated from a population in the Mediterranean Sea and genetically less diverse. Conserv Genet. 8(3):727–732. doi:10.1007/s10592-006-9205-y.

Gaspari S, Marsili L, Natali C, Airoldi S, Lanfredi C, Deeming C, Moura AE. 2019. Spatio-temporal patterns of genetic diversity in the Mediterranean striped dolphin (*Stenella coeruleoalba*). J Zool Syst Evol Res. 57(3):721–734. doi:10.1111/jzs.12265.

Gkafas GA, Exadactylos A, Rogan E, Raga JA, Reid R, Hoelzel AR. 2017. Biogeography and temporal progression during the evolution of striped dolphin population structure in European waters. J Biogeogr. 44(12):2681–2691. doi:10.1111/jbi.13079.

Golikov A V., Sabirov RM, Lubin PA, Jørgensen LL. 2013. Changes in distribution and range structure of Arctic cephalopods due to climatic changes of the last decades. Biodiversity. 14(1):28–35. doi:10.1080/14888386.2012.702301.

Gomerčić H, Duras Gomerčić M, Gomerčić T, Lucić H, Dalebout ML, Galov A, Škrtić D, Ćurković S, Vuković S, Huber D. 2006. Biological aspects of Cuvier's beaked whale (*Ziphius cavirostris*) recorded in the Croatian part of the Adriatic Sea. Eur J Wildl Res. 52(3):182–187. doi:10.1007/s10344-006-0032-8.

Gonçalves da Silva A, Barendse W, Kijas J, England PR, Hoelzel AR. 2020. Genomic data suggest environmental drivers of fish population structure in the deep sea: A case study for the orange roughy (*Hoplostethus atlanticus*). J Appl Ecol. 57(2):296–306. doi:10.1111/1365-2664.13534.

Goudet J. 2005. HIERFSTAT, a package for R to compute and test hierarchical F-statistics. Mol Ecol Notes. 5:184–186. doi:10.1111/j.1471-8286.2004.00828.x.

Gowans S, Rendell LE. 1999. Head-butting in northern bottlenose whales (*Hyperoodon ampullatus*): a possible function for big heads? Mar Mammal Sci. 15(4):1342–1350.

Gowans S, Whitehead H, Hooker SK. 2001. Social organization in northern bottlenose whales, *Hyperoodon ampullatus*: not driven by deep-water foraging? Anim Behav. 62:369–377. doi:10.1006/anbe.2001.1756.

Graham RM, De Boer AM. 2013. The dynamical subtropical front. J Geophys Res Ocean. 118:5676–5685. doi:10.1002/jgrc.20408.

Gray JE. 1860. On the genus *Hyperoodon*: the two British kinds and their food. Proc Zool Soc London.:422–426.

de Greef E, Einfeldt AL, Miller PJO, Ferguson SH, Garroway CJ, Lefort KJ, Paterson IG, Bentzen P, Feyrer LJ. 2022. Genomics reveal population structure, evolutionary history, and signatures of selection in the northern bottlenose whale, *Hyperoodon ampullatus*. Mol Ecol.:3. doi:10.1111/mec.16643.

Gruber B, Unmack PJ, Berry OF, Georges A. 2018. dartr: An r package to facilitate analysis of SNP data generated from reduced representation genome sequencing. Mol Ecol Resour. 18(3):691–699. doi:10.1111/1755-0998.12745.

von Haast J. 1876. On a New Ziphioid Whale. Proc R Soc London.:7–13.

Hancock-Hanser BL, Frey A, Leslie MS, Dutton PH, Archer FI, Morin PA. 2013. Targeted multiplex next-generation sequencing: Advances in techniques of mitochondrial and nuclear DNA sequencing for population genomics. Mol Ecol Resour. 13(2):254–268. doi:10.1111/1755-0998.12059.

Hansson B, Westerberg L. 2002. On the correlation between heterozygosity and fitness in natural populations. Mol Ecol. 11:2467–2474. doi:10.1046/j.1365-294x.2002.01644.x.

Hardy GH. 1908. Mendelian proportions in a mixed population. Science.:49–50.

Harlin-Cognato AD, Markowitz T, Würsig B, Honeycutt RL. 2007. Multi-locus phylogeography of the dusky dolphin (*Lagenorhynchus obscurus*): Passive dispersal via the west-wind drift or response to prey species and climate change? BMC Evol Biol. 7:1–17. doi:10.1186/1471-2148-7-131.

Harrison RG. 1989. Animal mitochondrial DNA as a genetic marker in population and evolutionary biology. Trends Ecol Evol. 4(1):6–11. doi:10.1016/0169-5347(89)90006-2.

Harrison T. 2017. Miocene Primates. In: Fuentes A, editor. The International Encyclopedia of Primatology. John Wiley & Sons, Inc.

Harwood J, King S, Booth C, Donovan C, Schick RS, Thomas L, New L. 2016. Understanding the population consequences of acoustic disturbance for marine mammals. Adv Exp Med Biol. 875:417–423. doi:10.1007/978-1-4939-2981-8_49.

Hauser DDW, Laidre KL, Stafford KM, Stern HL, Suydam RS, Richard PR. 2017. Decadal shifts in autumn migration timing by Pacific Arctic beluga whales are related to delayed annual sea ice formation. Glob Chang Biol. 23(6):2206–2217. doi:10.1111/gcb.13564.

Hazen EL, Abrahms B, Brodie S, Carroll G, Jacox MG, Savoca MS, Scales KL, Sydeman WJ, Bograd SJ. 2019. Marine top predators as climate and ecosystem sentinels. Front Ecol Environ. 17(10):565–574. doi:10.1002/fee.2125.

Heithaus MR, Frid A. 2003. Optimal diving under the risk of predation. J Theor Biol. 223(1):79–92. doi:10.1016/S0022-5193(03)00073-0.

Heller R, Chikhi L, Siegismund HR. 2013. The confounding effect of population structure on Bayesian skyline plot inferences of demographic history. PLoS One. 8(5). doi:10.1371/journal.pone.0062992.

Henderson EE, Martin SW, Manzano-Roth R, Matsuyama BM. 2016. Occurrence and

habitat use of foraging Blainville's beaked whales (*Mesoplodon densirostris*) on a U.S. Navy range in Hawaii. Aquat Mamm. 42(4):549–562. doi:10.1578/AM.42.4.2016.549.

Henshaw MD, Leduc RG, Chivers SJ, Dizon AE. 1997. Identifying Beaked whales (family Ziphidae) using mtDNA sequences. Mar Mammal Sci. 13(3):487–495. doi:10.1111/j.1748-7692.1997.tb00656.x.

Heyning JE. 1984. Functional morphology involved in intraspecific fighting of the beaked whale, *Mesoplodon carlhubbsi*. Can J Zool. 62:1645–1654. doi:10.1139/z84-239.

Heyning JE. 1989. Cuvier's Beaked Whale. In: Ridgway SH, Harrison SR, editors. Handbook of Marine Mammals Volume 4: River Dolphins and Larger Toothed Whales. p. 289–308.

Heyning JE, Mead JG. 1996. Suction feeding in beaked whales: Morphological and observational evidence. Contrib Sci. 464. doi:https://doi.org/10.5962/p.226802.

Heyning JE, Mead JG. 2008. Cuvier's Beaked Whale, *Ziphius Cavirostris*. In: Perrin WF, Wursig B, Thewissen JGM, editors. Encyclopedia of Marine Mammals, 2nd Edition. Academic Press. p. 294–295.

Hildebrand JA. 2009. Anthropogenic and natural sources of ambient noise in the ocean. Mar Ecol Prog Ser. 395:5–20. doi:10.3354/meps08353.

Hoban S, Archer FI, Bertola LD, Bragg JG, Breed MF, Bruford MW, Coleman MA, Ekblom R, Funk WC, Grueber CE, et al. 2022. Global genetic diversity status and trends: towards a suite of Essential Biodiversity Variables (EBVs) for genetic composition. Biol Rev. 97(4):1511–1538. doi:10.1111/brv.12852.

Hoban S, Bertorelle G, Gaggiotti OE. 2012. Computer simulations: Tools for population and evolutionary genetics. Nat Rev Genet. 13(2):110–122. doi:10.1038/nrg3130.

Hoegh-Guldberg O, Cai R, Poloczanska ES, Brewer PG, Sundby S, Hilmi K, Fabry VJ, Jung S. 2014. The Ocean. In: Barros VR, Field CB, Dokken DJ, Mastrandrea MD, Mach KJ, Bilir TE, Chatterjee M, Ebi KL, Estrada YO, Genova RC, et al., editors. Climate Change 2014: Impacts, Adaptaiton and Vulnerability, Part B: Revional Aspects, Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA. p. 1655–1731.

Hoelzel AR. 2009. Evolution of population genetic structure in marine mammal species. In: Bertorelle G, Bruford MW, Hauffe HC, Rizzoli A, Vernesi C, editors. Population Genetics for Animal Conservation. Cambridge University Press. p. 294–318.

Hoelzel AR. 2018. Molecular Ecology. In: Wursig B, Thewissen JGM, Kovacs KM, editors. Encyclopedia of Marine Mammals, 3rd Edition. Academic Press. p. 613–618. http://doi.wiley.com/10.1002/9780470015902.a0003268.pub2.

Hohenlohe PA, Funk WC, Rajora OP. 2021. Population genomics for wildlife conservation and management. Mol Ecol. 30:62–82. doi:10.1111/mec.15720.

Hooker SK, Aguilar de Soto N, Baird RW, Carroll EL, Claridge DE, Feyrer LJ, Miller PJO, Onoufriou AB, Schorr GS, Siegal E, et al. 2019. Future directions in research on beaked whales. Front Mar Sci. 5:1–16. doi:10.3389/fmars.2018.00514.

Hooker SK, Baird RW. 1999a. Deep-diving behaviour of the northern bottlenose whale, *Hyperoodon ampullatus* (Cetacea: Ziphiidae). Proc R Soc B Biol Sci. 266:671–676. doi:https://doi.org/10.1098/rspb.1999.0688.

Hooker SK, Baird RW. 1999b. Observations of Sowerby's Beaked Whales, *Mesoplodon bidens*, in the Gully , Nova Scotia. Can F Nat. 113(2):273–277.

Hudson RR, Slatkin M, Maddison WP. 1992. Estimation of levels of gene flow from DNA sequence data. Genetics. 132(2):583–589. doi:10.1093/genetics/132.2.583.

Hurlbert SH. 1971. The nonconcept of species diversity: a critique and alternative parameters. Ecology. 52(4):577–586. doi:10.2307/1934145.

Hurrell JW, Deser C. 2010. North Atlantic climate variability: The role of the North Atlantic Oscillation. J Mar Syst. 79(3–4):231–244. doi:10.1016/j.jmarsys.2009.11.002.

Illumina. 2017. DNA Sequencing Methods Collection. In: Scientific Publication Reviews. www.illumina.com/pubreviews.

IUCN (International Union for Conservation of Nature). 2008a. *Hyperoodon ampullatus* (Spatial data). IUCN Red List Threat Species.

IUCN (International Union for Conservation of Nature). 2008b. *Mesoplodon bidens* (Spatial data). IUCN Red List Threat Species.

IUCN (International Union for Conservation of Nature). 2012a. *Ziphius cavirostris* (Spatial data). IUCN Red List Threat Species.

IUCN (International Union for Conservation of Nature). 2012b. *Mesoplodon densirostris* (Spatial data). IUCN Red List Threat Species.

IUCN Marine Mammal Protected Areas Task Force. 2017. Hellenic Trench IMMA. Full Accounts of Mediterranean IMMA Factsheet. IUCN Joint SSC/WCPA Marine Mammal Protected Areas Task Force, 2017. www.marinemammalhabitat.org/portfolio-item/hellenic-trench/.

IUCN Standards and Petitions Committee. 2019. Guidelines for using the IUCN Red List categories and criteria. http://intranet.iucn.org/webfiles/doc/SSC/RedList/RedListGuidelines.pdf.

Jackson JA, Carroll EL, Smith TD, Zerbini AN, Patenaude NJ, Baker CS. 2016. An integrated approach to historical population assessment of the great whales: case of the New Zealand southern right whale. R Soc Open Sci. 3. doi:https://doi.org/10.1098/rsos.150669.

Jasny M, Reynolds J, Horowitz C, Wetzler A. 2005. Sounding the Depths II: the rising toll of sonar, shipping and industrial ocean noise on marine life. Report to Natural Resources Defence Council.

Jefferson TA. 2018. Rough-Toothed Dolphin, *Steno bredanensis*. In: Wursig B, Thewissen JGM, Kovacs KM, editors. Encyclopedia of Marine Mammals, 3rd Edition. Academic Press. p. 838–840.

Jenkins TL, Castilho R, Stevens JR. 2018. Meta-analysis of northeast Atlantic marine taxa shows contrasting phylogeographic patterns following post-LGM expansions. PeerJ. 2018(9). doi:10.7717/peerj.5684.

Jepson PD, Arbelo M, Deaville R, Patterson IAP, Castro P, Baker JR, Degollada E, Ross HM, Herráez P, Pocknell AM, et al. 2003. Gas-bubble lesions in stranded cetaceans. Nature. 425:575. http://dx.doi.org/10.1038/425575a.

Johnson MP, Madsen PT, Zimmer WMX, Aguilar de Soto N, Tyack PL. 2004. Beaked whales echolocate on prey. Proc R Soc B Biol Sci. 271:S383–S386.

doi:10.1098/rsbl.2004.0208.

Johnson MP, Tyack PL. 2003. A digital acoustic recording tag for measuring the response of wild marine mammals to sound. IEEE J Ocean Eng. 28(1):3–12. doi:10.1109/JOE.2002.808212.

Johnson NA, Lachance J. 2012. The genetics of sex chromosomes: evolution and implications for hybrid incompatibility. Ann N Y Acad Sci. 1256:E1–E22. doi:10.1111/j.1749-6632.2012.06748.x.

Jombart T. 2008. Adegenet: A R package for the multivariate analysis of genetic markers. Bioinformatics. 24(11):1403–1405. doi:10.1093/bioinformatics/btn129.

Jombart T, Collins C. 2015. A tutorial for Discriminant Analysis of Principal Components (DAPC) using adegenet 2.0.0. :1–43.

Jombart T, Collins C. 2017. A tutorial for analysing genome-wide SNP data using adegenet 2.1.0. :1–37.

Jombart T, Devillard S, Balloux F. 2010. Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. BMC Genet. 11(94):1–15. doi:10.1186/1471-2156-11-94.

Jones AT, Ovenden JR, Wang YG. 2016. Improved confidence intervals for the linkage disequilibrium method for estimating effective population size. Heredity (Edinb). 117(4):217–223. doi:10.1038/hdy.2016.19.

de Jong MJ, Li Z, Qin Y, Quéméré E, Baker K, Wang W, Hoelzel AR. 2020. Demography and adaptation promoting evolutionary transitions in a mammalian genus that diversified during the Pleistocene. Mol Ecol. 29(15):2777–2792. doi:10.1111/mec.15450.

Jost L. 2008. GST and its relatives do not measure differentiation. Mol Ecol. 17(18):4015–4026. doi:10.1111/j.1365-294X.2008.03887.x.

Jost L, Archer FI, Flanagan SP, Gaggiotti OE, Hoban S, Latch E. 2018. Differentiation measures for conservation genetics. Evol Appl. 11(7):1139–1148. doi:10.1111/eva.12590.

Joyce TW, Durban JW, Claridge DE, Dunn CA, Hickmott LS, Fearnbach H, Dolan K, Moretti D. 2020. Behavioral responses of satellite tracked Blainville's beaked whales (*Mesoplodon densirostris*) to mid-frequency active sonar. Mar Mammal Sci. 36(1):29–46. doi:10.1111/mms.12624.

Kamvar ZN, Brooks JC, Grünwald NJ. 2015. Novel R tools for analysis of genome-wide population genetic data with emphasis on clonality. Front Genet. 6(JUN):1–10. doi:10.3389/fgene.2015.00208.

Kasuya T, Nishiwaki M. 1971. First record of *Mesoplodon densirostris* from Formosa. Sci Reports Whales Res Inst. 23:129–137, 5 pls.

Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, et al. 2012. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics. 28(12):1647–1649. doi:10.1093/bioinformatics/bts199.

Keller LF, Waller DM. 2002. Inbreeding effects in wild populations. Trends Ecol Evol. 17(5):230–241. doi:10.1016/S0169-5347(02)02489-8.

Kemp AES, Grigorov I, Pearce RB, Naveira Garabato AC. 2010. Migration of the

Antarctic Polar Front through the mid-Pleistocene transition: Evidence and climatic implications. Quat Sci Rev. 29(17–18):1993–2009. doi:10.1016/j.quascirev.2010.04.027.

Kenyon KW. 1961. Cuvier Beaked Whales Stranded in the Aleutian Islands. J Mammal. 42(1):71–76.

Kimura M. 1991. The neutral theory of molecular evolution: A review of recent evidence. Japanese J Genet. 66:367–386.

Kirino T. 1956. On the Number of Teeth and Its Variability in *Berardius bairdi*, a Genus of the Beaked Whale. Okajimas Folia Anat Jpn. 28:429–434.

Knaus BJ, Grünwald NJ. 2017. vcfr: a package to manipulate and visualize variant call format data in R. Mol Ecol Resour. 17(1):44–53. doi:10.1111/1755-0998.12549.

Korneliussen TS, Albrechtsen A, Nielsen R. 2014. ANGSD: Analysis of Next Generation Sequencing Data. BMC Bioinformatics. 15(356):1–13. doi:10.1186/s12859-014-0356-4.

Kraemer P, Gerlach G. 2017. Demerelate: calculating interindividual relatedness for kinship analysis based on codominant diploid genetic markers using R. Mol Ecol Resour. 17(6):1371–1377. doi:10.1111/1755-0998.12666.

Kraft S, Pérez-Álvarez MjJ, Olavarría C, Poulin E. 2020. Global phylogeography and genetic diversity of the long-finned pilot whale *Globicephala melas*, with new data from the southeastern Pacific. Sci Rep. 10(1769):1–13. doi:10.1038/s41598-020-58532-3.

Krützen M, Barré LM, Möller LM, Heithaus MR, Simms C, Sherwin WB. 2002. A biopsy system for small cetaceans: Darting success and wound healing in *Tursiops* spp. Mar Mammal Sci. 18(4):863–878. doi:10.1111/j.1748-7692.2002.tb01078.x.

Lacy RC. 1993. Vortex: a computer simulation model for Population Viability Analysis. Wildl Res. 20:45–65.

Lacy RC. 2000. Structure of the VORTEX simulation model for population viability analysis. Ecol Bull. 48:191–203.

Lacy RC, Miller PS, Traylor-Holzer K. 2017. Vortex 10 User's Manual. 6 September 2017 update.

Lacy RC, Pollak JP. 2017. Vortex: A Stochastic Simulation of the Extinction Proccess. Version 10.2.9.

Lah L, Trense D, Benke H, Berggren P, Gunnlaugsson P, Lockyer C, Öztürk AA, Öztürk B, Pawliczka I, Roos A, et al. 2016. Spatially explicit analysis of genome-wide SNPs detects subtle population structure in a mobile marine mammal, the harbor porpoise. PLoS One. 11(10):1–23. doi:10.1371/journal.pone.0162792.

Lambertsen RH. 1987. A biopsy system for large whales and it's use for cytogenetics. J Mammal. 68(2):443–445.

Lanfear R, Frandsen PB, Wright AM, Senfeld T, Calcott B. 2017. Partitionfinder 2: New methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. Mol Biol Evol. 34(3):772–773. doi:10.1093/molbev/msw260.

Leroy G, Carroll EL, Bruford MW, Dewoody JA, Strand A, Waits L, Wang J. 2017. Next-generation metrics for monitoring genetic erosion within populations of conservation concern. Evol Appl.(October):1–18. doi:10.1111/eva.12564.

Leslie MS, Archer FI, Morin PA. 2018. Mitogenomic differentiation in spinner (*Stenella longirostris*) and pantropical spotted dolphins (*S. attenuata*) from the eastern tropical

Pacific Ocean. Mar Mammal Sci. 00(00):1–30. doi:10.1111/mms.12545.

Leslie MS, Morin PA. 2018. Structure and phylogeography of two tropical predators, spinner (*Stenella longirostris*) and pantropical spotted (*S. attenuata*) dolphins, from SNP data. R Soc Open Sci. 5. doi:http://dx.doi.org/10.1098/rsos.171615.

Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. :1–3. doi:arXiv:1303.3997 [q-bio.GN].

Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 25(14):1754–1760. doi:10.1093/bioinformatics/btp324.

Li H, Handsaker RE, Wysoker A, Fennell T, Ruan J, Homer N, Marth GT, Abecasis G, Durbin R, Subgroup 1000 Genome Project Data Processing. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 25(16):2078–2079. doi:10.1093/bioinformatics/btp352.

Lindenmayer DB, Clark T, Lacy RC, Thomas V. 1993. Population viability analysis as a tool in wildlife conservation policy: With reference to Australia. Environ Manage. 17(6):745–758. doi:10.1007/bf02393895.

Liu X, Fu YX. 2015. Exploring population size changes using SNP frequency spectra. Nat Genet. 47(5):555–559. doi:10.1038/ng.3254.

Liu X, Fu YX. 2020. Stairway Plot 2: demographic history inference with folded SNP frequency spectra. Genome Biol. 21(1):1–9. doi:10.1186/s13059-020-02196-9.

Loog L. 2021. Sometimes hidden but always there: the assumptions underlying genetic inference of demographic histories. Philos Trans R Soc B Biol Sci. 376(20190719):1–10. doi:http://dx.doi.org/10.1098/rstb.2019.0719.

Lopes da Silva Ferrette B, Coelho R, Peddemors VM, Ovenden JR, De Franco BA, Oliveira C, Foresti F, Mendonça FF. 2021. Global phylogeography of the smooth hammerhead shark: Glacial refugia and historical migration patterns. Aquat Conserv Mar Freshw Ecosyst. 31(9):2348–2368. doi:10.1002/aqc.3629.

Louis M, Fontaine MC, Spitz J, Schlund E, Dabin W, Deaville R, Caurant F, Cherel Y, Guinet C, Simon-Bouhet B. 2014. Ecological opportunities and specializations shaped genetic divergence in a highly mobile marine top predator. Proc R Soc B Biol Sci. 281(1795). doi:10.1098/rspb.2014.1558.

Lowry DB, Hoban S, Kelley JL, Lotterhos KE, Reed LK, Antolin MF, Storfer A. 2017. Breaking RAD: an evaluation of the utility of restriction site-associated DNA sequencing for genome scans of adaptation. Mol Ecol Resour. 17(2):142–152. doi:10.1111/1755-0998.12635.

Lu F, Lipka AE, Glaubitz JC, Elshire R, Cherney JH, Casler MD, Buckler ES, Costich DE. 2013. Switchgrass genomic diversity, ploidy, and evolution: Novel insights from a network-based SNP discovery protocol. PLoS Genet. 9(1):e1003215. doi:10.1371/journal.pgen.1003215.

Lucena-Perez M, Marmesat E, Kleinman-Ruiz D, Martínez-Cruz B, Węcek K, Saveljev AP, Seryodkin I V., Okhlopkov I, Dvornikov MG, Ozolins J, et al. 2020. Genomic patterns in the widespread Eurasian lynx shaped by Late Quaternary climatic fluctuations and anthropogenic impacts. Mol Ecol. 29(4):812–828. doi:10.1111/mec.15366.

Luikart G, England PR, Tallmon D, Jordan S, Taberlet P. 2003. The power and promise of population genomics: From genotyping to genome typing. Nat Rev Genet. 4:981–994.

doi:10.1038/nrg1226.

Lusher AL, Hernandez-Milian G, O'Brien J, Berrow S, O'Connor I, Officer R. 2015. Microplastic and macroplastic ingestion by a deep diving, oceanic cetacean: the True's beaked whale *Mesoplodon mirus*. Environ Pollut. 199:185–191. doi:10.1016/j.envpol.2015.01.023.

Macleod CD. 2000. Review of the distribution of *Mesoplodon* species (order Cetacea, family Ziphiidae) in the North Atlantic. Mamm Rev. 30(1):1–8. doi:10.1046/j.1365-2907.2000.00057.x.

MacLeod CD. 2005. How big is a beaked whale? A review of body length and sexual size dimorphism in the family Ziphiidae. J Cetacean Res Manag. 7(3):301.

MacLeod CD. 2009. Global climate change, range changes and potential implications for the conservation of marine cetaceans: a review and synthesis. Endanger Species Res. 7(2):125–136. doi:10.3354/esr00197.

MacLeod CD. 2018. Beaked Whales, Overview. In: Wursig B, Thewissen JGM, Kovacs KM, editors. Encyclopedia of Marine Mammals, 3rd Edition. Academic Press. p. 80–83.

MacLeod CD, Bannon SM, Pierce GJ, Schweder C, Learmonth JA, Herman JS, Reid RJ. 2005. Climate change and the cetacean community of north-west Scotland. Biol Conserv. 124(4):477–483. doi:10.1016/j.biocon.2005.02.004.

MacLeod CD, D'Amico A. 2006. A review of beaked whale behaviour and ecology in relation to assessing and mitigating impacts of anthropogenic noise. J Cetacean Res Manag. 7(3):211–221.

Macleod CD, Herman JS. 2004. Development of tusks and associated structures in *Mesoplodon bidens* (Cetaceae, Mammalia). Mammalia. 68(2–3):175–184. doi:10.1515/mamm.2004.018.

MacLeod CD, Mitchell G. 2006. Key areas for beaked whales worldwide. J Cetacean Res Manag. 7(3):309–322.

MacLeod CD, Perrin WF, Pitman RL, Barlow JP, Ballance LT, D'Amico A, Gerrodette T, Joyce G, Mullin KD, Palka DL, et al. 2006. Known and inferred distributions of beaked whale species (Cetacea : Ziphiidae). J Cetacean Res Manag. 7(3):271–286.

MacLeod CD, Santos MB, Pierce GJ. 2003. Review of data on diets of beaked whales: Evidence of niche separation and geographic segregation. J Mar Biol Assoc United Kingdom. 83(3):651–665. doi:10.1017/S0025315403007616h.

MacLeod CD, Zuur AF. 2005. Habitat utilization by Blainville's beaked whales off Great Abaco, northern Bahamas, in relation to seabed topography. Mar Biol. 147(1):1–11. doi:10.1007/s00227-004-1546-9.

Mangiafico S. 2020. rcompanion: Functions to Support Extension Education Program Evaluation. R package version 2.3.25.

Markus T, Sánchez PPS. 2018. Managing and Regulating Underwater Noise Pollution. In: Saloman M, Markus T, editors. Handbook on Marine Evironment Protection. Springer International Publishing. p. 971–995.

Marques TA, Jorge PA, Mouriño H, Thomas L, Moretti DJ, Dolan K, Claridge DE, Dunn CA. 2019. Estimating group size from acoustic footprint to improve Blainville's beaked whale abundance estimation. Appl Acoust. 156:434–439. doi:10.1016/j.apacoust.2019.07.042. https://doi.org/10.1016/j.apacoust.2019.07.042.

Martien KK, Lang AR, Taylor BL, Rosel PE, Simmons SE, Oleson EM, Boveng PL, Hanson MB. 2019. The DIP dilineation handbook: A guide to using multiple lines of evidence to delineate demographically independent populations of marine mammals. U.S. Department of Commerce, NOAA Technical Memorandum NMFS-SWFSC-622.

Martien KK, Taylor BL, Chivers SJ, Mahaffy SD, Gorgone AM, Baird RW. 2019. Fidelity to natal social groups and mating within and between social groups in an endangered false killer whale population. Endanger Species Res. 40:219–230. doi:10.3354/ESR00995.

Martín V, Servidio A, Garcia S. 2003. Mass strandings of beaked whales in the Canary Islands. In: Evans PGH, Miller LA, editors. Proceedings of the Workshop on Active Sonar and Cetaceans. European Cetacean Society Newsletter, No. 42-Special Issue. http://www.seawatchfoundation.org.uk/wp-content/uploads/2012/08/Active_Sonar_Workshop.pdf#page=18.

Maruki T, Lynch M. 2015. Genotype-frequency estimation from high-throughput sequencing data. Genetics. 201(2):473–486. doi:10.1534/genetics.115.179077.

Maruki T, Lynch M. 2017. Genotype calling from population-genomic sequencing data. G3 Genes, Genomes, Genet. 7(5):1393–1404. doi:10.1534/g3.117.039008.

Mazet O, Rodríguez W, Chikhi L. 2015. Demographic inference using genetic data from a single individual: Separating population size variation from population structure. Theor Popul Biol. 104:46–58. doi:10.1016/j.tpb.2015.06.003.

McClymont EL, Sosdian SM, Rosell-Melé A, Rosenthal Y. 2013. Pleistocene sea-surface temperature evolution: Early cooling, delayed glacial intensification, and implications for the mid-Pleistocene climate transition. Earth-Science Rev. 123:173–193. doi:10.1016/j.earscirev.2013.04.006.

McGowen MR, Tsagkogeorga G, Álvarez-Carretero S, dos Reis M, Struebig M, Deaville R, Jepson PD, Jarman SN, Polanowski AM, Morin PA, et al. 2019. Phylogenomic resolution of the cetacean tree of life using target sequence capture. Syst Biol. 0(0):1–24. doi:10.1093/sysbio/syz068.

McSweeney DJ, Baird RW, Mahaffy SD. 2007. Site fidelity, associations, and movements of Cuvier's (*Ziphius cavirostris*) and Blainville's (*Mesoplodon densirostris*) beaked whales off the island of Hawai'i. Mar Mammal Sci. 23(3):666–687. doi:10.1111/j.1748-7692.2007.00135.x.

Mead JG. 1984. Survey of reproductive data for the beaked whales (Ziphiidae). Rep Int Whal Comm.(Special Issue 6):91–96.

Mead JG. 1989. Beaked Whales of the Genus *Mesoplodon*. In: Ridgway SH, Harrison SR, editors. Handbook of Marine Mammals Volume 4: River Dolphins and Larger Toothed Whales. p. 349–430.

Mead JG. 2008. Beaked Whales, Overview. In: Perrin WF, Wursig B, Thewissen JGM, editors. Encyclopedia of Marine Mammals, 2nd Edition. Academic Press. p. 94–97.

Mejri R, Lo Brutto S, Hassine OKB, Arculeo M. 2009. A study on *Pomatoschistus tortonesei* Miller 1968 (Perciformes, Gobiidae) reveals the Siculo-Tunisian Strait (STS) as a breakpoint to gene flow in the Mediterranean basin. Mol Phylogenet Evol. 53:596–601. doi:10.1016/j.ympev.2009.04.018.

Meyer-Gutbrod EL, Greene CH, Sullivan PJ, Pershing AJ. 2015. Climate-associated changes in prey availability drive reproductive dynamics of the North Atlantic right whale

population. Mar Ecol Prog Ser. 535:243–258. doi:10.3354/meps11372.

Mignucci-Giannoni AA. 1996. Marine mammal strandings in Puerto Rico and the United States and British Virgin Islands. PhD Thesis. University of Puerto Rico, Puerto Rico.

Mignucci-Giannoni AA, Rosario-Delestre RJ. 1998. Mass strandings of goosebeak whales in Puerto Rico. Whale World (Nature Alert). 1(2).

Miller JM, Cullingham CI, Peery RM. 2020. The influence of *a priori* grouping on inference of genetic clusters: Simulation study and literature review of the DAPC method. Heredity (Edinb). 125(5):269–280. doi:10.1038/s41437-020-0348-2.

Miller MR, Dunham JP, Amores A, Cresko WA, Johnson EA. 2007. Rapid and cost-effective polymorphism identification and genotyping using restriction site associated DNA (RAD) markers. Genome Res. 17:240–248. doi:10.1101/gr.5681207.

Miller PJO, Johnson MP, Tyack PL, Terray EA. 2004. Swimming gaits , passive drag and buoyancy of diving sperm whales *Physeter macrocephalus*. J Exp Biol. 207:1953–1967. doi:10.1242/jeb.00993.

Miller PS. 2016. A population viability analysis for the Chinese white dolphin (*Sousa chinensis*) in the Pearl River estuary: Final report. Apple Val MN IUCN/SSC Conserv Breed Spec Gr.

Monsen-Collar KJ, Dolcemascolo P. 2010. Using molecular techniques to answer ecological questions. Nat Educ Knowl. 3(10):1.

Moore JE, Barlow JP. 2013. Declining abundance of beaked whales (family Ziphiidae) in the California Current large marine ecosystem. PLoS One. 8(1):e52770. doi:10.1371/journal.pone.0052770.

Moors-Murphy HB. 2018. Bottlenose Whales. In: Wursig B, Thewissen JGM, Kovacs KM, editors. Encyclopedia of Marine Mammals, 3rd Edition. Academic Press. p. 130–132.

Morin PA, Archer FI, Foote AD, Vilstrup J, Allen EE, Wade PR, Durban JW, Parsons KM, Pitman RL, Li L, et al. 2008. Complete mitochondrial genome phylogeographic analysis of killer whales (*Orcinus orca*) indicates multiple species. Genome Res. 20(858):908–916. doi:10.1101/gr.102954.109.908.

Morin PA, Baker CS, Brewer RS, Burdin AM, Dalebout ML, Dines JP, Fedutin ID, Filatova OA, Hoyt E, Jung JL, et al. 2017. Genetic structure of the beaked whale genus *Berardius* in the North Pacific, with genetic evidence for a new species. Mar Mammal Sci. 33(1):96–111. doi:10.1111/mms.12345.

Morin PA, Dizon AE. 2018. Genetics, Management. In: Wursig B, Thewissen JGM, Kovacs KM, editors. Encyclopedia of Marine Mammals, 3rd Edition. Academic Press. p. 410–416.

Morin PA, Duchêne S, Lee N, Durban JW, Claridge DE. 2012. Preliminary analysis of mitochondrial genome phylogeography of Blainville's, Cuvier's and Gervais' beaked whales. SC/64/SM14. Int Whal Comm Sci Meet 64.:1–17.

Morin PA, Foote AD, Baker CS, Hancock-Hanser BL, Kaschner K, Mate BR, Mesnick SL, Pease VL, Rosel PE, Alexander AM. 2018. Demography or selection on linked cultural traits or genes? Investigating the driver of low mtDNA diversity in the sperm whale using complementary mitochondrial and nuclear genome analyses. Mol Ecol. 27:2604–2619. doi:10.1111/mec.14698.

Morin PA, Forester BR, Forney KA, Crossman CA, Hancock-Hanser BL, Robertson KM, Barrett-Lennard LG, Baird RW, Calambokidis J, Gearin P, et al. 2021. Population structure in a continuously distributed coastal marine species, the harbor porpoise, based on microhaplotypes derived from poor quality samples. Mol Ecol. 00:1–20. doi:10.1111/mec.15827.

Moulins A, Rosso M, Nani B, Würtz M. 2007. Aspects of the distribution of Cuvier's beaked whale (*Ziphius cavirostris*) in relation to topographic features in the Pelagos Sanctuary (north-western Mediterranean Sea). J Mar Biol Assoc United Kingdom. 87(1):177–186. doi:10.1017/S0025315407055002.

Moura AE, Janse van Rensburg C, Pilot M, Tehrani A, Best PB, Thornton M, Plön S, De Bruyn PJN, Worley KC, Gibbs RA, et al. 2014. Killer whale nuclear genome and mtDNA reveal widespread population bottleneck during the last glacial maximum. Mol Biol Evol. 31(5):1121–1131. doi:10.1093/molbev/msu058.

El Mousadik A, Petit RJ. 1996. High level of genetic differentiation for allelic richness among populations of the argan tree [*Argania spinosa* (L.) Skeels] endemic to Morocco. Theor Appl Genet. 92:832–839. doi:10.1007/BF00221895.

Mueter FJ, Broms C, Drinkwater KF, Friedland KD, Hare JA, Hunt GL, Melle W, Taylor M. 2009. Ecosystem responses to recent oceanographic variability in high-latitude Northern Hemisphere ecosystems. Prog Oceanogr. 81:93–110. doi:10.1016/j.pocean.2009.04.018.

Nei M. 1978. Estimation of average heterozygosity and genetic distance from a small number of individuals. Genetics. 89(3):583–590.

Nei M. 1987a. Chapter 7: Genes in Populations. In: Molecular Evolutionary Genetics. Columbia University Press. p. 149–175.

Nei M. 1987b. Molecular evolutionary genetics. Columbia University Press.

Nelms SE, Alfaro-Shigueto J, Arnould J, Avila I, Bengtson Nash S, Campbell E, Carter M, Collins T, Currey R, Domit C, et al. 2021. Marine mammal conservation: over the horizon. Endanger Species Res. 44:291–325. doi:10.3354/esr01115.

New LF, Moretti DJ, Hooker SK, Costa DP, Simmons SE. 2013. Using energetic models to investigate the survival and reproduction of beaked whales (family Ziphiidae). PLoS One. 8(7). doi:10.1371/journal.pone.0068725.

Nielsen R, Paul JS, Albrechtsen A, Song YS. 2011. Genotype and SNP calling from next-generation sequencing data. Nat Rev Genet. 12(6):443–451. doi:10.1038/nrg2986.

Nijs V. 2022. radiant: Business analytics using R and Shiny.

Nikaido M, Matsuno F, Hamilton H, Brownell Jr. RL, Cao Y, Ding W, Zuoyan Z, Shedlock AM, Fordyce RE, Hasegawa M, et al. 2001. Retroposon analysis of major cetacean lineages: the monophyly of toothed whales and the paraphyly of river dolphins. Proc Natl Acad Sci. 98(13):7384–7389. doi:10.1073/pnas.121139198.

NOAA Fisheries. 2022. Sowerby's Beaked whale. NOAA Fish Species Dir. https://www.fisheries.noaa.gov/species/sowerbys-beaked-whale.

Noren DP, Mocklin JA. 2012. Review of cetacean biopsy techniques: Factors contributing to successful sample collection and physiological and behavioral impacts. Mar Mammal Sci. 28(1):154–199. doi:10.1111/j.1748-7692.2011.00469.x.

Notarbartolo-di-Sciara G, Agardy T, Hyrenbach D, Scovazzi T, Van Klaveren P. 2008.

The Pelagos Sanctuary for Mediterranean marine mammals. Aquat Conserv Mar Freshw Ecosyst. 18(4):367–391. doi:10.1002/aqc.855.

Nowacek DP, Thorne LH, Johnston DW, Tyack PL. 2007. Reponses of cetaceans to anthropogenic noise. Mamm Rev. 37(2):81–115. doi:10.1111/j.1365-2907.2007.00104.x.

O'Brien K, Whitehead H. 2013. Population analysis of Endangered northern bottlenose whales on the Scotian Shelf seven years after the establishment of a Marine Protected Area. Endanger Species Res. 21(3):273–284. doi:10.3354/esr00533.

O'Corry-Crowe G, Suydam R, Quakenbush L, Smith TG, Lydersen C, Kovacs KM, Orr J, Harwood L, Litovka D, Ferrer T. 2020. Group structure and kinship in beluga whale societies. Sci Rep. 10(1):1–21. doi:10.1038/s41598-020-67314-w.

O'Dea A, Lessios HA, Coates AG, Eytan RI, Restrepo-Moreno SA, Cione AL, Collins LS, de Queiroz A, Farris DW, Norris RD, et al. 2016. Formation of the Isthmus of Panama. Sci Adv. 2(8):1–11. doi:10.1126/sciadv.1600883.

O'Leary SJ, Puritz JB, Willis SC, Hollenbeck CM, Portnoy DS. 2018. These aren't the loci you'e looking for: Principles of effective SNP filtering for molecular ecologists. Mol Ecol.(June):3193–3206. doi:10.1111/mec.14792.

OBIS. 2021. Ocean Biodiversity Information System. Intergovernmental Oceanic Commission of UNESCO. www.obis.org.

Oliver WRB. 1937. *Tasmacetus shepherdi*: a new genus and species of beaked Whale from New Zealand. Proc Zool Soc London.:371–382. doi:10.1111/j.1469-7998.1937.tb00015.x.

Omura H, Fujino K, Kimura S. 1955. Beaked whale *Berardius bairdi* of Japan, with notes on *Ziphius cavirostris*. Sci Reports Whales Res Inst. 10:89–132.

Onoufriou AB, Gaggiotti OE, Aguilar de Soto N, McCarthy ML, Morin PA, Rosso M, Dalebout ML, Davison NJ, Baird RW, Baker CS, et al. 2022. Biogeography in the deep: Hierarchical population genomic structure of two beaked whale species. Glob Ecol Conserv. 40:1–18. doi:10.1016/j.gecco.2022.e02308.

Osman MB, Das SB, Trusel LD, Evans MJ, Fischer H, Grieman MM, Kipfstuhl S, McConnell JR, Saltzman ES. 2019. Industrial-era decline in subarctic Atlantic productivity. Nature. 569(7757):551–555. doi:10.1038/s41586-019-1181-8.

Palsbøll PJ, Bérubé M, Allendorf FW. 2007. Identification of management units using population genetic data. Trends Ecol Evol. 22(1):11–16. doi:10.1016/j.tree.2006.09.003.

Pante E, Simon-Bouhet B. 2013. marmap: a package for importing, plotting and analyzing bathymetric and topographic data in R. PLoS One. 8(9):e73051. doi:10.1371/journal.pone.0073051.

Parsons ECM. 2016. Why IUCN should replace 'Data Deficient' conservation status with a precautionary 'Assume Threatened' status- a cetacean case study. Front Mar Sci. 3:1–3. doi:10.3389/fmars.2016.00193.

Parsons ECM. 2017. Impacts of Navy Sonar on Whales and Dolphins: Now beyond a Smoking Gun? Front Mar Sci. 4:1–11. doi:10.3389/fmars.2017.00295.

Pastene LA, Goto M, Kanda N, Zerbini AN, Kerem D, Watanabe K, Bessho Y, Hasegawa M, Nielsen R, Larsen F, et al. 2007. Radiation and speciation of pelagic organisms during periods of global warming: the case of the common minke whale, *Balaenoptera acutorostrata*. Mol Ecol. 16(7):1481–1495. doi:10.1111/j.1365-294X.2007.03244.x.

Patarnello T, Volckaert FAMJ, Castilho R. 2007. Pillars of Hercules: is the Atlantic-Mediterranean transition a phylogeographical break? Mol Ecol. 16(21):4426–4444. doi:10.1111/j.1365-294X.2007.03477.x.

Patel S, Thompson KF, Santure AW, Constantine R, Millar CD. 2017. Genetic kinship analyses reveal that Gray's beaked whales strand in unrelated groups. J Hered. 108(4):456–461. doi:10.1093/jhered/esx021.

Peart CR, Tusso S, Pophaly SD, Botero-Castro F, Wu CC, Aurioles-Gamboa D, Baird AB, Bickham JW, Forcada J, Galimberti F, et al. 2020. Determinants of genetic variation across eco-evolutionary scales in pinnipeds. Nat Ecol Evol. doi:10.1038/s41559-020-1215-5.

Pereira HM, Ferrier S, Walters M, Geller GN, Jongman RHG., Scholes RJ, Bruford MW, Brummitt N, Butchar SHM, Cardoso AC, et al. 2013. Essential Biodiversity Variables. Science. 339:277–278. doi:10.1126/science.1229931.

Pérez-Espona S, CryoArks Consortium. 2021. Conservation-focused biobanks: a valuable resource for wildlife DNA forensics. Forensic Sci Int Anim Environ. 1:1–5. doi:10.1016/j.fsiae.2021.100017.

Pershing AJ, Stamieszkin K. 2020. The North Atlantic ecosystem, from plankton to whales. Ann Rev Mar Sci. 12:339–359. doi:10.1146/annurev-marine-010419-010752.

Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE. 2012. Double digest RADseq: An inexpensive method for *de novo* SNP discovery and genotyping in model and non-model species. PLoS One. 7(5):e37135. doi:10.1371/journal.pone.0037135.

Pitman RL. 2018. *Mesoplodon* Beaked Whales. In: Wursig B, Thewissen JGM, Kovacs KM, editors. Encyclopedia of Marine Mammals, 3rd Edition. Academic Press. p. 595–602.

Pitman RL, Brownell Jr. RL. 2020a. *Mesoplodon densirostris,* Blainville's Beaked Whale. IUCN Red List Threat Species. e.T13244A5. doi:10.2305/IUCN.UK.2020-3.RLTS.T13244A50364253.en.

Pitman RL, Brownell Jr. RL. 2020b. *Mesoplodon bidens*, Sowerby's Beaked Whale. IUCN Red List Threat Species. e.T13241A5.

Pitman RL, Palacios DM, Brennan PLR, Brennan BJ, Balcomb KC, Miyashita T. 1999. Sightings and possible identity of a bottlenose whale in the tropical Indo-Pacific: *Indopacetus pacificus*? Mar Mammal Sci. 15(2):531–549. doi:10.1111/j.1748-7692.1999.tb00818.x.

Podestà M, Azzellino A, Cañadas A, Frantzis A, Moulins A, Rosso M, Tepsich P, Lanfredi C. 2016. Cuvier's beaked whale, *Ziphius cavirostris*, distribution and occurrence in the Mediterranean Sea: High-use areas and conservation threats. Adv Mar Biol. 75:103–140. doi:10.1016/bs.amb.2016.07.007.

Podestà M, D'Amico A, Pavan G, Drougas A, Komnenou A, Portunato N. 2006. A review of Cuvier's beaked whale strandings in the Mediterranean Sea. J Cetacean Res Manag. 7(3):251–261.

Pritchard JK, Stephens M, Donnelly P. 2000. Inference of population structure using multilocus genotype data. Genetics. 155(2):945–959. doi:10.1111/j.1471-8286.2007.01758.x.

Puckett EE, Magnussen E, Khlyap LA, Strand TM, Lundkvist Å, Munshi-South J. 2020.

Genomic analyses reveal three independent introductions of the invasive brown rat (*Rattus norvegicus*) to the Faroe Islands. Heredity (Edinb). 124(1):15–27. doi:10.1038/s41437-019-0255-6.

Puritz JB, Hollenbeck CM, Gold JR. 2014. dDocent: a RADseq, variant-calling pipeline designed for population genomics of non-model organisms. PeerJ.:2e431. doi:10.7717/peerj.431.

Quick NJ, Cioffi WR, Shearer JM, Fahlman A, Read AJ. 2020. Extreme diving in mammals: First estimates of behavioural aerobic dive limits in Cuvier's beaked whales. J Exp Biol. 223(18):jeb222109. doi:10.1242/jeb.222109.

R Core Team. 2019. R: A Language and Environment for Statistical Computing.

Ramp C, Delarue J, Palsbøll PJ, Sears R, Hammond PS. 2015. Adapting to a warmer ocean - Seasonal shift of baleen whale movements over three decades. PLoS One. 10(3):1–15. doi:10.1371/journal.pone.0121374.

Ran M, Chen L. 2019. The 4.2 ka BP climatic event and its cultural responses. Quat Int. 521(April):158–167. doi:10.1016/j.quaint.2019.05.030.

Reece JS, Castoe TA, Parkinson CL. 2005. Historical perspectives on population genetics and conservation of three marine turtle species. Conserv Genet. 6(2):235–251. doi:10.1007/s10592-004-7821-y.

Reed DH, Frankham R. 2003. Correlation between fitness and genetic diversity. Conserv Biol. 17(1):230–237. doi:10.1046/j.1523-1739.2003.01236.x.

Reeves IM, Totterdell JA, Barceló A, Sandoval-Castillo J, Batley KC, Stockin KA, Betty EL, Donnelly DM, Wellard R, Beheregaray LB, et al. 2022. Population genomic structure of killer whales (*Orcinus orca*) in Australian and New Zealand waters. Mar Mammal Sci. 38(1):151–174. doi:10.1111/mms.12851.

Reeves RR. 2018. Conservation. In: Wursig B, Thewissen JGM, Kovacs KM, editors. Encyclopedia of Marine Mammals, 3rd Edition. Academic Press. p. 215–229.

Reid JB, Evans PGH, Northridge SP. 2003. Atlas of Cetacean, distribution in north-west European waters. http://www.ncbi.nlm.nih.gov/pubmed/17578592.

Reyes C. 2018. Abundance estimate, survival and site fidelity patterns of Blainville's (*Mesoplodon densirostris*) and Cuvier's (*Ziphius cavirostris*) beaked whales off El Hierro (Canary Islands). MPhil Thesis. University of St Andrews, Scotland.

Reyes C, Schiavi A, Aguilar de Soto N. 2012. CETABASE: a bilingual tool to enhance data sharing and public outreach on endangered species. www.cetabase.info.

Reyes JC, Mead JG, Van Waerebeek K. 1991. A new species of beaked whale *Mesoplodon peruvianus* Sp. N. (Cetacea: Ziphiidae) from Peru. Mar Mammal Sci. 7(1):1–24.

Riccialdelli L, Natalia Paso Viola M, Panarello HO, Goodall RNP. 2017. Evaluating the isotopic niche of beaked whales from the southwestern South Atlantic and Southern Oceans. Mar Ecol Prog Ser. 581:183–198. doi:10.3354/meps12296.

Rivera-Colón AG, Catchen J. 2022. Population Genomics Analysis with RAD, Reprised: Stacks 2. In: Verde C, Giordano D, editors. Marine Genomics. Methods in Molecular Biology vol 2498. Humana. p. 99–149.

Rochette NC, Rivera-Colón AG, Catchen JM. 2019. Stacks 2: Analytical methods for

paired-end sequencing improve RADseq-based population genomics. Mol Ecol. 28(21):4737–4754. doi:10.1111/mec.15253.

Rogan E, Cañadas A, Macleod K, Santos MB, Mikkelsen B, Uriarte A, Van Canneyt O, Vázquez JA, Hammond PS. 2017. Distribution, abundance and habitat use of deep diving cetaceans in the North-East Atlantic. Deep Sea Res Part II. 141:8–19. doi:10.1016/j.dsr2.2017.03.015.

Rosel PE, Hancock-hanser BL, Archer FI, Robertson KM, Martien KK, Leslie MS, Berta A, Cipriano F, Viricel A, Viaud-Martinez KA, et al. 2017. Examining metrics and magnitudes of molecular genetic differentiation used to delimit cetacean subspecies based on mitochondrial DNA control region sequences. Mar Mammal Sci. 33(Special Issue):76–100. doi:10.1111/mms.12410.

Rosel PE, Taylor BL, Hancock-Hanser BL, Morin PA, Archer FI, Lang AR, Mesnick SL, Pease VL, Perrin WF, Robertson KM, et al. 2017. A review of molecular genetic markers and analytical approaches that have been used for delimiting marine mammal subspecies and species. Mar Mammal Sci. 33(Special Issue):56–75. doi:10.1111/mms.12412.

Rossby T. 1996. The North Atlantic Current and surrounding waters: At the crossroads. Rev Geophys. 34(4):463–481. doi:10.1029/96RG02214.

Rosso M, Ballardini M, Moulins A, Würtz M. 2011. Natural markings of Cuvier's beaked whale *Ziphius cavirostris* in the Mediterranean Sea. African J Mar Sci. 33(1):45–57. doi:10.2989/1814232X.2011.572336.

Rozas J, Ferrer-Mata A, Sanchez-DelBarrio JC, Guirao-Rico S, Librado P, Ramos-Onsins SE, Sanchez-Gracia A. 2017. DnaSP 6: DNA sequence polymorphism analysis of large data sets. Mol Biol Evol. 34(12):3299–3302. doi:10.1093/molbev/msx248.

Rugh DJ, Shelden KEW, Schulman-Janiger A. 2001. Timing of the gray whale southbound migration. J Cetacean Res Manag. 3(1):31–39.

Salmona J, Heller R, Lascoux M, Shafer A. 2017. Inferring Demographic History Using Genomic Data. In: Rajora OP, editor. Population Genomics. Springer Nature. p. 511–537.

Sambrook J, Fritsch EF, Maniatis T. 1989. Molecular Cloning, A Laboratory Manual. Second Edition. Cold Spring Harbror Laboratory Press, Cold Spring Harbor, New York.

Schipper J, Chanson JS, Chiozza F, Cox NA, Hoffmann M, Katariya V, Lamoreux J, Rodrigues ASL, Stuart SN, Temple HJ, et al. 2008. The status of the world's land and marine mammals: Diversity, threat and knowledge. Science. 322:225–230. doi:10.1126/science.1165115.

Schmittner A. 2005. Decline of the marine ecosystem caused by a reduction in the Atlantic overturning circulation. Nature. 434(7033):628–633. doi:10.1038/nature03476.

Schorr GS, Falcone EA, Moretti DJ, Andrews RD. 2014. First long-term behavioral records from Cuvier's beaked whales (*Ziphius cavirostris*) reveal record-breaking dives. PLoS One. 9(3). doi:10.1371/journal.pone.0092633.

Secchi ER, Zarzur S. 1999. Plastic debris ingested by a Blainville's beaked whale, *Mesoplodon densirostris*, washed ashore in Brazil. Aquat Mamm. 25.1:21–24.

Shafer ABA, Gattepaille LM, Stewart REA, Wolf JBW. 2015. Demographic inferences using short-read genomic data in an approximate Bayesian computation framework: *in silico* evaluation of power, biases and proof of concept in Atlantic walrus. Mol Ecol. 24(2):328–345. doi:10.1111/mec.13034.

Shafer ABA, Peart CR, Tusso S, Maayan I, Brelsford A, Wheat CW, Wolf JBW. 2017. Bioinformatic processing of RAD-seq data dramatically impacts downstream population genetic inference. Methods Ecol Evol. 8(8):907–917. doi:10.1111/2041-210X.12700.

Shaffer ML. 1981. Minimum population sizes for species conservation. Bioscience. 31(2):131–134. doi:10.2307/1308256. https://academic.oup.com/bioscience/article-lookup/doi/10.2307/1308256.

Shannon G, McKenna MF, Angeloni LM, Crooks KR, Fristrup KM, Brown E, Warner KA, Nelson MD, White C, Briggs J, et al. 2016. A synthesis of two decades of research documenting the effects of noise on wildlife. Biol Rev. 91(4):982–1005. doi:10.1111/brv.12207.

da Silva DMP, Azevedo AF, Secchi ER, Barbosa LA, Flores PAC, Carvalho RR, Bisi TL, Lailson-Brito J, Cunha HA. 2015. Molecular taxonomy and population structure of the rough-toothed dolphin *Steno bredanensis* (Cetartiodactyla: Delphinidae). Zool J Linn Soc. 175(4):949–962. doi:10.1111/zoj.12301.

Simmonds MP, Eliott WJ. 2009. Climate change and cetaceans: Concerns and recent developments. J Mar Biol Assoc United Kingdom. 89(1):203–210. doi:10.1017/S0025315408003196.

Simmonds MP, Lopez-Jurado LF. 1991. Whales and the military. Nature. 351:448. doi:10.1038/351448a0.

Simonis AE, Brownell RL, Thayre BJ, Trickey JS, Oleson EM, Huntington R, Baumann-Pickering S. 2020. Co-occurrence of beaked whale strandings and naval sonar in the Mariana Islands, Western Pacific. Proc R Soc B Biol Sci. 287(1921). doi:10.1098/rspb.2020.0070.

Sims DW, Genner MJ, Southward AJ, Hawkins SJ. 2001. Timing of squid migration reflects North Atlantic climate variability. Proc R Soc B Biol Sci. 268(1485):2607–2611. doi:10.1098/rspb.2001.1847.

Skovrind M, Louis M, Westbury M V., Garilao C, Kaschner K, Castruita JAS, Gopalakrishnan S, Knudsen SW, Haile JS, Dalén L, et al. 2021. Circumpolar phylogeography and demographic history of beluga whales reflect past climatic fluctuations. Mol Ecol. 30(11):2543–2559. doi:10.1111/mec.15915.

Smith KJ, Mead JG, Peterson MJ. 2021. Specimens of opportunity provide vital information for research and conservation regarding elusive whale species. Environ Conserv.:1–9. doi:10.1017/S0376892920000521.

Smith KJ, Trueman CN, France CAM, Sparks JP, Brownlow AC, Dähne M, Davison NJ, Guðmundsson G, Khidas K, Kitchener AC, et al. 2021. Stable isotope analysis of specimens of opportunity reveals ocean-scale site fidelity in an elusive whale species. Front Conserv Sci. 2:1–11. doi:10.3389/fcosc.2021.653766.

Sowerby J. 1804. Extracts from the Minute-Book of the Linnean Society. Trans Linn Soc London.

Spear SF, Crisafulli CM, Storfer A. 2012. Genetic structure among coastal tailed frog populations at Mount St. Helens is moderated by post-disturbance management. Ecol Appl. 22(3):856–869. doi:10.1890/11-0627.1.

Spitz J, Cherel Y, Bertin S, Kiszka J, Dewez A, Ridoux V. 2011. Prey preferences among the community of deep-diving odontocetes from the Bay of Biscay, Northeast Atlantic. Deep Sea Res Part I. 58:273–282. doi:10.1016/j.dsr.2010.12.009.

Stafford KM. 2019. Increasing detections of killer whales (*Orcinus orca*), in the Pacific Arctic. Mar Mammal Sci. 35(2):696–706. doi:10.1111/mms.12551.

Storrie L, Lydersen C, Andersen M, Wynn RB, Kovacs KM. 2018. Determining the species assemblage and habitat use of cetaceans in the Svalbard Archipelago, based on observations from 2002 to 2014. Polar Res. 37(1). doi:10.1080/17518369.2018.1463065.

Sundby S, Drinkwater KF, Kjesbu OS. 2016. The North Atlantic spring-bloom system-where the changing climate meets the winter dark. Front Mar Sci. 3(MAR):1–12. doi:10.3389/fmars.2016.00028.

Sutton TT, Milligan RJ. 2019. Aquatic ecology: Deep-sea ecology. In: Fath BD, editor. Encyclopedia of Ecology, 2nd Edition. Vol. 1. Elsevier B. W. p. 35–45.

Taguchi M, Chivers SJ, Rosel PE, Matsuishi T, Abe S. 2010. Mitochondrial DNA phylogeography of the harbour porpoise *Phocoena phocoena* in the North Pacific. Mar Biol. 157(7):1489–1498. doi:10.1007/s00227-010-1423-7.

Tajima F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics. 123:585–595.

Taylor BL, Archer FI, Martien KK, Rosel PE, Hancock-Hanser BL, Lang AR, Leslie MS, Mesnick SL, Morin PA, Pease VL, et al. 2017. Guidelines and quantitative standards to improve consistency in cetacean subspecies and species delimitation relying on molecular genetic data. Mar Mammal Sci. 33(Special Issue):132–155. doi:10.1111/mms.12411.

Taylor BL, Baird RW, Barlow JP, Dawson SM, Ford JKB, Mead JG, Notarbartolo di Sciara G, Wade PR, Pitman RL. 2008a. *Ziphius cavirostris* (Global), Cuvier's Beaked Whale. IUCN Red List Threat Species 2008 eT23211A9429826.

Taylor BL, Baird RW, Barlow JP, Dawson SM, Ford JKB, Mead JG, Notarbartolo di Sciara G, Wade PR, Pitman RL. 2008b. *Mesoplodon densirostris* (Blainville's Beaked Whale). IUCN Red List Threat Species 2008 eT13244A3426474.

Taylor BL, Chivers SJ, Larese J, Perrin WF. 2007. Generation length and percent mature estimates for IUCN assessments of cetaceans. Adm Rep LJ-07-01 Natl Mar Fish.:24.

Taylor BL, Martien KK, Morin PA. 2010. Identifying units to conserve using genetic data. In: Boyd IL, Bowen WD, Iverson SJ, editors. Marine Mammal Ecology and Conservation: A handbook of techniques. Oxford University Press, University of Oxford, UK. p. 306–324.

Taylor BL, Perrin WF, Reeves RR, Rosel PE, Wang JY, Cipriano F, Baker CS, Brownell Jr. RL. 2017. Why we should develop guidelines and quantitative standards for using genetic data to delimit subspecies for data-poor organisms like cetaceans. Mar Mammal Sci. 33(Special Issue):12–26. doi:10.1111/mms.12413.

Tejedor M, Martín V, Perez-Gil M, Perez-Gil E, Ruiz L, Reyes M, Servidio A, Castrillon J, Brederlau B, Neves S. 2010. Importance of oriental Canary Islands as breeding ground for beaked whales, based on sighting and stranding data. Mar Mammal Sci. 26(3):602–613. doi:10.1111/j.1748-7692.2010.00370.x.

Thakur M, Schättin EW, McShea WJ. 2018. Globally common, locally rare: Revisiting disregarded genetic diversity for conservation planning of widespread species. Biodivers Conserv. 27(11):3031–3035. doi:10.1007/s10531-018-1579-x.

Thompson KF, Baker CS, van Helden AL, Patel S, Millar CD, Constantine R. 2012. The world's rarest whale. Curr Biol. 22(21):R905–R906. doi:10.1016/j.cub.2012.08.055.

Thompson KF, Millar CD, Scott Baker C, Dalebout ML, Steel DJ, van Helden AL, Constantine R. 2013. A novel conservation approach provides insights into the management of rare cetaceans. Biol Conserv. 157:331–340. doi:10.1016/j.biocon.2012.07.017.

Thompson KF, Patel S, Baker CS, Constantine R, Millar CD. 2016. Bucking the trend: Genetic analysis reveals high diversity, large population size and low differentiation in a deep ocean cetacean. Heredity (Edinb). 116(3):277–285. doi:10.1038/hdy.2015.99.

Thompson KF, Patel S, Williams L, Tsai P, Constantine R, Baker CS, Millar CD. 2016. High coverage of the complete mitochondrial genome of the rare Grays beaked whale (*Mesoplodon grayi*) using Illumina next generation sequencing. Mitochondrial DNA. 27(1):128–129. doi:10.3109/19401736.2013.878908.

Thompson KF, Ruggiero K, Millar CD, Constantine R, Van Helden AL. 2014. Large-scale multivariate analysis reveals sexual dimorphism and geographic differences in the Gray's beaked whale. J Zool. 294:13–21. doi:10.1111/jzo.12151.

Toonen RJ, Bowen BW, Iacchei M, Briggs JC. 2016. Biogeography, Marine. In: Kliman RM, editor. Encyclopedia of Evolutionary Biology, Volume 1. Oxford: Academic Press. p. 166–178.

Törnqvist TE, Hijma MP. 2012. Links between early Holocene ice-sheet decay, sea-level rise and abrupt climate change. Nat Geosci. 5(9):601–606. doi:10.1038/ngeo1536.

True FW. 1910. An account of the beaked whals of the family Ziphiidae in the collection of the United States Natioinal Museum, with remarks on some specimens in other American Museums. United States Natl Museum Bull 73. 89.

Tyack PL, Johnson MP, Aguilar de Soto N, Sturlese A, Madsen PT. 2006. Extreme diving of beaked whales. J Exp Biol. 209(21):4238–4253. doi:10.1242/jeb.02505.

Tyack PL, Zimmer WMX, Moretti DJ, Southall BL, Claridge DE, Durban JW, Clark CW, D'Amico A, DiMarzio N, Jarvis S, et al. 2011. Beaked whales respond to simulated and actual navy sonar. PLoS One. 6(3). doi:10.1371/journal.pone.0017009.

U.S. Department of the Navy. 2018. Final environmental impact statement/overseas environmental impact statement Hawaii-Southern California training and testing. https://www.hstteis.com/portals/hstteis/files/hstteis_p3/feis/section/HSTT_FEIS_3.07_Marine_Mammals_October_2018.pdf.

Vachon F, Hersh TA, Rendell L, Gero S, Whitehead H. 2022. Ocean nomads or island specialists? Culturally driven habitat partitioning contrasts in scale between geographically isolated sperm whale populations. R Soc Open Sci. 9(5). doi:10.1098/rsos.211737.

VanCise AM, Baird RW, Baker CS, Cerchio S, Claridge DE, Fielding R, Hancock-Hanser BL, Marrero J, Martien KK, Mignucci-Giannoni AA, et al. 2019. Oceanographic barriers, divergence, and admixture: Phylogeography and taxonomy of two putative subspecies of short-finned pilot whale. Mol Ecol.:1–17. doi:10.1111/mec.15107.

VanCise AM, Martien KK, Mahaffy SD, Baird RW, Webster DL, Fowler JH, Oleson EM, Morin PA. 2017. Familial social structure and socially driven genetic differentiation in Hawaiian short-finned pilot whales. Mol Ecol. 26(23):6730–6741. doi:10.1111/mec.14397.

VanHelden AL, Baker AN, Dalebout ML, Reyes JC, Van Waerebeek K, Baker CS. 2002. Resurrection of *Mesoplodon traversii* (Gray, 1874), senior synonym of M. bahamondi

Reyes, Van Waerebeek, Cardenas and Yanez, 1995 (Cetacea: Ziphiidae). Mar Mammal Sci. 18(July):609–621.

Víkingsson GA, Pike DG, Valdimarsson H, Schleimer A, Gunnlaugsson T, Silva T, Elvarsson BP, Mikkelsen B, Øien N, Desportes G, et al. 2015. Distribution, abundance, and feeding ecology of baleen whales in Icelandic waters: Have recent environmental changes had an effect? Front Ecol Evol. 3:1–18. doi:10.3389/fevo.2015.00006.

Visser F, Oudejans MG, Keller OA, Madsen PT, Mark JP. 2022. Sowerby's beaked whale biosonar and movement strategy indicate deep-sea foraging niche differentiation in mesoplodont whales. J Exp Biol. 255(9). doi:10.1242/jeb.243728.

Waples RS. 1995. Evolutionarily significant units and the conservation of biological diversity under the Endangered Species Act. Am Fish Soc Symp. 17(1):8–27. doi:10.1111/j.1523-1739.1998.97524.x.

Waples RS. 2006. A bias correction for estimates of effective population size based on linkage disequilibrium at unlinked gene loci. Conserv Genet. 7(2):167–184. doi:10.1007/s10592-005-9100-y.

Waples RS, Antao T, Luikart G. 2014. Effects of overlapping generations on linkage disequilibrium estimates of effective population size. Genetics. 197(2):769–780. doi:10.1534/genetics.114.164822.

Waples RS, Gaggiotti OE. 2006. What is a population? An empirical evaluation of some genetic methods for identifying the number of gene pools and their degree of connectivity. Mol Ecol. 15:1419–1439. doi:10.1111/j.1365-294X.2006.02890.x.

Waples RS, Rus Hoelzel A, Gaggiotti OE, Tiedemann R, Palsbøll PJ, Cipriano F, Jackson JA, Bickham JW, Lang AR. 2018. Guidelines for genetic data analysis. J Cetacean Res Manag. 18:33–80.

Waring GT, Palka DL, Evans PGH. 2009. North Atlantic Marine Mammals. In: Perrin WF, Wursig B, Thewissen JGM, editors. Encyclopedia of Marine Mammals, 2nd Edition. Academic Press. p. 773–781.

Warmuth VM, Ellegren H. 2019. Genotype-free estimation of allele frequencies reduces bias and improves demographic inference from RADSeq data. Mol Ecol Resour. 19(3):586–596. doi:10.1111/1755-0998.12990.

van Weelden C, Towers JR, Bosker T. 2021. Impacts of climate change on cetacean distribution, habitat and migration. Clim Chang Ecol. 1(November 2020):100009. doi:10.1016/j.ecochg.2021.100009.

Weilgart LS. 2007. The impacts of anthropogenic ocean noise on cetaceans and implications for management. Can J Zool. 85:1091–1116. doi:10.1139/Z07-101.

Weinberg W. 1908. Über den Nachweis der Vererbung beim Menschen. Jahreshefte des Vereins Varterländische Naturkdunde Württemb. 64:369–382.

Weir BS, Cockerham CC. 1984. Estimating F-statistics for the analysis of population structure. Evolution (N Y). 38(6):1358–1370. doi:10.2307/2408641. https://www.jstor.org/stable/2408641?origin=crossref.

West KL, Walker WA, Baird RW, Mead JG, Collins PW. 2017. Diet of Cuvier's beaked whales *Ziphius cavirostris* from the North Pacific and a comparison with their diet world-wide. Mar Ecol Prog Ser. 574:227–242. doi:10.3354/meps12214.

Westbury M V., Thompson KF, Louis M, Cabrera AA, Skovrind M, Castruita JAS,

Constantine R, Stevens JR, Lorenzen ED. 2021. Ocean-wide genomic variation in Gray's beaked whales, *Mesoplodon grayi*. R Soc Open Sci. 8(201788):1–14.

Wetterstrand K a. 2021. DNA Sequencing Costs: Data from the NHGRI Genome Sequencing Program (GSP). [accessed 2022 Nov 8]. www.genome.gov/sequencingcosts.

Whitehead H, Hooker SK. 2012. Uncertain status of the northern bottlenose whale *Hyperoodon ampullatus*: Population fragmentation, legacy of whaling and current threats. Endanger Species Res. 19:47–61. doi:10.3354/esr00458.

Whitehead H, MacLeod CD, Rodhouse P. 2003. Differences in niche breadth among some teuthivorous mesopelagic marine mammals. Mar Mammal Sci. 19(2):400–406. doi:10.1111/j.1748-7692.2003.tb01118.x.

Whitehead H, McGill B, Worm B. 2008. Diversity of deep-water cetaceans in relation to temperature: Implications for ocean warming. Ecol Lett. 11(11):1198–1207. doi:10.1111/j.1461-0248.2008.01234.x.

Whitehead H, Reeves R, Feyrer LJ, Brownell Jr. RL. 2021. *Hyperoodon ampullatus*, Northern Bottlenose Whale. IUCN Red List Threat Species. e.T10707A5.

Williamson MJ, ten Doeschate M, Deaville R, Brownlow AC, Taylor NL. 2021. Cetaceans as sentinels for informing climate change policy in UK waters. Mar Policy. 131:104634. doi:10.1016/j.marpol.2021.104634.

Winkelmann I, Campos PF, Strugnell JM, Cherel Y, Smith PJ, Kubodera T, Allcock L, Kampmann M-L, Schroeder H, Guerra A, et al. 2013. Mitochondrial genome diversity and population structure of the giant squid *Architeuthis*: Genetics sheds new light on one of the most enigmatic marine species. Proc R Soc B Biol Sci. 280:20130273–20130273. doi:10.1098/rspb.2013.0273.

Worm B, Lotze HK, Myers RA. 2003. Predator diversity hotspots in the blue ocean. Proc Natl Acad Sci U S A. 100(17):9884–9888. doi:10.1073/pnas.1333941100.

Wright S. 1931. Evolution in Mendelian populations. Genetics. 16(2):97–159.

Wright S. 1938. Size of a population and breeding structure in relation to evolution. Science. 87(430–431).

Wright S. 1951. The genetical structure of populations. Ann Hum Eugen. 15:323–354. doi:10.2307/2407273.

Wright S. 1965. The interpretation of population structure by F-statistics with special regard to systems of mating. Evolution (N Y). 19:395–420.

Wright S. 1978. Evolution and the genetics of populations: a treatise in four volumes: vol. 4: viability within and among natural populations. University of Chicago Press.

Xue AT, Hickerson MJ. 2017. Multi-Dice: R package for comparative population genomic inference under hierarchical co-demographic models of independent single-population size changes. Mol Ecol Resour. 17(6):e212–e224. doi:10.1111/1755-0998.12686.

Yamada TK, Kitamura S, Abe S, Tajima Y, Matsuda A, Mead JG, Matsuishi TF. 2019. Description of a new species of beaked whale (*Berardius*) found in the North Pacific. Sci Rep. 9(12723):1–14. doi:10.1038/s41598-019-46703-w.

Yu G, Smith DK, Zhu H, Guan Y, Lam TTY. 2017. GGTREE: an R package for visualization and annotation of phylogenetic trees with their covariates and other

associated data. Methods Ecol Evol. 8:28–36. doi:10.1111/2041-210X.12628.

Zhou X, Guang X, Sun D, Xu S, Li M, Seim I, Jie W, Yang L, Zhu Q, Xu J, et al. 2018. Population genomics of finless porpoises reveal an incipient cetacean species adapted to freshwater. Nat Commun. 9(1):1–8. doi:10.1038/s41467-018-03722-x.

Zimmer WMX, Tyack PL. 2007. Repetitive shallow dives pose decompression risk in deep-diving beaked whales. Mar Mammal Sci. 23(4):888–925. doi:10.1111/j.1748-7692.2007.00152.x.

Appendix A: Full sample dataset

The full dataset of all samples sequenced for this thesis including sample information and metadata as well as the sample contributor and the permits/ethical considerations required to collect the sample is stored in the University of St Andrews PURE research database with the title: “Onoufriou_DNAintheDeep_AppendixA”

Appendix B: Outcomes of optimising 'gstacks' parameters

Results of the 'gstacks' parameter optimisation described in 2.8.1: the chapter that the parameters were optimised for, the species of the ddRAD or GBS sequences, original sample size of the optimised dataset, species that the sequenced were aligned to, the four 'gstacks' parameters (min-mapq: minimum mapping quality score to consider a read, max-clipped: maximum soft-clipping level as a fraction of the read length, var-alpha: SNP discovery threshold and gt-alpha: genotype calling threshold), and the resulting SNP dataset data (number of variable loci, mean missingess, mean site depth). The rows highlighted in green represent the parameter sets that were selected.

Chapter	Species	*n*	Alignment	mapq	sclip	var_alpha	gt_alpha	Loci (ind)	Mean missingness	Mean site depth
3	Cuvier's	40	Cuvier's	20	0.1	0.01	0.05	211309	0.0133	81.6
3	Cuvier's	40	Cuvier's	20	0.2	0.01	0.05	212407	0.0133	81.6
3	Cuvier's	40	Cuvier's	20	0.1	0.05	0.05	211660	0.0133	81.5
3	Cuvier's	40	Cuvier's	20	0.2	0.05	0.05	212795	0.0133	81.5
3	Cuvier's	40	Cuvier's	10	0.1	0.01	0.05	212480	0.0134	81.6
3	Cuvier's	40	Cuvier's	10	0.2	0.01	0.05	213629	0.0134	81.7
3	Cuvier's	40	Cuvier's	10	0.1	0.05	0.05	212855	0.0134	81.5
3	Cuvier's	40	Cuvier's	10	0.2	0.05	0.05	214010	0.0134	81.6
3	Cuvier's	40	Cuvier's	20	0.1	0.01	0.01	201891	0.0144	83.8
3	Cuvier's	40	Cuvier's	20	0.1	0.05	0.01	202259	0.0145	83.7
3	Cuvier's	40	Cuvier's	20	0.2	0.01	0.01	202975	0.0145	83.8
3	Cuvier's	40	Cuvier's	20	0.2	0.05	0.01	203368	0.0145	83.7
3	Cuvier's	40	Cuvier's	10	0.1	0.01	0.01	203026	0.0145	83.8
3	Cuvier's	40	Cuvier's	10	0.1	0.05	0.01	203419	0.0146	83.7
3	Cuvier's	40	Cuvier's	10	0.2	0.01	0.01	204157	0.0146	83.8
3	Cuvier's	40	Cuvier's	10	0.2	0.05	0.01	204551	0.0146	83.7

Chapter	Species	*n*	Alignment	mapq	sclip	var_alpha	gt_alpha	Loci (ind)	Mean missingness	Mean site depth
3	Blainville's	56	Sowerby's	10	0.2	0.05	0.05	185884	0.0129	52.2
3	Blainville's	56	Sowerby's	10	0.2	0.01	0.05	185482	0.0129	52.3
3	Blainville's	56	Sowerby's	20	0.2	0.05	0.05	184510	0.0128	52.1
3	Blainville's	56	Sowerby's	10	0.1	0.05	0.05	184194	0.0129	52.2
3	Blainville's	56	Sowerby's	20	0.2	0.01	0.05	184112	0.0128	52.2
3	Blainville's	56	Sowerby's	10	0.1	0.01	0.05	183784	0.0129	52.2
3	Blainville's	56	Sowerby's	10	0.2	0.05	0.01	183095	0.0150	52.5
3	Blainville's	56	Sowerby's	20	0.1	0.05	0.05	182888	0.0128	52.1
3	Blainville's	56	Sowerby's	10	0.2	0.01	0.01	182682	0.0150	52.5
3	Blainville's	56	Sowerby's	20	0.1	0.01	0.05	182486	0.0128	52.2
3	Blainville's	56	Sowerby's	20	0.2	0.05	0.01	181803	0.0149	52.4
3	Blainville's	56	Sowerby's	10	0.1	0.05	0.01	181464	0.0150	52.4
3	Blainville's	56	Sowerby's	20	0.2	0.01	0.01	181396	0.0149	52.5
3	Blainville's	56	Sowerby's	10	0.1	0.01	0.01	181043	0.0150	52.5
3	Blainville's	56	Sowerby's	20	0.1	0.05	0.01	180239	0.0149	52.4
3	Blainville's	56	Sowerby's	20	0.1	0.01	0.01	179825	0.0149	52.4
4	N. bottlenose	58	Cuvier's	10	0.1	0.01	0.01	964	0.0363	10.9
4	N. bottlenose	58	Cuvier's	10	0.1	0.01	0.05	974	0.0172	11.0
4	N. bottlenose	58	Cuvier's	10	0.1	0.05	0.01	1118	0.0316	10.6
4	N. bottlenose	58	Cuvier's	10	0.1	0.05	0.05	1128	0.0151	10.7
4	N. bottlenose	58	Cuvier's	10	0.2	0.01	0.01	1725	0.0351	10.4
4	N. bottlenose	58	Cuvier's	10	0.2	0.01	0.05	1747	0.0154	10.5
4	N. bottlenose	58	Cuvier's	10	0.2	0.05	0.01	2038	0.0299	10.2

Chapter	Species	*n*	Alignment	mapq	sclip	var_alpha	gt_alpha	Loci (ind)	Mean missingness	Mean site depth
4	N. bottlenose	58	Cuvier's	10	0.2	0.05	0.05	2060	0.0133	10.2
4	N. bottlenose	58	Cuvier's	20	0.1	0.01	0.01	919	0.0360	10.6
4	N. bottlenose	58	Cuvier's	20	0.1	0.01	0.05	928	0.0168	10.7
4	N. bottlenose	58	Cuvier's	20	0.1	0.05	0.01	1070	0.0311	10.4
4	N. bottlenose	58	Cuvier's	20	0.1	0.05	0.05	1079	0.0147	10.4
4	N. bottlenose	58	Cuvier's	20	0.2	0.01	0.01	1662	0.0347	10.2
4	N. bottlenose	58	Cuvier's	20	0.2	0.01	0.05	1681	0.0151	10.3
4	N. bottlenose	58	Cuvier's	20	0.2	0.05	0.01	1974	0.0294	10.1
4	N. bottlenose	58	Cuvier's	20	0.2	0.05	0.05	1995	0.0130	10.1
4	Sowerby's	38	Cuvier's	10	0.1	0.05	0.05	57926	0.0119	68.9
4	Sowerby's	38	Cuvier's	10	0.1	0.05	0.01	57617	0.0184	69.1
4	Sowerby's	38	Cuvier's	10	0.1	0.01	0.05	57460	0.0118	69.2
4	Sowerby's	38	Cuvier's	10	0.1	0.01	0.01	57205	0.0184	69.3
4	Sowerby's	38	Cuvier's	10	0.2	0.05	0.05	58495	0.0119	69.0
4	Sowerby's	38	Cuvier's	10	0.2	0.05	0.01	58183	0.0184	69.2
4	Sowerby's	38	Cuvier's	10	0.2	0.01	0.05	58118	0.0119	69.2
4	Sowerby's	38	Cuvier's	10	0.2	0.01	0.01	57762	0.0185	69.4
4	Sowerby's	38	Cuvier's	20	0.1	0.05	0.05	57286	0.0117	68.8
4	Sowerby's	38	Cuvier's	20	0.1	0.05	0.01	56989	0.0182	69.0
4	Sowerby's	38	Cuvier's	20	0.1	0.01	0.05	56846	0.0116	69.1
4	Sowerby's	38	Cuvier's	20	0.1	0.01	0.01	56547	0.0182	69.3
4	Sowerby's	38	Cuvier's	20	0.2	0.05	0.05	57835	0.0117	68.9
4	Sowerby's	38	Cuvier's	20	0.2	0.05	0.01	57538	0.0182	69.1

Chapter	Species	*n*	Alignment	mapq	sclip	var_alpha	gt_alpha	Loci (ind)	Mean missingness	Mean site depth
4	Sowerby's	38	Cuvier's	20	0.2	0.01	0.05	57388	0.0116	69.1
4	Sowerby's	38	Cuvier's	20	0.2	0.01	0.01	57088	0.0182	69.3
4	Blainville's	32	Cuvier's	10	0.1	0.05	0.05	46869	0.0141	49.7
4	Blainville's	32	Cuvier's	10	0.1	0.05	0.01	46603	0.0227	49.8
4	Blainville's	32	Cuvier's	10	0.1	0.01	0.05	46439	0.0141	49.8
4	Blainville's	32	Cuvier's	10	0.1	0.01	0.01	46172	0.0227	49.9
4	Blainville's	32	Cuvier's	10	0.2	0.05	0.05	47364	0.0142	49.7
4	Blainville's	32	Cuvier's	10	0.2	0.05	0.01	47090	0.0228	49.8
4	Blainville's	32	Cuvier's	10	0.2	0.01	0.05	46921	0.0141	49.9
4	Blainville's	32	Cuvier's	10	0.2	0.01	0.01	46649	0.0227	49.9
4	Blainville's	32	Cuvier's	20	0.1	0.05	0.05	46114	0.0137	49.6
4	Blainville's	32	Cuvier's	20	0.1	0.05	0.01	45898	0.0223	49.7
4	Blainville's	32	Cuvier's	20	0.1	0.01	0.05	45701	0.0137	49.8
4	Blainville's	32	Cuvier's	20	0.1	0.01	0.01	45487	0.0222	49.8
4	Blainville's	32	Cuvier's	20	0.2	0.05	0.05	46591	0.0138	49.6
4	Blainville's	32	Cuvier's	20	0.2	0.05	0.01	46366	0.0223	49.7
4	Blainville's	32	Cuvier's	20	0.2	0.01	0.05	46171	0.0137	49.8
4	Blainville's	32	Cuvier's	20	0.2	0.01	0.01	45951	0.0223	49.9
4	Cuvier's	89	Cuvier's	10	0.1	0.05	0.05	223406	0.0205	54.6
4	Cuvier's	89	Cuvier's	10	0.1	0.05	0.01	221872	0.0253	54.7
4	Cuvier's	89	Cuvier's	10	0.1	0.01	0.05	222440	0.0205	54.7
4	Cuvier's	89	Cuvier's	10	0.1	0.01	0.01	220975	0.0252	54.8
4	Cuvier's	89	Cuvier's	10	0.2	0.05	0.05	224614	0.0206	54.6

Chapter	Species	*n*	Alignment	mapq	sclip	var_alpha	gt_alpha	Loci (ind)	Mean missingness	Mean site depth
4	Cuvier's	89	Cuvier's	10	0.2	0.05	0.01	223023	0.0254	54.7
4	Cuvier's	89	Cuvier's	10	0.2	0.01	0.05	223656	0.0206	54.7
4	Cuvier's	89	Cuvier's	10	0.2	0.01	0.01	222131	0.0253	54.8
4	Cuvier's	89	Cuvier's	20	0.1	0.05	0.05	221885	0.0204	54.5
4	Cuvier's	89	Cuvier's	20	0.1	0.05	0.01	220424	0.0251	54.6
4	Cuvier's	89	Cuvier's	20	0.1	0.01	0.05	220915	0.0203	54.6
4	Cuvier's	89	Cuvier's	20	0.1	0.01	0.01	219522	0.0251	54.7
4	Cuvier's	89	Cuvier's	20	0.2	0.05	0.05	223057	0.0205	54.6
4	Cuvier's	89	Cuvier's	20	0.2	0.05	0.01	221536	0.0252	54.7
4	Cuvier's	89	Cuvier's	20	0.2	0.01	0.05	222091	0.0204	54.6
4	Cuvier's	89	Cuvier's	20	0.2	0.01	0.01	220631	0.0252	54.7

Appendix C: Supplementary materials for “Biogeography in the deep”

The following pages contain the electronic supplementary material (ESM) and supplementary spreadsheet tables (SST) that were published alongside the manuscript "Biogeography in the deep: Hierarchical population genomic structure of two beaked whale species" in *Global Ecology and Conservation* (https://doi.org/10.1016/j.gecco.2022.e02308).

C.1 ESM 1: TISSUE ARCHIVE, DNA EXTRACTION AND SAMPLE SELECTION

Tissue samples were either collected specifically for this project or provided on loan from archives maintained by contributors. A large set of samples was provided by Dr. Merel Dalebout, who collated samples for investigations into beaked whale systematics and genetic diversity (Dalebout et al. 2002; Dalebout 2002; VanHelden et al. 2002; Dalebout et al. 2003; Dalebout et al. 2004; Dalebout et al. 2005; Dalebout et al. 2006; Gomerčić et al. 2006; Dalebout et al. 2007; Dalebout et al. 2008; Dalebout et al. 2014). All samples were either skin biopsies sampled directly from free-swimming animals (Lambertsen 1987; Krützen et al. 2002), tissue collected from dead beach-cast or ship-strike individuals, or already extracted DNA provided by NOAA's Southwest Fisheries Science Centre Marine Mammal and Turtle Molecular Research Sample Collection as already extracted DNA (https://www.fisheries.noaa.gov/west-coast/science-data/marine-mammal-and-sea-turtle-research-tissue-collection). Freshly collected tissue samples were typically stored in either DMSO or 70-99% ethanol and stored at -20°C.

The purpose of the current study was to develop an understanding of the global baseline genetic diversity and structure of Blainville's and Cuvier's beaked whales (*Mesoplodon densirostris* and *Ziphius cavirostris,* respectively and henceforth 'Blainville's' and 'Cuvier's'), providing context for investigating the impacts of anthropogenic activities on resident populations in future studies. Balancing the coverage of samples across their respective distributions and budgetary constraints, it was decided that five lanes of sequencing on a HiSeq 2500 (Illumina) would be used for the current study. The pilot study by Carroll et al. (2016) concluded that sequencing up to 50 individuals per HiSeq2500 lane would generate ~10k variable SNPs per individual, a number likely to detect structure in these two beaked whale species.

Of the 89 Blainville's individuals in the sample archive, n=67 were available as tissue, n=21 were already extracted DNA, and one individual had both a tissue and DNA sample. Of the 340 Cuvier's individuals, n=289 were available as tissue, n=29 were DNA, n=22 had both tissue and already extracted DNA. DNA for the ddRAD and mitogenome analyses was extracted from approximately 30-50mg of tissue, using the Phenol:Chloroform:Isoamyl Alcohol method described by Sambrook, Fritsch, & Maniatis (1989) and modified for use in small tissue samples by Baker et al. (1994). In addition, a subset of the Cuvier's samples used in the mitogenome analysis were extracted using a KingFisher Duo™ (Thermo Scientific™) automated extraction and purification instrument, following the manufacturer's instructions. Extracted DNA was checked for quality using a NanoDrop™ (Thermo Scientific™) spectrophotometer and gel electrophoresis and quantified using a Qubit (Invitrogen) fluorometer.

As the protocol for double-digest restriction site-associated DNA sequencing (ddRAD) requires both high molecular weight and high concentration DNA (>20ng/ul), a scoring system was developed to rank samples (Table S1.1) prior to preparing libraries for pooling and sequencing. DNA was run on 1.2% agarose gels to assess the overall quality

of the sample and the concentration was measured using spectrophotometry (NanoDrop) to obtain an approximate value. Some samples that yielded poor quality DNA were extracted up to two more times (*n*=4 Blainville's, *n*=68 Cuvier's). More precise measurements of DNA quantity were made using fluorometry (Qubit) for *n*=88 Blainville's and *n*=302 Cuvier's that either had visible amounts of DNA on the agarose gel or quantifiable amounts of DNA on the spectrophotometer. Samples selected based on their DNA score were pooled into libraries with individuals that shared the same score whenever possible.

Samples were also selected to ensure every geographical location possible was covered, and to fill in the rest of the sequencing lanes, samples from well-studied resident populations were prioritised (Hawai'i, Canary Islands, Bahamas, and Ligurian Sea). Although of poorer quality, DNA samples with scores lower than 'Good' were included in the libraries as many of them came from poorly sampled areas. Table S1.2 shows the geographical origin for the 170 Cuvier's and 55 Blainville's samples that were sequenced. Following sequencing and the bioinformatic steps outlined in Supplementary 3 and 4, several individuals was removed from the analysis that failed to pass quality control (QC). Table S1.3 shows the number of individuals from both Cuvier's and Blainville's that either passed or failed QC according to the score assigned to them based on DNA quality/quantity before sequencing.

The final list of individuals used in this study is found in Supplementary Table 1 (SST1). Once duplicate individuals were removed, the number of unique individuals for ddRAD sequencing remaining was *n*=161 Cuvier's and *n*=55 Blainville's.

Table S1.1. The scoring system developed to rank DNA samples before ddRAD library preparation based on the amount of DNA in the sample (measured using a Qubit fluorometer) and the molecular weight according to 1.2% agarose gels (HMW =High Molecular Weight, Smear=degraded DNA of varying sizes, LMW= Low Molecular Weight).

Score	DNA Concentration (Qubit)	DNA Gel Result
Great	>20ng/ul	HMW
Good	>20ng/ul	HMW + smear
Good	>20ng/ul	Smear
Good	>20ng/ul	No visible DNA
Good	15-20ng/ul	HMW
OK	>20ng/ul	LMW
OK	15-20ng/ul	Smear
OK	<15ng/ul	HMW
Some	<15ng/ul	Faint HMW
Some	<15ng/ul	Smear
Some	<15ng/ul	No visible DNA

Table S1.2. Broad geographic origin of the samples selected for ddRAD library preparation and sequencing for *n*=170 Cuvier's (*Ziphius cavirostris*) and *n*=55 Blainville's beaked whales (*Mesoplodon densirostris*).

	Cuvier's (*n*=170)	Blainville's (*n*=55)
North Atlantic	89	34
North Pacific	28	8
South Pacific	15	8
South Africa	2	5
Mediterranean	36	Not present

Table S1.3. The number of Cuvier's (*Ziphius cavirostris*) and Blainville's beaked whales (*Mesoplodon densirostris*) that passed or failed the ddRAD quality control steps based on the DNA quality/quantity score described in Table S1.1.

	Cuvier's (*n*=170)		Blainville's (*n*=55)	
Quality Score	No. Passed QC	No. Failed QC	No. Passed QC	No. Failed QC
Great	102	29	39	9
Good	19	15	2	0
OK	2	3	1	3
Some	0	0	1	0

C.2 ESM2: DDRAD AND MITOGENOME LIBRARY PREPARATION AND SEQUENCING

ddRAD builds upon the earlier RADseq method (N.A. Baird et al. 2008) by adding a second restriction enzyme (RE) to the digest and an explicit size-selection step, allowing researchers to have more control over the fraction of the genome that is sequenced (Peterson et al. 2012). In this protocol, samples were digested with one RE that targeted a commonly occurring motif (MspI, 4bp) and a rarely occurring motif (HindIII, 6bp). Unique P1 adaptors containing individual barcodes and PCR primers, and universal P2 adaptors with PCR primers, were ligated to both ends of the digested DNA. The samples were then cleaned, pooled and size selected using a Pippin Prep (Sage Science). A PCR step added a secondary identifier (reverse index) to the P2 end and Illumina flow cell annealing sequences on both ends. After this step, samples were cleaned, pooled, and sequenced. By adding a second RE digest, and eliminating random mechanical shearing and broad size selection, individual studies are much more reproducible and precise. Limiting the DNA window that is sequenced and precisely selecting for size, means that the sequenced fragments from different individuals are more likely to be recovered from the same region of the genome (Peterson et al. 2012).

The following ddRAD protocol was optimised for beaked whale tissue by Carroll et al., (2021, 2016). Samples selected for sequencing were grouped based on their score, normalised to 20ng/µl and a total of 250ng of DNA per individual underwent an overnight digestion at 37°C with MspI and HindIII. After a 20-minute heat kill step at 65°C, adaptors were ligated with one of 10 forward barcodes per sample, using the temperature profile of 22°C for 2 hours followed by 65°C for 20 minutes. With unique barcodes now ligated, up to 10 samples, grouped according to quality classification, were pooled to form a library, and cleaned using three PureLink PCR Micro Kit columns (Invitrogen) per library. Following the final elution step, 30µl of cleaned ligate underwent size selection

to a 300-400 bp range using a Pippin Prep (Sage Science). The resulting size-selected ligate was divided into 8 wells and library-specific reverse indices were annealed during low-cycle number PCR using a Phusion High-Fidelity PCR kit (ThermoFisher). PCR products were pooled and cleaned using AMPURE-XP (Beckman-Coulter) beads and eluted to a final volume of 15µl in EB buffer (Qiagen). The final libraries were sent to the National High-Throughput DNA Sequencing Centre at the University of Copenhagen where the quantities and quality were determined with qPCR and Bioanalyzer (Agilent Genomics). Finally, the libraries were normalised and up to five libraries (~50 samples) were pooled into sequencing lanes and sequenced on a HiSeq2500 V4 chemistry (Illumina®).

For the mitogenome sequencing, we used the Carøe et al. (2018) protocol. Briefly, DNA was fragmented to approximately 350 bp, using the M220 Focused-Ultrasonicator™ (Covaris), according to the manufacturer's protocol. After fragmentation, samples were quantified on the Agilent 2200 TapeStation according to the protocol for genomic DNA, to verify fragmentation success. DNA libraries were built, using the blunt-end single-tube protocol described by Carøe et al. (2018) with a few modifications. To each library, 2 µl of 10 µM Illumina® adapters were added to the fragmented DNA, followed by a MiniElute (Qiagen) clean-up step before indexing with P5 and P7 indices. Libraries were sequenced using paired-end (PE) 150 bp chemistry on two lanes of Illumina® HiSeq 4000 at the National High-Throughput Sequencing Centre at University of Copenhagen, Denmark. In addition, 16 libraries characterized by average fragment lengths <300bp were sequenced on a single lane of HiSeq4000 using single-end 80 bp chemistry.

C.3 ESM3: DDRAD STACKS PARAMETER OPTIMISATION

The Stacks SNP discovery pipeline (v. 2, Rochette et al. 2019) implements the Bayesian genotype caller (BGC) algorithm of Maruki & Lynch (2015, 2017) in the command 'gstacks'. This algorithm uses a Bayesian genotype-frequency prior that takes into account population-level allele frequencies, does not assume Hardy-Weinberg equilibrium and estimates error rates directly from the sequence data (not from read quality scores) when calling genotypes (Maruki and Lynch 2015; Maruki and Lynch 2017). The program first estimates significant polymorphic loci from the read alignments with confidence set by the parameter "var-alpha" in 'gstacks' (Maruki and Lynch 2015; Rochette et al. 2019). Genotypes for each individual are called at these loci, using a method that takes into account allele balance and read depth (Maruki and Lynch 2017). Confidence in the genotype calling is done using a likelihood ratio test which compares the likelihood of the two most likely genotypes (Maruki and Lynch 2015).

A subset of demultiplexed and quality-controlled (QC) reads from Cuvier's (n=40) and Blainville's (n=55) individuals was selected to optimize the parameters used in the Stacks SNP discovery pipeline. The subset of high-quality samples (>98% retained reads and >1,000,000 retained reads) were selected to cover the widest geographical range of Cuvier's and reduce computing time, while all Blainville's individuals were selected. In summary, the following parameters were modified to optimise the 'gstacks' command for each dataset: "min-mapq" (minimum mapping quality score to consider a read; 10,20), "max-clipped" (maximum soft-clipping level as a fraction of the read length; 0.1, 0.2), "var-alpha" (SNP discovery threshold; 0.05, 0.01) and "gt-alpha" (genotype calling threshold; 0.05, 0.01). Samples were selected for optimisation based on the number and proportion of retained reads and to ensure an even distribution from all geographical

regions. The best combination of parameters was selected based on resulting datasets with the highest number of SNP loci and the lowest amount of missing data.

At the end of all Stacks and filtering steps, the optimal parameters were selected based on maximizing the total number of final SNP loci and reducing the amount of missing data per species. The final optimised parameters were the same for both Cuvier's and Blainville's samples: mapq=10, sclip=0.2, var_alpha=0.05, gt_alpha=0.05.

C.4 ESM4: DDRAD STACKS PROTOCOL, LOCI QUALITY CONTROL AND FILTERING STEPS

Following the discovery and genotyping of SNPs in 'gstacks', individuals and their genotyped loci can be analysed in a framework incorporating some sort of group assignment (such as geographical origin or sex) and then filtered according to minor allele frequency or locus frequency within the entire population in the Stacks 'populations' module. To reduce bias from potentially arbitrary population designations, no such population framework was provided in the current study. Massively parallel sequencing can lead to high error rates and genotypic uncertainties that can be introduced at any step throughout the analysis (O'Leary et al. 2018). Fortunately, many errors can be overcome by employing rigorous filtering to identify and reduce errors before analysing the final dataset (O'Leary et al. 2018). In this study, we took a tiered approach to filtering, starting with low cut-off values for missing data (applied separately per locus and individual) and finalizing the dataset with higher thresholds. This alternative and iterative filtering method, whereby you increase the cut-off threshold, has been shown to retain more loci and individuals as poor-quality individuals can deflate genotype call rates in otherwise acceptable loci, while poor-quality loci can increase the amount of missing data in otherwise acceptable individuals (O'Leary et al. 2018). Below we describe each of the steps that were implemented using R v. 3.6.0 (R Core Team 2019) and VCFtools v. 0.1.12a (Danecek et al. 2011) to filter individuals and loci based on the amount of missing data, read depth, and quality score (Table S4.1). In Table S4.2, each bioinformatic step is listed, with the resulting number of loci and individuals remaining throughout the process for both Cuvier's and Blainville's datasets.

Table S4.1. List of filtering commands and steps used in the program VCFtools to filter loci and individuals based on locus depth, genotype quality, minor allele frequency (MAF) and missingness. Low stringency indicates that lower cut-off values are used to filter out missing data before iteratively increasing cut-off values, a strategy shown by O'Leary et al. (2018) to increase the proportion of retained loci and individuals.

Stringency	VCFtools Command	Description
Low	--minDP 5 –minGQ20	Recode genotypes with quality <20 and depth <5 to zero
	--maf 0.001	Remove the sites made monomorphic by previous step.
	--max-missing 0.5	Remove sites with >50% missing data
	--missing-indv	Calculate missingness per individual, write a list of individuals with >50% missing data
	--remove	Remove individuals on list with >50% missing data
High	--site-depth	Calculate site depth, list loci with m >3x the overall mean
	--exclude-positions	Remove sites with site depth >3x the overall mean
	--max-missing 0.75	Remove sites with more than 75% missing data
	--missing-indv	Calculate missingness per individual, write a list of individuals with >25% missing data
	--remove	Remove individuals on list with >25% missing data

Table S4.2. Summary of each bioinformatic step to discover, genotype and filter loci based on the steps described in Table 1. Data are presented for the Cuvier's and Blainville's beaked whale (*Ziphius cavirostris* and *Mesoplodon densirostris,* respectively) datasets with a summary of each step and the program that was used.

	Species	**Cuvier's**	**Blainville's**
	starting sample size	170	56
optimised 'gstacks' parameters	mapq	10	10
	sclip	0.2	0.2
	var_alpha	0.05	0.05
	gt_alpha	0.05	0.05
'process_ radtags' (Stacks)	Total PE reads	1112692388	339961638
	Retained PE reads	1063993201	322949310
	Mean PE retained reads/sample	6258784	5766952
	% Retained (across all samples)	95.6%	95.0%
'gstacks' (Stacks)	Individuals Remaining	154	54
	Alignments Read	1085038314	334888850
	Alignments Kept	778173558	249685545
	% Alignments Kept	71.7%	74.6%
	Loci Built and Genotyped	1801998	979828
'populations' (Stacks)	Loci Kept	1795750	977060
	No. Sites	477045540	184528823
	No. Polymorphic Sites	2087305	1054577
Filtering (VCFtools)	Sites with >5x depth, >20 genotype quality, >0.001 MAF	1326391	735883
	Sites with <50% missing data	327270	340095
	Sites with depth <3x overall mean depth	326459	339441
	Sites with <75% missing data	262482	296250
'populations' (Stacks)	Loci passed filtering (whitelist)	34264	37617
	Loci present in >80% individuals with >0.01 MAF	31734	32610
	No. Sites	9994609	9527357
	No. Polymorphic Sites	30479	271983
	No. Genomic Sites	9610872	9504054
	No. Individuals remaining	123	49
"glPlot", Duplicates (*adegenet*)	Final no. Individuals	123	43
	Final no. Loci	30479	13988

C.5 ESM5: PHYLOGENETIC TREES WITH SOUTHERN RIGHT WHALE OUTGROUPS

Phylogenetic trees using the ddRAD SNP data were generated for both Cuvier's and Blainville's incorporating data from Southern right whales (SRW, *Eubalaena australis*) as the outgroup. The full SNP discovery pipeline as described in SM4 was repeated for the final *n*=123 Cuvier's and *n*=43 Blainville's, each time including sequence data from six SRWs. The Cuvier's + SRW dataset was aligned to the same Cuvier's genome (NCBI Genbank database accession: PRJNA399469) and the Blainville's + SRW sequences were aligned to the same Sowerby's genome (*Mesoplodon bidens*: PRJNA399476). All bioinformatic steps were followed as before, with one exception.

The final VCF file was uploaded into R and converted to a "genlight" file as before and BIONJ trees with bootstrap support (in %, based on 100 bootstraps) were produced in R ('poppr' v2.8.5; Kamvar, Brooks, & Grünwald, 2015 and 'ggtree'v.2.0.2; Yu, Smith, Zhu, Guan, & Lam, 2017). The trees were rooted using one of the SRW sequences and re-plotted. The SRW individuals were then dropped from the tree, to better visualize the ocean-level phylogenetic patterns of Cuvier's and Blainville's. The original BIONJ trees (without dropping the SRW outgroup are found in Figures S5.1 and S5.2. The final datasets (including 6 SRW samples) were *n*=118 Cuvier's (*n*=33137 SNPs) and *n*=42 Blainville's (*n*=29904 SNPs).

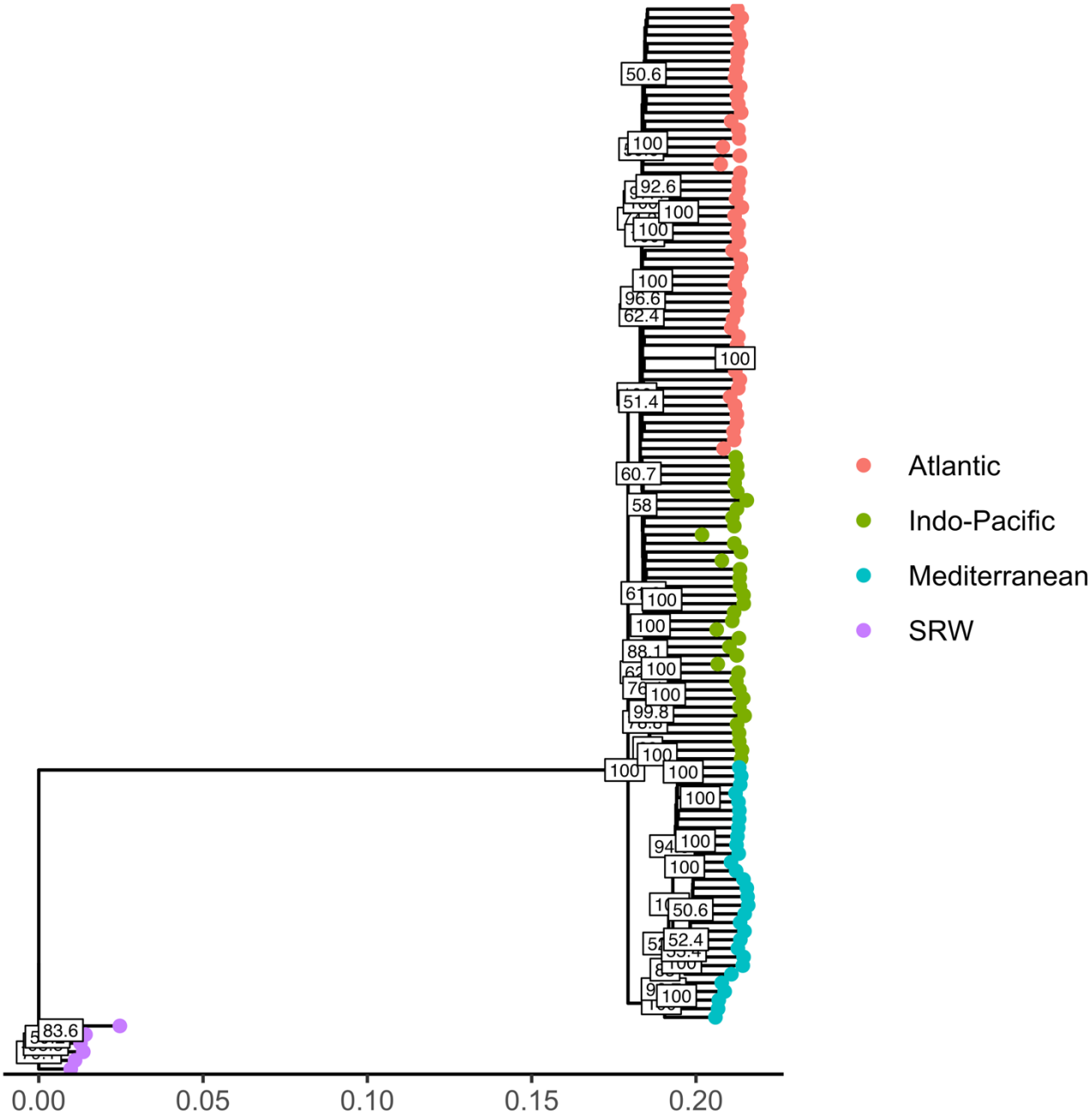


Figure S5.1. BIONJ phylogenetic tree of 118 Cuvier’s beaked whales (*Ziphius cavirostris*) and 6 Southern right whales (*Eubalaena australis*) as outgroups, generated using *n*=33137 ddRAD SNPs.

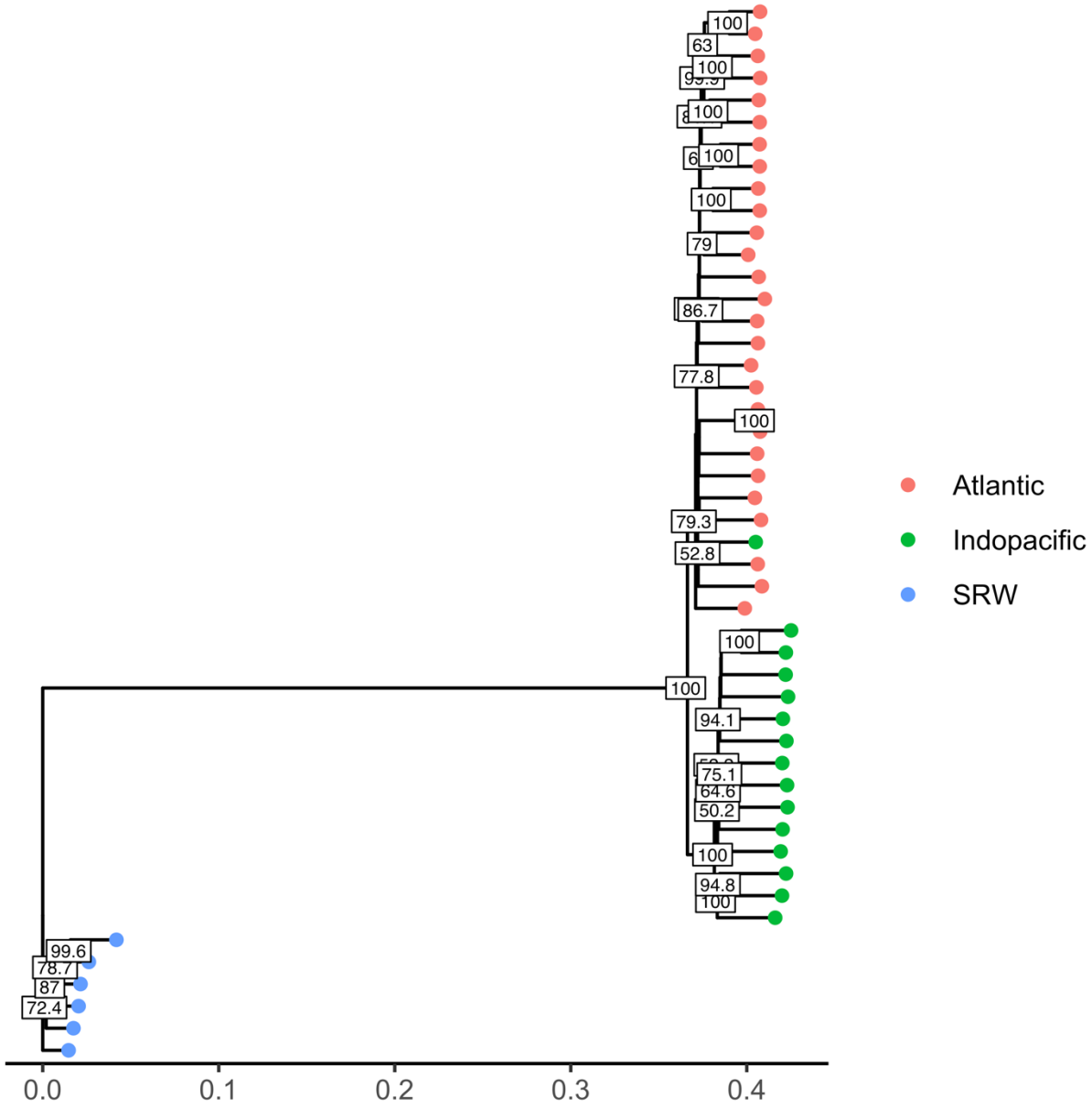


Figure S5.2. BIONJ phylogenetic tree of 42 Blainville’s beaked whales (*Mesoplodon densirostris*) and 6 Southern right whales (*Eubalaena australis*) as outgroups, generated using *n*=329904 ddRAD SNPs.

C.6 ESM6: ISOLATION-BY-DISTANCE

Isolation by distance (IBD) was calculated per species, and within ocean basin per species, using a Mantel test in *ade4* v1.7-16 in R (Dray and Dufour 2007) and geographical distances calculated via the least cost (LC) path distance over seawater in *marmap* v1.0.5 (Pante and Simon-Bouhet 2013). To calculate the LC distance over seawater, sampling locations were plotted on a global bathymetry map (with 4-minute resolution) using the "getNOAA.bathy" function in *marmap*. The resolution of the world bathymetry map resulted in some sampling locations of stranded individuals to be on land, and therefore incur a great coast in the LC path. Sample coordinates were therefore adjusted to the nearest -200m isobath using the "dist2isobath" function in *marmap*. The updated sample coordinates and bathymetry map were used to calculate a transition matrix using "trans.mat" in *marmap*, requiring the LC path to have a minimum depth of 200m. Finally, the LC path distance between each individual was calculated using "lc.dist" in *marmap*. The resulting pairwise matrix of geographic distances was used in combination with pairwise genetic distance (Euclidean) to run a Mantel test using "mantel.randtest" with 999 permutations. Mantel tests were conducted based on 999 replicates for all Cuvier's or Blainville's combined, and for individual ocean basins (Atlantic, Indo-Pacific, and Mediterranean- Cuvier's only) (Table S6.1). All correlation values were positive and all but the Indo-Pacific Blainville's were significant ($p<0.05$). The genetic and geographic distance matrices were plotted with a 2-dimensional kernel density estimation to visualize whether the apparent IBD was the result of a continuous cline or population clustering (Figures S6.1 and S6.2).

Table S6.1. The observation correlation and associated p-value of Mantel tests for Isolation by distance.

Species	Ocean Basin	*n*	Observation correlation (r)	P-value
Cuvier's	All	123	0.308	0.001
	Atlantic	54	0.154	0.001
	Indo-Pacific	36	0.162	0.028
	Mediterranean	33	0.218	0.002
Blainville's	All	43	0.665	0.001
	Atlantic	28	0.110	0.03
	Indo-Pacific	15	0.014	0.427

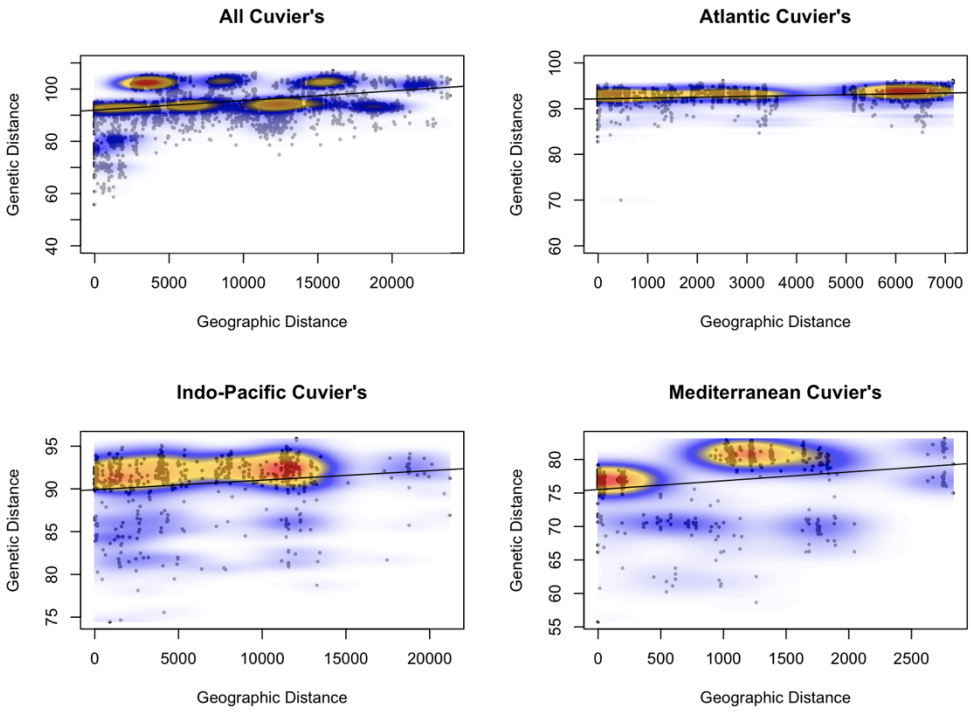


Figure S6.1. Scatterplots of genetic distance (Euclidian) and geographic distance (least cost (LC) path distance over seawater) overlaid with 2-dimensional kernel density estimation from *n*=123 Cuvier’s beaked whales (*Ziphius cavirostris, n*=30479 SNPs).

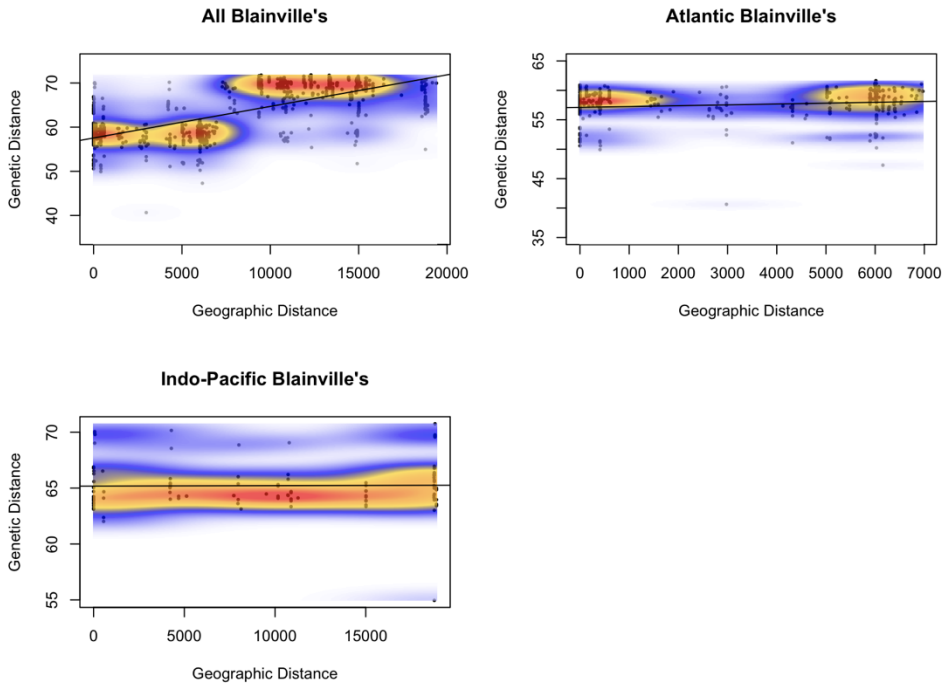


Figure S6.2 Scatterplots of genetic distance (Euclidian) and geographic distance (least cost (LC) path distance over seawater) overlaid with 2-dimensional kernel density estimation from *n*=43 Blainville’s beaked whales (*Mesoplodon densirostris, n*=13988 SNPs).

C.7 ESM7: TESS3R CROSS-ENTROPY SCORES

The R package *tess3r* (Caye et al. 2018) incorporates genotypic and geographical information (latitude and longitude coordinates for each sample) in a spatially explicit, least-squares optimization approach to estimate ancestry. The user defines and compares results from a range of ancestral population clusters (*K*) with the resulting bar plots displaying ancestry coefficients reflecting the probability of population membership and estimates of admixture. As opposed to the initial version of TESS (Chen et al. 2007; Durand et al. 2009), there is no biological model underlying this version however, the model does expect that individuals sampled in close geographical proximity are more likely to share ancestry than those sampled from further away.

For both Cuvier's and Blainville's, *tess3r* was run for *K*=2-10 and cross-entropy scores were plotted against *K* values to infer the most likely number of genetic clusters. In cross-entropy plots, smaller values indicate better fit with the best estimate of *K* corresponding to the value at which the curve reaches a plateau or starts to increase. In cases where a clear minimum or plateau is not observed, the *K* value that leads to the most parsimonious assignment of individuals (least amount of admixture) to populations can be considered as selection criteria. The figures below display the cross-entropy scores for: Cuvier's (Figure S7.1) and Blainville's (Figure S7.2).

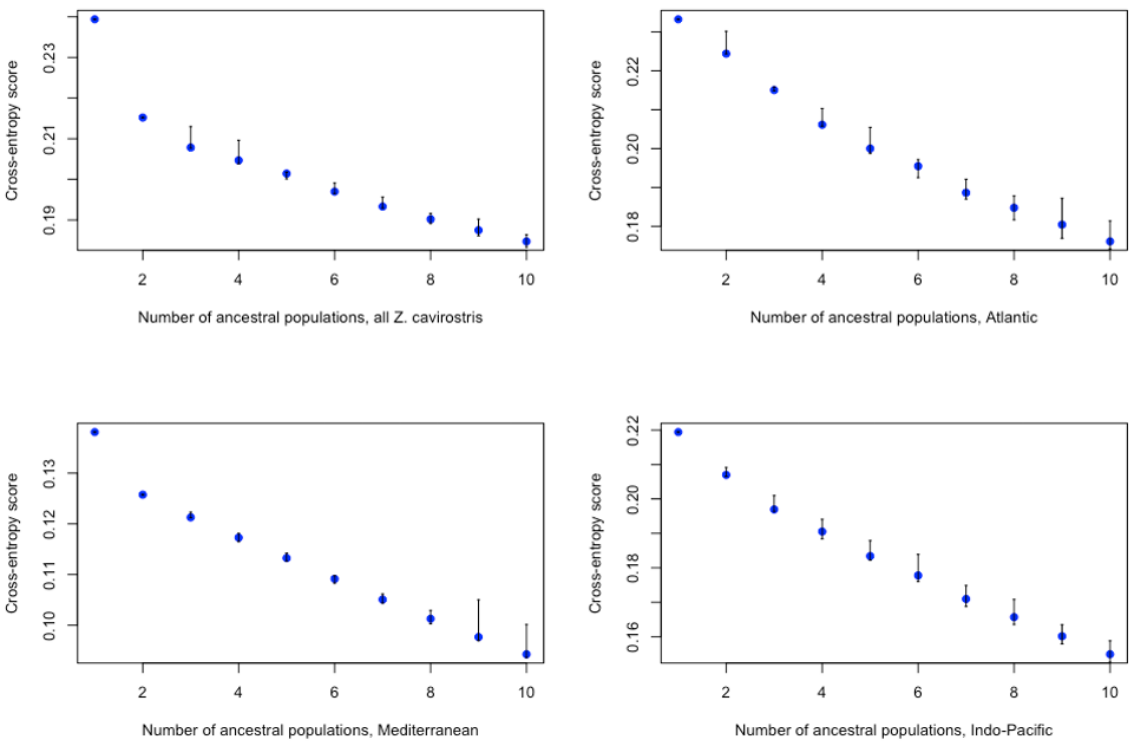


Figure S7.1. Cross-entropy scores of *K*=2-10 genetic clusters generated using *tess3r* for *n*=123 Cuvier’s beaked whales (*Ziphius cavirostris*) sampled from across their global range (top left), the Atlantic (top right), Mediterranean (bottom left) and Indo-Pacific (bottom right).

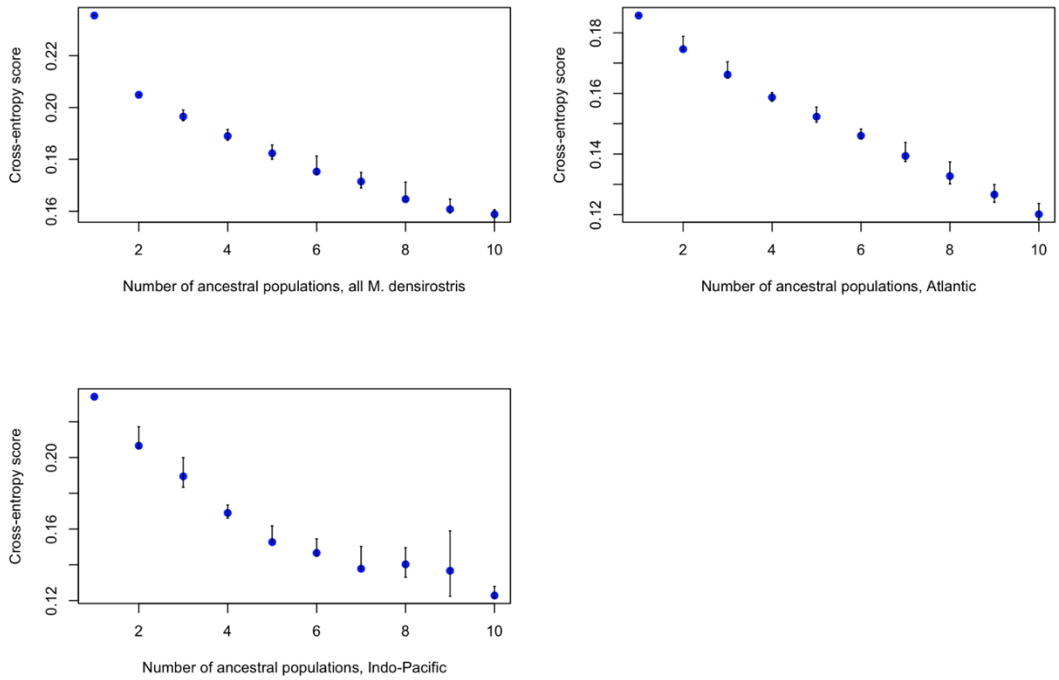


Figure S7.2. Cross-entropy scores of *K*=2-10 genetic clusters generated using *tess3r* for *n*=43 Blainville’s beaked whales (Mesoplodon densirostris) sampled from across their global range (top left), the Atlantic (top right), and Indo-Pacific (bottom left).

C.8 ESM8: DISCRIMINANT ANALYSIS OF PRINCIPLE COMPONENTS

Discriminant Analysis of Principle Components (DAPC) was conducted in the R package, *adegenet* (Jombart et al. 2010). DAPC is a useful tool to summarise the amount of genetic differentiation between groups (either determined *a priori* or *de novo* using K-means clustering) while ignoring the amount of variation within groups. To optimise the number of principle components (PCs) to retain from the analysis, *adegenet* offers a cross-validation tool with "xvalDapc". This command subsets the data to use as a training set, runs the analysis over a pre-determined number of repeats (*n*=30), and determines the best number of PCs to retain based on whichever yields the highest predictive success of the training data with the lowest root mean squared error (RMSE). The resulting DAPC can be plotted to observe the spatial structure of SNP genotypes across clusters and the function "assignplot" can be used to visualize the proportion of successful reassignment to the prior groups.

DAPC with cross-validation was conducted for *n*=118 Cuvier's (excluding Atl_Spain and Indo_Mix). The highest mean success and lowest MSE was achieved when 20 PCs were retained. The resulting scatterplot and assignment plot are found in figure S8.1. DAPC with cross-validation was conducted for *n*=43 Blainville's and the optimal number of PCs to retain was 10. The resulting scatter and assignment plots are in figure S8.2.

The presence of hierarchical structure requires investigation of more than the first and second axes to resolve finer scales. Scatter plots of the 2nd vs 3rd (Figure S8.3) and 3rd vs 4th (Figure S8.4) axes help to discriminate between the genetic clusters found in Cuvier's within the Atlantic and Indo-Pacific.

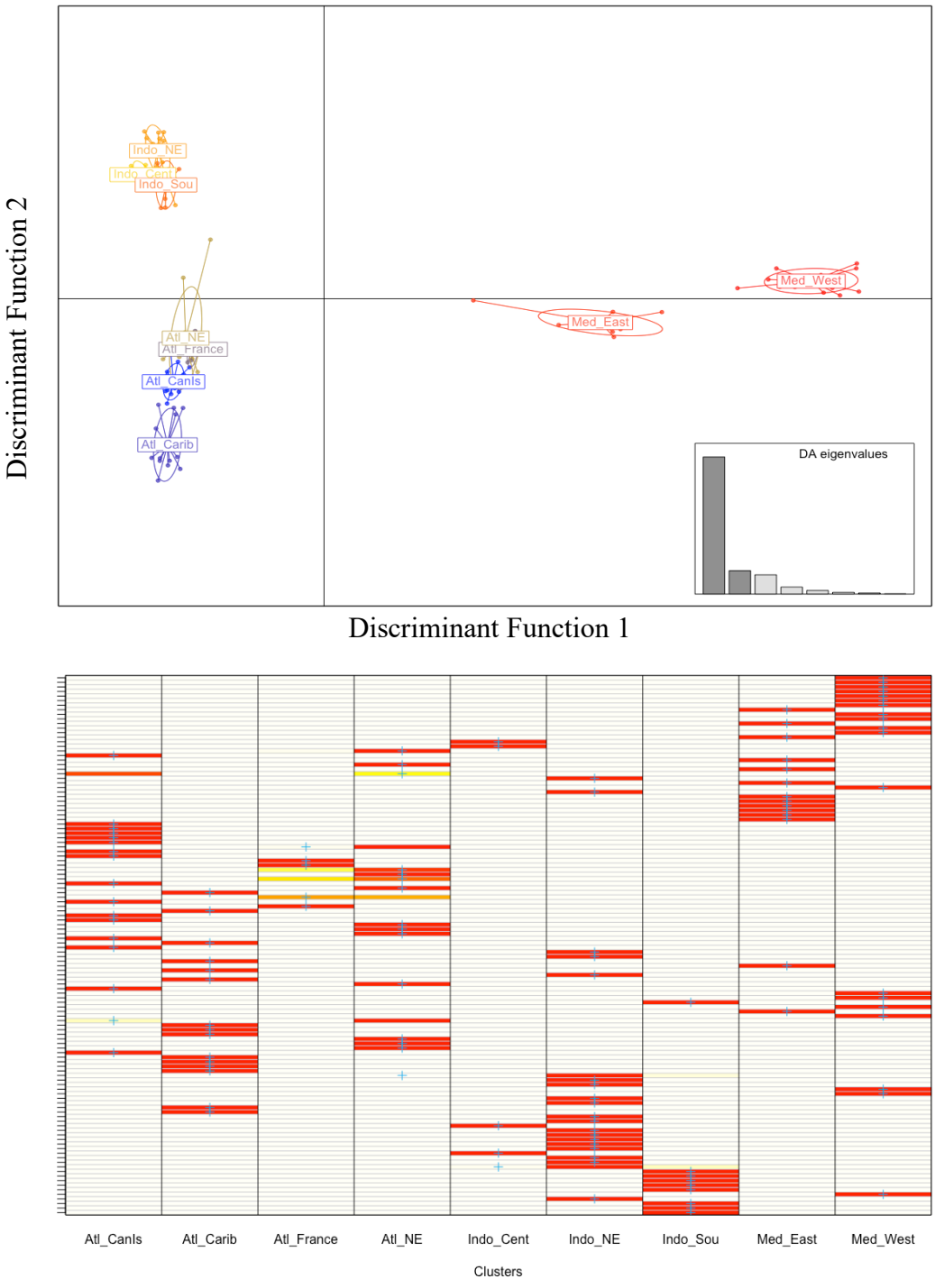


Figure S8.1. DAPC scatter (top) and assignment (bottom) plot for *n*=118 Cuvier’s beaked whales (*Ziphius cavirostris*) generated using cross-validation and retaining 20 PCs. In the assignment plot, each row represents an individual, the blue cross indicates the prior cluster assignment, and the colours represent membership probability (red=1, white=0).

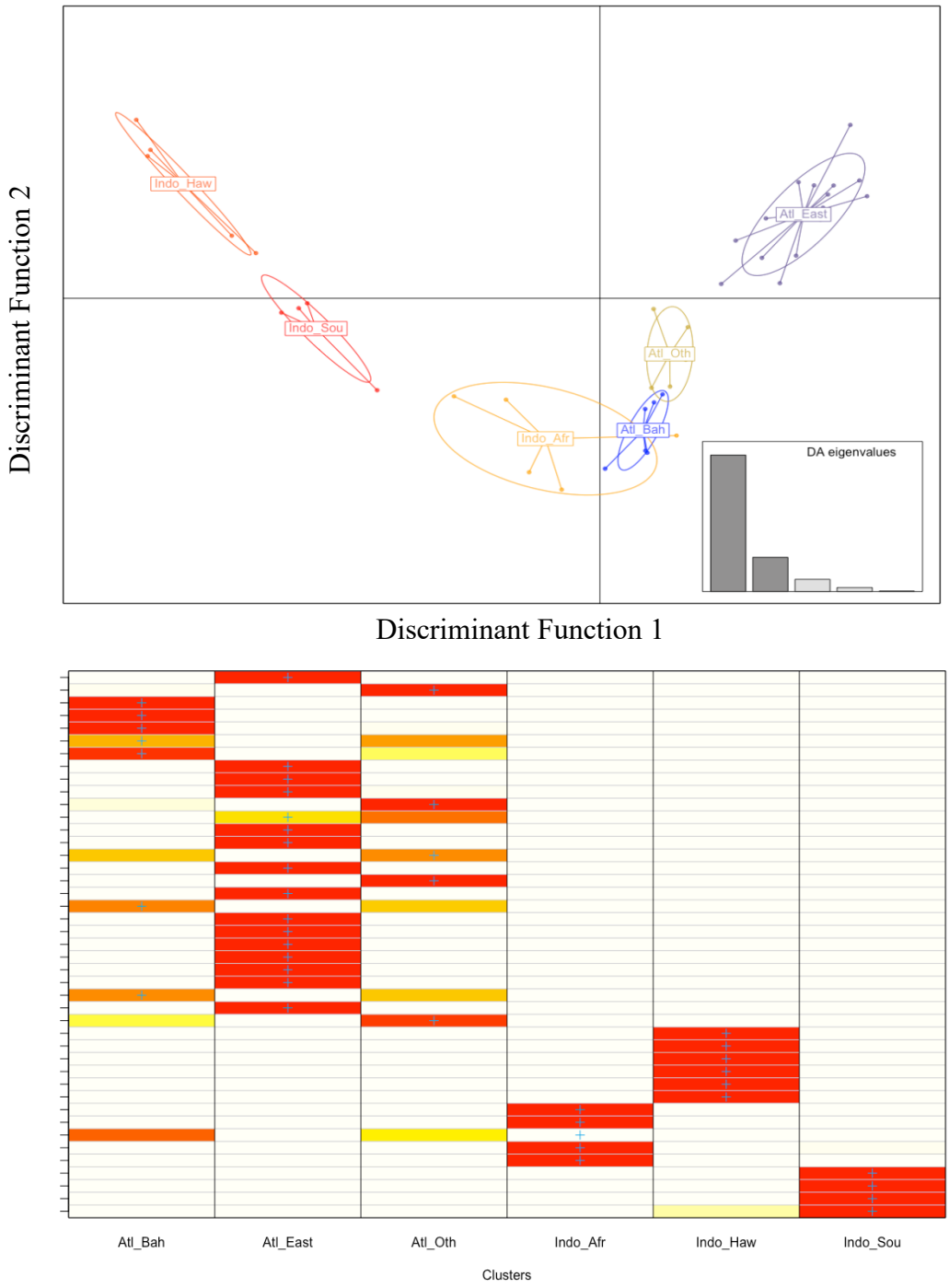


Figure S8.2. DAPC scatter (top) and assignment (bottom) plot for *n*=43 Blainville’s beaked whales (*Mesoplodon densirostris*) generated using cross-validation and retaining 10 PCs. In the assignment plot, each row represents an individual, the blue cross indicates the prior cluster assignment, and the colours represent membership probability (red=1, white=0).

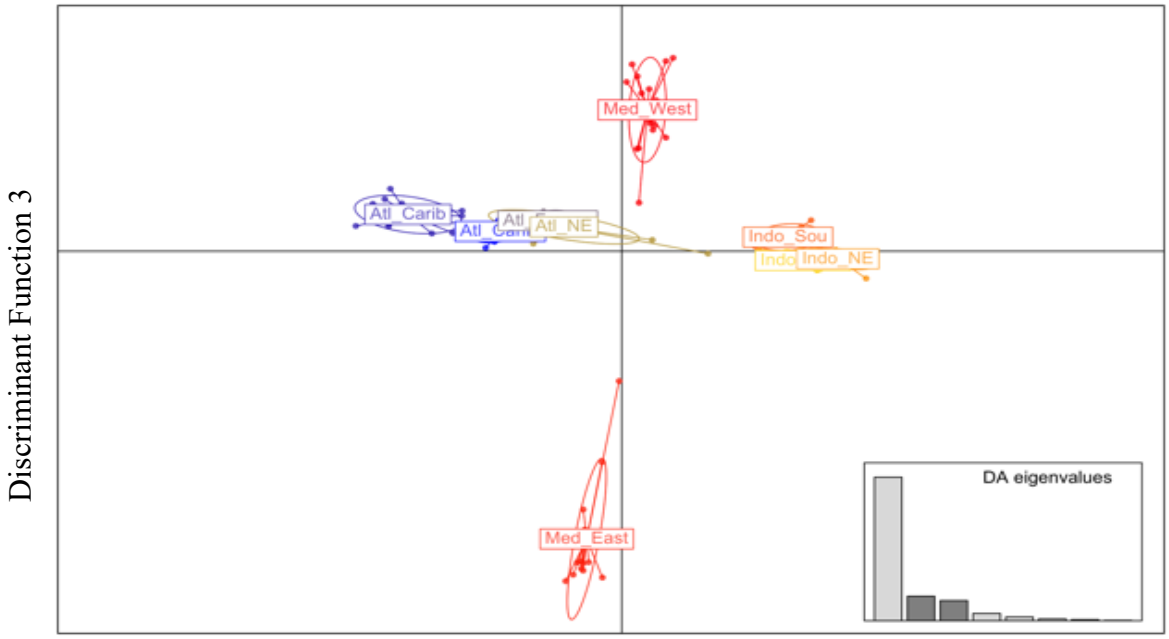


Figure S8.3. DAPC scatter plot (2nd and 3rd axes) for *n*=118 Cuvier's beaked whales (*Ziphius cavirostris*) generated using cross-validation and retaining 20 PCs.

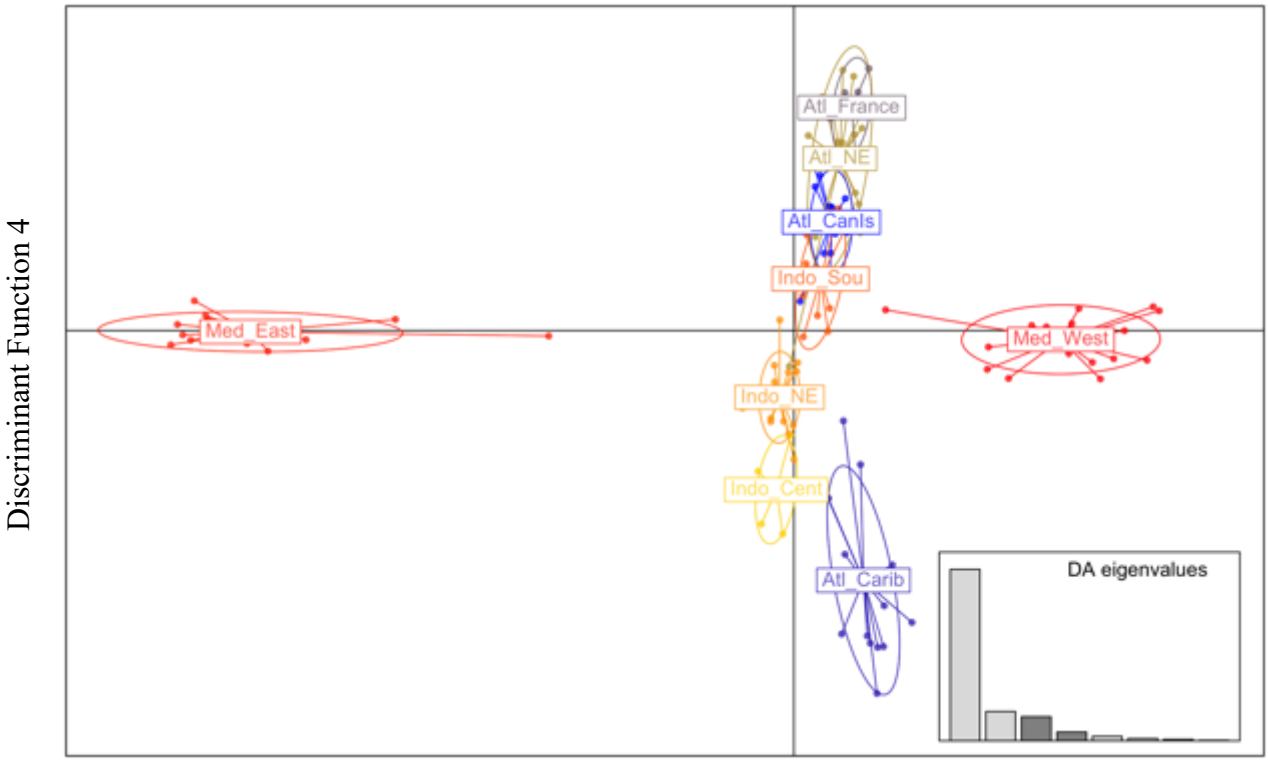


Figure S8.4. DAPC scatter plot (3rd and 4th axes) for *n*=118 Cuvier's beaked whales (*Ziphius cavirostris*) generated using cross-validation and retaining 20 PCs.

C.9 ESM9: DDRAD SEQUENCING RESULTS

From the ITABW collection, 225 individuals were selected for ddRAD analysis (Cuvier's n=170, Blainville's n=55), balancing DNA quality and quantity, and covering as much of each species' broad geographical ranges as possible (See ESM1 for sample selection process). The samples were run across five HiSeq 2500 lanes, generating a total of 340 million and 1.113 billion PE reads across libraries of Blainville's and Cuvier's samples, respectively. Following demultiplexing and initial QC, 323 million and 1.064 billion PE reads were retained for Blainville's and Cuvier's samples, respectively. The number of reads and proportion of those retained were consistent across libraries. In the Blainville's libraries (n=7), the mean number of retained reads was 5.77 million per sample (97% retained, standard error of the mean (SE)=1.4%) and in the Cuvier's libraries (n=18), the mean number of retained reads was 6.26 million per sample (94.6% retained, SE=1.0%).

The final ddRAD QC dataset included 123 Cuvier's individuals and 30,479 (72.4% individuals retained) and 43 Blainville's individuals and 13,988 SNPs (76.8% individuals retained) (Figure 1). Each Cuvier's individual was genotyped at an average of 29,697 (SE=90.5) SNPs with a mean per locus read depth of 59x (SE=3.6). Each Blainville's individual was genotyped at an average of 13,760 (SE=63.59) SNPs with a mean per locus read depth of 53x (SE=4.4). Overall, both datasets had low levels of missing data (2.6% in Cuvier's and 1.6% in Blainville's). Only SNPs with a genotype quality greater than 20 were kept (99% base call accuracy).

C.10 SST 1: FULL SAMPLE LIST”

List of all samples included in both the ddRAD and mitogenome analyses from Chapter 3. Sample names are in the column headed “ITABW-ID”, indicating the identifier used in the newly established ITAWB database. The following are also provided per sample when available: species, sampling date, sampling location, sampling region, sampling ocean, whether or not they were included in the ddRAD or mitogenome analysis, the population defined using either method, and a detailed description of the sample’s origin (who contributed the sample, the origin of the sample, sample type, any permits or considerations for collection of the sample and any impact minimisation or assessment steps taken during sample collection).

This dataset sample is stored in the University of St Andrews PURE research database as “Onoufriou_DNAintheDeep_AppendixC

C.11 SST2. BEAST PARTITIONS

List of partitions inferred with PARTITIONFINDER and analysed in a Bayesian phylogenetic framework implemented in BEAST.

Species	Subset	Best Model	# Sites	Partition names	BEAUTI input
Cuvier's	1	HKY+I+X	4322	16s_rRNA, ATP8_2ndpos, ATP6_1stpos, ND2_1stpos, ATP8_1stpos, 12s_rRNA, ND4_1stpos, ND5_1stpos	HKY
	2	TRNEF	433	ND3_1stpos, ND1_1stpos	TRN
	3	HKY+I+G+X	3552	COX1_2ndpos, ND1_2ndpos, ATP6_2ndpos, COX3_2ndpos, COX2_2ndpos, ND5_2ndpos, ND4L_2ndpos, ND2_2ndpos, ND3_2ndpos, ND4_2ndpos, CYTB_2ndpos	HKY
	4	TRN+G+X	2485	ND1_3rdpos, ND2_3rdpos, ND3_3rdpos, CYTB_3rdpos, COX3_3rdpos, ND4_3rdpos, ND5_3rdpos	TRN
	5	K80+I	1481	COX3_1stpos, COX1_1stpos, CYTB_1stpos, COX2_1stpos, ND4L_1stpos	HKY
	6	HKY+I+X	1133	ATP6_3rdpos, ATP8_3rdpos, COX1_3rdpos, COX2_3rdpos, ND4L_3rdpos	HKY
	7	HKY+I+X	350	ND6_2ndpos, ND6_1stpos	HKY
	8	HKY+G+X	175	ND6_3rdpos	HKY
Blainville's	1	TRN+I	2933	CYTB_1stpos, 16s_rRNA, 12s_rRNA	TRN
	2	TRNEF+G	1437	COX3_1stpos, COX2_1stpos, ND1_1stpos, ND3_1stpos, COX1_1stpos	TRN
	3	HKY+I+X	2235	ATP6_2ndpos, ND4L_2ndpos, ND2_2ndpos, ND3_2ndpos, ND1_2ndpos, ND5_2ndpos, ND4_3rdpos, ATP8_2ndpos	HKY
	4	TRN+I+X	3554	COX1_3rdpos, ND4L_3rdpos, COX3_3rdpos, ND5_3rdpos, ND1_3rdpos, ND4_1stpos, COX2_3rdpos, ATP6_3rdpos, CYTB_3rdpos, ND3_3rdpos, ND2_3rdpos	TRN
	5	TRN+G+X	1867	ND4L_1stpos, ATP8_1stpos, ATP8_3rdpos, ND2_1stpos, ND4_2ndpos, ATP6_1stpos, ND5_1stpos	TRN
	6	HKY+X	1558	ND6_2ndpos, COX1_2ndpos, COX2_2ndpos, CYTB_2ndpos, COX3_2ndpos	HKY
	7	HKY+I+X	175	ND6_1stpos	HKY
	8	HKY+I+X	175	ND6_3rdpos	HKY

C.12 SST3. NCBI MTDNA LIST

List of the already published mitogenome sequences used in the mitogenome analysis by NCBI accession number, sample ID, species, and sampling locality.

Species	Sample ID	Other ID	Genbank ID	Geographic Location	Ocean
Blainville's	z0004010	NOAA_4010 (MdeSW4010, ZZZ0031)	KF032860.2	Santa Barbara, CA,USA	Pacific
	z0033736	NOAA_33736 (RWB070503.W02)	KF032862.2	HI, USA	Pacific
	z0033737	NOAA_33737 (RWB070503.W03)	KF032863.2	HI, USA	Pacific
	z0050723	NOAA_50723 (LL1904-130208)	KF032864.2	North HI, USA	Pacific
	z0074424	NOAA_74424 (080611_Md04)	KF032867.2	Bahamas	Atlantic
	z0074425	NOAA_74425 (080611_Md05)	KF032868.2	Bahamas	Atlantic
	z0074263	NOAA_74263 (050126_Md1)	KF032869.2	Bahamas	Atlantic
	z0074264	NOAA_74264 (050128_Md1)	KF032870.2	Bahamas	Atlantic
	z0008681	NOAA_8681	KF032871.2	FL, USA	Atlantic
	z0079824	NOAA_79824 (080613_Md1c)	KF032872.2	Bahamas	Atlantic
	z0079835	NOAA_79835 (MdenBahamas2, 090508_Md1)	KF032873.1	Bahamas	Atlantic
	z0079838	NOAA_79838 (090530_Md1)	KF032875.2	Bahamas	Atlantic
	z0079839	NOAA_79839 (090531_Md1)	KF032876.2	Bahamas	Atlantic
	z0079840	NOAA_79840 (090602_Md1)	KF032877.2	Bahamas	Atlantic
	z0094563	NOAA_94563 (100613_Md2ac)	KF032878.2	Bahamas	Atlantic

Species	Sample ID	Other ID	Genbank ID	Geographic Location	Ocean
Cuvier's	Mden1	NOAA_79837 (090526_Md1)	KF032874.2	Bahamas	Atlantic
	z0004472	NOAA_4472 (duplicate ZcaSW4472 (NE9521), S-95-ZC-21)	KC776696.1	Florida, USA	Atlantic
	z0014950	NOAA_14950 (duplicate ZcaNEPST382)	KC776697.1	Puerto Rico	Atlantic
	z0074262	NOAA_74262 (040403_Zc1, Zca01BMMS)	KC776698.1	Bahamas	Atlantic
	z0079887	NOAA_79887 (080610_Zc2c)	KC776699.1	Bahamas	Atlantic
	z0079893	NOAA_79893 (080611_Zc3c)	KC776700.1	Bahamas	Atlantic
	z0079898	NOAA_79898 (090506_Zc1)	KC776701.1	Bahamas	Atlantic
	z0079900	NOAA_79900 (090528_Zc1)	KC776702.1	Bahamas	Atlantic
	z0094591	NOAA_94591 (100614_Zc2ac)	KC776703.1	Bahamas	Atlantic
	z0094595	NOAA_94595 (100616_Zc1ac)	KC776704.1	Bahamas	Atlantic
	z0007445	NOAA_7445 (duplicate ZcaSW7445, S-96-ZC-24)	KC776705.1	Florida, USA	Atlantic
	z0004967	NOAA_4967 (duplicate LACM91908 (ZcaSW4967))	KC776706.1	Hawaii, USA	Indo-Pacific
	z0024816	ZcaSW24816	KC776707.1	Alaska, USA	Indo-Pacific
	z0026279	NOAA_26279 (duplicate ZcaSW26279, DSJ011024.04)	KC776708.1	West Coast, USA	Indo-Pacific
	z0030065	NOAA_30065 (duplicate ZcaSW30065, RWB270902.02)	KC776709.1	Hawaii, USA	Indo-Pacific
	z0061950	NOAA_61950 (CRC20060817-02)	KC776710.1	California, USA	Indo-Pacific
	z0068606	NOAA_68606 (Z. cavirostris 02/ZGCMay06-ZC)	KC776711.1	Baja California, Mexico	Indo-Pacific
	z0072279	NOAA_72279 (DBW-3)	KC776712.1	ETP-Mexico?	Indo-Pacific
	z0079632	NOAA_79632 (ZCAV090419)	KC776713.1	Alaska, USA	Indo-Pacific
	z0087481	NOAA_87481 (ZCAV090823.01)	KC776714.1	California, USA	Indo-Pacific
	z0005565	NOAA_5565 (duplicate ZcaSW5565, WFP0832)	KC776715.1	Philippines	Indo-Pacific
	z0009122	NOAA_9122 (ZCA02)	KC776716.1	New Zealand	Indo-Pacific
	z0009561	NOAA_9561 (duplicate ZcaSW9561, TT9601)	KC776717.1	Taiwan	Indo-Pacific

C.13 SST4. MTDNA SEQUENCING STATS

Summary of shotgun sequencing statistics for the samples used in the mitogenome analyses. For each species, Copenhagen sample ID, corresponding ITABW ID, sampling locality, number of total reads, reads mapped to the reference mitogenome and coverage are given. Summaries of each read statistic are given overall for each species.

Species	ExtID	ITABW ID	Locality	Reads	ReadsMap	Coverage
Cuvier's	Zca144	Zca_311204_TF	Canary Islands	28868279	28552	248
	Zca145	U15-108	New Zealand	6506257	11201	93
	Zca149	NOAA_133994 (110613_Zc2)	Bahamas	8129669	20542	181
	Zca151	ULLEHZc04 (EH15Zc04, ull_EH_Zc04)	El Hierro, Spain	5952949	17068	149
	Zca16	ZcaRNP2094	Argentina	12096270	17126	150
	Zca166	NMS.Z.2006.11 (M327/05,SW2005/301)	Scotland	4896862	5907	41
	Zca170	Zca.IRL.21.CBW	Ireland	6143384	15753	136
	Zca172	IRLX2	Ireland	35702987	32614	296
	Zca98	U15-109	New Zealand	13655271	10811	118
	Zca99	SECACMCC0106 (Zca_140705_FV, Zca_0106)	Canary Islands	10336778	8882	85
	ZcaCN1	D1 (CN1)	Liguria, Italy	1270940	6627	42
	ZcaD4	D4	Liguria, Italy	8841974	23432	206
	ZcaD5	D5	Liguria, Italy	12565710	31823	287
			Min	1270940	5907	41
			Max	35702987	32614	296
			Average	11920564	17718	156
			SD	9762564	9193	84
			SE	2707648	2550	23

Species	ExtID	ITABW ID	Locality	Reads	ReadsMap	Coverage
Blainville's	Mden1	NOAA_79837 (090526_Md1)	Bahamas	77687598	24727	203
	Mden12	NOAA_106825 (Hubbs-1105-Md)	Florida, USA	5964668	18582	165
	Mden13	NOAA_33738 (RWB070503.W04)	Hawai'i, USA	7029725	22343	195
	Mden15	MdeSAM.ZM.84/11 (PBB1984/011_ATS0581, ZM-040045)	South Africa	5834952	12098	105
	Mden18	NOAA_132613 (RWB2011MAY09.02)	Hawai'i, USA	9437039	14994	121
	Mden3	MdeZL1	Canada	6185745	14569	128
	Mden4	ULLEHMd03 (EH15MD03, MdH2)	Canary Island	12810881	26531	212
	Mden6	MdeCH9901	Chile	9997180	22363	181
	Mden7	MdenCRMM102-05-207	France	11167589	9963	81
	Mden8	Mde02FP04	French Polynesia	12145640	22986	197
	Mden9	Mbow01/MdeNZ04	New Zealand	19709108	32417	289
	ZP013	CNX1	Car Nicobar	10331217	3273	14
			Min	5834952	3273	14
			Max	77687598	32417	289
			Average	15691779	18737	158
			SD	19905273	8088	72
			SE	5746157	2335	21

C.14 SST5. MTDNA DIV-DIFF

Diversity and differentiation statistics for the mtDNA data including whole mitogenomes and extracted control regions (CR) with total length. Ocean-basin-level statistics are provided including sample size (N), segregating sites (S), haplotypes (h), haplotype diversity (Hd), nucleotide diversity (Nu) and fixed differences (FixDiff). Total and pair-wise MtDNA-based measurements of F_{ST} (Hudson et al. 1992)and d_A (Equation 10.21; Nei 1987) are also provided. All F_{ST} estimates were significant at P<0.01 with 1000 replicates.

Species/data	Region	N	S	h	Hd	π	FixDiff	F_{ST}	Da
Cuvier's mitogenome 15219 bp	Atlantic	18	462	17	0.993	0.0097			
	IndoPacificSouthern	14	344	13	0.989	0.0079			
	Mediterranean	3	47	3	1.000	0.0021			
	Total	35	645	32	0.993	0.0101	na	0.41012	na
Cuvier's CR 860 bp	Atlantic	18	14	9	0.882	0.0048			
	Indo-Pacific	14	9	6	0.747	0.0041			
	Mediterranean	3	5	3	1.000	0.0039			
	Total	35	22	17	0.914	0.0055	na	0.39163	na
Blainville's mitogenome 14147 bp	Atlantic	16	164	13	0.975	0.0028			
	Indo-Pacific	11	312	10	0.982	0.0050			
	Total	27	444	23	0.989	0.0086	1	0.7112	0.0097
Blainville's CR 852 bp	Atlantic	16	9	7	0.800	0.0030			
	Indo-Pacific	11	11	9	0.964	0.0040			
	Total	27	17	16	0.926	0.0042	0	0.3276	0.0017

Within-ocean basin estimates of Cuvier's F_{ST} (above diagonal) and d_A (below diagonal) calculated using the whole mitogenome.

Cuvier mitogenomes	Atlantic	Mediterranean	IndoPacificSouthern
Atlantic		0.37534 (p<0.01)	0.17756 (p<0.01)
Mediterranean	0.00354		0.62313 (p<0.01)
IndoPacificSouthern	0.00191	0.00826	

Within-ocean basin estimates of Cuvier's F_{ST} (above diagonal) and d_A (below diagonal) calculated using the control region (CR).

Cuvier CR	Atlantic	Mediterranean	IndoPacificSouthern
Atlantic		0.34683 (p<0.01)	0.24109 (p<0.01)
Mediterranean	0.0023		0.53064 (p<0.01)
IndoPacificSouthern	0.0014	0.00448	

C.15 SST6. DDRAD F_{ST}

Pairwise genetic differentiation of Cuvier's beaked whale (Ziphius cavirostris) populations defined using *tess3r* and based on 30479 SNPs. Above the diagonal are the F_{ST} 95% confidence intervals. Below the diagonal are the F_{ST} point estimates and the p-value in parentheses. All estimates are significant at $p<0.05$, except the bold estimates, which are $p>0.05$.

Ocean Basins	*n*	Atlantic	Indo-Pacific	Mediterranean
Atlantic	54		0.017 - 0.018	0.178 - 0.188
Indo-Pacific	36	0.018 (0.0099)		0.191 - 0.202
Mediterranean	33	0.184 (0.0099)	0.197 (0.0099)	

Populations	*n*	Atl_CanIs	Atl_Carib	Atl_France	Atl_NE	Indo_Cent	Indo_NE	Indo_South	Med_East	Med_West
Atl_CanIs	15		0.011 - 0.013	0.007 - 0.011	0.006 - 0.007	0.027 - 0.033	0.028 - 0.031	0.015 - 0.018	0.224 - 0.234	0.191 - 0.200
Atl_Carib	17	0.012 (0.0099)		0.011 - 0.016	0.009 - 0.012	0.030 - 0.035	0.031 - 0.035	0.019 - 0.022	0.231 - 0.242	0.198 - 0.209
Atl_France	5	**0.009 (0.0693)**	0.014 (0.0099)		0.002 - 0.005	0.030 - 0.038	0.027 - 0.032	0.014 - 0.019	0.261 - 0.275	0.215 - 0.226
Atl_NE	15	0.006 (0.0099)	0.011 (0.0099)	0.004 (0.01)		0.025 - 0.030	0.022 - 0.024	0.009 - 0.012	0.224 - 0.235	0.190 - 0.200
Indo_Cent	5	0.030 (0.0101)	0.033 (0.0099)	0.034 (0.011)	0.028 (0.01)		0.014 - 0.018	0.015 - 0.020	0.277 - 0.288	0.229 - 0.239
Indo_NE	19	0.029 (0.0099)	0.033 (0.0099)	0.029 (0.01)	0.023 (0.0099)	0.016 (0.0099)		0.010 - 0.013	0.230 - 0.241	0.199 - 0.210
Indo_South	9	0.017 (0.0099)	0.021 (0.0099)	0.017 (0.01)	0.011 (0.0099)	0.018 (0.01)	0.012 (0.0099)		0.242 - 0.253	0.203 - 0.214
Med_East	14	0.229 (0.0099)	0.237 (0.0099)	0.268 (0.012	0.229 (0.0099)	0.283 (0.012)	0.236 (0.0099)	0.248 (0.0099)		0.088 - 0.096
Med_West	19	0.195 (0.0099)	0.203 (0.0099)1	0.220 (0.012)	0.194 (0.0099)	0.234 (0.013)	0.204 (0.0099)	0.208 (0.0099)	0.092 (0.0099)	

Pairwise genetic differentiation of Blainville's beaked whale (Mesoplodon densirostris) populations defined using *tess3r* and based on 13988 SNPs. Above the diagonal are the F_{ST} 95% confidence intervals. Below the diagonal are the F_{ST} point estimates and the p-values in parentheses. All estimates are significant at $p<0.05$, except the bold estimates, which are $p>0.05$.

Ocean Basins				
Atlantic	Indo-Pacific	F_{ST}	95% CI	p-value
n=28	*n*=14	0.119	0.113 - 0.124	0.001

Populations	***n***	**Atl_Bah**	**Atl_East**	**Atl_Oth**	**Indo_Afr**	**Indo_Haw**	**Indo_Sou**
Atl_Bah	7		0.0370-0.44	0.010-0.019	0.094-0.106	0.153-0.170	0.153-0.170
Atl_East	16	0.04 (0.001)		0.012-0.019	0.100-0.110	0.154-0.168	0.153-0.170
Atl_Oth	5	**0.015 (0.256)**	0.016 (0.007)		0.065-0.075	0.125-0.139	0.125-0.140
Indo_Afr	5	0.1 (0.005)	0.105 (0.001)	0.07 (0.047)		0.023-0.031	0.004-0.014
Indo_Haw	6	0.161 (0.002)	0.162 (0.001)	0.133 (0.003)	0.027 (0.001)		0.008-0.018
Indo_Sou	3	0.162 (0.017)	0.163 (0.002)	0.134 (0.021)	**0.009 (0.149)**	**0.013 (0.078)**	

C.16 SST7. HAPLOTYPES

List of samples and their mitogenome haplotypes.

Species	Sequence_ID	SWFSC_ID	Ocean Basin	Haplotype
Cuvier's	KC776696	z0004472	Atlantic	1
Cuvier's	KC776698	z0074262	Atlantic	2
Cuvier's	KC776701	z0079898	Atlantic	2
Cuvier's	Zca_98_NC_021435_50bpPad		Indo-Pacific	3
Cuvier's	Zca145_NC_021435_50bpPad		Indo-Pacific	3
Cuvier's	Zca16_NC_021435_50bpPad		Atlantic	3
Cuvier's	KC776705.1	z0007445	Atlantic	4
Cuvier's	KC776706	z0004967	Indo-Pacific	5
Cuvier's	KC776717	z0009561	Indo-Pacific	6
Cuvier's	KC776709.1	z0030065	Indo-Pacific	7
Cuvier's	KC776716.1	z0009122	Indo-Pacific	8
Cuvier's	KC776697	z0014950	Atlantic	9
Cuvier's	KC776699.1	z0079887	Atlantic	10
Cuvier's	KC776704.1	z0094595	Atlantic	11
Cuvier's	KC776702.1	z0079900	Atlantic	12
Cuvier's	KC776707	z0024816	Indo-Pacific	13
Cuvier's	KC776711.1	z0068606	Indo-Pacific	14
Cuvier's	KC776712	z0072279	Indo-Pacific	15
Cuvier's	KC776713.1	z0079632	Indo-Pacific	16
Cuvier's	KC776715	z0005565	Indo-Pacific	17
Cuvier's	KC776708	z0026279	Indo-Pacific	18
Cuvier's	KC776710	z0061950	Indo-Pacific	19
Cuvier's	KC776714.1	z0087481	Indo-Pacific	20
Cuvier's	KC776700.1	z0079893	Atlantic	21
Cuvier's	KC776703.1	z0094591	Atlantic	22
Cuvier's	Zca166_NC_021435_50bpPad		Atlantic	23
Cuvier's	Zca170_NC_021435_50bpPad		Atlantic	24
Cuvier's	Zca172_NC_021435_50bpPad		Atlantic	25
Cuvier's	Zca_99_NC_021435_50bpPad		Atlantic	26
Cuvier's	Zca144_NC_021435_50bpPad		Atlantic	27
Cuvier's	Zca149_NC_021435_50bpPad		Atlantic	28
Cuvier's	Zca151_NC_021435_50bpPad		Atlantic	29
Cuvier's	ZcaCN1_NC_021435_50bpPad		Mediterranean	30
Cuvier's	ZcaD4_NC_021435_50bpPad		Mediterranean	31
Cuvier's	ZcaD5_NC_021435_50bpPad		Mediterranean	32
Blainville's	KF032869	z0074263	Atlantic	1
Blainville's	KF032870	z0074264	Atlantic	1

Blainville's	Mden1_NC_021974_Atlantic		Atlantic	2
Blainville's	KF032871_Atlantic	z0008681	Atlantic	3
Blainville's	Mden12_NC_021974_Atlantic		Atlantic	4
Blainville's	Mden4_NC_021974_Atlantic		Atlantic	5
Blainville's	KF032876_Atlantic	z0079839	Atlantic	6
Blainville's	KF032877	z0079840	Atlantic	6
Blainville's	KF032878_Atlantic	z0094563	Atlantic	7
Blainville's	KF032867_Atlantic	z0074424	Atlantic	8
Blainville's	KF032868_Atlantic	z0074425	Atlantic	8
Blainville's	Mden7_NC_021974_Atlantic		Atlantic	9
Blainville's	Mden3_NC_021974_Atlantic		Atlantic	10
Blainville's	KF032872_Atlantic	z0079824	Atlantic	11
Blainville's	KF032875_Atlantic	z0079838	Atlantic	12
Blainville's	KF032873_Atlantic	z0079835	Atlantic	13
Blainville's	ZP013_NC_021974_Indian		Indo-Pacific	14
Blainville's	KF032862_Pacific	z0033736	Indo-Pacific	15
Blainville's	Mden18_NC_021974_Pacific		Indo-Pacific	15
Blainville's	KF032863_Pacific	z0033737	Indo-Pacific	16
Blainville's	Mden9_NC_021974_Pacific		Indo-Pacific	17
Blainville's	KF032864_Pacific	z0050723	Indo-Pacific	18
Blainville's	KF032860_Pacific	z0004010	Indo-Pacific	19
Blainville's	Mden8_NC_021974_Pacific		Indo-Pacific	20
Blainville's	Mden15_NC_021974_South_Africa		Indo-Pacific	21
Blainville's	Mden6_NC_021974_Pacific		Indo-Pacific	22
Blainville's	Mden13_NC_021974_Pacific		Indo-Pacific	23

Appendix D: Summary of genomic resources from this thesis

The resources that I have developed in this PhD will be useful for many future studies that pertain to beaked whale genomics and population genomics analyses in general. A summary of the major resources that my thesis has contributed to this growing field is presented here.

D.1 INTERNATIONAL TISSUE ARCHIVE FOR BEAKED WHALES (ITABW)

This PhD was only possible thanks to the contribution of beaked whale tissue and DNA samples from all over the world. To ensure the security of this valuable dataset for future studies, the International Tissue Archive for Beaked Whales (2.3.3) was established between the Universities of Auckland, Copenhagen, and La Laguna.

Species	Atlantic	Indo-Pacific	Mediterranean	Total
Northern bottlenose whale	9			9
Sowerby's beaked whale	59			59
Blainville's beaked whale	55	34		89
Gervais' beaked whale	4			4
True's/Ramari's beaked whale	9	11		20
Cuvier's beaked whale	204	79	55	338
Unknown beaked whale	12			12
Total	352	124	55	531

D.2 DDRAD SEQUENCES

In this thesis I have generated the largest RRS dataset for beaked whales using ddRAD sequencing (2.6). The DNA sequences can be used in future analyses where they are aligned and filtered using different tools or parameters, allowing researchers to get finer and coarser levels of population structure, or datasets that are usable for answering different questions.

Species	Atlantic	Indo-Pacific	Mediterranean	Total
Sowerby's beaked whale	40			40
Blainville's beaked whale	58	21		79
Cuvier's beaked whale	134	45	55	234
Total	232	66	55	353

D.3 MITOGENOMES

Though not generated by me, the mitogenomes used in the Chapter 3 analysis were generated by colleagues at the University of Copenhagen for this project (2.5.2). These are a valuable resource for future beaked whale population studies.

Species	Atlantic	Indo-Pacific	Mediterranean	Total
Blainville's beaked whale	4	7		11
Cuvier's beaked whale	8	2	3	13
Total	12	9	3	24

D.4 R-SCRIPTS

One of the aims of my PhD was to conduct as many of the analyses as possible using the R programming language. SNP datasets are often very large (hundreds of megabytes) and transferring them between clusters and local devices, and between analysis programs can be very time consuming. Various analysis programs also require a time-consuming period of learning how the program is executed and formatting the data. There are many benefits to conducting population genomic analyses in R, including the fact that the large VCF files will only need to be imported once into the working environment where they can then be converted into many different formats for the various packages. Most population genomic analyses in R can either be done on a local device or through a cluster, both usually allowing parallel processing to speed up the analysis. Plotting functionality is also very advanced, fast, and user-friendly in R. Finally, using R as much as possible allows the analysis to be easily shared and reproducible. I have therefore documented many of the population genomic analyses that I conducted in R in a publicly available GitHub repository:
https://github.com/aono87/ddRAD-Diversity-Stats-in-R/blob/master/R_codes.